

THESE de DOCTORAT de l'UNIVERSITE PARIS 6
Ecole doctorale SMAE
Laboratoire d'Acoustique Musicale (UPMC, CNRS, Ministère de la Culture)

Spécialité :
Acoustique-Phonétique

Sujet de la thèse :

**Communiquer en environnement bruyant :
de l'adaptation jusqu'au forçage vocal**

Présentée par **Maëva Garnier**

Pour obtenir le grade de DOCTEUR de l'UNIVERSITE PARIS 6

Soutenue le 18 Juin 2007 devant un jury composé de :

Madame	Danièle DUBOIS	(IJLRDA-LAM/LCPE)	Directrice de thèse
Monsieur	Antoine GIOVANNI	(LAPC)	Examinateur
Madame	Nathalie HENRICH	(GIPSA-Lab)	Directrice de thèse
Monsieur	Jean-Dominique POLACK	(IJLRDA-LAM)	Directeur de thèse
Madame	Annie RIALLAND	(LPP)	Examinatrice
Monsieur	Jean-Luc SCHWARTZ	(GIPSA-Lab)	Rapporteur
Monsieur	Sten TERNSTRÖM	(KTH)	Rapporteur
Monsieur	Jean-Luc ZARADER	(LISIF)	Président du jury

RESUME

Ce travail part de l'observation que certaines personnes sont plus susceptibles que d'autres de développer des troubles de la voix lorsqu'elles sont amenées à communiquer fréquemment dans des conditions perturbées telles que des environnements bruyants. Le but de cette thèse est de comprendre l'origine de ces différences afin de prévenir le forçage vocal.

Jusqu'à maintenant, l'explication de ces différentes sensibilités a été principalement cherchée du côté de la variabilité des constitutions physiologiques et des facteurs environnementaux auxquels sont confrontées ces personnes. Au cours de ce travail, nous allons plutôt explorer l'hypothèse que ces différences puissent également provenir du comportement d'adaptation du locuteur à la situation de communication.

C'est pourquoi nous avons tout d'abord examiné en quoi les différentes modifications acoustiques et articulatoires peuvent être interprétées en terme de stratégies de communication visant à émerger du bruit ambiant, à faciliter la reconnaissance audiovisuelle des unités phonétiques pour l'interlocuteur, à renforcer les indices prosodiques de structuration de l'énoncé ou encore les mots de l'énoncé portant le plus d'information. Nous avons apporté plusieurs arguments en faveur d'une adaptation de la parole dans le bruit qui ne serait pas uniquement vocale, réflexe et subie, mais également en partie intentionnelle, communicationnelle et gérable par le locuteur. Pour cela, nous avons enregistré deux principales bases de données constituées de parole semi-spontanée en interaction avec un interlocuteur et pour une immersion des locuteurs dans le bruit par le biais de haut-parleurs. Nous avons au préalable montré l'apport de ces méthodologies pour pouvoir tester nos hypothèses sur le forçage vocal et sur l'adaptation de la parole dans le bruit.

L'examen individuel des comportements d'adaptation de la parole dans le bruit nous a permis de mettre en évidence des différences au niveau de l'adoption combinée des différentes stratégies de communication –celles-ci n'étant pas équivalentes vis à vis de la charge qu'elles font porter au larynx– ainsi qu'au niveau de la réorganisation de l'adaptation en fonction du contexte (type et niveau de bruit). Nous avons pu relier ces variabilités comportementales à des profils laryngés différents, ouvrant en cela de nouvelles perspectives pour l'étude et la prévention du forçage vocal.

Mots clés : Forçage vocal, Effet Lombard, Production de la parole, Bruit, Stratégies d'adaptation.

ABSTRACT

This work originates from the observation that some people are more likely than others to develop voice disorders when they often have to communicate in adverse conditions such as noisy environments. This thesis aims at understanding what induces these differences, in order to prevent vocal straining.

To account for these differences in sensibility, previous studies have essentially focused on the variability in physiological constitutions as well as on the environmental factors people subject to vocal straining are confronted with. The work presented here rather explores the hypothesis that these differences could also result from an adaptation behaviour of the speaker to the particular communicative situation.

Firstly, we examined how the different acoustic and articulatory speech modifications can be interpreted as indicators of communicative strategies adopted by the speaker to emerge from the ambient noise, to facilitate the audiovisual perception of phonetic units, to enhance the prosodic cues of discourse structure or to highlight words which are more informative. Several arguments are put forward for considering speech adaptation in noise not only as vocal, inflicted and reflex but also as cognitive, communicational and monitored by the speaker. In order to assess this point, two main databases of semi-spontaneous speech were recorded for which speakers were playing interactive games in noise played over loudspeakers. These methodologies were preliminarily assessed considering our hypotheses on vocal straining and speech adaptation in noisy environments.

The inter-individual comparison of adaptation behaviours lead us to point out some differences in the combined use of these different communicative strategies –not equivalent in terms of load for the larynx– and in the reorganization of adaptation strategies for varying noise types and levels. Finally these varying adaptation behaviours were related to different laryngeal profiles. Therefore these results offer a new perspective for the study and prevention of vocal straining.

Keywords : Vocal straining, Lombard effect, Speech production, Noise, Adaptation strategies.

- Remerciements -

Dans le but de mener des recherches sur le forçage vocal et de me former, à travers cette thèse, au métier de chercheur, je n'aurais difficilement pu rêver d'un meilleur environnement pluridisciplinaire que celui qui m'a été offert par le co-encadrement de Nathalie Henrich, acousticienne de la voix et de Danièle Dubois, psycholinguiste, et par le foisonnement d'idées du Laboratoire d'Acoustique Musical. Je suis également très reconnaissante à mes encadrants de m'avoir proposé ce sujet de thèse pluridisciplinaire pour lequel je me suis passionnée durant ces années, et qui m'ouvre aujourd'hui de nombreuses perspectives.

Je souhaite donc avant tout remercier Nathalie Henrich, sans qui cette belle aventure n'aurait pu avoir lieu. Je suis très heureuse qu'elle ait accepté de m'encadrer comme sa « première thésarde ». Je mesure la chance d'avoir eu une directrice de thèse aussi jeune, brillante et passionnée, ayant montré une grande implication dans cet encadrement, tant par son enthousiasme que par sa disponibilité. Elle représente pour moi un modèle de rigueur scientifique, de dynamisme et de gestion des relations humaines, qui continuera de me guider après cette thèse. Je voudrais en particulier lui témoigner ma grande reconnaissance pour la confiance qu'elle m'a accordée en me laissant développer très librement mes propres idées et m'orienter vers d'autres champs de recherche et disciplines que les siens. Avec franchise, elle m'a toujours exprimé son désaccord sur certains points mais elle a en même temps toujours témoigné d'une grande tolérance en respectant mes arguments et en cherchant à les comprendre. Au-delà de l'encadrement scientifique, j'ai énormément apprécié les grandes discussions que nous avons souvent eues, et qui m'ont tour à tour émue, fait rire, fait réfléchir. Je la remercie pour son amitié et pour m'avoir fait rencontrer Paul et Anna. Merci à eux également de m'avoir accueillie si souvent lors de mes déplacements sur Grenoble.

Je voudrais également remercier Danièle Dubois qui représente une rencontre très importante dans ma professionnelle. Ses positionnements épistémologiques et sa façon d'appréhender les problématiques humaines ont été une réelle révélation pour l'ingénieure de formation que j'étais, et ont fait résonner en moi beaucoup d'intuitions ou de doutes que je ressentais sans savoir clairement les exprimer. A son contact, je me suis initiée aux théories psychologiques et linguistiques. Mais surtout, elle m'a transmis les fondements d'une vraie démarche scientifique, où la rigueur ne réside pas toujours là où on le penserait (dans le contrôle ultra-paramétré de la situation expérimentale, la mise en place de paradigmes classiques ou la quantification à tout prix des phénomènes), mais avant tout dans la réflexion quant au sens de ce qu'on souhaite observer et de la démarche à adopter pour y parvenir. Je suis consciente du trésor que représente cet enseignement, indépendamment des problématiques que je serai susceptible d'aborder par la suite.

Une autre rencontre très marquante a été celle de Michèle Castellengo qui m'a fait découvrir le monde de la perception et de la voix. Je voudrais non seulement lui témoigner ma reconnaissance pour son soutien constant, depuis l'obtention d'une bourse de thèse MRT jusqu'à mes récentes demandes de financement pour des stages post-doctoraux, mais aussi mon respect pour l'originalité et la richesse de ses idées de recherche, conjuguées à sa grande culture, son art de vivre et son humour que j'ai beaucoup appréciés durant mes années de thèse.

Je tiens à remercier Jean-Dominique Polack en tant que directeur de thèse et directeur du Laboratoire d'Acoustique Musicale de m'avoir permis d'effectuer cette thèse dans un tel environnement scientifique, matériel et humain. Le LAM m'a en effet offert un cadre incomparable de rencontres avec des acousticiens, des psychologues, des linguistes, des phoniâtres, des orthophonistes, des phonéticiens, des chanteurs et des professeurs de chant, propices à l'acquisition d'une bonne vue d'ensemble sur la voix et la parole. Je suis également consciente du privilège d'avoir pu participer à plusieurs conférences internationales à l'étranger et d'avoir disposé de conditions et de matériels de grande qualité pour effectuer mes expériences. Enfin, j'ai énormément apprécié l'ambiance musicale, chaleureuse, voire presque familiale que j'ai trouvée au LAM. Je voudrais à ce propos remercier tous les doctorants et les permanents du laboratoire pour leur accueil, en particulier Guillaume Pellerin, Pierre Leveau, Adrien Mamou-Mani, Séverine Morange, Caroline Cance, Nicolas Montgermont, François Blanc (pour les pauses thé salvatrices en fin de rédaction), Laurent Daudet, Catherine Fourcin et Pascal Lesaëc pour leur attention et pour les bons moments passés en leur compagnie.

Je ne peux manquer de remercier très chaleureusement Marion Dohen, Hélène Loevenbruck et Pauline Welby du GIPSA-lab avec lesquelles j'ai eu le plaisir de mener l'étude articulatoire pilote et

auxquelles je dois énormément de ce que j'ai pu apprendre dans le domaine de la prosodie et de la phonétique articulatoire. Je remercie également beaucoup Lucie Bailly d'avoir contribué, par son excellent stage de DEA, à initier cette collaboration. Je souhaite de tout cœur que celle-ci se poursuive sur les nombreux points restant à explorer ainsi que sur la validation perceptive de nos premiers résultats.

Je veux maintenant manifester ma gratitude à l'égard d'Antoine Giovanni, Annie Rialland, Jean-Luc Zarader, Jean-Luc Schwartz et Sten Ternström pour avoir accepté avec autant d'enthousiasme de faire partie du jury de ma thèse. Je suis à la fois très honorée et intimidée d'avoir l'avis de ces experts de différents champs disciplinaires, que je respecte infiniment pour leur travaux scientifiques. Je remercie tout particulièrement J-L. Schwartz et S. Ternström d'avoir bien voulu lire et rapporter mon travail, car je sais combien leur temps est précieux et je mesure la charge que cela peut représenter en plus lorsque l'on doit lire un manuscrit de plus de 200 pages dans une langue étrangère.

Tout ce travail n'aurait pu être effectué sans l'aimable participation des locuteurs et locutrices ayant accepté de se prêter au(x) jeu(x) et surtout au bruit de ces expériences. Un grand merci donc à Anne B., Eric B., Lucie B., Claire C., Hadrien C., Antoine L., Pierre L., Manuel L., Séverine M., Aude N. et Marion S. Toute ressemblance avec des personnes ayant pu exister est purement fortuite, évidemment ! Je remercie également toutes les personnes ayant pris le temps de répondre à mes deux questionnaires sur les environnements bruyants et sur le forçage vocal.

Je voudrais également témoigner ma reconnaissance à toutes les personnes et des institutions qui m'ont aidée tout au long de ce travail :

- Tout d'abord l'ex Institut de la Communication Parlée, aujourd'hui GIPSA-lab pour m'avoir permis d'effectuer des enregistrements articulatoires avec leur système de suivi labiométrique, ainsi que Christophe Savariaux et Alain Arnal pour leur compétence, leur efficacité et leur gentillesse dans l'aide qu'il m'ont apportée pour la mise en place de ces enregistrements. Je suis particulièrement reconnaissante à Christophe Savariaux du temps qu'il a pris pour m'aider à l'extraction des signaux articulatoires.
- Merci également à Corsin Vogel pour m'avoir aidée à câbler la table de mixage pour l'enregistrement de la 1^{ère} base de données, et pour avoir été le premier à relire le début de mon travail.
- Je voudrais également remercier Jacques Prado et Maurice Charbit de l'ENST, ainsi que Mickaël Bohman du KTH pour m'avoir aidée avec patience dans l'implémentation des algorithmes de débruitage. Merci également à Boris Doval et à Gaël Richard pour m'avoir mise en contact avec les chercheurs de l'ENST.
- Je tiens également à remercier Martine Adda de m'avoir accordé de son temps précieux à essayer d'étiqueter automatiquement ma 1^{ère} base de données.
- Merci encore à Jacques Poitevineau pour avoir aussi gentiment répondu à toutes mes interrogations concernant les statistiques.
- Merci à René Carré, Dominique Rostolland, Annie Rialland, Marie Annick Morel, Michel Pitterman, Bernard Roubeau, Antoine Giovanni, Claire Pillot et Robert Expert pour les échanges scientifiques et bibliographiques que j'ai pu avoir avec eux.
- Je remercie aussi Pascal Perrier d'avoir présenté notre poster à ISSP.
- Enfin merci à Cédric Gendrot et à Gregory Beller pour m'avoir aidée et surtout rassurée en matière d'estimation formantique.

A présent, je voudrais remercier mes amis et mes proches pour leur soutien et leurs encouragements. Merci à Ethel et Franck pour leurs conseils avisés d'anciens. Merci à Audrey, Michèle et Albert, Robert et Lina pour leur soutien. Merci à Loïc, Sandrine et Jean-Alain, Fanette et Thierry, David et Rachel, Cyrille et Myriam, Corsin, Marie-Céline, Marion et Benoît pour leur amitié constante et leur compréhension, en dépit de mes silences ou mes négligences durant cette dernière année de thèse. Merci de ne pas m'avoir mis la pression, de m'avoir fait confiance et de m'avoir réconfortée dans les moments difficiles. Merci à mes parents de m'avoir transmis la curiosité, l'esprit critique, le goût de réfléchir par soi-même et le non-conformisme, ainsi que de m'avoir soutenue pendant toutes ces années d'études en acceptant mes changements de cap sans les juger. Je remercie en particulier ma mère pour avoir corrigé attentivement plusieurs chapitres de ce manuscrit.

Enfin, je voudrais terminer ces remerciements en témoignant ma reconnaissance à Benoît pour m'avoir évité de nombreuses crises de nerfs en sauvant plusieurs disques durs, pour sa relecture « significative » de plusieurs chapitres de ce manuscrit, mais surtout pour sa patience pendant cette dernière année monacale, sa confiance et son soutien moral quotidien.

- Sommaire -

INTRODUCTION

PARTIE I : ACQUIS ET HYPOTHESES SUR LE FORÇAGE VOCAL

Chapitre 1. Le forçage vocal

1. Définitions du forçage vocal.....	8
2. Etat des connaissances sur le forçage vocal	14
3. Positionnement et hypothèses de notre travail.....	25
4. Conclusion du chapitre 1	34

PARTIE II : METHODOLOGIES ET BASES DE DONNEES POUR L'ETUDE DE L'ADAPTATION DE LA PAROLE EN ENVIRONNEMENT BRUYANT

Chapitre 2. Mécanismes et facteurs de l'adaptation de la parole en environnement bruyant

1. Les acquis des travaux antérieurs.....	40
2. Les mécanismes de l'adaptation de la parole dans le bruit.....	45
3. Les méthodologies d'étude de l'adaptation de la parole dans le bruit	51
4. Positionnement de notre étude par rapport à ces différentes théories et méthodologies	53

Chapitre 3. Influence de la méthode d'immersion dans le bruit sur l'adaptation de la parole

1. Les éventuels problèmes liés au port d'un casque.....	58
2. Recherche d'une méthode de débruitage	60
3. Comparaison de trois méthodes d'immersion du locuteur dans le bruit	72

Chapitre 4. Influence du dialogue avec un interlocuteur sur l'adaptation de la parole dans le bruit

1. Recherche d'alternatives à la lecture par un locuteur isolé.....	88
2. Comparaison de deux tâches de parole avec et sans interaction avec un interlocuteur	95
3. Conclusion du chapitre 4	101

PARTIE III : CARACTERISATION DES STRATEGIES D'ADAPTATION DE LA PAROLE EN ENVIRONNEMENT BRUYANT

Chapitre 5. Principales observations

1. Quels sont les aspects de la parole significativement modifiés dans le bruit ?.....	108
2. Quelles sont les tendances partagées et les différences individuelles ?	115
3. Quels sont les paramètres dont l'évolution dans le bruit est dépendante du type de bruit ?....	123
4. Conclusion du chapitre 5	127

Chapitre 6. Une adaptation visant à émerger du bruit ambiant ?

1. Préservation d'un rapport signal sur bruit acceptable	130
2. Expansion ou compression de la dynamique vocale	141
3. Adaptation en fréquences.....	145
4. Conclusion du chapitre 6	154

Chapitre 7. Une adaptation visant à favoriser la reconnaissance des unités phonétiques ?

1. Renforcement du contraste acoustique entre les voyelles.....	156
2. Renforcement des indices visuels d'identification ou de discrimination des unités phonétiques	166
3. Allongement du temps de production ou de récupération de ces indices.....	171
4. Conclusion du chapitre 7	175

Chapitre 8. Une adaptation visant à renforcer des indices de structuration de l'énoncé ?

1. Indices de frontière syllabique	179
2. Indices de frontière entre un déterminant et une unité lexicale.....	181
3. Indices de début de l'énoncé.....	185
4. Indices de fin de l'énoncé.....	186
5. Conclusion du chapitre 8	188

Chapitre 9. Une adaptation visant à mettre en valeur l'information importante ?

1. Matériel et méthode	190
2. Résultats	191
3. Conclusion	194

Chapitre 10. Synthèse quant à l'utilisation des différentes stratégies d'adaptation en fonction des individus et des contextes

1. Premiers bilans	198
2. Réorganisation de l'adaptation en fonction du contexte.....	205
3. Conclusion	214

CONCLUSION GENERALE DE L'ETUDE

REFERENCES BIBLIOGRAPHIQUES

- Abréviations -

Liées aux bases de données

BD1	1 ^{ère} base de données
BD2	2 ^{ème} base de données
L1 → L11	Locuteurs et locutrices 1 à 11
Bb	Bruit blanc
Cktl	Bruit cocktail

Pour certains paramètres examinés

F0	Fréquence fondamentale
F1, F2, F3	1 ^{er} , 2 ^{ème} et 3 ^{ème} formants vocaliques
A	Etirement labial
B	Ouverture labiale
S	Aire intéro-labiale
P1	Protrusion de la lèvre supérieure
AR	Arrondi labial
PI	Pincement des lèvres sur les consonnes bilabiales
VLP	Vertical Larynx Position
Oq	Quotient ouvert
IdB	Intensité vocale en dB
ΔParam	Evolution du silence au bruit du paramètre « Param »
HF	Hautes fréquences
BF	Basses fréquences
RSB	Rapport Signal sur Bruit
SPL	Sound Pressure Level
EGG	Signal électroglottographique (DEGG sa dérivée)
ODG	Onde de débit glottique (ODGD sa dérivée)

En général

E.L Effet Lombard

Nous appellerons « Effet Lombard » l'effet l'adaptation de la parole en environnement bruyant, sans distinguer les mécanismes réflexes ou communicationnels qui interviennent dans cet effet.

Nous appellerons « Parole Lombard » la parole produite en situation d'effet Lombard, autrement dit la parole produite en environnement bruyant, différant de la parole conversationnelle produite en situation calme.

- Conventions -

Notations phonétiques

CV Syllabe de type Consonne-Voyelle

Voyelles orales :

[a]	de « là »
[e]	de « les »
[ɛ]	de « lait »
[i]	de « lit »
[y]	de « lu »
[u]	de « loup »
[o]	de « lot »
[ɔ]	de « lotte »
[ø]	de « leurrer »

Voyelles nasales :

[ã]	de « lent »
[õ]	de « long »

Consonnes occlusives :

[b]	de « bout »
[p]	de « poux »
[g]	de « goût »
[k]	de « cou »
[d]	de « doux »
[t]	de « tout »

Consonnes fricatives

[z]	de « zou »
[s]	de « sous »
[v]	de « vous »
[f]	de « fou »
[ʒ]	de « joue »
[ʃ]	de « choux »

Consonnes liquides

[l]	de « loup »
[ʀ]	de « roue »

Consonnes nasales

[m]	de « mou »
[n]	de « nous »

Notations statistiques

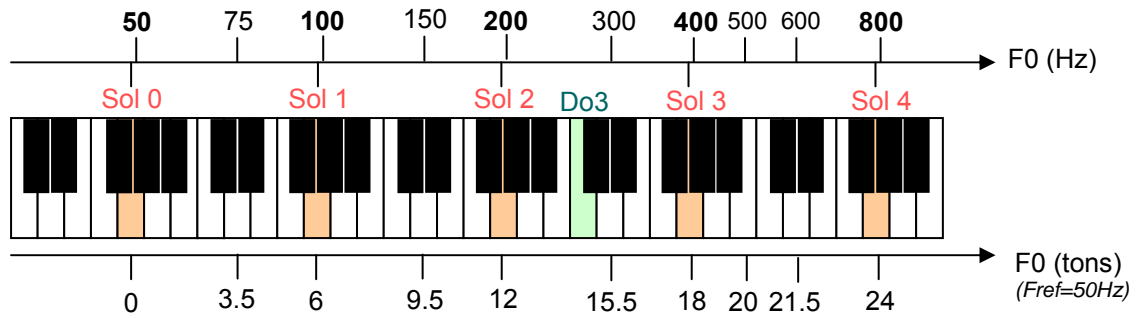
Pour les tests statistiques Anova, les résultats détaillés seront présentés en annexe. L'abréviation suivante sera utilisée dans le manuscrit pour synthétiser les résultats :

ns	non significatif
*	significatif, avec $p < 0.05$
**	significatif, avec $p < 0.01$
***	significatif, avec $p < 0.001$

Unités

La définition des paramètres examinés dans cette thèse ainsi que leur méthode de mesure est présentée en détail en Annexe.

La fréquence fondamentale sera fréquemment représentée en tons, par référence à la fréquence 50Hz. La Figure ci-dessous permet de se repérer dans la conversion tons \leftrightarrow hertz \leftrightarrow hauteur sur un clavier.



- Introduction -

La voix est un outil de travail pour de nombreuses professions. C'est le cas par exemple des enseignants (Rantala et al. 1999 ^[280] ; Sodersten et al. 2002 ^[321]), des annonceurs de télévision ou de radio, des hommes politiques, des avocats, des prêtres, des instructeurs d'aérobic (Long et al. 1998 ^[225]), des instructeurs de l'armée (Mann et al. 1999 ^[232]), des opérateurs de renseignements téléphoniques ou de télémarketing (Lehto et al. 2003 ^[206]) ou des chanteurs et acteurs professionnels (Kitch et al. 1994 ^[181]). Dans les pays occidentaux, plus de 30% des personnes actives se servent de leur voix de façon importante dans le cadre de leur profession (Eluard 2004 ^[84]). Par ailleurs, la voix est de plus en plus sollicitée dans nos sociétés modernes avec l'avènement des téléphones mobiles, des logiciels et automates à commande vocale, ajouté aux communications verbales interpersonnelles (Haxer et al. 2001 ^[137]).

Cet usage important de la voix n'est pas sans conséquence sur la santé du larynx chez certains individus, et est souvent aggravé par des conditions adverses de communication, obligeant les individus à augmenter encore plus leur effort vocal (mauvaise acoustique de salles, absence de sonorisation, bruit de fond important). Ainsi, des études ont relevé chez les enseignants de maternelle et d'élémentaire un usage moyen de la parole de 5 heures par jour en continu, à une intensité supérieure à 80 dB pour la moitié du temps, à cause d'un bruit de fond de la classe de l'ordre de 70 dB (Sodersten et al. 2002 ^[321] ; Szabo et al. 2003 ^[340]). De nombreuses personnes ressentent une gêne ou une irritation après de tels efforts. Elles peuvent également vivre des épisodes d'aphonie ou d'enrouement important ponctuels ou récurrents, voire développer sur le moyen terme des problèmes vocaux plus graves avec lésion des cordes vocales (nodules, polypes, oedèmes,...).

Ces troubles de la voix ne sont pas seulement « anecdotiques » ou liés au confort des individus, mais ils constituent un réel problème de santé publique ayant des conséquences économiques. Ainsi, des problèmes vocaux empêchent en moyenne 2 jours par an les enseignants des Etats Unis d'assurer leurs cours, ce qui coûte 638 millions de Dollars annuels pour les soigner et les remplacer (Autesserre et al. 2006 ^[16]). Les pays scandinaves sont les premiers à avoir pris mesure de l'ampleur de ce problème et mené depuis une quinzaine d'années un grand nombre de travaux sur la dysphonie et sur le forçage vocal, aussi bien d'un point de vue économique, épidémiologique que préventif. En France, la Mutuelle Générale de l'Education Nationale (MGEN) a récemment sollicité l'INSERM (Institut National de la Santé Et de la Recherche Médicale) pour effectuer un bilan des connaissances scientifiques concernant les troubles de la voix chez les personnes l'utilisant dans le cadre de leur travail, en particulier chez les enseignants (Autesserre et al. 2006 ^[16]). Ce problème est donc de plus en plus pris au sérieux par les services de santé.

Si certaines professions sont considérées comme « à risque » vis à vis des troubles de la voix et ont fait l'objet de nombreuses études (Rantala et al. 1999 ^[280] ; Sodersten et al. 2002 ^[321] ; Long et al. 1998 ^[225] ; Mann et al. 1999 ^[232] ; Lehto et al. 2003 ^[206] ; Kitch et al. 1994 ^[181]), il existe également des différences interindividuelles face au problème du forçage vocal : les femmes sont ainsi davantage touchées que les hommes, de même que certains types de personnalités (Vilkman 2004 ^[365] ; Roy et al. 2000 ^[295] ; Giovanni 2007 ^[114]). Il semble également que le risque augmente avec l'âge (Autesserre et al. 2006 ^[16]).

Pourquoi certains individus sont-ils plus sujets que d'autres à développer des troubles de la voix ? Telle est la question très générale qui nous a motivés à mener ce travail, dans le but si possible de prévenir ces troubles ou d'apporter une aide aux personnes qui en souffrent.

Jusqu'à maintenant, ces disparités ont principalement été expliquées par des différences physiologiques de constitution ou de tonicité du larynx grâce à la pratique d'exercices vocaux, permettant à certains individus d'être plus endurants que d'autres face à un même effort vocal. L'origine de ces différences a également été recherchée au niveau des habitudes des individus (tabagisme, consommation d'alcool, etc.) ou des facteurs environnementaux auxquels ils sont confrontés (exposition à de la poussière, à des vapeurs irritantes, à un air sec, etc.). Pourtant, on constate que certaines personnes communiquent sans grand effort apparent en situation adverse (mauvaise acoustique de salle, bruit, absence de sonorisation) tandis que d'autres s'épuisent sans succès. De même, certaines techniques vocales semblent permettre aux acteurs ou aux chanteurs qui les utilisent de réaliser des efforts vocaux importants sans se faire mal. Dans ce travail, nous faisons donc l'hypothèse que les différences interindividuelles face au forçage vocal puissent également en partie provenir de différences de gestion de l'effort vocal par l'individu lui-même, et pas uniquement de

différences physiologiques. Nous nous intéresserons donc à caractériser le forçage vocal à travers les différences interindividuelles de gestion de l'effort vocal.

Pour cela, nous pourrions essayer de caractériser les différences de comportements phonatoires en situation d'effort vocal et tenter de les relier à l'apparition ou non de symptômes, à leur apparition plus ou moins rapide, ou au statut plus ou moins grave de ces symptômes. Cependant, cette démarche ne permettrait pas de distinguer ce qui serait lié à la gestion de chaque individu de ce qui serait lié à sa constitution.

C'est pourquoi nous avons préféré considérer les différences interindividuelles de gestion de l'effort vocal du point de vue de leur efficacité. Mais en quoi consiste cette efficacité et comment l'évaluer ?

- Une première façon d'envisager cette efficacité consiste à évaluer le rapport entre l'énergie acoustique du son produit et l'énergie dépensée par les muscles laryngés, ou en évaluant le rapport entre l'énergie acoustique du son produit et le produit de la pression sous-glottique par le débit d'air expiré (Isshiki 1981^[164]).
- Cependant, si les symptômes ou les douleurs consécutives à un effort vocal sont majoritairement localisées au niveau du larynx ou des cordes vocales, l'acte phonatoire ne se résume pas au seul fonctionnement du larynx mais à la gestion combinée des muscles respiratoires, articulatoires et phonatoires. Aussi, l'inefficacité d'une gestion de l'effort vocal par un individu ne provient pas nécessairement d'un dysfonctionnement au niveau du larynx mais peut provenir du dysfonctionnement d'un autre endroit de cette chaîne de production, faisant alors par compensation porter trop de poids sur le larynx. Une autre façon d'envisager l'efficacité de la phonation consiste dès lors à caractériser le fonctionnement respiratoire, phonatoire et articulatoire en situation d'effort et à examiner comment ces différents « étages » se répartissent l'effort de la production du son.
- Enfin, il ne faut pas perdre de vue lorsque la voix est utilisée dans un cadre professionnel, comme cela est le cas pour les enseignants, les acteurs ou les opérateurs téléphoniques, l'effort vocal n'est pas « subi » mais régulé par ces individus. Ils ne produisent en effet pas de sons ni d'effort sans but, mais avant tout pour communiquer un message ou une émotion à un interlocuteur ou à une assistance. L'efficacité d'une gestion individuelle de l'effort ne nous semble donc pas tant reliée au rapport entre l'effort dépensé et les caractéristiques acoustiques du son produit, mais davantage au rapport entre l'effort dépensé et l'intelligibilité perçue par un interlocuteur, dans un certain contexte de communication. Certaines personnes pourraient alors être plus efficaces que d'autres dans leur adaptation à une situation de communication :
 - en ajustant de façon adéquate leur effort aux besoins de la situation de communication, tandis que d'autres pourraient surestimer les exigences de la situation ou au contraire sous-estimer leur propre intelligibilité, et dans tous les cas dépenseraient plus d'effort que ce qui est nécessaire.
 - en adoptant des stratégies de communication moins coûteuses en énergie, pour une même intelligibilité perçue par l'auditeur (par exemple des stratégies linguistiques de renforcement de certains indices de la parole plutôt que d'augmenter uniquement l'intensité vocale de façon globale)
 - en optimisant la réalisation « technique » de ces stratégies, afin de minimiser l'énergie dépensée et de répartir autant que possible l'effort entre le larynx et les autres organes de la parole, pour un même résultat produit (par exemple en accolant bien les cordes vocales pendant la phonation pour éviter des fuites d'air, ou en augmentant l'intensité vocale par des techniques résonnantes et par une bonne gestion du souffle, plutôt que par serrage laryngé).

Nous ferons donc l'hypothèse durant ce travail que les différences interindividuelles face au forçage vocal peuvent en partie provenir de différences de gestion de l'effort. Cette gestion ne consisterait pas uniquement à utiliser certaines techniques vocales de placement de la voix ou de gestion du souffle, mais plus largement à adopter des stratégies de communication dont l'effort serait adapté ou non aux exigences d'intelligibilité de la situation et serait plus ou moins partagé entre le larynx et les autres étages de la phonation.

Nous avons donc décidé de nous intéresser durant cette thèse à une situation de communication particulière nécessitant une adaptation de l'effort à fournir par le locuteur pour être compris d'un interlocuteur. La communication en environnement bruyant représente une telle situation, et constitue en outre un des principaux facteurs environnementaux rencontrés par les enseignants et contribuant à leurs troubles vocaux. Nous chercherons donc à caractériser les différences interindividuelles d'adaptation de la parole à un environnement bruyant, dans le but de comprendre en quoi cette adaptation peut devenir forçage chez certains individus. Pour y parvenir, nous avons adopté la démarche suivante :

Etape méthodologique préliminaire et constitution de bases de données (Partie 2)

Avant toute chose, les aspects communicationnels ou adaptatifs du forçage vocal que nous proposons d'étudier dans cette thèse nous amènent à envisager de nouveaux paradigmes expérimentaux pour appréhender au mieux notre objet d'étude. Nous avons commencé par recenser les connaissances relatives aux différents mécanismes physiologiques et psychologiques mis en jeu dans l'adaptation de la parole en environnement bruyant, ainsi que les facteurs pouvant influencer ces mécanismes, afin d'en tenir compte dans nos protocoles expérimentaux (Chapitre 2).

- Nous avons alors examiné si le port d'un casque pour immerger les locuteurs dans le bruit en laboratoire avait une influence sur l'adaptation de leur parole dans le bruit, et si cela avait par conséquent un intérêt d'adopter une autre méthode pour appréhender notre objet d'étude (Chapitre 3). Pour cela, nous avons comparé la parole produite pour une immersion du locuteur dans le bruit par casque avec la parole produite pour une reconstitution d'un environnement bruyant par le biais de haut-parleurs. Cela nous a obligés à trouver une technique de débruitage pour pouvoir effectuer des mesures rigoureuses sur le signal acoustique de parole. Cette expérience nous a amenés à constituer une première grande base de données acoustiques et électroglottographiques (BD1) composée de 10 locuteurs immergés dans 2 types de bruit à 4 niveaux d'intensité et pour 3 moyens d'immersion des locuteurs dans le bruit. Nous avons par la suite de nouveau exploré une des sessions de cette base de données, où les locuteurs sont immergés dans le bruit par le biais de haut-parleurs, pour caractériser les stratégies individuelles d'adaptation de la parole dans le bruit.
- De même, nous avons examiné si l'interaction avec un interlocuteur influence l'adaptation des locuteurs en environnement bruyant, et si cela avait par conséquent un intérêt, par rapport à nos objectifs, de proposer aux locuteurs des tâches de production de parole autres que celle consistant à lire à voix haute un matériel linguistique imposé et connu par l'interlocuteur (Chapitre 4). Pour cela, nous avons proposé plusieurs jeux permettant d'enregistrer des locuteurs en situation de communication semi-spontanée interactive ou non, tout en contrôlant le matériel linguistique. Cette expérience nous a amenés à constituer un corpus pilote (présentée en Annexe Pil2) puis une deuxième base de données acoustiques, électroglottographiques et articulatoires (BD2) composée cette fois de 3 locutrices immergées dans 2 types de bruit et pour 2 conditions d'énonciation (avec et sans interaction avec un interlocuteur). Nous avons par la suite de nouveau exploré une des sessions de cette base de données, où les locutrices interagissent avec un interlocuteur, pour caractériser les stratégies individuelles d'adaptation de la parole dans le bruit.

Identification des différentes stratégies de communication adoptées en environnement bruyant, selon les locuteurs, le type et le niveau de bruit ambiant (Partie 3)

Nous testons dans ce travail l'hypothèse que l'adaptation des locuteurs en environnement bruyant est principalement motivée par la recherche d'intelligibilité vis à vis d'un interlocuteur. C'est pourquoi nous avons envisagé différentes stratégies de communication pouvant être potentiellement adoptées par les locuteurs pour compenser la dégradation de l'intelligibilité induite par un bruit ambiant, et permettant d'expliquer l'évolution des différents paramètres acoustiques et articulatoires de la parole observée lors du passage d'une condition calme à une condition bruyante (Chapitre 5).

- Pour commencer, nous avons examiné s'il existe des stratégies acoustiques adoptées par les locuteurs, visant à renforcer certaines caractéristiques globales de leur parole afin d'émerger par rapport au bruit ambiant (Chapitre 6). Nous avons en particulier exploré si les locuteurs cherchent à préserver un rapport signal sur bruit acceptable, à renforcer leur dynamique vocale ou plutôt à adapter leur fréquence fondamentale ou le timbre de leur voix aux caractéristiques spectrales du bruit ambiant.
- Nous avons ensuite examiné l'hypothèse de l'existence, chez certains individus, de stratégies de renforcement des indices acoustiques ou visuels de reconnaissance des phonèmes dans le bruit (Chapitre 7). Nous avons en particulier étudié si l'évolution des trois premiers formants vocaliques dans le bruit allait dans le sens d'un meilleur contraste perceptif entre les extrema [a], [i] et [u] du système vocalique. De même, nous avons examiné si l'évolution dans le bruit des mouvements articulatoires labiaux d'ouverture, d'étirement et de protrusion pouvait contribuer à une meilleure distinctivité visuelle de ces voyelles. Pour finir, nous avons également exploré s'il peut exister une stratégie de renforcement des indices visuels de reconnaissance des consonnes bilabiales.
- Nous nous sommes ensuite intéressés aux indices prosodiques de structuration de l'énoncé et à leur potentiel renforcement dans le bruit (Chapitre 8). Ainsi, nous avons examiné si l'évolution de la dynamique d'intensité dans le bruit pouvait contribuer au renforcement du marquage syllabique. Nous avons également exploré si les débuts et fins d'énoncés étaient davantage marqués dans le bruit au niveau acoustique et articulatoire, de même que la frontière entre les déterminants et les mots de contenu.

- Pour finir, nous avons examiné s'il existait des stratégies d'adaptation de la parole dans le bruit liées à la quantité d'information variable apportée par des mots dans l'énoncé (Chapitre 9). Ainsi, nous avons exploré si certains locuteurs renforcent davantage les mots de contenu dans le bruit que les déterminants.

Après avoir examiné ces différentes hypothèses, nous avons cherché à déterminer si les locuteurs adoptent ces différentes stratégies de façon sélective et combinée ou de façon exclusive. De même, nous avons exploré si les locuteurs adoptent les mêmes stratégies dans tous les types de bruit ou si leur adaptation en est dépendante. Enfin, nous avons cherché à savoir si l'adaptation des locuteurs évolue de façon graduelle en fonction du niveau de bruit ou s'il existe une réorganisation de l'adaptation en fonction du niveau de bruit (Chapitre 10).

Evaluation de l'efficacité de ces différentes stratégies de communication dans le bruit

L'identification des différences interindividuelles d'adaptation de la parole est une première étape dans notre démarche de caractérisation du forçage. Il est ensuite nécessaire d'évaluer l'efficacité de ces adaptations pour pouvoir les relier à un comportement de forçage ou à un comportement de gestion efficace de l'effort vocal.

Sur ce point, nous ne pourrions apporter dans cette thèse que de premiers éléments d'information et de premières considérations à développer lors de travaux futurs. Nous pourrions tout d'abord comparer l'effort dépensé par les différents locuteurs pour communiquer dans le silence, dans le bruit et pour s'adapter du silence au bruit. Nous examinerons également comment certaines stratégies impliquent une répartition des efforts sur les différents organes de la parole, ou comment les caractéristiques laryngées suggèrent une inefficacité ou une tension lors de la phonation.

Le but de cette thèse est d'apporter de nouvelles connaissances sur les mécanismes qui sous-tendent l'adaptation de la parole en environnement bruyant, afin tout d'abord de déterminer si les individus possèdent une marge de contrôle de cette adaptation et par conséquent une marge de gestion de l'effort vocal qu'elle nécessite. Outre l'intérêt théorique de tels résultats pour une meilleure compréhension de l'effet Lombard, cela pourrait ouvrir de nouvelles perspectives comportementales quant à la prévention du forçage vocal dans de tels environnements. La ligne directrice de ce travail concerne également la comparaison des comportements individuels d'adaptation de la parole dans le bruit, afin de déterminer si d'éventuelles différences dans la mise en oeuvre de stratégies de communication, dans l'utilisation de techniques vocales, ou dans la perception et l'analyse de la situation de communication, peuvent contribuer à ce que certains individus soient davantage sujets que d'autres à développer des troubles de la voix. L'existence et l'identification de telles différences seraient d'un grand intérêt pour la prévention et la rééducation du forçage vocal et permettrait d'apporter une aide concrète aux personnes touchées par ce problème ou à amenées à communiquer très fréquemment dans de tels environnements bruyants, en les amenant à travailler sur leur propre comportement et en leur enseignant certaines techniques, plutôt qu'en cherchant uniquement des solutions extérieures pour remédier à ce problème (sonorisation, hydratation, ...).

Partie I

Acquis et hypothèses sur le forçage vocal

Chapitre 1 :

Le forçage vocal

- 1. DEFINITIONS DU FORÇAGE VOCAL**
- 2. ÉTAT DES CONNAISSANCES SUR LE FORÇAGE VOCAL**
- 3. POSITIONNEMENT ET HYPOTHESES DE NOTRE TRAVAIL**
- 4. CONCLUSION DU CHAPITRE 1**

Chacun a une idée intuitive de ce que signifie : « forcer sur sa voix », souvent basée sur une propre expérience vécue en ayant parlé dans un bar bruyant par exemple, en ayant chanté dans une chorale, ou en ayant enseigné pendant toute une journée. Pour autant, il y a une grande différence entre avoir une idée globale du phénomène et en donner une définition précise, permettant de poser le forçage vocal en tant qu'objet scientifique et de l'étudier.

Quelle sont précisément les caractéristiques communes de ces expériences vécues qui en constituent l'aspect de forçage: le fait d'avoir fourni un effort important, inhabituel ? Le fait d'avoir eu la voix « cassée » après ces expériences ? Le fait d'avoir ressenti une difficulté ou une douleur pendant ces expériences? Tout cela à la fois ?

Mais surtout, par rapport à notre objectif, qui est de comprendre pourquoi certaines personnes sont davantage touchées que d'autres par ce problème, existe-t-il des caractéristiques individuelles de ces expériences évitables ou modifiables pour se prévenir du forçage? Si oui lesquelles ?

Voilà les questions qui nous ont amenés dans un premier temps, à chercher une définition de référence du forçage vocal dans la littérature scientifique, puis celle-ci faisant défaut, à déterminer par un travail d'enquête les attributs importants de ce concept.

Cela nous a permis de dégager différents axes d'expertise, de champs disciplinaires et de positionnements épistémologiques selon lesquels organiser et présenter les connaissances très diverses ayant été apportées sur le problème du forçage vocal.

Une fois cet état des connaissances posé, nous pourrions préciser notre positionnement vis à vis de cet ensemble de points de vue et de connaissances, relatif à notre propre champ disciplinaire (acoustique-phonétique, psychologie) et aux objectifs de cette thèse. Nous avons en particulier identifié certains aspects du forçage vocal non abordés dans la littérature dont nous effectuons l'hypothèse qu'ils puissent avoir une influence déterminante sur les différences interindividuelles. Nous argumenterons pourquoi il semble intéressant vis à vis de nos objectifs de prendre en compte le rôle actif du locuteur dans le phénomène de forçage ainsi que l'aspect communicationnel qui est la source de la motivation du locuteur à adapter son effort vocal.

1. DEFINITIONS DU FORÇAGE VOCAL

Avant toute chose, nous avons cherché à comprendre plus précisément la notion de forçage vocal, en explorant tout d'abord sa signification dans la littérature. Devant l'absence d'une définition consensuelle adoptée en référence, nous avons alors cherché à déterminer les attributs de ce concept en explorant les conceptions de divers experts de ce domaine sur cette thématique.

1.1. Définitions du forçage vocal dans la littérature

1.1.1. Le forçage vocal dans la littérature francophone

Dans le sens commun, le mot « forçage » est défini par le TLF^[1] comme « l'action d'insérer, de bloquer par une vive pression » ou comme un « effort excessif ». Mais le terme plus spécifique de « forçage vocal », utilisé principalement en phoniatry et en orthophonie, est apparu à partir des années 80 dans la littérature francophone.

Pour plusieurs auteurs, le forçage vocal est globalement un comportement vocal adopté pour faire face à une situation ponctuelle, provoquant des conséquences pathologiques pour la voix. Ainsi, Papon 2006^[265] le définit comme « une utilisation excessive de la voix dans le cadre de situations de communication particulières » et Ormezzano 2000^[262] comme « une mauvaise réponse dans l'utilisation de la voix face à un effort ou une situation d'agression du système phonatoire lui-même ». Parallèlement, Giovanni 2006^[113] parle d'« un comportement vocal complexe susceptible d'entraîner l'apparition sur les cordes vocales de lésions dites dysfonctionnelles et notamment de nodules ».

Pour d'autres auteurs, le forçage vocal est lui-même « la conséquence d'un geste malencontreux fait consciemment ou non par l'orateur insatisfait de sa voix telle qu'il la perçoit ou croit la percevoir » (Guerin 2004^[128]) ou « un mode phonatoire pathologique produit par un mauvais comportement vocal » (Guerrero Lopez 2004^[130])

Enfin, certains auteurs définissent le forçage vocal non par comme un comportement ponctuel mais comme un cercle vicieux de comportement d'effort (Ormezzano 2000^[262] ; Le Huche 1984^[202] ; Guerin 2004^[128]). La figure 1 représente cette dernière conception.

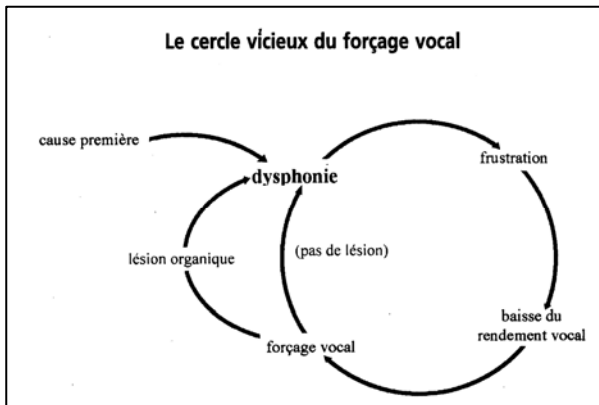


Figure 1. Représentation et conception du forçage vocal comme un cercle vicieux, qui débute en général par une cause extérieure (laryngite, surmenage, ...). La dysphonie qui en résulte gêne le sujet (frustration), ce qui entraîne une baisse de son rendement vocal, induisant un forçage vocal, d'où une nouvelle dysphonie. Une éventuelle lésion des cordes vocales induite par le forçage vocal peut participer au cercle vicieux et augmenter encore la dysphonie. (D'après Ormezzano 2000^[262]).

Nous n'avons pas trouvé de définition de référence dans cette littérature francophone. Et bien que ces différentes définitions partagent l'aspect pathologique du forçage vocal, elles divergent quant au fait qu'il s'agisse d'un comportement ponctuel ou chronique, ou encore d'un état de symptômes.

1.1.2. Le forçage vocal dans la littérature anglophone

Il n'existe pas de traduction exacte du terme de forçage vocal en anglais, ce qui ne simplifie pas le problème du manque de consensus autour de ce concept. Les notions de « vocal loading », « vocal load », ou « vocal strain » sont celles qui s'en rapprochent le plus. Le premier article traitant de la notion de « vocal strain », d'un point de vue biomécanique (Rumsey 1950^[296] et Sonninen et al. 1972^[324]), fut publié au milieu du XX^{ème} siècle.

Le terme de « vocal loading » correspond plutôt à un usage prolongé de la voix, et à des études menées par des orthophonistes ou des acousticiens de la voix. Ainsi, Vintturi 2001^[368] définit le « Early stage of vocal loading (ESVL) » comme les 45 premières minutes de phonation et Rantala et al. 1999^[280] introduisent un « vocal loading index » correspondant au nombre moyen de vibrations des cordes vocales par unité de temps sur une journée de travail.

Tandis que le terme « vocal strain » renvoie davantage à un usage excessif ou à forte intensité de la voix, et est utilisé aussi bien dans des études biomécaniques, orthophoniques ou psychologiques. Ainsi, Sonninen et al. 1998^[325] définissent le paramètre « vocal strain » comme le pourcentage d'augmentation de la longueur des cordes vocales par rapport à leur longueur au repos. Mais le terme « Strain » constitue également un critère d'évaluation de la qualité des voix pathologiques sur l'échelle perceptive GRBAS (Hirano 1981^[147]). Il est d'ailleurs traduit en français par « voix forcée ». Enfin, les « voice straining factors » recensés dans plusieurs études anglophones (Vintturi 2001^[368] ; Vilkmán 2004^[365] ; Buekers 1998^[38]) sont semblables aux facteurs de forçage vocal présentés dans les études francophones (Le Huche 1984^[202] ; Papon 2006^[265] ; Ormezzano 2000^[262] ; Giovanni 2006^[113]).

L'équivalent anglophone du forçage vocal pourrait donc être « vocal straining » ou « voice straining », bien qu'aucune étude n'ait à notre connaissance encore été publiée sous cette dénomination.

1.1.3. Les autres notions connexes au forçage vocal

Lorsqu'on effectue une recherche bibliographique par mot-clé sur le *forçage vocal*, sur *vocal loading* ou *vocal strain*, ou par auteurs ayant déjà publié sur ces notions, les articles rassemblés – parus essentiellement durant les quinze dernières années – font émerger une multitude d'autres termes et notions connexes du forçage, tels que *l'effort vocal*, *la fatigue vocale*, *la dysphonie*, *le malmenage vocal*, *le surmenage vocal*, *l'efficacité vocale*, *l'échauffement vocal*, *le repos vocal* (ou en anglais *vocal effort*, *vocal fatigue*, *dysphonia*, *misuse*, *overuse of the voice*, *vocal efficiency*, *warm-up*, *vocal rest*). Le recouvrement ou la distinction de ces différentes notions n'est pas toujours évidente. C'est pourquoi nous avons cherché à mieux comprendre les relations qu'entretiennent ces différents termes avec la notion de forçage vocal, et ainsi à déterminer les attributs et les frontières du concept de forçage vocal.

1.2. Enquête auprès de différentes personnes concernées par le forçage vocal

Nous avons alors interrogé plusieurs personnes sur leur conception de ces différents termes rencontrés dans la littérature. Ces personnes sont toutes concernées par le forçage vocal, mais avec une expertise différente :

- 4 médecins (M1 - M4), 5 orthophonistes (O1 - O5) et 5 professeurs de chant (PC1 - PC5) ont une expérience pratique du diagnostic, de la prévention ou de la rééducation du forçage vocal.
- 3 enseignants (E1- E3) ont une autre expérience pratique du forçage vocal, cette fois en tant que sujets à ce problème.
- enfin 4 chercheurs en phonétique (CP1 – CP4) ont une expérience théorique du forçage vocal.

Ces 21 personnes ont accepté de répondre à notre questionnaire écrit consistant à expliciter leurs propres conceptions du *forçage vocal*, de *l'effort vocal*, d'une *voix forcée*, du *malmenage vocal*, de la *fatigue vocale*, de *l'efficacité vocale* et de *l'échauffement vocal*. L'intitulé exact des questions de ce questionnaire ainsi que le détail des réponses individuelles, sont exposés en Annexe Res1. Nous ne présenterons ici qu'une synthèse des résultats obtenus.

1.2.1. Le concept de « Forçage vocal »

1.2.1.1. Un excès, une inadéquation, voire une pathologie

Pour seulement quelques personnes, le forçage vocal est un concept relativement neutre, proche de la notion d'« effort vocal ». Il désigne alors une utilisation importante ou inhabituelle de la voix, c'est à dire un **écart à une « norme »** de phonation, sans autre jugement de valeur. La norme dont le locuteur s'écarte en forçant sur sa voix est rarement explicitée : il s'agirait de « *la passivité du larynx* », de « *l'équilibre énergétique entre pression sous-glottique, résistance de la glotte, et impédance des résonateurs* », du « *soutien diaphragmatique* » ou de « *l'utilisation d'un mécanisme de voix projetée* » (en référence implicite à Le Huche 1984^[202]).

Excepté dans ces définitions, le forçage vocal est quasi unanimement conçu comme un phénomène négatif pour trois raisons différentes:

- en premier lieu, la négativité du forçage vocal est associée à la notion d'**excès**. Il n'y a plus seulement l'idée d'un écart par rapport à une « norme » de phonation, mais également le concept d'un « seuil » dépassé. Celui-ci n'est également pas souvent précisé. CP3 néanmoins, écrit clairement que « *le seuil [est] le niveau d'intensité minimale de la voix audible en situation conversationnelle* » et PC4 que le forçage est le fait d'« *aller au-delà de la limite garantissant la préservation du bon état de la voix* ».
- le deuxième point concerne la conception **pathologique** du forçage vocal, très souvent défini comme un symptôme vocal ou comme une cause de symptômes.
- enfin, l'aspect négatif du forçage vocal correspond souvent une **inadéquation** du comportement vocal. L'objectif ou la situation auxquels la phonation n'est pas adaptée ne sont encore que très rarement explicités. M4 explique néanmoins que l'« *énergie musculaire n'est pas en rapport avec l'efficacité acoustique du son produit* » et PC2 que ce n'est « *pas adéquat en rapport avec l'effet recherché* ».

1.2.1.2. Un objet à différentes temporalités

On note dans les réponses au questionnaire une ambivalence vis à vis de la temporalité du forçage vocal.

Ainsi, il est défini la plupart du temps comme un **comportement**, un geste ou une action effectué **ponctuellement pendant** la phonation, ou comme la **conséquence** d'un comportement. Pour d'autre, il s'agit surtout de la **répétition** d'un comportement, pouvant prendre la forme d'un engrenage. Enfin pour certains, le forçage vocal est un objet **atemporel**, correspondant à la fois à un processus et à son résultat (*altération; perturbation; dysfonctionnement*).

L'aspect ponctuel ou récurrent du forçage vocal est donc ambigu, et se retrouve dans la notion de *malmenage vocal*. De même, l'aspect concomitant ou conséquent à la phonation du forçage vocal se retrouve dans la notion de *fatigue vocale*. Le terme de *comportement vocal* semble lui-même ambigu puisqu'il désigne parfois un mode de phonation ponctuel (qui ne se limite pas pour autant à un geste laryngé ou respiratoire), et d'autres fois rend compte d'une attitude ou

d'un état vocal plus général du locuteur, incluant ses antécédents, son hygiène vocale, sa personnalité ou sa façon de réagir à la situation.

1.2.1.3. Des cordes vocales à la communication globale

De nombreuses personnes interrogées conçoivent le forçage vocal comme un phénomène localisé au niveau du **larynx** ou des **cordes vocales**. PC1 explique clairement que « *Le terme [de forçage vocal] est impropre et amène à l'erreur. Le forçage n'est pas vocal mais laryngé* ». Pourtant, plus de la moitié des définitions traduisent une conception plus large du forçage vocal étendue à la **voix** et à l'appareil phonatoire en général. 24% des définitions évoquent le rôle de la **respiration** dans le forçage. Et 14% définissent également le forçage vocal par des tensions musculaires au niveau d'autres parties du **corps** (*Musculature cervicale ; articulaire ; postural ; région du cou ; de la nuque ; du maxillaire*). Enfin, quelques définitions seulement soulèvent un éventuel aspect **psychologique** (*aphonie psychogène ; sans respect du repos de l'organisme ; on s'oppose volontairement*) ou **communicationnel** du forçage vocal (*lors d'un chant et de l'émission de la parole ; situation conversationnelle ; pour se faire entendre ; pour mieux faire passer un message*).

Par conséquent, le forçage vocal est dans la majorité des définitions un concept objectivé et décontextualisé : La présence d'un locuteur est très rarement mentionnée, sous-entendue alors à travers ses actions, essentiellement de contrôle, et à travers son ressenti. De même, le forçage n'est presque jamais défini par une situation ou par des motivations. Seul E1 évoque l'intention de « se faire entendre ; [de] mieux faire passer un message », PC2 d' « obtenir un son » et PC4 la « recherche d'efficacité ».

1.2.2 Le concept de voix forcée

Une voix forcée est pour 81% des interrogés une voix **en situation** de forçage et pour 19% une voix **après** situation de forçage. Elle est définie par des **symptômes cliniques** mais surtout par des symptômes perceptifs au niveau de la **qualité vocale**, en particulier par une impression de **manque de contrôle** ou de **perte d'efficacité**.

1.2.3 Le concept d'effort vocal

On peut distinguer clairement à partir des réponses deux significations distinctes du terme d'effort vocal : la première est neutre, voire positive (43%), tandis que la deuxième est négative (28.5%).

- la première signification de l'effort vocal correspond à une augmentation de l'**énergie musculaire** et respiratoire nécessaire à la phonation, sans que cela ne soit considéré comme pathologique. Il s'agit simplement d'un mode de phonation qui **s'écarte d'une norme** (toujours non explicitée).
- la deuxième signification de l'effort vocal correspond à une utilisation **excessive** du larynx au-delà de ses possibilités. On retrouve l'idée d'un « seuil » dépassé, déjà rencontrée dans la définition du forçage vocal, tout comme l'aspect **nocif** d'un tel comportement, **inadapté** ou **inefficace**. On remarque que l'idée d' « énergie » émergeant dans la première signification devient celle de **tension** dans la deuxième et l'on passe d'une mobilisation musculaire générale à une focalisation de l'effort au niveau des **cordes vocales**.

28.5% mentionnent d'eux même qu'ils font la distinction entre ces deux significations. L'explication qu'ils en donnent permet de mieux comprendre la nature du « seuil » dépassé lors du forçage vocal :

- le forçage vocal commencerait à partir du degré d'effort vocal où le locuteur ne possède plus les ressources techniques ou physiologiques pour pouvoir résister et continuer à être efficace (*au-delà de ses possibilités acoustiques et anatomiques ; l'effort n'est pas synonyme de forçage quand le larynx PEUT résister à la pression sous glottique ; pression sous glottique trop importante [devant] l'insuffisance de la protection ; donner plus que ce que nous avons*).
- le forçage vocal commencerait à partir du moment où l'effort vocal est mal géré ou de façon inefficace par rapport aux objectifs de la phonation (*utilisation d'un geste brutal [...] pour peu d'efficacité vocale ; ça devient pathologique si l'utilisation de l'appareil vocal est inadaptée ; c'est l'effort contrôlé qui permet d'obtenir l'efficacité vocale ; le forçage, c'est l'effort mal géré*).

1.2.4. Le concept de malmenage vocal

En surface, le malmenage vocal est conçu de façon unanime comme une **mauvaise utilisation** de sa voix.

Cependant, on peut distinguer 4 principales causes pour lesquelles cette utilisation est jugée « mauvaise ». Celles-ci permettent toutes d'établir un lien entre le malmenage et le forçage vocal, mais avec des nuances différentes.

- Cette utilisation peut d'abord être mauvaise car **excessive**, **inadaptée**, **inefficace** ou entraînant des **symptômes** pour la voix. Par ses caractéristiques, le malmenage pourrait donc être synonyme de forçage vocal. Certaines définitions d'ailleurs le mentionnent explicitement.
- 48% des définitions suggèrent l'idée que cet usage de la voix est mauvais du fait de sa **répétition** dans le temps.
- Pour 4 personnes interrogées, le malmenage est un stade avancé du forçage dans lequel le locuteur **s'acharne** à mal utiliser sa voix malgré l'apparition de symptômes qui nécessiteraient au contraire un repos vocal.
- Au contraire, pour O4, le malmenage vocal est un « *stade avant le forçage vocal, qui n'est pas encore accompagné de signes vocaux entravant la vie professionnelle et sociale* », autrement dit une mauvaise utilisation de la voix **sans** que des **symptômes** ne soient encore apparus.

Il est intéressant de noter que les personnes interrogées évoquent particulièrement au travers de cette définition du malmenage vocal la notion de **conscience/inconscience** du locuteur d'être en train de forcer ou de fatiguer. M1 définit même le malmenage comme un « équivalent du forçage, avec la nuance de traficotage volontaire surajouté ». C'est la seule personne à penser que le forçage puisse être effectué délibérément. Par contre, d'autres sujets interrogés mentionnent le fait que le locuteur puisse être conscient de forcer sur sa voix, sans pour autant pouvoir ou vouloir s'arrêter ou au contraire, que son problème consiste à ne pas être attentif à la façon dont il gère sa voix et aux signes de fatigue.

1.2.5. Le concept de fatigue vocale

La fatigue vocale correspond pour certains à un synonyme du forçage vocal mais est le plus souvent définie comme la **conséquence** du forçage vocal ou de l'effort vocal.

De façon similaire à la voix forcée, la fatigue vocale est un état de **symptômes** observables par un bilan clinique, ressentis subjectivement par le locuteur ou perceptibles auditivement au travers de sa **qualité vocale**, avec une **perte de contrôle et d'efficacité** de la voix.

La fatigue vocale est peu localisée mais plutôt **diffuse** et ressentie par le locuteur à travers l'envie ou le besoin de **repos vocal** (*émission vocale demandant un effort de volonté, manque d'appétence à la communication, Le patient a envie de se taire à ces moments-là, Le corps "signale" au sujet [...] que le repos est nécessaire*). Seul PC1 précise qu'« il faut dire fatigue du larynx (ce ne sont ni les zygomatiques ni les transverses qui sont fatigués, et encore moins la voix) ». Lorsque des parties du corps sont citées, il s'agit davantage de douleurs et non plus seulement de fatigue (*gêne laryngée, gorge sèche, mal à la gorge*).

1.2.6. Bilan

Il existe un consensus certain sur le caractère positif ou négatif des différents concepts examinés ci-dessus. Par contre, l'exploration plus en détail de ces termes montre une pluralité de conceptions, souvent complémentaires, mais parfois aussi contradictoires. Le Tableau 1 récapitule les caractéristiques les plus importantes de ces différents concepts ayant émergé des réponses données à notre questionnaire.

Ce tableau met en évidence le recouvrement de différents concepts par leur partage de certains attributs. Il souligne également les frontières entre les concepts ou les spécificités de certains concepts. Ainsi, le forçage et la fatigue vocale se rapprochent sur le fait qu'ils sont tantôt considérés comme un comportement vocal, tantôt comme un symptôme, tandis que le forçage vocal et le malmenage ont en commun l'aspect éventuellement répétitif de l'utilisation d'un mauvais geste. Les concepts de forçage vocal, d'effort vocal, de fatigue vocale ont également tous en communs de générer des symptômes, contrairement au malmenage qui est considéré par certains comme un mauvais geste ne provoquant pas obligatoirement de symptômes. L'effort vocal peut être synonyme du forçage vocal ou le devenir seulement lorsqu'il dépasse un certain seuil ou lorsqu'il est mal géré. Enfin, le malmenage se distingue pour certains par un manque d'attention

porté à la fatigue vocale et à l'apparition de symptômes, ou pour d'autres par une absence de symptômes, pour d'autres comme un forçage répété de la voix.

	Fatigue vocale	Forçage vocal	Malmenage vocal	Effort vocal	Voix forcée
négatif					
neutre					
Comportement causant des symptômes					
Symptôme causé par un comportement					
Symptôme entraînant un comportement					
Répétition d'un comportement					
Continuation d'un comportement					
Prévention d'un comportement					
Cercle vicieux					
Absence de symptômes					
Symptômes acoustiques					
Symptômes cliniques					
Ecart à une norme					
Franchissement d'un seuil / Excès					
Caractère inadapté					
Caractère inefficace					
Besoin de repos vocal					
Non-respect du repos vocal					
Perte de contrôle					
Laryngé					
Corporel					
Psychologique					
Communicationnel					

Tableau 1. Récapitulatif des attributs de chacun des termes examinés, mettant en évidence les recouvrements et frontières entre les différentes conceptions auxquelles ces termes renvoient.

Le forçage vocal s'avère être un objet complexe et envisagé de multiples points de vues. Ceux-ci proviennent principalement de la pluralité des domaines d'expertise qui abordent ce problème :

- d'un côté, les phoniâtres voient chaque jour arriver des patients touchés par ce problème et ont besoin de définir un certain nombre d'observables (symptômes ou comportements) qui accompagnent ce dysfonctionnement et qui peuvent permettre de le diagnostiquer.
- d'un autre côté, les orthophonistes sont confrontés à la rééducation de ces personnes et cherchent à trouver des techniques vocales, des traitements (hydratation) ou des supports techniques (microphone) pour supprimer ou réduire les symptômes du patient.
- d'un point de vue épidémiologique, les phoniâtres et les orthophonistes s'intéressent également à la prévention du forçage vocal et à déterminer les facteurs favorisant ce dysfonctionnement (environnements, humidité de l'air, personnalités à risque, constitutions anatomiques, etc...)
- enfin, les chercheurs en physiologie de la voix, en biomécanique du larynx, en acoustique ou en phonétique, s'intéressent à caractériser l'effort vocal lui-même, dans le but de comprendre le fonctionnement de modes de phonation extrêmes et en quoi ils peuvent entraîner une fatigue vocale ou des lésions anatomiques, ou encore dans le but de déterminer des indicateurs objectifs du forçage vocal utilisables par les phoniâtres et les orthophonistes.

L'intérêt de ces différents domaines pour le forçage vocal n'est pas le même. C'est pourquoi les différentes études portant sur le forçage vocal ne le définissent pas ni ne l'observent au même niveau (cordes vocales, ressenti du locuteur, etc.), ni au même instant (pendant, après), ni par le biais des mêmes champs disciplinaires (biomécanique, acoustique, psychologie, etc.)

2. ETAT DES CONNAISSANCES SUR LE FORÇAGE VOCAL

Après avoir tenté de mieux cerner le concept de forçage, nous allons maintenant nous intéresser à l'état des connaissances sur cette problématique. Etant donné la multiplicité des conceptions et des façons de caractériser le forçage vocal recensées précédemment, il nous a semblé plus clair de ne pas présenter les connaissances acquises sur le forçage vocal de façon mélangées, et sur un même plan, mais en les situant précisément dans l'espace

- des différents positionnements conceptuels et épistémologiques précédemment recensés de cette problématique, en particulier en ce qui concerne le degré d' « humanité » auquel le forçage est défini.
- des différents champs disciplinaires ayant caractérisé ce phénomène à travers le « tamis » de leurs indicateurs et de leurs théories
- des divers domaines d'expertise ayant motivé, par leurs objectifs différents (diagnostic, prévention, etc.), la caractérisation du forçage vocal à plusieurs niveaux de temporalité et de causalité.

2.1. Un forçage défini selon les études d'un niveau physiologique à un niveau intentionnel

Une première organisation des différentes conceptions du forçage vocal et des études ayant été réalisées sur ce sujet peut être effectuée selon le niveau du système sur lequel porte le forçage (cf. Figure 2).

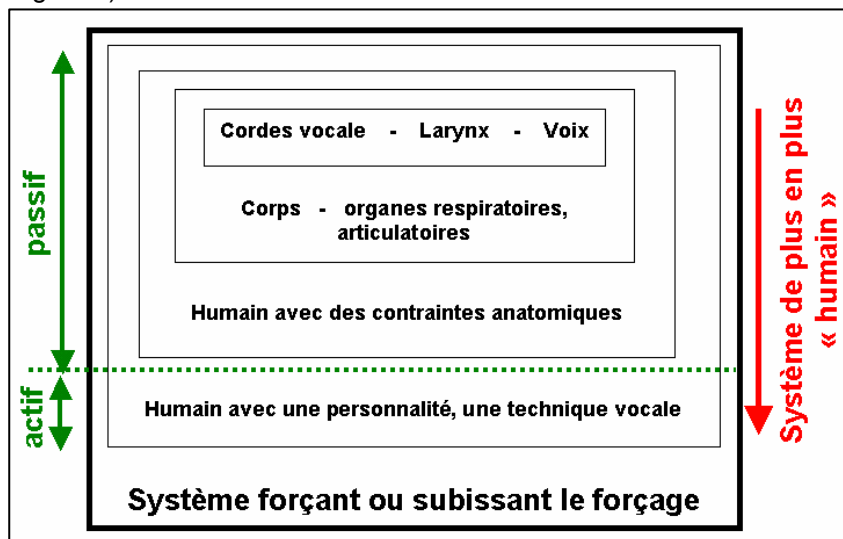


Figure 2. Différents niveaux de définition du forçage vocal

2.1.1. Un problème exclusivement laryngé

Un certain nombre d'auteurs pensent, à l'instar de Ormezzano 2000 ^[262], que « le forçage vocal est avant tout d'origine laryngée, à l'image du rôle sphinctérien du larynx ».

Cette conception a motivé un certain nombre d'études sur la description des patterns de vibration des cordes vocales en relation avec les conséquences pathologiques qu'ils peuvent avoir. De même, de nombreuses études ont été menées sur la voix à forte intensité, pour laquelle la force de collision importante des cordes vocales peut mener à des lésions membranaires, et pour laquelle des tensions musculaires telles que la pression importante des aryténoïdes, peuvent mener à une fatigue neuromusculaire. D'autres études se sont plutôt portées sur le lien entre la fréquence fondamentale et l'apparition de symptômes vocaux. Celle-ci est en effet liée au nombre de collisions des cordes vocales par unité de temps, ainsi qu'à la raideur et à l'élongation des cordes vocales.

De nombreuses autres études ont cherché à montrer que certains facteurs pouvaient affecter les caractéristiques laryngées et créer ainsi un terrain propice au forçage vocal et à l'apparition de symptômes : C'est le cas de maladies ORL mal soignées, de modifications hormonales du

tabagisme ou la forte consommation d'alcool, de l'exposition à la poussière ou à l'air conditionné qui sont susceptibles de créer une irritation des cordes vocales et de mener au forçage vocal. L'influence de l'humidité de l'air a été particulièrement étudiée (Sihvo et al. 1996^[317]; Vilkman et al. 1997^[366]; Vilkman et al. 1999^[367]; Hemler et al. 1997^[138]; Vintturi 2001^[368]) et a montré un effet significatif en défaveur des ambiances sèches. De même, Verdolini et al. 1994^[363] et Chan et al. 2003^[48] ont montré les bénéfices d'une hydratation régulière en buvant de l'eau. L'influence de l'humidité serait reliée à la viscosité de la muqueuse des cordes vocales qui, sans lubrification, va davantage subir de friction et d'échauffement (Verdolini et al. 1994^[363]). Ainsi une expérience de lubrification (avec un liquide plus ou moins visqueux) sur un larynx canin excisé a montré une augmentation du quotient ouvert avec une viscosité croissante. De la même façon, quelques études seulement ont pu montrer que l'échauffement vocal avant une performance avait un effet positif sur l'état de la voix après l'effort (Stemple et al. 1994^[328]). Celui-ci permettrait d'améliorer la mobilité des cordes vocales en augmentant la température des muscles laryngés, et par conséquent en diminuant la viscosité des cordes vocales (Safran et al. 1989^[298]).

2.1.2. Un problème corporel plus global

D'autres auteurs définissent le forçage d'un point de vue toujours physiologique mais plus global, englobant l'ensemble des parties du corps participant à l'acte phonatoire. Ainsi, Morrison 1997^[246] définit la notion de « muscle misuse voice disorders (MMVD) » comme un ensemble de gestes respiratoires, glottiques, articulatoires ou posturaux jugés nocifs et conduisant à la dysphonie. De même, Lozanski 2004^[226] parle d'une « utilisation incorrecte des systèmes musculaires engagés dans la production de la parole et de la voix, incluant une mauvaise utilisation posturale et des muscles du système respiratoire, du larynx, de la langue, de la mâchoire, du visage et de la gorge ».

Des études se sont ainsi attachées à montrer que certains patterns respiratoires ou posturaux n'étaient pas seulement contemporains du forçage mais le constituaient. Ainsi, une respiration thoracique supérieure est jugée propice au forçage vocal par rapport à une respiration abdominale car elle oblige le larynx à assumer, en plus de sa fonction de vibreur, le rôle de régulateur du débit d'air normalement assumé par le diaphragme, induisant une tension laryngée plus importante.

Certaines postures favorisent donc indirectement le forçage vocal puisqu'elles affectent la façon dont la personne respire (Hoit 1995^[151]). Vilkman et al. 1997^[366], Vilkman et al. 1999^[367], Vintturi 2001^[368] ont ainsi montré que la station debout était plus favorable aux hommes qu'une position assise, et que les douleurs dans la nuque, les épaules et le dos se trouvaient alors minimisées.

De même certaines situations émotionnelles de trac (Kitch et al. 1996^[182]) ou d'émotion intense, peuvent avoir une influence sur la respiration et sur la tension générale et sont ainsi recensées parmi les facteurs potentiels du forçage vocal.

2.1.3. Un problème individuel, mais toujours physiologique

Certains auteurs se sont intéressés au fait que toutes les personnes ne sont pas touchées de façon équivalente par le forçage vocal. Elles ont cherché des éléments de compréhension de ces différences interindividuelles au niveau des anatomiques de constitution ou de résistance.

Ainsi, le forçage vocal touche largement plus les femmes que les hommes (Sodersten et al. 2002^[321]; Russell et al. 1998^[297]; Vintturi 2001^[368]; Novak et al. 1991^[257]; Smith et al. ^[320]). Plusieurs études ont tenté d'expliquer cette disparité par la fréquence fondamentale plus élevée chez les femmes, associée par conséquent à un nombre de collisions plus important des cordes vocales par unité de temps (Rantala et al. 1999^[280]; Titze 1984^[350]). Sarfati 1998^[300] soulève la possibilité que les femmes et les enfants soient plus touchés par la dysphonie du fait que leur zone fréquentielle de parole se situe dans la zone de recouvrement des mécanismes laryngés 1 et 2, ce qui n'est pas le cas chez les hommes.

D'autres études ont constaté des durées assez variables d'une personne à l'autre lors d'un usage prolongé de la voix, en particulier une endurance accrue de sujets « entraînés » par rapport à des sujets « non entraînés » (Gelfer et al. 1991^[111]; Scherer et al. 1991 cité dans Stemple et al. 1995^[329]), de la même manière que chez les personnes faisant des exercices vocaux réguliers par rapport à des sujets témoins (Laukkanen et al. 2004^[197], Stemple et al. 1994^[328] et Laukkanen et al. 1996^[199]). Neils et al. 1987^[252] interprètent ces différences par le fait que certaines personnes

ont des muqueuses plus sensibles ou une musculature laryngée plus solide que d'autres. Le travail vocal plus régulier qu'un simple échauffement semble également davantage considéré du point de vue de la tonicité et de la résistance laryngée qu'il améliore, que du point de vue de l'apprentissage d'une technique vocale pour mieux gérer l'effort.

2.1.4. Un problème individuel, lié à la personnalité ou à l'expérience du sujet

Enfin, certaines études conçoivent que les différences interindividuelles face au forçage vocal soient d'origine psychologique ou liée à l'expérience/inexpérience du sujet

Des types de personnalité ont ainsi pu être reliés à différents troubles de la voix, dont la dysphonie fonctionnelle (Giovanni 2007 ^[114]). Roy et al. 2000 ^[295] ont ainsi avancé que les personnalités extraverties ne tiendraient pas compte des signaux de fatigue ou de douleur de l'organisme et seraient ainsi prédisposées au forçage. A l'inverse, les personnalités introverties et anxieuses, seraient sujettes aux aphonies psychogènes. De même, McHugh-Munier et al. 1997 ^[236] ont montré que les sujets avec polypes et nodules utilisent davantage des stratégies d'adaptation émotionnelle au stress et moins des stratégies cognitives que les sujets sans problèmes vocaux. Cette implication émotionnelle se traduirait par des modifications respiratoires, phonatoires et articulatoires conduisant au malmenage, ainsi que par un contrôle interne plus grand pouvant induire un état de tension, en particulier au niveau vocal.

Le comportement phonatoire des chanteurs et des acteurs a également été comparé à celui de sujets « non entraînés », mais cette fois interprété vis à vis de leur expertise de production et non de leurs capacités de résistance. Sonninen et al. 1998 ^[325] ont ainsi observé une tension des cordes vocales avec l'augmentation de la fréquence fondamentale plus grande chez les non chanteurs que chez les chanteurs. Des études sur la voix projetée ont montré la présence d'un renforcement spectral autour de 4kHz chez les acteurs, appelé « speaker's formant » n'existant pas chez les sujets témoins, et pouvant contribuer à leur efficacité (Raphael et al. 1987 ^[282] et Acker 1987 ^[4]).

2.2. Un forçage observé à différents niveaux et par différents champs disciplinaires

Le niveau de définition du forçage vocal n'est pas nécessairement identique à son niveau de description. Le forçage vocal est ainsi par essence un objet pluridisciplinaire puisque les indicateurs qui permettent de le caractériser ne sont pas toujours du même champ disciplinaire que les modèles et les théories qui cherchent à les interpréter (cf. Figure 3). Ainsi, des indicateurs perceptifs peuvent être utilisés pour caractériser un dysfonctionnement biomécanique du larynx quand au contraire des indicateurs physiologiques peuvent être utilisés pour caractériser le comportement de gestion de l'effort vocal par des locuteurs présentant une certaine personnalité. Cette pluridisciplinarité du forçage vocal en fait toute sa richesse et son intérêt scientifique, mais peut également constituer un danger si on l'étudie sans acquérir les connaissances des autres champs disciplinaires que le sien, ou sans s'entourer d'experts de ces autres domaines.

Dans ce paragraphe, nous allons chercher à bien distinguer les niveaux de définition et de description du forçage dans la littérature, en reprenant chacun des niveaux de définition exposés dans les paragraphes précédents et en recensant pour chacun d'entre eux les descripteurs et les disciplines associées ayant cherché à caractériser le forçage vocal.

Ainsi, quel que soit le niveau auquel le forçage est envisagé, il se définit dans tous les cas par son aspect pathologique et par l'apparition de symptômes au niveau du larynx. Ces symptômes peuvent être caractérisés de façon directe par une observation clinique (irritation, présence de lésions, etc.) ou par la variation d'indicateurs biomécaniques, tels que la température, la viscosité, ou le mode de vibration des cordes vocales. Ils peuvent également être caractérisés de façon indirecte par la variation d'indicateurs aérodynamiques (débit d'air, seuil de phonation, temps de phonation, etc.), d'indicateurs acoustiques (pente spectrale, Jitter, rapport signal sur bruit, etc.), d'indicateurs perceptifs équivalents (perception d'une voix criarde, instable, soufflée, etc.) ou encore d'indicateurs psychologiques (perception de gêne, de douleur, de fatigue, capacités de concentration, de contrôle, etc.)

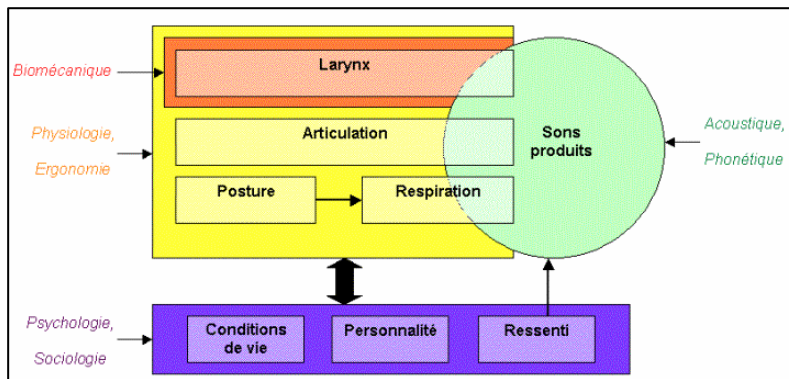


Figure 3. Différents niveaux de description du forçage vocal associés à différents champs disciplinaires

Lorsque le forçage vocal est considéré comme un dysfonctionnement du larynx (cf. § 2.1.1), il peut être caractérisé de manière directe par des indicateurs biomécaniques/physiologiques tels que la tension, la raideur, la température ou la viscosité des cordes vocales, les caractéristiques de leur vibration (force de collision, quotient ouvert, quotient de vitesse, amplitude de vibration, etc.) ou encore l'activité musculaire des muscles laryngés (par EMG par exemple). Des indicateurs aérodynamiques tels que le débit d'air, le temps de phonation maximum ou la résistance glottique peuvent également apporter des informations indirectes sur ce fonctionnement du larynx, par exemple sur la constriction glottique, les phases ouvertes de la glotte ou le mécanisme laryngé utilisé. Des indicateurs acoustiques et perceptifs peuvent également être reliés à des caractéristiques du fonctionnement laryngé. Ainsi, l'intensité est liée à la force de collision des cordes vocales, la fréquence fondamentale à leur nombre de collisions par unité de temps, le Jitter à l'instabilité de la vibration, la pente spectrale à la tension vocale, ou encore le rapport signal sur bruit à l'accolement incomplet des cordes vocales. Enfin, le sujet peut être questionné sur sa sensation de gêne, de serrage ou de douleur dans la gorge, de crispation dans le cou, ou sur sa sensation de contrôle de sa phonation.

Lorsque le forçage vocal est conçu comme un geste inadapté au niveau postural respiratoire ou articulaire (cf. § 2.1.2), il peut être de la même manière caractérisé de façon directe

- par des indicateurs physiologiques posturaux, tels que le centre de gravité du corps, les mouvements du buste ou du menton, l'activité de différents muscles liés à la posture
- par des indicateurs visibles de la respiration (thoracique, abdominale, etc.) ou aérodynamiques (Pression sous-glottique, débit d'air, temps de phonation maximum, résistance glottique, rendement vocal)
- par des indicateurs articulaires (mouvements de la langue, lèvres, mâchoire, déplacements verticaux du larynx)

Des indices acoustiques et perceptifs peuvent également caractériser ces gestes de façon indirecte. Une inspiration sonore est par exemple l'indicateur d'une respiration thoracique. L'intensité et l'attaque « en coup de glotte » des sons sont liées à la pression sous-glottique, l'irrégularité de l'intensité ou la perception d'essoufflement du sujet renseignent également sur sa gestion du débit d'air. Enfin, les formants vocaliques sont reliés à la configuration du conduit vocal et à la position des articulateurs.

Enfin, le sujet peut être questionné sur sa sensation de gêne, de douleur ou de crispation dans la mâchoire, la nuque et les épaules, sur sa sensation de fatigue générale.

Pour finir, lorsque le forçage vocal est considéré comme un comportement de gestion de l'effort lié au vécu du sujet ou à sa personnalité (cf. § 2.1.4), une première possibilité peut être de caractériser ce vécu ou cette personnalité à l'aide de questionnaires psychologiques ou sociologiques (tests de personnalité, tests socio-émotionnels, etc.). Une autre possibilité consiste à caractériser expérimentalement sa façon de réagir ou de s'adapter à différentes conditions de gestions de l'effort.

Certains des indicateurs acoustiques, glottiques et articulaires évoqués dans les paragraphes précédents ont été utilisés dans cette thèse. Ils sont définis plus précisément en Annexe Met1.

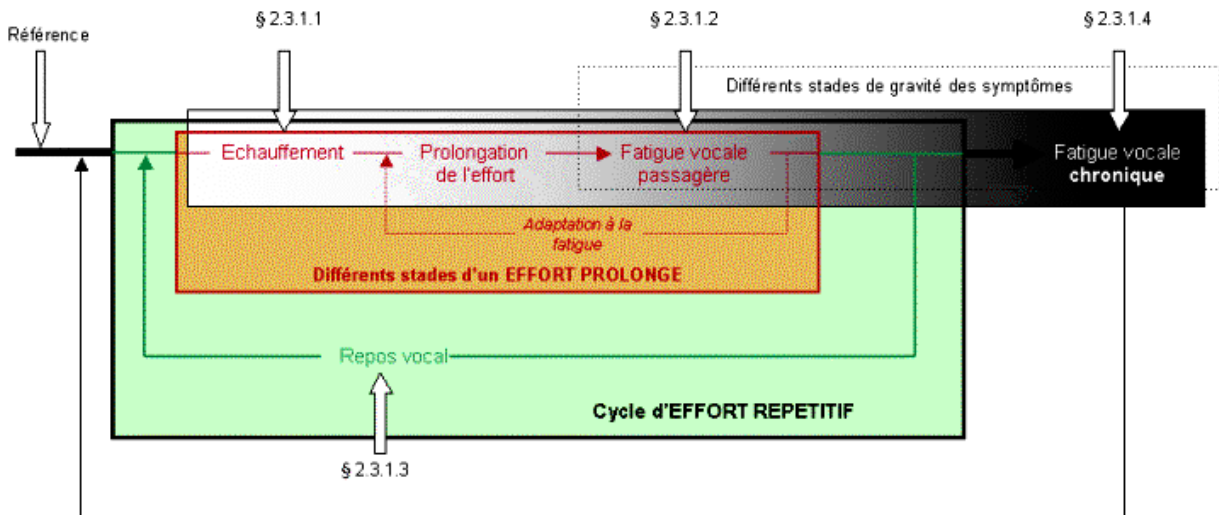


Figure 5. Représentation du cycle de forçage vocal comme imbrication des cycles d'effort prolongé et d'effort répétitif. Les différents stades d'un effort prolongé forment un continuum depuis l'échauffement vocal à la fatigue vocale, en intégrant l'adaptation à la fatigue. Le cycle d'efforts répétitifs consiste en une alternance de ces efforts prolongés avec des phases de repos vocal. La non-récupération de la fatigue vocale passagère peut entraîner un passage à une fatigue vocale chronique.

2.3.1.1. Conséquences d'un échauffement vocal

Les phénomènes physiologiques mis en jeu dans l'échauffement vocal ne sont pas encore totalement connus. Certains auteurs ont avancé que l'augmentation de la température des muscles serait accompagnée d'une diminution de la viscosité des cordes vocales, augmentant par conséquent la mobilité des cordes vocales (Safran et al. 1989 ^[298]). D'autres se sont également intéressés aux changements intervenant au niveau de la circulation sanguine ou des gestes respiratoires (Titze 1984 ^[350])

La majorité des études relatent que le locuteur se sent plus alerte et que sa voix est plus stable après un échauffement vocal. Cela est tout à fait logique puisque la fin de l'étape d'échauffement vocal est définie dans la plupart des études par un ressenti positif de la part des locuteurs ou chanteurs. Par contre, la caractérisation acoustique ou glottique des effets de l'échauffement vocal n'est pas toujours positive (Vintturi 2001 ^[368]). Au niveau glottique, Laukkanen et al. 1996 ^[199] observent que l'activité musculaire laryngée décroît au bout de quelques minutes d'exercices vocaux simples, la résistance glottique diminue, accompagnant une augmentation du débit glottique (Laukkanen et al. 1996 ^[199]). Elliot et al. 1995 ^[82] constatent également que le seuil de phonation (PTP) varie significativement avec l'échauffement vocal, mais de façon différente selon les individus. Laukkanen et al. 1996 ^[199] observent également que la pente spectrale devient moins raide, le rapport signal sur bruit augmente, en rapport avec la perception d'une voix plus solide, moins dure. Enfin, Neils et al. 1987 ^[252] ; Rantala et al. 2002 ^[281] ; Laukkanen et al. 1996 ^[199] ne constatent pas d'effet de l'échauffement vocal sur l'intensité et la F0, quand au contraire Sihvo et al. 1996 ^[317] observent une augmentation de l'étendue vocale.

La limite principale de ces caractérisations réside dans la définition problématique de la fin de l'échauffement vocal. Un temps fixe valable pour tous les locuteurs serait aberrant du fait des différences interindividuelles de résistance vocale. Mais un degré de ressenti subjectif par le locuteur nous semble également problématique du fait que l'estimation de la gêne ou de la douleur est très variable selon les personnes, et que le forçage vocal consiste peut être justement en une sous-estimation de la douleur ou de la fatigue par les locuteurs.

2.3.1.2. Conséquences d'un effort important ou prolongé ponctuel

L'évaluation subjective de la fatigue vocale par les locuteurs eux-mêmes a fait l'objet de plusieurs études (Buekers 1998 ^[38] ; Gotaas et al. 1993 ^[117] ; Rantala et al. 1999 ^[280] ; Kitch et al. 1994 ^[181] ; Kostyk et al. 1998 ^[186]). Ces symptômes sont de trois ordres :

- tout d'abord, il s'agit de douleurs ou de gênes : irritation, sécheresse, picotements, brûlures ou sensation de serrage au niveau de la gorge, douleurs dans le dos, la nuque et les épaules, maux de tête, douleurs au niveau de la mâchoire, aphonie ou voix qui s'enroue.
- les symptômes de la fatigue concernent également une perte de contrôle de la voix : difficulté à maintenir l'intensité, manque de contrôle du souffle (Koufman et al. 1988 ^[187]), difficultés à maintenir une F0 basse (Stemple et al. 1995 ^[329]) et à maintenir une hauteur stable tout court (Kitch et al. 1994 ^[181] ; Smith et al. ^[320])
- enfin, la fatigue peut être générale, s'accompagnant d'une chute du moral, d'ennui ou de problèmes de concentration.

L'évaluation perceptive de la fatigue vocale par l'auditeur est également très importante à des fins de diagnostic. Ainsi, Stemple et al. 1995 ^[329] ont rapporté un manque de modulation de la voix en situation de fatigue, et introduisent l'idée de « monopitch » pour décrire une intonation qui devient monocorde. Au niveau du timbre, la voix peut devenir soufflée (Stemple et al. 1995 ^[329]), rauque, tendue, dure ou criarde, et instable. L'échelle GRBAS d'évaluation des voix pathologiques est controversée mais beaucoup utilisée pour caractériser la fatigue vocale d'un point de vue perceptif (Hirano 1981 ^[147] ; Dejonckere et al. 1993 ^[64]). Cette échelle comprend 5 paramètres et 4 niveaux de graduation (de 0, pour une voix « normale » à 3, pour une altération sévère). L'échelle G (pour « Grade ») permet d'évaluer l'impression générale d'anormalité de la voix. L'échelle R (pour « Rough »), évalue la raucité et l'instabilité de la voix en fréquence et amplitude. L'échelle B (pour « Breathy »), évalue le caractère soufflé de la voix, ses fuites d'air. L'échelle A (pour « Asthenic »), concerne le manque de puissance ou d'harmoniques élevées de la voix. Enfin, l'échelle S (pour « Strain ») évalue le caractère forcé et hyperfonctionnel de la voix.

Au niveau acoustique, ces modifications de qualité vocale se traduisent par une augmentation de l'intensité vocale (Gelfer et al. 1991 ^[111] ; Vilkmán et al. 1999 ^[367] ; Buekers 1998 ^[38]) ainsi que de la fréquence fondamentale moyenne au bout d'une heure de phonation en moyenne (Rantala et al. 1999 ^[280] ; Gelfer et al. 1991 ^[111] ; Stemple et al. 1995 ^[329]). Le seuil de phonation (PTP) augmente également, mais seulement après au moins 2h de phonation (Maclagan et al. 2004 ^[227]). L'ambitus de F0 et la dynamique d'intensité diminuent pour la plupart des chanteurs, mais s'étend pour quelques-uns (Kitch et al. 1996 ^[182]). Enfin, le jitter et le shimmer évoluent de façon différente selon les études (Rantala et al. 1999 ^[280] ; Kitch et al. 1996 ^[182] ; Stemple et al. 1995 ^[329] ; Mann et al. 1999 ^[232] ; Gelfer et al. 1991 ^[111]). La phonation devient globalement plus pressée, allant vers l'hyperfonction (Lauri et al. 1997 ^[201] ; Vilkmán et al. 1997 ^[366] ; Vilkmán et al. 1999 ^[367]). Ainsi, la fermeture des cordes vocales devient plus abrupte (Vilkmán et al. 1999 ^[367] ; Vintturi 2001 ^[368]), et l'énergie des harmoniques de la voix augmente (Rantala et al. 1998 ^[279] ; Kitch et al. 1996 ^[182]). A noter que certaines études ne trouvent pas de changements significatifs après un usage prolongé (Verstraete et al. 1993 ^[364] ; Neils et al. 1987 ^[252] ; Novak et al. 1991 ^[257] ; Buekers 1998 ^[38])

Au niveau physiologique, la façon dont vibrent les cordes vocales se modifie au bout de seulement 15 min d'effort vocal (Linville 1995 ^[222]), mais avec une très grande variabilité interindividuelle. Stemple et al. 1995 ^[329] observent également par vidéostroboscopie l'apparition de fentes glottiques antérieures après une tâche d'effort chez 6 sujets sur 10. L'observation clinique montre souvent une irritation de la muqueuse laryngée, ainsi qu'une élévation de la viscosité et de la raideur des cordes vocales, pouvant être accompagnée d'œdèmes.

La comparaison de ces différentes études est assez difficile du fait que les tâches d'effort imposées aux sujets sont très variables (cf. Tableau1 de l'annexe Bib1). Les différents auteurs semblent considérer toute tâche d'effort comme équivalente, du moment où elle pousse le système à générer des symptômes. Cela nous semble poser plusieurs problèmes :

- tout d'abord cela semble sous-entendre que les symptômes sont induits par L'effort ou LE forçage en général, et ne sont pas spécifiques d'UN effort ou d'UN forçage en particulier. Pourtant, le manque de résultats significatifs de certaines études et la grande variabilité des résultats d'une étude à l'autre ont motivé certaines études à comparer l'influence de la tâche d'effort vocal sur l'importance et la rapidité d'apparition des symptômes (cf. § 2.3.3.). A notre connaissance, aucune étude n'a exploré s'il existait une différence de nature des symptômes selon la tâche d'effort effectuée. Pourtant, le fait de parler fort ou aigu ne fait pas subir les mêmes tensions ou traumatismes aux cordes vocales. Il ne nous semble donc pas évident que toutes les tâches d'effort conduisent aux mêmes conséquences pour la voix.

- par ailleurs, cela semble considérer la fatigue vocale comme un état « stable » une fois qu'il est atteint ou du moins comme un état qui évolue dans le prolongement de l'échauffement et de l'effort. Pourtant, les conséquences générées par le système « en forçage » ne lui sont pas extérieures mais l'affectent lui-même. Aussi, il faudrait tenir compte d'une rétroaction des symptômes sur le système, influençant l'aggravation de ses symptômes, ou l'apparition de nouveaux symptômes (cf. Figure 4). Et dans une conception du forçage où le locuteur ne se contente pas que de subir l'effort, on peut penser, comme Stemple et al. 1995^[329] le suggèrent, que les locuteurs s'adaptent à la fatigue (cf. Figure 5) en effectuant des ajustements vocaux subtils compensatoires qui pourraient réduire ou au contraire augmenter l'aggravation des symptômes. Cette idée de surenchère ou de manque d'attention aux signes de fatigue est également souvent évoquée dans la littérature ou dans les définitions du malmenage vocal de notre questionnaire (cf. § 1.2.4). Pour autant, elle n'a à notre connaissance jamais été caractérisée expérimentalement ni prise en compte dans les protocoles d'études de la fatigue vocale. Pour revenir à notre propos, ces réflexions impliquent que lorsqu'on observe le comportement vocal à un unique stade de la fatigue vocale, il est possible que les symptômes que l'on croit caractériser du forçage vocal correspondent en fait à un comportement compensatoire d'adaptation à ces symptômes.

2.3.1.3. Conséquences d'un repos vocal après un effort important ou prolongé

Les études ayant étudié les effets d'une pause vocale (Rantala et al. 1999^[280] ; Neils et al. 1987^[252] ; Viikman et al. 1997^[366] ; Lauri et al. 1997^[201] ; Sherman et al. 1962^[313]) montrent des résultats divergents, mais aussi que des groupes d'individus réagissent de façon contraire, certains montrant une amélioration de la voix après une pause vocale quand d'autres au contraire ressentent une aggravation de la gêne.

2.3.1.4. Conséquences d'un effort important ou prolongé répété

L'évaluation de la sévérité d'une dysphonie a motivé la création de plusieurs questionnaires d'évaluation subjective par le locuteur de ses symptômes. Le « Voice handicap index » (ou VHI) a été introduit par Jacobson et al. 1997^[165]. Il consiste en une trentaine de questions concernant le ressenti du locuteur, ses symptômes et ses habitudes de vie. Les dix questions ayant montré le plus de corrélation avec une description acoustique et clinique de la voix ont été retenues pour former le VHI. Le Voice-Related Quality of Life (V-RQOL) consiste de la même façon en une dizaine de questions physiques et socio-émotionnelles (Hogikyan et al. 1999^[150]). Rantala et al. 1999^[280] a également mis au point un questionnaire sur la gêne ressentie par les locuteurs et leur fréquence.

Au niveau de la qualité vocale, la perception d'une fatigue chronique peut être comparable à celle d'une fatigue passagère (raucité, dureté, souffle, instabilité...). Les voix dysphoniques sont également évaluées à l'aide de l'échelle GRBAS (Hirano 1981^[147]).

Au niveau acoustique, Rantala et al. 1999^[280] ont réussi à corréliser la sévérité des symptômes ressentis par les locuteurs à l'augmentation de la F0 moyenne, de l'intensité, du Jitter et du Shimmer, ainsi qu'à la réduction de l'écart type en fréquence et en intensité. Eustace et al. 1996^[85] ont également montré sur des patients un débit d'air élevé et un temps de phonation maximum réduit. Aronsson et al. 2005^[12] ont aussi observé une pression sous-glottique plus importante chez les sujets dysphoniques. Enfin Eustace et al. 1996^[85] ont mis en évidence des modes de vibration laryngée particuliers chez les sujets dysphoniques, présentant une fente glottique antérieure et/ou postérieure, ou encore en « fermeture-éclair ». Morrison 1997^[246] définit également plusieurs modes vibration des cordes vocales dans son essai de classification des dysphonies. Enfin, Mann et al. 1999^[232] observent également par vidéostroboscopie une augmentation des irrégularités sur le bord des cordes vocales après 5 jours d'efforts excessifs, des oedèmes et une réduction significative de l'amplitude de vibration.

La dysphonie se concrétise malheureusement bien souvent par des lésions organiques observables lors d'un bilan clinique (Giovanni 2006^[113]). Les principales lésions observées sont les nodules. Il s'agit d'un épaississement de la muqueuse semblable à la corne que pourrait présenter un ouvrier là où il tient son outil (Ormezzano 2000^[262]), détecté chez 9 à 13 % des enseignants. Viennent ensuite les polypes, correspondant à une lésion de la corde vocale en

dessous de la muqueuse. Elle survient chez des locuteurs qui utilisent le 1^{er} mécanisme laryngé avec une pression d'accolement importante.

2.3.2. Caractérisation des facteurs du forçage vocal

D'autres études se sont portées sur le fait que les conséquences du système face à un effort n'étaient pas toujours identiques. Le recensement des facteurs de forçage ou de prévention du forçage a été principalement réalisé par le biais de questionnaires, ainsi que par les observations empiriques des phoniâtres et orthophonistes. Quelques études ont commencé à vérifier la validité de l'effet de ces facteurs. A effort vocal constant en entrée du système, elles ont cherché à montrer que certains facteurs pouvaient augmenter ou réduire l'importance des symptômes ou de leur rapidité d'apparition. Ou bien au contraire, elles ont cherché à montrer qu'un effort vocal en entrée plus ou moins important était nécessaire pour provoquer le même degré de symptômes en sortie.

Au niveau environnemental et contextuel, l'exposition à la poussière, aux vapeurs irritantes, ou à une ambiance sèche est considérée comme un facteur de forçage vocal (Sihvo et al. 1996^[317] ; Viikman et al. 1997^[366] ; Viikman et al. 1999^[367] ; Hemler et al. 1997^[138] ; Vintturi 2001^[368]). Une mauvaise posture peut également entraîner un forçage (Hoit 1995^[151] ; Viikman et al. 1997^[366], Viikman et al. 1999^[367], Vintturi 2001^[368]). Les situations de trac (Kitch et al. 1996^[182]) et de conflit durables sont recensées parmi les facteurs potentiels du forçage vocal. Enfin, la présence d'un dysphonique dans l'entourage pourrait induire un phénomène de mimétisme.

Au niveau individuel, les femmes sont beaucoup plus sujettes au forçage vocal que les hommes (Sodersten et al. 2002^[321] ; Russell et al. 1998^[297] ; Vintturi 2001^[368] ; Novak et al. 1991^[257] ; Smith et al.^[320]), de même que des personnalités extraverties (Roy et al. 2000^[295] ; Giovanni 2007^[114]). Le travail vocal régulier est un facteur prévenant du forçage et augmente l'endurance des locuteurs lorsqu'un usage excessif est nécessaire (Gelfer et al. 1991^[111] ; Laukkanen et al. 2004^[197], Stemple et al. 1994^[328] ; Laukkanen et al. 1996^[199]). De même, l'échauffement vocal avant une performance a montré un effet positif sur l'état de la voix après l'effort (Stemple et al. 1994^[328]). Mais la majorité n'a pas pu montrer la significativité de ce facteur (Milbrath et al. 2003^[239] ; Donohue 2005^[75]), voire au contraire un effet négatif (Motel et al. 2003^[247]). Enfin, le tabagisme, la forte consommation d'alcool ou de café, une affection de la sphère ORL ou des modifications hormonales peuvent prédisposer à forcer sur la voix, quand au contraire une hydratation régulière est préconisée aux orateurs pour augmenter leur endurance (Verdolini et al. 1994^[363] et Chan et al. 2003^[48]). Ce traitement ne s'est cependant pas avéré significatif dans d'autres études (Donohue 2005^[75] et Solomon et al. 2000^[323]).

2.3.3. Influence du type d'effort sur les symptômes qu'il provoque

Certaines études ont cherché à comparer différentes tâches d'effort du point de vue de l'importance des symptômes qu'elles provoquent ou de leur rapidité d'apparition. Ainsi, Stone et al. 1973^[332] ont montré qu'une tâche de phonation avec élévation de la hauteur de la voix amenait une fatigue plus rapide qu'une tâche de phonation avec un abaissement de la hauteur de la voix. Holbrook 1974, cité dans Kitch et al. 1994^[181], a montré que le fait de parler fort induisait une fatigue vocale plus importante que le fait de parler pendant un temps très long à un niveau conversationnel, au contraire de Sander et al. 1983^[299] qui affirment que la parole forte n'accélère pas la fatigue. Buekers 1998^[38], après n'avoir observé aucune différence significative entre des sujets sains et dysphoniques au bout de 30min de parole prolongée, en conclut que les tests de parole excessive ou non naturelle sur de courtes périodes ne permettent pas de simuler et d'étudier le phénomène de forçage vocal vécu par les personnes devant utiliser leur voix de façon prolongée, dans la réalité.

2.3.4. Caractérisation du fonctionnement du système en situation d'effort

D'autres études se sont enfin intéressées à caractériser le fonctionnement du système pendant différentes tâches d'effort, en particulier pendant une utilisation de la voix à forte intensité, ou à une hauteur plus aiguë ou plus grave que la parole « confortable » habituelle du locuteur.

On peut cependant distinguer deux objectifs différents recherchés au travers de cette caractérisation. Certaines études ont plutôt cherché à caractériser le fonctionnement aérodynamique et vibratoire du larynx en situation d'effort, dans le but de comprendre en quoi ce fonctionnement peut générer des symptômes. D'autres études se sont plutôt penchées sur les caractéristiques gestuelles, posturales ou acoustiques du comportement d'effort, de façon à

déterminer des indicateurs objectifs de reconnaissance d'un comportement de forçage pouvant être utilisés par des phoniâtres et des orthophonistes lors de leurs consultations.

2.3.4.1. En quoi ce fonctionnement peut-il provoquer des symptômes ?

Sonninen et al. 1998^[325] ont ainsi pu montrer que l'évolution de la tension des cordes vocales n'était pas linéaire en fonction de la hauteur de la voix mais suivant la courbe de tension des tissus organiques. Sonninen et al. 1998^[325] n'évoquent pas la notion de mécanisme laryngé mais on sait maintenant que la tension des cordes vocales augmente avec la hauteur de la voix en mécanisme 1 jusqu'au passage en mécanisme 2 où la fréquence fondamentale continue de croître du fait d'ajustements différents, davantage liés à l'étirement des cordes vocales (Henrich 2001^[139] ; Roubeau et al. 1993^[293]). La fréquence fondamentale de la voix est également liée au nombre de collisions des cordes vocales par unité de temps. Par conséquent, une fréquence fondamentale plus élevée fait subir aux cordes vocales davantage de microtraumatismes.

L'augmentation de l'intensité vocale est principalement la résultante de l'augmentation de la pression sous-glottique (Gauffin et al. 1989^[109]) et/ou de l'augmentation de l'accolement des cordes vocales (Isshiki 1964^[163]). Une forte pression des aryténoïdes caractérise une phonation pressée, entraînant une constriction de la glotte et la diminution du débit glottique. Le quotient d'amplitude est typiquement faible pour ce type de phonation (Alku et al. 1996^[5]). Au contraire, les cordes vocales s'accolement alors plus longuement et avec plus de force, ce qui se traduit par la diminution du quotient ouvert (Hanson 1997^[135], Henrich 2001^[139] l'augmentation de la vitesse d'accolement (Gauffin et Sundberg 89 ; Vintturi 2001^[368]) et du quotient de vitesse (Sundberg et al. 1993^[338]).

Plusieurs hypothèses physiologiques ont été recensées pour tenter d'expliquer comment ces modes de phonation pourraient entraîner une fatigue vocale et des lésions anatomiques (d'après Titze 1984^[350]):

- la contraction importante des muscles laryngés pourrait induire une fatigue neuromusculaire et une réduction de leur capacité à résister à une stimulation répétée. Cette hypothèse est controversée. En effet, les muscles laryngés ont une structure spécifique différente des autres muscles, adaptée à un usage continu (Titze 1984^[350]) et des recherches sur des larynx excisés ont ainsi montré que les muscles laryngés étaient résistants à la fatigue (Cooper et al. 1990^[58]). Il serait donc discutable de penser que la fatigue vocale puisse être musculaire, de même que l'échauffement vocal puisse être bénéfique à un niveau strictement neuromusculaire. Cependant, d'autres études au contraire ont observé des cordes vocales arquées après un effort très violent (Lira 1971 dans Vilkmann 2004^[365]), ce qui pourrait expliquer une moins bonne adduction des cordes vocales et l'apparition de fuites d'air glottiques avec la fatigue vocale.
- le maintien d'une pression sous-glottique importante pourrait finir par épuiser les muscles respiratoires, entraînant une moins bonne gestion du souffle et contraignant le larynx à jouer à la fois le rôle de vibreur et de régulateur du débit d'air.
- la collision importante et répétée des cordes vocales pourrait induire un stress des tissus ligamentaires et des cartilages du larynx, et ainsi provoquer une inflammation des tissus, voire des lésions telles que des nodules ou des polypes. Elle pourrait également induire une augmentation de la viscosité et de la raideur des cordes vocales, entraînant des forces de friction accrues et une irritation des tissus. Cela pourrait expliquer que l'hydratation régulière lors d'un effort vocal prolongé ait un effet bénéfique, puisqu'elle joue un rôle important de régulation de la viscosité des cordes vocales.
- enfin, une crispation laryngée ou une trop grande tension pourrait réduire la circulation sanguine dans les muscles laryngés. Cela gênerait à la fois l'évacuation de l'acide lactique et l'approvisionnement des muscles en oxygène et en glucose, mais aussi la dissipation de la chaleur hors des cordes vocales. Par conséquent, cela pourrait expliquer les oedèmes formés après des efforts vocaux importants, ainsi que des lésions des tissus induits par une augmentation de la température.

2.3.4.2. Quels sont les manifestations observables concomitantes du forçage vocal ?

Au niveau postural, les personnes qui forcent sur leur voix montrent des crispations importantes au niveau des muscles sterno-cléido-mastoïdiens, voire de l'ensemble du corps (Papon 2006^[265]). Morrison 1997^[246] dans son listage des « Muscle Misuse Voice disorders (MMVD) » rapporte également des tensions et une restriction de la liberté de mouvement au niveau de la mâchoire, avec une éventuelle pression de la langue contre les dents. Cette crispation générale peut alors s'accompagner d'une turgescence des veines jugulaires externes et d'une augmentation de l'écart type de la vitesse des mouvements du centre de gravité du corps, due à son raidissement (La Machhia 2005^[189]). Grini et al. 1998^[124], Morrison 1997^[246] et Papon 2006^[265] ont également montré que le forçage vocal s'accompagnait d'un déplacement global du centre de gravité du corps vers l'avant, ainsi que d'une avancée locale du buste et du menton. Cette perte de verticalité et d'alignement postural, avec flexion du rachis cervico-thoracique, arrondissement du dos et déplacement du menton vers l'avant pour maintenir un contact visuel avec l'interlocuteur a déjà été décrit par Le Huche 1984^[202] comme caractéristique de la voix de détresse ou d'insistance.

Hoit 1995^[151] a souligné l'influence de la posture du locuteur sur sa phonation, en particulier au niveau de la gestion respiratoire. Le Huche 1984^[202] et Faure 1988^[90] ont également observé que les sujets forçant sur leur voix n'augmentent pas leur intensité vocale par la mise en place d'une respiration abdominale, mais par une respiration thoracique supérieure (élévation des épaules et du thorax) ou une respiration que Le Huche 1984^[202] qualifie de « vertébrale », s'accompagnant d'une flexion vertébrale et d'un affaissement du thorax. Le passage à ce mode respiratoire se manifeste chez le locuteur par des respirations rapides et buccales, et un temps de phonation réduit pour chaque inspiration, contraignant le locuteur à accroître la tension laryngée pour prolonger la vibration des cordes vocales. Par ailleurs, ce mode respiratoire sans soutien abdominal et diaphragmatique fait alors jouer au larynx le double rôle de vibreur et de régulateur du débit d'air expiré.

Les attaques du son se font alors en « coup de glotte », avec une très forte augmentation de la pression sous-glottique en début de phonation (Papon 2006^[265]). Auditivement, la voix des « forceurs » donne l'impression de manque d'efficacité, plus soufflée, rauque ou criarde que la voix de personnes « saines » (Revis 2004^[283]). Les patients témoignent d'ailleurs souvent d'une sensation de serrage laryngé ou d'oppression respiratoire lors de la phonation. Ce serrage s'accompagne d'une atténuation de l'énergie du fondamental par rapport aux harmoniques aiguës correspondant à un changement de pente spectrale (Klatt et al. 1990^[183] ; Holmberg et al. 1995^[154] ; Gauffin et al. 1989^[109] ; Childers et al. 1991^[50] ; Sluijter 1996^[319]).

Lienard et al. 1999^[215] et Rostolland 1982^[289] ont de la même manière montré une forte corrélation entre l'augmentation de l'effort vocal et l'augmentation de F0 et F1, tandis que F2 et F3 n'évoluent pas significativement. Ces résultats formantiques concordent avec l'augmentation de l'ouverture de la bouche observée en situation d'effort vocal par Schulman 1989^[304]. Traunmüller 1981^[355] a confirmé que cette augmentation de F0 et F1 était suffisante d'un point de vue perceptif pour percevoir l'effort dans une voix.

3. POSITIONNEMENT ET HYPOTHESES DE NOTRE TRAVAIL

Le paysage conceptuel et méthodologique d'étude du forçage vocal étant présenté, nous allons pouvoir positionner notre propre contribution par rapport à notre propre champ disciplinaire et à notre objectif qui est de comprendre les différences interindividuelles vis à vis du forçage vocal dans le but d'aider si possible les personnes y étant davantage sujettes que d'autres.

Nous avons vu que dans la littérature, ces différences interindividuelles sont quasiment toujours étudiées ou expliquées du point de vue de la constitution physiologique des individus et par conséquent de leur fragilité face à l'effort. Il existe pourtant une autre hypothèse complémentaire qui n'est, étonnamment, que très rarement explorée : l'idée que les individus ne soient pas passifs vis à vis du forçage mais y jouent un rôle actif déterminant par leur façon de gérer l'effort. Cela est d'autant plus étonnant que comme nous l'avons vu au cours du questionnaire, le forçage vocal ou le malmenage vocal sont souvent définis comme des « comportements inadaptés » ou des « mauvais gestes », « un effort mal géré ». Cette idée que le locuteur participe au forçage vocal semble donc assez partagée aussi bien dans le sens commun que par les phoniâtres et les orthophonistes. Certaines études ont caractérisé des différences acoustiques, respiratoires ou posturales entre des individus au cours de leur phonation (cf. § 2.3.4.2). Pour autant, ces caractéristiques sont la plupart du temps envisagées comme des manifestations concomitantes du forçage vocal pouvant servir à le diagnostiquer, et non comme des actions du locuteur inadaptées pour gérer l'effort vocal, le conduisant à forcer sur sa voix. Seules les études sur les types de personnalité cherchent à comprendre en quoi certains caractères peuvent avoir une influence sur la façon de réagir à l'effort, et par conséquent de favoriser ou de limiter le forçage vocal.

Nous tenterons donc dans cette thèse d'explorer les différences interindividuelles de gestion de l'effort pendant la phonation et d'identifier ces différentes stratégies de gestion possibles.

Dans le but de comprendre quelles sont les stratégies de gestion de l'effort associées au forçage vocal, et celles qui permettent de l'éviter, nous devons trouver un critère pour évaluer le caractère forcé ou non de chaque stratégie. Au regard des études antérieures, un premier critère qui s'impose est l'apparition ou non de conséquences pathologiques à la suite de l'utilisation de ces stratégies. Nous discuterons pourquoi cette solution ne nous semble pas totalement satisfaisante. Un autre critère pourrait provenir des caractéristiques des locuteurs observés, par exemple leur genre le prédisposant au forçage vocal, leur expertise de techniques vocales, ou leur dysphonie déjà installée. Nous argumenterons également les limites que nous semblent présenter une telle hypothèse. Un troisième critère nous convainc davantage : il s'agit de la notion d'efficacité. Mais qu'entend-on exactement par efficacité ?

- En premier lieu, lorsque nous parlons de « stratégie de gestion de l'effort », nous évoquons le fait qu'il puisse exister plusieurs configurations phonatoires possibles, intuitives ou apprises par la technique vocale, pour réaliser une même tâche vocale. Certaines de ces stratégies sont plus efficaces que d'autres vis à vis du larynx
 - du fait qu'elles dépendent davantage d'énergie pour un même résultat.
 - ou bien du fait que, toujours pour un même résultat produit, elles répartissent les efforts à différents étages de la chaîne de production du son plutôt que de concentrer la charge uniquement sur le larynx. Nous tenterons de recenser les différentes théories et techniques vocales existantes allant dans le sens d'une telle efficacité de répartition des efforts.
- Mais surtout, lorsqu'on cherche à expliciter ce « résultat produit », par rapport auquel se définit l'efficacité, on se rend compte que pour les personnes dont la voix est un outil de travail, la finalité de la production (et donc son résultat) ne réside pas tant dans les caractéristiques acoustiques de la voix (intensité, richesse spectrale, etc.) que dans la transmission d'un message à un interlocuteur. Une grande absente de la littérature sur le forçage vocal et de notre propos jusqu'à maintenant est la communication. Cela implique qu'en plus de ne pas être passif vis à vis de la gestion de l'effort, le locuteur n'est pas non plus passif vis à vis de l'effort lui-même. Nous avons jusqu'à maintenant représenté l'effort vocal comme une « entrée » du système de forçage, ce qui sous-entendait que cet effort était « imposé » au locuteur. En réalité, si les locuteurs acceptent de réaliser (et non de subir) un effort pour la phonation, c'est qu'ils y trouvent une motivation supérieure à la fatigue que cela occasionne : le désir de communiquer et d'être compris. L'effort vocal dépensé, puis géré, est donc toujours en rapport avec l'estimation de l'intelligibilité nécessaire à la situation de

communication. Nous faisons alors l'hypothèse dans cette étude que les différences interindividuelles face au forçage vocal puissent ne pas seulement être liées à différentes configurations vocales pour produire un même résultat acoustique mais puissent être aussi plus largement liées à différentes stratégies de communication pour produire un même niveau résultat perceptible en terme d'intelligibilité. Différentes stratégies peuvent en effet être adoptées pour se faire comprendre dans une situation. Certaines seront plus ou moins efficaces

- selon l'énergie dépensée pour une même intelligibilité produite.
- selon la mobilisation laryngée nécessaire à la réalisation de ces stratégies. L'intelligibilité peut par exemple être augmentée en parlant plus fort, ce qui nécessite une augmentation plus ou moins importante de la tension laryngée, selon la technique vocale mise en oeuvre pour cela. Tandis que l'augmentation de l'intelligibilité par une meilleure précision articulatoire ou par le ralentissement du débit de parole ne nécessite pas en soi d'augmentation de la tension laryngée (même si nous nous n'excluons pas qu'il puisse exister des tensions laryngées compensatoires à la réalisation de ces stratégies, ou que toutes ces stratégies soient réalisées simultanément).

Nous recenserons différentes théories et techniques vocales allant dans le sens d'une telle efficacité de communication.

3.1. Des stratégies individuelles de gestion de l'effort

La sensibilité de certains individus à développer des troubles vocaux est très souvent interprétée comme d'origine directement ou indirectement physiologique. Ainsi, les femmes seraient plus sujettes au forçage vocal que les hommes du fait de leur fréquence fondamentale plus haute et se situant dans la zone de passage des deux mécanismes laryngés. Les sujets entraînés seraient plus résistants à l'effort du fait de leur tonicité accrue du larynx. Les fumeurs seraient davantage exposés au forçage à cause de l'irritation que leur habitude provoque au niveau des muqueuses.

Les différences individuelles de constitution physiologique tout comme les facteurs environnementaux influençant l'état de la muqueuse laryngée jouent sans aucun doute un rôle dans le forçage vocal. Cependant, cette conception passive du locuteur subissant le forçage vocal ne nous semble pas expliquer totalement ce phénomène : nous faisons l'hypothèse que le comportement du sujet puisse également jouer un rôle prédominant. Et surtout, cette conception limite les possibilités de prévention ou de traitement des troubles de la voix au contrôle de facteurs extérieurs (acoustique de salles, humidité, poussières, sonorisation par un microphone) et à des règles d'hygiène vocale (hydratation, exercices de tonification du larynx, évitement du café ou du tabac, etc.). Nous pensons que la prévention du forçage vocal pourrait également résider dans l'éducation ou la rééducation du comportement de gestion de l'effort vocal de chaque individu.

Un effort vocal très long ou très important entraîne inmanquablement des conséquences pour la voix, chez n'importe quel individu, au bout d'un temps plus ou moins long. Personne ne peut fournir d'effort indéfiniment et cela est valable en particulier au niveau vocal. C'est pourquoi de très nombreux chercheurs définissent le forçage vocal par l'apparition de symptômes ou par le dépassement d'un seuil d'effort. Dans cette conception, tous les locuteurs forcent donc sur leur voix lorsqu'ils hurlent, par exemple. Cette vision des choses se défend tout à fait. Malheureusement, elle ne permet pas de prendre en compte l'activité du sujet dans le forçage ni d'expliquer les différences interindividuelles face aux troubles de la voix, excepté par des causes physiologiques. Elle ne permet pas non plus de prévenir le forçage vocal pour les personnes qui ne peuvent de toutes façons pas faire autrement dans leur profession que de parler longtemps ou à forte intensité. La seule prévention possible consiste alors à retarder l'apparition de symptômes engendrés par ce forçage, grâce au contrôle de facteurs extérieurs (sonorisation, hydratation, etc.).

C'est pourquoi dans ce travail, nous testerons l'hypothèse que le locuteur ne joue pas un rôle uniquement passif vis à vis du forçage mais peut également jouer un rôle actif, par la façon dont il gère l'effort et que cette gestion peut être elle-aussi déterminante, au même titre que les facteurs extérieurs, dans l'accélération ou le retardement de l'apparition de symptômes. Pour reprendre l'exemple précédent du hurlement, nous ne nous intéressons donc pas dans ce travail à caractériser le fonctionnement laryngé en général pour ce type de phonation, ni à comprendre en quoi ce fonctionnement peut générer des symptômes. Nous chercherons plutôt à identifier les différentes « techniques » utilisées par les individus pour hurler, et en quoi celles-ci sont plus ou moins efficaces, et surtout à comprendre pourquoi, dans une même situation, certains individus hurlent pour se faire comprendre tandis que d'autres parviennent au même résultat tout économisant leur voix. Autrement dit, nous n'envisagerons pas le forçage vocal comme l'équivalent d'un effort extrême, mais plutôt comme une « mauvaise gestion » individuelle de cet effort. Cette idée entraîne automatiquement le

fait que la cognition du locuteur (son ressenti, son expertise, ses représentations mentales, sa personnalité) soit impliquée dans le phénomène de forçage et pas uniquement son corps. Ainsi, La Figure 6 présente une première proposition de réadaptation de la modélisation classique du forçage vocal précédemment schématisée sur la Figure 4, en tenant compte cette fois des aspects de gestion de l'effort par le sujet et par son implication cognitive et pas uniquement corporelle dans le phénomène du forçage. (La Figure 9 en proposera une deuxième réadaptation supplémentaire de cette modélisation).

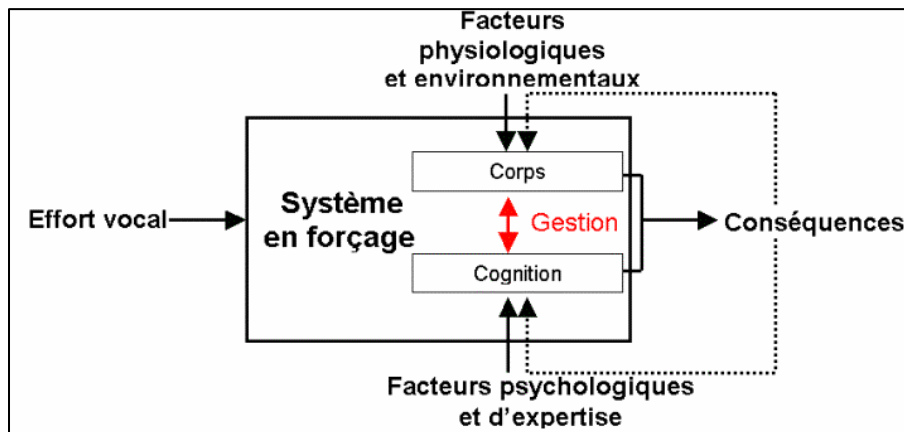


Figure 6. Première réadaptation de la modélisation du forçage vocal sous la forme d'une mauvaise gestion de l'effort vocal par le système-locuteur (corps et cognition), et plus seulement comme la soumission du corps du locuteur à un effort excessif.

Partant de ce principe, nous envisagerons tout au long de ce travail le forçage vocal sous la forme d'une « mauvaise » gestion de l'effort vocal par certains individus seulement et dans certaines circonstances. Nous ne pensons d'ailleurs pas qu'il existe UNE unique façon de bien gérer l'effort ni UNE unique façon de forcer sur sa voix, mais un panel de stratégies différentes que nous chercherons donc à caractériser puis à évaluer du point de vue de leur caractère forcé. Nous allons examiner dans les paragraphes suivants les différents critères que nous avons envisagé pour évaluer ce caractère forcé.

3.2. L'évaluation de ces stratégies de gestion à partir des symptômes qu'elles provoquent

Comment déterminer que la gestion de l'effort par une personne est « bonne » tandis que celle d'une autre est « mauvaise » ?

Les conceptions du forçage vocal dans la littérature nous conduisent assez spontanément à envisager l'aspect pathologique comme critère d'évaluation d'une « mauvaise » stratégie. Ainsi, une mauvaise gestion de l'effort vocal serait une gestion qui provoque des symptômes, plus de symptômes que d'autres, ou encore plus rapidement que d'autres.

Ce critère nous semble cependant problématique. En effet, les symptômes peuvent être dépendants à la fois de la constitution vocale d'un locuteur et de sa gestion de l'effort : Certains individus vont se fatiguer moins vite que d'autres, grâce à une constitution plus résistante du larynx, et indépendamment de leur comportement vocal. Il est par conséquent envisageable que certaines personnes puissent tenir plus longtemps avec un « mauvais » geste que d'autres avec un bon geste.

On remarque que l'on continue à parler de « bon » ou de « mauvais » geste même s'il ne provoque pas de symptômes. Il existe en effet un certain consensus sur la perception acoustique d'une voix forcée par exemple ou sur les signes visuels de crispation du locuteur, qui font unanimement penser que le locuteur n'adopte pas un geste sain, indépendamment de l'état clinique de son larynx. C'est donc que le caractère « bon » ou « mauvais » de la gestion de l'effort se situe à un autre niveau.

3.3. L'évaluation de ces stratégies de gestion à partir des individus qui les adoptent

La deuxième possibilité pour déterminer quelles stratégies de gestion correspondent à du forçage consiste à définir des groupes de locuteurs supposés être prédisposés ou sujets au forçage, et de comparer leur gestion de l'effort vocal à celle d'un autre groupe témoin, supposé « sain ».

Ainsi, plusieurs études ont comparé le comportement vocal des hommes et des femmes en situation d'effort vocal, afin d'essayer de comprendre pourquoi les femmes sont plus touchées que les hommes par le forçage vocal. Holmberg et al. 1995^[154] et Vilkmann et al. 1999^[367] ont observé que les hommes avaient une intensité vocale plus importante en voix criée, s'accompagnant d'une vitesse de fermeture plus élevée des cordes vocales et d'une amplitude de voisement plus importante. De même, Alku et al. 1996^[5] ont observé des quotients d'amplitude plus grands chez les hommes que chez les femmes. Et Lauri et al. 1997^[201] ont mesuré des quotients ouverts plus faibles chez les hommes en voix normale et forte que chez les femmes. Ces résultats seraient plutôt associés à des microtraumatismes plus importants subis par les cordes vocales des hommes que des femmes. La seule explication de la sensibilité accrue des femmes au forçage vocal semble, dans l'état des connaissances, liée à leur fréquence fondamentale plus élevée, entraînant un nombre de collisions plus important des cordes vocales, et par conséquent des micro-traumatismes plus répétés que chez les hommes (Rantala et al. 1999^[280] ; Titze 1984^[350]).

De même, des sujets « entraînés » (acteurs ou chanteurs selon les études) ont été supposés comme moins sujets au forçage que des sujets « non entraînés ». Brown et al. 2000^[37] n'ont cependant pas observé de différence significative entre des chanteurs et des non chanteurs au niveau de l'analyse acoustique de leur voix parlée et projetée. Seuls 5 chanteurs ont montré en voix parlée une variation plus importante de la fréquence fondamentale, des segments vocaliques plus longs et des phases d'occlusion plus courtes lors de la production de consonnes occlusives. Raphael et al. 1987^[282] et Acker 1987^[4] ont observé un renforcement spectral autour de 4kHz chez les acteurs lorsqu'ils utilisaient une voix projetée.

Enfin des patients dysphoniques, donc au forçage vocal avéré par leurs symptômes, ont été comparés à des sujets sains, et ont montré des différences posturales (arrondissement du dos et déplacement du menton vers l'avant, crispation générale) et respiratoires (respiration thoracique supérieure ou « vertébrale » au sens de Le Huche 1984^[202]). Des types de personnalité, en particulier une tendance à l'extraversion, ont également pu être associées au groupe des dysphoniques.

La classification de locuteurs a priori est intéressante pour tester des hypothèses, mais doit être prise avec précaution. Par exemple, il semble intuitif de penser que les acteurs ou les chanteurs se servent plus efficacement et sainement de leur voix, et donc à les considérer comme des locuteurs experts en comparaison de locuteurs témoins. Pourtant, il existe une proportion importante de chanteurs présentant des dysfonctionnements vocaux (Koufman et al. 1988^[187]). De même, dans leur enquête auprès de 10 acteurs et 10 chanteurs, Kitch et al. 1994^[181] relèvent que tous leurs sujets ont vécu des épisodes de fatigue vocale plus ou moins sévères. Enfin, il est connu que certains grands chanteurs placent très mal leur voix parlée. Ainsi, contre toute attente, Rothman et al. 2002^[291] ont observé que les chanteurs étaient plus sensibles que d'autres au changement d'acoustique de la salle en voix parlée, contrairement à leur plus grande maîtrise du retour kinesthésique observée en voix chantée (Scotto di Carlo 194^[311]). Brown et al. 2000^[37] montrent également que les chanteurs sont unanimement distingués perceptivement des non chanteurs sur des extraits chantés mais que cela n'est pas le cas sur des extraits parlés. Ainsi, il est important de rappeler que la fonction ne fait pas nécessairement l'expertise de gestion. Et surtout que l'expertise de gestion n'est valable que pour le mode de phonation dont le sujet est spécialiste (voix projetée, voix lyrique, voix dans le bruit, etc.) et non pas pour tous les modes de phonation. Ainsi, ce n'est pas parce qu'une personne est capable de chanter pendant des heures sur une scène d'opéra qu'elle saura communiquer de façon efficace dans une discothèque bruyante. Le travail de la voix chantée apporte certainement des atouts techniques (maîtrise de la respiration) ou physiologiques (tonification du larynx et augmentation de l'endurance vocale), mais pas nécessairement l'expertise nécessaire à la gestion de toutes les situations.

De même, il peut y avoir un problème à comparer le comportement vocal de dysphoniques avec des locuteurs sains. Tout d'abord, rien ne garantit que les locuteurs « sains » n'ont pas également un mauvais geste et qu'ils ne vont pas devenir également dysphoniques. Mais surtout, il y a un risque à caractériser comme du forçage vocal un comportement qui est en fait une adaptation compensatoire aux symptômes déclarés, et non le comportement de forçage lui-même qui les a générés.

3.4. L'évaluation de ces stratégies de gestion à partir du degré de répartition des efforts

Quel critère utiliser alors pour évaluer le caractère « bon » ou « mauvais » d'une stratégie de gestion de l'effort, si la présence de symptômes ou l'appartenance à une catégorie de locuteurs ne sont pas satisfaisantes ?

Nous proposons la notion d'efficacité, qui ressort des réponses à notre questionnaire comme un concept opposé de celui du forçage vocal. Un geste sans forçage serait alors celui qui atteint son but en consommant un minimum d'effort, en particulier pour la partie la plus fragile mais surtout la plus sollicitée de la chaîne de production : le larynx.

Hors un même but peut être atteint par différentes mobilisations de la source glottique, des articulateurs ou des muscles respiratoires. Ainsi, une même position articuloire peut être réalisée par différentes activations musculaires (Maeda et al. 1994 ^[229] ; Abbs et al. 1984 ^[3]), plusieurs positions des articuloires et formes du conduit vocal peuvent générer les mêmes signaux acoustiques (Atal et al. 1971 ^[13] ; Maeda 1990 ^[228], Boë et al. 1992 ^[29]). Enfin plusieurs modèles acoustiques peuvent être associés à la perception d'un même phonème (Perkell et al. 1986 ^[267]). Cette idée d'« équivalence moteur » implique l'existence d'une pluralité de stratégies pour réaliser une même intention dans le domaine physique. Par exemple, un son de même et forte intensité peut être produit en augmentant la pression sous-glottique, en augmentant l'accolement des cordes vocales, en renforçant davantage l'énergie spectrale de la voix dans la zone sensible de l'oreille ou en adoptant des stratégies résonnantes d'ajustement du premier formant sur un harmonique de la voix (formant tuning), etc. Ces différentes stratégies ne correspondent pas au même niveau d'effort laryngé selon que le larynx assume tout seul cette production ou qu'il en partage les efforts avec les autres étages de la production.

Cette première notion d'efficacité vocale du point de vue de la répartition des efforts s'est imposée à nous suite à plusieurs observations : d'un côté, le forçage vocal a été décrit par des modifications acoustiques, posturales, respiratoires ou articuloires (cf. paragraphe 3.1.1). De l'autre côté, les personnes forçant sur leur voix témoignent de douleurs principalement laryngées, et également dans la nuque ou les épaules. Mais aucune ne se plaint par exemple de douleurs à la langue ou à la mâchoire, ni de douleurs aux muscles abdominaux. Cela pourrait amener à conclure rapidement que le forçage se situe donc aux endroits où le patient a mal (en particulier le larynx), et qu'il est inutile de se pencher sur les autres aspects (la respiration, l'articulation, la posture). Si on envisage le forçage vocal d'un point de vue plutôt comportemental, et la phonation comme un système « modulaire » (la « soufflerie », le « vibreur », les résonateurs, les articulateurs,...) l'apparition de symptômes au niveau de certains modules ne signifie pas que le comportement de forçage se situe exclusivement à cet endroit. Il traduit plutôt le fait que ces parties du corps subissent l'effet du comportement de forçage, soit par leur fragilité comparée à d'autres parties, soit par le fait qu'un module supporte trop de charge à cause du désistement ou du dysfonctionnement des autres modules. Ainsi, on sait par exemple qu'un manque de soutien respiratoire a des conséquences sur l'efficacité et la santé du larynx (Le Huche 1984 ^[202] ; Faure 1988 ^[90]). Si des symptômes peuvent se manifester au niveau laryngé, le geste nocif à réduire se situe pourtant au niveau respiratoire. Nous faisons donc ici l'hypothèse que le comportement de forçage n'est pas obligatoirement une mauvaise gestion du larynx mais peut être davantage, à même résultat produit, une mauvaise répartition des efforts sur l'ensemble des organes contribuant à l'acte de parole, ce qui ferait porter trop d'effort sur le larynx. Au cours de nos lectures, nous avons recensé 2 notions : celle d'équilibre pneumo-phonique et celle d'impédance ramenée sur le larynx, faisant intervenir une répartition de l'effort vocal. Certaines techniques vocales de recherche d'efficacité ou de protection du larynx peuvent être expliquées par ces notions, et méritent notre attention.

3.4.1. La notion d'équilibre pneumophonique

La gestion du souffle et la notion de soutien (ou appoggio) sont les bases d'un travail de la technique vocale en voix parlée ou chantée. De nombreux auteurs et pédagogues ont souligné sa nécessité pour produire des notes aiguës ou maintenir un débit d'air régulier à l'expiration (Miller 1990 ^[243] ; Amy de la Bretèque 2003 ^[10]). Mais un bon équilibre entre le soutien respiratoire et l'effort laryngé est également recommandé pour éviter ou diminuer la fatigue engendrée par une phonation à forte intensité (Sander et al. 1983 ^[299]). Le soutien consiste globalement à maîtriser la régulation du débit d'air expiré et la pression sous-glottique par l'action antagoniste contrôlée des muscles inspireurs et expirateurs, afin de libérer le larynx de cette fonction de régulateur du débit d'air et de lui laisser seulement remplir le rôle de vibreur.

Au-delà de sa signification originelle, le soutien en est venu à désigner un idéal d'équilibre et de coordination entre la gestion du souffle et la résonance de la voix (Miller 1990 ^[243]). En effet, un « étage » du système phonatoire est mis de côté dans la conception initiale du soutien : il s'agit des résonateurs, dont l'action peut être contrôlée à l'aide des articulateurs (mâchoire, lèvres, langue, voile du palais,...), et a une influence importante non seulement sur le son rayonné mais

également sur l'équilibre de phonation. Si on représente donc la phonation sous forme d'un système à 3 étages : le souffle, le vibrateur, les résonateurs (cf. Figure 7), le larynx apparaît comme le maillon le plus sollicité de la chaîne. Pour garantir son bon état de fonctionnement, de nombreux auteurs recommandent un équilibre entre la pression sous-glottique, le timbre de la voix et la résistance du larynx (Amy de la Bretèque 2003^[10] ; Guerin 2004^[128]).

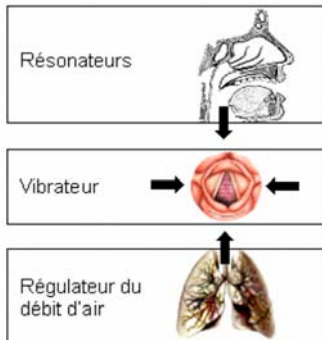


Figure 7. Schématisation de l'accord pneumophonique entre la pression sous-glottique, le timbre de la voix et la résistance du larynx. Inspiré de Guerin 2006^[129].

L'expérience de rééducation de Guerin 2006^[129] l'amène à répertorier plusieurs déséquilibres fréquents du fonctionnement vocal :



Le larynx peut compenser un déficit de pression sous-glottique en prenant en charge le rôle de régulateur du débit d'air en plus de celui de vibrateur, amenant crispation et fatigue à moyen terme.

Au contraire, le larynx va se contracter pour résister à un excès de pression sous-glottique, entraînant une fatigue et d'éventuelles fuites d'air. La dégradation du timbre peut alors amener au comportement suivant.

Le manque de timbre ou de résonance de la voix peut conduire le locuteur ou le chanteur à augmenter la pression sous-glottique pour se faire entendre, ramenant au cas précédent de dysfonctionnement vocal.

Cette influence du timbre de la voix sur l'équilibre phonatoire a déjà fait l'objet d'études et peut être expliquée par la notion d'impédance ramenée sur le larynx (Husson 1962^[162]).

3.4.2. La notion d'impédance ramenée sur le larynx

Cette théorie, introduite par Husson 1962^[162], se fonde sur la théorie des pavillons développée par Rocard. Le conduit vocal est en effet formé de cavités (pharyngée, buccale, nasale) fortement convergentes et très étroites par endroit, et peut ainsi être assimilé à un pavillon convergent (cf. Figure 8). Cette analogie est plus nette lorsqu'on représente le conduit vocal de façon redressée (cf. Figure 8).

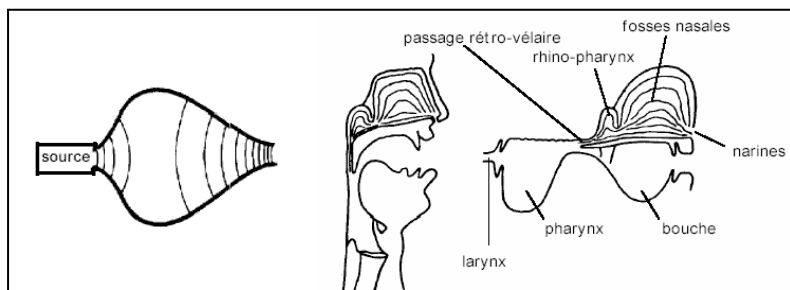


Figure 8. La figure de gauche représente un pavillon convergent. Celle de droite représente le conduit vocal de façon habituelle et redressée. D'après Husson 1962^[162].

La propriété principale d'un pavillon convergent est d'opposer à la propagation des ondes une impédance acoustique correspondant au rapport entre la pression et le déplacement d'air qui « charge » la source du pavillon. Cette charge ne doit pas être prise au sens d'effort mais au contraire comme un retour d'énergie entretenant la source. Ainsi, en ramenant de l'impédance sur le larynx, le conduit vocal provoque un accroissement de la pression intra-glottique, s'accompagnant d'une chute légère du tonus glottique, et d'une légère augmentation du quotient ouvert, aidant alors les cordes vocales dans leur effort de résistance à la pression sous-glottique et diminuant le risque de lésions (Husson 1962^[162]). L'impédance ramenée sur le larynx peut donc être utilisée comme mécanisme protecteur du larynx.

De façon générale, l'impédance ramenée sur la source excitatrice est d'autant plus importante que le pavillon est long, étroit ou convergent. Dans le cas du pavillon bucco-pharyngé, l'impédance ramenée croît lorsqu'on allonge le conduit vocal ou lorsqu'on abaisse le vélum pour nasaliser un son, ce qui revient à ajouter un deuxième conduit en dérivation. Certaines techniques vocales peuvent ainsi être interprétées en terme d'élévation de l'impédance ramenée sur le larynx :

- le « tubage » est ainsi une technique utilisée principalement par certains chanteurs russes ou de l'Europe de l'est, et consiste à allonger les lèvres en protrusion.
- l'abaissement du larynx préconisé par certaines méthodes de chant consiste également à allonger le conduit vocal.
- la technique de placement de la voix « dans le nez » ou « dans le masque », permettrait également d'augmenter considérablement l'impédance ramenée sur le larynx grâce à l'action d'un deuxième pavillon nasal.
- le « sombrage », perçu comme une déformation des voyelles ouvertes vers des voyelles fermées, est parfois interprété comme une dilatation du pharynx.
- la « couverture », parfois confondue avec le sombrage, et déjà étudiée comme technique de protection du larynx au niveau des zones de passage (Chuberre 2000 ^[53]), pourrait également être interprétée comme une augmentation de l'impédance ramenée sur le larynx puisque sa réalisation consiste, selon Miller 1990 ^[243], à rétrécir l'ouverture buccale, avancer les lèvres et abaisser le larynx.

Enfin, la facilité d'émission dont témoignent les chanteurs et orateurs dans certaines acoustiques de salle pourrait s'expliquer par la notion d'impédance ramenée sur le larynx. Celle-ci augmente en effet avec l'impédance de rayonnement du pavillon. Les propriétés réverbérantes de la pièce dans laquelle se trouve le locuteur vont ainsi avoir une influence sur l'impédance de rayonnement et donc également sur l'impédance ramenée sur le larynx.

3.5. L'évaluation de ces stratégies de gestion par l'adéquation de l'effort aux exigences de la situation de communication

Nous avons envisagé jusqu'à maintenant les stratégies individuelles de gestions de l'effort comme des configurations laryngées, articulatoires ou posturales différentes pour parvenir à un résultat semblable. Mais quel est au plus précisément ce « résultat » ? Dans la majorité des études sur l'efficacité vocale, ce résultat est objectivé par l'intensité du son produit, ou par son contenu spectral. Cependant, le but final de l'enseignant ou du vendeur sur un marché, lorsqu'ils forcent sur leur voix, est-il de « produire un son » ? En réalité, leur but est plus large et consiste à communiquer un message au travers de ce son. Leur production ne consiste donc pas seulement à générer des sons vocaux mais plus globalement de la parole, à l'attention d'un ou de plusieurs destinataires.

Cette finalité de la production vocale est complètement occultée de la littérature et des protocoles d'étude du forçage vocal. Nous faisons au contraire l'hypothèse dans ce travail que la communication est la clé de voûte du problème du forçage vocal. En effet, l'humain est par nature économe vis à vis de son énergie. C'est la motivation à obtenir quelque chose d'important pour lui qui peut le pousser à faire des efforts. Dans notre cas, le locuteur va fournir un effort (et non le subir), dans le but d'être compris ou de faire passer une émotion à un auditeur. Mais son objectif premier, au-delà de transmettre un message, est avant tout de provoquer chez l'interlocuteur, par la signification de ce message, une conduite désirée (Bougnoux 2002 ^[33]). Sa gestion de l'effort ne s'arrête donc pas à adopter telle ou telle configuration corporelle pour optimiser l'efficacité de la production, mais consiste également à adapter son effort aux exigences de la situation de communication, en particulier aux exigences d'intelligibilité lors de la communication parlée (théorie hypo-hyper (H&H) de Lindblom 1990 ^[217]). Ainsi, on aura davantage tendance à articuler en parlant à un bébé ou à une personne de langue étrangère, ou encore à davantage crier en parlant à une personne un peu sourde. De même, Le Huche 1984 ^[202] propose une typologie des comportements vocaux allant dans le sens d'une adaptation de l'effort vocal au degré de persuasion recherché par le locuteur. Ainsi, ce qu'il nomme la « voix d'expression simple » et la « voix projetée » se distinguent selon lui par le fait que la première n'attend pas de réaction de la part de l'interlocuteur (parler tout seul, faire un commentaire,...) tandis que la deuxième appelle une réponse ou un comportement de la part du destinataire (question, ordre,...). Dans ce cas, une absence de réaction de l'interlocuteur peut entraîner l'utilisation de la « voix d'insistance », caractéristique selon lui du forçage vocal.

Le locuteur ne se contente donc pas de gérer un effort qui lui serait imposé, mais régule lui-même cet effort par rapport à la situation de communication, en effectuant un compromis permanent entre le

désir d'être compris et d'agir sur son interlocuteur, et la minimisation de l'énergie dépensée pour y parvenir (cf. Figure 9). Le forçage vocal pourrait alors consister chez certains locuteurs en un dysfonctionnement de cette régulation, les menant à surestimer l'effort nécessaire pour être compris dans une situation donnée. Cette surestimation pourrait provenir d'une mauvaise évaluation de la situation de communication ou d'une sous-estimation de leur propre intelligibilité. La prédisposition des personnalités extraverties vis à vis du forçage vocal (Roy et al. 2000^[295] ; Giovanni 2007^[114]), de même que des personnes ne prêtant pas attention aux signes de fatigue ou de douleur, va tout à fait dans le sens de cette hypothèse.

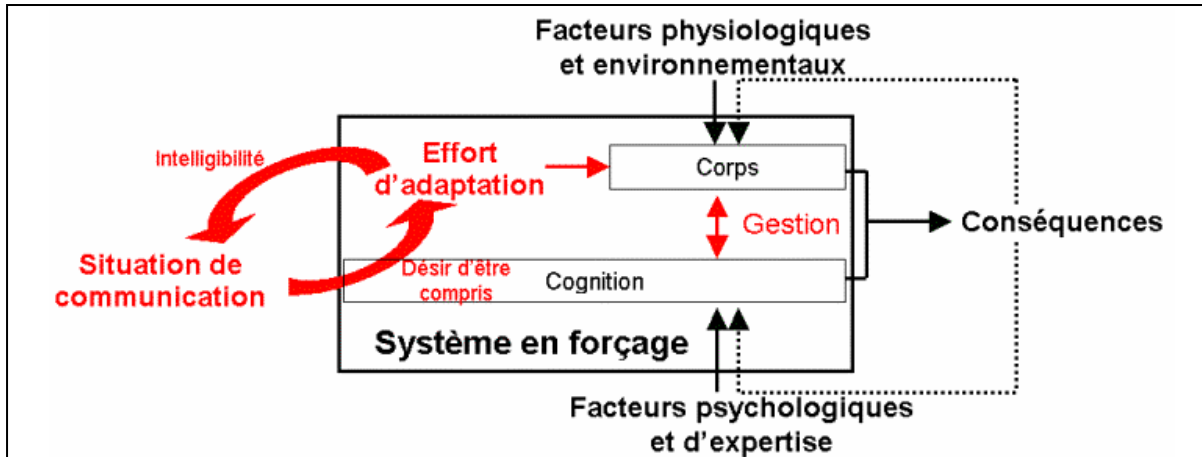


Figure 9. Deuxième réadaptation de la modélisation du forçage vocal (dans la continuité des Figure 4 et Figure 6), à la fois comme une mauvaise gestion de l'effort par le système-locuteur (corps + cognition), mais également comme une régulation de cet effort vocal par rapport aux exigences d'intelligibilité de la situation de communication. L'entrée du système, subie par le locuteur, n'est alors plus l'effort vocal mais la situation de communication.

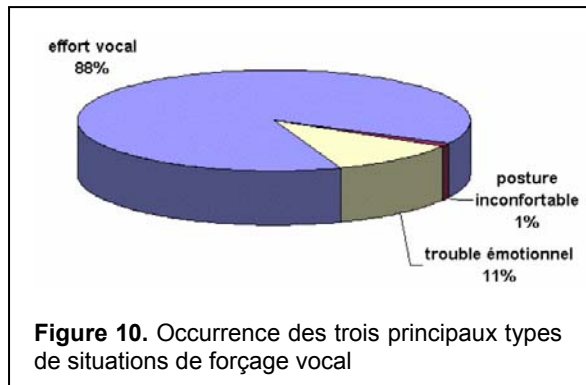
De ce point de vue, plusieurs techniques vocales peuvent être considérées comme efficaces vis-à-vis de la communication : Les formants du chanteur et de l'acteur, correspondant à renforcement énergétique du spectre dans la zone sensible de l'oreille (Sundberg 1972^[337] ; Bele 2006^[23]), permettaient de faire émerger la voix au milieu d'un orchestre ou de la porter à distance, sans pour autant augmenter l'intensité vocale globale. De même, une augmentation de la précision articulatoire peut accroître considérablement l'intelligibilité. Mais surtout, d'autres techniques de communication non vocales peuvent participer à l'efficacité de la communication : le prototypage articulatoire ou l'utilisation du langage des mains peuvent également servir à l'interlocuteur d'un point de vue visuel pour compléter l'information auditive (Benoit et al. 1994^[24] ; Schwartz et al. 2004^[309]), et ainsi peut-être permettre au locuteur de réduire son effort au niveau acoustique. Par ailleurs, la capacité à développer sa présence et à gérer les tours de parole dans une conversation peut éviter au locuteur de faire des efforts intenses uniquement pour prendre la parole. Enfin, la concision, ou l'art d'exprimer une idée en peu de mots, peut également rentrer dans les stratégies d'efficacité de communication.

Ces réflexions nous amènent à formuler l'hypothèse qu'il existe des stratégies linguistiques pour communiquer dans certaines situations de communication adverses et qu'elles participent au même titre que d'autres stratégies physiologiques ou acoustiques, à la gestion de l'effort vocal et par conséquent au forçage vocal ou à sa prévention. Cela nous amènera à étendre notre champ disciplinaire de l'acoustique aux sciences du langage, et à la phonétique en particulier.

Enfin, plusieurs situations de communication sont recensées dans la littérature comme des facteurs prévenant ou favorisant le forçage vocal. Certaines professions dont la voix est un outil de travail sont en effet considérées comme « à risque » vis à vis du forçage vocal : les enseignants (Rantala et al. 1999^[280] ; Sodersten et al. 2002^[321]), les instructeurs d'aérobic (Long et al. 1998^[225]), les instructeurs de l'armée (Mann et al. 1999^[232]), les opérateurs de renseignements téléphoniques ou de télémarketing (Lehto et al. 2003^[206]) ou encore les chanteurs et acteurs professionnels (Kitch et al. 1994^[181]). De même, la présence de bruit ambiant, l'éloignement de l'interlocuteur, une mauvaise acoustique de salle ou l'absence d'amplification de la voix (Long et al. 1998^[225]) sont considérées comme des facteurs de forçage. Toutes ces situations ne sont pas pour nous des « facteurs » mais l'entrée même du système de forçage (cf. Figure 9), qui oblige le locuteur à augmenter son effort vocal en adaptation à la situation de communication, dans le but de se faire comprendre par un interlocuteur

dans une situation extrême, ou dans le but de tenir un engagement professionnel, malgré les signes de fatigue qu'il peut ressentir.

Nous avons alors cherché à recenser les différentes situations de communication pouvant pousser le locuteur à adapter son effort et l'amener à éventuellement forcer sur sa voix. Les 21 personnes interrogées quant à leur conception du forçage vocal (cf. paragraphe 1) ont également répondu à la question « pouvez vous citer des situations de forçage vocal ? ». Le détail des réponses est consultable en Annexe Res1.



Parmi les situations citées, 1% représentent des prises de parole en posture inconfortable et 11% des situations de trouble émotionnel inhibant (trac, timidité), tandis que 88% représentent des situations d'effort vocal (cf. Figure 11). Cet effort est induit à 8% (des 88%) par une mauvaise acoustique de salle (amphithéâtre, plein air), à 12% par l'usage d'une hauteur de voix inhabituelle (principalement dans le chant), à 19% par un usage prolongé de la voix (enseignement la plupart du temps) et à 53% pour un usage de la voix à forte intensité (cf. Figure 11).

Enfin, l'usage de la voix forte ou criée est exigé dans 3% des cas (des 53% précédents) par une situation d'urgence (avertissement, demande d'aide), dans 3% par l'éloignement de l'interlocuteur, dans 13% par une émotion (colère, excitation, joie) et dans 63% par la présence de bruit ambiant (salle de classe, lieux public de divertissement, usine, milieu urbain et transports en commun) (cf. Figure 11).

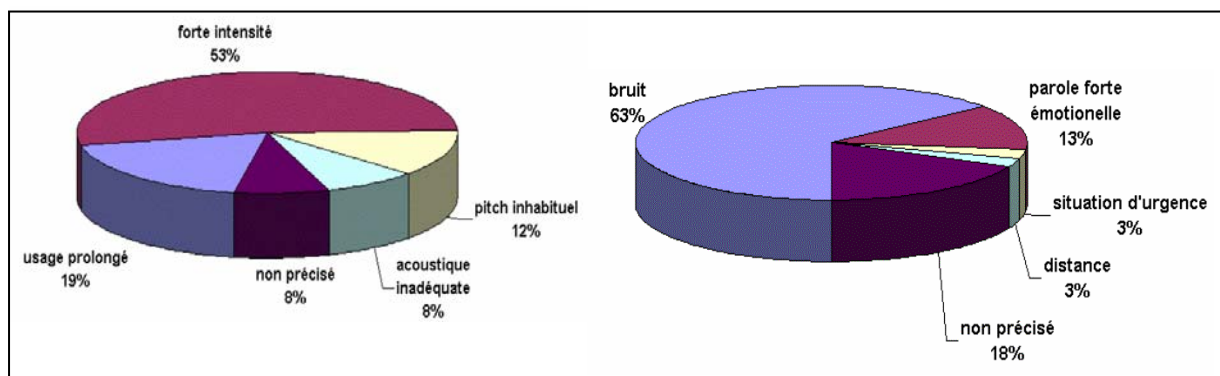


Figure 11. A gauche, répartition des différentes situations d'effort vocal. A droite, répartition des différentes situations d'usage de la voix à forte intensité.

4. CONCLUSION DU CHAPITRE 1

Le forçage vocal s'avère être un objet complexe et appréhendé de différentes façons selon les domaines d'expertise. Qu'il soit défini au niveau laryngé, gestuel ou individuel, le forçage a le plus souvent été appréhendé comme un phénomène subi par ce système (laryngé, gestuel ou individuel), menant à des conséquences pathologiques pour ce système, et variant dans une marge très étroite par réaction à une variation du degré d'effort et à la variation de certains facteurs extérieurs (hydratation, poussière, posture, etc.)

Nous rappelons que notre objectif est de comprendre pourquoi, face à une même situation, certaines personnes réussissent à communiquer sans mal tandis que d'autres s'épuisent jusqu'au dysfonctionnement vocal. Pour tenter de répondre à cette question, nous effectuons l'hypothèse que les différences interindividuelles face au forçage vocal ne sont pas seulement d'ordre physiologique mais également dues à des différences de gestion de l'effort, impliquant le fonctionnement cognitif du locuteur et son rôle actif dans le phénomène de forçage. Nous testerons la validité d'une telle hypothèse en cherchant à caractériser les différences interindividuelles de comportement vocal en situation d'effort.

Par ailleurs, nous effectuons l'hypothèse que ces différences de gestion peuvent être de deux ordres :

- de l'ordre de différences de configuration au niveau du geste de production (laryngé, mais également respiratoire, postural ou articulatoire), pour réaliser une même intention. Certaines stratégies étant plus coûteuses que d'autres et faisant porter plus ou moins de charge au larynx.
- de l'ordre de différences d'adaptation de l'effort à la situation de communication. Certaines stratégies pour être intelligibles étant plus économes que d'autres en énergie.

L'immersion dans un environnement bruyant étant apparue comme la situation de forçage la plus fréquente dans notre questionnaire, nous avons donc choisi de commencer par étudier cette situation au cours de cette thèse, tout en étant conscients que nos résultats ne seront valables que pour ce cas particulier.

Pour résumer, nous envisagerons durant ce travail que **le forçage vocal est une stratégie (non unique) inefficace de gestion de l'effort utilisée par certains individus pour faire face à une situation de communication. L'inefficacité de cette stratégie provient en premier lieu de la concentration des efforts pour la produire sur le larynx au lieu de leur répartition sur l'ensemble de la chaîne de production. Mais elle correspond également à une mauvaise adaptation (en particulier à une surestimation) par le locuteur de l'effort à fournir par rapport aux exigences d'intelligibilité de la situation de communication.** Nous chercherons durant cette thèse à explorer de façon expérimentale la validité d'une telle hypothèse.

Plusieurs étapes seront nécessaires:

- (1) La conception du forçage vocal que nous proposons et dont nous allons tester la validité nous oblige à apporter également des modifications quant aux paradigmes expérimentaux d'étude du forçage vocal, de façon à tenir compte de certains aspects tels que l'activité du locuteur dans ce phénomène ou l'aspect d'interaction communicationnelle avec un interlocuteur. Avant toute chose, nous chercherons dans le prochain chapitre à comprendre les mécanismes physiologiques et cognitifs qui sous-tendent l'adaptation de la parole en environnement bruyant, puisque telle est la situation de communication que nous avons choisie d'étudier pour commencer. Cela nous permettra d'identifier les facteurs influençant ces mécanismes d'adaptation (chapitre 2) et ainsi de les prendre en compte dans nos protocoles expérimentaux (chapitres 3 et 4).
- (2) Nous pourrions alors caractériser les différences interindividuelles d'adaptation de plusieurs locuteurs pendant une situation d'exposition au bruit. Rappelons que nous prendrions prioritairement en compte le forçage vocal comme un phénomène cognitif et communicationnel, se traduisant par un comportement gestuel et acoustique observable et pouvant éventuellement induire des conséquences pathologiques pour le larynx. Aussi, nous caractériserons dans les chapitres 5 à 10 les différences individuelles et contextuelles d'adaptation à une situation de communication à l'aide de paramètres physiologiques (laryngés et articulatoires) et acoustiques que nous considérerons comme des indicateurs de stratégies (cognitives) de communication (linguistique, acoustique et visuelle).
- (3) Une fois ces stratégies identifiées, il nous faudra en évaluer leur efficacité vis à vis de la répartition des efforts et de l'adaptation de l'effort aux exigences d'intelligibilité de la situation de communication. Ce point ne sera pas abordé pendant cette thèse, faute de temps. Quelques pistes seront données à explorer plus en détail dans des travaux futurs

Partie II

***Méthodologies et bases
de données pour l'étude
de l'adaptation de la
parole dans en
environnement bruyant***

- Présentation de la Partie II -

Dans la partie précédente, nous avons précisé la façon dont nous concevons le forçage vocal dans ce travail et notre positionnement par rapport aux différents champs disciplinaires ayant déjà contribué à une meilleure connaissance de ce problème. En particulier, le fil conducteur de notre travail sera de caractériser les différences interindividuelles d'adaptation de la parole dans le bruit.

Puisque notre objet d'étude est d'abord défini dans le champ psychologique et communicationnel, nous devons en tenir compte dans nos protocoles expérimentaux. Quels sont les mécanismes physiologiques et cognitifs mis en jeu dans l'adaptation de la parole en environnement bruyant ? Nous verrons dans le Chapitre 2 comment la production de parole est régulée par le retour auditif et comment les locuteurs adaptent leur effort vocal dans le but d'être intelligibles pour leur interlocuteur. Nous ferons également le point sur les méthodologies utilisées dans les études antérieures, afin d'en déterminer les intérêts et les limites par rapport à nos objectifs.

Dans le Chapitre 3, nous examinerons alors si le port d'un casque pour immerger les locuteurs dans le bruit en laboratoire constitue un biais à l'adaptation dans le sens où il rajoute une atténuation supplémentaire du retour auditif. Nous envisagerons comme alternative de reconstituer un environnement bruyant par le biais de haut-parleurs, ce qui nous obligera à trouver une technique de débruitage pour pouvoir effectuer des mesures rigoureuses sur le signal acoustique de parole. Nous mènerons ensuite une expérience de comparaison de trois méthodologies d'immersion d'un locuteur dans le bruit, et constituerons pour ces besoins une première grande base de données acoustiques et électroglottographiques (BD1) composée de 10 locuteurs, deux types de bruit, quatre niveaux de bruit et trois moyens d'immersion des locuteurs dans le bruit.

De même, nous nous poserons la question dans le Chapitre 4 de savoir si une tâche de lecture est toujours appropriée à notre objectif de caractériser la façon dont les individus adaptent leur effort à une situation de communication. Nous avons alors imaginé plusieurs jeux permettant d'enregistrer des locuteurs en situation de communication semi-spontanée tout en contrôlant le matériel linguistique, en interaction avec un interlocuteur ou de façon isolée. Nous avons pu alors tester l'intérêt de tels jeux et l'influence du dialogue avec un interlocuteur sur l'adaptation de la parole dans le bruit, en constituant une deuxième base de données acoustiques, électroglottographiques et articulatoires (BD2) composée de trois locutrices, deux types de bruit et deux conditions d'énonciation (avec et sans interaction avec un interlocuteur).

Chapitre 2 :

Mécanismes et facteurs de l'adaptation de la parole en environnement bruyant

- 1. LES ACQUIS DES TRAVAUX ANTERIEURS**
- 2. LES MECANISMES DE L'ADAPTATION DE LA PAROLE DANS LE BRUIT**
- 3. LES METHODOLOGIES D'ETUDE DE L'ADAPTATION DE LA PAROLE DANS LE BRUIT**
- 4. POSITIONNEMENT DE NOTRE ETUDE PAR RAPPORT A CES DIFFERENTES THEORIES ET METHODOLOGIES**

Dans le premier chapitre, nous avons posé l'hypothèse que le forçage est un dysfonctionnement de la régulation de l'effort et de sa gestion face à une situation de communication. Nous avons retenu dans un premier temps les environnements bruyants comme situation amenant les locuteurs à adapter leur effort.

Notre but est donc maintenant de chercher à caractériser le forçage vocal à travers les différences interindividuelles de stratégies d'adaptation à un environnement bruyant. Pour cela, nous allons commencer par explorer l'état des connaissances sur ce sujet dans la littérature, par mieux comprendre les mécanismes qui régissent l'adaptation de la parole dans le bruit et par recenser les facteurs influençant ces mécanismes à prendre en compte au niveau méthodologique dans nos protocoles d'étude.

1. LES ACQUIS DES TRAVAUX ANTERIEURS

Avant toute chose, nous allons faire le bilan du travail qui a déjà été accompli dans la littérature en ce qui concerne les différences interindividuelles d'adaptation de la parole dans le bruit.

En fait, nous verrons que l'adaptation au bruit a déjà été largement étudiée dans divers objectifs (diagnostiquer la surdité, caractériser le forçage vocal, modéliser les variations de la parole, améliorer la robustesse de la reconnaissance automatique face à des conditions perturbées, perception de la parole dans des conditions d'intelligibilité dégradée) mais que ces différents buts ont plutôt conduit leurs auteurs à chercher une unique stratégie d'adaptation au bruit dont nous présenterons les caractéristiques principales. Les importantes différences interindividuelles observées sont alors très souvent considérées comme une variabilité gênante vis à vis de la modélisation de cette adaptation. Quelques études se sont cependant intéressées à cette variabilité au niveau du genre du locuteur et de sa langue.

1.1. Des objectifs différents selon les champs disciplinaires

Toujours comme pour le problème de l'effet Lombard, le problème de l'adaptation de la parole a été envisagé de différentes façons dans la littérature, tant au niveau de l'interprétation des mécanismes qui la régissent qu'au niveau des méthodologies mises en oeuvre pour l'étudier. De la même façon, ces différences proviennent avant tout des différentes disciplines et domaines d'expertise s'étant intéressés à cette adaptation et à leurs différents objectifs. Nous commencerons donc par dresser le cadre disciplinaire et épistémologique dans lequel ont été menées les différentes études sur l'adaptation de la parole dans le bruit, afin de mieux en comprendre les différences de points de vue.

1.1.1. L'effet lombard en audiologie et en psychoacoustique

L'adaptation de la voix en environnement bruyant a en premier lieu été mise en évidence par Etienne Lombard, oto-rhino-laryngologiste, dans le but de trouver un test de détection de la surdité (unilatérale ou bilatérale). Il constate en effet que :

- les sourds par lésion de l'appareil labyrinthique élèvent anormalement la voix.
- de façon comparable, un locuteur sans problème auditif élève anormalement la voix lorsqu'il est immergé dans du bruit intense. « Il se crie en quelque sorte à lui-même comme on crie à l'oreille d'un sourd », et présente des caractéristiques vocales identiques à celles des sourds labyrinthiques bilatéraux (débit élevé, voix criarde, monotone, sans couleur...).
- l'immersion des sujets sourds dans du bruit n'ajoute pas d'effet sur leur façon de crier.
- l'application d'un bruit intense sur une oreille, puis sur les deux, d'un sourd unilatéral, ne provoque une élévation de la voix que lorsque l'oreille saine est assourdie.

Lombard déclare donc être en mesure de détecter avec certitude une surdité unilatérale et publie en 1911 « le signe de l'élévation de la voix » (Lombard 1911 ^[224]), article de référence sur l'« effet Lombard ». Egan 1971 ^[79] a ensuite développé les observations de Lombard au cours de sa thèse portant sur la détection de la surdité et sa simulation, ainsi que sur la régulation de la parole par l'audition (Egan 1972 ^[80]). L'effet Lombard devient alors l'exemple typique de l'étude de la boucle audio-phonatoire en tant que système asservi (Badaoui 1997 ^[18] ; Korn 1954 ^[185] ; Tonkinson 1994 ^[353])

1.1.2. L'effet Lombard-Tarneaud en phoniatry

Toujours dans le domaine médical, le test de type Lombard a été réutilisé par les phoniatres et les orthophonistes cette fois, afin de diagnostiquer ou de rééduquer certaines pathologies vocales, en particulier les dysphonies fonctionnelles (Tarneaud 1935^[343]). L'effet Lombard est mis à profit pour provoquer chez le patient une situation d'effort vocal plus naturelle qu'en lui demandant simplement de crier (Aronsson et al. 2005^[12] ; Dejonckere 1979^[63] ; Le Huche 1984^[202] ; Giovanni 2006^[113] ; Papon 2006^[265] ; La Machhia 2005^[189], Neils et al. 1987^[252] ; Schultz-Coulon et al. 1976^[305]).

1.1.3. La reconnaissance automatique et la modélisation phonétique de la parole Lombard

Ces modifications de la parole en environnement bruyant ne se limitent pourtant pas à une simple élévation de la voix et posent des problèmes au domaine de la reconnaissance de la parole. Le problème de la reconnaissance automatique de la parole Lombard (i.e. de la parole produite en environnement bruyant) réside en effet bien moins dans la détérioration du signal acoustique par le bruit ambiant que dans l'ensemble des adaptations acoustiques et phonétiques effectuées par rapport à la parole normale, diminuant considérablement la performance des systèmes automatiques basés sur des modèles de parole normale (Stanton 1988^[327] ; Junqua 1993^[171] ; Takizawa et al. 1990^[342]). Par conséquent, de nombreux projets ont été développés en collaboration avec des phonéticiens dans le but de modéliser ces modifications de la parole pour les intégrer aux systèmes de reconnaissance de la parole existant et pour améliorer leur robustesse (Stanton 1988^[327], Hanson et al. 1990^[134]). Le but n'est alors pas de décrire une adaptation individuelle ni de comprendre les mécanismes ou les motivations de cette adaptation. Cela explique pourquoi ces études se sont orientées vers la recherche d'invariants et de profils moyens entre les locuteurs (Junqua 1993^[171] ; Castellanos et al. 1996^[46], Mokbel 1992^[245]...), ainsi que sur la constitution de bases de données de grande taille (European Esprit project ROARS (Robust Analytical Speech Recognition System), BDBruit (Zeiliger et al. 1994^[382]), Avicar (Lee et al. 2004^[204]), etc.)

1.1.4. La communication et l'intelligibilité en environnement bruyant

Enfin, de nombreuses études s'intéressent à la communication en environnement bruyant du point de vue du locuteur ou de l'auditeur : comment l'interlocuteur s'adapte-t-il pour être compris, et y parvient-il ? Au contraire, quels sont les indices robustes au bruit, sur lesquels l'auditeur se base pour percevoir le message, malgré la sérieuse détérioration du signal acoustique ?

Dans le premier cas, les caractéristiques de la parole Lombard sont interprétées sous un angle phonologique, afin de comprendre en quoi elles renseignent sur les intentions de communication du locuteur (Welby 2006^[375] ; Davis et al. 2006^[62]), ainsi que sur les mécanismes réflexes ou haut-niveau mis en jeu dans cette adaptation (Lane et al. 1971^[192]). Dans le deuxième cas, la parole étudiée n'est pas toujours produite dans du bruit, mais un bruit additionnel est ajouté aux enregistrements afin de mener des tests perceptifs d'intelligibilité auprès d'auditeurs (Miller et al. 1955^[242], Dohen 2005^[74]).

1.2. A la recherche d'une modélisation unique des modifications de la parole dans le bruit

Ces différents domaines d'expertise et leurs objectifs permettent de comprendre pourquoi les modifications de la parole Lombard ont jusqu'à maintenant plutôt été caractérisées en recherchant une même tendance commune ou moyenne à tous les locuteurs. Les résultats des études antérieures sont détaillés dans le tableau 3 de l'annexe Bib2. Nous en présentons ici une synthèse.

1.2.1. Caractéristiques acoustiques globales de la parole Lombard

- *Intensité vocale*

L'intensité vocale moyenne est le principal paramètre de description de l'effet Lombard et elle augmente significativement dans toutes les études (Bond et al. 1989^[31] ; Castellanos et al. 1996^[46] ; Davis et al. 2006^[62] ; Dejonckere et al. 1983^[65] ; Egan 1972^[80] ; Junqua 1992^[173] ; Junqua 1993^[171] ; Kadiri 1998^[176] ; Kim et al. 2005^[179] ; Lane et al. 1970^[194] ; Mokbel 1992

[245] ; Pisoni et al. 1985 [272] ; Stanton 1988 [327] ; Ternström et al. 2002 [346] ; Ternström et al. 2006 [345] ; Van Summers et al. 1988 [361] ; Webster 1962 [372]). Cette augmentation de l'intensité est plus importante sur les voyelles que sur les consonnes (Junqua et al. 1990 [174] ; Fonagy et al. 1966 [94] ; Schulman 1989 [304] ; Draeger 1951 [76] ; Junqua 1993 [171] ; Castellanos et al. 1996 [46]).

- *Fréquence fondamentale*

De même, une élévation de la fréquence fondamentale moyenne est observée dans toutes les études. Boril et al. 2005 [32] ont également examiné plus en détail la distribution des valeurs de fréquence fondamentale et observé un déplacement de cette distribution vers les fréquences aiguës.

- *Durée*

Un ralentissement du débit de parole accompagne fréquemment ces modifications (Junqua 1993 [171] ; Kim 2005 [180] ; Kadiri 1998 [176] ; Pisoni et al. 1985 [272]), et concerne davantage l'allongement des voyelles que celui des consonnes. Cependant, certaines études ont observé un ralentissement non significatif, voire une accélération chez certains locuteurs (Kim 2005 [180]).

- *Répartition de l'énergie en fréquences*

Au niveau du spectre de la voix, de nombreuses études rapportent un aplatissement de la pente spectrale dans les basses fréquences sur les voyelles, et une augmentation de la pente spectrale hautes-fréquences (Castellanos et al. 1996 [46] ; Van Summers et al. 1988 [361] ; Junqua 1993 [171] ; Mokbel 1992 [245] ; Pisoni et al. 1985 [272]). D'autres études examinent plutôt l'énergie spectrale par bandes de fréquences et notent un renforcement du spectre dans les médiums, ainsi que dans la zone 4-5 kHz pour les femmes (Castellanos et al. 1996 [46] ; Junqua 1993 [171] ; Stanton 1988 [327]).

1.2.2. Caractéristiques articulatoires globales de la parole Lombard

Peu d'études se sont intéressées à l'articulation en environnement bruyant. Plusieurs auteurs ont écrit que l'articulation était peu affectée par le bruit (Schliesser et Coleman 1968, Ringel et Steer 1963, Gammon et al. 1971 cités dans Garber et al. 1980 [101]). L'étude de Schulman 1989 [304] sur la voix forte reste une référence pour dire qu'une augmentation de l'ouverture de la mâchoire accompagne en général l'augmentation de l'intensité vocale. Depuis le début de notre travail, deux études ont montré que l'amplitude des mouvements articulatoires augmentait effectivement aussi pour la parole Lombard en particulier (Davis et al. 2006 [62] ; Kim et al. 2005 [179]).

1.2.3. Caractéristiques phonétiques de la parole Lombard

1.2.3.1. Du point de vue des indices intonatifs

Au niveau syllabique, aucune différence significative n'a été observée entre l'augmentation de la fréquence fondamentale sur la première ou la deuxième syllabe des mots (Bond et al. 1989 [31]). Welby 2006 [375] a en revanche émis l'hypothèse que le « coude » de la courbe mélodique (pouvant marquer la frontière entre le déterminant et le début d'un mot de contenu) soit renforcé dans le bruit, de même que l'accentuation secondaire sur des syntagmes de plus de 3 syllabes. Ces tendances ont en effet été observées, mais de façon non significative.

1.2.3.2. Du point de vue des caractéristiques acoustiques des différents phonèmes

- *Formants vocaliques*

Toutes les études observent une augmentation du premier formant des voyelles (F1) dans le bruit ou durant un effort vocal. Cette élévation du F1 a également été observée pour les liquides et les nasales. En revanche, les résultats sont variables selon les études en ce qui concerne le deuxième et le troisième formant (respectivement F2 et F3).

Ainsi, Castellanos et al. 1996 [46], Davis et al. 2006 [62] observent une augmentation de F2 pour toutes les voyelles, Junqua 1993 [171] rapporte une augmentation de F2 seulement chez les femmes, et Junqua 1992 [173] et Boril et al. 2005 [32] une augmentation de F2 pour toutes les voyelles excepté pour les [i]. Enfin, Van Summers et al. 1988 [361] et Lienard et al. 1999 [215] n'observent pas d'évolution significative de F2, tandis que Pisoni et al. 1985 [272] observent même une diminution de F2.

La plupart des études ne prêtent pas attention au 3^{ème} formant ou bien ne constatent pas d'évolution particulière (Traunmüller et al. 2000 ^[356]). Au contraire, Davis et al. 2006 ^[62] rapportent une augmentation importante de F3 dans le bruit, et cela pour toutes les voyelles quand Hansen et al. font cette même observation mais pour certaines voyelles seulement.

Certaines études ont alors moins regardé l'évolution isolée de chaque formant que l'évolution du triangle vocalique dans le plan F1*F2. Junqua 1992 ^[173] constate ainsi que l'espace vocalique dans le plan (F2-F1)*F1 est plus important dans le bruit, tandis qu'au contraire, Rostolland 1982 ^[289] observe une réduction de l'espace vocalique en voix criée, bien que le deuxième formant reste distinctif entre les voyelles. En revanche, cet auteur mentionne que les évolutions de F3 vont dans le sens d'une uniformisation de toutes les voyelles pour cette dimension formantique, excepté pour les [i] où le F3 reste plus important et pour le [u] où il reste plus faible.

- *Répartition de l'énergie en fréquences*

L'exploration des modifications phonétiques par catégorie phonétique a fait l'objet de quelques études (Junqua 1993 ^[171] ; Castellanos et al. 1996 ^[46] ; Kim 2005 ^[180]).

La pente spectrale basses fréquences (en dessous de 3kHz) diminue dans le bruit pour les voyelles, les liquides et les nasales (Castellanos et al. 1996 ^[46] ; Junqua 1993 ^[171]), c'est-à-dire que l'énergie est renforcée dans les médiums. Castellanos et al. 1996 ^[46] observe également une diminution de cette pente spectrale pour les plosives, mais pas pour les fricatives non voisées.

La pente spectrale hautes fréquences (entre 3 et 8kHz) diminue pour la plupart des voyelles et des liquides (Junqua 1993 ^[171]), ainsi que pour les plosives voisées et les nasales dans Castellanos et al. 1996 ^[46], mais augmente au contraire pour les fricatives non voisées (Castellanos et al. 1996 ^[46]).

Lorsque l'énergie est examinée par bande de fréquences, les nasales voient leur énergie atténuée dans toutes les bandes de fréquences chez Stanton 1988 ^[327], tandis que Junqua 1993 ^[171] observe un renforcement de l'énergie pour ces consonnes entre 250 et 500Hz et une diminution entre 3 et 8kHz.

L'énergie spectrale des liquides diminue également entre 0 et 500 Hz, augmente entre 500 et 3000 Hz, et diminue entre 5 et 8kHz (Junqua 1993 ^[171]).

L'énergie des fricatives comme celle des plosives est atténuée entre 0 et 500 Hz, et augmente entre 5 et 8kHz (Junqua 1993 ^[171]). En revanche, les fricatives sont aussi renforcées dans les médiums (entre 1 et 2,5kHz) tandis que les plosives sont plutôt renforcées entre 2 et 4kHz (Castellanos et al. 1996 ^[46]).

- *Durée*

Au niveau de la durée des segments phonétiques, un raccourcissement des consonnes non voisées a été observé, en particulier des plosives non voisées (Kim 2005 ^[180], Stanton 1988 ^[327]). Kim 2005 ^[180] n'observe aucune modification significative de la durée des autres consonnes voisées, nasales et liquides, au contraire de Stanton 1988 ^[327] qui note un allongement significatif de la durée des liquides en parole forte et Lombard.

1.3. Une « variabilité » individuelle importante

Nous pouvons voir que les différentes études sur la parole Lombard observent des résultats légèrement différents, voire parfois contraires. Les auteurs eux-mêmes soulignent la grande variabilité des adaptations entre leurs locuteurs (Holmberg et al. 1988 ^[153] ; Junqua 1993 ^[171]). Van Summers et al. 1988 ^[361] observent ainsi de grandes différences au niveau du premier formant selon les locuteurs. Kim 2005 ^[180] rapporte également l'existence de tendances d'évolution différentes de la durée des nasales et des liquides dans le bruit, allongées ou raccourcies selon les locuteurs. Enfin, Egan 1972 ^[80] constate que la variabilité individuelle est plus marquée à forte intensité de bruit ambiant, et davantage chez les femmes.

Certains auteurs ont cherché à interpréter ces différences interindividuelles et à comparer en particulier les adaptations fournies par les locuteurs masculins et féminins, et par les locuteurs de langues différentes.

1.3.1 Selon le genre du locuteur

L'augmentation de l'intensité est souvent plus importante chez les hommes que chez les femmes (8 dB d'écart dans Rostolland 1982^[289]). Junqua 1992^[173] rapporte une augmentation de la fréquence fondamentale également plus importante pour les hommes, et plus faible pour les enfants. Tandis qu'au contraire, Kadiri 1998^[176] observe que l'augmentation du couple F0* Intensité est plus importante chez les femmes que chez les hommes, et Ternström et al. 2002^[346] et Sodersten et al. 2005^[322] constatent que l'étendue vocale des femmes est plus importante dans le bruit que celle des hommes.

Au niveau de la durée des segments phonétiques, Kim 2005^[180] observe que celle des consonnes diminue davantage dans le bruit pour les hommes que pour les femmes, tandis que celle des voyelles augmente davantage pour les femmes que pour les hommes. Ternström et al. 2002^[346] et Sodersten et al. 2005^[322] remarquent que le pourcentage de voisement augmente davantage dans le bruit chez les femmes que chez les hommes, bien que le temps de lecture augmente de façon comparable pour les deux genres. A l'inverse, Traunmüller et al. 2000^[356] observent que les hommes et les enfants marquent moins de pauses lorsque la distance entre les interlocuteurs augmente, contrairement aux femmes. Cependant, à très grande distance, les hommes finissent par augmenter également le nombre de pauses.

Au niveau spectral, de nombreuses études sur la voix criée ou sur la parole Lombard retrouvent un renforcement spectral chez les femmes autour de 4-5 kHz sur les voyelles, les consonnes nasales ou liquides, tandis que les hommes au contraire présentent un antiformant dans cette région, mais un renforcement spectral plus bas, entre 2 et 4kHz (Castellanos et al. 1996^[46] ; Junqua 1993^[171]). Ternström et al. 2002^[346] constatent que le barycentre spectral reste constant chez les hommes, au-dessus de 600 Hz tandis qu'il est proportionnel à la fréquence fondamentale pour les femmes. Enfin, Junqua 1993^[171] s'est intéressé au souffle sur la voix, relié à la différence d'amplitude des deux premières harmoniques, et a observé que les femmes réduisent beaucoup plus cette caractéristique en parole Lombard que les hommes.

Enfin au niveau formantique, Junqua 1993^[171] a observé que le deuxième formant augmentait dans le bruit seulement pour les femmes. Dans une étude précédente, il a également constaté des valeurs formantiques plus élevées dans le bruit chez les femmes, pouvant être interprétées comme le signe d'une parole plus « ouverte » (Junqua 1992^[173]).

1.3.2. Selon la langue du locuteur

La plupart des études concernent l'anglais américain (Junqua 1993^[171] ; Stanton 1988^[327]). Cependant, quelques études sur le français (Kadiri 1998^[176] ; Mokbel 1992^[245]), sur l'espagnol (Castellanos et al. 1996^[46]), sur l'allemand (Mixdorff et al. 2006^[244]), sur le japonais (Takizawa et al. 1990^[342]), sur le coréen (Kim 2005^[180]), sur le suédois (Ternström et al. 2002^[346] ; Sodersten et al. 2005^[322] ; Ternström et al. 2006^[345]) et sur le tchèque (Boril et al. 2005^[32]) sont également recensées.

Aucune étude, à notre connaissance, n'a explicitement comparé les adaptations acoustiques et phonétiques selon la langue.

Les études sur le français ou l'espagnol semblent montrer quelques différences par rapport aux études anglophones en ce qui concerne la variation du second formant.

L'allongement de la durée des syllabes est également variable selon les langues étudiées. Ainsi, Bond et al. 1989^[31] observent sur l'anglais que la durée de la première voyelle d'un mot de structure CVCV n'augmente pas significativement dans le bruit, tandis que la deuxième devient significativement plus longue. Rostolland 1982^[288] a de façon comparable observé que la deuxième syllabe était davantage allongée en français que la première. Au contraire, Kim 2005^[180] note qu'en Coréen, la plus grande augmentation de durée concerne la première syllabe, et que le mot est d'autant moins allongé qu'il est constitué d'un nombre élevé de syllabes.

2. LES MECANISMES DE L'ADAPTATION DE LA PAROLE DANS LE BRUIT

Nous venons de voir les principales modifications de la parole produite en environnement bruyant. Mais quels sont au juste les mécanismes induisant ces modifications ? En effet, si nous voulons caractériser le forçage vocal par les différences interindividuelles face à cette adaptation, il nous faut en comprendre les mécanismes ainsi que les facteurs qui les influencent, afin de mettre au point des protocoles expérimentaux permettant de s'approcher au mieux de notre objet d'étude.

2.1. Une régulation basée sur le retour auditif de sa propre voix

Nous avons vu précédemment que l'adaptation de la parole observée en environnement bruyant est aussi appelée Effet Lombard, en référence à celui qui le mit en évidence. Cet effet est en particulier caractérisé par l'augmentation de l'intensité vocale lorsque le retour de la voix est atténué, en particulier par l'immersion dans du bruit (Lombard 1911^[224] ; Lane et al. 1970^[194] ; Sinnott et al. 1975^[318]). Il existe l'effet inverse, appelé effet Sidetone, correspondant au fait qu'un locuteur diminue son intensité vocale lorsqu'il a un retour amplifié de sa propre voix, par exemple dans un casque (Lane et al. 1970^[194]).

Mais un tel comportement d'adaptation à une modification du retour auditif a également été observé pour d'autres caractéristiques telles que la hauteur modifiée du retour auditif (« *pitch-shift reflex* » étudié par Burnett et al. 1998^[39] ; Elman 1981^[83] ou Ternström et al. 1988^[347]), le timbre modifié par filtrage (« *filtered auditory feedback* » étudié par Tomatis 1987^[352], Garber et al. 1981^[99] ou Burzynski et al. 1985^[40]) ou encore un délai différé du retour auditif (« *delayed auditory feedback* », étudié par Lee 1950^[203], Howell 92 ; Stuart et al. 2002^[333] ; Kawahara94)

Cordemoy, en 1668, est le premier à avancer l'hypothèse d'une relation directe entre la production vocale et l'audition. Lombard 1911^[224], Fletcher et Fairbanks 1954^[86] s'y sont ensuite intéressés plus scientifiquement. Plusieurs observations étayant en effet cette hypothèse :

- une modification comportementale est non seulement toujours observée lors de la modification du retour sensoriel, mais la rapidité et l'amplitude de cette réponse sont proportionnelles à celles du stimulus (Lane et al. 1970^[194] ; Larson et al. 2000^[196])
- il semble exister une corrélation entre la précision du contrôle moteur et la sensibilité perceptive du sujet. Ainsi, Ringel et al. 1970^[286] ou Fucci 1972^[98] ont observé de meilleures capacités articulatoires chez les enfants possédant une sensibilité buccale développée. Dans le même sens, Martone, cité par Baum et al. 1997^[22], constatent que l'adaptation à une perturbation de la parole est plus difficile pour des personnes présentant des troubles auditifs. Enfin, de nombreux auteurs soulignent l'importance de la modalité auditive dans le développement de la parole chez l'enfant (Siegel et al. 1976^[316]), un développement anormal permettant justement de détecter d'éventuels problèmes auditifs. De façon comparable, un adulte confronté à une nouvelle situation, provoquée par une perturbation articulatoire par exemple, arrive très rarement à la compenser totalement dès le premier essai, mais améliore cette compensation au bout de plusieurs tentatives, allant de quelques minutes à quelques heures (Savariaux et al. 1999^[302], Jones et al. 2003^[169]). Une fois que les sujets ont trouvé et assimilé une solution adéquate à un nouveau problème, certains arrivent à ne plus avoir besoin de retour sensoriel pour s'adapter. En revanche, il semble indéniable que ce dernier soit nécessaire à l'apprentissage de cette adaptation.

Ces observations ont amené plusieurs chercheurs, en particulier Fairbanks 1954^[86], à modéliser cette relation entre la production vocale et l'audition sous la forme d'un système cybernétique de régulation dans lequel la perception de sa propre voix forme une boucle de rétroaction, et où la variable régulée correspond aux paramètres de la phonation (intensité, hauteur, timbre). Le système auditif est alors considéré comme le capteur de ce système, et le point de consigne envoie des commandes motrices au système phonatoire (tension des muscles, mouvement de cartilages, pression sous-glottique) (cf. Figure 12).

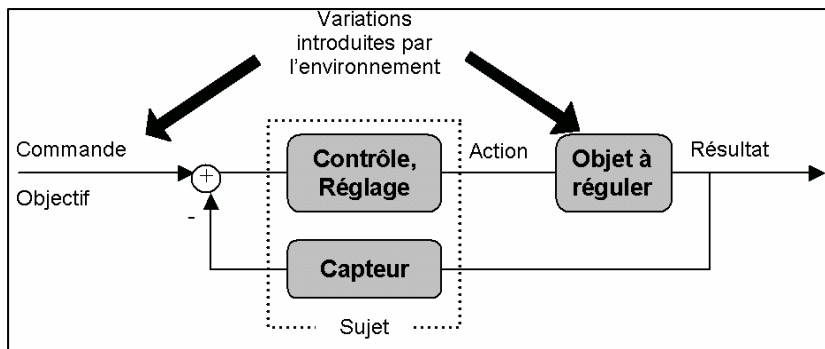


Figure 12. Représentation d'un système de régulation.

L'adaptation de la parole dans le bruit est donc principalement considérée comme une régulation de la production à partir des retours auditifs de sa propre voix. Les retours auditifs dont dispose le locuteur de sa propre voix et leurs caractéristiques psychoacoustiques sont donc des facteurs importants de l'adaptation de la parole dans le bruit dont nous aurons à tenir compte au niveau des protocoles expérimentaux.

2.1.1. Les différents retours auditifs de sa propre voix

Le locuteur dispose de 2 retours auditifs principaux pour évaluer sa propre voix : un retour auditif par voie aérienne, un autre par voie osseuse.

La principale voie de retour auditif pour la plupart des locuteurs en situation de communication quotidienne est celle de perception par l'oreille externe du son rayonné au niveau de la bouche et conduit par l'air jusqu'à l'oreille. Le trajet du son de la bouche à l'oreille est à la fois direct et indirect, grâce aux réflexions du son sur les parois de la salle ou les objets alentours. Ce retour aérien dépend donc de deux caractéristiques :

- la morphologie du visage du locuteur. Selon la forme des lèvres, la rondeur de la joue, ou encore l'angle et la forme du pavillon de l'oreille, la fonction de transfert modélisant le canal bouche-oreille ne va pas être identique d'un locuteur à l'autre.
- les caractéristiques de la salle (taille, forme, réverbération) ainsi que la place du locuteur dans cette salle. On voit ainsi souvent des chanteurs se déplacer vers des murs pour améliorer le retour de leur voix.

Le retour auditif peut être également interne par conduction osseuse et tendino-cartilagineuse, et ne dépend alors plus des conditions extérieures, ce qui rend son étude ou sa suppression très difficile. Le son peut être conduit du larynx jusqu'à l'oreille interne par différentes voies internes, que Tonndorf 1966^[354] regroupe en 3 catégories, selon que la vibration parvient à l'oreille externe, à l'oreille moyenne ou directement à l'oreille interne. La composante de la conduction relative à l'oreille externe résulte du fait que la vibration du crâne entraîne des déformations du canal auditif externe et ainsi de légères variations de pression (Von Békésy 1960^[370] ; Hood 1962^[155]). Enfin, certains auteurs suggèrent que les vibrations pourraient atteindre la cochlée directement (Tonndorf 1966^[354] ; Stenfelt et al. 2002^[330]). En temps normal, la contribution de la conduction osseuse a été mesurée comme équivalente à celle du retour auditif aérien (Von Békésy 1960^[370] ; Stenfelt et al. 2002^[330] ; Pörschmann 2000^[274]). Elle devient au contraire particulièrement importante lorsqu'il y a occlusion de l'oreille externe, par le port de protections auditives ou d'un casque fermé par exemple. Ce phénomène est appelé « effet d'occlusion ».

Enfin, nous devons mentionner l'existence d'un retour perceptif tactile et non plus auditif, de sa propre voix, par le biais de sensations vibratoires. Les chanteurs sont particulièrement entraînés à se baser autant voire davantage sur ce retour kinesthésique que sur leur retour auditif externe pour contrôler leur voix, car ils sont habitués à s'exprimer avec un accompagnement musical, dans des acoustiques de salles très différentes ou dans un chœur (Scotto di Carlo 194^[311], Ternström 1994^[344]).

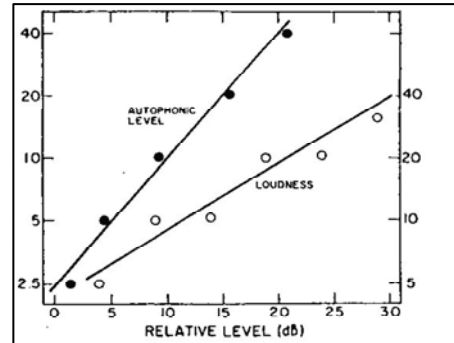
2.1.2. La perception de sa propre voix à partir de ces retours auditifs

La perception de sa propre voix est le meilleur indice dont le locuteur dispose pour estimer la façon dont son interlocuteur le perçoit. Pourtant, les perceptions intérieure et extérieure de la voix sont très différentes, au point qu'il est fréquent que des locuteurs ne se reconnaissent pas lorsqu'ils ont l'occasion d'écouter leur voix enregistrée. Ce phénomène a motivé plusieurs études à comprendre

les mécanismes de perception de sa propre voix et surtout les raisons de cette perception intérieure différente.

En 1961, Lane et al. 1961^[193] ont cherché à évaluer la courbe de perception de l'intensité de sa propre voix par un locuteur, appelée « échelle subjective de réponse autophonique ». Lorsque le locuteur pense doubler son intensité, sa voix n'est perçue que légèrement plus intense (d'un facteur 1.2) par un auditeur extérieur (cf. Figure 13).

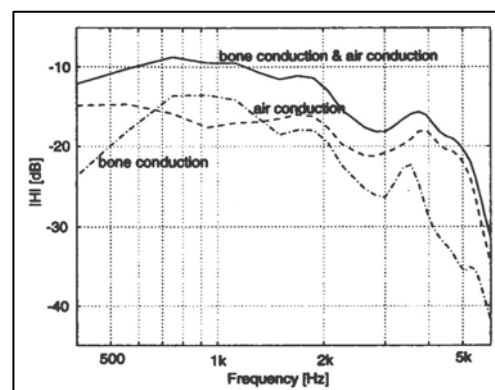
Figure 13. Echelle autophonique et échelle de sonie (D'après Lane et al. 1970^[194]).



Mais ce qui différencie la voix perçue par un auditeur extérieur, de la voix perçue en interne par le locuteur lui-même, n'est pas qu'une question d'intensité mais aussi de timbre. Le renforcement des basses fréquences dans la perception de sa propre voix provient de plusieurs facteurs :

- tout d'abord, un premier effet de filtrage a lieu lors de la conduction aérienne du son de la bouche vers l'oreille. Le rayonnement du son est directif en hautes fréquences et omnidirectionnel en basses-fréquences, ce qui implique que les fréquences aiguës sont atténuées vers l'arrière, tandis que les basses fréquences se propagent beaucoup mieux jusqu'au pavillon de l'oreille externe. La fonction de transfert bouche-oreille peut déjà être modélisée par un filtre passe-bas (cf. Figure 14).
- un deuxième filtrage important est dû à la conduction osseuse, atténuant les fréquences supérieures à 2kHz, ce qui explique le caractère très sourd de la voix perçue de l'intérieur en se bouchant les oreilles (Pörschmann 2000^[274] ; Von Békésy 1960^[370] ; Maurer et al. 1990^[234]). Au contraire, la conduction osseuse participe de façon plus importante à la perception de sa propre voix que le retour aérien pour les fréquences entre 700 et 1200 Hz, c'est-à-dire dans la zone du premier formant (cf. Figure 14).
- enfin, l'oreille moyenne dispose d'un réflexe de protection de l'oreille aux fortes intensités, appelé réflexe stapédien (ou « acoustic reflex » en anglais) qui est également activé automatiquement lors de la phonation (Olsen et al. 1976^[261]). Ce réflexe permet de réduire la sensibilité de l'oreille juste avant la phonation, et se traduit concrètement par une atténuation auditive au niveau des basses fréquences.

Figure 14. Représentation des différentes contributions du retour aérien ou de la conduction osseuse à la perception de sa propre voix en fonction des fréquences (D'après Pörschmann 2000^[274]). Les basses fréquences seraient davantage régulées à l'aide de la conduction osseuse, et les hautes-fréquences à l'aide du retour aérien.



Ces différents retours perceptifs peuvent expliquer l'influence différente qu'exerce un retour filtré de sa propre voix sur la régulation vocale, selon la bande de fréquences affectée par le filtrage. Ainsi, Garber et al. 1981^[99] ont montré dans une expérience que l'effet Sidetone (diminution de l'intensité vocale lorsque le retour aérien est amplifié) est diminué lorsque le retour auditif est également filtré par un passe-bas, tandis qu'il reste inchangé par un passe-haut. Ces auteurs en concluent que les retours auditifs aériens et osseux ne sont pas « additifs » mais complémentaires: les basses fréquences de la voix seraient principalement régulées à l'aide du retour par conduction osseuse, et

les hautes fréquences à l'aide du retour aérien (cf. Figure 14). Ainsi, un filtrage des basses fréquences du retour aérien n'aurait pas de grande influence sur la perception de sa propre voix, entraînant une absence d'effet Sidetone, tandis qu'un filtrage des hautes fréquences du retour aérien entraînerait un changement important de perception de sa propre voix, et donc un effet Sidetone.

2.2. Une régulation physiologique réflexe

Si la majorité des études s'accordent sur le fait de concevoir l'adaptation de la parole dans le bruit comme une régulation de la voix par l'audition, un débat existe depuis la mise en évidence de l'effet Lombard sur l'interprétation des mécanismes qui le gouvernent : certains le considèrent comme une régulation physiologique d'ordre réflexe face à l'atténuation du retour auditif de sa propre voix. D'autres l'envisagent comme une régulation cognitive et communicationnelle, motivée par le désir de compenser la dégradation de l'intelligibilité que le bruit ambiant induit pour l'interlocuteur.

De nombreuses études apportent des arguments pour considérer la régulation de la voix par l'audition comme d'ordre réflexe, ou du moins comme non volontaire : une telle régulation audio-phonatoire est en effet également observée chez des animaux (Sinnott et al. 1975^[318] ; Manabe et al. 1998^[231]) ou chez de jeunes enfants (Siegel et al. 1976^[316] ; Garber et al. 1980^[100], Amazi et al. 1982^[8]). Par ailleurs, les temps de réaction très courts sont considérés comme des preuves que seuls des mécanismes réflexes et bas niveau participent à la régulation de la voix (Leydon et al. 2003^[211]). Enfin, certains auteurs ont montré que cette régulation n'était pas possible à inhiber par la simple volonté (Pick et al. 1989^[271] ; Siegel et al. 1974^[315]). Un apprentissage de l'inhibition est cependant rendu possible en utilisant une modalité sensible autre que l'audition. Dans le cas de l'effet Lombard, cet apprentissage semble en réalité davantage correspondre à un déplacement de l'échelle des intensités vocales utilisées qu'à un apprentissage de l'inhibition : le sujet arrive effectivement à ne plus crier lorsqu'on l'expose à un bruit intense. Cependant, il se met également à parler anormalement doucement en situation calme (Pick et al. 1989^[271]).

La première interprétation neurologique réflexe de la boucle audio-phonatoire est apportée par Garde (cité par Husson 1962^[162]). Qualifiée de « cochléo-résonnante », cette théorie supporte l'idée que « la stimulation d'un noyau cochléaire au niveau du bulbe se diffuserait dans la substance réticulée puis dans les noyaux moteurs, pouvant augmenter le tonus glottique ». L'effet Lombard n'est alors pas seulement une réaction à l'atténuation du retour auditif, mais essentiellement dû à une hyperstimulation cochléo-résonnante par le bruit ambiant. Dans une expérience de régulation du timbre, Tomatis 1987^[352] observe que le sujet produit une voix dont le contenu spectral correspond à celui de la voix filtrée retournée au locuteur : par exemple, si les hautes fréquences sont filtrées dans son retour auditif, le locuteur a également tendance à atténuer ces hautes fréquences en production. Reprenant la théorie cochléo-résonnante, Tomatis en conclut que « la voix d'un sujet ne contient que les harmoniques que son oreille peut entendre ». Ces affirmations sont évidemment plus que discutables : un sourd complet ne pourrait alors absolument pas émettre de son. De plus, cette théorie s'appuie sur la théorie chronaxique de la phonation de Husson 1962^[162], depuis totalement infirmée par la communauté scientifique, qui supportait l'idée que la vibration des cordes vocales était commandée activement par des influx nerveux au niveau laryngé.

D'autres études neurophysiologiques menées sur des animaux décérébrés ont par la suite prouvé que des mécanismes purement neurologiques pouvaient suffire à induire un effet Lombard (Nonaka et al. 1997^[256]). Le développement récent de l'imagerie cérébrale a ouvert de nouvelles possibilités d'investigation. Il existe indéniablement une interaction réflexe entre la phonation et l'audition (Houde et al. 1998^[156] ; Eliades et al. 2003^[81]). En revanche, les mécanismes neuronaux sous-tendant le contrôle sensori-moteur et la fonction du système auditif pendant la phonation restent encore mal connus (Houde et al. 1998^[156] ; 2002^[157]). Il semblerait que les régions tempo-pariétales soient impliquées dans un système de régulation automatique de la parole par l'audition (Hashimoto et al. 2003^[136]).

2.3. Une régulation psychologique et communicationnelle

Plusieurs auteurs soutiennent que les locuteurs peuvent inhiber cette régulation automatique (Lane et al. 1961^[193], Tonkinson 1994^[353]) et que l'adaptation de la parole est d'abord une « régulation sociale », liée au désir d'être compris par l'interlocuteur, et non pas liée seulement au besoin de s'entendre soi-même pour arriver à contrôler sa parole (Lane et al. 1961^[193]). Ainsi, les modifications de la parole observées en environnement bruyant ne consistent pas seulement en une augmentation

de l'intensité vocale, mais en un ensemble de réorganisations acoustiques, phonétiques ou articulatoires. Il est difficile d'interpréter celles-ci comme les conséquences réflexes d'une diminution du retour auditif aérien, tandis qu'au contraire elles contribuent à une augmentation de l'intelligibilité pour l'auditeur (Junqua 1993^[171]). L'adaptation de la parole semble donc effectuée non pour soi, mais pour l'autre.

D'autres auteurs adoptent une conception intermédiaire en reconnaissant la contribution réflexe des retours sensoriels sur l'adaptation de la parole, mais en considérant que la recherche d'intelligibilité prime dans l'adaptation. Lane et al. 1961^[193] ont ainsi montré qu'une plus grande compensation de la parole était observée en parole spontanée par rapport à une tâche de lecture, témoignant selon eux du fait qu'une contribution réflexe existe mais que l'effet Lombard est avant tout provoqué par la nécessité que le locuteur a de se faire entendre et comprendre. De la même façon, Amazi et al. 1982^[8] ont montré que l'effet Lombard était également observé chez les enfants, identique quelle que soit la tâche de parole, tandis que l'effet Lombard est plus important chez les adultes lorsqu'ils doivent raconter une histoire que lorsqu'ils doivent la lire. Ces observations sont corroborées par le fait que la pente de la droite représentant l'intensité vocale d'un locuteur en fonction du niveau de bruit, souvent utilisée pour « quantifier » l'effet Lombard, est effectivement plus grande dans les études portant sur la parole spontanée ou en situation d'interaction, que dans les études portant sur de la lecture.

L'adaptation de la parole ne relève donc pas seulement du réflexe, mais fait également intervenir des mécanismes haut-niveau tels que la prédiction de l'intelligibilité pour l'auditeur, relativement à un contexte. Dans sa théorie de la mutualité ou H&H (pour Hypo&Hyper), Lindblom 1990^[217] propose que le locuteur adapte son effort articulatoire aux exigences de la situation de communication, c'est-à-dire au milieu de transmission (bruit, distance, acoustique de salle) ainsi qu'à l'interlocuteur (interaction sociale, personne malentendante, personne étrangère). Cette adaptation de l'effort ne se fait pas sur une unique dimension (par exemple l'intensité vocale) ni de façon nécessairement globale. Selon Jurafsky et al. 2001^[175] le locuteur réduit son effort sur les mots prévisibles par l'auditeur. De la même façon, Fonagy et al. 1966^[94] émettent l'hypothèse que l'allongement des voyelles par rapport aux consonnes dans la parole criée serait motivé perceptivement, par le fait que les voyelles sont les sons de la parole les plus audibles à grande distance et dans des environnements bruyants et qu'elles porteraient par conséquent selon eux la plus grande information à l'auditeur.

Il existe d'autres situations où le locuteur doit adapter son effort de production pour être compris par son interlocuteur, sans subir de perturbations au niveau de son retour auditif. Les modifications observées dans ces types de parole vont dans le même sens que ceux observés pour la parole Lombard. Ainsi, lorsque l'interlocuteur est éloigné, Traunmüller et al. 2000^[356] ont observé une augmentation de l'intensité et de la fréquence fondamentale, ainsi qu'un ralentissement du débit de parole lorsque la distance est très importante, la parole devenant alors « hachée » par syllabes et le spectre de la voix se déplaçant vers les fréquences médium. Une hyper-articulation et un ralentissement de la parole sont également observés dans la parole « claire » à destination d'une personne malentendante, étrangère, ainsi que dans la parole adressée à un enfant de bas âge (Dodane et al. 2006^[72]; Lindblom et al. 1992^[219]). Le ralentissement du débit n'est pas uniforme mais dépend des phonèmes et ne consiste pas seulement en un allongement des phonèmes mais également en l'insertion de pauses (Picheny et al. 1985^[269]). Dans ce mode de parole, les voyelles tendent à atteindre davantage les cibles formantiques qu'en parole conversationnelle et l'intensité des sons obstruents, en particulier des plosives, est plus importante, ce qui argumente en faveur d'un prototypage des indices de reconnaissance acoustiques des phonèmes. La parole claire et Lombard diffèrent au niveau du rapport d'intensité consonne/voyelle qui diminue dans le bruit tandis qu'il est augmente en parole claire.

Ces différentes modifications de la parole observées en parole Lombard comme dans les autres types de parole « hyper » ne sont pas seulement supposées être effectuées dans le but d'augmenter l'intelligibilité pour l'interlocuteur : Des études perceptives ont montré que dans la majorité des cas, ces différents types de parole sont plus intelligibles que la parole « normale » (Dreher et al. 1957^[77], Van Summers et al. 1988^[361], Tolhurst 1957, cité dans Nabelek 1990^[249]). Junqua 1993^[171] a également observé que la parole produite dans le bruit par des femmes est plus intelligible que celle des hommes (cela n'est pas contradictoire avec le fait que la parole « normale » des femmes soit moins intelligible que celle des hommes comme nous l'avons présenté au paragraphe 2.3.3.2). Picheny et al. 1985^[269] ont montré que certains locuteurs arrivaient mieux que d'autres à augmenter leur intelligibilité en parole claire. Il existe néanmoins quelques cas où l'augmentation de l'effort de production par le locuteur est contre-productive. Ainsi, Rostolland et al. 1973^[290] ont montré que l'intelligibilité décroît lorsque l'effort vocal atteint un niveau de parole très fortement criée. L'énergie

acoustique est certes accrue, mais l'information phonétique est alors trop déformée. Ce phénomène a également été observé en parole claire où les fricatives deviennent moins intelligibles qu'en parole normale, à l'inverse des autres phonèmes (Picheny et al. 1985^[269]).

Ces différentes observations nous conduisent donc à envisager l'interaction avec un interlocuteur, sa distance ou encore la prévisibilité du discours comme des facteurs pouvant influencer l'adaptation de la parole en environnement bruyant, et dont nous devons par conséquent prendre en compte au niveau de la mise au point de protocoles expérimentaux pour caractériser les différences interindividuelles de stratégies d'adaptation.

2.4. Une adaptation plus complexe qu'une simple régulation

Nous avons présenté jusqu'à maintenant la conception majoritaire de l'adaptation de la parole en environnement bruyant sous la forme d'une régulation de la production par l'audition, que celle-ci soit de l'ordre physiologique réflexe ou de l'ordre cognitif et communicationnel. Cependant, d'autres travaux sur la parole compensatoire ou sur l'adaptation sensori-motrice en général laissent penser que l'adaptation de la parole dans le bruit serait plus complexe qu'une simple régulation.

En effet, de nombreuses études mentionnent que l'entraînement du sujet pour améliorer une compensation (Jones et al. 2003^[169] ; Savariaux et al. 1999^[302]) ou apprendre à inhiber une compensation (Pick et al. 1989^[271]), n'a pas seulement des conséquences pendant l'exposition à la perturbation, mais également après, lors du retour à une situation normale. Ce phénomène est désigné dans la littérature anglophone par « after-effect » ou « over-compensation ». Par ailleurs, l'apprentissage d'une nouvelle stratégie ne se limite pas seulement au problème particulier sur lequel le sujet s'est entraîné mais est également étendu à d'autres problèmes similaires (Baum et al. 1997^[22]). L'adaptation de la parole ne semble donc pas être réduite à une régulation mais être un réel processus d'apprentissage, ou de « recalibration neuromotrice » vers un nouvel état d'équilibre.

De plus, certaines observations laissent penser que l'adaptation n'est pas seulement un mécanisme de réglage des paramètres d'un système d'action, mais peut consister en une réorganisation globale de l'action, voire en un changement de système. C'est le cas par exemple lorsqu'on passe de la marche à la course (Berthoz 1997^[25]). L'adaptation correspond alors davantage à la recherche d'une solution adaptée qu'à un simple ajustement.

Ces réflexions nous amènent à remettre en question une description unique, graduelle ou linéaire du comportement d'adaptation en fonction de l'importance croissante de la perturbation imposée au système (dans notre cas le bruit ambiant). Une modélisation catégorielle de l'adaptation en fonction des individus et des caractéristiques de la perturbation pourrait s'avérer plus pertinente.

3. LES METHODOLOGIES D'ETUDE DE L'ADAPTATION DE LA PAROLE DANS LE BRUIT

Nous avons vu comment plusieurs champs disciplinaires se sont intéressés à certains aspects de l'adaptation de la parole dans le bruit et ont cherché à en comprendre les mécanismes par des modèles différents, en particulier sous la forme d'une régulation physiologique réflexe ou plutôt sous la forme d'une régulation « sociale » motivée par la recherche d'intelligibilité. Ces divers intérêts pour l'effet Lombard et les hypothèses d'explication des mécanismes qui le sous-tendent, expliquent les différences de protocole observées entre les différentes études, en particulier en ce qui concerne l'environnement bruyant et la situation de communication.

3.1. L'environnement bruyant

Le tableau 2 de l'annexe Bib2 récapitule les différents types de bruit, niveaux de bruit et moyens de restitution du bruit utilisés dans plusieurs protocoles d'études portant sur l'effet Lombard. La grande majorité est menée en laboratoire. Elles utilisent alors la plupart du temps du bruit blanc large bande (Egan 1972^[80]), filtré par bandes de fréquences (Mokbel 1992^[245] ; Egan 1972^[80]) ou avec une enveloppe spectrale comparable à un bruit de conversations (Siegel et al. 1974^[315] ; Egan 1972^[80]). Les études plus récentes utilisent maintenant également des bruits plus réalistes, en particulier des bruits de conversations (Zeiliger et al. 1994^[382] ; Sodersten et al. 2005^[322] ; Davis et al. 2006^[62]) ou le bruit ambiant à l'intérieur d'un habitacle automobile (Egan 1972^[80] ; Mokbel 1992^[245] ; Boril et al. 2005^[32] ; Lee et al. 2004^[204]). Des études in situ sont également allées à la rencontre de personnes confrontées quotidiennement au bruit sur leur lieu de travail : institutrices de maternelle (Sodersten et al. 2002^[321] ; Dieroff et al. 1966^[70]), aiguilleurs ferroviaires (Dohalska et al. 2000^[73]), marins dans la salle des machines (Borsuk et Klajman, 1967), ouvrier dans une usine (Pruszewicz et al. 1974^[277]).

Les niveaux de diffusion du bruit en laboratoire sont très difficiles à comparer entre les différentes études, tantôt exprimés en dB SPL, en dB (A), en dB (C), ou encore sans précision. Ils vont globalement de 60 à 100 dB SPL. La restitution du bruit se fait très largement à l'aide d'un casque. Quelques études récentes commencent à envisager d'autres solutions (Ternström et al. 2002^[346] ; Ternström et al. 2006^[345] ; Mixdorff et al. 2006^[244]) dont nous parlerons au chapitre 3.

3.2. La situation de communication

L'étude audiolinguistique de l'effet Lombard s'intéresse à sa part réflexe physiologique, sans impliquer le sujet au niveau cognitif, tandis que l'étude phonologique de l'effet Lombard se base plutôt sur le fait que l'adaptation au bruit est principalement motivée par la compensation de l'intelligibilité dégradée pour l'interlocuteur, prenant en compte la situation de communication et les représentations mentales du sujet. Ces deux conceptions opposées impliquent des méthodologies différentes pour accéder à des objets d'étude différents, bien que regroupés sous la même appellation « effet Lombard » :

- dans le premier cas, le locuteur peut être seul, sa tâche est avant tout d'émettre des *sons vocaux* et pas nécessairement de la parole. La lecture de listes de mots convient alors tout à fait. Son *système auditif* est assourdi à l'aide de bruit intense, bien calibré au niveau de son contenu spectral (bruit blanc filtré par bandes ou large bande) et de son intensité. L'exposition au bruit est alors considérée comme équivalente à une atténuation du retour auditif.
- dans le deuxième cas, le locuteur s'adresse à un interlocuteur réel ou imaginaire. Sa tâche est de lui transmettre des informations (mots, phrases). L'exposition au bruit sert principalement à dégrader l'intelligibilité entre les interlocuteurs (réels ou imaginaires), et est jugée comme équivalente à un allongement des distances entre les interlocuteurs (Traunmüller et al. 2000^[356]).

Un biais méthodologique important provient du fait que de nombreuses études continuent à utiliser des protocoles dérivés de l'audiologie pour étudier l'effet Lombard d'un point de vue phonologique. Nous nous pencherons plus en détail sur ce problème au chapitre 4.

Enfin, lorsque l'effet Lombard est étudié non pas comme une fin mais comme un moyen pour contraindre le locuteur à un effort vocal, la parole Lombard est alors souvent confondue avec d'autres types de paroles à forte intensité comme la voix criée ou la parole à distance.

3.3. La tâche de parole

Les études in situ portent sur de la parole spontanée, mais quelques études en laboratoire mettent également en jeu des interlocuteurs en situation de dialogue totalement libre (Korn 1954^[185] ; Pick et al. 1989^[271]). Elles n'examinent alors que des paramètres acoustiques très généraux mais ne peuvent pas examiner de paramètres plus précis qui exigeraient un contrôle expérimental. Sinon, dans la grande majorité des études, il est demandé au locuteur de lire des listes de mots ou de phrases prédéfinies, sans interaction avec un interlocuteur (Junqua 1993^[171] ; Castellanos et al. 1996^[46] ; Welby 2006^[375]). Il existe cependant quelques études dans lesquelles il est demandé aux locuteurs de lire ces listes « comme si » ils s'adressaient à un interlocuteur (Ternström et al. 2002^[346] ; Kim 2005^[180]), ou bien dans lesquelles un interlocuteur est effectivement présent et donne un retour sur le succès ou non de la communication (Papon 2006^[265]).

En anglais, les listes sont souvent constituées de mots à deux syllabes accentuées de façon équivalente (ou "spondées", Hirsh et al. 1952^[148]), de vocabulaire aéronautique ou de listes de chiffres (Van Summers et al. 1988^[361] ; Junqua 1996^[172]), ou encore de commandes vocales à destination d'interfaces homme-machine (Kim 2005^[180]). Le vocabulaire de Miller et al. 1955^[242] ou les phrases phonétiquement équilibrées de 1969^[2] sont également largement utilisées dans les expériences de production ou de perception de la parole dans le bruit. En français plus particulièrement, la base DB bruit est constituée de logatomes CVC, de listes de nombres et de phrases (Zeiliger et al. 1994^[382]). Les tests d'audiométrie et d'intelligibilité se basent souvent sur les listes de Fournier, constituées de mots mono et bi-syllabiques et de phrases phonétiquement équilibrées (Portmann et al. 1959^[275]).

3.4. Le traitement des données expérimentales

Le but de l'étude de la parole Lombard, que ce soit à des fins de diagnostic, de compréhension ou de modélisation, entraîne un traitement des données très différent :

- dans un cas, l'individu est considéré comme type, et des traitements statistiques sont effectués sur des paramètres issus de grandes bases de données, à la recherche d'un profil moyen. La variabilité individuelle prend alors une connotation négative puisqu'elle perturbe la modélisation.
- dans un autre cas, la considération d'individus particuliers permet de distinguer les traits communs entre les locuteurs, des caractéristiques qui relèvent de stratégies individuelles. Ce genre d'étude ne peut pas être totalement automatisé et est effectué par conséquent sur des corpus de plus petite taille.

4. POSITIONNEMENT DE NOTRE ETUDE PAR RAPPORT A CES DIFFERENTES THEORIES ET METHODOLOGIES

Notre problématique est de caractériser la façon dont un locuteur s'adapte à un environnement bruyant et gère l'effort vocal requis par cette adaptation. Nous avons d'abord rappelé les mécanismes mis en jeu lors de l'adaptation de la parole en environnement bruyant. Nous avons vu en particulier que l'adaptation de la parole en environnement bruyant, qualifiée d' « effet Lombard », est imputée selon les études à un mécanisme réflexe de régulation de la voix par l'audition et/ou à une volonté de la part du locuteur de compenser la dégradation de l'intelligibilité induite par le bruit ambiant.

4.1. Bilan sur les mécanismes gouvernant l'adaptation de la parole dans le bruit et sur les facteurs qui les influencent

Nous faisons ici l'hypothèse que l'adaptation de la parole en environnement bruyant fait intervenir une régulation réflexe du niveau d'intensité de la voix mais elle est également et surtout motivée par la volonté d'être compris par un interlocuteur. Nous explorerons au cours de la thèse plusieurs points pour apporter de nouveaux arguments à la validité de cette hypothèse.

Nous ne rentrerons pas plus dans le détail neurologique de la régulation réflexe de la voix par l'audition et considérerons seulement que les retours auditifs aériens et par conduction osseuse conditionnent cette régulation. Aussi, nous serons attentifs à la façon dont la situation expérimentale peut affecter ces retours. Nous prendrons également en considération l'influence probable des caractéristiques spectrales du bruit.

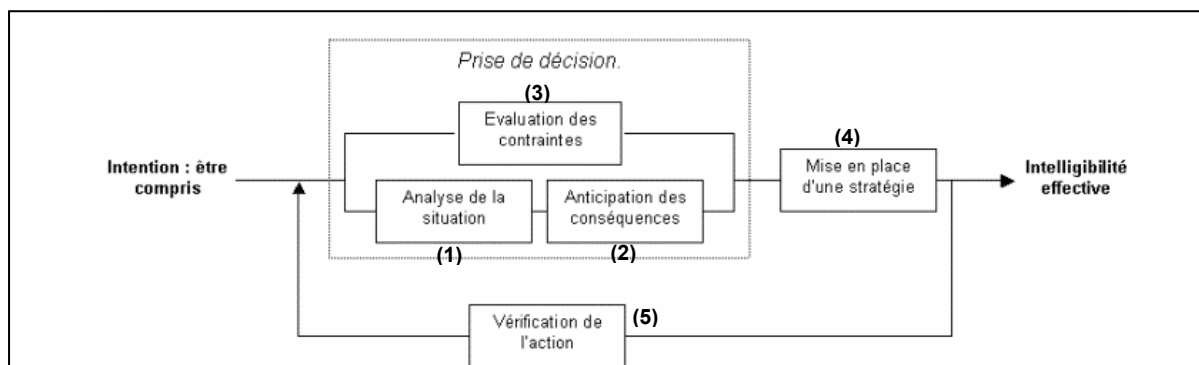


Figure 15. Schéma récapitulatif de notre conception de l'adaptation de la parole en environnement bruyant, sous forme d'une résolution de problème par boucle de régulation fermée. Ce schéma est explicité dans le paragraphe suivant.

En ce qui concerne la part communicationnelle de l'adaptation au bruit, nous effectuons l'hypothèse qu'elle fait intervenir le fonctionnement cognitif du locuteur à plusieurs niveaux (cf. Figure 15):

- (1) Tout d'abord au niveau de la perception et de l'analyse de la situation de communication, faisant intervenir ses représentations mentales afin d'évaluer de quelles modalités sensorielles (auditives, visuelles) il dispose pour interagir avec l'interlocuteur, sa distance, les caractéristiques de l'environnement bruyant dans lequel il est immergé, le niveau de compréhension de la langue par l'interlocuteur ou encore ses éventuels problèmes de déficit auditif.
- (2) Ensuite, au niveau de la comparaison de cette situation à des expériences passées semblables stockées en mémoire. Cette catégorisation de la situation de communication peut alors lui permettre d'anticiper la façon dont l'interlocuteur peut le comprendre dans cette situation, ainsi que la façon dont il doit s'adapter pour maintenir un niveau correct d'intelligibilité.
- (3) Enfin au niveau de la connaissance et de l'évaluation des contraintes limitant son action. Ces contraintes peuvent être de plusieurs ordres :
 - la première contrainte réside dans le principe général de limiter autant que possible l'énergie dépensée pour l'adaptation.

- le locuteur est également vocalement limité dans sa capacité d'adaptation par son anatomie (forme du conduit vocal, longueur des cordes vocales, forme des lèvres), ses possibilités physiologiques (ouverture de la mâchoire), sa technique vocale (gestion du souffle ou des résonances), son hygiène de vie (tabagisme, alcool) ou sa santé vocale (affection de la sphère ORL ou lésion organique des cordes vocales).
 - Ces possibilités vocales sont elles-mêmes limitées par les sensations de gêne, de confort ou de fatigue ressenties par le locuteur.
 - les contraintes sociales limitent également l'adaptation du locuteur. Pour éviter que la parole ne soit jugée comme inconvenante ou anormale, certaines caractéristiques telles que la F0 ou la qualité vocale doivent rester dans une certaine marge imposée par des conventions sociales implicites et partagées par le locuteur et l'interlocuteur.
 - les possibilités d'adaptation sont encore limitées par les contraintes phonologiques. Les indices acoustiques et articulatoires de reconnaissance des segments phonétiques ne peuvent pas s'écarter au-delà d'une certaine limite, au risque que la parole ne soit plus intelligible.
- (4) A partir de l'intégration de ces différentes informations (1, 2 et 3), le locuteur peut alors prendre une décision et mettre en œuvre une stratégie de compromis entre les objectifs de l'action, les exigences de la situation et les contraintes à cette action.
- (5) Enfin, une dernière étape de cette régulation consiste à vérifier l'action et la réajuster par rapport aux conséquences anticipées de l'action. Cette vérification est rendue possible dans une certaine mesure par la perception de sa propre voix (malheureusement perturbée) permettant d'estimer la façon dont un auditeur extérieur peut la percevoir, mais aussi par l'interprétation des indices visuels ou verbaux émis par l'interlocuteur sur sa compréhension ou non du message.

Cette hypothèse sur les mécanismes mis en jeu dans l'adaptation de la parole dans le bruit nous conduit à soigner la prise en compte des facteurs qui les influencent dans la mise au point de nos protocoles expérimentaux. Nous avons vu que ces facteurs sont principalement

- les retours auditifs que le locuteur a de sa propre voix
- la situation de communication, intégrant le bruit ambiant, mais également l'interlocuteur.

Par conséquent, nous serons attentifs dans la suite de cette thèse à bien contrôler les conditions expérimentales d'immersion du locuteur dans le bruit, afin de s'approcher d'une situation la plus naturelle possible pour l'auditeur et surtout de ne pas induire d'atténuation auditive supplémentaire à celle imposée par le bruit. La validité d'utiliser un casque pour immerger le locuteur dans le bruit, ainsi que la proposition de nouvelles méthodologies, fera ainsi l'objet du chapitre 4. Nous serons également particulièrement soucieux de prendre en compte la présence d'un interlocuteur (à une certaine distance, disposant de certaines modalités sensorielles pour interagir avec le locuteur, et d'une certaine prévisibilité quant au discours du locuteur) dans la situation de communication, ainsi que son influence interactive (retours sur le succès de la communication, etc.). Ces considérations feront l'objet du chapitre 4.

4.2. Hypothèses sur les différences interindividuelles d'adaptation de la parole dans le bruit

Nous avons vu au paragraphe 1 les différentes manifestations observables de l'adaptation de la parole dans le bruit déjà caractérisées par les études antérieures. Cependant, ces travaux ont surtout cherché à recenser objectivement ces modifications, dans le but par exemple de les intégrer à des modèles de langage pour les systèmes de reconnaissance automatique de la parole, mais pas tant de les comprendre en terme de stratégie de communication du locuteur.

Notre hypothèse est justement que les différences interindividuelles d'adaptation de l'effort vocal à une situation de communication se situent au niveau de l'adoption de stratégies de communication différentes pour faire face à cette situation perturbée, et que ces différentes stratégies ne sont pas toutes de la même efficacité en ce qui concerne la répartition des efforts sur le larynx et le niveau d'effort dépensé pour un même niveau d'intelligibilité par l'interlocuteur.

Aussi, notre démarche ne consistera pas seulement à examiner de façon paramétrique toutes les différences acoustiques, articulatoires ou glottiques existant entre les locuteurs lorsqu'ils s'expriment en environnement bruyant mais plutôt à envisager différentes hypothèses de stratégies de communication pouvant être adoptées dans le bruit. Nous examinerons alors si ces stratégies sont propres à certains individus ou partagés par tous.

Puisqu'un mécanisme principal de l'adaptation de la parole dans le bruit semble consister en la compensation de la dégradation de l'intelligibilité induite par le bruit ambiant (Lane et al. 1971^[192]), nous avons recensé dans la littérature les différents facteurs dégradant ou améliorant l'intelligibilité de la parole en perception, de façon à élaborer des hypothèses de stratégies de communication potentiellement adoptées dans le bruit (donc du point de vue de la production) que nous testerons au cours des chapitres suivants.

En premier lieu, le rapport signal sur bruit est déterminant sur l'intelligibilité. La dégradation de l'intelligibilité d'enregistrements audio par leur mixage avec du bruit est un paradigme classique pour réaliser des tests perceptifs (Sumbly et al. 1954^[335], Nicely et al. 1955^[255]).

L'intelligibilité de la parole peut être également affectée de façon différente selon le contenu spectral du bruit. Un bruit de conversations aux caractéristiques semblables à la parole du locuteur dégrade davantage son intelligibilité que d'autres types de bruit, en particulier si les conversations à l'intérieur de ce bruit restent intelligibles et ajoutent un effet de compétition lexicale pour l'auditeur. On connaît déjà l'existence de stratégies de renforcement spectral autour de 3kHz chez les chanteurs pour « émerger » par rapport à la masse sonore d'un orchestre (Sundberg 1972^[337]; Oliveira Barrichelo et al. 2001^[260]).

On peut donc envisager l'existence de stratégies « acoustiques » chez les locuteurs immergés dans un bruit intense, visant à émerger du bruit ambiant par la préservation du rapport signal sur bruit, ou par le renforcement spectral de la voix dans des zones pertinentes (zone sensible de l'oreille, zone où l'énergie spectrale du bruit ambiant est la plus faible ou au contraire la plus forte). Nous examinerons ces différentes hypothèses au cours du chapitre 6.

L'acoustique d'une salle a également montré des effets sur l'intelligibilité puisqu'une réverbération trop forte peut entraîner un recouvrement des indices acoustiques spectro-temporels et perturber la reconnaissance des segments (Cox et al. 1987^[59]; Nabelek 1990^[249]). De même, tous les phonèmes ne sont pas perturbés de la même façon par le bruit et par différents types de bruit. L'accélération du débit de parole perturbe également l'intelligibilité car elle restreint à la fois le temps de récupération de l'information par l'auditeur, mais surtout introduit des réorganisations temporelles à l'intérieur des segments et ne permet plus au locuteur d'atteindre les cibles articulatoires. Enfin, la disponibilité de la modalité visuelle accroît considérablement l'intelligibilité de la parole, en aidant à discriminer 40 à 60 % des phonèmes d'une langue et 10 à 20 % des mots (Schwartz et al. 2004^[309]). Sumbly et al. 1954^[335] ont ainsi montré que l'ajout de la modalité visuelle était équivalent à une augmentation du rapport signal sur bruit de 15dB.

Ces différentes observations laissent envisager que les locuteurs puissent adopter dans le bruit des stratégies de renforcement de certains indices spectro-temporels et/ou visuels des segments pour augmenter leur intelligibilité. Nous explorerons cette possibilité dans le chapitre 7.

La prosodie joue également un rôle déterminant dans l'intelligibilité, en particulier par sa fonction démarcative permettant à l'auditeur de segmenter le flux de parole en unités de sens. Des études antérieures ont déjà montré que le renforcement de certains de ces indices de segmentation améliorait l'intelligibilité en perception (Christophe 1993^[52]; Bagou et al. 2006^[20]; Dodane et al. 2006^[72]; Welby 2003^[373]) et étaient renforcés en production dans certains types de parole « hyper » (Dodane et al. 2006^[72]; Welby 2006^[375]).

Il est donc fort probable que les locuteurs adoptent des stratégies semblables dans le bruit. C'est ce que nous examinerons dans le chapitre 8.

Pour finir, l'intelligibilité ne dépend pas seulement des indices audio ou visuels extérieurs, mais est fortement influencée par les mécanismes descendants (top-down) de traitement de ces indices par le sujet. En particulier, l'auditeur est capable de prévoir ou de reconstituer le message à partir des indices perçus et du contexte de communication (Dohalska et al. 2000^[73]). Ainsi, l'intelligibilité augmente avec le nombre de syllabes des mots, diminue lorsque la taille du vocabulaire possible augmente (Howes 1957^[159]), augmente lorsqu'un mot est plus probable par rapport au contexte. Des études antérieures ont déjà montré que les locuteurs n'adaptent pas seulement leur effort à la situation de communication, comme le propose Lindblom 1990^[217], mais réduisent également leur effort sur les mots de leur discours n'apportant pas d'information « nouvelle » ou étant facilement reconstituables par l'auditeur (Gravano et al. 2006^[122]).

Nous envisageons par conséquent qu'il puisse exister dans le bruit des stratégies de renforcement plus ou moins important des différents constituants de l'énoncé par rapport à leur prévisibilité ou au

degré d'information qu'ils apportent à l'interlocuteur. Nous testerons cette hypothèse au cours du chapitre 9.

Dans cette thèse, nous nous placerons toujours du point de vue de la production de la parole, en cherchant à déduire les stratégies de chaque locuteur pour communiquer dans le bruit à partir de leurs comportements observables. Nous prenons en compte la perception par l'interlocuteur dans le sens où nous envisageons la production du locuteur comme celle d'indices audibles et visibles son attention. L'adaptation ne réside pas tant dans la modification de la production du locuteur que dans la façon dont cette modification influence l'intelligibilité perçue par l'auditeur. Ce positionnement a des conséquences tant au niveau de l'interactivité des tâches de parole dans les protocoles expérimentaux que de la représentation des données mesurées, à l'aide d'unités perceptives plutôt que d'unités physiques (par exemples des barks plutôt que des hertz pour les modifications formantiques), ou à l'aide d'unités relatives plutôt que d'unités absolues lorsque cela est davantage pertinent du point de vue de la perception d'un paramètre (par exemple la fréquence fondamentale sera exprimée en tons plutôt qu'en Hertz).

Chapitre 3 :

Influence de la méthode d'immersion dans le bruit sur l'adaptation de la parole

- 1. LES EVENTUELS PROBLEMES LIES AU PORT
D'UN CASQUE**
- 2. RECHERCHE D'UNE METHODE DE
DEBRUITAGE**
- 3. COMPARAISON DE TROIS METHODES
D'IMMERSION DU LOCUTEUR DANS LE BRUIT**

Nous avons vu au chapitre précédent que l'adaptation de la parole en environnement bruyant est en partie liée à la régulation automatique de la voix à partir du retour auditif. Au niveau des protocoles d'étude de l'effet Lombard, il nous semble par conséquent important de bien contrôler le retour auditif dont le locuteur dispose de sa propre voix, non seulement en intensité mais également au niveau des contributions aérienne et osseuse de ce retour. C'est pourquoi, l'usage d'un casque pour immerger le locuteur dans le bruit nous semble discutable. Au-delà du fait que l'immersion du locuteur dans l'environnement sonore par le biais d'un casque s'éloigne d'une situation écologique, le port du casque induit surtout une atténuation et une modification spectrale du retour auditif pouvant avoir des conséquences sur l'adaptation du locuteur, et donc introduire un biais expérimental non négligeable.

Pour pouvoir tester cette hypothèse, il est donc nécessaire de pouvoir comparer cette condition expérimentale avec une situation où le retour auditif du locuteur est géré de façon écologique, c'est-à-dire dans une situation bruyante réelle, ou reconstituée en laboratoire à l'aide de haut-parleurs. Le problème avec cette alternative est que l'enregistrement audio de la parole du locuteur se trouve fortement bruité par le bruit ambiant, rendant alors impossible toute analyse acoustique précise, en particulier du contenu spectral ou de l'intensité vocale. Cet obstacle a d'ailleurs été la principale motivation des études antérieures à utiliser un casque pour pouvoir malgré tout étudier les caractéristiques acoustiques et phonétiques ce type de parole. Cependant, si cet obstacle paraissait insurmontable par le passé, de nouvelles techniques mécaniques, électroacoustiques ou numériques semblent maintenant pouvoir être mises en œuvre pour débruiter ces enregistrements en conservant un signal de parole de très bonne qualité. Nous présenterons dans ce chapitre l'implémentation et la comparaison de plusieurs de ces techniques, afin de faire le choix d'une méthodologie appropriée et de proposer une alternative à l'immersion du locuteur dans le bruit par le biais d'un casque. Nous examinerons ensuite l'influence du port du casque sur l'effet Lombard, par comparaison à une restitution du bruit par haut-parleurs.

1. LES EVENTUELS PROBLEMES LIES AU PORT D'UN CASQUE

1.1. Perception de l'environnement sonore

La perception d'un environnement sonore dépend de ses caractéristiques d'intensité, de contenu spectral, mais également de la localisation des sources sonores, de la sensation d'immersion et d'espace (Guastavino 2003^[126] ; Vogel 1999^[369] ; Maffiolo 1999^[230] ; Raimbault 2002^[278] ; Dubois et al. 2001^[78]). C'est pourquoi la prise de son d'un environnement sonore n'est pas sans conséquence sur sa perception par l'auditeur. De même, la reconstitution d'un environnement sonore en laboratoire est influencée par l'acoustique de la salle, par la qualité, le nombre et la position des enceintes (Guastavino 2003^[126] ; Polack et al. 1988^[273]) en encore par les caractéristiques électroacoustiques du casque (Larcher et al. 2001^[195] ; Briolle et al. 1993^[34]). C'est pourquoi la prise de son d'un environnement et sa restitution en laboratoire doivent pouvoir conserver ses caractéristiques, d'autant plus lorsqu'il s'agit d'une étude portant sur la perception de cet environnement, ou sur un phénomène influencé par cette perception.

Plusieurs études ont comparé différents moyens d'enregistrer et de restituer les ambiances sonores, dans le but de recréer des situations écologiquement valides en laboratoire. En particulier, Hugonnet et al. 1995^[161] préconisent pour la prise de son d'ambiances sonores en intérieur l'utilisation de deux microphones cardioïdes à 20 cm d'écart avec un angle d'ouverture d'une centaine de degrés. Pour eux, la stéréophonie d'intensité fournit des images bien localisées mais ne rend pas suffisamment compte de l'espace et de la profondeur. Selon ces auteurs, des configurations omnidirectionnelles ou ORTF sont pour eux adaptées à des prises de son d'intérieurs ou de sources fixes. Maffiolo 1999^[230] et Vogel 1999^[369] ont exploré la validité écologique de plusieurs techniques de prise de son pour des environnements en extérieur, caractérisés par une faible réverbération, la présence de vent et de sources en mouvement. Ils ont montré que la configuration perçue comme la plus naturelle par les auditeurs consiste en 2 microphones cardioïdes électrostatiques écartés de 60 cm avec un angle de 100°. Amphoux 1991^[9] utilise quant à lui une tête artificielle, ou place des microphones au niveau de ses oreilles afin d'enregistrer des ambiances urbaines. Guastavino et al. 2005^[127] se sont intéressés plus particulièrement à la prise de son et à la restitution des basses fréquences, et ont montré qu'une prise de son et une restitution ambisonique augmentent la sensation d'immersion du sujet, au détriment de la localisation des sources. Enfin, Polack et al. 1988^[273] ont également observé que la

sensation d'immersion est dépendante de l'intensité de l'environnement et qu'à très fort niveau sonore, la place et le nombre d'enceintes n'a plus d'influence significative sur la perception d'intensité et sur la sensation d'immersion.

L'écoute au casque n'est pas sans conséquence sur la perception d'un environnement sonore. L'effet le plus important concerne la localisation de l'image auditive : Avec une restitution stéréophonique classique, l'auditeur perçoit une image sonore « au milieu » de sa tête, entre les deux oreilles, lorsque le son est restitué par casque, tandis que la source sonore est localisée dans l'espace plan devant l'auditeur lorsque le même son est restitué par haut-parleurs. De nombreux travaux de recherche en réalité virtuelle s'intéressent depuis plusieurs années à développer des techniques de spatialisation pour l'écoute au casque (Blum et al. 2004 ^[27]).

1.2. Perception de l'interlocuteur

Le port d'un casque n'a pas seulement une conséquence sur la perception du son restitué dans le casque, mais en a une également sur les sons réels à l'extérieur du casque. Selon le modèle de casque (ouvert, fermé, plus ou moins isolant...), l'atténuation du son extérieur peut aller de 2 à 30 dB, et concerne davantage les hautes fréquences que les basses fréquences. La communication peut être gênée du fait que l'on perçoit moins bien l'interlocuteur.

1.3. Retour auditif de sa propre voix

De la même façon, le port d'un casque atténue et filtre le retour aérien que le locuteur a de sa propre voix. Nous avons vu dans le deuxième chapitre que l'effet lombard est en partie interprété comme un phénomène automatique de régulation de l'intensité vocale face à une atténuation du retour auditif. C'est pourquoi il est envisageable que le port d'un casque induise en soi un effet Lombard supplémentaire, même sans restitution de bruit.

Par ailleurs, le port d'un casque provoque également une occlusion du conduit auditif d'autant plus importante que le casque est collé à l'oreille du locuteur. Nous avons vu au chapitre 3 que l'occlusion du conduit auditif entraîne un renforcement du retour auditif par conduction osseuse. Le retour auditif que le locuteur a de sa propre voix n'est donc pas seulement atténué mais modifié, au bénéfice des basses fréquences. Nous avons vu au chapitre 2 que la modification spectrale du retour de sa propre voix peut également entraîner une régulation automatique de la composition spectrale de la parole. Il est donc également possible que le port du casque entraîne en soi une adaptation du locuteur à la modification spectrale du retour de sa propre voix, constituant un biais à la description de l'effet Lombard.

1.4. Modifications conséquentes de la parole

Plusieurs études ont montré que les locuteurs baissent le niveau de leur voix de 1 à 6 dB lorsqu'ils parlent dans du bruit en portant des protections auditives (Frank et al. 2003 ^[97], Kryter 1946 ^[188]). Ce phénomène est interprété par le fait que les protections auditives atténuent le bruit ambiant extérieur et par la même occasion l'effet Lombard qu'il provoque. Cependant, d'autres effets semblent se combiner, car des modifications sont également observées lorsque les locuteurs parlent avec des protections auditives dans le silence : Frank et al. 2003 ^[97] observent que la voix est moins forte lorsque le locuteur porte des protections auditives, et que le barycentre spectral de sa parole est plus bas, ce qui est pour eux le signe d'un effet Sidetone induit par l'amplification du retour auditif interne par effet d'occlusion du conduit auditif et d'un renforcement de la perception des basses fréquences de sa propre voix. Au contraire, Kryter 1946 ^[188] observe une augmentation de l'intensité vocale dans le silence lorsque le sujet porte des protections auditives, ce qui serait plutôt le signe que l'atténuation du retour auditif aérien provoque un effet Lombard. Enfin, Martin et al. 1976 (cités dans Frank et al. 2003 ^[97]) n'observent pas de différence significative au niveau du contenu spectral de la parole avec ou sans protection auditive. La nature des protections auditives utilisées dans ces différentes études peut expliquer la différence des effets observés.

2. RECHERCHE D'UNE METHODE DE DEBRUITAGE

Les différents points soulevés précédemment laissent supposer que l'immersion des locuteurs dans le bruit par le biais d'un casque puisse introduire un biais dans l'étude de l'effet Lombard. Afin de tester la validité de cette méthode, nous avons cherché à comparer la parole Lombard produite lorsque le locuteur est immergé dans le bruit par le biais d'un casque et lorsqu'il est immergé dans le bruit de façon plus « naturelle », par exemple par le biais de haut-parleurs.

Mais cette alternative soulève un problème de taille, qui a justement motivé les auteurs d'études antérieures à utiliser un casque dans leurs protocoles : lorsque le bruit est diffusé par haut-parleurs, il se superpose à la parole sur l'enregistrement, perturbant les mesures acoustiques lorsqu'il est diffusé à faible ou moyenne intensité, et les rendant totalement impossibles à forte intensité. Ainsi, les bruits non stationnaires perturbent davantage les mesures acoustiques que les bruits stationnaires, et davantage les voix de femmes que celles des hommes (Deliyski et al. 2005^[68]). De façon générale, le NCVS, cité dans Deliyski et al. 2005^[68], recommande un rapport signal sur bruit (RSB) d'au moins 15 dB sur l'enregistrement pour réaliser des analyses valides. Carson et al. 2003^[45] recommandent plus de 25 dB(A) de RSB et Deliyski et al. 2005^[68] un RSB de 42dB pour avoir 99% de précision.

C'est pourquoi nous avons avant toute chose cherché des méthodes pour atténuer au maximum le bruit ambiant des enregistrements de parole, tout en gardant un signal acoustique le plus fidèle possible à la voix, afin que les analyses acoustiques représentent de façon rigoureuse la production de la parole. Nous avons recensé plusieurs techniques mécaniques, électroacoustiques ou numériques que nous nous proposons d'exposer ci-dessous. Puis nous avons implémenté et comparé trois d'entre elles afin d'examiner leur performance et surtout leur adéquation par rapport aux exigences de notre situation expérimentale.

2.1. Revue de différentes méthodes d'atténuation du bruit

Plusieurs domaines de recherche ont déjà cherché à réduire ou atténuer un bruit ambiant. Tout d'abord, le domaine de la reconnaissance automatique de la parole où la performance des algorithmes est dégradée par le bruit (Lee et al. 2004^[204] ; Murakami et al. 2006^[248]). De même, des recherches ont été menées pour améliorer la compréhension dans le bruit des personnes portant des prothèses auditives (Hagerman et al. 2004^[132] ; Prado et al. 1994^[276]). A l'inverse, certaines protections auditives « actives » ont été développées pour atténuer le bruit environnant des personnes confrontées à un environnement bruyant dans leur activité professionnelle (Zimpfer et al. 2006^[384]). Ces différents domaines s'intéressent surtout à augmenter la performance d'un algorithme ou le confort des personnes en éliminant le bruit au maximum, mais la conservation « exacte » des caractéristiques du signal utile n'est pas nécessairement leur priorité. Dans notre cas, la recherche d'une méthode d'atténuation du bruit est justement motivée par la volonté de réaliser des mesures acoustiques fiables. Nous souhaitons donc éviter que la méthode de réduction du bruit n'introduise un autre biais dans les mesures. Nous avons donc envisagé plusieurs types de méthodes mécaniques, électroacoustiques ou numériques de réduction du bruit.

2.1.1. Systèmes d'isolation

Une première atténuation du bruit peut être réalisée à l'aide de cloisons séparatrices ou de panneaux absorbants.

Meltzner et al. 2003^[237] ont été ainsi confrontés à un problème comparable en voulant étudier la voix électrolaryngée, i.e. le son émanant du conduit vocal lorsque la source de vibration n'est pas glottique, mais un buzzer électrique placé au niveau du cou. Afin de pouvoir enregistrer le signal rayonné par la bouche et le nez, sans le son du buzzer, ils ont imaginé une cloison avec une ouverture permettant au locuteur d'insérer la région bucco-nasale (cf. Figure 16).

De même, Feng et al. 2006^[91], s'intéressant à la production des voyelles nasales, ont imaginé un système composé d'une planche séparatrice insérée entre le nez et la bouche d'un locuteur, dans le plan sagittal de sa tête, et permettant d'enregistrer de façon isolée le son rayonné par le nez et celui rayonné par la bouche (cf. Figure 17).

Enfin, Ternström et al. 2006 ^[345] ont étudié la parole produite en environnement bruyant, restitué par haut-parleurs, et ont dans un premier temps placé des panneaux absorbants autour du locuteur afin d'atténuer le bruit ambiant dans lequel il est immergé, avant d'appliquer une autre méthode numérique de débruitage (cf. paragraphe 2.1.3).



Figure 16. Cloison permettant d'enregistrer la parole du locuteur de façon isolée par rapport à l'environnement sonore qui l'entoure. D'après Meltzner et al. 2003 ^[237].

Figure 17. Système permettant d'enregistrer de façon isolée le son rayonné par le nez et celui rayonné par la bouche, grâce à une planche séparatrice. D'après Feng et al. 2006 ^[91].

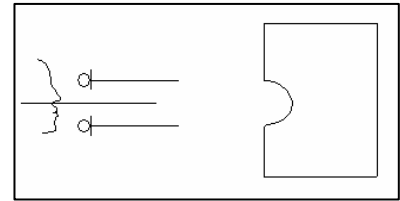
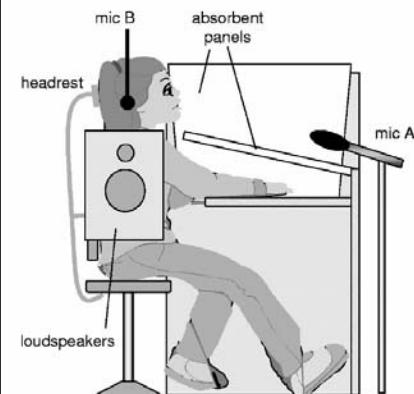


Figure 18. Première atténuation du bruit ambiant à l'aide de panneaux absorbants placés autour du sujet. D'après Ternström et al. 2006 ^[345].

Ces systèmes s'avèrent intéressants et ne distordent pas le signal de parole. En revanche, ils ne permettent pas de supprimer totalement le bruit ambiant, surtout lorsque celui-ci est de très forte intensité. De plus, les cloisons de séparation ne permettent pas une communication très naturelle avec un interlocuteur.

2.1.2. Systèmes électroacoustiques

D'autres systèmes ont cherché à améliorer le rapport signal sur bruit en renforçant le signal utile plutôt qu'en atténuant le bruit. Ainsi, Graciarena et al. 2004 ^[118] ont proposé d'enregistrer le signal audio par un microphone placé sur le cou du sujet, simultanément au signal audio au niveau des lèvres, afin de renforcer le signal de parole avant de le débruiter par une méthode numérique (cf. paragraphe 2.1.3). Cette méthode modifie évidemment beaucoup les caractéristiques spectrales de la parole.

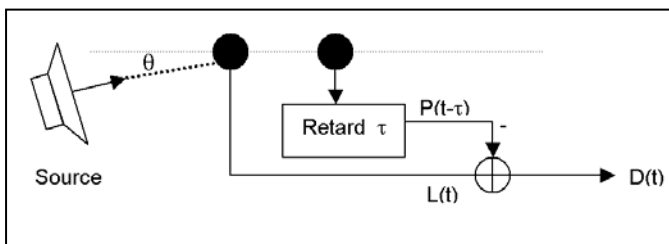


Figure 19. Schéma de principe d'un microphone autodirectif.

Une autre possibilité pour renforcer le rapport signal sur bruit dès l'enregistrement consiste à jouer sur la directivité du microphone. C'est le principe des microphones qualifiés de « close talking » ou de « noise cancelling » (Goldin 2004 ^[116]) : l'effet de proximité et la directivité de tels microphones sont utilisés pour atténuer les sons distants et sur les cotés, et renforcer les sons dans l'axe de la capsule. De tels microphones sont en fait constitués d'au moins 2 microphones omnidirectionnels, placés sur un même axe, et éloignés d'une distance d (cf. Figure 19). Le signal audio $D(t)$ correspond à la différence des signaux acquis par les différents microphones, entre lesquels on introduit un délai temporel τ . La transformée de Fourier du signal audio, $D(f)$, dépend donc de l'angle θ de la source par rapport à l'axe des microphones et s'exprime par l'équation (1).

$$(1) |D_{f,\theta}| = \frac{2\pi f d}{c} P(\theta) \left\{ \begin{array}{l} \text{, où } P(\theta) = \left| \frac{\tau c}{d} + \cos\theta \right| \text{ représente la fonction de directivité du microphone.} \\ c \text{ représente la vitesse de propagation du son dans l'air.} \end{array} \right.$$

Lorsque le délai τ correspond à d/c , le microphone a une directivité cardioïde. Lorsque τ correspond à $d/2c$, la directivité est supercardioïde. Enfin, un délai nul correspond à un microphone bidirectionnel (cf. Figure 20). Les microphones omnidirectionnels sont sensibles à toutes les directions de sources, tandis qu'un microphone hyper-directif montre au contraire un angle de captation très restreint et atténue considérablement les sources situées latéralement. En revanche, la directivité super-cardioïde présente un lobe également vers l'arrière, signifiant une bonne sensibilité à une source située derrière lui. Enfin, un microphone cardioïde capte avec un angle assez ouvert les sources situées vers l'avant mais atténue celles placées vers l'arrière (cf. Figure 20). Il faut également mentionner que la directivité des microphones a une incidence sur le contenu spectral du signal audio enregistré. En particulier, les microphones cardioïdes renforcent légèrement les basses fréquences tandis que les microphones omnidirectionnels ont une réponse relativement plate en fréquence.

Ce principe de base a été utilisé pour développer l'idée de microphone « autodirectif ». Il s'agit d'un microphone adaptatif, à l'intérieur duquel le délai τ est réévalué à chaque instant de façon à modifier sa directivité et obtenir grâce à cela le meilleur rapport signal sur bruit possible (Goldin 2004 ^[116]). Cette méthode est donc particulièrement utile pour éliminer des sources de bruit en déplacement constant.

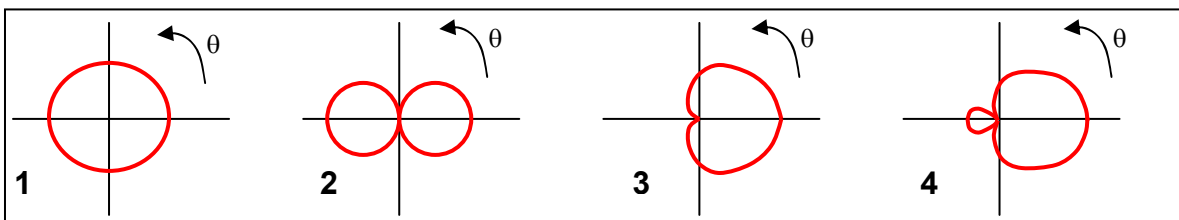
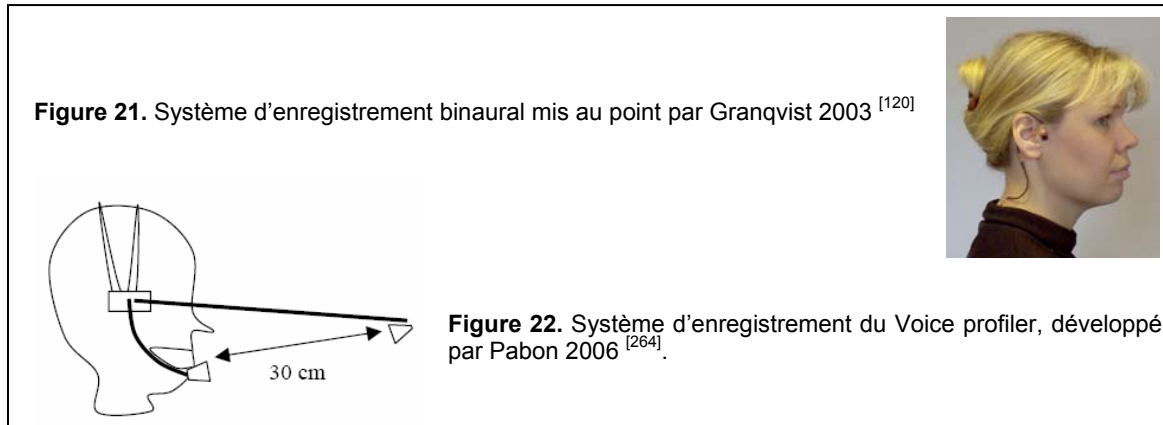


Figure 20. Diagrammes polaires de différentes directivités de microphones: 1-omnidirectionnel, 2-bidirectionnel, 3-cardioïde, 4-hypercardioïde.

D'autres techniques d'atténuation du bruit se basent également sur l'utilisation de plusieurs microphones, en utilisant cette fois les informations de phase des différentes sources : Granqvist 2003 ^[120] a ainsi mis au point une méthode d'enregistrement binaural, consistant à placer deux microphones condensateurs à électret au niveau des oreilles du locuteur (cf. Figure 21). Lorsque le sujet parle, les 2 signaux sont en phase, leur soustraction est donc très faible tandis que leur addition est au contraire très forte. Lorsque le sujet ne parle pas, les différentes sources ambiantes ne sont pas symétriques par rapport aux microphones et ne sont pas corrélées. La soustraction et l'addition des deux signaux sont donc d'intensités moyennes comparables. Granqvist 2003 ^[120] définit alors le « Self to other ratio », correspondant à la différence entre l'addition et la soustraction de ces deux signaux. Dans le cas où le sujet parle, cet indice est donc important tandis qu'il devient très faible lorsque le sujet ne parle pas. Cette méthode a ainsi été utilisée par Sodersten et al. 2002 ^[321] et Ternström 1994 ^[344] pour réaliser des enregistrements in situ et mesurer des temps de phonation ainsi que des niveaux d'intensité vocale. Cependant, cette méthode sert davantage à détecter les instants de phonation qu'à débruiter le signal de façon vraiment efficace. Par ailleurs, les sources situées face au locuteur posent problème dans la mesure où elles sont situées à équidistance des deux microphones et peuvent donc être prises en compte au même titre que la parole du locuteur.

Dans le même esprit, Pabon 2006 ^[264] a équipé un dispositif commercial d'analyse de la voix en temps réel (« Voice Profiler »), de deux microphones, situés dans l'axe de la bouche et éloignés d'une distance de 30cm (cf. Figure 22). Le logiciel de traitement élimine ensuite les sons dont l'intensité est plus grande au niveau du microphone éloigné qu'au niveau du microphone proche des lèvres, ainsi que les sons dont le délai entre les deux microphones ne correspond pas à d/c (c étant la vitesse de propagation du son dans l'air).



De façon plus précise mais également plus complexe, des antennes ou rangées de microphones peuvent être disposées autour du locuteur et utilisées pour reconstituer le champ acoustique en un point de l'espace (dans notre cas, il s'agirait des lèvres du locuteur). Ces techniques d'holographie ou de « beamforming » sont valables en champ stationnaire comme en champ non stationnaire (Haddad et al. 2002 ^[131]). C'est pourquoi elles ont déjà été utilisées pour le débruitage et l'étude de la parole produite dans une automobile (Lee et al. 2004 ^[204]). Pour augmenter la qualité de la reconstruction du champ acoustique, les microphones doivent être les plus nombreux possibles, couvrir une bonne partie de l'espace et être situés à proximité de la source.

Ces différents systèmes sont donc très intéressants car ils permettent d'atténuer le bruit dans des conditions où les sources sont en mouvement ou lorsque le locuteur se déplace. Cependant, les méthodes n'utilisant que deux microphones ne sont pas suffisantes pour supprimer totalement le bruit ambiant. L'augmentation du nombre de microphones s'accompagne d'une augmentation du coût et de la complexité de l'expérience.

2.1.3. Systèmes numériques de traitement du signal

2.1.3.1. Soustraction spectrale d'un profil de bruit

Certains logiciels de traitement du son, tels qu'Adobe Audition ®, proposent des modules de débruitage (ou « denoising » en anglais) permettant de supprimer des bruits de souffle ou d'atténuer avec efficacité des bruits stationnaires. Le principe de ces modules se base sur la soustraction spectrale du bruit de l'enregistrement. Pour cela, il est nécessaire de capturer le profil spectral du bruit à atténuer sur une partie de l'enregistrement où il n'y a aucun signal « utile ». Ce profil spectral est alors soustrait du spectre de l'enregistrement audio, calculé sur chaque fenêtre temporelle d'analyse. Le signal audio débruité est alors reconstitué par addition et recouvrement de chaque fenêtre. La taille de la FFT a donc une influence déterminante sur la qualité du débruitage : une taille de FFT importante améliore la précision fréquentielle, mais diminue au contraire la précision temporelle. Un compromis est donc à trouver en fonction des caractéristiques du bruit et du signal utile. Cependant, lorsque le bruit varie rapidement, et si son profil spectral n'est pas constant sur l'ensemble de l'enregistrement, la soustraction du profil spectral estimé du bruit peut entraîner un son indésirable métallique appelé « bruit musical », en particulier lorsque le rapport signal sur bruit est faible. C'est pourquoi les algorithmes de débruitage basés sur le principe de soustraction spectrale se sont grandement améliorés et proposent des solutions pour atténuer cette métalllicité ou mieux gérer des variations rapides du spectre de bruit (Murakami et al. 2006 ^[248]).

2.1.3.2. Soustraction temporelle d'un bruit connu

Lorsque le bruit que l'on souhaite supprimer est parfaitement connu, comme c'est le cas en laboratoire lorsqu'on diffuse un bruit calibré par haut-parleurs, plusieurs techniques de soustraction temporelle échantillon par échantillon sont envisageables. Il suffit pour cela d'estimer le bruit tel qu'il serait acquis au microphone si le locuteur restait silencieux ($h(t)*x(t)$ sur la Figure 23). Dans ce cas, on peut alors soustraire cette estimation du bruit à l'enregistrement bruité de parole ($y(t)$ sur la Figure 23), afin de ne récupérer que le signal utile et débruité de parole ($v(t)$ sur la Figure 23). L'estimation de ce bruit correspond au filtrage du fichier source de bruit par un « canal ». Ce canal est composé du convertisseur numérique/analogique, des haut-parleurs, de l'acoustique de la salle, du microphone et du convertisseur analogique/numérique (cf. Figure 23).

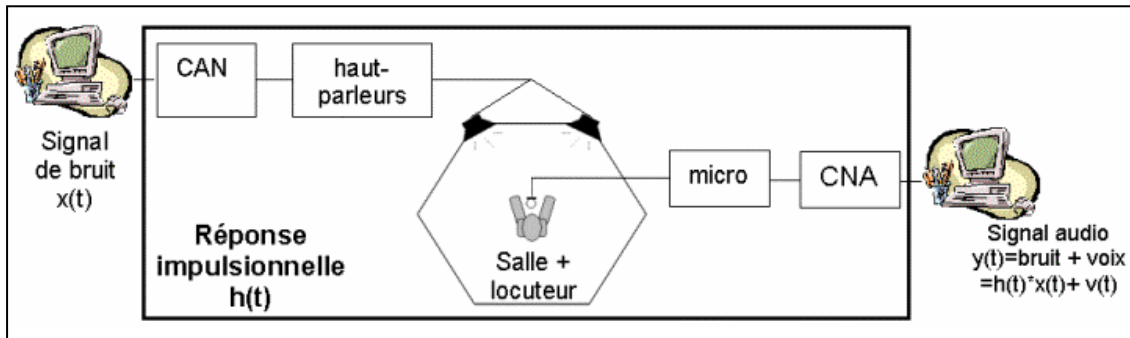


Figure 23. Schéma du canal à estimer pour pouvoir débruitier un enregistrement par soustraction temporelle du bruit.

Pour estimer le bruit acquis au microphone, deux méthodes sont envisageables :

- la première consiste à réellement enregistrer ce bruit au microphone ($h(t)*x(t)$ sur la Figure 23), en demandant d'abord au locuteur de rester silencieux (Mixdorff et al. 2006 ^[244]). Cette méthode a l'avantage d'être très simple d'implémentation mais nécessite d'exposer les locuteurs au bruit davantage que ce qui est nécessaire pour l'enregistrement de leur parole produite en environnement bruyant.
- la deuxième consiste à estimer la réponse impulsionnelle du canal ($h(t)$ sur la Figure 23), et ainsi à estimer le bruit acquis au microphone en convoluant le fichier source de bruit ($x(t)$ sur la Figure 23) à cette réponse impulsionnelle (Ternström et al. 2002 ^[346]). Pour cela, il est nécessaire de diffuser un extrait de bruit blanc pendant une dizaine de secondes, et de l'enregistrer au microphone en demandant au locuteur de rester silencieux. La réponse impulsionnelle du canal peut alors être connue en comparant le fichier source de bruit blanc au signal acquis au microphone ($h(t) = TF^{-1}[Y(f)/X(f)]$), et réutilisée ensuite pour estimer le signal acquis au microphone pour n'importe quel fichier source (cf. Figure 24).

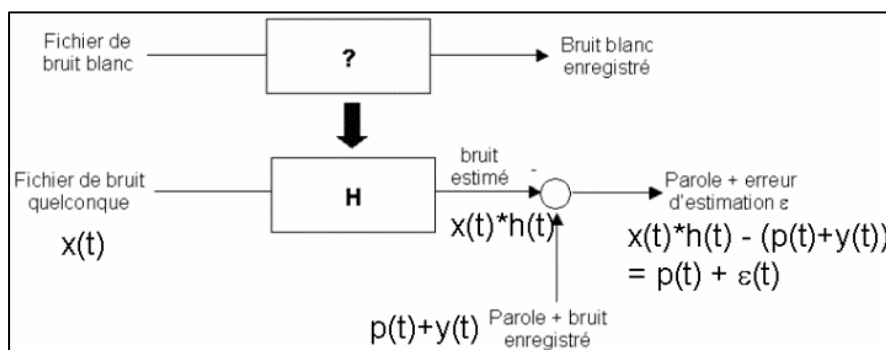


Figure 24. Principe du débruitage par soustraction temporelle à partir d'une estimation préalable du canal.

Ces deux méthodes obtiennent des résultats très intéressants, même sur des niveaux de bruit très importants, et sans distordre le signal ni introduire de métallicité parasite. Une fois que le canal a été estimé, ces méthodes peuvent en outre agir en temps réel. Leur inconvénient est en revanche de ne pouvoir fonctionner que sur un bruit parfaitement connu, d'exposer le locuteur à davantage de bruit que ce qui est nécessaire pour les enregistrements, et surtout de le contraindre à rester le plus immobile possible. En effet, l'estimation du bruit qui serait acquis au microphone dépend dans les deux cas de l'estimation préalable du canal. Celui-ci dépend lui-même de l'acoustique de la salle à laquelle le locuteur et son positionnement participent. Un

changement de position du locuteur induit alors un changement dans la réponse impulsionnelle du canal, détériorant rapidement les performances de débruitage (Mixdorff et al. 2006^[244]; Ternström et al. 2002^[346]). Etant donné que le locuteur ne peut physiquement pas rester immobile pendant un temps infini, l'usage de ces méthodes nécessite donc une ré-estimation du canal à chaque nouvelle condition d'enregistrement.

2.1.3.3. Filtrage adaptatif

En cela, le filtrage adaptatif s'avère intéressant puisqu'il n'estime plus le canal préalablement à l'enregistrement, et de façon invariante dans le temps, mais au contraire à chaque instant. Le canal peut donc changer légèrement au cours de l'enregistrement, et le locuteur peut alors bouger et être plus décontracté. Par ailleurs, aucune estimation préalable du canal n'est nécessaire. Le locuteur n'a donc pas besoin d'être exposé au bruit en dehors de l'expérience. Le principe de base de cette méthode consiste à estimer le canal à chaque instant en minimisant l'erreur quadratique de ε , correspondant à la différence entre le bruit réellement enregistré au microphone et son estimation (cf. Figure 25).

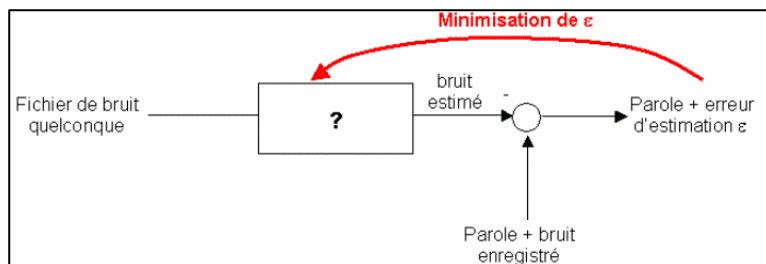


Figure 25. Principe du filtrage adaptatif par estimation de canal.

Une première implémentation assez simple du filtrage adaptatif peut être effectuée à l'aide de l'algorithme BLMS (Block Least Mean Square) (Clark et al. 1981^[54]). Dans cet algorithme, le signal enregistré est fenêtré par blocs de N échantillons temporels, pour lesquels la réponse impulsionnelle du canal est considérée comme constante. Pour le $k^{\text{ième}}$ bloc, la minimisation de l'erreur quadratique de ε_k correspond concrètement à l'annulation de la dérivée de son espérance mathématique. Cette condition se traduit par l'équation (1) $\mathbf{h}_k = \mathbf{h}_{k-1} + \mu \cdot \mathbf{X}_k \cdot \varepsilon_k$

- où - $\mathbf{h}_k = [h_{k;0}, \dots, h_{k;L-1}]^T$ est un vecteur de longueur L qui représente la réponse impulsionnelle du canal à estimer, à l'instant $k.N$
- \mathbf{X}_k est une matrice de taille $L \cdot N$. Les colonnes de cette matrice sont les vecteurs χ_{kN} à χ_{kN+N-1} , où $\chi_i = [x_i, \dots, x_{i-L+1}]^T$ représente L échantillons consécutifs du signal source de bruit x , pris à partir de l'instant i .
 - μ est le "pas d'adaptation" de l'algorithme qui détermine la convergence de l'algorithme, empiriquement ajusté. Lorsque ce pas est trop faible, l'algorithme ne s'adapte pas assez rapidement et ne réussit pas à atténuer le bruit si celui-ci change trop rapidement. Au contraire, un pas d'adaptation trop important peut entraîner une divergence immédiate de l'algorithme (cf. Figure 26)
 - $\varepsilon_k = \text{BruitEnregistré}_k - \text{BruitEstimé}_k = \text{BruitEnregistré}_k - \mathbf{X}_k^T \cdot \mathbf{h}_k$ est un vecteur de taille N .

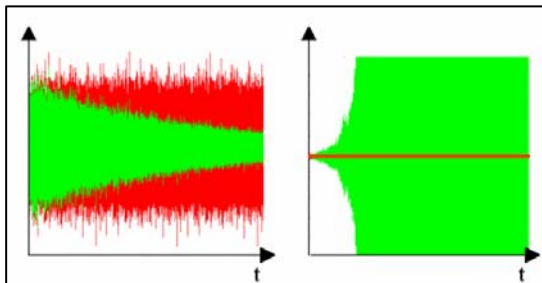


Figure 26. Réglage empirique du pas d'adaptation μ d'un algorithme de filtrage adaptatif. Le signal de bruit est représenté en rouge (foncé) et le résultat de l'algorithme de débruitage en vert (clair). À gauche, représentation d'un cas où le pas d'adaptation est trop faible. À droite, représentation d'un cas où celui-ci est trop important et où l'algorithme diverge

De nombreuses améliorations ont été proposées à cet algorithme de base au fil des années. En particulier, Ferrara 1980^[92] a implémenté cet algorithme dans le domaine fréquentiel (Fast LMS), permettant de réduire la quantité de calcul et de mieux optimiser le taux de convergence de l'algorithme lorsque le spectre de bruit présente des variations importantes.

D'autres extensions de cette implémentation fréquentielle ont été proposées sous la dénomination MDF (Multi Delay Filters). Dans ces algorithmes, la réponse impulsionnelle,

initialement de longueur L , est divisée en plusieurs segments, permettant d'implémenter l'équation (1) dans le domaine spectral mais sur des vecteurs de taille plus réduite que celle, parfois importante, de la réponse impulsionnelle (Soo 1990^[326]).

Enfin, une généralisation du MDF a été proposée par Prado et al. 1994^[276] sous la dénomination $\text{GMDF}\alpha$ (Generalized Multi Delay Filters). Cet algorithme, en cours de brevetage, a été appliqué à l'annulation d'écho dans les prothèses auditives.

2.2. Comparaison de 3 méthodes numériques de débruitage

Dans le but d'évaluer et de comparer ces méthodes, une étude a été menée sur 3 principales méthodes numériques d'atténuation du bruit :

- une méthode de soustraction temporelle avec estimation de canal (Ternström et al. 2002^[346]), que nous appellerons « Méthode NC » pour reprendre la dénomination « Noise Cancellation Method » de l'auteur.
- une méthode simple de filtrage adaptatif (BLMS), basée sur Clark et al. 1981^[54].
- une des dernières propositions en matière de filtrage adaptatif ($\text{GMDF}\alpha$), aimablement mise à disposition par l'auteur (Prado et al. 1994^[276]).

Le but de cette étude comparative n'est pas seulement de comparer les performances de ces méthodes en terme d'amélioration du rapport signal sur bruit, mais également d'examiner l'influence qu'elles exercent sur le signal de parole et déterminer s'il peut être valide de réaliser des mesures acoustiques à partir des signaux débruités par ces méthodes. Par ailleurs, nous évaluons également leurs avantages et inconvénients par rapport à différentes situations expérimentales. C'est pourquoi nous comparerons d'abord ces méthodes en situation de performance optimale, puis dans des conditions plus réelles correspondant à différentes situations expérimentales.

2.2.1. Choix préalable d'un microphone adapté

Avant même de chercher à réduire le bruit par des techniques de traitement de signal, une première réduction du bruit peut être obtenue en plaçant un microphone le plus près possible de la bouche du locuteur et le plus loin possible des sources de bruit. Cependant, dans le cas de l'étude de la voix produite dans un environnement bruyant, l'intensité vocale peut atteindre jusqu'à 110 dB au niveau des lèvres (Pabon 2006^[264]), intensité bien supérieure à la limite à partir de laquelle la plupart des microphones saturent (environ 90 dB en moyenne). Aussi, il faut trouver un compromis entre le rapprochement de la bouche et la non-saturation du microphone.

De plus, le rapport signal sur bruit peut être considérablement amélioré en choisissant un microphone dont la directivité est adaptée par rapport au placement du locuteur et des sources de bruit. Dans le cas d'une reconstitution d'un environnement bruyant en laboratoire, il est également envisageable de placer les sources de bruit de façon adéquate pour optimiser la réjection du bruit en fonction de la directivité du microphone.

Plusieurs essais ont été effectués avec différents types de microphones et différents placements des enceintes acoustiques par rapport au locuteur et au microphone. Nous avons obtenu le meilleur rapport signal sur bruit à l'enregistrement grâce à un microphone-casque cardioïde Beyerdynamic (Opus 54), pour des angles de disposition des enceintes de 60° face au sujet (donc à l'arrière de la capsule) jusqu'à 180° (latéralement au sujet et à la capsule). Nous avons donc retenu ce microphone pour la suite de l'étude comparative ainsi que pour l'enregistrement de la première base de donnée (BD1).

2.2.2. Optimisation et comparaison des trois méthodes dans des conditions idéales

Avant de comparer leurs performances, nous avons commencé par optimiser ces différentes méthodes. Certains paramètres de ces algorithmes sont réglables de façon objective et prévisible. Mais la plupart nécessite un réglage empirique en fonction du type de bruit et de la condition d'enregistrement. Il est possible qu'une différence de réglage existe selon le locuteur. Toutefois, nous n'en avons pas remarqué entre les 4 locuteurs que nous avons enregistrés.

Par ailleurs, dans le cas des deux algorithmes de filtrage adaptatif, l'optimisation de l'algorithme consiste essentiellement à trouver un compromis entre l'atténuation du bruit et la conservation d'une bonne qualité de la parole débruitée. Un bon rapport signal sur bruit n'est pas la garantie

d'une qualité sonore de l'enregistrement. De même, nous avons observé que le profil moyen du spectre ne permettait pas de rendre compte de toutes les distorsions de la parole, en particulier des effets de réverbération très gênants perceptivement mais transparents au niveau spectral. C'est pourquoi nous avons ici choisi d'optimiser ces trois méthodes sur la base de l'augmentation du rapport signal sur bruit qu'elles apportent, de leur fidélité spectrale et perceptive par rapport au signal de parole de référence (i.e. sans bruit).

2.2.2.1. Matériel et méthode

Pour optimiser les différentes méthodes, nous avons choisi de nous placer en condition idéale d'enregistrement, c'est à dire sans déplacement du locuteur et pour une unique source de bruit située face au locuteur. De façon à supprimer les déplacements, même minimes, du locuteur, nous avons utilisé un locuteur « virtuel », c'est à dire un haut-parleur (TANOY Reveal) placé à l'endroit du locuteur, équipé du microphone-casque Beyerdynamic (Opus 54) et diffusant la parole préenregistrée du locuteur. Un haut-parleur (Studer A1) était placé à 2 m en face du locuteur virtuel, et diffusait deux types de bruit (bruit blanc et bruit de conversations), de même intensité en dB SPL mesurée sur les signaux d'entrée. La chaîne de diffusion-acquisition ajoute une atténuation spectrale hautes-fréquences au-delà de 12kHz (cf. Figure 27). Par conséquent, les deux bruits sont enregistrés au sonomètre avec une différence d'intensité mesurée à 5.8dB en faveur du bruit de conversations (84.9 dB pour le bruit de conversations et 79.1 dB pour le bruit blanc).

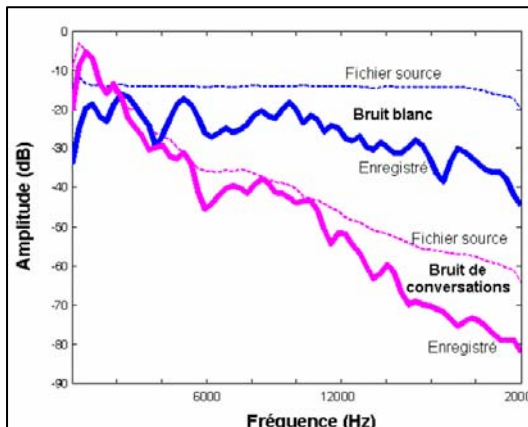


Figure 27. Spectre moyenné des deux types de bruit utilisés pour cette étude comparative : un bruit blanc et un bruit de conversations.

Le rapport signal sur bruit a été calculé en calculant l'intensité du signal audio en dB lorsque le locuteur parle et en la soustrayant à l'intensité du signal audio en dB durant les pauses entre les énoncés, durant lesquelles le bruit continue d'être diffusé. Sur les figures suivantes, ce n'est pas le rapport signal sur bruit lui-même qui est représenté en ordonnées mais son gain, entre le signal audio enregistré et le signal audio débruité.

2.2.2.2. Optimisation de la « méthode NC »

Cette méthode ne dépend globalement que de la résolution fréquentielle de l'estimation de la réponse impulsionnelle du canal. L'atténuation du bruit augmente avec le nombre de points de calcul de la FFT, et augmente faiblement au-delà de $NFFT=4096$. C'est pourquoi nous avons retenu cette valeur pour optimiser cette méthode. Avec cette méthode, le contenu spectral de la parole débruitée est uniquement perturbé par les résidus de bruit que l'algorithme n'a pu totalement supprimer, et aucune distorsion spectrale ou temporelle n'est ajoutée.

2.2.2.3. Optimisation de l'algorithme BLMS

Cette méthode dépend principalement du pas d'adaptation μ et de la longueur L d'estimation de la réponse impulsionnelle.

L'atténuation du bruit augmente avec des valeurs croissantes de μ , jusqu'à une valeur limite où l'algorithme diverge. Cette valeur limite est située vers $\mu = 0.3$ pour le bruit blanc, et $\mu = 0.2$ pour le bruit de conversations (cf. Figure 28). L'algorithme BLMS est inefficace pour des valeurs de L inférieures à 700 échantillons (environ 16ms). Il atténue ensuite d'autant mieux le bruit que cette longueur L est importante, puis atteint un palier à partir d'une valeur de L d'autant plus petite que μ est faible (cf. Figure 28). L'atténuation maximale du bruit blanc est obtenue pour

$\mu=0.2$ et $L=1400$ (environ 14 dB de gain sur le rapport signal sur bruit). Celle du bruit cocktail est obtenue pour $\mu=0.1$ et $L=1300$ (environ 12 dB de gain).

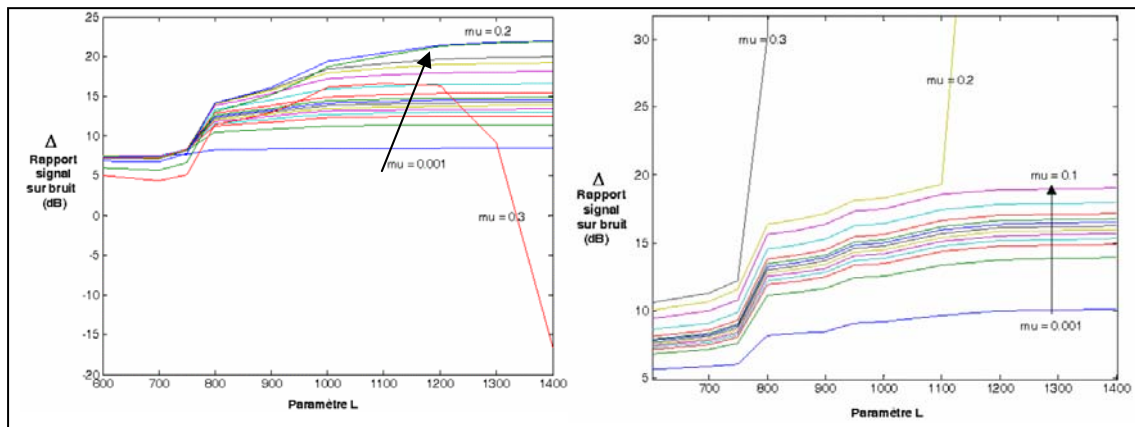


Figure 28. Augmentation du rapport signal sur bruit en fonction des paramètres μ et L , pour l'algorithme BLMS, dans le bruit blanc (à gauche) et le bruit de conversations (à droite).

Pour les valeurs de μ inférieures à 0.1, le contenu spectral de la parole débruitée se rapproche de celui du signal de référence pour des valeurs moyennes de L , entre 1000 et 1300. Au vu des différentes combinaisons de paramètres, nous avons retenu les valeurs $\mu = 0.05$ et $L = 1200$ pour optimiser les performances de l'algorithme BLMS dans le bruit blanc, et les valeurs $\mu = 0.01$ et $L = 1200$ dans le bruit de conversations.

2.2.2.4. Optimisation de l'algorithme GMDF_α

Cette méthode dépend également du pas d'adaptation μ_0 et de la longueur d'estimation de la réponse impulsionnelle. Dans le cas de l'algorithme GMDF_α , celle-ci est découpée en K blocs de longueur L . Ces deux paramètres ont donc une influence sur les performances de l'algorithme.

On observe en tout premier lieu que le rapport signal sur bruit augmente, puis décroît avec des valeurs croissantes de μ_0 (cf. Figure 29). Le maximum d'atténuation du bruit est atteint pour des valeurs de μ_0 proches de 0.01, plus ou moins important selon les valeurs de K et L (cf. Figure 29).

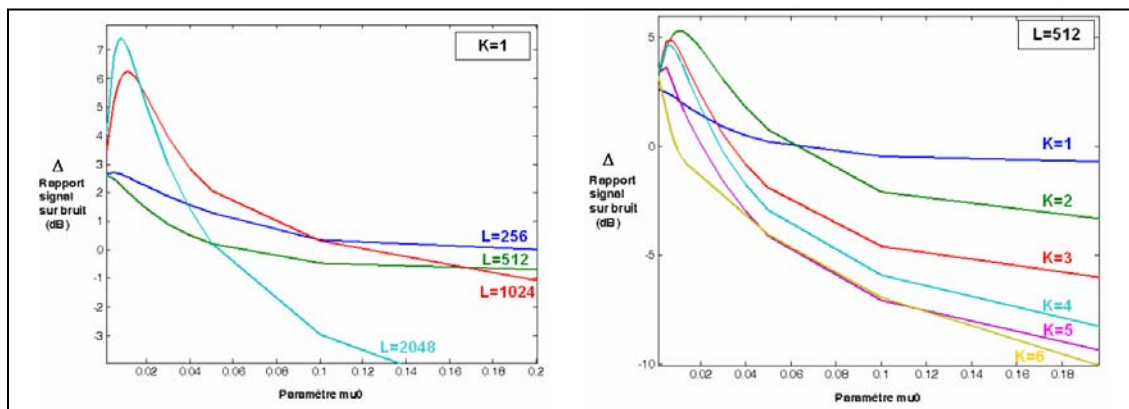


Figure 29. Influence du pas d'adaptation μ_0 sur l'augmentation du rapport signal sur bruit réalisée par l'algorithme GMDF_α , ici dans le bruit de conversations. La figure de gauche représente également l'influence de la variation de la longueur L des blocs pour un seul bloc. La figure de droite représente au contraire l'influence d'un nombre de blocs K croissant, pour une longueur de bloc L fixe de 512 échantillons.

Pour $\mu_0 = 0.01$, on observe que le rapport signal sur bruit est maximum non pas pour une valeur fixe de L ou de K , mais plutôt pour une taille de réponse impulsionnelle ($L \cdot K$) comprise entre 1024 et 2048 échantillons (soit entre 23 et 46 ms) (cf. Figure 30).

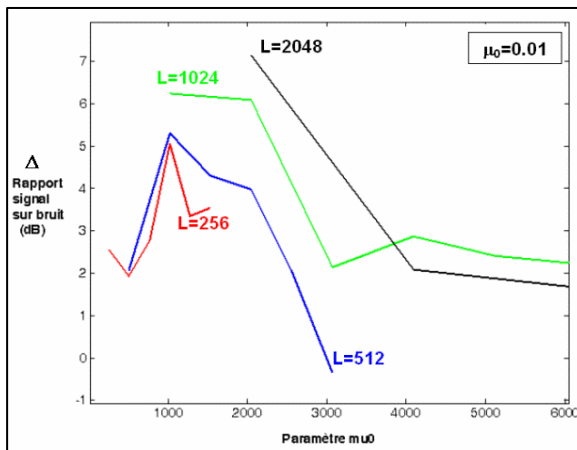


Figure 30. Influence des paramètres L et K sur l'augmentation du rapport signal sur bruit réalisée par l'algorithme GMDF_α , ici dans le bruit de conversations et pour une valeur fixe de μ_0 égale à 0.01.

La Figure 31 représente l'augmentation du rapport signal sur bruit en fonction de μ_0 pour les différentes combinaisons de L et K telles que $L \cdot K$ soit compris entre 1024 et 2048. On observe que le meilleur rapport signal sur bruit est atteint dans le bruit de conversations pour $L=2048$ et $K=1$, et pour une valeur de $\mu_0=0.008$.

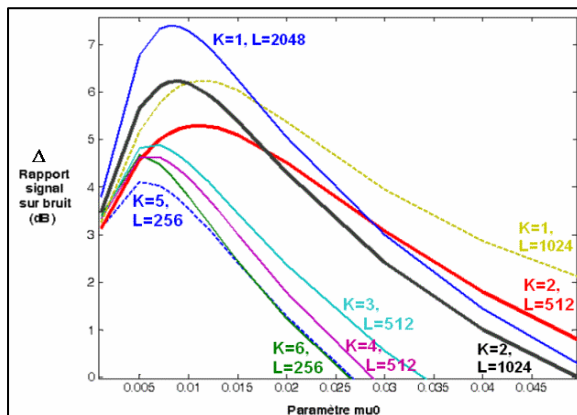


Figure 31. Influence du pas d'adaptation μ_0 sur l'augmentation du rapport signal sur bruit réalisée par l'algorithme GMDF_α , ici dans le bruit de conversations, pour différentes combinaisons des paramètres L et K tels que $1024 < LK < 2048$.

Cependant, la parole débruitée est légèrement distordue pour cette valeur de μ_0 . Le meilleur compromis entre l'atténuation du bruit et la conservation de la qualité de la voix est obtenu pour $K=1$, $L=2048$, et $\mu_0=0.01$ dans le bruit de conversations (cf. Figure 31). Dans le bruit blanc, l'atténuation maximale du bruit est obtenue pour $\mu_0=0.1$, $K=1$ et $L=1024$. Cependant, aucun compromis satisfaisant n'a été trouvé : l'algorithme atténue très efficacement le bruit, mais au prix d'un fort effet cathédrale, ou bien conserve la qualité de la voix, mais sans apporter aucune amélioration du rapport signal sur bruit. On peut cependant se poser la question de savoir si les mesures spectrales effectuées sur de la parole avec effet « cathédrale » ne peuvent pas être considérées malgré tout comme valides. En effet, cet effet, très gênant au niveau perceptif, semble avoir des conséquences sur les caractéristiques temporelles du signal de parole davantage que sur ses caractéristiques spectrales. La Figure 32 (schéma de droite) montre en effet que le contenu spectral de la parole fortement débruitée par l'algorithme GMDF, avec effet « cathédrale » prononcé, est très proche du contenu spectral du signal de référence, i.e. enregistré dans le silence.

2.2.2.5. Comparaison des 3 méthodes optimisées

En condition idéale, le rapport signal sur bruit est considérablement amélioré (cf. Tableau 2). L'algorithme GMDF et la Méthode NC parviennent à augmenter le RSB de 13.7 et 15.4 dB dans le bruit blanc. La Méthode NC est en revanche plus performante face au bruit de conversations et parvient à augmenter le RSB de 13.2 dB. Nous rappelons que les deux algorithmes de filtrage adaptatifs permettent de réduire le bruit davantage que ce qui est présenté ici, mais au prix de distorsions de la parole que nous souhaitons éviter par rapport à notre problématique.

	Signal de référence (sans bruit)	Enregistrement	Débruitage BLMS	Débruitage GMDF α	Débruitage Méthode NC
Bruit blanc	40.9 dB	12.5 dB	16.1 dB	26.3 dB	27.9 dB
Bruit de conversations		16.6 dB	24.3 dB	20.4 dB	29.9 dB

Tableau 2. Rapport signal sur bruit avant et après débruitage par les différentes méthodes.

L'estimation de l'intensité vocale sur les voyelles est très fiable. En revanche, le débruitage des consonnes est moins efficace, induisant une surestimation de l'intensité vocale moyenne pouvant atteindre 1.7 dB. L'estimation de la F0 semble assez robuste au bruit : l'erreur d'estimation n'excède pas 0.15 ton et n'est pas amoindrie par le débruitage. L'estimation de l'énergie spectrale est très fiable en dessous de 6kHz, quel que soit le type de bruit pour les enregistrements débruités par l'algorithme GMDF α et par la méthode NC. Le bruit est très efficacement atténué entre 6 et 10 kHz, en particulier par la Méthode NC (moins de 0.2 dB d'erreur d'estimation de l'intensité dans cette zone). Enfin, le bruit blanc est beaucoup atténué entre 10 et 20 kHz par l'algorithme GMDF α et la Méthode NC, par rapport à l'enregistrement bruité, mais il reste néanmoins un résidu de bruit non négligeable dans cette bande de fréquences (cf. Figure 32).

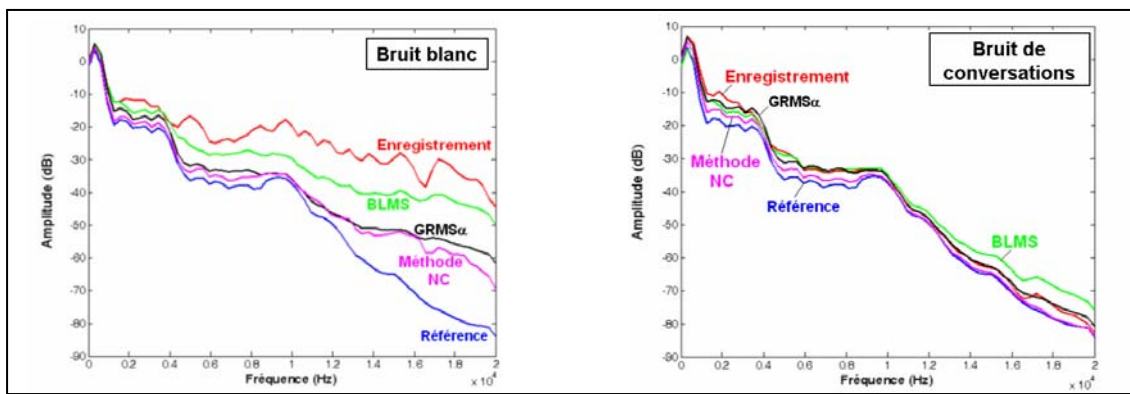


Figure 32. Comparaison des spectres moyennés sur la phrase « J'ai un ami qui travaille dans le quartier » pour l'enregistrement de référence, i.e. dans le silence, pour l'enregistrement bruité et pour les 3 différentes méthodes de débruitage (algorithmes BLMS et GMDF α , Méthode NC). A gauche, dans le bruit blanc, à droite, dans le bruit de conversations.

2.2.4. Comparaison en situations non idéales

Afin de savoir si ces méthodes restent efficaces dans des conditions moins idéales, correspondant davantage à des conditions expérimentales réelles, nous avons comparé de nouveau ces trois algorithmes mais cette fois pour un bruit très intense (diffusion toujours mono et locuteur immobile), pour un déplacement léger et important du locuteur (diffusion mono et d'intensité forte identique à celle de la situation idéale), ainsi que pour une diffusion stéréo (locuteur immobile et intensité forte identique à celle de la situation idéale).

2.2.4.1. Matériel et méthode

Afin d'examiner l'influence du niveau de bruit ambiant à supprimer de l'enregistrement, nous avons diffusé les deux types de bruit à un niveau très intense, de même intensité en dB SPL mesurée sur les signaux d'entrée (1^{er} test), avec un locuteur immobile et une restitution mono du bruit. Les deux bruits montrent une différence d'intensité mesurée à 6.1dBC en faveur du bruit de conversations (91.2 dBC pour le bruit de conversations et 85.1 dBC pour le bruit blanc), due comme précédemment à l'atténuation spectrale hautes-fréquences induite par la chaîne de diffusion-acquisition.

Afin d'examiner l'influence du déplacement du locuteur, le haut-parleur représentant le locuteur virtuel est porté par une personne réelle bougeant légèrement (2^{ème} test) puis beaucoup (3^{ème} test), dans un bruit d'intensité identique à la situation idéale (cf. § 2.2.2.1, 84.9 dBC pour le bruit de conversations et 79.1 dBC pour le bruit blanc).

Pour explorer l'influence de la restitution stéréo, le locuteur virtuel reste immobile, comme dans l'expérience précédente, mais deux haut-parleurs (Studer A1) diffusent le bruit latéralement

(stéréo, avec un angle de 180°) (4^{ème} test), dans un bruit d'intensité identique à la situation idéale (cf. § 2.2.2.1)

Nous avons toujours utilisé le même microphone, les deux mêmes types de bruit, et les mêmes enregistrements des 4 locuteurs avec un locuteur virtuel.

2.2.4.2. Résultats

L'estimation de la F0 n'est pas beaucoup affectée par le bruit ni par les conditions d'enregistrement, excepté dans le cas où le locuteur bouge de façon importante. L'erreur peut alors excéder le demi-ton dans le cas de la Méthode NC (cf. Figure 33). Deliyski et al. 2005^[68] et Carson et al. 2003^[45] ont montré que la F0 était résistante aux effets de bruit, davantage que le Jitter.

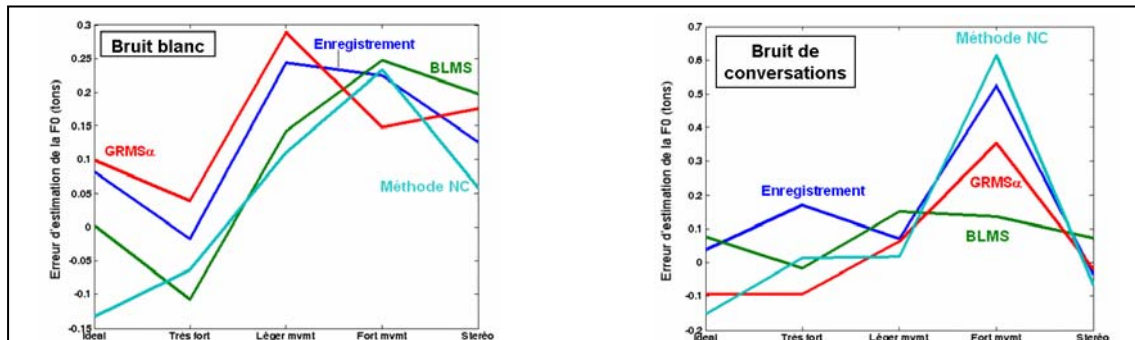


Figure 33. Différence d'estimation de la F0 entre l'enregistrement bruité et le signal de référence, i.e. enregistrée dans le silence, ainsi qu'entre les signaux débruités par les 3 différentes méthodes (algorithmes BLMS et GMDF α , Méthode NC) et le signal de référence. A gauche : enregistrements dans le bruit blanc, à droite, dans le bruit de conversations.

De façon générale, la performance des trois méthodes est amoindrie lorsque le locuteur bouge de façon importante, en particulier pour la Méthode NC pour un bruit de conversations. Il est donc préférable de minimiser autant que possible le déplacement du locuteur. Dans le cas où cela n'est pas possible, l'algorithme GMDF α s'avère le plus efficace, mais entraîne cependant des erreurs d'estimation non négligeables.

Dans les autres conditions d'enregistrement (fort niveau de bruit, léger mouvement du locuteur, restitution stéréo du bruit), les différentes méthodes se valent pour la réduction du bruit sur les parties voisées. Malgré sa performance de débruitage, la Méthode NC induit une erreur d'estimation de l'intensité pouvant aller jusqu'à 0.6 dB dans le bruit blanc et jusqu'à 1.6 dB dans le bruit de conversations (cf. Figure 34).

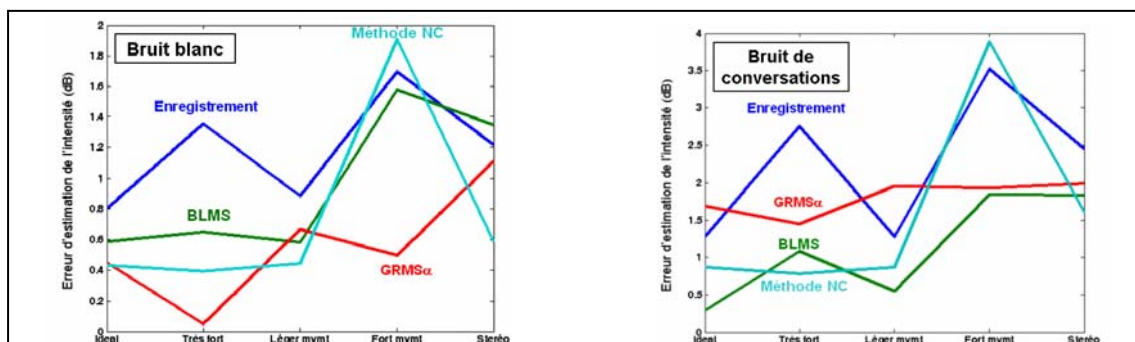


Figure 34. Différence d'estimation de l'intensité entre l'enregistrement bruité et la parole et le signal de référence, i.e. enregistrée dans le silence, ainsi qu'entre les signaux débruités par les 3 différentes méthodes (algorithmes BLMS et GMDF α , Méthode NC) et le signal de référence.. A gauche : enregistrements dans le bruit blanc, à droite, dans le bruit de conversations.

Cette erreur d'estimation de l'intensité est liée à un débruitage plus ou moins efficace en fonction des bandes de fréquences. De ce point de vue, on observe que la Méthode NC est la méthode dont le débruitage est perceptivement le plus proche de la parole produite dans le silence, excepté lorsque le locuteur bouge de façon importante. Cela se traduit en partie par

une erreur d'estimation de l'énergie spectrale assez faible dans la bande 50-6000Hz (cf. Figure 35).

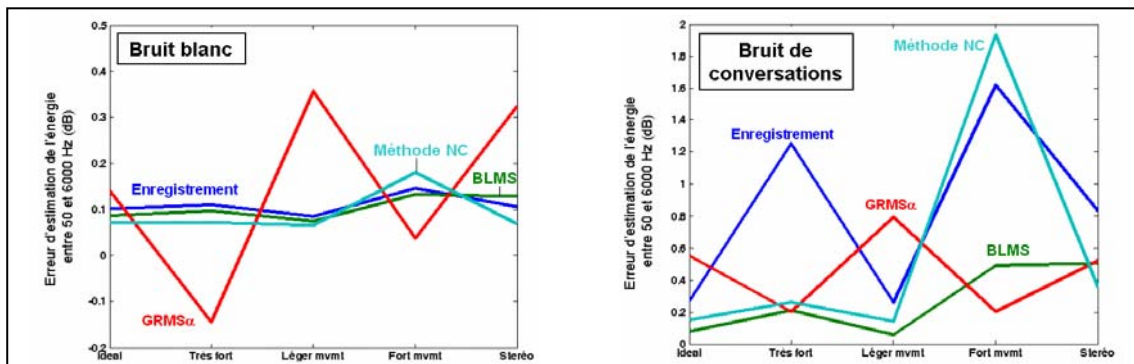


Figure 35. Différence d'estimation de l'énergie spectrale entre 50 et 6000Hz, entre l'enregistrement bruité et le signal de référence, i.e. enregistrée dans le silence, ainsi qu'entre les signaux débruités par les 3 différentes méthodes (algorithmes BLMS et $GMD\alpha$, Méthode NC) et le signal de référence. A gauche : enregistrements dans le bruit blanc, à droite, dans le bruit de conversations.

2.2.5. Conclusions

Après avoir implémenté et comparé 3 différentes méthodes numériques de débruitage, il s'avère que les méthodes de filtrage adaptatif, peu contraignantes pour le locuteur, sont par ailleurs les plus efficaces en ce qui concerne la réduction du bruit sur l'enregistrement. Malheureusement, cette atténuation du bruit s'accompagne d'une distorsion importante de la voix problématique pour les analyses spectrales de la parole. Nous avons donc déterminé empiriquement les paramètres de réglages des différentes méthodes afin d'optimiser le compromis entre la meilleure réduction du bruit et la conservation de la qualité vocale. La Méthode NC, non adaptative, s'avère alors la plus adaptée à nos objectifs pour des conditions d'enregistrements idéales (sujet immobile, restitution mono, bruit à 80 dBC) ou pour un faible déplacement du locuteur. En revanche, les algorithmes adaptatifs, en particulier l'algorithme $GMD\alpha$, sont plus efficaces dans le cas où le locuteur n'est pas immobile.

D'un point de vue acoustique, l'utilisation de la Méthode NC pour réaliser des mesures acoustiques sur le signal de parole débruité peut être valide, à condition que l'élévation de la F_0 dans le bruit soit supérieure au demi-ton, que l'augmentation de l'intensité vocale soit supérieure à 1.5 dB, et que la modification de la parole concerne davantage les fréquences inférieures à 10kHz.

3. COMPARAISON DE TROIS METHODES D'IMMERSION DU LOCUTEUR DANS LE BRUIT

En s'appuyant sur la méthode de débruitage sélectionnée, nous avons pu mener une expérience afin de tester la validité de l'utilisation d'un casque pour immerger le locuteur en environnement bruyant lors de l'étude de la parole Lombard. Nous souhaitons tout d'abord examiner si le port d'un casque (sans retour auditif) induit un changement significatif dans l'adaptation des locuteurs pour un même type et un même niveau de bruit. Dans un second temps, nous examinerons si le fait de rajouter un retour auditif de sa propre voix dans le casque du locuteur peut combler l'atténuation du retour auditif aérien imposée par le casque, et ainsi annihiler l'influence du port du casque.

3.1. Matériel et méthode

3.1.1. Constitution d'une première base de données (BD1)

Pour examiner cette question, nous avons enregistré une importante première base de données (BD1), explorée ensuite plus en détail pour tester d'autres hypothèses de la thèse. Cette base de données est constituée de la parole semi-spontanée en Français de 10 locuteurs (5 hommes et 5 femmes) produite dans une chambre assourdie, en condition « silencieuse » (bruit de fond mesuré à 40 dBC) puis dans différentes conditions de bruit, et cela pour différentes méthodes d'immersion des locuteurs dans le bruit.

Locuteurs

Les locuteurs (L3, L4, L5, L7 et L8) et locutrices (L1, L2, L6, L9 et L10) étaient âgés de 20 à 28 ans, de langue maternelle française. Exceptée une locutrice, ils étaient naïfs vis à vis des problématiques de l'effet Lombard ou du forçage vocal, et ne présentaient pas de troubles vocaux ni auditifs. Un audiogramme de chaque locuteur a été réalisé avant l'enregistrement du corpus.

Tâche de parole et matériel linguistique

Ces locuteurs et locutrices ont été enregistrés par binôme (L1 avec L2, L3 avec L4, etc.) en situation de dialogue interactif au cours d'un jeu (exposé au paragraphe 1.2.1 du chapitre 4) nécessitant l'utilisation de 16 mots cibles de forme CVCV (bijou, chausson, cochon, dauphin, fusil, guenon, gitans, lagon, mairie, navet, panda, requin, sommet, toupie, vallée, zébu). En dehors de ces mots-cibles, le discours des locuteurs était libre et donc variable. Il correspondait à un temps de parole d'environ 1min30 pour chaque condition.

Types de bruit

Les conditions de bruit ont été simulées à l'aide de deux types de bruit, issus de la base de données BD_Bruit (Zeiliger et al. 1994^[382]):

- un bruit blanc (bb), plat en fréquence jusqu'à 10kHz puis atténué au-delà de 10kHz (cf. Figure 36).
- un bruit de conversations, ou bruit de « cocktail-party » (cktl) à 8 voix mixtes, inintelligibles, et d'énergie plutôt basse-fréquences, située en dessous de 800Hz (cf. Figure 36).

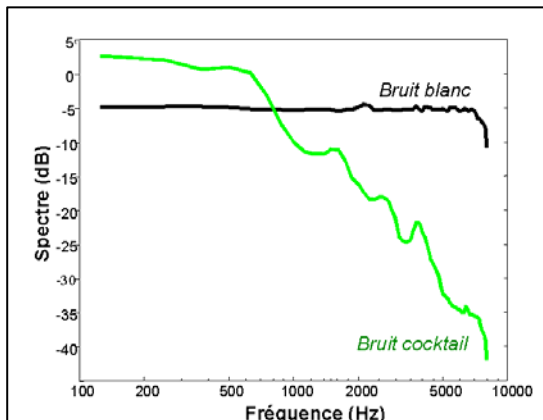


Figure 36. Représentation du spectre des deux types de bruit étudiés.

Niveaux de bruit

Ces deux types de bruit ont été diffusés à 4 niveaux d'intensité (62, 70, 78 et 86 dBC). La condition silencieuse et les 8 conditions bruyantes ont été enregistrées dans un ordre aléatoire, mais identique pour tous les locuteurs.

Moyens d'immersion des locuteurs dans le bruit

Enfin, pour pouvoir tester la validité de l'immersion du locuteur dans le bruit par le biais d'un casque, par comparaison à d'autres méthodes expérimentales, nous avons réalisé 3 sessions d'enregistrement de ces 9 conditions, en modifiant le dispositif d'immersion des locuteurs dans le bruit à chaque session :

- dans la première session, les locuteurs étaient immergés dans le bruit par le biais d'un casque fermé Sennheiser HD250 Linear II. Ce casque est de bande passante 50-40 kHz et de réponse

en fréquence quasiment plate, avec une légère atténuation de -3dB autour de 200Hz et également de -3dB environ autour de 3kHz . La voix du partenaire était redirigée dans le casque du locuteur à un niveau d'intensité compensant l'atténuation introduite par le port du casque.

- dans la deuxième session, les locuteurs étaient toujours immergés dans le bruit par le biais d'un même casque fermé, mais leur propre voix était cette fois retournée dans le casque en plus de celle de leur partenaire, à un niveau d'intensité sensé compenser l'atténuation introduite par le port du casque.
- enfin dans la troisième session, le bruit ambiant était diffusé par deux haut-parleurs (Tanoy System 600) situés à $1,5\text{m}$ des locuteurs (cf. Figure 37).

Les détails de calibration et d'implémentation technique de cette expérience sont présentés en Annexe Met2.

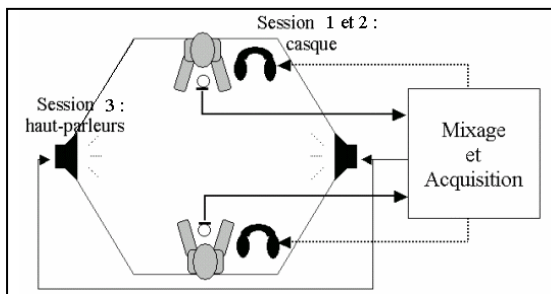


Figure 37. Placement des locuteurs et des enceintes dans la chambre assourdie. Les locuteurs sont immergés dans le bruit par le biais d'un casque (sessions 1 et 2) puis par haut-parleurs (session 3).

Signaux enregistrés

Pour chaque session et chaque condition, nous avons enregistré :

- le **signal audio** de parole à l'aide d'un microphone serre-nuque (Beyerdynamic Opus 54). Pour calibrer le niveau d'intensité acquis au microphone, nous avons enregistré le locuteur en début d'enregistrement sur un [a] tenu stable et mesuré cette intensité vocale de référence au sonomètre au niveau des lèvres du locuteur.
- le **signal électroglottographique** (EGG) à l'aide d'un électroglottographe (Glottal Enterprise EG2).

Ces deux signaux ont été préamplifiés à l'aide d'un RME Octamic puis numérisés sur 16 bits à 44.1kHz à l'aide d'un convertisseur (RME ADI 8 Pro) et d'une carte son (RME DIGI 9652 HDSP).

Questionnaire perceptif

Par ailleurs, après l'enregistrement de chaque condition, et ce pour les 3 sessions, nous avons demandé aux locuteurs d'évaluer sur des échelles à 5 points leur ressenti correspondant aux questions suivantes :

1- Comment entendiez-vous globalement votre propre voix lorsque vous parliez ?				
très bien	assez bien	pas très bien	difficilement	presque plus
2- Comment estimez-vous que votre interlocuteur vous a entendu ? (d'après ses mimiques faciales ou ses éventuelles demandes de répétition)				
très bien	assez bien	pas très bien	difficilement	presque plus
3-Comment entendiez-vous personnellement votre interlocuteur ?				
très bien	assez bien	pas très bien	difficilement	presque plus
4- Avez vous ressenti un inconfort lorsque vous parliez?				
Pas du tout	très légèrement	un peu	assez	beaucoup

Questionnaire sur le rapport de chaque locuteur avec sa voix

Par ailleurs, nous avons demandé à chaque locuteur à l'issue de l'enregistrement de répondre par écrit aux questions suivantes, afin de connaître leur parcours et leurs habitudes de vie relatives à la voix.

- (1) Avez-vous déjà travaillé votre voix ou pratiqué une activité vocale (chant, théâtre, orthophonie, autre...) ?
- (2) Utilisez-vous beaucoup votre voix au cours d'une journée ? Si oui, pouvez-vous décrire pour quel(s) usage(s) ?
- (3) Avez-vous déjà dans votre vie ressenti une gêne vocale importante (perte de voix, enrrouement important ou autres) ? Si oui, pouvez-vous décrire ce que vous avez ressenti, dans quelles situations et à quelle fréquence cela vous est-il arrivé ?

Le Tableau 3 présente un résumé des informations individuelles importantes.

	Genre	Age	(1) Travail de la voix	(2) Utilisation de la voix	(3) Problèmes vocaux
L1	Femme	24 ans	Un peu de chant	Peu importante	Non
L2	Femme	27 ans	Théâtre (semi-professionnelle)	Communicationnelle + théâtre	Voix qui décroche très souvent en répétition
L3	Homme	25 ans	Non	Peu importante	Plusieurs épisodes d'aphonie durable non expliqués
L4	Homme	25 ans	Chant lyrique	Communicationnel, se décrit comme bavard	Non
L5	Homme	23 ans	Non	Enseignement	Non
L6	Femme	22 ans	Un peu de chant choral et de théâtre	Communicationnelle, se décrit comme bavarde	Aphonie après concerts
L7	Homme	21 ans	Chant choral et lyrique	Communicationnelle, peu importante	Irritation après des concerts
L8	Homme	24 ans	non	Peu importante	Irritations et aphonie après concerts
L9	Femme	30 ans	non	Enseignement, se décrit comme bavarde	Après plusieurs heures d'enseignement
L10	Femme	28 ans	non	Peu importante	Non, mis à part des épisodes de bégaiement

Tableau 3. Caractéristiques individuelles des 10 locuteurs de la première base de données.

3.1.2. Analyse et traitement statistique des données

Pour examiner l'influence des différents moyens d'immersion dans le bruit sur la parole produite, nous avons considéré plusieurs paramètres classiquement utilisés pour décrire la parole Lombard. Certains ont été mesurés pour chaque locuteur et chaque condition sur les 16 mots-cibles. C'est le cas de la durée, de l'intensité moyenne et de la fréquence fondamentale moyenne. D'autres paramètres ont été mesurés sur les voyelles de ces mots-cibles. C'est le cas de la différence d'énergie entre les zones [1-2kHz], [2-4kHz] et [4-6kHz] par rapport à la zone du fondamental [0-1kHz], du quotient ouvert (Oq) de la différence entre les deux premiers harmoniques du spectre de l'onde de débit glottique dérivée (H1-H2), de la vitesse de fermeture des cordes vocales (E) et de la fréquence du premier formant vocalique. Enfin, des paramètres ont été mesurés sur l'ensemble de chaque condition : la dynamique d'intensité, l'ambitus de F0, la perception par le locuteur de sa propre voix (Question 1), de l'interlocuteur (Question 2), l'estimation par le locuteur de sa propre intelligibilité (Question 3) et de son inconfort (Question 4).

Pour la condition de silence et pour les 8 conditions de bruit (2 types de bruit x 4 niveaux de bruit), nous avons comparé la différence entre la parole produite en portant un casque et celle produite en étant immergé dans le bruit par le biais de haut-parleurs (cf. § 3.2) et la différence entre la parole produite en portant un casque avec ou sans retour auditif de sa propre voix réinjecté dans le casque (cf. § 3.3). Nous avons examiné la significativité de ces différences par un test ANOVA à un facteur (dans le § 3.2, ce facteur est le port du casque, dans le § 3.3, le facteur est l'introduction d'un retour auditif dans le casque). Ces résultats statistiques sont exposés en détail en Annexe Res3.

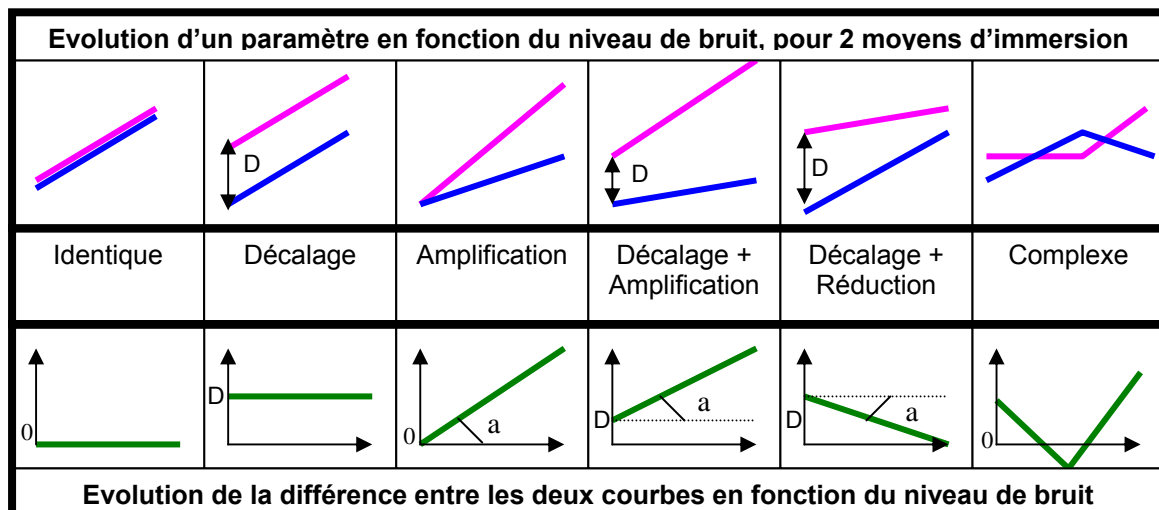


Figure 38. Représentation de 6 types de différence d'évolution d'un paramètre en fonction du niveau de bruit, pour 2 moyens d'immersion.

Afin d'alléger la lecture des résultats, nous proposons 6 types de profils pour modéliser la différence d'évolution d'un paramètre dans un bruit d'intensité croissante entre deux moyens d'immersion dans le bruit (cf. Figure 38) :

- (Identique) : Il n'existe pas de différence significative entre les deux moyens d'immersion pour ce paramètre, quel que soit le niveau de bruit : la différence entre les deux courbes d'évolution du paramètre est donc quasiment nulle, quel que soit le niveau de bruit.
- (Décalage) : L'évolution du paramètre est semblable entre les deux moyens d'immersion, à un décalage vertical « D » près. La différence entre les deux courbes d'évolution est donc significative et semblable pour chaque niveau de bruit. Le décalage D est alors estimé comme la moyenne des différences observées entre les deux courbes d'évolution pour chaque niveau de bruit. La significativité de cette différence globale entre les deux moyens d'immersion est testée en effectuant un test ANOVA entre les deux moyens d'immersion, cette fois tous niveaux de bruit confondus.
- (Amplification) : L'évolution du paramètre est semblable entre les deux moyens d'immersion, à un facteur d'amplification « a » près. La différence entre les deux courbes d'évolution n'est donc pas toujours significative dans le silence ou à faibles niveaux de bruit, mais est dans tous les cas linéairement croissante et significative pour les forts niveaux de bruit. La linéarité de cette

amplification en fonction du niveau de bruit est testée en effectuant une régression linéaire de la différence entre les deux courbes d'évolution en fonction du niveau de bruit et en observant son coefficient de corrélation « r ». Le facteur d'amplification « a » est alors estimé comme la pente de cette régression linéaire.

- (Décalage + Amplification ou Réduction) : Combinaison des deux points précédents.
- (Complexe) : L'évolution du paramètre est significativement différente entre les deux moyens d'immersion, pour chaque niveau de bruit ou pour certains niveaux de bruit seulement. Mais cette différence n'est ni constante ni linéairement croissante. Aucune relation simple ne se dégage entre ces deux évolutions.

3.2. Résultats concernant l'influence du port du casque

Les Tableau 4 à Tableau 8 récapitulent les différences observées entre l'immersion du locuteur par le biais d'un casque et l'immersion par le biais de haut-parleurs.

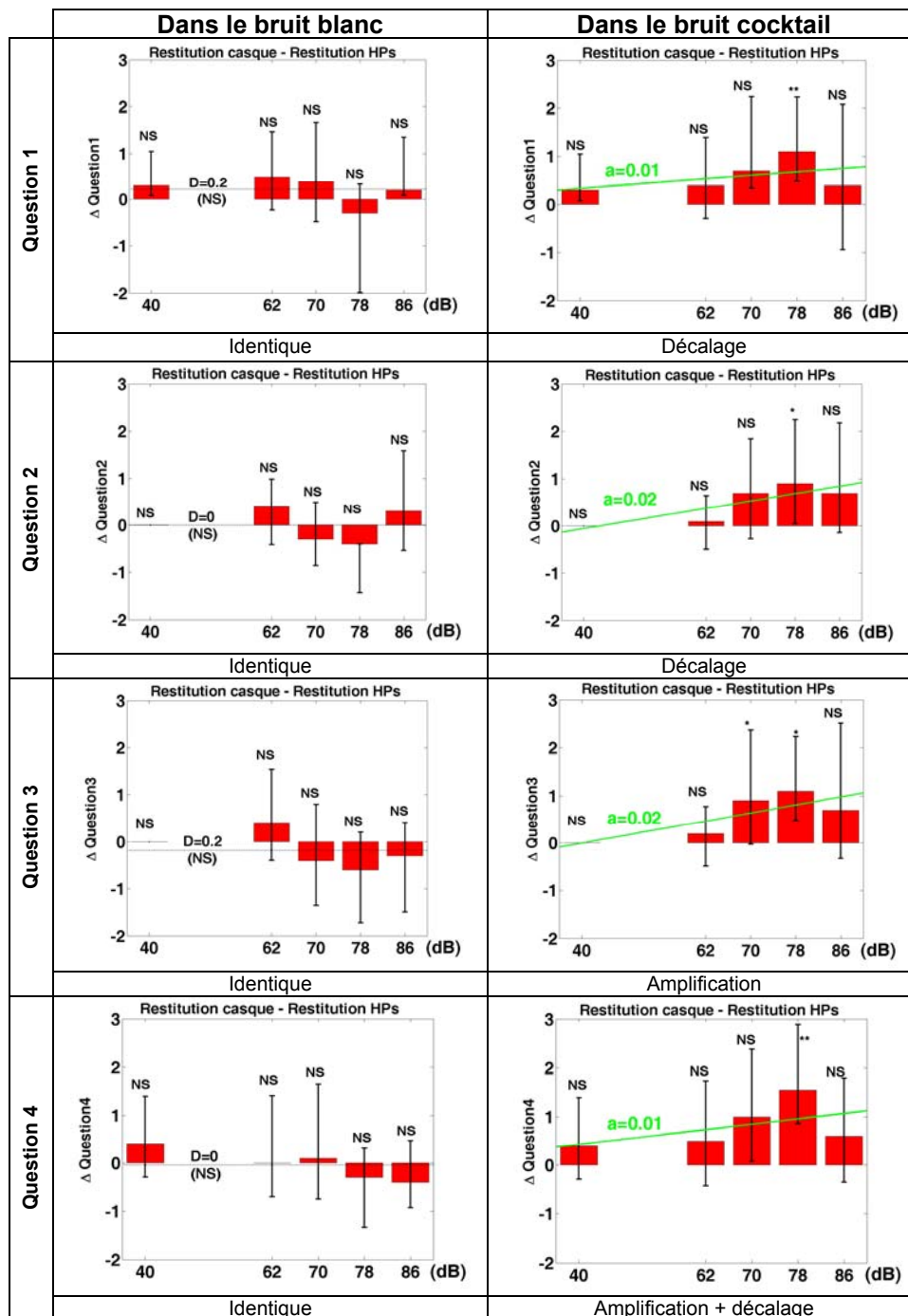


Tableau 4. Effet du port du casque par rapport à une restitution haut-parleurs sur la perception du locuteur de sa propre voix (Q1), sur son estimation de sa propre intelligibilité (Q2), sur sa perception de l'interlocuteur (Q3) et sur son ressenti de gêne (Q4), en fonction du niveau de bruit (silence, 62, 70, 78, 86 dB), pour le bruit blanc (à gauche) et le bruit cocktail (à droite).

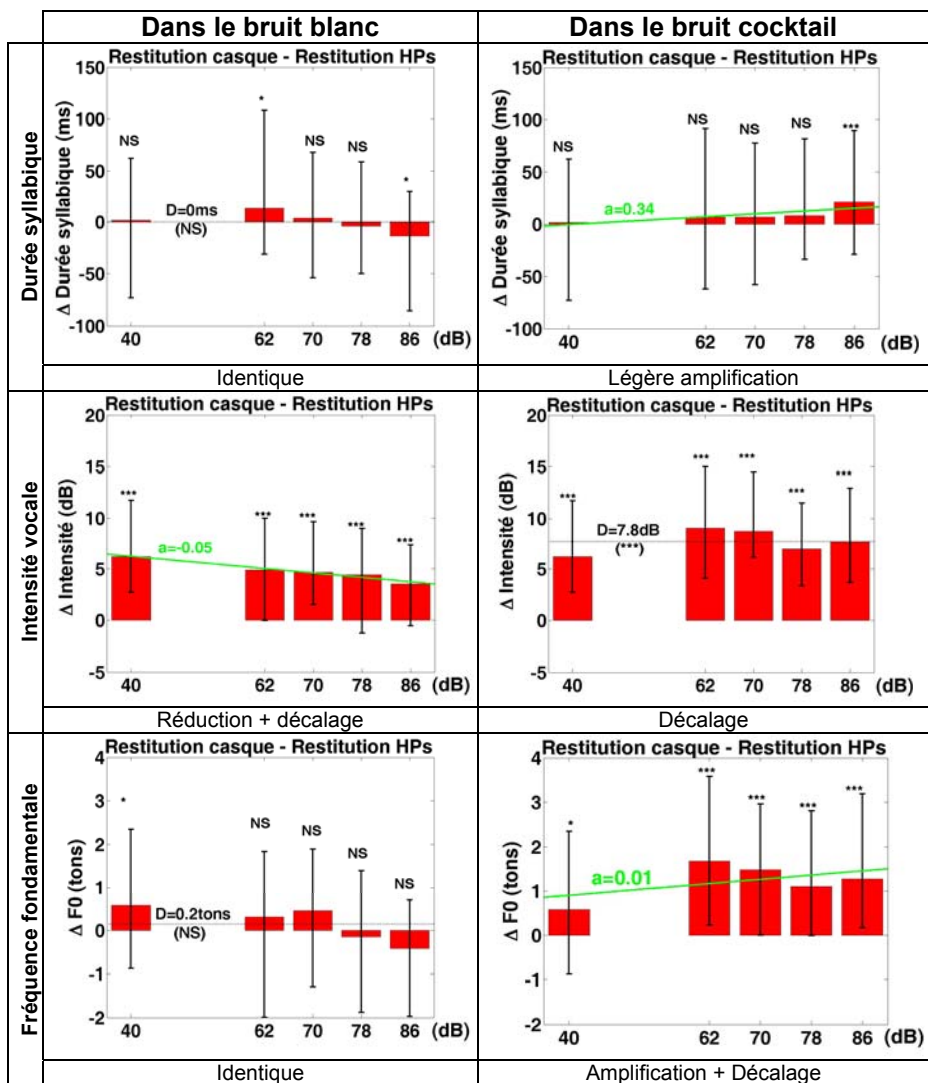


Tableau 5. Effet du port du casque par rapport à une restitution haut-parleurs sur l'évolution de la durée syllabique, de l'intensité vocale et de la F0 en fonction du niveau de bruit (silence, 62, 70, 78 et 86 dB), pour le bruit blanc (à gauche) et le bruit cocktail (à droite).

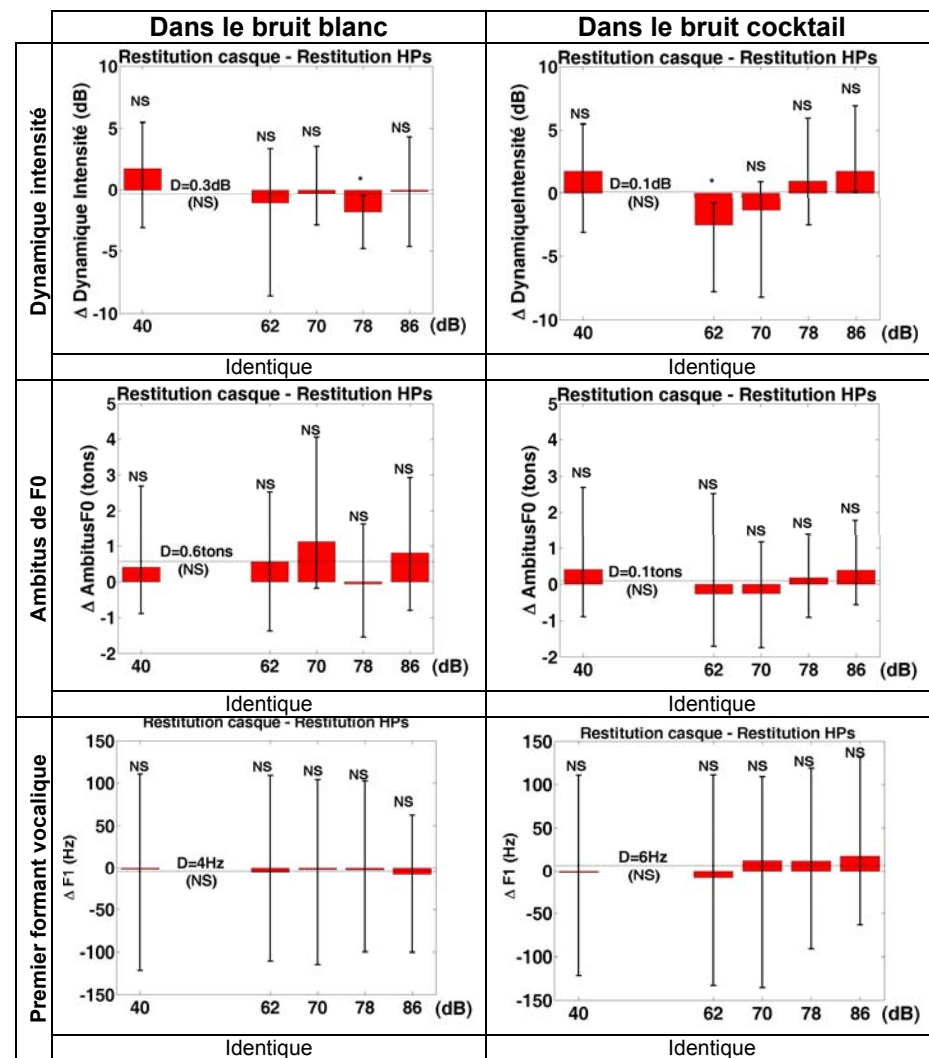


Tableau 6. Effet du port du casque par rapport à une restitution haut-parleurs sur l'évolution de la dynamique d'intensité, de l'ambitus de F0 et de la fréquence du premier formant vocalique en fonction du niveau de bruit (silence, 62, 70, 78 et 86 dB), pour le bruit blanc (à gauche) et le bruit cocktail (à droite).

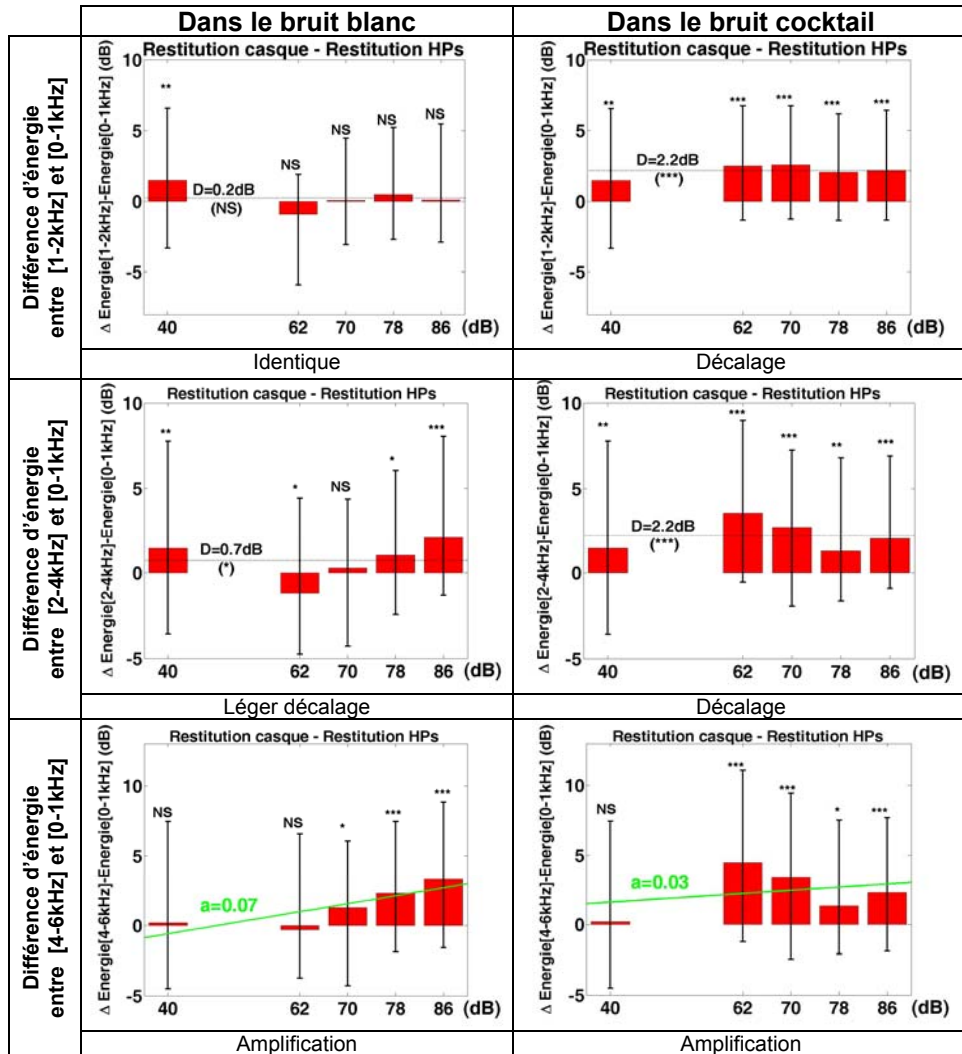


Tableau 7. Effet du port du casque par rapport à une restitution haut-parleurs sur l'évolution de l'énergie spectrale de la voix en fonction du niveau de bruit (silence, 62, 70, 78 et 86 dB), pour le bruit blanc (à gauche) et le bruit cocktail (à droite).

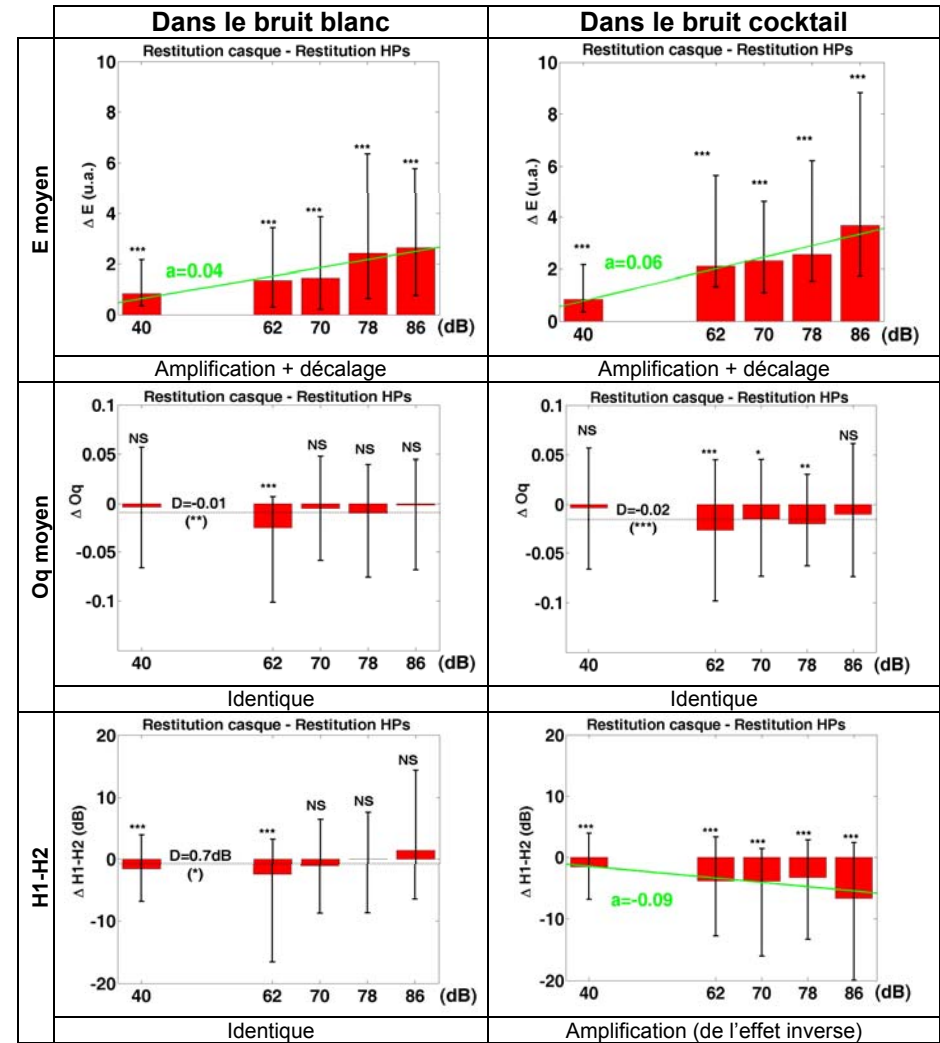


Tableau 8. Effet du port du casque par rapport à une restitution haut-parleurs sur l'évolution de la vitesse de fermeture laryngée, du quotient ouvert, de la différence (H1-H2) en fonction du niveau de bruit (silence, 62, 70, 78 et 86 dB), pour le bruit blanc (à gauche) et le bruit cocktail (à droite).

Il existe plusieurs paramètres sur lesquels le port du casque n'a pas d'effet significatif. Ainsi, la dynamique d'intensité, l'ambitus de F0, la fréquence du premier formant vocalique et le quotient ouvert ne montrent pas de différences significatives lorsque le locuteur porte ou non un casque. De même, dans le bruit blanc, la durée des mots, la fréquence fondamentale moyenne, la différence (H1-H2), la différence d'énergie entre [1-2kHz] et [0-1kHz], ainsi que le ressenti du locuteur ne sont pas significativement influencés par le port d'un casque.

D'autres paramètres montrent des différences significatives lorsque le locuteur est immergé dans le bruit par le biais d'un casque :

- c'est le cas de l'intensité vocale, de l'énergie spectrale entre 1 et 2kHz et entre 2 et 4kHz dans le bruit cocktail, globalement plus importants lorsque le locuteur porte un casque, aussi bien dans le silence que dans le bruit, sans pour autant que cet effet du port du bruit n'évolue avec le niveau de bruit.
- c'est également le cas du ressenti du locuteur, de la durée syllabique, de la fréquence fondamentale dans le bruit cocktail, et de la vitesse de fermeture laryngée comme de l'énergie spectrale de la voix entre 4 et 6 kHz pour les deux types de bruit, plus importants lorsque le locuteur porte un casque, avec un effet croissant avec le niveau de bruit. Au contraire, la différence (H1-H2) est moins importante lorsque le locuteur est immergé dans le bruit par haut-parleurs que par casque, et cet effet s'amplifie avec le niveau de bruit ambiant croissant. Enfin, on observe également une intensité vocale toujours plus importante dans le bruit blanc lorsque le locuteur porte un casque mais cet effet tend à diminuer avec un niveau de bruit ambiant croissant.

Les paramètres influencés par le port d'un casque et croissant en fonction du niveau de bruit sont plus forts en immersion dans le bruit par casque que par haut-parleurs et cela, même dans le silence. De la même manière, Davis et al. 2006^[62] ont récemment constaté que les mouvements articulatoires sont plus amples lorsque le locuteur est immergé dans le bruit par le biais d'un casque plutôt que par haut-parleurs. Il semble donc que le port d'un casque induise un effet Lombard supplémentaire. Cet effet, quantifiable grâce au décalage entre les deux courbes d'évolution de l'intensité moyenne de la voix lorsque le locuteur est immergé dans le bruit par casque et par haut-parleurs, est plus important dans le bruit cocktail (7.8 dB, ***) que dans le bruit blanc (4.7 dB, ***). Il semble donc que l'effet du port du casque ne concerne pas seulement l'atténuation du retour auditif aérien (de 10 dB dans le cas du casque utilisé ici), mais exerce également une influence sur la façon dont le bruit est perçu et perturbe le retour auditif du locuteur. Ces résultats peuvent s'expliquer par le fait que le locuteur a un retour essentiellement basses-fréquences de sa propre voix lorsqu'il porte un casque, du fait que son retour aérien est coupé et sa conduction osseuse au contraire amplifiée. L'énergie du bruit de conversations étant justement concentrée dans les basses-fréquences, il est donc possible qu'il perturbe alors plus le contrôle de sa propre voix que le bruit blanc, dont l'énergie est répartie sur l'ensemble des fréquences. Les résultats des évaluations perceptives vont également dans ce sens puisque le locuteur ne perçoit pas sa voix ou celle de son interlocuteur de façon significativement différente dans le bruit blanc lorsque celui-ci est restitué par casque ou par haut-parleurs, tandis qu'au contraire sa perception est nettement dégradée par le port d'un casque lorsque le bruit diffusé est un bruit de conversations.

3.3. Résultats concernant l'influence de l'ajout d'un retour auditif dans le casque

Les Tableau 9 et Tableau 10 récapitulent les différences observées lorsqu'un retour auditif de sa propre voix est ajouté dans le casque, seulement pour les paramètres ayant montré dans le paragraphe précédent une sensibilité au port d'un casque par le locuteur.

Pour la majorité des paramètres, le rajout d'un retour auditif n'induit pas de changements significatifs sur la production de parole et ne permet pas, par conséquent, de compenser l'effet du port d'un casque par rapport à une immersion dans le bruit par le biais de haut-parleurs: c'est le cas pour l'énergie spectrale de la voix entre 2 et 4kHz et entre 4 et 6kHz pour les deux types de bruit, et pour le bruit cocktail, de l'énergie spectrale de la voix entre 1 et 2kHz, de la différence H1-H2, de la durée syllabique, de la F0, de l'estimation de sa propre intelligibilité, de la perception de l'interlocuteur et du ressenti de gêne.

Quelques paramètres sont cependant influencés par l'ajout d'un retour auditif dans le casque, cette influence compense partiellement l'effet induit par le port d'un casque sur l'adaptation au bruit par rapport à une immersion par haut-parleurs : c'est le cas de l'intensité vocale et de la vitesse de fermeture laryngée pour les deux types de bruit. Au contraire, et très paradoxalement, l'ajout d'un

retour auditif dans le casque n'améliore pas la perception que le locuteur a de sa propre voix dans le bruit, et au contraire amplifie l'effet lié au port d'un casque pour un niveau de bruit ambiant croissant.

	Dans le bruit blanc	Dans le bruit cocktail
Intensité		
	Décalage compensant partiellement l'effet du port du casque	Réduction + décalage compensant partiellement l'effet du port du casque
Différence d'énergie entre [2-4kHz] et [0-1kHz]		
	Identique. Pas de compensation de l'effet du port du casque	Identique. Pas de compensation de l'effet du port du casque
Différence d'énergie entre [4-6kHz] et [0-1kHz]		
	Identique. Pas de compensation de l'effet du port du casque	Identique. Pas de compensation de l'effet du port du casque
Vitesse de fermeture laryngée E		
	Amplification compensant partiellement l'effet du port du casque	Amplification compensant partiellement l'effet du port du casque

Tableau 9. Effet du rajout d'un retour auditif dans le casque sur l'évolution de l'intensité vocale, de l'énergie spectrale de la voix entre [2-4kHz] et [4-6kHz] et de la vitesse de fermeture glottique E en fonction du niveau de bruit (silence, 62, 70, 78 et 86 dB), pour le bruit blanc (à gauche) et le bruit cocktail (à droite). La courbe bleue superposée représente l'effet du port même d'un casque par rapport à une restitution haut-parleurs, (cf. §3.2) afin de visualiser si l'ajout d'un retour auditif dans le casque permet de compenser cet effet du port d'un casque.

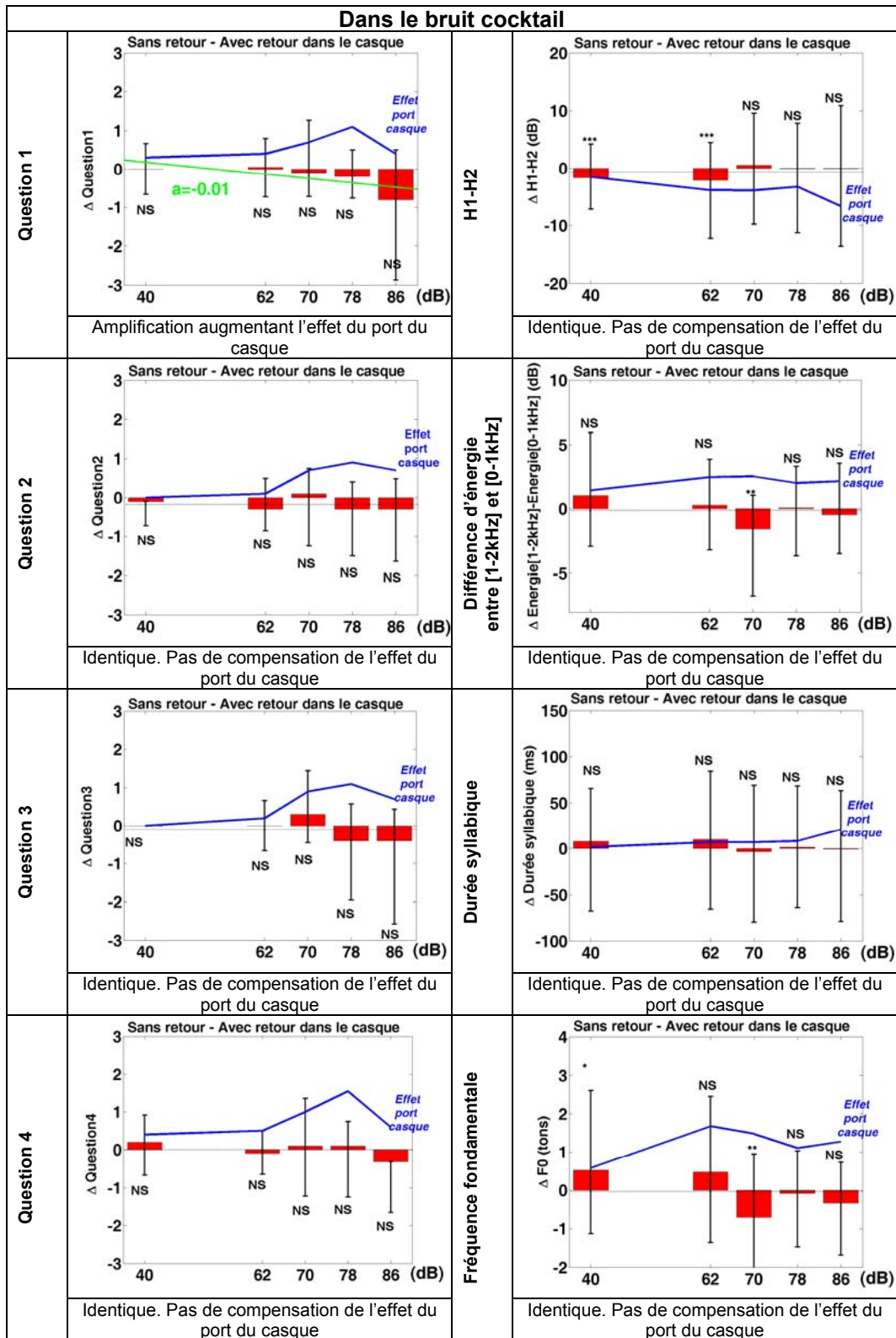


Tableau 10. Effet du rajout d'un retour auditif dans le casque sur la perception de sa propre voix (Q1), sur l'estimation de sa propre intelligibilité (Q2), sur la perception de l'interlocuteur (Q3), sur le ressenti de gêne (Q4), sur l'énergie spectrale de la voix entre [1-2kHz], sur la différence H1-H2, sur la durée syllabique et sur la F0, en fonction du niveau de bruit (silence, 62, 70, 78 et 86 dB), et seulement pour le bruit cocktail. La courbe bleue superposée représente l'effet du port même d'un casque par rapport à une restitution haut-parleurs (cf. §3.2), afin de visualiser si l'ajout d'un retour auditif dans le casque permet de compenser cet effet du port d'un casque.

Ces résultats confirment ceux de Kadiri 1998^[176] ayant noté une absence d'effet du retour auditif dans le casque sur la durée des mots et sur le premier formant des voyelles. Nos résultats vont également dans le même sens que ceux de Laukkanen et al. 2004^[200], observant une diminution de l'intensité et de la tension vocale lorsque les locuteurs ont un retour de leur voix dans un casque, avec une diminution du quotient ouvert et de la pression orale. Au contraire, Kadiri 1998^[176] a noté une augmentation de l'intensité vocale plus importante dans le bruit lorsque le locuteur a un retour de sa propre voix dans le casque, ce qui semble contre-intuitif et va à l'encontre des résultats des études menées sur l'effet Sidetone observant au contraire une diminution de l'intensité vocale en compensation de l'augmentation du retour auditif de sa propre voix (Lane et al. 1961^[193]; Lane et al. 1970^[194]).

En revanche, Kadiri 1998^[176] rapporte que la fréquence fondamentale augmente moins dans le bruit lorsque le locuteur dispose d'un retour de sa propre voix dans le casque et Laukkanen et al. 2004^[200] note que les locuteurs témoignent d'une meilleure perception de leur voix, leur paraissant plus forte et plus brillante. Nous n'avons pas observé de tels résultats dans notre étude.

Pour la majorité des descripteurs de la parole examinés, l'ajout d'un retour auditif conserve ou compense partiellement l'effet induit par le port d'un casque par rapport à une immersion dans le bruit par haut-parleurs. Il est donc toujours plus valide d'ajouter un retour auditif dans le casque si l'on souhaite utiliser cette méthode. Cependant, la diminution de l'écart qu'apporte l'ajout d'un retour auditif n'est pas suffisante et il existe toujours une différence significative par rapport à l'immersion dans le bruit par haut-parleurs. Cette différence n'est pas négligeable :

- non seulement par rapport à l'erreur d'estimation liée au débruitage. L'effet du port du casque (avec retour) par rapport à une immersion naturelle est en effet de l'ordre de 5dB (***) dans le bruit de conversations tandis que l'erreur d'estimation de l'intensité peut atteindre 1.5 dB au grand maximum avec la Méthode NC.
- mais aussi par rapport à l'effet Lombard que l'on souhaite observer par ce protocole : l'augmentation de l'intensité dans un bruit de conversations de 86dB est en effet de 12-14 dB par rapport à l'intensité dans une condition calme, tandis que l'effet du port du casque dans le bruit de conversations est de 5dB (***) avec retour auditif, et de 7.8dB (***) sans.

Il serait intéressant de renouveler cette expérience en faisant varier de façon croissante le retour auditif réinjecté dans le casque du locuteur afin de voir s'il existe un niveau plus important de ce retour pour lequel l'effet du port du casque peut être totalement compensé ou si le port d'un casque exerce dans tous les cas une influence significative sur le comportement vocal d'un locuteur.

3.4. Conclusion sur les méthodes d'immersion du locuteur dans le bruit

Les résultats obtenus montrent que le port du casque est valide en l'état pour observer la modification dans le bruit de certains paramètres de la parole, tels que la dynamique d'intensité, l'ambitus de F0, le quotient ouvert ou le premier formant vocalique. De même, la fréquence fondamentale ou la durée syllabique ne sont pas influencées dans le bruit blanc par le port d'un casque.

Pour tous les autres paramètres, le port du casque, même avec rajout d'un retour auditif, introduit des modifications significatives, davantage pour un bruit de conversations que pour un bruit blanc. Cependant, l'immersion de locuteurs par le biais d'un casque peut être considérée comme valide, à condition de prendre en compte son effet, pouvant être modélisé simplement pour certains paramètres par un facteur correctif d'amplification et/ou un facteur correctif de décalage. C'est le cas par exemple pour l'étude de l'intensité moyenne ou de la vitesse de fermeture laryngée.

Cependant, la modélisation de l'effet du port du casque sous forme d'une amplification et/ou d'un décalage est une première approximation. En réalité, il semble que l'évolution de certains paramètres de la voix en fonction du niveau de bruit ne soit pas linéaire. Nous reviendrons en détail sur ce point au chapitre 10. Cette constatation a pour conséquence qu'une transformation de type homothétie et translation ne permet pas de modéliser précisément l'effet du port du casque et de prévoir la valeur d'un paramètre en condition naturelle à partir de sa valeur mesurée en condition d'immersion par casque. Cette imprécision n'est pas vraiment gênante dans le cas où l'effet Lombard est étudié davantage pour l'effort vocal qu'il induit que pour lui-même. En revanche, cette approximation n'est pas rigoureuse dans le cas où l'on s'intéresse à l'effet Lombard d'un point de vue psychoacoustique ou pour caractériser l'évolution des paramètres de description de la parole en fonction du niveau de bruit ambiant.

Ces différentes remarques nous amènent à effectuer certains choix méthodologiques pour ce travail :

- lorsque cela ne nuit pas à la rigueur de l'étude, nous cherchons à nous rapprocher autant que possible d'une situation de communication naturelle. Les résultats de cette étude montrent que le port du casque a non seulement une influence significative sur la perception que le locuteur a de sa propre voix, mais également sur la perception de la voix de l'interlocuteur. Nous verrons dans le chapitre suivant que l'interaction entre les interlocuteurs joue un rôle non négligeable sur l'adaptation du locuteur en environnement bruyant. Aussi, ces différents points nous incitent à privilégier une immersion du locuteur dans le bruit par le biais de haut-parleurs.
- dans cette étude, nous ne nous intéressons pas seulement à provoquer un effort vocal mais à étudier les stratégies d'adaptation d'un locuteur à une situation de communication (type de bruit, niveau de bruit). L'effet Lombard n'est pas pour nous nécessairement un phénomène linéaire en fonction du niveau de bruit qui traduirait l'existence d'une unique stratégie d'adaptation, graduelle. Au contraire, nous nous intéressons aux possibles changements de stratégies d'adaptation en fonction du niveau de bruit (cf. chapitre 10). Dans cette optique, il semble donc plus prudent de contrôler précisément le niveau de bruit et donc d'immerger le locuteur dans le bruit par haut-parleurs.

Cependant, les résultats de ce chapitre nous amène à effectuer un bilan à plus long terme : Que l'effet Lombard soit étudié grâce à une immersion du sujet par casque ou par haut-parleurs, l'imprécision des mesures engendrée par ces deux alternatives est importante : l'effet du port du casque est aussi grand que la moitié de l'effet étudié, tandis que l'erreur d'estimation introduite par le débruitage peut atteindre 10% de l'effet étudié. Nous sommes conscients de l'incertitude de nos mesures mais nous ne disposons malheureusement pas pour l'instant de meilleurs moyens techniques. L'effort de recherche dans ce domaine ne doit cependant pas s'arrêter là. En particulier, en parallèle de l'amélioration des techniques de débruitage du signal, il serait également intéressant d'améliorer l'écologie de la restitution d'environnements sonores par casque. Plusieurs travaux œuvrent ainsi à développer la spatialisation au casque (Blum et al. 2004 ^[27]). De même, nous avons récemment eu connaissance de l'existence de casques d'écoute n'atténuant pas le retour auditif. Il serait intéressant refaire cette étude comparative sur le port du casque avec de telles techniques.

La comparaison de ces différentes méthodologies d'immersion d'un locuteur dans le bruit nous a amené à enregistrer une première base de donnée (BD1) conséquente, constituées de 10 locuteurs, de 2 types de bruit et de 4 niveaux de bruit. Aussi, nous explorerons de nouveau et plus en détail la 3^{ème} session de cette base de données, où l'environnement bruyant a été diffusé par haut-parleurs, pour caractériser les différences interindividuelles de stratégies d'adaptation à un environnement bruyant (dans les chapitres 6, 7, 8 et 11).

Chapitre 4 :

Influence du dialogue avec un interlocuteur sur l'adaptation de la parole dans le bruit

- 1. RECHERCHE D'ALTERNATIVES A LA LECTURE PAR UN LOCUTEUR ISOLE**
- 2. COMPARAISON DE DEUX TACHES DE PAROLE AVEC ET SANS INTERACTION AVEC UN INTERLOCUTEUR**
- 3. CONCLUSION DU CHAPITRE 4**

Dès le premier chapitre, nous avons défendu l'idée que dans la plupart des situations, en particulier celles vécues par les enseignants ou les personnes dont la voix est un outil de travail, l'effort vocal n'est pas subi ni fourni sans but, mais dans l'intention de communiquer un message à un destinataire. En plus de tous ces aspects physiologiques et biomécaniques, le problème du forçage vocal pourrait également consister en une mauvaise estimation chez certaines personnes de l'effort à fournir par rapport aux exigences d'intelligibilité requises par la situation de communication. C'est sur cet aspect du forçage vocal, parmi d'autres, que nous avons choisi de nous focaliser dans cette thèse.

L'estimation de l'effort de production de parole résulte, comme nous l'avons vu au chapitre 2, d'une analyse de la situation par le locuteur, en particulier du bruit dans lequel est immergé l'interlocuteur et qui perturbe sa perception, mais aussi de sa distance, de son niveau de compréhension (est-ce un enfant, une personne de langue étrangère, etc), des modalités sensorielles dont les deux partenaires disposent pour communiquer, ou encore de la prévisibilité du message à transmettre en fonction du contexte. Ensuite, le réajustement constant de la stratégie mise en place est rendu possible par la vérification que le locuteur effectue de son action, grâce à ses retours sensoriels, mais également grâce aux signes de compréhension ou non de l'interlocuteur. L'interlocuteur peut donc participer à la fois de façon « passive » à l'adaptation du locuteur en environnement bruyant, dans le sens où sa présence, sa distance, son niveau de compréhension font partie intégrante de la situation de communication, et vont être évaluées par le locuteur pour mettre en place une stratégie adaptée. Mais il peut également participer de façon active à cette adaptation par le retour qu'il fournit au locuteur de la réussite ou non de son action.

C'est pourquoi nous avons cherché à mettre en place de nouvelles tâches de production de parole en laboratoire, alternatives à la lecture par un locuteur isolé, et permettant malgré tout de contrôler une partie du matériel linguistique tout en laissant de l'imprévisibilité au discours et de la spontanéité aux locuteurs. Cette initiative résulte de la volonté de reconstituer en laboratoire une situation écologique d'adaptation au bruit, en se rapprochant d'une situation naturelle vécue quotidiennement par les sujets, mais surtout de tenir compte de mécanismes cognitifs et communicationnels conditionnant l'adaptation de la parole dans le bruit, et sans lesquels notre étude n'aurait donc pas de sens.

Grâce à ces nouvelles méthodologies, nous avons ensuite pu mener une étude comparative entre plusieurs tâches de parole, afin d'examiner l'influence du dialogue avec un interlocuteur sur l'adaptation dans le bruit. Nous présenterons dans le 2^{ème} paragraphe les résultats de cette étude et concluons sur la validité ou non des tâches non interactives pour l'étude de la parole Lombard.

1. RECHERCHE D'ALTERNATIVES A LA LECTURE PAR UN LOCUTEUR ISOLE

La dimension de communication et de dialogue avec un interlocuteur est pour nous au centre du problème de l'adaptation en environnement bruyant et du forçage vocal. Pourtant, lorsqu'on effectue une revue de la littérature sur le forçage vocal ou sur la parole Lombard, on s'aperçoit que très peu d'études antérieures la prennent en compte au niveau de la situation expérimentale (cf. tableau 2 de l'annexe Bib2). Mis à part les études in situ (Dohalska et al. 2000^[73] ; Dieroff et al. 1966^[71] ; Van Heusden et al. 1979^[360]), les autres études en laboratoires enregistrent toutes des locuteurs en tâche de lecture. La plupart du temps, aucun interlocuteur n'est présent ni suggéré : le locuteur est alors isolé dans une pièce et lit des listes de mots ou de phrases (Junqua 1993^[171] ; Castellanos et al. 1996^[46] ; Van Summers et al. 1988^[361]). Dans certaines études, un interlocuteur est physiquement présent ou suggéré au locuteur, mais il ne donne aucun retour de sa compréhension au locuteur (Kim 2005^[180]). Une seule étude a tenu à donner au locuteur un retour de son intelligibilité, par l'intermédiaire d'un vu-mètre contrôlé de l'extérieur par l'expérimentateur (Ternström et al. 2006^[345] ; Ternström et al. 2006^[345]). Cependant, les phrases lues étant totalement prévisibles par l'expérimentateur, il est dans ce genre de cas difficile de parler de recherche d'intelligibilité de la part du locuteur.

De tels protocoles nous semblent à réadapter par rapport à nos objectifs et notre conception de la problématique :

- Tout d'abord, indépendamment de l'effet Lombard, il est connu que la parole lue diffère de la parole spontanée (Lieberman et al. 1985^[213], Ayers 1994^[17]). Celle-ci est plus lente, davantage articulée, et atteint par conséquent davantage les cibles articulatoires. Cela se traduit au niveau formantique

par une expansion du triangle vocalique dans l'espace F1/F2 par rapport à la parole « naturelle » (Lindblom 1990 ^[217], Meunier et al. 2006 ^[238]). La validité ou non d'utiliser une telle tâche dépend de la problématique étudiée : Lorsqu'il s'agit d'études sur des propriétés spécifiques de la langue ou sur les capacités articulatoires du locuteur, le recours à de la parole lue peut être tout à fait justifié. De même, si l'effet Lombard est étudié d'un point de vue purement psychoacoustique, le type de parole n'est pas tant important (de nombreuses études se sont d'ailleurs intéressées à l'effet Lombard dans le chant ou chez les animaux) puisque seule l'évolution relative silence-bruit est finalement considérée. Par contre, si le but de l'étude est de modéliser un certain type de parole afin qu'un système de reconnaissance vocale puisse ensuite le traiter, il est plus approprié de travailler sur le mode d'énonciation spontané auquel seront finalement confrontés ces systèmes automatiques par la suite. De même dans notre cas, notre but est d'étudier des stratégies individuelles d'adaptation à une situation de communication, et nous faisons l'hypothèse que ces stratégies peuvent être communicationnelles et pas uniquement vocales. Aussi, il nous paraît nécessaire de laisser de la liberté au locuteur afin de ne pas restreindre son champ d'adaptation à une situation expérimentale trop rigide, au risque de passer alors à côté de certains aspects importants de son adaptation.

- Mais le problème fondamental de ces tâches ne réside pas tant dans la lecture que dans l'isolation du locuteur et dans l'absence d'interlocuteur. Dans l'absolu, cette situation peut paraître absurde et non naturelle : le locuteur n'est jamais confronté à l'expérience de devoir parler à voix haute sans s'adresser à personne. Mais là encore, le naturel n'est pas le principal problème. Le plus gênant est que la notion de communication implique fondamentalement un interlocuteur et une interaction, et qu'on ne peut pas parler d'intelligibilité hors situation de communication. Dans notre cas où nous postulons que l'adaptation du locuteur est motivée par la recherche d'intelligibilité, la présence d'un interlocuteur est donc essentielle. Par ailleurs, la présence « imaginée » d'un interlocuteur n'est pas suffisante. Le locuteur ne fait pas que s'adresser à un interlocuteur, il s'adapte à sa distance, à son niveau de compréhension, aux modalités sensorielles dont il dispose pour le percevoir. Sans interlocuteur réel et physiquement présent, le locuteur n'a pas de moyen d'évaluer la situation et ne dispose pas de représentations mentales de ce genre de situation, construites à partir d'expériences passées, qui lui permettent d'anticiper les conséquences de sa stratégie de communication. Enfin, la recherche d'intelligibilité n'a de sens que si la prévisibilité du message n'est pas totale. A partir du moment où l'interlocuteur connaît le message avant que le locuteur ne le produise, ce dernier n'a alors pas de raison de fournir d'effort pour être intelligible.

Pour satisfaire ces différents points, l'étude de la parole totalement spontanée serait bien évidemment idéale. Cependant, elle a été principalement abandonnée au profit de tâches de lecture, pour satisfaire à une démarche scientifique imposant la paramétrisation d'un phénomène et le contrôle des conditions expérimentales, nécessaires à la répétitivité des mesures et à la comparaison de plusieurs conditions ou de plusieurs individus. Nous pensons également que les conditions expérimentales doivent être contrôlées au maximum, mais non au point de mettre de côté les facteurs constitutifs du phénomène à étudier, et par là même de passer à côté de l'objet d'étude. Aucune solution n'existant pour maximiser à la fois la spontanéité du discours et le contrôle du matériel linguistique, l'étude de la parole Lombard nécessite selon nous un compromis adapté au niveau de détail des questions que l'on souhaite étudier. Par exemple, des mesures de F0 ou d'intensité moyenne sur des temps relativement longs, comme ce qui est réalisé dans des études in situ (Sodersten et al. 2002 ^[321]), ne nécessitent pas un contrôle précis du matériel linguistique, et peuvent être réalisées de façon rigoureuse sur de la parole spontanée. Au contraire, l'étude de caractéristiques phonétiques et prosodiques plus précises nécessite une syntaxe contrôlée des énoncés, un contexte phonémique et syllabique contrôlé, qui vont obligatoirement restreindre le champ de liberté du locuteur. Nous considérons ce compromis à trouver, comme un choix méthodologique sérieux d'un point de vue épistémologique, et non pas comme un relâchement sur la rigueur de l'étude. Un tel compromis nécessite d'inventer de nouvelles tâches intermédiaires entre la parole spontanée et la parole lue, permettant de contrôler une partie du matériel linguistique tout en laissant de l'imprévisibilité au discours et de la spontanéité aux locuteurs. Nous commencerons dans le premier paragraphe par effectuer une revue des premières initiatives dans ce sens, répertoriées dans la littérature, dont nous nous sommes inspirés pour imaginer deux jeux interactifs correspondant à des niveaux de compromis différents entre le contrôle des conditions expérimentales et la spontanéité du discours. Leurs principes, avantages et inconvénients seront exposés au deuxième paragraphe.

1.1. Les premières initiatives dans la littérature

De telles tâches semi-spontanées ont commencé à être développées initialement pour l'étude du dialogue entre les individus (collaborations, négociation, etc). La « Map task » en particulier (Brown et al. 1983 ^[36]), consiste en un jeu faisant intervenir deux partenaires qui doivent se communiquer des informations concernant un chemin dessiné sur une carte. Les deux locuteurs se répartissent un rôle de meneur et un rôle de suiveur : le meneur dispose de l'intégralité d'une carte sur laquelle sont disposés des objets et d'un chemin passant autour de ces objets. Le suiveur, lui, ne dispose que de la carte avec seulement certains des objets. Le meneur doit lui décrire le chemin sur la carte de façon à ce que le suiveur arrive à tracer le chemin sur sa propre carte. Les deux locuteurs doivent donc collaborer puisque le chemin n'est pas évident et certains objets manquent sur la carte du suiveur. Par la suite, de telles tâches de description de route (Levelt et al. 1983 ^[210]) ou de jeu de tangram (Clark et al. 1986 ^[55]) ont été utilisées pour l'étude intonative de la parole, complétées par d'autres jeux (Schafer et al. 2001 ^[303] ; Peng et al. ^[266]). Enfin, le domaine de la parole émotionnelle connaît un problème similaire au nôtre sur la parole Lombard : Par constitution, l'émotion réelle ne peut être que spontanée, et par conséquent l'étude de l'émotion actée au cours de la lecture de phrases est certes bien contrôlée, mais il s'agit d'un tout autre objet d'étude que l'émotion naturelle. Aubergé et al. 2005 ^[14] ont ainsi réutilisé des dispositifs de « magicien d'Oz », initialement dédiés à l'étude de la communication homme-machine, pour provoquer des émotions naturelles chez des locuteurs tout en contrôlant le matériel linguistique. Enfin, depuis le début de ma thèse, une étude très récente a enregistré un locuteur en environnement bruyant en situation de jeu du type « Pictionary® » (Papon 2006 ^[265]). Puis le locuteur devait communiquer à un interlocuteur des non-mots, que ce dernier devait écrire sur une ardoise.

1.2. Nos propositions. Intérêts et limites

Nous nous sommes inspirées de ces tâches interactives pour concevoir deux nouveaux jeux plus adaptés aux points précis que nous souhaitons examiner concernant la parole Lombard. Le premier jeu, assez libre, a été conçu pour l'étude globale, au niveau des énoncés et des mots, des stratégies individuelles d'adaptation acoustique et glottique en environnement bruyant (BD1). L'étude de caractéristiques articulatoires et phonétiques plus précises nécessitant le contrôle de la structure syntaxique des énoncés et du contexte segmental a motivé la conception du deuxième jeu plus contrôlé (BD2).

1.2.1. Un premier jeu assez libre

1.2.1.1. Contraintes

En s'inspirant de la « Map task » (Brown et al. 1983 ^[36]), nous avons cherché à construire une tâche de jeu interactif nécessitant l'usage d'un certain nombre de mots cibles (15 dans notre cas) au cours d'un dialogue non attendu à l'avance par l'interlocuteur, de façon à ce que le locuteur recherche réellement à être intelligible dans le cadre d'un vocabulaire limité.

Les niveaux de bruit étant parfois très forts (jusqu'à 85dB SPL), les conditions étudiées étant assez nombreuses, nous avons cherché à éviter que l'expérimentateur ne soit obligé de donner la réplique à chaque locuteur et de subir le bruit pour chaque enregistrement. C'est pourquoi nous avons imaginé une tâche où les locuteurs interagissent par binômes (dans le cas de la BD1, 5 binômes ont été enregistrés), ce qui permet d'acquérir deux corpus en parallèle. Malheureusement, la « Map Task » pose le problème que l'on ne contrôle pas parfaitement le fait que les locuteurs vont prononcer tous les mots cibles. Par ailleurs, les rôles du meneur et du suiveur sont très dissymétriques, en particulier le suiveur n'est pas obligé de parler.

Pour les mêmes raisons, nous avons minimisé autant que possible la durée du jeu afin de ne pas exposer de façon inutile les locuteurs au bruit. Le jeu de « map task » ne convenait pas non plus sur ce point puisque chaque « partie » pouvait être relativement longue, et le temps de parole analysé (les mots-cibles) ne représentait finalement qu'un dixième de la durée d'une partie.

C'est pourquoi nous avons imaginé un nouveau jeu, fortement inspiré de la « map task », mais mieux adapté aux contraintes de notre expérience.

1.2.1.2. Principe

Ce jeu inclue deux partenaires : un "meneur" et un "suiveur" (cf. Figure 39). Le meneur dispose d'une carte sur laquelle sont disposées des images correspondant aux mots-cibles. Un chemin est tracé sur cette carte et relie la moitié de ces items (tracé en pointillée sur la Figure 40). L'autre moitié des items est libre. Le "suiveur" dispose d'une liste d'associations mettant en relation les items déjà reliés avec chacun des items libres (cf. Figure 40). Le but du jeu consiste à ce que les deux interlocuteurs échangent les informations complémentaires dont ils disposent pour parvenir à reconstituer, dans l'ordre, un deuxième chemin reliant les huit items libres (tracé en gras sur la Figure 40).



Figure 39. Enregistrement des locuteurs par binômes. L'un joue le rôle de « meneur », l'autre de « suiveur ».

Le "meneur" doit décrire librement au "suiveur" le chemin tracé sur sa carte (représenté en pointillés sur la figure). Il peut dire par exemple: "*Je pars du requin*", puis "*Je vais ensuite vers le cochon*", etc. Pour chaque nouvelle étape du chemin citée par le "meneur", le "suiveur" doit lui répondre en lui indiquant quel est l'item libre correspondant sur la liste d'association (cf. Figure 40). Il va donc dire par exemple "*Le requin est associé au chausson*", puis "*le cochon va avec la vallée*", etc. Au fur et à mesure, le "meneur" va pouvoir tracer sur sa feuille le nouveau chemin (représenté en gras sur la Figure 40) et le "suiveur" va pouvoir noter sur sa feuille l'ordre dans lequel les items libres s'organisent pour constituer le nouveau chemin (cf. Figure 40). Lorsque le chemin de la carte du "meneur" a été parcouru jusqu'au bout, le nouveau chemin à trouver est donc normalement découvert par les deux partenaires. Le "meneur" doit alors le récapituler et s'assurer de la confirmation du "suiveur".

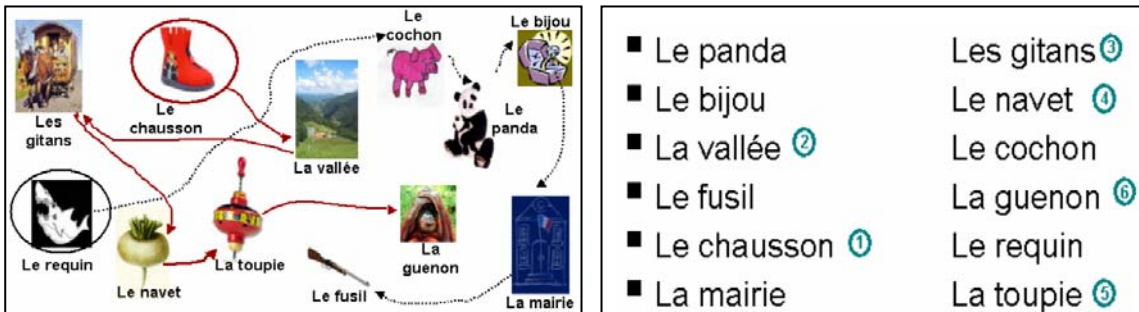


Figure 40. Exemple pour 12 mots cibles de la carte du "meneur" et de la liste d'associations du "suiveur" à la fin du jeu. Le chemin en pointillé est fourni au "meneur" en début de jeu. Le chemin en gras est tracé au cours du jeu par le "meneur". La liste d'associations est fournie au suiveur en début de jeu. Les chiffres (de 1 à 6) sont notés au cours du jeu par le suiveur et indiquent l'ordre des items constituant le nouveau chemin à découvrir.

1.2.1.3. Intérêts et limites

Si deux sujets sur dix ont montré des difficultés de compréhension de la tâche, celles-ci ont totalement disparu après quelques parties d'entraînement. Aucun locuteur n'a montré d'hésitation liée à une charge cognitive trop importante.

Une petite partie du corpus (1,5% des mots cibles) n'a pu être exploitée car les mots correspondants étaient bafouillés ou hésités. Cette proportion de données inexploitable nous semble tout à fait raisonnable. Cela implique toutefois la nécessité d'enregistrer un corpus semi-spontané au moins en double pour éviter les mots-cibles manquants. Cela peut par ailleurs être intéressant pour pallier la plus grande variabilité de la parole spontanée par une plus grande consistance du corpus.

Par ailleurs, nous devons mentionner que l'automatisation de l'analyse d'un corpus spontané se trouve légèrement complexifiée du fait que les phrases et l'ordre des mots ne sont pas toujours les mêmes. Cela nécessite la mise au point d'un système d'indexation pour tenir compte des différents cas, positions, ordres, etc.

Le protocole de ce premier jeu est approprié pour explorer la parole Lombard au niveau global des énoncés ou des mots-cibles, mais ne permet pas de réaliser des mesures rigoureuses de durée ou d'intonation, du fait que la position des mots-cibles dans les énoncés n'est pas contrôlée. De même, l'étude détaillée de certains segments phonétiques particuliers nécessiterait le contrôle du contexte vocalique ou consonantique de ces segments.

1.2.2. Un deuxième jeu plus contrôlé

Ces considérations nous ont amenées à imaginer un deuxième jeu plus contrôlé, dans le but d'explorer en détail l'articulation de plusieurs voyelles et plusieurs consonnes bilabiales dans le bruit, et l'influence de la position des syntagmes dans l'énoncé.

1.2.2.1. Contraintes

Le dispositif expérimental de suivi des mouvements labiaux étant déjà assez lourd pour un locuteur, il n'était pas envisageable de pouvoir enregistrer deux locuteurs simultanément comme dans l'expérience précédente. Par conséquent, l'expérimentatrice a donné la réplique à chaque locutrice.

Les questions que nous souhaitons examiner à travers la constitution de ce corpus (cf. Chapitres 7, 8 et 9) nous ont contraints à imposer une syntaxe Sujet-Verbe-Objet, et à contrôler la composition des syntagmes cibles avec peu de marge de liberté. L'impossibilité de résoudre ces différentes contraintes combinatoires à l'aide de noms du lexique français nous a conduit à construire le corpus à partir de logatomes.

De la même façon que la parole lue diffère de la parole spontanée, l'énonciation de mots n'ayant aucun sens pour le locuteur diffère de la parole naturelle (Stephenson 2004^[331]).

- Cependant, le fait qu'un mot ait ou non du sens pour le locuteur n'est pas lié à son appartenance à un dictionnaire mais davantage à son contexte d'utilisation et au fait qu'un consensus s'est établi avec d'autres interlocuteurs. Ainsi, nous pouvons sans problème apprendre de nouveaux mots argotiques, des noms propres de lieux ou de personnes, communiquer à partir d'un lexique codé, sans que ces mots ne soient répertoriés dans un quelconque dictionnaire. Pour qu'il puisse y avoir communication, l'essentiel est qu'une signification soit associée au mot et que les différents interlocuteurs soient en accord sur cette signification. C'est donc dans cet esprit que nous avons présenté les logatomes aux locuteurs comme des noms propres de rivières africaines.
- De plus, le fait qu'un mot inconnu ne soit pas prononcé de façon naturelle peut être lié à ses caractéristiques phonotactiques inhabituelles pour le locuteur (par exemple enchaîner 4 consonnes d'affilée). Une vision exemplariste des représentations de la parole pourrait également interpréter cette différence de production par le fait que le locuteur n'en a jamais eu l'expérience. Dans notre cas, les logatomes choisis respectent tout à fait les règles phonotactiques du français. Par ailleurs, pour éviter tout éventuel effet d'apprentissage articulaire de ces nouveaux mots, nous avons demandé aux locuteurs de lire une dizaine de fois ces mots à voix haute avant le commencement de l'expérience.

1.2.2.2. Principe

Le jeu conçu pour cette base de données est une fois encore dérivé d'un principe de « map task ». Le locuteur dispose d'une carte sur laquelle sont disposées 17 rivières représentant les 17 logatomes que nous souhaitons lui faire prononcer (cf. Figure 41). Bien que les logatomes soient traités comme des noms propres de rivières, nous avons préféré ne pas les écrire avec une majuscule sur la carte de façon à éviter une focalisation sur la première syllabe.

Le but du jeu consiste à imaginer librement un chemin reliant ces rivières. Ce chemin doit néanmoins satisfaire à plusieurs critères de construction :

- aucune rivière ne doit être laissée de côté.
- le chemin doit être linéaire et par conséquent, une rivière ne peut pas participer au chemin par plus d'une flèche entrante et d'une flèche sortante.

- le chemin doit être fermé, autrement dit toutes les rivières doivent participer au chemin par une flèche entrante et une flèche sortante : il n'y a pas de point de départ ou d'arrivée.
- le chemin doit être construit le plus aléatoirement possible, et non pas de part en part. Pour construire le chemin représenté Figure 41, le locuteur doit éviter de tracer La bala→la lalou, puis ensuite la lalou→la léla, puis encore la léla→la lama mais plutôt tracer par exemple la bala→la lalou, puis ensuite la léla→la lama, puis ensuite la lalou→la léla.

Selon la condition expérimentale, le locuteur dispose d'une carte sur un pupitre à côté de lui, sur laquelle il doit tracer tout seul ce chemin et décrire à voix haute les flèches qu'il trace, ou bien la carte est représentée sur un tableau et il doit donner des consignes à l'expérimentateur pour que celui-ci trace les flèches souhaitées au tableau. Dans tous les cas, la structure des phrases que le locuteur doit employer pour décrire la construction du chemin est imposée : le locuteur doit suivre la forme : «La Rivière_1 longe la Rivière_2» pour décrire une flèche qui part de la Rivière_1 pour aller à la Rivière_2.

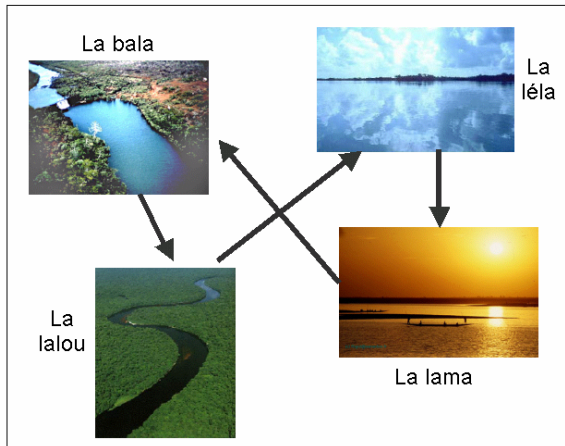


Figure 41. Exemple sur seulement 4 logatomes de la carte dont pourrait disposer le locuteur et sur laquelle il pourrait tracer les flèches en noir pour relier les différentes rivières.

Grâce à ces différentes contraintes, ce jeu permet au locuteur de prononcer tous les logatomes, en sujet/position initiale et en objet/position finale des énoncés. Le discours reste libre et non prévisible, tout en contrôlant le contexte segmental de la première consonne et de la dernière voyelle des mots-cibles. Enfin, il peut être réalisé avec ou sans interaction avec un interlocuteur.

1.2.2.2. Intérêts et limites

Une des 3 locutrices enregistrées (L11) a montré des difficultés de compréhension puis de réalisation de la tâche pendant l'enregistrement. Même en tâche de lecture, cette locutrice avait déjà témoigné dans une étude antérieure de difficultés de concentration dues selon elle au bruit ambiant. Il est donc possible qu'un jeu un peu compliqué ne soit pas adapté à une situation déjà stressante en soi. Cependant, les 2 autres locutrices n'ont pas montré de difficultés particulières.

Nous avons donné cette fois comme consigne aux locutrices de répéter leur phrase si elles bafouillaient. Aussi, nous n'avons recensé aucune perte dans le corpus comme dans l'expérience précédente : tous les logatomes ont été prononcés « correctement » dans toutes les conditions et dans toutes les positions de l'énoncé. Dans certains cas, les locutrices ont inséré des pauses au sein de l'énoncé. Nous ne considérons pas que celles-ci soient liées à une hésitation quelconque.

2. COMPARAISON DE DEUX TACHES DE PAROLE AVEC ET SANS INTERACTION AVEC UN INTERLOCUTEUR

Ayant mis au point des jeux interactifs permettant de reconstituer en laboratoire des situations de communication interactive et relativement spontanée, tout en contrôlant le matériel linguistique, nous pouvons désormais mener une expérience afin de tester si l'interaction avec un interlocuteur influence de façon significative ou non l'adaptation de la parole d'un locuteur en environnement bruyant.

2.1. Matériel et méthode

2.1.1. Constitution d'une deuxième base de données (BD2)

Locutrices

Pour examiner cette question, nous avons enregistré une deuxième base de données (BD2).

Celle-ci est constituée de la parole semi-spontanée en Français de 3 locutrices (L2, L6 et L11). Les locutrices L2 et L6 avaient déjà participé à la première base de données et L11 à l'étude pilote préalable à la constitution de cette deuxième base de données (cf. Annexe Pil2). Les 3 locutrices étaient de langue maternelle française.

Conditions de bruit

Nous avons enregistré ces locutrices en condition silencieuse de référence puis dans les deux types de bruit déjà utilisés pour la première base de données (BD1): un bruit blanc (bb) et un bruit de « cocktail-party » (cktl), tirés de la base de données BD_Bruit (Zeiliger et al. 1994^[382]). Ces deux environnements bruyants ont été amplifiés par un amplificateur (JVC Ax40) puis diffusés à un niveau de 85dBc par le biais de deux haut-parleurs (A2t, de puissance maximale 200 Watts) placés à une hauteur de 1m, c'est à dire au niveau des oreilles de la locutrice en position assise, éloignés de 1.4 m de la locutrice et espacés entre eux de 1.5m (cf. Figure 42). Le niveau d'intensité des bruits a été calibré à l'aide d'un sonomètre au niveau des oreilles de la locutrice.

Tâche de parole et matériel linguistique

Le deuxième jeu, présenté en détail au paragraphe 1.2.2 a été utilisé de façon différente pour enregistrer deux sessions de ces 3 conditions : la première fois, les locutrices devaient effectuer ce jeu de façon isolée, et décrire à voix haute en Français leurs actions, nécessitant l'utilisation contrôlée de 17 logatomes-cibles de forme CVCV ([lale], [lali], [laly], [lalu], [lalā], [lela], [lila], [lyla], [lula], [lāla],[pala], [bala], [mala] et [lapa], [laba], [lama], choisis pour examiner d'autres hypothèses dans le chapitre 6). La deuxième fois, les locutrices effectuaient ce jeu en collaboration avec l'expérimentatrice, située à 2m50 en face d'elle (cf. Figure 42), en lui donnant des consignes pour qu'elle exécute ces actions au tableau. Dans les deux cas, le jeu était identique, utilisait les mêmes logatomes cibles et la même phrase porteuse (celle-ci a été choisie pour explorer d'autres hypothèses dans les chapitres 6 à 8). Seule variait l'interaction avec l'expérimentatrice entre les deux sessions. Les locutrices n'étaient pas informées à l'avance du déroulement expérimental. Lors de la première session où elles devaient jouer seules, elles ne s'attendaient donc pas à devoir ensuite interagir avec quelqu'un dans une deuxième session.

Signaux enregistrés

Afin de réaliser des analyses acoustiques et phonétiques de la parole des locutrices, nous avons enregistré leur signal audio de parole à l'aide d'un microphone à électret (AKG C1000S), situé à 50cm de la bouche de la locutrice. Pour calibrer le niveau d'intensité acquis au microphone, nous avons enregistré le locuteur en début d'enregistrement sur un [a] tenu stable et mesuré cette intensité vocale de référence au sonomètre au niveau des lèvres du locuteur. Ce signal a ensuite été numérisé sur 16 bits à 44.1 kHz sur la première voie d'une carte son externe (Edirol M-100FX)

Afin d'étudier la vibration des cordes vocales et les mouvements verticaux du larynx, nous avons acquis à l'aide d'un électroglottographe à deux paires d'électrodes (Glottal Enterprise EG2) :

- le signal d'accolement des cordes vocales, que nous désignerons par la suite par l'abréviation EGG. Il a été directement numérisé sur 16 bits à 44.1 kHz sur la deuxième voie de la carte son externe (Edirol M-100FX).
- le signal de déplacement vertical du larynx (ou LT pour « Larynx Tracking »). Ce signal modulé a été numérisé sur 16 bits à 44.1kHz à l'aide d'un enregistreur numérique (Korg D1200).

Afin d'explorer les mouvements articulatoires labiaux, nous avons effectué des enregistrements vidéo de face et de profil du visage des locutrices et extrait leurs mouvements articulatoires au niveau des lèvres grâce à un système de « Lip Tracking » explicité en détail en Annexe Met1. Il nous a semblé plus intéressant de normaliser l'amplitude des mouvements articulatoires par rapport au maximum « physiologique » des locutrices plutôt que de présenter des résultats dans l'absolu en cm. C'est pourquoi à la fin du protocole nous avons enregistré les mouvements articulatoires maximaux que pouvaient effectuer les locutrices en ouverture de la mâchoire, en étirement des lèvres et en protrusion, hors contexte de parole.

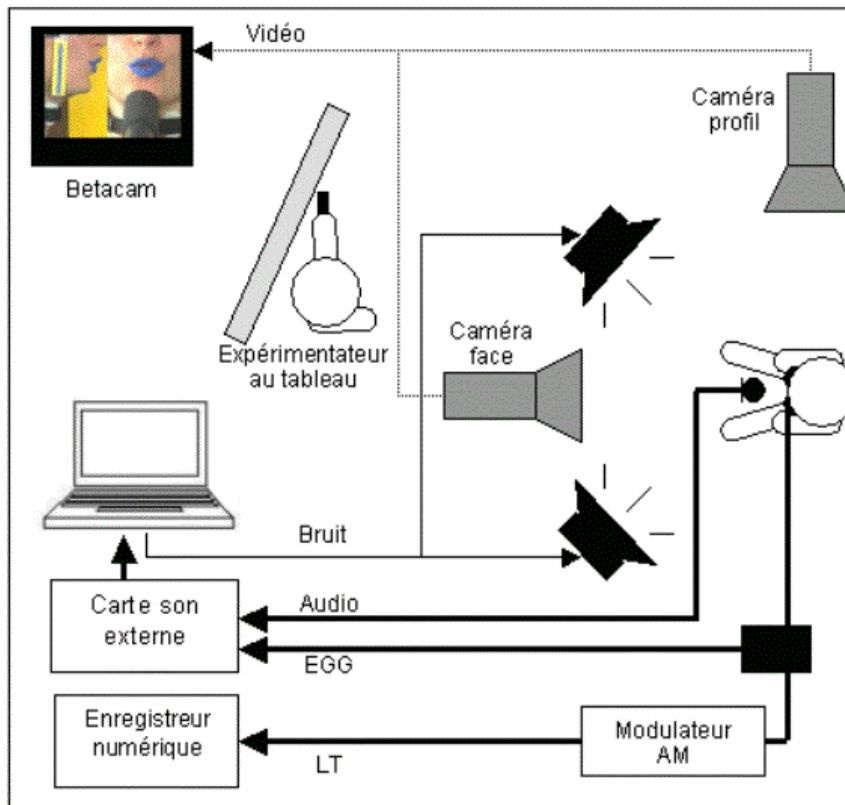


Figure 42 : Schéma du dispositif expérimental pour la constitution de la deuxième base de données.

2.1.2. Analyse et traitement statistique des données

Pour examiner l'influence du dialogue avec un interlocuteur sur l'adaptation de la parole des locutrices dans le bruit, nous avons considéré plusieurs descripteurs de la voix et de la parole. Certains ont été mesurés pour les 3 locutrices sur les syllabes des 17 syntagmes de chaque condition, en position initiale et finale de l'énoncé: l'intensité moyenne, la F0 moyenne, la durée syllabique, l'amplitude maximale des gestes articulatoires labiaux d'ouverture et de protrusion pour les voyelles [a], [i] et [u]. D'autres paramètres ont été mesurés sur les voyelles de ces syntagmes : c'est le cas de la différence d'énergie spectrale [2-4kHz] et [0-1kHz], du quotient ouvert et de la fréquence centrale du premier formant vocalique. Enfin, la présence de pauses dans l'énoncé a été mesurée sur l'ensemble de chaque condition.

De la même façon que dans le chapitre précédent, nous proposons de simplifier la présentation de ces résultats en adoptant plusieurs profils pour modéliser la différence d'évolution d'un paramètre du silence au bruit en tâche de dialogue ou de jeu isolé (cf. Figure 43).

Effet du bruit significatif avec et sans dialogue			Effet du bruit non significatif sans dialogue	
Identique	Décalage	Amplification + éventuel décalage	Aucun effet	Amplification + éventuel décalage

Figure 43. Représentation de 6 types de différences d'évolution d'un paramètre avec le bruit, avec ou sans dialogue du locuteur avec l'expérimentatrice. La courbe rose (du dessus) représente l'évolution avec dialogue et la courbe bleue (du dessous), sans.

Nous avons tout d'abord distingué les paramètres évoluant significativement avec le bruit aussi bien en situation isolée qu'en situation de dialogue, des paramètres évoluant de façon significative dans le bruit uniquement en situation de dialogue. Pour cela, nous avons testé pour chaque paramètre la significativité de la différence entre sa valeur dans le bruit et sa valeur dans le silence (Test Anova à un facteur : BRUIT), et cela pour les deux sessions avec ou sans dialogue avec l'expérimentatrice. Pour chacune de ces 2 catégories, nous avons distingué plusieurs profils d'évolutions d'un paramètre.

- (**Identique ou Aucun effet**) : Le dialogue n'induit aucune différence que ce soit dans le silence ou dans le bruit, et par conséquent pas non plus de différence au niveau de l'évolution du silence au bruit.
- (**Décalage**) : L'évolution du paramètre est semblable avec et sans dialogue, à un décalage vertical D près.

Ces deux profils se caractérisent par une différence non significative entre l'évolution du silence au bruit pour la session sans dialogue et l'évolution pour la session avec dialogue (Test Anova à un facteur : DIALOGUE, sur l'évolution $\Delta P = P_{\text{bruit}} - P_{\text{sil}}$ de chaque paramètre du silence au bruit). Ils se distinguent par un effet significatif du dialogue (Décalage) ou non significatif (Identique) dans le silence ET dans le bruit (Tests Anova à un facteur : DIALOGUE, sur les valeurs de chaque paramètre dans le silence, dans le bruit).

- (**Amplification**) : L'évolution du paramètre est semblable avec et sans dialogue, à un facteur d'amplification « a » près. Ce profil se caractérise par un effet non significatif du dialogue dans le silence par contre significatif dans le bruit (Tests Anova à un facteur : DIALOGUE, sur les valeurs de chaque paramètre dans le silence, dans le bruit) et par un effet « a » significatif du dialogue sur l'évolution du paramètre du silence au bruit (Test Anova à un facteur : DIALOGUE, sur l'évolution $\Delta P = P_{\text{bruit}} - P_{\text{sil}}$ de chaque paramètre du silence au bruit).

- (**Décalage + Amplification**) : Combinaison des deux profils précédents. Ce profil se caractérise par un effet « *D* » significatif du dialogue dans le silence, et par un effet du dialogue également significatif dans le bruit (Tests Anova à un facteur : DIALOGUE, sur les valeurs de chaque paramètre dans le silence, dans le bruit), ainsi que par un effet significatif du dialogue sur l'évolution du paramètre du silence au bruit (Test Anova à un facteur : DIALOGUE, sur l'évolution $\Delta P = P_{\text{bruit}} - P_{\text{sil}}$ de chaque paramètre du silence au bruit).

2.2. Résultats

Les Figure 44 à Figure 48 récapitulent les différences observées pour plusieurs descripteurs de la parole lorsque les locutrices jouent seules ou en situation de dialogue avec l'expérimentatrice. Ces différences sont catégorisées relativement aux différents profils présentés ci-dessus.

Tout d'abord on observe quelques rares paramètres sur lesquels ni le dialogue ni l'exposition au bruit n'ont d'influence (Aucun effet). C'est le cas de la protrusion pour les voyelles [i] et [u] (cf. Figure 44). Dans le cas des voyelles [u], on observe cependant une tendance de l'effet de dialogue à être plus important dans le bruit et l'effet Lombard à être plus important en situation de dialogue, bien que cela ne soit pas significatif.

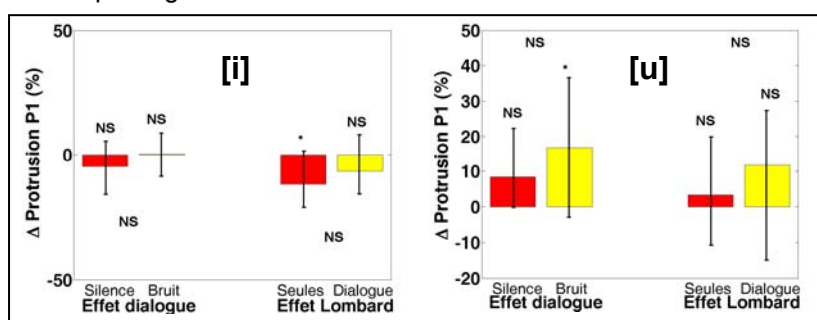
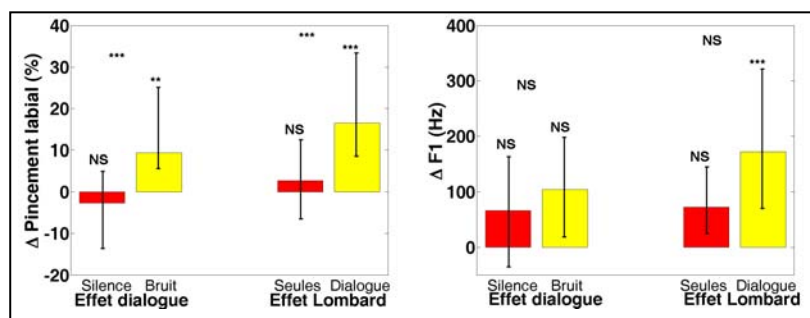


Figure 44. Effet du dialogue dans le silence et dans le bruit et de l'exposition au bruit en situation interactive ou non sur la protrusion des voyelles [i] et [u].

On observe ensuite plusieurs paramètres pour lesquels l'exposition au bruit n'entraîne pas de modification significative de la parole en situation non interactive, mais provoque au contraire une adaptation significative en situation de dialogue. C'est le cas de l'amplitude du pincement labial sur les consonnes bilabiales [m], [p], [b] ainsi que de la fréquence du premier formant vocalique en général pour l'ensemble des voyelles (cf. Figure 45).

Figure 45. Effet du dialogue dans le silence et dans le bruit et de l'exposition au bruit en situation interactive ou non sur le pincement labial et la fréquence du 1^{er} formant vocalique.



Ensuite, il existe plusieurs paramètres évoluant dans le bruit de façon semblable, que ce soit avec ou sans dialogue avec un interlocuteur. Autrement dit, le dialogue n'a aucun effet sur ces paramètres, que ce soit dans le silence, dans le bruit, ou sur l'évolution du paramètre du silence au bruit (Effet Lombard). C'est le cas de l'aire intéro-labiale sur les voyelles [i] (cf. Figure 46). On observe cependant une tendance de l'effet de dialogue à être plus important dans le bruit et l'effet Lombard à être plus important en situation de dialogue, bien que cela ne soit pas significatif.

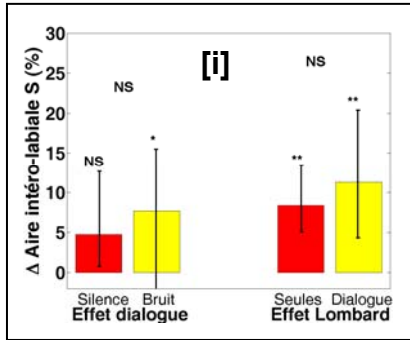


Figure 46. Effet du dialogue dans le silence et dans le bruit et de l'exposition au bruit en situation interactive ou non sur l'aire intéro-labiale des voyelles [i].

Nous n'avons pas recensé de cas où un descripteur de la parole augmente dans le bruit de façon comparable, avec ou sans dialogue, à un décalage vertical près, correspondant à un effet constant du dialogue, indépendant de l'exposition au bruit.

En revanche, la grande majorité des paramètres, en particulier ceux classiquement utilisés pour décrire l'effet Lombard, évoluent dans le bruit davantage en situation de dialogue que sans interaction. La plupart montrent un effet non significatif du dialogue dans le silence, le devenant dans le bruit : c'est le cas de la durée syllabique, de l'intensité vocale, de la fréquence fondamentale moyenne, du quotient ouvert, de la différence d'énergie entre [2-4kHz] et [0-1kHz] ainsi que de l'aire intéro-labiale et de la protrusion des voyelles [a] (cf. Figure 47).

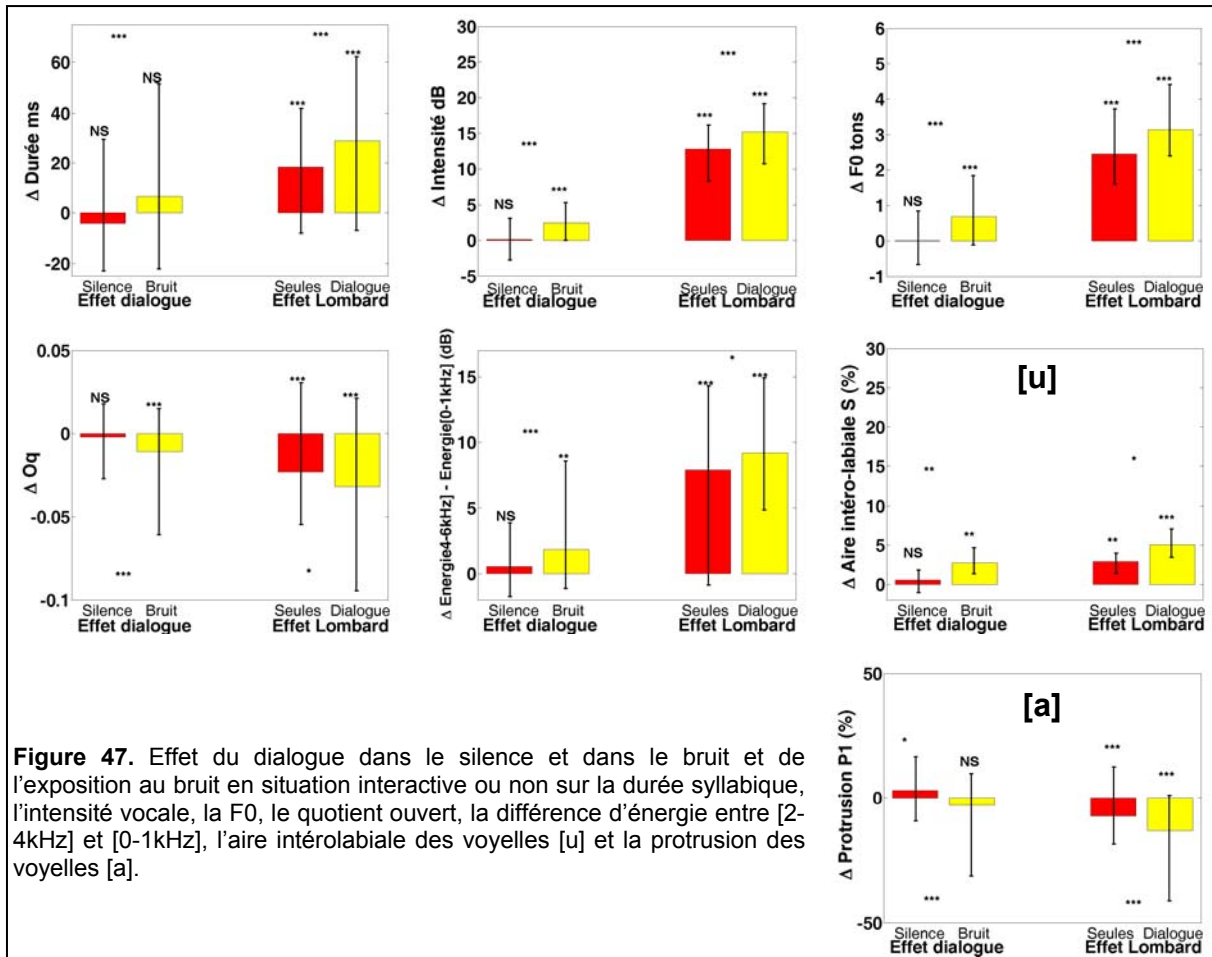


Figure 47. Effet du dialogue dans le silence et dans le bruit et de l'exposition au bruit en situation interactive ou non sur la durée syllabique, l'intensité vocale, la F0, le quotient ouvert, la différence d'énergie entre [2-4kHz] et [0-1kHz], l'aire intéro-labiale des voyelles [u] et la protrusion des voyelles [a].

Dans certains cas, il existe déjà un effet du dialogue significatif dans le silence, mais moindre par rapport à dans le bruit. C'est le cas de l'aire intéro-labiale pour les voyelles [a] ainsi que de la fréquence d'insertion de pauses dans l'énoncé (cf. Figure 48)

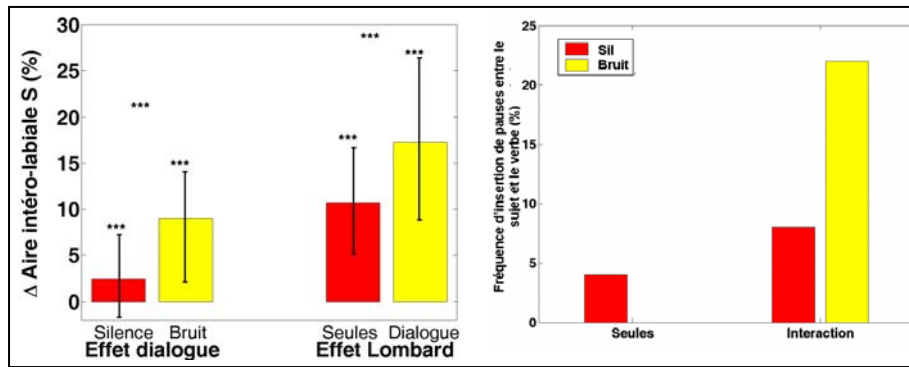


Figure 48. Effet du dialogue dans le silence et dans le bruit et de l'exposition au bruit en situation interactive ou non sur l'aire intérolabiale des voyelles [a] et sur la fréquence d'insertion de pauses entre le sujet et le verbe de l'énoncé.

2.3. Conclusion sur la comparaison des deux tâches de parole

Les résultats de cette étude montrent donc tout d'abord qu'il existe indéniablement un effet significatif de l'exposition du locuteur au bruit, même sans dialogue avec un interlocuteur, excepté pour le pincement labial et pour la fréquence du 1^{er} formant (qui montre quand même une augmentation notable, mais non significative). Cela n'est pas un résultat surprenant. Il ne fait que vérifier les observations de la grande majorité des études antérieures sur l'effet Lombard ayant exploré ce phénomène en tâche de lecture, ou du moins sans dialogue avec un interlocuteur.

L'évolution dans le bruit, en dehors de toute interaction avec un interlocuteur, de paramètres acoustiques et articulatoires pouvant être liés à l'augmentation de l'intensité vocale (fréquence fondamentale, ouverture de la bouche, fréquence du premier formant, quotient ouvert, enrichissement du spectre en hautes fréquences), peut être en premier lieu interprétée comme la conséquence d'une régulation réflexe de l'intensité de la voix du fait de l'atténuation du retour auditif par le bruit ambiant, tel que le suggère Lombard 1911^[224], Fairbanks 1954^[86] ou Tonkinson 1994^[353].

Il est également possible que même en l'absence d'interlocuteur physiquement présent, et même sans aucun enjeu à devoir être intelligible, les locuteurs continuent à reproduire des schémas de communication en s'adressant à un interlocuteur virtuel et en cherchant à compenser la dégradation de leur intelligibilité par le bruit. On remarque bien par exemple que des personnes parlant au téléphone continuent à accompagner leur parole de gestes des mains ou de mimiques faciales bien que celles-ci ne puissent pas être perçues par leur interlocuteur. De la même façon, même si le locuteur est censé jouer tout seul dans la première session, il se sait écouté par l'expérimentateur qui vérifie le déroulement de l'expérience. Ce seul auditoire peut le motiver à maintenir un certain degré d'intelligibilité. Aussi, notre paradigme de comparaison de la parole avec ou sans dialogue avec un interlocuteur ne rendrait pas compte de la recherche ou non d'intelligibilité par le locuteur, mais plutôt de différents degrés de recherche d'intelligibilité.

Enfin, les résultats de cette étude montrent que de nombreux paramètres sont significativement plus renforcés dans le bruit en situation de dialogue : à la fois les paramètres acoustiques et articulatoires pouvant être liés à l'augmentation de l'intensité et de la fréquence fondamentale, mais également d'autres paramètres linguistiques tels que la fréquence d'insertion de pauses dans l'énoncé. Il ne semble donc pas que le locuteur cherche à augmenter son intelligibilité dans le bruit uniquement en parlant plus fort, ni uniquement de façon réflexe. Les rares études antérieures à avoir exploré l'effet Lombard en parole spontanée ou en interaction avec un partenaire, ont déjà observé une augmentation de l'intensité vocale du silence au bruit plus importante par rapport aux autres études concernant la parole lue. Ainsi, le coefficient de régression de la droite représentant l'évolution de l'intensité vocale moyenne en fonction de l'intensité du bruit ambiant a été estimé à 0.15 par Egan 1972^[80] ou à 0.12 par Lane et al. 1970^[194] en parole lue. Tandis que Korn 1954^[185], Gardner, ou Kryter 1946^[188] ont mesuré une pente respectivement de 0.38, 0.30 et 0.33 en parole spontanée. De même, nous observons ici une pente de 0.26 en parole spontanée mais sans interaction avec un interlocuteur, et de 0.33 avec interaction. L'avantage dans notre étude est d'avoir enregistré les mêmes locuteurs avec le même matériel linguistique, en faisant uniquement varier l'interaction ou non de la tâche. Nous pouvons donc conclure avec davantage de certitude que l'augmentation de la pente est due à l'effet de dialogue. Par ailleurs, nous observons une pente de 0.22 pour le corpus articulatoire pilote (cf. Annexe Pil2) où la tâche consistait à lire une liste de phrases, et de 0.36 pour la première base de données (BD1, cf. chapitre 3) où la tâche correspond au premier jeu interactif semi-spontané, exposé au paragraphe 1.2.1, et beaucoup plus libre que ce deuxième jeu.

Les résultats de cette étude montrent donc que l'effet Lombard est en partie motivé par la recherche d'intelligibilité et que l'interaction avec un interlocuteur n'influence pas cette adaptation pour tous les paramètres, mais en tout cas pour tous les paramètres classiquement utilisés pour décrire l'effet Lombard (Intensité, F0, durée, mouvements articulatoires, ...). Afin d'appréhender ce phénomène dans son ensemble, il semble donc nécessaire de prendre en compte l'interaction avec un interlocuteur dans la mise au point de protocoles expérimentaux.

3. CONCLUSION DU CHAPITRE 4

Le but de ce chapitre était avant tout méthodologique. Ainsi, nous avons proposé de nouveaux jeux interactifs permettant d'enregistrer des locuteurs dans le bruit en situation de discours semi-spontané et non entièrement prévisible, en alternative aux tâches de lecture par un locuteur isolé classiquement utilisées pour étudier l'effet Lombard.

Nous avons montré plusieurs intérêts directs et indirects de telles tâches pour l'étude des stratégies d'adaptation en environnement bruyant, et avons validé leur utilisation, en montrant que l'interaction avec un interlocuteur avait un effet significatif sur l'adaptation du locuteur pour certains paramètres de la parole, en particulier pour les paramètres classiquement considérés de la parole Lombard (Intensité moyenne, F0 moyenne, etc.).

Enfin, l'examen de l'influence du dialogue avec un interlocuteur sur l'adaptation de la parole dans le bruit a motivé la constitution d'une deuxième base de donnée (BD2). Celle-ci a également été constituée dans le but d'explorer en détail certaines hypothèses quant à l'existence de stratégies de renforcement des indices acoustiques et visuels de reconnaissance des voyelles (Chapitre 7), de renforcement des indices de structuration des énoncés (Chapitre 8), ou encore de renforcement des mots présentant un poids sémantique plus important dans l'énoncé (Chapitre 9). Nous explorerons donc à nouveau et plus en détail la 2ème session de cette base de données, où les locutrices interagissent avec l'expérimentatrice, pour caractériser les différences interindividuelles de stratégies d'adaptation à un environnement bruyant (Chapitres 5 à 10).

Partie III

Caractérisation des stratégies d'adaptation de la parole dans le bruit

- Présentation de la Partie III -

Dans cette partie, nous chercherons à caractériser les différences interindividuelles d'adaptation de la parole en environnement bruyant en nous appuyant sur les approches méthodologiques discutées précédemment.

Nous commencerons dans le Chapitre 6 par examiner les tendances principales de cette adaptation au niveau acoustique, articulatoire et glottique, afin d'avoir une vue d'ensemble du phénomène et de comparer nos résultats obtenus avec de nouvelles méthodologies expérimentales à ceux de la littérature. Par ailleurs, ce chapitre sera l'occasion de commencer à dégager, parmi l'ensemble des modifications observées, quelles sont celles qui montrent une variabilité interindividuelle notable et quelles sont celles qui dépendent du type de bruit.

Des chapitres 7 à 8, nous ne chercherons pas à caractériser de façon exhaustive et paramétrique toutes les différences existant entre la parole produite dans le silence et la parole produite dans le bruit. En effet, les modifications acoustiques, articulatoires et glottiques ne nous intéressent pas tant pour elle-mêmes que parce qu'elles représentent des indicateurs de stratégies adoptées par les locuteurs pour s'adapter à un environnement bruyant. Aussi, puisque nous considérons dans ce travail que l'adaptation des locuteurs en environnement bruyant est principalement motivée par la recherche d'intelligibilité vis à vis d'un interlocuteur, nous envisagerons tour à tour plusieurs hypothèses de stratégies de communication potentiellement adoptées dans le bruit et examinerons en quoi certaines évolutions conjointes et non plus isolées, de certaines paramètres acoustiques et articulatoires peuvent argumenter en faveur ou non de leur utilisation par les locuteurs.

Nous explorerons en particulier s'il existe des stratégies purement acoustiques adoptées par les locuteurs, visant à renforcer de façon globale certaines caractéristiques acoustiques de leur voix afin d'émerger par rapport au bruit ambiant (Chapitre 6). Nous explorerons également l'existence possible de stratégies de renforcement dans le bruit des indices acoustiques ou visuels de reconnaissance des phonèmes (Chapitre 7). Nous nous intéresserons ensuite à de potentielles stratégies de renforcement dans le bruit des indices prosodiques de structuration de l'énoncé (Chapitre 8). Enfin, nous déterminerons s'il existe des stratégies d'adaptation de la parole dans le bruit liées au renforcement plus important dans le bruit des mots de l'énoncé apportant une information plus importante (Chapitre 9).

Contrairement à la partie méthodologique précédente, nous ne chercherons plus ici de comportement moyen ou de stratégie commune à tous les locuteurs mais nous serons au contraire intéressés par leurs différences. En particulier, nous dresserons dans le Chapitre 10 un bilan des différences interindividuelles d'adoption des précédentes stratégies, en fonction également du type de bruit et du niveau de bruit, afin d'examiner si certains locuteurs s'adaptent au bruit toujours avec une même et unique stratégie, ou en adoptant de façon sélective et combinée plusieurs stratégies. De même, nous étudierons si les locuteurs adoptent les mêmes stratégies avec une magnitude croissante en fonction du niveau de bruit ou si leur adaptation est davantage catégorielle et consiste en une réorganisation de l'action en fonction de niveaux de perturbation du bruit ambiant.

Chapitre 5 :

Principales observations

- 1. QUELS SONT LES ASPECTS DE LA PAROLE SIGNIFICATIVEMENT MODIFIES DANS LE BRUIT ?**
- 2. QUELLES SONT LES TENDANCES PARTAGEES ET LES DIFFERENCES INDIVIDUELLES ?**
- 3. QUELS SONT LES PARAMETRES DONT L'EVOLUTION DANS LE BRUIT EST DEPENDANTE DU TYPE DE BRUIT ?**
- 4. CONCLUSION DU CHAPITRE 5**

Avant de rentrer dans le détail des différentes hypothèses de stratégies pouvant être adoptées pour communiquer dans le bruit, nous allons commencer par poser à plat les grandes tendances générales observées dans nos bases de données en ce qui concerne l'adaptation de la parole en environnement bruyant.

Nous présenterons les aspects de la parole significativement modifiés par l'exposition des locuteurs à un environnement bruyant et comparerons ces résultats, obtenus avec de nouvelles méthodologies (restitution du bruit par le biais de haut-parleurs, situations de communication interactives et semi-spontanées), à ceux des études antérieures.

Nous examinerons ensuite parmi ces modifications quelles sont celles étant communément produites par tous les locuteurs et au contraire quelles sont celles présentant le plus de différences entre les locuteurs. N'oublions pas que notre objectif principal est de caractériser les différences interindividuelles d'adaptation de la parole dans le bruit, afin de mieux comprendre le phénomène du forçage vocal. Aussi, nous chercherons à voir s'il est possible de distinguer des groupes de locuteurs aux comportements semblables, notamment par rapport à leur genre.

Enfin, nous observerons parmi ces modifications quelles sont celles qui montrent une différence entre les deux types de bruit et au contraire quelles sont celles qui sont uniquement liées à la présence de bruit.

Pour cela, nous avons repris la première base de données acoustiques et électroglottographiques (BD1) présentée au paragraphe 3.1.1 du chapitre 3, constituée par la parole semi-spontanée de 5 locuteurs (L3, L4, L5, L7 et L8) et 5 locutrices (L1, L2, L6, L9, L10) produite au cours d'un jeu interactif nécessitant l'utilisation de 16 mots-cibles de forme CVCV (bijou, chausson, cochon, dauphin, fusil, guenon, gitans, lagon, mairie, navet, panda, requin, sommet, toupie, vallée, zébu). En dehors de ces mots-cibles, le discours des locuteurs était libre et correspondait à un temps de parole d'environ 1mn30 pour chaque condition. Nous avons ici exploré la condition silencieuse de la 3^{ème} session de cette base de données, où les locuteurs s'exprimaient en situation calme (mesurée au sonomètre à un niveau de 40dBC) et sans porter de casque, ainsi que les deux conditions de bruit blanc et de bruit cocktail diffusés à 86dBC par le biais de haut-parleurs.

Nous avons également repris la deuxième base de données acoustiques, articulatoires et électroglottographiques (BD2) présentée au paragraphe 2.1.1 du chapitre 4, consistant en la parole semi-spontanée de 3 locutrices (L2, L6 et L11) produite au cours d'un jeu nécessitant l'utilisation de 17 mots-cibles de forme CVCV ([lale], [lali], [laly], [lalu], [lalā], [lela], [lila], [lyla], [lula], [lāla], [pala], [bala], [mala] et [lapa], [laba], [lama]) dans le contexte d'une phrase porteuse « la __ longue la ____ ». Nous avons ici exploré la condition silencieuse et les deux conditions de bruit blanc et de bruit cocktail de la 2^{ème} session de cette base de données, où les locutrices s'exprimaient en situation de dialogue interactif avec l'expérimentatrice. Pour les deux conditions de bruit, celui-ci était diffusé à 85dBC par le biais de haut-parleurs.

1. QUELS SONT LES ASPECTS DE LA PAROLE SIGNIFICATIVEMENT MODIFIÉS DANS LE BRUIT ?

Pour commencer, nous allons examiner parmi différents paramètres acoustiques, articulatoires et glottiques permettant de décrire le comportement vocal, quels sont ceux qui varient de façon notable dans le bruit, en examinant l'amplitude de leur variation du silence au bruit pour les différents locuteurs et pour les différents types de bruit.

Pour caractériser ces évolutions, nous avons extrait chaque paramètre des signaux acoustiques, articulatoires ou électroglottographiques :

- sur chaque syllabe des 16 mots-cibles pour la BD1, ce qui représente donc 32 mesures par condition et par locuteur
- sur la syllabe du déterminant [la] et sur chaque syllabe des 17 logatomes-cibles, en position initiale et finale de la phrase porteuse, pour la BD2. Cela représente donc 102 mesures par condition et par locutrice.

A partir de ces données, nous avons effectué pour chaque base de données, pour chaque locuteur et pour chaque type de bruit, un test Anova à un facteur (BRUIT), afin d'évaluer la significativité de la

différence entre la valeur des paramètres dans le bruit et leur valeur dans le silence. Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res5. Nous en présenterons ici seulement une synthèse.

1.1. Au niveau acoustique

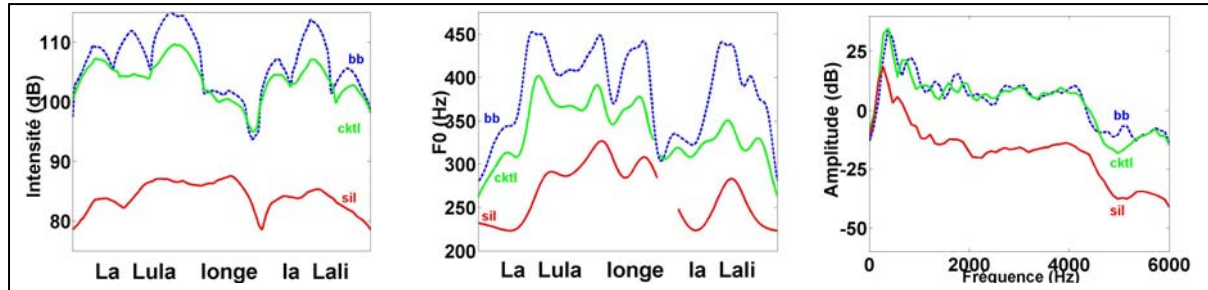


Figure 49. Représentation de l'intensité vocale, de la fréquence fondamentale et du spectre de la voix au cours d'un même énoncé de la BD2 produit par une locutrice en condition silencieuse (sil), dans un bruit blanc de 85dB (bb) et dans un bruit cocktail de 85dB (cktl). La durée des énoncés a été normalisée sur cette figure de façon à faciliter la comparaison visuelle des signaux.

La première observation principale et très classique concerne **l'augmentation de l'intensité vocale** dans le bruit, de même que **l'élévation de la fréquence fondamentale** (cf. Figure 49).

Pour les deux bases de données, l'augmentation de l'intensité vocale et de la fréquence fondamentale est très significative pour tous les locuteurs et locutrices et ce, quel que soit le type de bruit considéré. L'augmentation moyenne de l'intensité vocale est très semblable entre les deux bases de données : de 16.4dB (ou 20% d'augmentation) pour la BD1 et de 16.8dB (ou 18% d'augmentation) pour la BD2 (cf. Tableau 11).

L'augmentation de la fréquence fondamentale est en revanche plus importante dans la 1^{ère} base de données que dans la 2^{ème} : de 4.7 tons dans la BD1 (soit 77% d'augmentation pour la F0 exprimée en Hz) contre 3.7 tons dans la BD2 (soit 54% d'augmentation).

Dans tous les cas, l'effet du bruit sur l'augmentation de l'intensité vocale ou de la fréquence fondamentale est bien supérieure à la variabilité intra-individuelle moyenne pour ces mêmes paramètres (correspondant à la moyenne des écart-types observés chez chaque locuteur).

		BD1	BD2
Δ Intensité	Diff	16.4 dB	16.8dB
	Taux	20%	18%
Δ F0	Diff	4.7 tons	3.7 tons
	Taux	77%	54%

Tableau 11. Tendence moyenne d'évolution du silence au bruit de l'intensité vocale et de la fréquence fondamentale (F0) pour les locuteurs et locutrices de la 1^{ère} et de la 2^{ème} base de données (BD1 et BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

Ces observations corroborent celles de toutes les études antérieures sur l'effet Lombard (Lombard 1911 ^[224] ; Van Summers et al. 1988 ^[361] ; Castellanos et al. 1996 ^[46] ; Junqua 1993 ^[171], etc.). Cependant, les augmentations de l'intensité vocale et de la F0 observées dans notre étude sont supérieures à celles des études antérieures (par exemple de 12 à 15dB dans 82-85dB de bruit chez Ternström et al. 2002 ^[346]). Par ailleurs, la pente de régression linéaire de l'intensité vocale en fonction de l'intensité du bruit ambiant est en moyenne de 0.33 dans notre étude, ce qui est supérieur aux valeurs des études effectuées en parole lue (Egan 1972 ^[80] ; Lane et al. 1970 ^[194]) et conforme à celles des études effectuées en parole spontanée (Kryter 1946 ^[188] ; Korn 1954 ^[185]).

Le déplacement du spectre de la voix vers les médiums est une deuxième observation très classique concernant la parole criée ou la parole produite en environnement bruyant (cf. Figure 49). Ainsi, nous obtenons des résultats semblables à ceux de Junqua 1993 ^[171], Castellanos et al. 1996 ^[46], Pisoni et al. 1985 ^[272] ou Stanton 1988 ^[327] rapportant une élévation du centre de gravité spectral, de même que Rostolland 1982 ^[288] notant une augmentation de l'énergie spectrale entre 4 et 5kHz, et parfois au-delà. Ce déplacement ne correspond pas à l'augmentation de l'énergie acoustique uniquement dans les médiums : L'augmentation de l'intensité vocale dans le bruit correspond à une augmentation significative de l'énergie acoustique dans toutes les bandes de fréquences, quel que soit le type de bruit et les locuteurs considérés (cf. Tableau 12). Néanmoins, cette augmentation de l'énergie acoustique est particulièrement importante dans les zones [1-2kHz] et [2-4kHz], davantage

que dans les autres bandes, ce qui se traduit par une réorganisation de la répartition d'énergie acoustique en fréquences : Ainsi, l'énergie acoustique augmente en moyenne dans la BD1 de 27.7dB entre 2 et 4kHz contre 16.2dB entre 0 et 1kHz et dans la BD2 de 25.1dB entre 2 et 4kHz contre 15.3dB entre 0 et 1kHz. Cette évolution est donc très semblable entre les deux bases de données.

Enfin, quelle que soit la bande de fréquences considérée, l'augmentation de l'énergie acoustique induite par l'exposition du locuteur à un environnement bruyant de 85 et 86dBC est bien supérieure à la variabilité intra-individuelle.

	BD1	BD2
Δ Energie entre 0 et 1 kHz	16.2 dB	15.3 dB
Δ Energie entre 1 et 2 kHz	27.3 dB	25.2 dB
Δ Energie entre 2 et 4 kHz	27.7 dB	25.1 dB
Δ Energie entre 4 et 6 kHz	25.7 dB	23.0 dB

Tableau 12. Tendence moyenne d'évolution de l'énergie acoustique dans les zones [0-1kHz], [1-2kHz], [2-4kHz] et [4-6kHz] pour les locuteurs et locutrices de la 1^{ère} et de la 2^{ème} base de données (BD1 et BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

1.2. Au niveau articulatoire

1.2.1. Durée syllabique

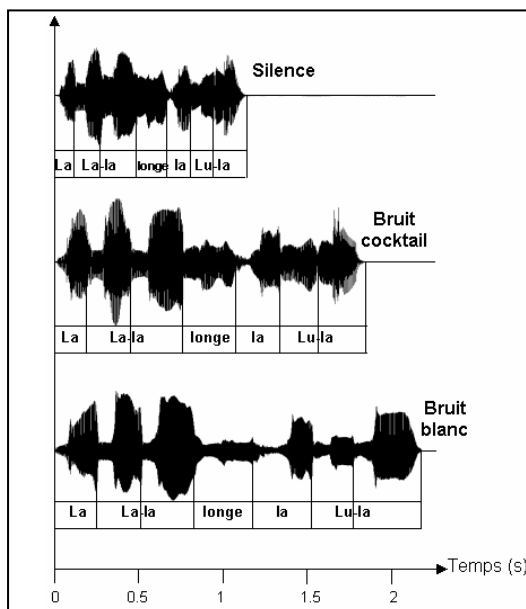


Figure 50. Représentation des signaux acoustiques d'un même énoncé de la BD2 produit par une locutrice en condition silencieuse (sil), dans un bruit blanc de 85dBC (bb) et dans un bruit cocktail de 85dBC (cktl).

	BD1	BD2
Δ Durée	26 ms	39 ms
	17%	26%

Tableau 13. Tendence moyenne d'évolution de la durée syllabique pour les locuteurs et locutrices de la 1^{ère} et de la 2^{ème} base de données (BD1 et BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

La littérature mentionne très fréquemment un **ralentissement du débit de parole** en environnement bruyant (cf. Figure 50). C'est le cas de Davis et al. 2006^[62], Hanley et al. 1949^[133], Junqua 1993^[171], Kadirî 1998^[176], Kim 2005^[180] ou Pisoni et al. 1985^[272].

Nous avons également observé cette tendance dans la première base de données, mais pas de façon majoritaire : une augmentation significative de la durée syllabique a été constatée dans 8 cas sur 20 (les 20 cas correspondants aux 10 locuteurs dans les 2 types de bruit). Les 12 cas sur 20 restants montrent plutôt une évolution non significative de la durée syllabique dans le bruit, et deux d'entre eux montrent même une tendance à l'accélération du débit de parole dans le bruit cocktail. Pour la deuxième base de données, toutes les locutrices augmentent significativement la durée syllabique dans le bruit, et cela quel que soit le type de bruit.

Il est par conséquent logique d'observer un allongement syllabique moyen dans le bruit plus faible dans la BD1 (26ms ou de 17% d'augmentation) que dans la BD2 (39ms ou 26% d'augmentation) (cf. Tableau 13). De même, on observe que la variabilité intra-individuelle moyenne est assez importante vis à vis de l'évolution de la durée syllabique dans le bruit, proche de l'effet lui-même, voire supérieur dans le cas de la 1^{ère} base de données. Nous avons vu que le ralentissement du débit de parole n'est pas une stratégie commune à tous les locuteurs, mais il semble qu'elle ne soit pas toujours adoptée non plus par un même locuteur dans tous les cas.

1.2.2. Amplitude et vitesse des mouvements articulaires

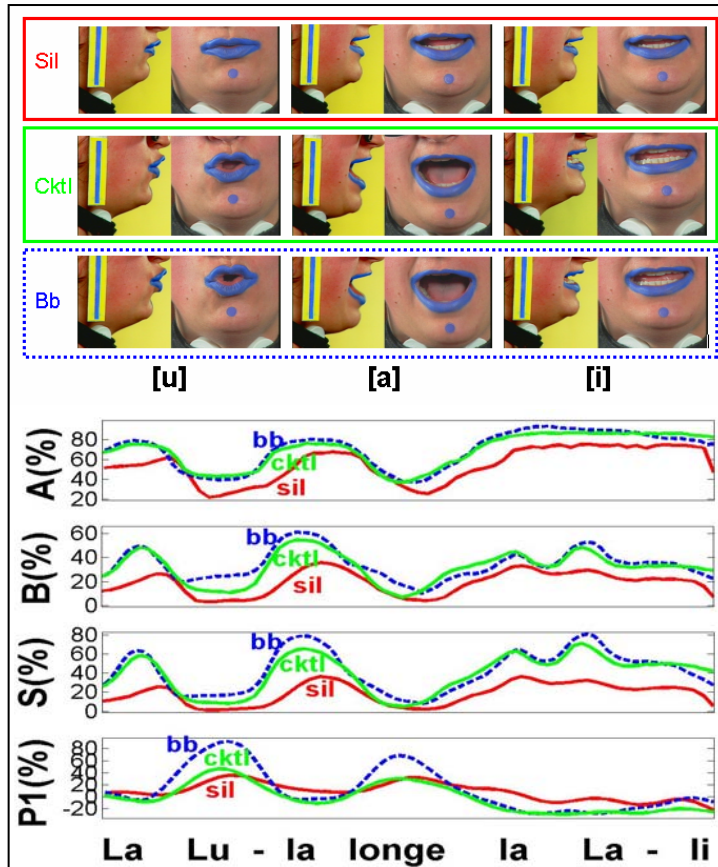


Figure 51. Représentation des mouvements articulaires labiaux d'étirement A, d'ouverture B, de protrusion P1 et de l'aire intéro-labiale S au cours d'un même énoncé de la BD2 produit par une locutrice en condition silencieuse (sil), dans un bruit blanc de 85dBC (bb) et dans un bruit cocktail de 85dBC (cktl). La durée des énoncés a été normalisée sur cette figure de façon à faciliter la comparaison visuelle des signaux.

Les images de face et de profil de la locutrice ont été extraites des enregistrements vidéo de ces énoncés pour le [u] de [Lula], à l'instant où la protrusion était maximale, pour le [a] de [Lula], à l'instant où l'aire intéro-labiale était maximale, et pour le [i] de [Lali], à l'instant où l'étirement était maximal.

Nous avons également observé une quatrième tendance importante de modification de la parole dans le bruit, concernant **l'amplification des mouvements articulaires d'ouverture de la bouche**. L'étirement, l'ouverture et l'aire intéro-labiale augmentent significativement dans le bruit pour les voyelles [a] de toutes les locutrices et quel que soit le type de bruit. Ils augmentent aussi significativement pour les voyelles [i], mais seulement pour 2 locutrices sur 3. Enfin ils augmentent significativement pour les voyelles [u] et pour toutes les locutrices, mais seulement dans le bruit cocktail. Ces résultats corroborent l'étude de Schulman 1989^[304] constituant la référence en ce qui concerne la corrélation entre l'augmentation de l'intensité vocale et celle de l'ouverture de la mâchoire. Plus récemment et plus spécifiquement à la parole Lombard, Davis et al. 2006^[62] ont rapporté à partir de mesures Optotrack que 90% des évolutions dans le bruit quant aux mouvements du visage peuvent être expliqués par l'amplification des mouvements d'ouverture/fermeture de la bouche et l'amplification de la protrusion labiale.

	Moyenne		
	[a]	[i]	[u]
$\Delta\text{Max A}$	7.6% de l'étirement maximal	7.1% de l'étirement maximal	21.2% de l'étirement maximal
$\Delta\text{Max B}$	17.7% de l'ouverture maximale	7.8% de l'ouverture maximale	8.3% de l'ouverture maximale
$\Delta\text{Max S}$	25.1% de l'aire maximale	13.6% de l'aire maximale	4.9% de l'aire maximale
$\Delta\text{Max P1}$	-23.5% de la protrusion maximale	-14% de la protrusion maximale	12.2% de la protrusion maximale

Tableau 14. Tendance moyenne d'évolution de l'amplitude maximale de l'étirement labial A, de l'ouverture labiale B, de l'aire intéro-labiale S et de la protrusion labiale P1 pour les voyelles [a], [i] et [u] produites par les locutrices de la 2^{ème} base de données (BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

Dans notre cas, la protrusion diminue dans le bruit pour les voyelles [a] et [i] de façon conjointe à une plus grande ouverture de la bouche sur ces voyelles. Cette diminution n'est significative que pour 2 locutrices sur 3. La protrusion a au contraire tendance à augmenter pour les voyelles [u], mais de façon non significative pour toutes les locutrices et quel que soit le type de bruit.

Dans l'étude de Schulman 1989 ^[304] sur la parole forte, une augmentation de la vitesse des mouvements articulatoires est observée conjointement à l'augmentation de leur amplitude. Dans notre étude, lorsque les mouvements de protrusion montrent des pics de vitesse, ceux-ci n'augmentent pas significativement dans le bruit pour aucune locutrice et aucun type de bruit. Lorsque l'aire intéro-labiale montre des pics de vitesse, ceux-ci augmentent peu significativement (*) dans la moitié des cas pour les voyelles [a] et [i], et dans un seul cas pour les voyelles [u]. L'augmentation de la vitesse des mouvements articulatoires ne semble donc pas être une caractéristique importante de l'adaptation de la parole dans le bruit. Ce résultat peut être expliqué par le fait que les mouvements articulatoires sont certes plus amples dans le bruit, mais également réalisés sur une durée allongée. Leur vitesse n'est donc pas nécessairement beaucoup plus importante.

	Moyenne		
	[a]	[i]	[u]
Δ Pic de vitesse de S	2.7 %/s	1.3 %/s	1.3 %/s
Δ Pic de vitesse de P1			1.0 %/s

Tableau 5. Tendence moyenne d'évolution des pics de vitesse (lorsqu'ils existent) de l'aire intéro-labiale S et de la protrusion labiale P1 pour les voyelles [a], [i] et [u] produites par les locutrices de la 2^{ème} base de données (BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

1.2.3. Evolution formantique

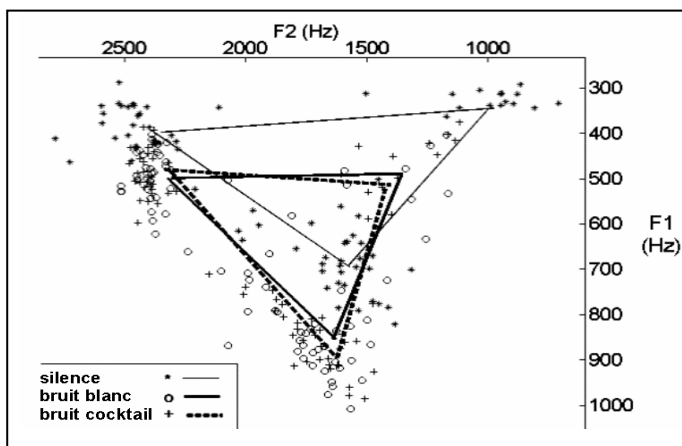


Figure 52. Représentation dans le plan F1*F2 des voyelles produites par une locutrice de la BD1 en condition silencieuse (sil), dans un bruit blanc de 86dBC (bb) et dans un bruit cocktail de 86dBC (cktl). Le triangle représenté a été construit à partir des valeurs moyennes de F1 et de F2 dans chaque condition pour les voyelles [a], [i] et [u].

L'augmentation de la fréquence centrale du premier formant représente une 5^{ème} tendance principale d'évolution de la parole dans le bruit, en particulier chez les femmes.

La fréquence de F1 augmente très significativement dans le bruit sur les voyelles [a], pour tous les locuteurs et locutrices de la BD1 comme pour les locutrices de la BD2, et cela quel que soit le type de bruit. Elle augmente également très significativement sur les voyelles [i] et [u] produites par toutes les locutrices de BD1 et de BD2, quel que soit le type de bruit. En revanche, elle n'évolue jamais significativement sur les voyelles [u] produites par les hommes de BD1, et n'évolue significativement sur les voyelles [i] produites par des hommes de BD1 que dans la moitié des cas.

La fréquence centrale du deuxième formant n'évolue jamais significativement sur les voyelles [a], pour aucun locuteur ou locutrice de BD1 comme pour BD2. Cette absence de significativité générale concerne également les voyelles [u] produites par tous les locuteurs et locutrices de la BD1, tandis qu'au contraire le 2^{ème} formant des voyelles [u] augmente très significativement dans le bruit pour 2 locutrices sur 3 de la BD2. Enfin, le deuxième formant des voyelles [i] évolue de façon significative pour seulement 4 cas sur 20 dans la BD1, et en aucun cas dans la BD2. Les rares cas significatifs montrent d'ailleurs des tendances d'évolution opposées.

Quant à la fréquence centrale du troisième formant, elle augmente de façon significative sur les [a] produits par les locuteurs et locutrices de la BD1 dans 8 cas sur 20, et sur les [a] de la BD2 pour une locutrice sur 3. Elle augmente également significativement sur les [i] de la BD1 dans 7 cas sur 20 mais jamais pour la BD2. Enfin, on observe une augmentation significative du 3^{ème} formant sur

les [u] produits par une locutrice et deux locuteurs de la BD1, dans les deux types de bruit, mais aucune évolution significative pour les autres locuteurs, ni pour les locutrices de la BD2.

		BD1			BD2		
		[a]	[i]	[u]	[a]	[i]	[u]
$\Delta F1$	Diff	247 Hz	88 Hz	84 Hz	198 Hz	89 Hz	86 Hz
	Taux	39%	36%	32%	25%	34%	31%
$\Delta F2$	Diff	41 Hz	-16 Hz	21 Hz	24 Hz	1 Hz	205 Hz
	Taux	3%	-1%	7%	2%	0%	31%
$\Delta F3$	Diff	51 Hz	-81 Hz	35 Hz	83 Hz	-15 Hz	23 Hz
	Taux	2%	-2%	1%	3%	0%	1%

Tableau 15. Tendence moyenne d'évolution des 3 premiers formants vocaliques pour les voyelles [a], [i] et [u] produites par les locuteurs et locutrices de la 1^{ère} et de la 2^{ème} base de données (BD1 et BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

L'évolution dans le bruit des premiers formants vocaliques est celle qui montre le plus de variabilité entre les différentes études antérieures, entre les genres et entre les langues étudiées. Ainsi, la quasi totalité des études s'accordent sur l'augmentation du 1^{er} formant dans le bruit ou en situation d'effort vocal (Mixdorff et al. 2006^[244]; Pisoni et al. 1985^[272]; Traunmüller et al. 2000^[356]; Ternström et al. 2006^[345]; Van Summers et al. 1988^[361]; Castellanos et al. 1996^[46]; Davis et al. 2006^[62]; Junqua 1992^[173]; Lienard et al. 1999^[215]). Cependant, Bond et al. 1989^[31] rapportent que le F1 augmente surtout pour les voyelles [u] et Rostolland 1982^[289] que le F1 augmente pour toutes les voyelles mais peu pour les [a]. Nous n'observons pas de tels résultats, mais sommes au contraire en accord avec ceux de Boril et al. 2005^[32] notant une augmentation du F1 plus importante pour les voyelles ouvertes [a]. Il est possible que cette augmentation plus importante sur les voyelles [a] contribue à une meilleure distinctivité des voyelles du Français selon leur trait phonologique d'ouverture/fermeture. Nous examinerons cette hypothèse dans le chapitre 7.

En ce qui concerne le deuxième formant, nos résultats sont en accord avec ceux de Boril et al. 2005^[32], Bond et al. 1989^[31] et Mixdorff et al. 2006^[244] rapportant légère augmentation pour la majorité des voyelles, surtout pour les [u], mais au contraire pas d'augmentation voire une diminution pour les voyelles [i]. Pisoni et al. 1985^[272] observent une diminution de F2 sur toutes les voyelles, quand Castellanos et al. 1996^[46] et Davis et al. 2006^[62] notent au contraire une augmentation pour toutes les voyelles, Lienard et al. 1999^[215] et Van Summers et al. 1988^[361] ne remarquent pas d'évolution significative du deuxième formant, et enfin Junqua 1993^[171] rapporte une augmentation de F2 seulement pour les femmes.

Pour finir, peu d'études se sont intéressées à l'évolution du 3^{ème} formant dans le bruit ou en situation d'effort vocal. Nos résultats corroborent plutôt ceux de Traunmüller et al. 2000^[356] et de Lienard et al. 1999^[215] sur la parole forte, n'observant pas d'évolution significative de F3. Ils vont au contraire à l'encontre de ceux de Davis et al. 2006^[62] sur la parole Lombard rapportant que le 3^{ème} formant augmente dans le bruit, davantage que F1 et F2.

1.3. Au niveau glottique

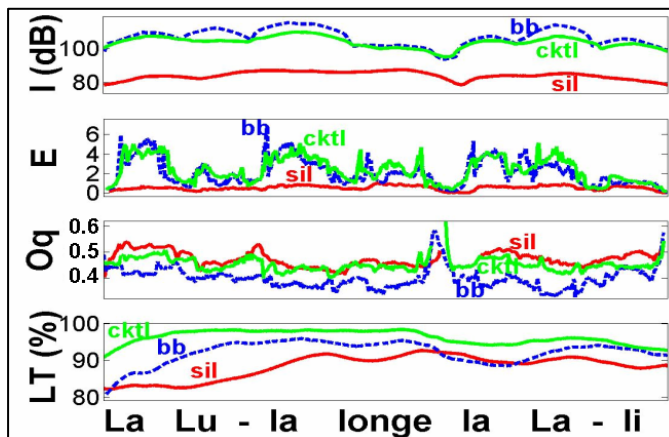


Figure 53. Représentation de la vitesse de fermeture des cordes vocales E, du quotient ouvert Oq et des mouvements verticaux du larynx au cours d'un même énoncé de la BD2 produit par une locutrice en condition silencieuse (sil), dans un bruit blanc de 85dBC (bb) et dans un bruit cocktail de 85dBC (cktl). La durée des énoncés a été normalisée sur cette figure de façon à faciliter la comparaison visuelle des signaux. L'énoncé n'est constitué que de segments voisins. C'est pourquoi les courbes sont continues.

La vitesse de fermeture des cordes vocales augmente de façon très significative avec le bruit pour tous les locuteurs et tous les types de bruit. Cela est tout à fait logique puisque ce paramètre glottique est corrélé à l'intensité vocale.

L'augmentation de l'effort vocal dans le bruit s'accompagne également d'une **diminution très significative du quotient ouvert** pour 8 locuteurs sur 10 de la BD1 et 2 locutrices sur 3 de la BD2. Une locutrice montre au contraire un quotient ouvert tendant à augmenter dans le bruit dans les deux bases de données et quel que soit le type de bruit. L'existence de tendances opposées d'évolution du quotient ouvert dans le bruit a pour effet de donner une évolution moyenne du quotient ouvert plus faible qu'elle ne l'est (0.06 en valeur absolue au lieu de 0.03 pour la BD1 et 0.07 en valeur absolue au lieu de 0.04 pour la BD2). De même, la forte variabilité intra-individuelle sur l'évolution du quotient ouvert dans le bruit laisse penser que ces différentes tendances d'évolution du quotient ouvert n'existent pas seulement entre les individus mais également au sein d'un même individu (cf. Tableau 16).

Enfin, **le larynx monte de façon très significative** pour 2 locutrices sur 3 dans les deux types de bruit. Pour la troisième, il monte de façon peu significative dans le bruit cocktail et a tendance à descendre de façon non significative dans le bruit blanc. Cette dernière locutrice est justement celle pour laquelle le quotient ouvert a tendance à augmenter dans le bruit.

	BD1	BD2
Δ Quotient ouvert Oq	-0.03	-0.04
Δ Hauteur du larynx (LT)		17% de la hauteur maximale

Tableau 16. Tendence moyenne d'évolution du quotient ouvert, de la vitesse de fermeture des cordes vocales, de l'amplitude maximale des mouvements verticaux du larynx pour les locuteurs et locutrices de la 1ère et de la 2ème base de données (BD1 et BD2). Ces résultats sont présentés de façon plus détaillée en Annexe Res5.

Ces différents résultats corroborent ceux de nombreux auteurs, en particulier de Huang et al. 1995^[160], notant une augmentation du quotient fermé (donc équivalent à une diminution du quotient ouvert) avec l'effort vocal. Ce même auteur observe également une diminution du Jitter et du Shimmer conjointement à une augmentation de l'intensité vocale, ce que nous avons également constaté dans l'exploration de notre base de données.

2. QUELLES SONT LES TENDANCES PARTAGEES ET LES DIFFERENCES INDIVIDUELLES ?

Nous avons vu au paragraphe précédent que certains paramètres évoluent dans le bruit avec la même tendance pour tous les locuteurs tandis que d'autres présentent des évolutions suivant la même tendance mais plus ou moins significatives selon les individus, quand d'autres encore présentent des tendances d'évolution posées.

Dans le but de caractériser les différences interindividuelles d'adaptation de la parole dans le bruit, il est particulièrement intéressant de distinguer les adaptations communes à tous les locuteurs pour s'adapter à un environnement bruyant des adaptations individuelles pouvant expliquer le fait que certaines personnes développent davantage de symptômes que d'autres à l'issue d'un effort vocal.

Nous allons donc maintenant examiner les différences interindividuelles d'évolution dans le bruit pour les paramètres ayant montré une variation significative au paragraphe précédent. Nous tâcherons dans un premier temps de comprendre ces différences interindividuelles par rapport au comportement vocal des locuteurs dans le silence, par rapport à leur genre, et par rapport à leur technique vocale (cf. Tableau 3 du paragraphe 3.1.1 dans le chapitre 3).

Pour chaque paramètre, nous avons calculé l'évolution de chaque syllabe (toujours les mêmes qu'au paragraphe 1) du silence au bruit blanc et du silence au bruit cocktail. Nous avons ensuite comparé ces évolutions, sans distinguer les deux types de bruit, entre les différents locuteurs. Pour évaluer la significativité des différences interindividuelles, nous avons effectué un test Anova à un facteur (INDIVIDU) et 10 niveaux pour la BD1 (L1 à L10) ou à 3 niveaux pour la BD2 (L2, L6 et L11). Cependant, une différence peut être statistiquement significative et pour autant très faible, sans aucun impact perceptif. Aussi, nous n'avons pas uniquement considéré la significativité des tests Anova mais également comparé la grandeur des différences interindividuelles relativement à la grandeur de l'effet moyen d'évolution du silence au bruit et relativement à la variabilité intra-individuelle.

Les détails de ces différents résultats sont consultables en Annexe Res5.

2.1. Tendance partagée sans grande variabilité interindividuelle

Parmi les paramètres évoluant significativement du silence au bruit, seule **la fréquence fondamentale** montre assez peu de variabilité inter-individuelle, que ce soit dans la 1^{ère} ou dans la deuxième base de données (cf. Figure 54). La différence interindividuelle moyenne d'évolution de la F0 avec le bruit est pourtant statistiquement significative dans les deux cas mais n'atteint jamais plus de 2.3 tons dans la BD1 (différence maximale entre les locuteurs L4 et L8, suivie de près par la différence entre les locuteurs L4 et L7) et 1.2 ton dans la BD2 (entre les locutrices L2 et L11), ce qui est presque toujours inférieur au tiers de l'effet d'évolution du silence au bruit (de 4.7 tons). Par ailleurs, cette différence interindividuelle est inférieure à la variabilité intra-individuelle dans la grande majorité des cas pour la BD1. Au contraire, la locutrice L2 de la BD2 se distingue des deux autres par une différence d'évolution supérieure à la variabilité intra-individuelle moyenne (cf. Annexe Res5 pour les détails de ces résultats).

Dans la BD1, on remarque également que l'évolution de la fréquence fondamentale (en tons) n'est pas significativement différente entre les hommes et les femmes, contrairement aux observations de Junqua 1993^[171] et Castellanos et al. 1996^[46] rapportant une augmentation de la F0 plus importante pour les hommes que pour les femmes. Cette différence quant à nos résultats ne peut pas être mise sur le compte de l'unité choisie : en Tons dans notre étude et en Hertz dans les leurs. Il ne semble pas non plus y avoir de lien particulier entre l'évolution de la F0 du silence au bruit et sa valeur dans le silence.

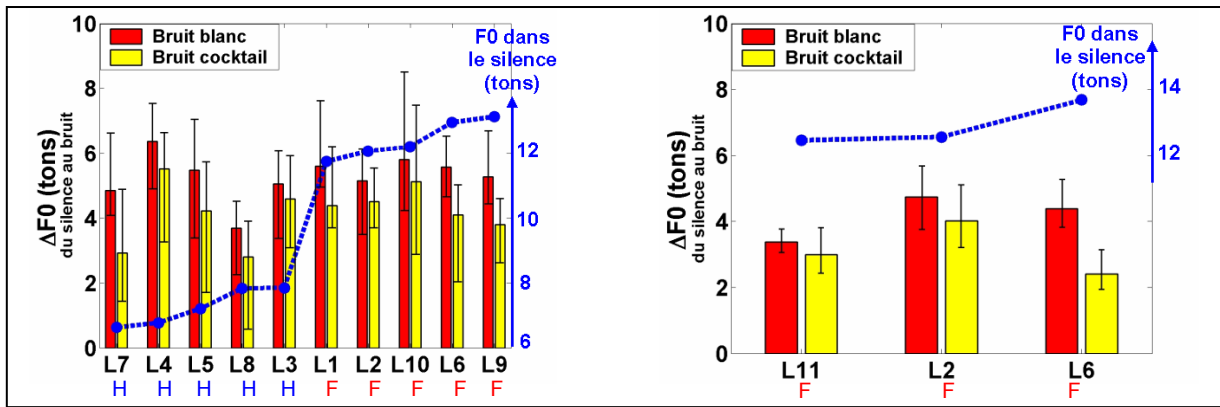


Figure 54. Représentation des différences interindividuelles d'évolution de la fréquence fondamentale (F0) dans le bruit pour la 1^{ère} base de données (à gauche) et la 2^{ème} base de données (à droite). La F0 est exprimée en tons par rapport à la fréquence de référence 50Hz. La courbe bleue en pointillés représente la F0 moyenne de chaque locuteur en condition silencieuse. Nous avons ordonné les locuteurs par ordre croissant de F0 en condition silencieuse. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution de sa F0 entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

2.2. Tendence unique avec différentes magnitudes

Mis à part la fréquence fondamentale, tous les autres paramètres montrent une évolution du silence au bruit assez variable selon les locuteurs. Pour la plupart, il s'agit quand même d'une même tendance d'évolution mais d'importance plus ou moins grande selon les individus.

Ainsi, l'intensité vocale, l'énergie spectrale, le premier formant vocalique et l'amplitude des mouvements articulatoires labiaux varient avec la même tendance pour tous les locuteurs, mais de façon non significative pour certains, moyennement significative pour d'autres ou encore très significatives pour d'autres.

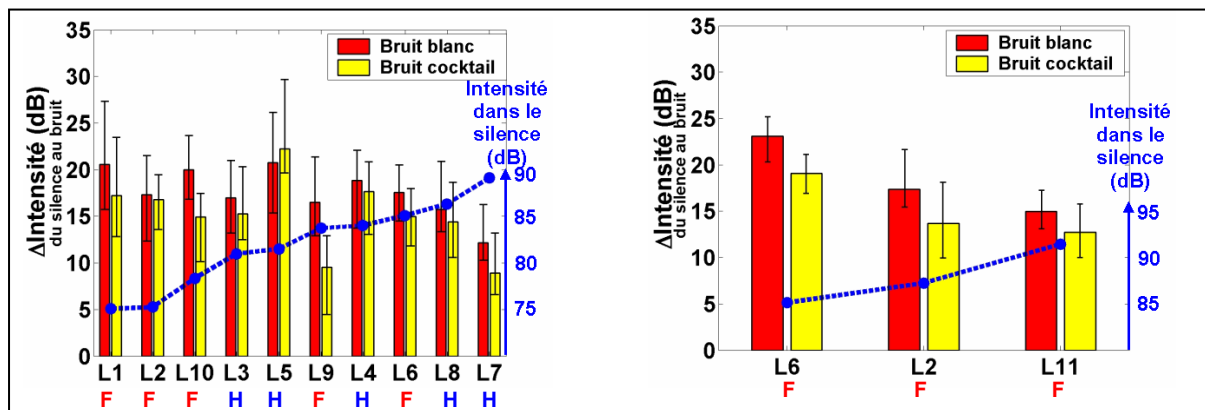


Figure 55 . Représentation des différences interindividuelles d'évolution de l'intensité vocale dans le bruit pour la 1^{ère} base de données (à gauche) et la 2^{ème} base de données (à droite). La courbe bleue en pointillés représente l'intensité vocale moyenne de chaque locuteur en condition silencieuse. Nous avons ordonné les locuteurs par ordre croissant d'intensité vocale en condition silencieuse. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution de son intensité vocale entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

Ainsi, la différence interindividuelle moyenne d'évolution de l'intensité vocale avec le bruit est statistiquement significative pour la BD1 comme pour la BD2, allant jusqu'à 9.2 dB pour la BD1 (entre les locuteurs L5 et L7) et jusqu'à 7.0 dB (entre les locutrices L6 et L11) pour la BD2. Tous les locuteurs et locutrices ne se distinguent pas pour autant de façon significative, mais leur différence

dépasse la variabilité intra-individuelle dans 31% des cas pour la BD1 et dans 2 cas sur 3 pour la BD2. En particulier, le locuteur L5 augmente davantage son intensité vocale dans le bruit que les autres locuteurs et locutrices, tandis que le locuteur L7 et la locutrice L9 l'augmentent un peu moins. De même, la locutrice L6 augmente davantage son intensité vocale dans le bruit que les deux autres locutrices de la BD2. Dans la BD1, on remarque une différence interindividuelle très importante d'intensité vocale dans le silence, allant jusqu'à 15 dB. Les intensités les plus faibles correspondent à des locutrices et les intensités les plus importantes à des locuteurs, mais il n'y a pas pour autant de séparation des genres aussi nette que pour la F0. Quoi qu'il en soit, cette variabilité importante d'intensité vocale dans le silence ne semble pas influencer systématiquement la magnitude de l'évolution de l'intensité vocale lors de l'exposition au bruit. On remarque cependant que la locutrice L1 et le locuteur L5 augmentant le plus leur intensité dans le bruit présentent une faible intensité dans le silence, tandis que la locutrice L9 et le locuteur L7, augmentant moins que les autres leur intensité dans le bruit, présentent une intensité vocale importante dans le silence. De même dans la BD2, les locutrices augmentent d'autant moins leur intensité vocale dans le bruit que celle-ci est importante dans le silence. Ce phénomène peut correspondre à un effet plafond chez les locuteurs ou locutrices parlant déjà fort dans le silence, et disposant par conséquent d'une marge moins grande pour surenchérir dans le bruit. Ce phénomène peut être également le signe d'une surestimation constante chez ces personnes de l'effort à fournir. Quelle que soit la situation de communication, ces locuteurs ou locutrices seraient « par défaut » dans un mode de fonctionnement excessif.

Contrairement à Egan 1972^[80] observant une augmentation de l'intensité vocale plus importante chez les femmes, ou à Junqua 1993^[171] et Rostolland 1982^[288] observant une augmentation plus importante chez les hommes, nous ne remarquons pas dans notre étude de différence significative entre les deux genres concernant l'évolution de leur intensité vocale dans le bruit. Nous n'observons pas non plus de variabilité plus importante chez les femmes, comme le note Egan 1972^[80].

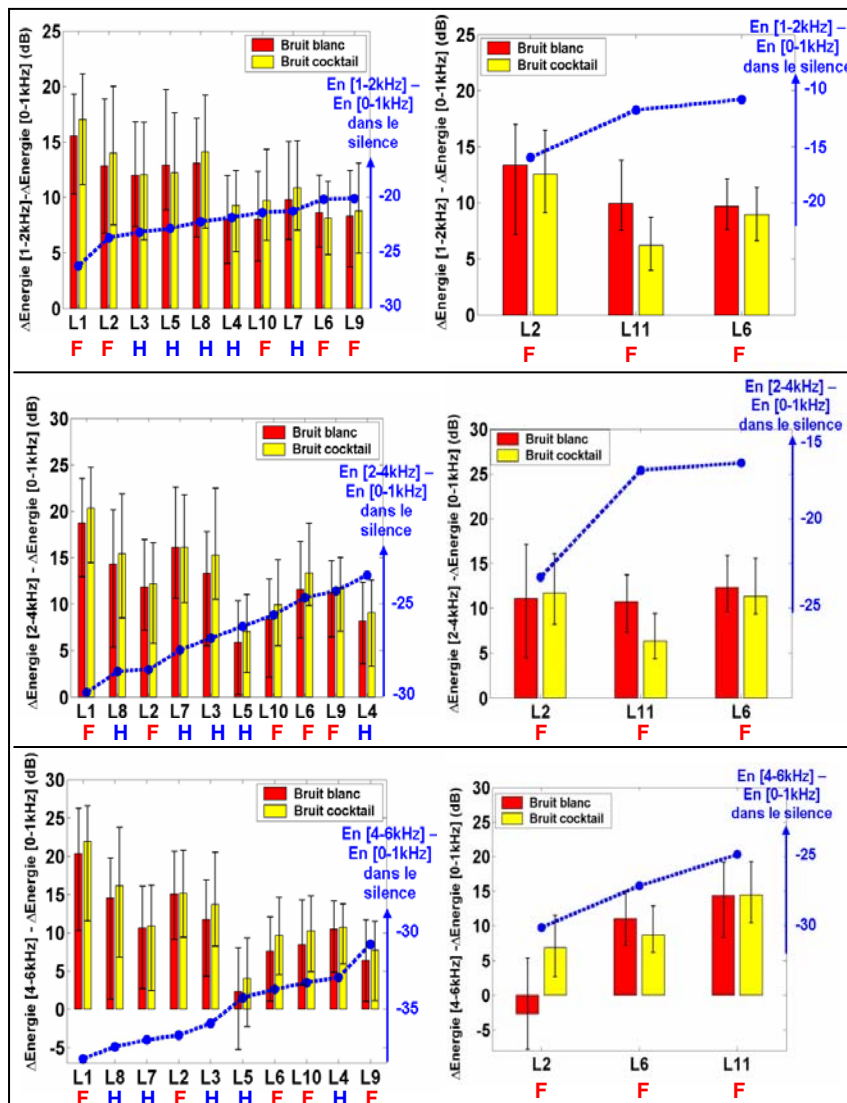


Figure 56. Représentation des différences d'évolution de l'énergie spectrale dans le bruit pour la 1^{ère} base de données (à gauche) et la 2^{ème} base de données (à droite). La courbe bleue en pointillés représente l'énergie moyenne de chaque locuteur en condition silencieuse pour les différentes bandes de fréquences [0-1kHz], [1-2kHz], [2-4kHz] et [4-6kHz]. Nous avons ordonné les locuteurs par ordre croissant de cette énergie en condition silencieuse. Cet ordre n'est pas le même pour toutes les bandes de fréquences. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution de son énergie spectrale entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

L'évolution de l'énergie spectrale montre une différence interindividuelle moyenne d'évolution avec le bruit statistiquement significative pour la BD1 comme pour la BD2 (cf. Figure 56), et pour les 3 bandes de bruit considérées ([1-2kHz], [2-4kHz] et [4-6kHz]).

- Pour la bande [1-2kHz], il se distingue 2 groupes de locuteurs : les locuteurs L1, L2 (femmes), L3, L5 et L8 (hommes) augmentent beaucoup l'énergie spectrale de leur parole dans cette zone par rapport à l'énergie dans la zone du fondamental tandis que les locuteurs L4, L7 (hommes), L6, L10 et L9 (femmes) augmentent également l'énergie de leur parole dans cette zone, mais de façon moindre. La différence entre ces deux groupes est supérieure à la variabilité intra-individuelle moyenne ainsi qu'au tiers de l'effet global d'exposition au bruit. Il atteint un maximum de 7.9 dB entre les locutrices L1 et L9. De même, la différence interindividuelle d'évolution de l'énergie entre 1 et 2 kHz peut atteindre 5.1dB dans la BD2 entre les locutrices L2 et L11. Malgré une variabilité interindividuelle assez faible dans le silence, on observe que les locuteurs qui augmentent le plus leur énergie entre 1 et 2kHz dans le bruit sont également ceux qui présentent une énergie plus faible dans cette zone en condition silencieuse.
- Pour la bande [2-4kHz], il ne se dégage pas dans la BD1 deux groupes distincts comme précédemment, mais de fortes différences interindividuelles correspondant en moyenne au tiers de l'effet global d'exposition au bruit, et pouvant atteindre jusqu'à 11.6dB (entre la locutrice L1 et le locuteur L5). La locutrice L1 et les locuteurs L3 et L8, qui renforçaient déjà beaucoup leur énergie entre 1 et 2kHz, continuent à renforcer davantage que d'autres leur énergie entre 2 et 4kHz. De même, le locuteur L4 et la locutrice L10 qui renforçaient moins que les autres leur énergie entre 1 et 2 kHz, continuent à renforcer moins que d'autres leur énergie entre 2 et 4kHz. Au contraire, le locuteur L5 timbre peu sa voix entre 2 et 4kHz tandis qu'il la timbrait beaucoup entre 1 et 2kHz, et le locuteur L7 timbre beaucoup sa voix entre 2 et 4kHz alors qu'il la timbrait peu entre 1 et 2 kHz. Pour la BD2, les différences interindividuelles sont très faibles entre les 3 locutrices et ne dépassent pas les 3.2dB.
- Pour la bande [4-6kHz], on observe des résultats ressemblants à ceux de la bande [2-4kHz], avec des différences encore plus marquées entre les locuteurs : L1, L2 (femmes), L8 et L3 (hommes) timbrent davantage leur voix entre 4 et 6 kHz que L6, L9, L10 (femmes), L7 et L4 (hommes). Enfin le locuteur L5 se démarque nettement avec un renforcement non significatif de son énergie dans cette zone, relativement à celle du fondamental, voire le contraire dans certains cas. De même, la locutrice L2 a un comportement similaire au locuteur L5 dans la BD2. Ce résultat est plutôt étonnant car elle fait au contraire partie des locuteurs et locutrices de BD1 renforçant leur énergie entre 4 et 6kHz de façon moyennement importante. Cependant, on remarque que la voix des 3 locutrices de BD2 est beaucoup plus timbrée en moyenne dans le silence que celles des locuteurs et locutrices de la BD1, quelle que soit la bande de fréquences considérées. Ce changement de comportement de la locutrice L2 observé entre les 2 bases de données peut être dû au fait qu'elle est déjà à son maximum dans le silence dans le cas de la BD2 et qu'elle ne peut pas surenchérir dans le bruit.

Plusieurs études antérieures ont observé chez les femmes une augmentation importante de l'énergie spectrale entre 4 et 5kHz ou autour de 5kHz, tandis que l'énergie augmente surtout entre 2 et 4kHz chez les hommes (Castellanos et al. 1996^[46] ; Junqua 1993^[171]). De même, Ternström et al. 2006^[345] rapportent que le centre de gravité spectrale de la voix reste stable dans le bruit chez les hommes tandis qu'il s'élève proportionnellement à la F0 pour les femmes. Dans notre cas, nous n'observons pas de différence significative entre les hommes et les femmes au niveau de l'évolution de l'énergie spectrale de leur voix dans le bruit : Ainsi, parmi les locuteurs augmentant le plus leur énergie spectrale dans le bruit dans la zone [4-6kHz], il y a autant de femmes (L1 et L2) que d'hommes (L3 et L8). De même il n'y a pas de nette distinction entre les deux genres quant à l'augmentation de l'énergie acoustique dans la zone [1-2kHz] ou [2-4kHz].

L'évolution du premier formant dans le bruit permet également de distinguer des groupes de locuteurs aux évolutions différentes.

- Sur les voyelles [a], le premier formant est classiquement plus élevé chez les femmes que chez les hommes. C'est bien ce que l'on observe au niveau des valeurs du premier formant dans le silence (cf. Figure 57). En contrepartie, l'évolution du premier formant avec le bruit est beaucoup plus importante chez les hommes que chez les femmes, contrairement à l'étude de Junqua 1993^[171] notant une augmentation plus importante chez les femmes. La différence interindividuelle moyenne d'augmentation du 1^{er} formant est de 143Hz, ce qui représente la moitié de l'effet global d'évolution du premier formant dans le bruit et est largement au-dessus de la variabilité intra-individuelle. Au sein de chaque genre, on observe peu de variabilité entre les 5 hommes. En revanche, la locutrice L2 se distingue des autres femmes de la BD1 par une évolution plus importante de F1. Cela n'est

pas le cas dans la BD2, où c'est plutôt la locutrice L11 qui montre une forte augmentation de F1 par rapport aux autres.

- Sur les voyelles [i] et [u], on distingue également deux groupes de locuteurs, avec une différence interindividuelle moyenne de 84 Hz, équivalente à l'effet global moyen d'augmentation du F1 dans le bruit. Il ne s'agit pas cette fois-ci d'une séparation des locuteurs par genre : L1, L2 (femmes), L3, L5 et L7 (hommes) augmentent davantage leur premier formant dans le bruit tandis que L4, L8 (hommes), L6, L9 et L10 (femmes) l'augmentent peu. Ce comportement différent des individus ne peut pas s'expliquer par la valeur de leur premier formant dans le silence, très semblable entre tous les individus. Les locutrices de BD2 ne montrent pas de différences notables concernant l'évolution de leur premier formant dans le bruit sur ces deux voyelles.

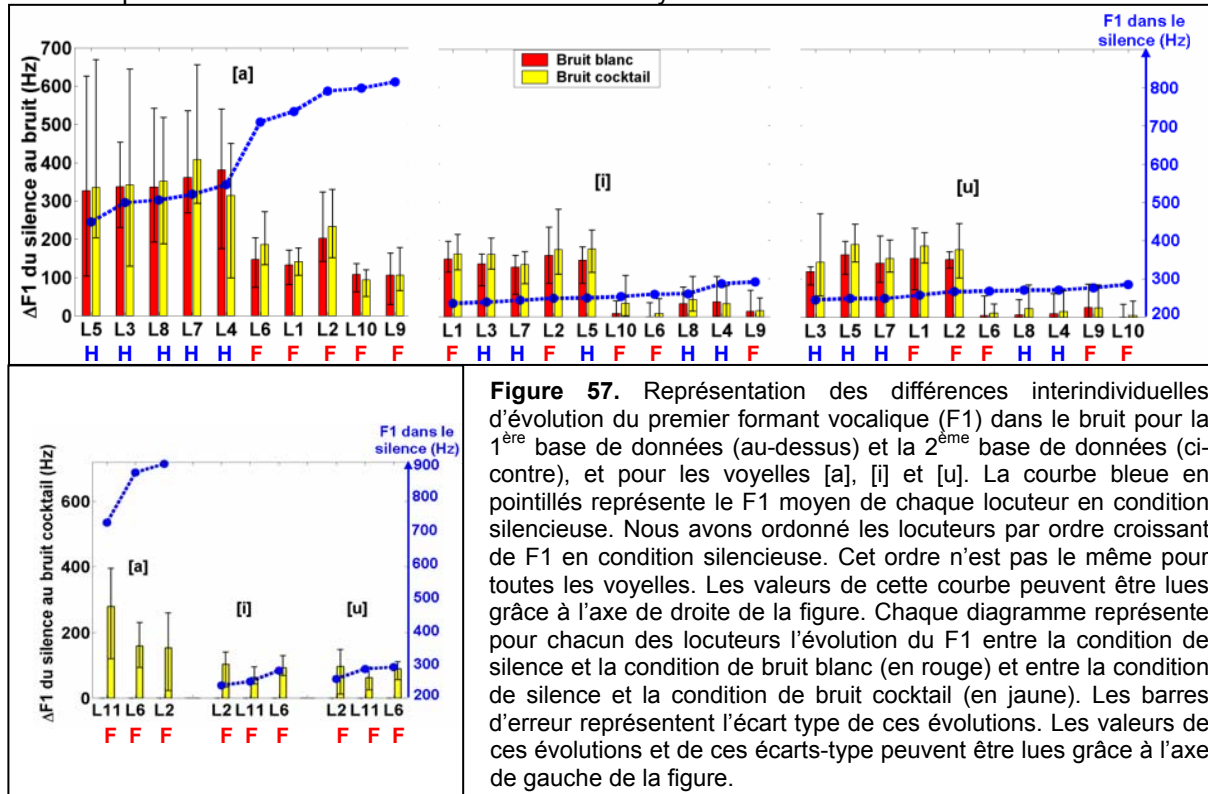


Figure 57. Représentation des différences interindividuelles d'évolution du premier formant vocalique (F1) dans le bruit pour la 1^{ère} base de données (au-dessus) et la 2^{ème} base de données (ci-contre), et pour les voyelles [a], [i] et [u]. La courbe bleue en pointillés représente le F1 moyen de chaque locuteur en condition silencieuse. Nous avons ordonné les locuteurs par ordre croissant de F1 en condition silencieuse. Cet ordre n'est pas le même pour toutes les voyelles. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution du F1 entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

Enfin, l'évolution de certains mouvements articulatoires labiaux dans le bruit montre également des différences interindividuelles non négligeables.

- Ainsi, l'étirement labial augmente dans le bruit avec une variabilité interindividuelle équivalente à la moitié de l'effet global d'évolution du silence au bruit. Celui-ci n'est pas très important (bien que significatif) pour les voyelles [a] et [i]. Mais il est beaucoup plus important pour les voyelles [u]. La différence interindividuelle atteint alors 14% pour ces voyelles (entre les locutrices L2 et L6). On observe que l'augmentation de l'étirement sur les [u] est d'autant plus importante que celui-ci est faible dans le silence.
- L'ouverture labiale évolue de façon semblable pour les 3 locutrices sur les voyelles [i] et [u]. En revanche, L2 a tendance à davantage ouvrir les lèvres dans le bruit sur les voyelles [a] que L6, elle-même davantage que L11. Cette différence interindividuelle est déjà valable dans le silence et se trouve donc d'autant plus renforcée dans le bruit.
- La protrusion ne diminue pas de façon différente sur les [a] pour les 3 locutrices. En revanche, elle diminue davantage sur les [i] pour L2 que pour L6, avec une différence interindividuelle moyenne de 10% équivalente à l'ordre de grandeur de l'effet global de diminution sur les [i]. Mais la variabilité interindividuelle la plus importante est observée pour l'augmentation de la protrusion dans le bruit sur les voyelles [u], où la différence entre les locutrices atteint 13% entre L11 et L2, et 8% en moyenne, soit la moitié de l'effet global d'augmentation de la protrusion sur les [u]. On remarque que cette augmentation est d'autant plus faible avec le bruit qu'elle était déjà importante dans le silence.

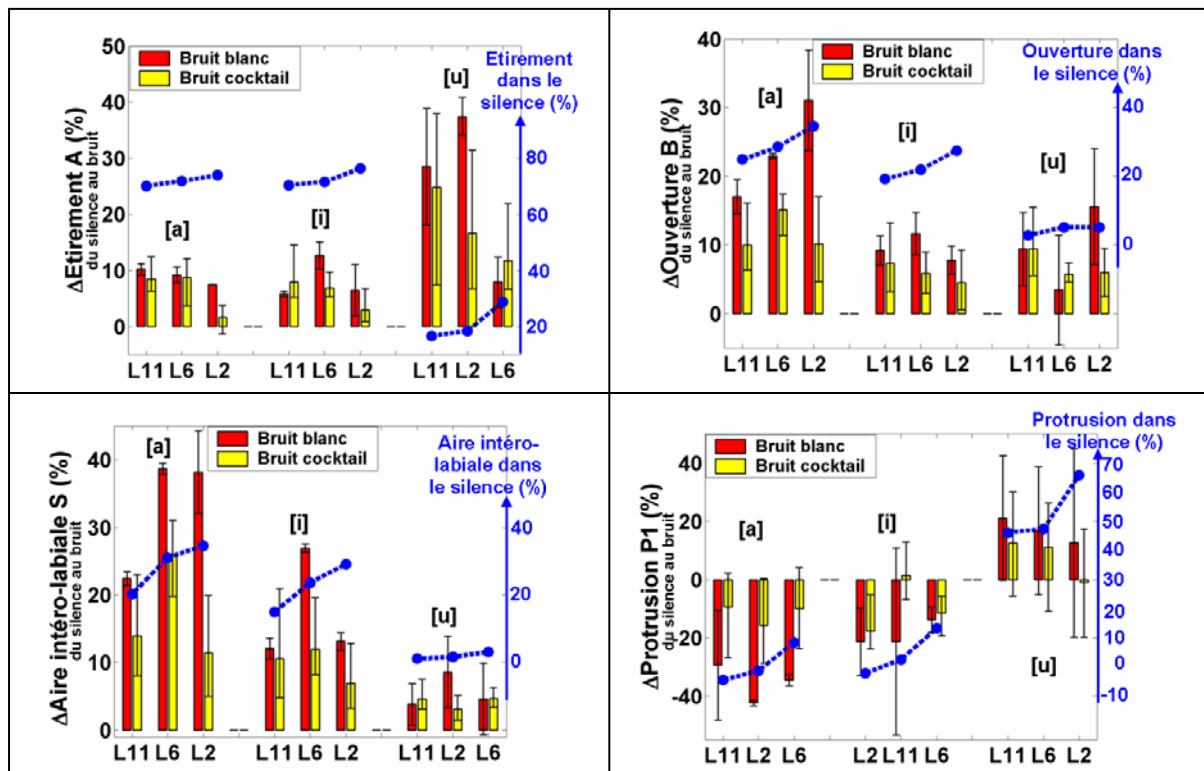


Figure 58. Représentation des différences interindividuelles d'évolution de l'amplitude des mouvements articulatoires labiaux dans le bruit pour la 2^{ème} base de données, et pour les voyelles [a], [i] et [u]. La courbe bleue en pointillés représente l'amplitude moyenne des mouvements d'éirement, d'ouverture, de l'aire intéro-labiale et des mouvements de protrusion de chaque locuteur en condition silencieuse. Nous avons ordonné les locuteurs par ordre croissant de cette amplitude des mouvements articulatoires en condition silencieuse. Cet ordre n'est pas le même pour tous les mouvements labiaux ni pour toutes les voyelles. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution de l'amplitude de ses mouvements articulatoires labiaux entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

2.3. Tendances opposées

Il reste encore quelques paramètres pour lesquels on observe une évolution significative dans le bruit mais avec des tendances opposées selon les locuteurs.

Ainsi, **le débit de parole** est parfois ralenti, parfois accéléré selon les locuteurs, et même au sein d'un même locuteur ou d'une même locutrice. Nous remarquons également que cette variabilité intra-individuelle est beaucoup plus importante dans la première base de données que dans la deuxième base de données. Cela provient sans aucun doute de la plus grande liberté de placement des mots-cibles dans l'énoncé dans la BD1, conduisant à un biais non négligeable lorsqu'on calcule l'évolution de la durée syllabique du silence au bruit. Dans la BD2, le débit de parole est significativement ralenti pour toutes les locutrices. Cependant, L6 et L2 allongent beaucoup plus leurs syllabes dans le bruit que L11, avec une différence interindividuelle moyenne de 22ms, équivalente à la moitié de l'effet global d'allongement des syllabes dans le bruit. Dans une étude antérieure, Ternström et al. 2002^[346] rapporte que le temps de phonation dans le bruit est plus important pour les femmes que pour les hommes. Dans notre 1^{ère} base de données, nous observons également une tendance (non significative) chez les femmes à davantage ralentir leur débit de parole que les hommes, mis à part le locuteur L5 augmentant sa durée syllabique dans le bruit de façon très semblable aux locutrices.

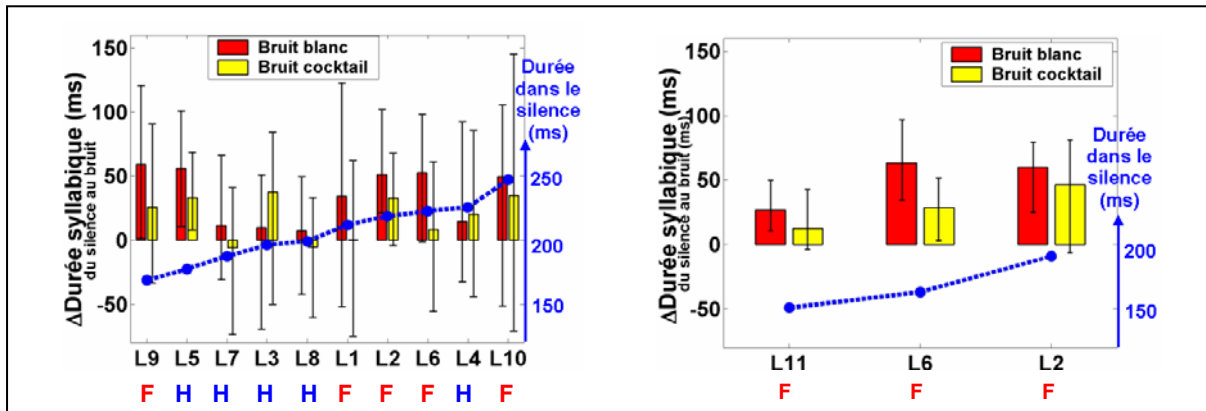


Figure 59. Représentation des différences interindividuelles d'évolution de la durée syllabique dans le bruit pour la 1^{ère} base de données (à gauche) et la 2^{ème} base de données (à droite). La courbe bleue en pointillés représente la durée syllabique moyenne de chaque locuteur en condition silencieuse. Nous avons ordonné les locuteurs par ordre croissant de cette durée syllabique en condition silencieuse. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution de sa durée syllabique entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

De même, on observe une tendance majoritaire de diminution du **quotient ouvert** dans le bruit (cf. Figure 60). C'est le cas dans la BD1 de L7, L8, L5, L3 (hommes) et L9 (femme), qui montre la plus forte diminution. D'autres locuteurs cependant ne présentent pas d'évolution significative du quotient ouvert dans le bruit (L4 (homme), L10, L1, L6 (femmes)). Enfin, une locutrice, L2, se démarque nettement des autres par une augmentation conséquente du quotient ouvert. Ce comportement est de nouveau observé dans la BD2. Les différences interindividuelles d'évolution du quotient ouvert ne semblent pas liées à la valeur du quotient ouvert dans le silence.

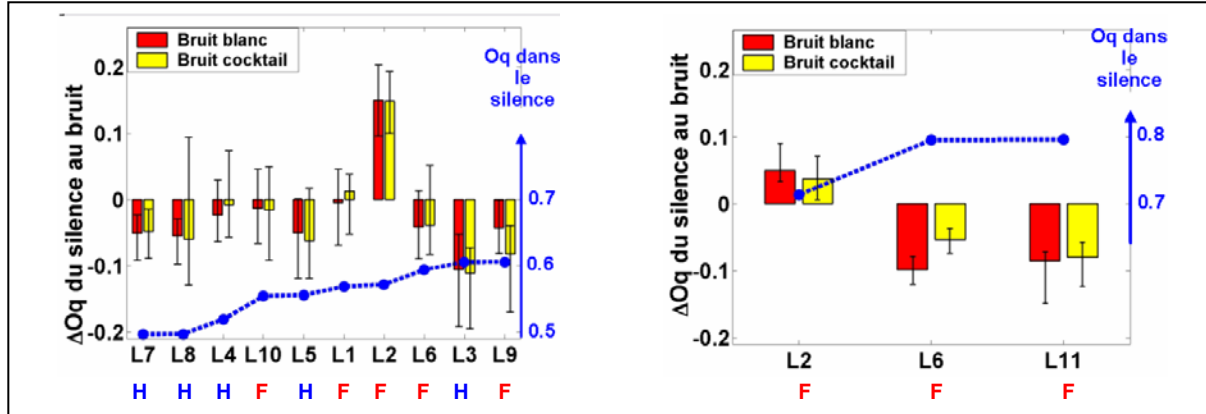


Figure 60. Représentation des différences interindividuelles d'évolution du quotient ouvert laryngé dans le bruit pour la 1^{ère} base de données (à gauche) et la 2^{ème} base de données (à droite). La courbe bleue en pointillés représente le quotient ouvert moyen de chaque locuteur en condition silencieuse. Nous avons ordonné les locuteurs par ordre croissant de ce quotient ouvert en condition silencieuse. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacun des locuteurs l'évolution de son quotient ouvert entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

Enfin, **l'élévation du larynx** dans le bruit montre également une grande différence entre les locutrices de la BD2. Ainsi, seules L6 et L11 montent davantage leur larynx dans le bruit que dans le silence, et L6 de façon plus importante (20% d'écart). Nous remarquons que L6 est la locutrice maintenant son larynx le plus bas en condition silencieuse. Au contraire, L2 conserve sensiblement la même hauteur de larynx dans le bruit, voire même parfois une hauteur moins importante.

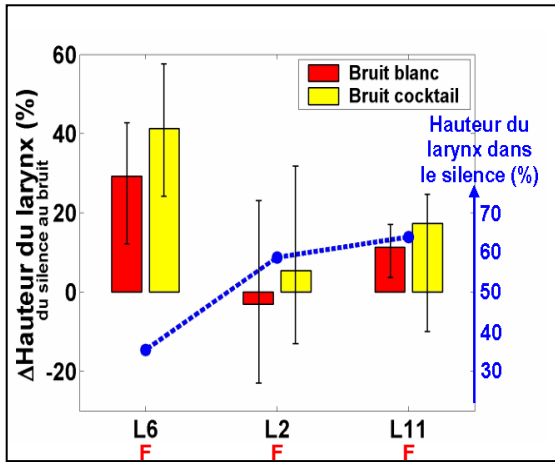


Figure 61. Représentation des différences interindividuelles d'évolution de l'amplitude des mouvements verticaux du larynx dans le bruit pour la 2^{ème} base de données. La courbe bleue en pointillés représente le quotient ouvert moyen de chaque locuteur en condition silencieuse. Nous avons ordonné les locutrices par ordre croissant de leur élévation du larynx en condition silencieuse. Les valeurs de cette courbe peuvent être lues grâce à l'axe de droite de la figure. Chaque diagramme représente pour chacune des locutrices l'évolution l'amplitude de ses mouvements verticaux de son larynx entre la condition de silence et la condition de bruit blanc (en rouge) et entre la condition de silence et la condition de bruit cocktail (en jaune). Les barres d'erreur représentent l'écart type de ces évolutions. Les valeurs de ces évolutions et de ces écarts-type peuvent être lues grâce à l'axe de gauche de la figure.

3. QUELS SONT LES PARAMETRES DONT L'EVOLUTION DANS LE BRUIT EST DEPENDANTE DU TYPE DE BRUIT ?

Nous avons vu au cours des paragraphes précédents qu'une partie de variabilité d'adaptation pouvait provenir du type de bruit dans lequel le locuteur est immergé. Nous allons examiner ici plus en détail l'effet du type de bruit sur l'évolution des descripteurs de la voix dans le bruit.

Pour chaque paramètre, nous avons calculé l'évolution de chaque syllabe (toujours les mêmes qu'au paragraphe 1) du silence au bruit blanc et du silence au bruit cocktail. Nous avons ensuite comparé ces évolutions pour chaque locuteur, et tous locuteurs confondus. Pour évaluer la significativité de l'effet du type de bruit, nous avons effectué un test Anova à un facteur (TYPE DE BRUIT).

Cependant, une différence peut être statistiquement significative et pour autant très faible, sans aucun impact perceptif. Aussi, nous n'avons pas uniquement considéré la significativité des tests Anova mais également comparé pour chaque locuteur la grandeur de l'effet du type de bruit relativement à la grandeur de l'effet d'évolution du silence au bruit et relativement à la variabilité intra-individuelle de chaque locuteur. Les détails de ces différents résultats sont consultables en Annexe Res5.

3.1. Paramètres dont l'évolution dans le bruit ne dépend pas du type de bruit

Tout d'abord, plusieurs paramètres ne montrent pas de sensibilité significative au type de bruit dans leur évolution du silence au bruit. C'est le cas du **quotient ouvert** (excepté pour la locutrice L6 montrant un quotient ouvert significativement plus grand dans le bruit cocktail que dans le bruit blanc), du **premier formant** ainsi que de **l'étirement labial**.

3.2. Paramètres dont l'évolution dans le bruit montre la même dépendance au type de bruit pour tous les locuteurs

D'autres paramètres montrent une évolution dans le bruit significativement différente selon le type de bruit, avec une différence allant dans le même sens pour tous les locuteurs.

Ainsi, on observe que l'augmentation de **l'intensité vocale** dans le bruit est moins importante dans le bruit cocktail que dans le bruit blanc pour la moitié des locuteurs de la BD1 et pour L6 dans la BD2 (cf. Figure 62). Cette différence peut aller jusqu'à -6.9dB chez L9. L'augmentation de l'intensité vocale est sinon comparable pour les deux types de bruit chez les autres locuteurs. Ces résultats peuvent être rapprochés de ceux obtenus par Egan 1972^[80], observant une plus grande augmentation de l'intensité vocale dans le bruit blanc large bande et médium que dans le bruit blanc filtré basses-fréquence ou hautes-fréquences.

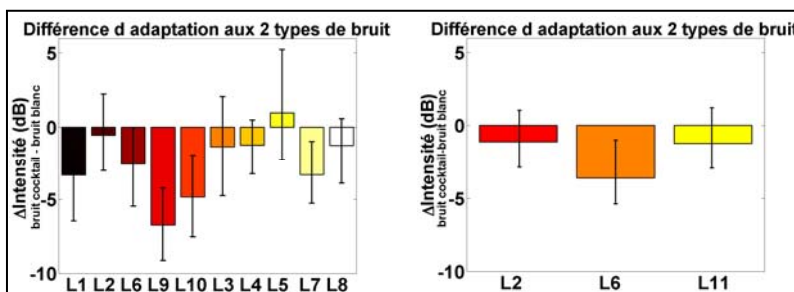


Figure 62. Représentation pour chaque locuteur et locutrice de la BD1 (à gauche) et de la BD2 (à droite), de la différence d'évolution de l'intensité avec le bruit entre un bruit de cocktail party et un bruit blanc.

De même, **la fréquence fondamentale** augmente moins dans le bruit cocktail que dans le bruit blanc pour la majorité des locuteurs de la BD1, et pour la locutrice L6 de la BD2 (cf. Figure 63). Cette différence peut atteindre jusqu'à 1.9 tons pour L7 dans la BD1 et 2 tons pour L6 dans la BD2. Au contraire, la locutrice L10 ne montre pas de différence entre les deux types de bruit, de même que L2 et L11 dans la BD2. Ces observations vont dans le même sens que Kadiri 1998^[176] observant une augmentation plus importante de la F0 dans le bruit blanc que dans un bruit de conversations.

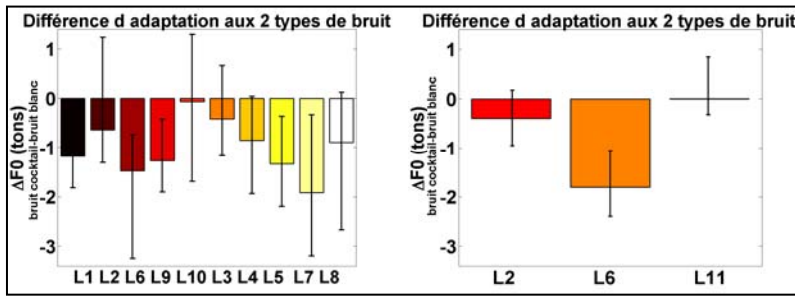


Figure 63. Représentation pour chaque locuteur et locutrice de la BD1 (à gauche) et de la BD2 (à droite), de la différence d'évolution de l'intensité avec le bruit entre un bruit de cocktail party et un bruit blanc.

De façon similaire, l'augmentation de **l'énergie spectrale dans les bandes [0-1kHz] et [1-2kHz]** est moins importante dans le bruit cocktail que dans le bruit blanc. Cette différence peut devenir significative chez certains locuteurs de la BD1 (L1, L6, L9, L10), mais reste négligeable pour les autres locuteurs de la BD1 ainsi que pour les locutrices de la BD2.

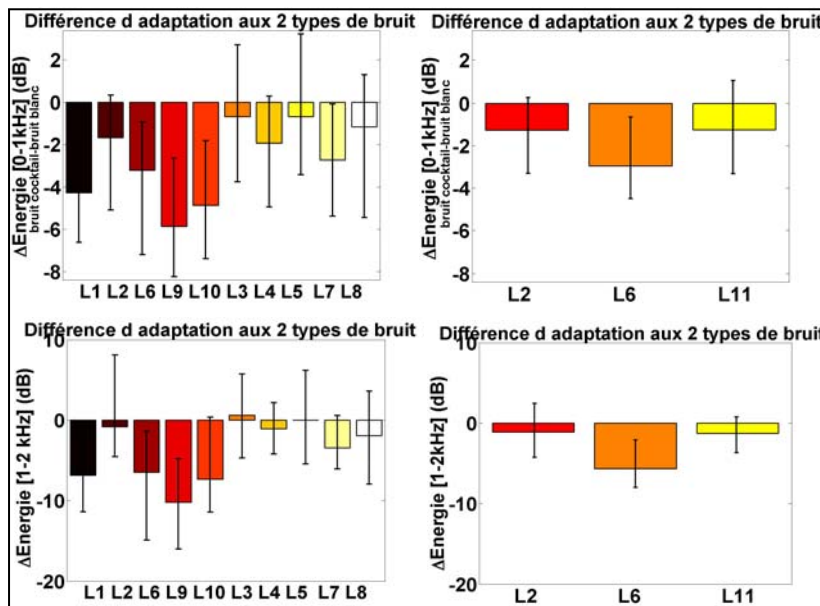


Figure 64. Représentation pour chaque locuteur et locutrice de la BD1 (à gauche) et de la BD2 (à droite), de la différence d'évolution avec le bruit de l'énergie spectrale entre 0 et 1kHz et entre 1 et 2 kHz, entre un bruit de cocktail party et un bruit blanc.

Cette adaptation moins importante dans le bruit cocktail que dans le bruit se traduit également au niveau de l'augmentation de **l'ouverture labiale** et de **l'aire intéro-labiale** dans le bruit sur les voyelles [a] et [i]. On note une différence d'évolution entre les deux types de bruit pouvant atteindre, chez la locutrice L6, -21% sur l'ouverture labiale des voyelles [a] et -27% sur l'aire intéro-labiale des voyelles [a] (cf. Figure 65). De même, la diminution de **la protrusion** dans le bruit est moins importante dans le bruit cocktail que dans le bruit blanc, avec une différence moyenne de 35%. Nos résultats confirment les observations de Junqua 1992^[173] notant une différence d'articulation dans un bruit blanc et un bruit de conversations. Ils vont au contraire à l'encontre de ceux de Davis et al. 2006^[62] rapportant une amplification des gestes articulatoires labiaux plus importante dans le bruit cocktail que dans le bruit blanc.

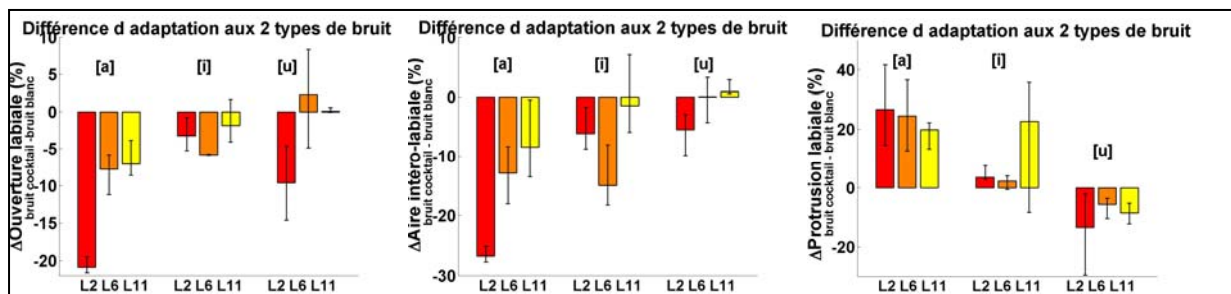


Figure 65. Représentation pour chaque locutrice de la BD2 de la différence d'évolution avec le bruit de l'amplitude des mouvements articulatoires d'ouverture et de protrusion, ainsi que de l'aire intéro-labiale, entre un bruit de cocktail party et un bruit blanc.

Au contraire, **l'élévation du larynx** dans le bruit est significativement plus importante dans le bruit cocktail que dans le bruit blanc, en moyenne de 8.8%, correspondant à la moitié de l'effet global d'élévation du larynx dans le bruit (cf. Figure 66).

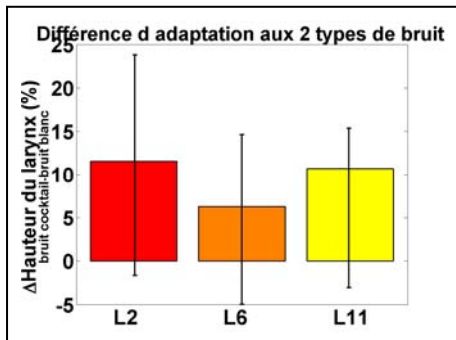


Figure 66. Représentation pour chaque locutrice de la BD2 de la différence d'évolution de l'intensité avec le bruit entre un bruit de cocktail party et un bruit blanc.

3.3. Paramètres dont l'évolution dans le bruit montre une dépendance au type de bruit différente selon les locuteurs

Pour finir, certains paramètres peuvent montrer chez certains locuteurs une différence d'évolution significative entre les deux types de bruit considérés, mais cette différence est de signe différent selon les locuteurs.

En particulier, les syllabes sont souvent moins allongées dans le bruit cocktail que dans le bruit blanc, avec des différences pouvant atteindre 45ms dans la BD1 et 34 ms dans la BD2 (pour L6 dans les deux cas). Au contraire, les syllabes sont plus allongées dans le bruit blanc que dans le bruit cocktail par L3, de 28ms en moyenne. Dans une étude antérieure, Kadiri 1998^[176] n'avait pas observé de différence de durée des mots entre une condition de bruit blanc et une condition de bruit de conversations.

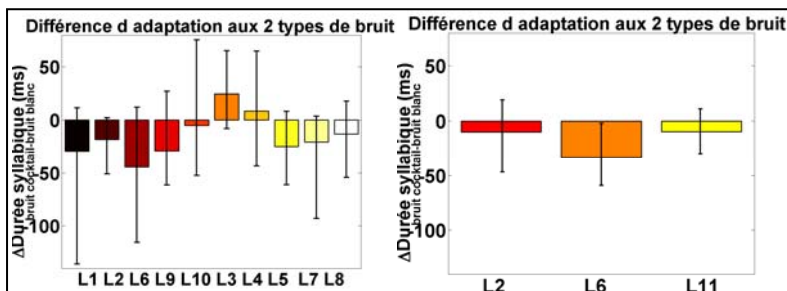


Figure 67. Représentation pour chaque locuteur et locutrice de la BD1 (à gauche) et de la BD2 (à droite), de la différence d'évolution avec le bruit de la durée syllabique entre un bruit de cocktail party et un bruit blanc

De même, l'augmentation de **l'énergie spectrale dans les bandes [2-4kHz] et [4-6kHz]** est moins importante dans le bruit cocktail que dans le bruit blanc - de la même façon que dans les basses fréquences - pour les locutrices L1, L6, L9 et L10 de BD1, ainsi que pour L6 dans la BD2. Au contraire, l'augmentation de l'énergie spectrale dans ces bandes de fréquences est plus importante dans le bruit cocktail que dans le bruit blanc pour les locuteurs L3, L4 et L5 bien que peu ou pas significative, ainsi que pour L2 dans la BD2, où la différence d'évolution entre les deux types de bruit atteint 7dB pour la bande de fréquences [4-6kHz].

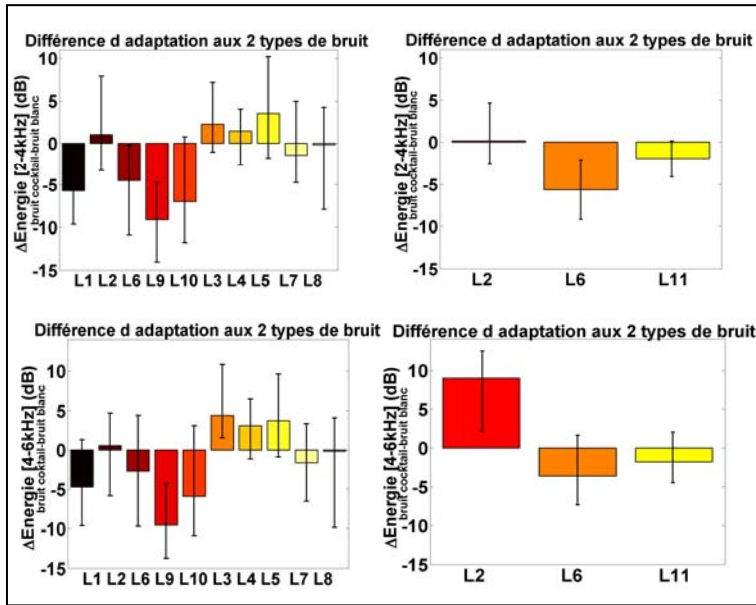


Figure 68. Représentation pour chaque locuteur et locutrice de la BD1 (à gauche) et de la BD2 (à droite), de la différence d'évolution avec le bruit de l'énergie spectrale entre 1 et 2 kHz et entre 4 et 6 kHz, entre un bruit de cocktail party et un bruit blanc.

4. CONCLUSION DU CHAPITRE 5

Nous avons recensé dans ce chapitre un certain nombre de modifications importantes de la parole produite dans le bruit observées chez la majorité des locuteurs: augmentation de l'intensité vocale, de la fréquence fondamentale, de l'énergie spectrale dans les fréquences médiums de la voix, ralentissement du débit de parole, augmentation du premier formant vocalique, augmentation de l'amplitude des mouvements articulatoires d'ouverture des lèvres, diminution du quotient ouvert et élévation verticale du larynx.

Ces modifications ont déjà été caractérisées de nombreuses fois par des études antérieures. Leur examen ici n'avait pas pour but de les mettre en évidence mais avant tout de comparer les données de nos corpus, constitués avec de la parole en Français et avec de nouvelles méthodologies (au niveau des tâches de parole : ici semi-spontanées et interactives, ainsi qu'au niveau de la méthode d'immersion du locuteur dans le bruit : ici par le biais de haut-parleurs). Nous avons donc retrouvé dans nos deux bases de données les mêmes tendances que dans les études antérieures, avec cependant des magnitudes d'évolution plus importantes que celles observées dans des études portant sur de la parole lue, et relativement comparables à celles observées dans des études portant sur de la parole spontanée. Au niveau des modifications formantiques, nous avons observé une élévation du F1 comme dans la plupart des études antérieures. Nos résultats, en ce qui concerne F2 et F3, vont dans le sens de certaines études antérieures et à l'encontre d'autres études. Ces différences entre les études peuvent provenir de particularités de chaque langue examinée. Enfin, contrairement à la plupart des études, nous n'avons pas observé de ralentissement significatif du débit de parole dans le bruit pour tous les locuteurs, mais seulement pour certains.

Par ailleurs, l'intérêt de ce chapitre était également d'aller au-delà des tendances globales pour tenter de distinguer les caractéristiques de l'adaptation dans le bruit communes à tous les locuteurs de celles montrant des différences de magnitude ou même de tendance entre les locuteurs. Nous avons ainsi observé que l'élévation de la fréquence fondamentale était assez peu variable entre les individus. Au contraire, les augmentations de l'intensité vocale, de l'énergie spectrale, du premier formant vocalique et de l'amplitude des mouvements articulatoires suivent la même tendance mais avec des différences interindividuelles non négligeables dans la magnitude de cette augmentation. Enfin, le quotient ouvert et la durée des syllabes montrent même des tendances d'évolution opposées entre les différents locuteurs.

Notre but dans cette thèse n'est pas de nous limiter à caractériser objectivement l'ensemble des modifications de la parole dans le bruit. Ces modifications ne nous intéressent pas tant en elle-même que par leur rôle d'indicateurs de stratégies de communication plus globales adoptées par les locuteurs pour adapter leur effort vocal à l'exigence d'intelligibilité d'un environnement bruyant.

Nous allons donc maintenant examiner plusieurs hypothèses de stratégies de communication pouvant être adoptées par les locuteurs en environnement bruyant. Nous envisagerons en particulier les potentielles stratégies de renforcement du rapport signal sur bruit (chapitre 6), de l'émergence de la voix par rapport au bruit ambiant (chapitre 6), des indices acoustiques et visuels de reconnaissance des phonèmes (chapitre 7), des indices de structuration du discours (chapitre 8) ou encore des mots de l'énoncé apportant davantage d'information que d'autres (chapitre 9). Nous explorerons en quoi les modifications globales de la parole observées dans ce chapitre et d'autres modifications plus spécifiques peuvent participer à ces stratégies, ou au contraire aller à leur rencontre.

Pour cela, nous distinguerons dans les chapitres suivants les descripteurs de la parole pouvant être directement perçus par l'interlocuteur et contribuer à l'intelligibilité de la parole (intensité, F0, répartition de l'énergie spectrale, durée syllabique, formants, amplitude des gestes articulatoires, etc.), des descripteurs pouvant également contribuer à l'intelligibilité, mais de façon indirecte, et caractérisant davantage la manière dont le locuteur gère sa production (quotient ouvert, mouvements verticaux du larynx, etc.)

Chapitre 6 :

Une adaptation visant à émerger du bruit ambiant ?

- 1. PRESERVATION D'UN RAPPORT SIGNAL SUR BRUIT ACCEPTABLE**
- 2. EXPANSION OU COMPRESSION DE LA DYNAMIQUE VOCALE**
- 3. ADAPTATION EN FREQUENCES**
- 4. CONCLUSION**

Tout au long de cette thèse, nous effectuons l'hypothèse que l'adaptation du locuteur en environnement bruyant est principalement motivée par la préservation de l'intelligibilité pour l'interlocuteur. Cela nous a amené à envisager l'existence potentielle de plusieurs stratégies de communication dans le bruit, visant à compenser la dégradation de l'intelligibilité induite par le bruit, par différentes façons.

En tout premier lieu, on peut considérer que le bruit perturbe l'intelligibilité du fait qu'il peut masquer la voix par sa puissance ou son contenu spectral. Aussi, il est envisageable que les locuteurs cherchent à compenser cette perturbation en renforçant les aspects acoustiques de leur voix leur permettant d'émerger perceptivement du bruit ambiant. La notion d'émergence sous-entend les idées de forme et de fond, la forme étant ici la parole du locuteur et le fond, le bruit ambiant. Une forme peut émerger d'un fond par cohérence d'une ou de plusieurs de ses caractéristiques, mais surtout par contraste de ces caractéristiques par rapport à celles du fond. Ainsi, de nombreuses études psychoacoustiques ont étudié les seuils de discrimination perceptive (en fréquence, en intensité, en timbre, etc.) au-delà desquels deux flux sonores entremêlés sont perçus comme distincts et en dessous desquels ils sont regroupés en un même flux auditif McCabe et al. 1997^[235]; Gaudrain et al. 2006^[108]). Ces phénomènes de fusion ou de séparation perceptive de flux auditifs sont au cœur de l'effet « cocktail-party », consistant à pouvoir suivre la conversation d'une personne au milieu d'autres conversations alentour (Cherry 1953^[49]; Arons 1992^[11]). Dans le cas de la parole produite en environnement bruyant, on peut donc envisager que celle-ci émerge du bruit ambiant de plusieurs façons : tout d'abord par un niveau d'intensité plus important que celui du bruit, par des caractéristiques dynamiques plus larges que celles du bruit, par une fréquence fondamentale très distincte du barycentre spectral du bruit ou encore par un timbre très distinct du contenu spectral du bruit. Nous allons examiner ces différentes hypothèses au cours de chapitre, en considérant les observations acoustiques ou articulatoires argumentant en faveur ou en défaveur de chacune d'entre elles. Pour examiner ces différentes hypothèses, nous avons exploré les deux bases de données, dont les détails sont présentés dans le paragraphe 3.1 du chapitre 3 et dans le paragraphe 2.1 du chapitre 4.

1. PRESERVATION D'UN RAPPORT SIGNAL SUR BRUIT ACCEPTABLE

1.1. Relations entre l'intensité vocale et l'intensité du bruit ambiant

La manière la plus simple de dégrader ou d'améliorer l'intelligibilité de la parole consiste à diminuer ou augmenter son rapport signal sur bruit, i.e. le rapport entre l'intensité vocale et l'intensité du bruit de fond. Ainsi, l'ajout de bruit additionnel à un enregistrement de parole est un paradigme classique de dégradation de l'intelligibilité pour réaliser des tests perceptifs (Sumbly et al. 1954^[335], Miller et al. 1955^[242], Neely 1956^[251], Benoit et al. 1994^[24]). Au contraire, Miller et al. 1951^[241] a montré qu'une augmentation du rapport signal sur bruit de 1 dB entraîne une amélioration de 5 à 10 % de l'intelligibilité. Pour ces mêmes raisons, Junqua 1993^[171] a ainsi avancé que les voix soufflées, montrant des composantes bruitées, seraient moins intelligibles que les autres. Javkin et al.^[166] et Henton et al. 1985^[143] ont cependant réfuté cette idée, argumentant que le bruit ambiant n'a rien de comparable avec le bruit d'une fuite glottique, et que l'aspect soufflé d'une voix n'est pas exempt d'information articulatoire puisqu'il est filtré par le conduit vocal.

Comme nous l'avons vu au chapitre précédent, l'augmentation de l'intensité vocale est une des caractéristiques principales de la voix produite en environnement bruyant, et a servi initialement à Lombard pour mettre en évidence l'effet du même nom. Cet auteur interprétait l'augmentation de l'intensité vocale comme une régulation de la voix par l'audition, en réaction à l'atténuation du retour auditif imposé par le bruit ambiant. Ce mécanisme réflexe participe sans aucun doute à l'augmentation de l'intensité vocale. Mais l'augmentation plus importante de l'intensité vocale en situation d'interaction avec un interlocuteur (cf chapitre 4) laisse également penser que cette augmentation de l'intensité est également liée à la recherche d'intelligibilité, et peut être interprétée en terme de compensation du rapport signal sur bruit dégradé pour l'interlocuteur. La Figure 69 représente ainsi le rapport signal sur bruit de chaque locuteur de la 1^{ère} base de données (BD1), en condition silencieuse (où le niveau d'intensité du fond sonore a été mesuré à 40dB) puis pour 4 niveaux d'intensité croissante du bruit ambiant (62, 70, 78 et 86 dB). Pour chaque locuteur, nous avons représenté le rapport signal sur bruit sans compensation, i.e. comme si les locuteurs

continuaient à s'exprimer dans le bruit de la même façon que dans le silence, et avec compensation, i.e. de la façon que nous avons observée expérimentalement. Dans le premier cas, le rapport signal sur bruit correspond donc à la différence d'intensité en dB entre la voix produite dans le silence et l'intensité des différents niveaux de bruit croissants. Dans le deuxième cas, il s'agit de la différence d'intensité en dB entre la voix produite dans ces bruits d'intensité croissante et l'intensité de ces bruits. On observe que l'augmentation de l'intensité vocale dans le bruit ne permet pas de compenser totalement la dégradation du rapport signal sur bruit, mais la limite malgré tout de façon efficace. Par ailleurs, on remarque que sans compensation, le rapport signal sur bruit serait négatif pour la majorité des locuteurs à partir d'un niveau de bruit supérieur à 78dB. Ce n'est pas le cas de la parole produite avec compensation, où le rapport signal sur bruit diminue certes avec le niveau d'intensité du bruit ambiant, mais reste dans tous les cas positif. A très fort niveau de bruit, les locuteurs arrivent à compenser la dégradation de 15dB en moyenne. Le rapport signal sur bruit reste alors encore de 5dB dans les pires cas (pour les locutrices L1 ou L2 par exemple), mais peut être maintenu jusqu'à 15-20dB par certains locuteurs (L6, L4, L5 ou L8).

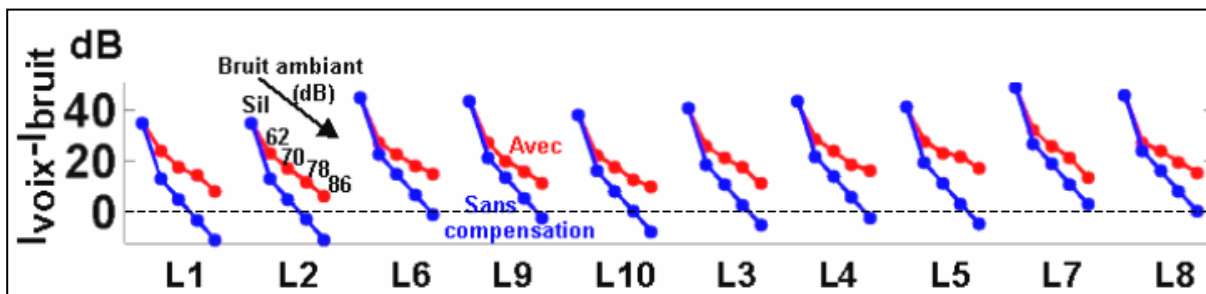


Figure 69 . Représentation de la diminution du rapport signal sur bruit (en dB) lorsque chaque locuteur de la 1^{ère} base de données (BD1) parle dans un bruit d'intensité croissante (calme \leftrightarrow 40dB, puis 62, 70, 78 et 86dB de bruit) en tentant de compenser cette perturbation (mesures effectuées sur la base de données (BD1) et représentées par la courbe rouge du haut). Nous avons représenté par la courbe bleue (du bas) le rapport signal sur bruit tel qu'il serait si le locuteur ne compensait pas et continuait à adopter la même façon de parler dans le bruit que dans le silence.

Par ailleurs, nous avons observé dans le chapitre précédent que l'augmentation de l'intensité vocale était pour la majorité des locuteurs moins importante dans le bruit cocktail que dans le bruit blanc. De prime abord, cette observation ne semble donc pas aller dans le sens d'une adaptation vocale dans le bruit visant à seulement préserver un rapport signal sur bruit acceptable. Pour autant, nous ne devons pas perdre de vue que la perturbation que le locuteur cherche à compenser n'est pas définie de façon « physique », mais de façon perceptive, par la dégradation de l'intelligibilité du point de vue de l'interlocuteur. Aussi, on voit bien sur la Figure 70 comment l'énergie acoustique du bruit cocktail est surtout concentrée dans les basses fréquences et décroît fortement à partir de 2kHz, tandis que celle du bruit blanc est répartie sur l'ensemble des fréquences inférieures à 10kHz. Comme la zone de sensibilité maximale de l'oreille se situe entre 2 et 4kHz, le bruit blanc est donc théoriquement perçu comme plus intense que le bruit cocktail, bien que leur intensité physique soit identique. Cette hypothèse reste à prouver perceptivement, mais pourrait assez bien expliquer la plus grande compensation de l'intensité vocale observée dans le bruit blanc que dans le bruit cocktail.

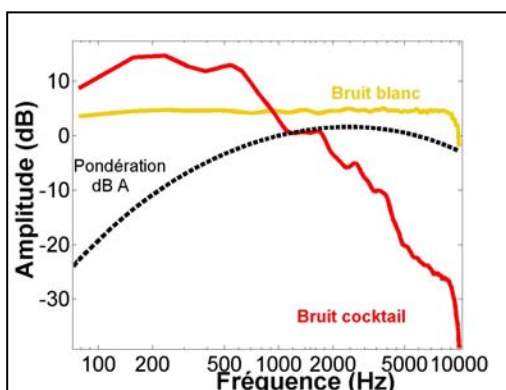


Figure 70 . Représentation des profils spectraux des deux types de bruits examinés au cours de la thèse, par comparaison à la courbe de pondération en dBA rendant compte de la sensibilité de la perception humaine dans les fréquences médiums et de l'atténuation de l'intensité perçue dans les basses et les hautes fréquences.

1.2. Des évolutions directement liées à l'augmentation de l'intensité vocale

Si nous considérons que la stratégie du locuteur pour communiquer dans le bruit consiste à augmenter son intensité vocale pour préserver un rapport signal sur bruit acceptable, comment expliquer les autres modifications de la parole observées dans le bruit, et présentées au chapitre précédent ? Nous allons examiner dans les paragraphes suivants si ces autres modifications principales, telles que l'élévation de la fréquence fondamentale, du premier formant vocalique, de l'amplification des mouvements d'ouverture des lèvres ou encore de l'aplatissement de la pente spectrale, ne sont pas finalement directement liées à cette augmentation de l'intensité vocale.

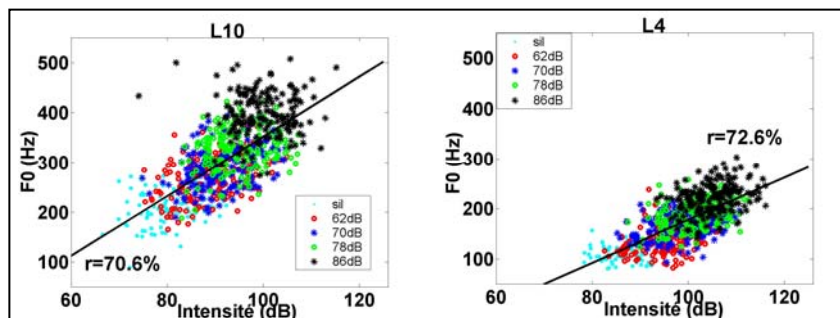
1.2.1. Augmentation conjointe de la fréquence fondamentale

Il est en effet très connu que l'intensité vocale et la fréquence fondamentale sont étroitement liées, même si elles peuvent être contrôlées séparément dans une certaine mesure, comme c'est le cas en particulier dans le chant.

Nous avons examiné cette potentielle corrélation entre l'intensité vocale et la fréquence fondamentale en traçant dans un même plan les valeurs d'intensité vocale en fonction des valeurs de fréquence fondamentale (cf. Figure 71). Ces valeurs ont été mesurées :

- dans la BD1, sur les 2 syllabes des 16 mots cibles prononcés dans chaque condition de bruit, et cela pour les 9 conditions (silence et 8 conditions de bruit à différents niveaux d'intensité). Cela représente donc au total 288 mesures par locuteur et locutrice.
- dans la BD2, sur la syllabe du déterminant [la] et sur les 2 syllabes des 17 logatomes cibles, prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 255 mesures par locutrice.

Figure 71. Exemples pour une locutrice (L10) et un locuteur (L4) de la corrélation entre leurs valeurs de F0 et d'intensité vocale en situation calme et bruyante. L'ensemble des figures pour tous les locuteurs et locutrices des deux bases de données sont consultables en Annexe Res6.



On observe effectivement une évolution conjointe et proportionnelle de la fréquence fondamentale et de l'intensité vocale avec un niveau de perturbation croissant du bruit ambiant. On remarque toutefois chez L2, L6 et L5 une sorte de plafond dans l'évolution de l'intensité vocale à très fort niveau de bruit tandis que la fréquence fondamentale continue à augmenter. Nous avons estimé la corrélation entre la fréquence fondamentale et l'intensité vocale à partir du coefficient de régression linéaire de ces données. Celui-ci est rapporté pour chaque locuteur et locutrice dans la colonne « En Absolu » du Tableau 17. Excepté pour le locuteur L8, la fréquence fondamentale est donc fortement corrélée dans l'absolu à l'intensité vocale, conformément aux observations de la littérature.

	En Absolu	En Evolution
L1	71.1 %	68.1 %
L2	59.1 %	28.8 %
L6	61.1 %	75.0 %
L9	66.3 %	69.0 %
L10	70.6 %	72.8 %
L3	64.9 %	4.0 %
L4	72.6 %	70.6 %
L5	64.2 %	68.2 %
L7	68.9 %	71.5 %
L8	52.5 %	64.6 %

BD1

	En Absolu	En Evolution
L2	89.8 %	38.4 %
L6	85.7 %	67.6 %
L11	77.2 %	60.2 %

BD2

Tableau 17. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs de fréquence fondamentale et celles d'intensité vocale, toutes conditions confondues de BD1 (à gauche) et de BD2 (à droite). La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit de la F0 et les valeurs d'évolution de l'intensité vocale.

Nous avons souhaité aller plus loin et examiner si l'évolution de la fréquence fondamentale du silence au bruit était toujours corrélée à l'évolution de l'intensité vocale. Nous avons donc cette fois tracé dans un même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence. L'observation de ces figures (cf. Annexe Res6) nous amène à plusieurs constatations :

- tout d'abord, l'intensité et la fréquence fondamentale augmentent conjointement dans le bruit pour la totalité des syllabes observées de la BD2, et pour toutes les locutrices. Cela est également valable pour la grande majorité des syllabes observées de la BD1. Il existe cependant quelques cas (à faible niveau de bruit pour la plupart) où la fréquence fondamentale diminue dans le bruit alors que l'intensité vocale augmente (cf. Tableau 18). L5 présente jusqu'à 12.5% de cas de ce type.
- par ailleurs, nous avons estimé la corrélation entre l'évolution de la fréquence fondamentale avec le bruit et celle de l'intensité vocale en calculant le coefficient de régression linéaire de ces données, rapporté pour chaque locuteur et locutrice dans la colonne « En Evolution » du Tableau 17. Excepté pour la locutrice L2, on observe effectivement que l'évolution de la fréquence fondamentale est fortement corrélée à celle de l'intensité vocale dans le bruit. La faible corrélation observée pour cette locutrice semble provenir du fait qu'elle ne s'adapte pas de façon très différente à faible ou fort niveau de bruit. Par conséquent, le nuage de point représentant l'évolution de sa F0 et de son intensité dans le bruit est concentré dans le plan $\Delta F0 * \Delta \text{Intensité}$, au lieu de former une ellipse autour d'un axe principal comme pour les autres locuteurs.

	F0 diminue	F0 augmente
L1	3.5 %	95.9 %
L2	5.4 %	93.1 %
L6	7.5 %	91.5 %
L9	3.7 %	94.0 %
L10	4.1 %	95.3 %
L3	2.2 %	96.5 %
L4	3.3 %	96.1 %
L5	12.5 %	85.5 %
L7	6.8 %	92.2 %
L8	7.9 %	89.0 %

Tableau 18. Pourcentage de cas où la fréquence fondamentale augmente ou diminue conjointement à une augmentation de l'intensité vocale, pour chaque locuteur et locutrice de la 1^{ère} base de données (BD1). La somme des deux colonnes ne correspond pas toujours à 100%. Il existe en effet quelques rares cas où l'intensité vocale diminue avec le bruit. L'effectif de ces cas correspond au pourcentage restant pour atteindre 100%.

L'observation d'une forte corrélation entre l'évolution de la fréquence fondamentale et celle de l'intensité vocale dans le bruit ne nous permet pas d'aller au-delà de la conclusion que ces deux évolutions font partie d'une même stratégie d'adaptation pour la majorité des locuteurs. Il est possible que l'augmentation de la fréquence fondamentale soit simplement la conséquence d'une intention du locuteur d'augmenter son intensité vocale, mais nos données ne permettent pas de le démontrer. Qu'elle soit ou non dépendante de l'intensité vocale, il n'est pas exclu que l'augmentation de la fréquence fondamentale contribue en soi à la préservation de l'intelligibilité dans le bruit (cf. § 2.1).

1.2.2. Augmentation conjointe de l'ouverture des lèvres

Dans le même esprit, l'augmentation de l'ouverture de la mâchoire accompagne très souvent l'augmentation de l'intensité vocale, sans que l'on sache pour autant si l'ouverture de la mâchoire facilite la phonation à forte intensité ou si, comme Schulman 1989^[304] le propose, elle en soit une conséquence, liée à la compensation des bruits de turbulence qui pourraient être créés si le débit d'air expiré augmentait face à une constriction constante du conduit vocal. Toujours est-il que cet auteur observe que l'ouverture labiale peut être modélisée de façon linéaire en fonction de l'intensité vocale, avec cependant une relation différente selon les catégories vocaliques.

Nous avons donc examiné cette potentielle corrélation entre l'intensité vocale et l'ouverture de la mâchoire, ici approximée à partir de l'ouverture labiale sur des voyelles [a]. Pour cela, nous avons tracé dans un même plan les valeurs d'intensité vocale en fonction des valeurs de l'ouverture labiale maximale (cf. Figure 72). Ces valeurs ont été mesurées dans la BD2 sur la syllabe du déterminant [la] et sur les 1^{ères} ou 2^{èmes} syllabes [la] des 17 logatomes cibles prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en

position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 175 mesures par locutrice.

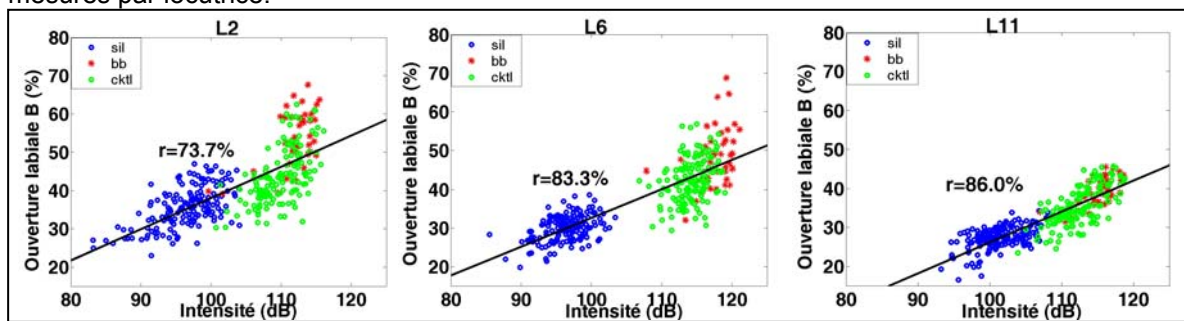


Figure 72. Représentation, pour les 3 locutrices de BD2, de leur intensité vocale en fonction de leur ouverture labiale, dans le silence (sil), dans le bruit blanc (bb) et dans le bruit de cocktail-party (cktl).

On observe effectivement une évolution conjointe et proportionnelle de l'ouverture labiale et de l'intensité vocale avec l'immersion du locuteur dans les deux types de bruit. Nous avons estimé la corrélation entre l'ouverture labiale et l'intensité vocale à partir du coefficient de régression linéaire de ces données. Celui-ci est rapporté pour chaque locutrice dans la colonne « En Absolu » du Tableau 19 et montre une forte corrélation entre l'ouverture labiale des voyelles [a] et leur intensité, pour toutes les locutrices de la BD2.

	[a]	
	En Absolu	En Evolution
L2	73.7 %	62.8 %
L6	83.3 %	53.8 %
L11	86.0 %	60.8 %

Tableau 19. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs d'ouverture labiale et celles d'intensité vocale, toutes conditions confondues. La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit de l'ouverture labiale et les valeurs d'évolution de l'intensité vocale.

De même que pour la fréquence fondamentale, nous avons examiné si l'évolution de l'ouverture labiale du silence au bruit était également corrélée à l'évolution de l'intensité vocale dans le bruit. Nous avons donc cette fois tracé dans un même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence (cf. Figure 73). Nous remarquons tout d'abord que l'ouverture labiale et l'intensité vocale augmentent conjointement dans le bruit dans tous les cas et pour les 3 locutrices considérées. Par ailleurs, nous avons estimé la corrélation entre leur évolution en calculant le coefficient de régression linéaire de ces données, rapporté pour chaque locutrice dans la colonne « En Evolution » du

Tableau 19. Les 3 locutrices montrent une corrélation supérieure à 50% entre ces 2 paramètres, mais qui n'excède jamais les 63%. Ce résultat peut être le signe que l'augmentation de l'ouverture labiale dans le bruit est effectivement en partie liée à celle de l'intensité vocale, mais que d'autres facteurs participent également à cette adaptation articulatoire. Il est également possible que les corrélations ne puissent pas être très importantes du fait que nous n'avons qu'un seul niveau de bruit dans la BD2 et par conséquent des évolutions de l'ouverture labiale et de l'intensité vocale peu dispersées dans le plan $\Delta B^* \Delta \text{Intensité}$.

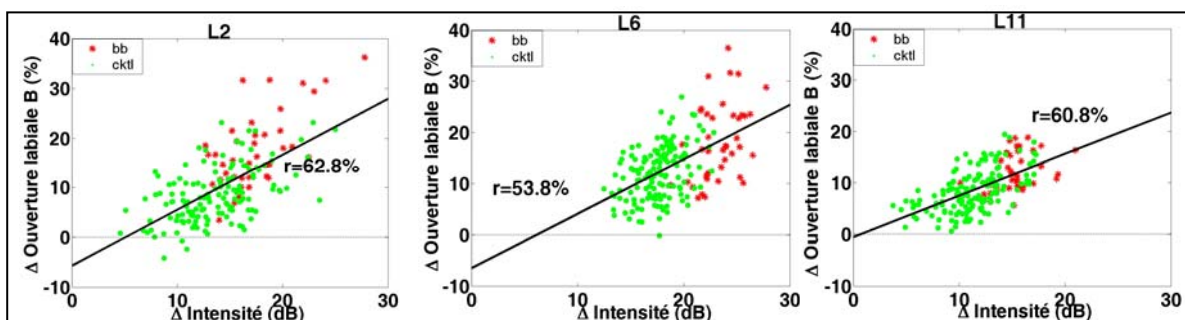


Figure 73. Représentation, pour les 3 locutrices de BD2, de l'évolution de leur intensité vocale du silence au bruit en fonction de l'évolution de leur ouverture labiale, pour une situation de bruit blanc (bb) et de bruit de cocktail-party (cktl).

1.2.3. Augmentation conjointe du premier formant vocalique

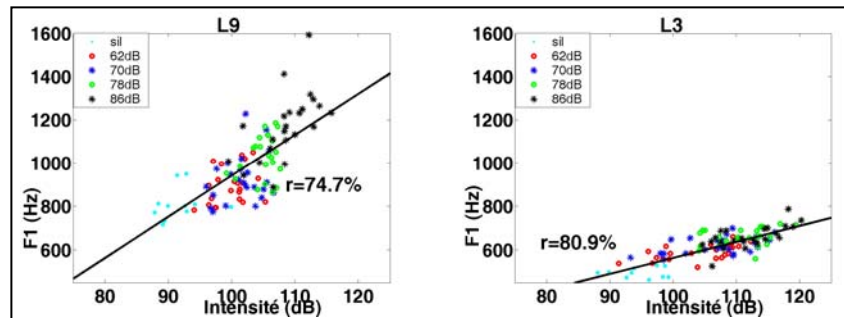
Il est également très classique d'observer une corrélation entre l'intensité vocale et le premier formant vocalique. Nous avons procédé de la même façon que dans les deux paragraphes précédents en traçant dans le même plan les valeurs du premier formant vocalique en fonction des valeurs d'intensité vocale (cf. Figure 74). Ces valeurs ont été mesurées

-dans la BD1, sur les syllabes des 16 mots cibles contenant un [a], un [i] ou un [u], prononcés dans chaque condition de bruit, et cela pour les 9 conditions (silence et 8 conditions de bruit à différents niveaux d'intensité). Cela représente donc au total 108 mesures par locuteur et locutrice pour les voyelles [a], 135 pour les [i] et 54 pour les [u].

-dans la BD2, sur la syllabe du déterminant [la] et sur les 1ères ou 2èmes syllabes [la] des 17 logatomes cibles prononcés en position initiale et finale de l'énoncé pour la condition silencieuse et la condition de bruit cocktail, et seulement en position finale pour la condition de bruit blanc. Cela représente donc au total 255 mesures par locutrice pour les voyelles [a]. Nous n'avons pas examiné les corrélations pour les autres voyelles [i] et [u], faute d'effectif suffisamment important de ces voyelles.

Les coefficients de régression linéaire de ces données sont rapportés dans la colonne « En Absolu » du Tableau 20. Nous avons également tracé dans le même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence. Les coefficients de régression linéaire de ces données sont également rapportés dans le Tableau 20, cette fois dans la colonne « En Evolution ».

Figure 74. Exemples pour une locutrice (L9) et un locuteur (L3) de la corrélation entre leurs valeurs de F1 et d'intensité vocale pour des voyelles [a] en situation calme et bruyante. L'ensemble des figures pour tous les locuteurs et locutrices des deux bases de données sont consultables en Annexe Res6.



	[a]		[i]		[u]		[a]		
	En Absolu	En Evolution	En Absolu	En Evolution	En Absolu	En Evolution	En Absolu	En Evolution	
L1	70.1 %	54.6 %	87.0 %	81.7 %	81.7 %	55.3 %	L2	64.4 %	-42.7 %
L2	65.4 %	6.4 %	83.6 %	86.4 %	86.4 %	58.8 %	L6	76.0 %	36.0 %
L6	55.3 %	72.2 %	66.4 %	67.2 %	67.2 %	71.9 %	L11	41.6 %	25.4 %
L9	74.7 %	55.6 %	75.6 %	79.0 %	79.0 %	31.6 %	BD2		
L10	66.6 %	43.7 %	84.6 %	84.0 %	84.0 %	68.3 %			
L3	80.9 %	75.0 %	27.5 %	9.4 %	9.4 %	26.2 %			
L4	72.2 %	58.9 %	7.3 %	17.7 %	17.7 %	46.2 %			
L5	72.4 %	79.1 %	33.1 %	42.5 %	42.5 %	23.9 %			
L7	71.2 %	63.9 %	35.3 %	59.3 %	59.3 %	34.0 %			
L8	47.5 %	35.1 %	15.8 %	10.8 %	10.8 %	13.7 %			

BD1

Tableau 20. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs du premier formant vocalique des voyelles [a], [i] et [u] et celles de leur intensité vocale, toutes conditions confondues de la BD1 (à gauche) et de la BD2 (à droite). La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit du premier formant vocalique et les valeurs d'évolution de l'intensité vocale.

L'examen de ces figures et des coefficients de régression linéaire montre effectivement une forte corrélation sur les voyelles [a] entre le premier formant et l'intensité vocale, excepté pour le locuteur L8 de la BD1 et la locutrice L2 de la BD2. Cependant les évolutions de ces 2 paramètres du silence au bruit ne sont pas toujours aussi bien corrélées, en particulier dans la BD2 où les faibles corrélations obtenues sont toujours difficiles à interpréter, du fait qu'elles peuvent provenir de l'unicité du niveau de bruit examiné et par conséquent d'un manque de dispersion des données dans l'espace $\Delta F1 \cdot \Delta \text{Intensité}$. L'examen plus global des tendances d'évolution dans le bruit montre pour la locutrice L11 un nombre de cas non négligeable où le premier formant diminue alors que l'intensité vocale augmente (cf. Tableau 21), ce qui peut expliquer la faible corrélation observée

entre les évolutions de ces deux paramètres dans le bruit. Au contraire, les deux autres locutrices de la BD2 le locuteur L8 de la BD1 voient leur premier formant augmenter dans le bruit de façon conjointe à l'intensité vocale, pour la quasi-totalité de leurs voyelles [a].

Pour les voyelles [i] et [u], le premier formant est également très corrélé à l'intensité vocale chez les femmes, mais peu chez les hommes. Cela se retrouve dans les tendances globales d'évolution où le premier formant des [i] et des [u] augmente de façon conjointe avec l'intensité vocale pour la grande majorité des syllabes produites par les femmes, quand au contraire il existe de nombreux cas chez les hommes où le premier formant des voyelles [i] et [u] diminue conjointement à l'augmentation de l'intensité vocale.

	[a]		[i]		[u]	
	F1 diminue	F1 augmente	F1 diminue	F1 augmente	F1 diminue	F1 augmente
L1	12.6 %	87.4 %	1.8 %	93.6 %	10.3 %	89.7 %
L2	4.6 %	95.4 %	5.6 %	93.5 %	10.6 %	89.4 %
L6	2.3 %	97.7 %	8.4 %	88.2 %	6.3 %	91.7 %
L9	2.4 %	97.6 %	4.7 %	92.5 %	2.7 %	91.9 %
L10	12.0 %	87.0 %	2.7 %	94.5 %	9.1 %	90.9 %
L3	0.0 %	97.5 %	36.1 %	61.4 %	48.9 %	51.1 %
L4	6.4 %	93.6 %	38.9 %	60.0 %	45.7 %	54.3 %
L5	1.1 %	97.7 %	14.4 %	83.9 %	43.8 %	56.3 %
L7	7.6 %	92.4 %	21.1 %	78.1 %	40.4 %	59.6 %
L8	4.2 %	94.7 %	32.8 %	61.3 %	52.1 %	35.4 %

BD1

	[a]	
	F1 diminue	F1 augmente
L2	0.0 %	100.0 %
L6	6.5 %	93.5 %
L11	15.0 %	85.0 %

BD2

Tableau 21. Pourcentage de cas où le premier formant vocalique augmente ou diminue conjointement à une augmentation de l'intensité vocale dans le bruit, pour chaque locuteur et locutrice de la 1^{ère} base de données (à gauche) et de la 2^{ème} base de données (à droite). La somme des deux colonnes ne correspond pas toujours à 100%. Il existe en effet quelques rares cas où l'intensité vocale diminue avec le bruit. L'effectif de ces cas correspond au pourcentage restant pour atteindre 100%.

La relation entre le premier formant vocalique et l'intensité vocale ne semble pas aussi nette que pour la fréquence fondamentale ou l'ouverture des lèvres.

- Pour les femmes, l'évolution du premier formant dans le bruit est assez liée à celle de l'intensité vocale sans pour autant que l'augmentation de l'intensité vocale explique totalement celle du 1^{er} formant. En effet, le premier formant est dans la littérature également corrélé à la fréquence fondamentale, à l'ouverture de la mâchoire et au quotient ouvert laryngé. Ces paramètres sont tous intimement liés à l'intensité vocale, tout en étant théoriquement contrôlables de façon relativement indépendante. Il est donc difficile de savoir si ces derniers paramètres peuvent contribuer à l'augmentation du premier formant de façon complémentaire à l'augmentation de l'intensité vocale ou uniquement de façon redondante. En dehors des liens très étroits qu'entretiennent ces différents paramètres acoustiques, glottiques et articulatoires avec l'intensité vocale, il est également envisageable qu'il existe une autre stratégie d'augmentation du 1^{er} formant dans le bruit, indépendante de l'augmentation de l'intensité vocale, mais liée au renforcement des indices de reconnaissance des voyelles (cf. chapitre 7).
- Pour les hommes, le premier formant ne semble corrélé à l'intensité vocale que dans le cas des voyelles [a] et pas pour les autres voyelles [i] et [u].

Cette différence entre les deux genres ainsi qu'entre les différents types de voyelles nous laisse penser qu'il puisse exister une relation entre la fréquence fondamentale (plus haute chez les femmes que chez les hommes) et le premier formant (plus haut sur les [a] que sur les [i] ou les [u]), comme Lienard et al. 1999^[215] et Traunmüller 1981^[355], entre autres, l'ont déjà remarqué.

Nous avons alors exploré ce lien en traçant dans un même plan les valeurs du premier formant vocalique en fonction des valeurs de fréquence fondamentale. Le but n'est pas ici d'explorer leur corrélation -très semblable à celle observée entre le premier formant et l'intensité vocale- mais plutôt d'examiner si le premier formant « accroche » un harmonique de la voix, autrement dit une fréquence multiple de la fréquence fondamentale.

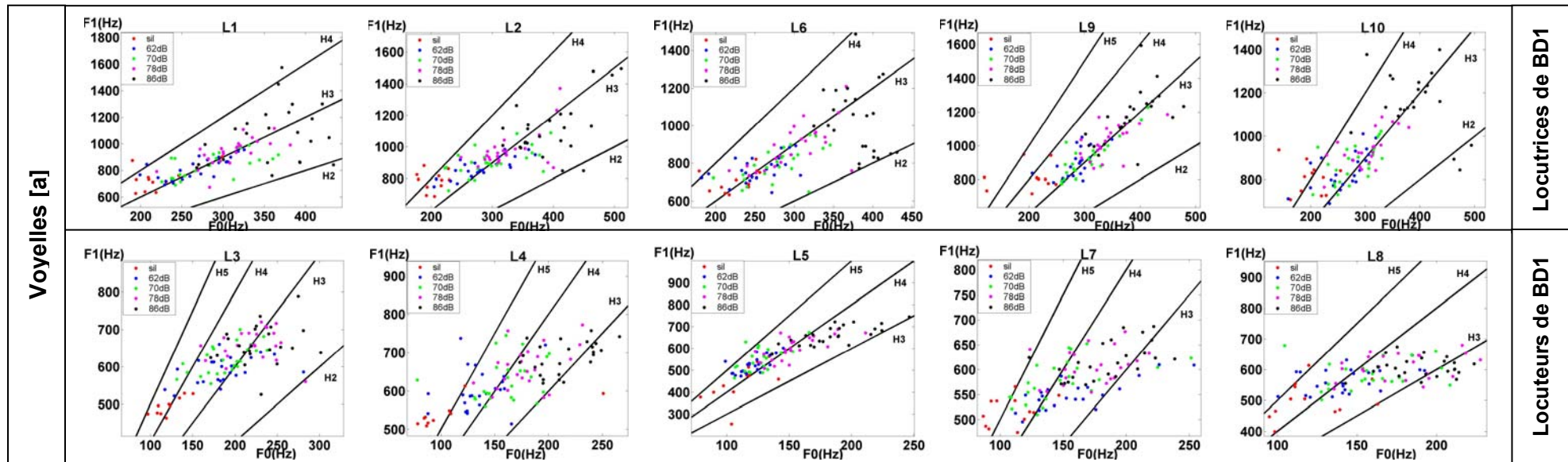


Figure 75. Evolution du premier formant vocalique sur des voyelles [a] en fonction de la fréquence fondamentale, pour chaque locutrice (en haut) et locuteur (en bas) de la 1^{ère} base de données (BD1) s'exprimant en condition silencieuse puis dans 4 niveaux de bruit croissants (62, 70, 78 et 86 dB). Les droites H2 à H5 représentent les harmoniques 2 à 5 de la voix pour une fréquence fondamentale croissante.

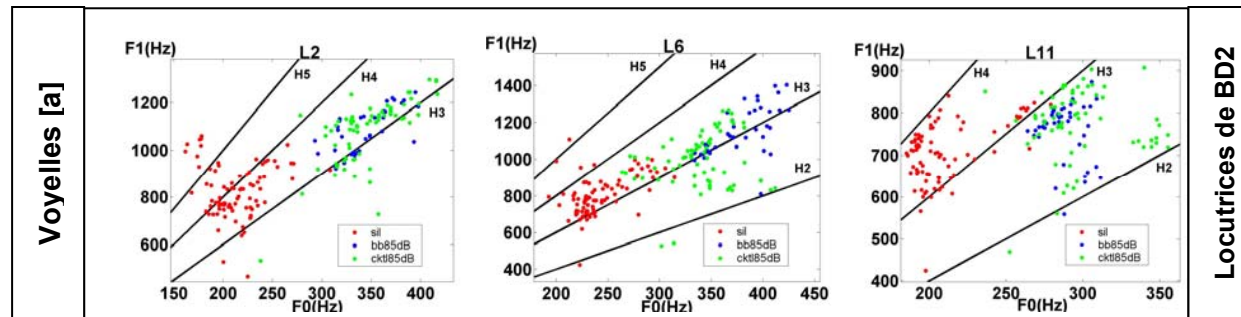


Figure 76. Evolution du premier formant vocalique sur des voyelles [a] en fonction de la fréquence fondamentale, pour chaque locutrice de la 2^{ème} base de données (BD2) s'exprimant en condition silencieuse puis dans 2 types de bruit à 85dB. Les droites H2 à H5 représentent les harmoniques 2 à 5 de la voix pour une fréquence fondamentale croissante.

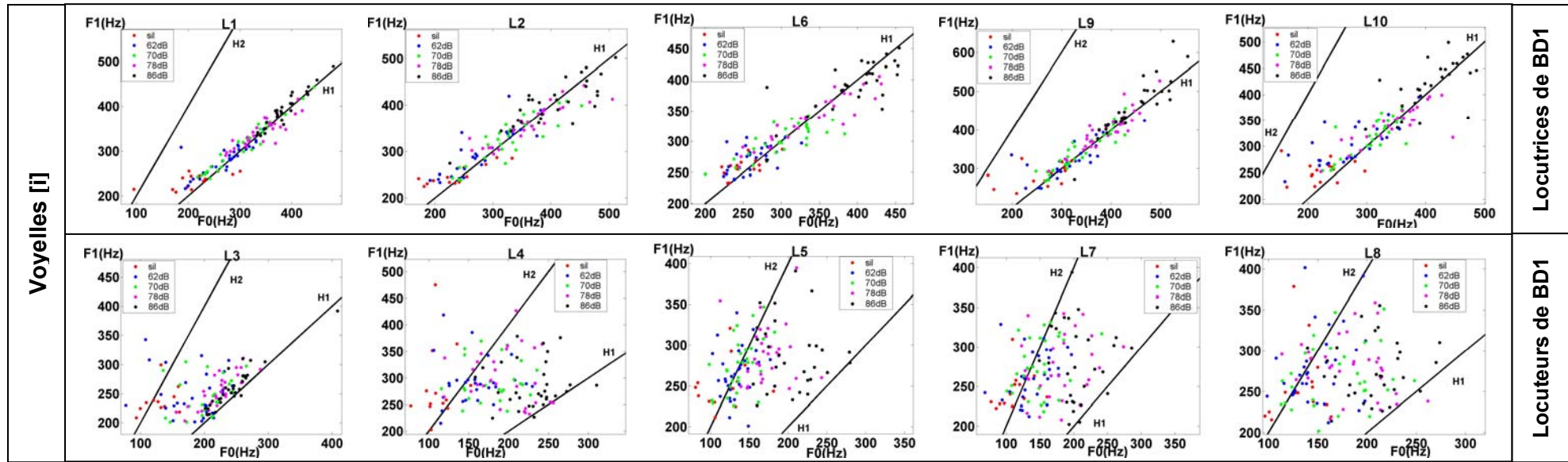


Figure 77. Evolution du premier formant vocalique sur des voyelles [i] en fonction de la fréquence fondamentale, pour chaque locutrice (en haut) et locuteur (en bas) de la 1^{ère} base de données (BD1) s'exprimant en condition silencieuse puis dans 4 niveaux de bruit croissants (62, 70, 78 et 86 dB). Les droites H2 à H5 représentent les harmoniques 2 à 5 de la voix pour une fréquence fondamentale croissante.

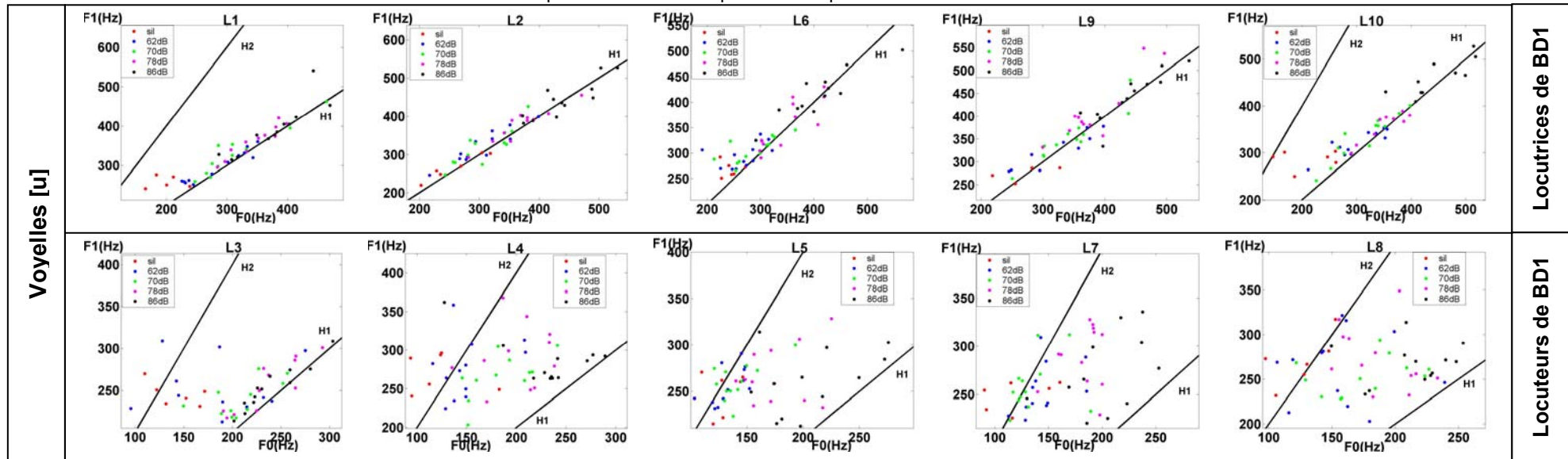


Figure 78. Evolution du premier formant vocalique sur des voyelles [u] en fonction de la fréquence fondamentale, pour chaque locutrice (en haut) et locuteur (en bas) de la 1^{ère} base de données (BD1) s'exprimant en condition silencieuse puis dans 4 niveaux de bruit croissants (62, 70, 78 et 86 dB). Les droites H2 à H5 représentent les harmoniques 2 à 5 de la voix pour une fréquence fondamentale croissante.

Sur les figures précédentes, on observe effectivement un ajustement du premier formant chez les femmes, surtout pour les forts niveaux de bruit, sur le 3^{ème} harmonique de la voix dans le cas des voyelles [a] et sur la fréquence fondamentale elle-même (1^{er} harmonique) dans le cas des voyelles [i] et [u]. Dans la 2^{ème} base de données, on voit plus nettement que le premier formant « accroche » le 3^{ème} harmonique ou le 4^{ème} harmonique dans le silence, et qu'il se situe plutôt sur le 3^{ème} harmonique à fort niveau de bruit, voire dans quelques cas sur le 2^{ème} harmonique. Ces observations permettent de comprendre pourquoi la corrélation entre le 1^{er} formant et l'intensité vocale n'est pas très importante, puisque le premier formant n'évolue pas de façon linéaire avec l'effort vocal mais par « saut » vers l'harmonique de rang inférieur lorsque la fréquence fondamentale dépasse un certain seuil.

Au contraire, on n'observe pas chez les hommes d'ajustement du 1^{er} formant sur un harmonique de la voix, excepté pour les [a] produits par L5 où le 1^{er} formant s'accroche également au 3^{ème} harmonique et pour les [i] de L3 où le 1^{er} formant accroche le fondamental. On observe cependant que le 1^{er} formant suit assez bien le 2^{ème} harmonique pour les [i] de L4, L5, L7 et L8, mais de façon assez dispersée. De même, on remarque une tendance du 1^{er} formant à suivre le fondamental sur les [u] produits par L3 et le deuxième harmonique sur les [u] produits par L5 et L7, mais toujours de façon relativement dispersée.

Ces résultats peuvent être interprétés de deux façons :

- en premier lieu, l'accrochage du premier formant sur un harmonique de la voix chez les locutrices peut être l'indicateur d'une stratégie d'augmentation de l'intensité vocale par « formant tuning » (Miller et al. 1990 ^[240] ; Carlsson et al. 1992 ^[43]). Cette technique résonnante a pour l'instant principalement été observée dans le cadre de certains types de chant (Henrich et al. 2006 ^[142] ; Joliveau et al. 2004 ^[168]) où elle permet d'augmenter l'intensité de la voix par le renforcement d'un de ses premiers harmoniques, et non pas seulement par l'augmentation de la pression sous-glottique ou de l'accolement des cordes vocales. L'utilisation de ce « formant tuning » par les locutrices dans le cadre de la communication parlée en environnement bruyant serait donc extrêmement intéressante puisqu'il s'agit d'une gestion efficace de l'effort vocal, en le partageant entre le larynx et les articulateurs.
- cependant, cet accrochage peut être un artéfact de la méthode d'estimation formantique par LPC. Le fait qu'il soit davantage observé chez les femmes, à la F0 plus élevée que les hommes, argumente dans ce sens

Afin d'éliminer le doute que cet accrochage du premier formant sur un harmonique de la voix soit dû à la méthode d'estimation formantique, il serait utile de réaliser des mesures physiques de résonances du conduit vocal chez des locuteurs communiquant dans le bruit, afin d'examiner si ces résonances coïncident effectivement ou non avec un harmonique de la voix.

1.2.4. Renforcement conjoint de l'énergie acoustique dans les médiums

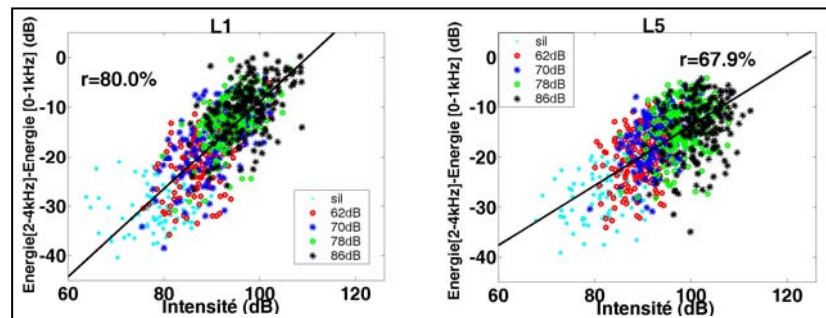
Enfin, il existe également très souvent une corrélation entre l'intensité vocale et la richesse spectrale de la voix dans les médiums et les hautes fréquences. En effet, l'augmentation de l'intensité vocale s'accompagne d'une augmentation de la force d'accolement des cordes vocales et souvent d'une fermeture glottique plus abrupte. Cela a pour conséquence un déplacement du formant glottique (cf. Annexe Met1) vers les fréquences aiguës et par conséquent un aplatissement de la pente spectrale à la fois basses-fréquences (en dessous de 3kHz) et hautes-fréquences entre 3 et 8kHz).

Nous avons donc voulu ici explorer cette potentielle corrélation afin de voir si le renforcement du spectre de la voix dans les médiums était directement lié à l'augmentation de l'intensité vocale ou s'il pouvait constituer une stratégie en soi d'émergence spectrale par rapport au bruit ambiant (cf. § 3.2). Pour cela, nous avons tracé dans le même plan les valeurs de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz] en fonction des valeurs d'intensité vocale (cf. Figure 79). Ces valeurs ont été mesurées :

- dans la BD1 sur les 2 syllabes des 16 mots cibles prononcés dans chaque condition de bruit, et cela pour les 9 conditions (silence et 8 conditions de bruit à différents niveaux d'intensité). Cela représente donc au total 288 mesures par locuteur et locutrice.
- dans la BD2 sur la syllabe du déterminant [la] et sur les 2 syllabes des 17 logatomes cibles prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 255 mesures par locutrice.

Les coefficients de régression linéaire de ces données sont rapportés dans la colonne « En Absolu » du Tableau 22. Nous avons également tracé dans le même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence. Les coefficients de régression linéaire de ces données sont également rapportés dans le Tableau 22, cette fois dans la colonne « En Evolution ».

Figure 79. Exemples pour une locutrice (L1) et un locuteur (L5) de la corrélation entre leur intensité vocale et le timbrage de leur voix entre 2 et 4kHz par rapport à la zone [0-1kHz], en situation calme et bruyante. L'ensemble des figures pour tous les locuteurs et locutrices des deux bases de données sont consultables en Annexe Res6.



	En Absolu	En Evolution
L1	80.0 %	36.2 %
L2	58.1 %	-12.0 %
L6	48.8 %	29.5 %
L9	65.2 %	33.2 %
L10	55.3 %	19.9 %
L3	41.0 %	11.9 %
L4	47.1 %	15.7 %
L5	67.9 %	56.8 %
L7	60.5 %	45.3 %
L8	37.9 %	26.8 %

BD1

	En Absolu	En Evolution
L2	68.2 %	-35.3 %
L6	72.2 %	37.3 %
L11	83.4 %	-0.5 %

BD2

Tableau 22. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs de pente spectrale basses fréquences et celles d'intensité vocale, toutes conditions confondues de BD1 (à gauche) et de BD2 (à droite). La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit de la pente spectrale BF et les valeurs d'évolution de l'intensité vocale.

L'examen de ces figures et des coefficients de régression linéaire montre une forte corrélation entre l'intensité vocale et la pente spectrale pour certains locuteurs seulement (L1 dans la BD2, L2 et L11 dans la BD2). D'autres locuteurs présentent un taux de corrélation entre 50 et 70% (L2, L9, L10, L5 et L7 dans la BD1 et L2 dans la BD2). Enfin, 4 locuteurs de la BD1 montrent une faible corrélation entre l'intensité vocale et la pente spectrale (L6, L3, L4 et L8). On remarque au passage que L6 n'a pas le même comportement entre les deux bases de données.

L'évolution de la pente spectrale dans le bruit est en revanche très peu corrélée à l'augmentation de l'intensité vocale dans le bruit, pour les deux bases de données. Pourtant, lorsqu'on examine les profils globaux d'évolution dans le bruit de l'intensité vocale et de la pente spectrale, on observe que la pente spectrale diminue (⇔ s'aplatit) conjointement à l'augmentation de l'intensité vocale dans la quasi-totalité des cas de la BD1 et de la BD2.

Il semble donc que le renforcement du spectre de la voix dans les médiums accompagne l'augmentation de l'intensité vocale dans le bruit, sans pour autant qu'il s'agisse d'une même stratégie. Nous avons effectivement vu dans le chapitre 5 l'existence d'une différence de renforcement spectral de la voix en fonction du type de bruit. Il est donc tout à fait envisageable qu'il existe une autre stratégie d'adaptation de la voix dans le bruit visant à émerger du bruit ambiant, cette fois par des caractéristiques spectrales (cf. § 3.2), qui pourrait expliquer l'évolution conjointe de l'intensité vocale et de la pente spectrale dans le bruit sans que celles-ci soient pour autant totalement corrélées.

2. EXPANSION OU COMPRESSION DE LA DYNAMIQUE VOCALE

Une autre stratégie potentielle d'émergence par rapport au bruit ambiant pourrait consister à élargir la « dynamique vocale » de façon à se distinguer du bruit ambiant, dont les caractéristiques spectro-temporelles sont ici relativement constantes pour les deux types de bruit examinés dans cette thèse. La « dynamique vocale » pourrait donc être augmentée au niveau de l'étendue vocale dans le plan F_0 *Intensité, ou plus particulièrement au niveau de la dynamique d'intensité ou de l'ambitus de fréquence fondamentale.

2.1. Evolution de l'étendue vocale dans le bruit

Les phonétogrammes de la page suivante représentent l'évolution de l'étendue vocale dans le plan F_0 *Intensité pour les locuteurs et locutrices de la BD1 (cf. Figure 80). Cette étendue vocale est calculée à partir de la distribution 3D des valeurs d'intensité et de fréquence fondamentale mesurées sur les 1min à 1min30 de voisement produits par chaque locuteur en condition silencieuse et dans les conditions de bruit blanc et de bruit cocktail à 85dB. Le calcul de l'étendue vocale est décrit plus en détail en Annexe Met1.

On observe tout d'abord chez l'ensemble des locuteurs la même tendance d'augmentation conjointe de l'intensité vocale et de la fréquence fondamentale dans le bruit, faisant se déplacer l'étendue vocale vers une zone complètement disjointe de la parole produite dans le silence, à part pour les locuteurs L7 et L8. Cependant, nous ne pouvons pas généraliser dans notre étude une augmentation en fréquence et en intensité plus importante des femmes que des hommes, contrairement à Kadiri 1998^[176].

Ce déplacement de l'étendue vocale ne s'accompagne pas d'un élargissement ni d'un rétrécissement notable de l'étendue vocale, excepté pour la locutrice L1 ou bien seulement dans un type de bruit pour les locutrices L2 et L6. Nos résultats ne semblent donc pas aller dans sens de ceux de Ternström et al. 2002^[346] constatant plutôt une diminution de l'aire du phonétogramme avec l'exposition au bruit ainsi qu'une étendue vocale plus grande chez les femmes que chez les hommes, que ce soit dans le silence ou dans le bruit. Cependant, ces différences peuvent provenir de l'unité choisie pour mesurer la fréquence fondamentale (en tons dans notre cas et en Hertz pour cet auteur). Elles peuvent également provenir de la méthode adoptée pour mesurer l'étendue vocale. Enfin, elles peuvent surtout être liées aux niveaux de bruit et aux types de bruit dans lesquels sont effectuées ces mesures. Dans une étude antérieure, Sodersten et al. 2005^[322] a en effet observé que l'aire du phonétogramme commençait par augmenter à faible et moyen niveau de bruit puis diminuait à fort niveau de bruit.

Nous allons maintenant regarder plus en détail s'il existe une augmentation de la dynamique vocale dans le bruit selon les dimensions distinctes de l'intensité vocale ou de la fréquence fondamentale.

2.2. Evolution de la dynamique d'intensité vocale dans le bruit

Nous avons alors tracé pour chaque locuteur et locutrice de la BD1 la distribution de ses valeurs d'intensité vocale en condition silencieuse puis dans les deux types de bruit à 86dB (cf. Figure 81). Cette distribution est calculée, de la même manière que l'étendue vocale, à partir des 1mn à 1mn30 de parole libre des locuteurs pour chaque condition.

En premier lieu, on observe donc que la dynamique vocale n'est pas plus grande dans le bruit que dans le silence. Il y a surtout un déplacement de la gamme d'intensités produites, mais pas d'élargissement ou de réduction particulière de cette distribution dans le bruit, excepté pour la locutrice L10 dans le bruit cocktail. L'allure de la distribution des valeurs d'intensité dans le bruit évoque cependant l'idée de compression audio utilisée en radiophonie ou en mastering de musiques actuelles (rock, RnB, etc.), consistant à amplifier une voix ou un instrument de musique et à en compresser la dynamique dans les fortes intensités. Cet effet est utilisé pour donner à la musique ou à la voix un aspect très énergique.

Aussi, contrairement à notre hypothèse de départ, le décalage de la dynamique vocale vers les fortes intensités pourrait permettre aux locuteur d'émerger du bruit ambiant de façon plus efficace que s'ils

élargissaient leur gamme d'intensités. On voit en effet sur la Figure 81 qu'un élargissement de la dynamique d'intensité conserverait des intensités vocales du même ordre de grandeur que l'intensité du bruit ambiant (86dB). Alors que dans le cas où la dynamique vocale est simplement décalée vers les fortes intensités, l'intensité vocale est à 95% du temps supérieure au niveau d'intensité du bruit ambiant.

2.3. Evolution de l'ambitus de fréquence fondamentale dans le bruit

Nous avons également examiné plus en détail l'évolution de l'ambitus de fréquence fondamentale du silence au bruit, en traçant pour chaque locuteur et locutrice de la BD1 la distribution de ses valeurs de F0 en condition silencieuse puis dans les deux types de bruit à 86dB (cf. Figure 82). Cette distribution est calculée, de la même manière que l'étendue vocale, à partir des 1mn à 1mn30 de parole libre des locuteurs pour chaque condition.

Contrairement à Boril et al. 2005^[32], nous n'observons pas seulement un décalage de la distribution de F0 vers les fréquences aiguës, mais également un élargissement important de cette distribution entre la condition silencieuse et les conditions de bruit. Cette tendance est plus marquée chez les femmes que chez les hommes, et va dans le sens de Ternström et al. 2002^[346] ayant observé une étendue vocale plus importante dans le bruit pour les femmes que pour les hommes. Cependant, il ne faut pas perdre de vue qu'il s'agit là d'un ambitus de F0 représenté en Hertz. En tons, cet ambitus n'est pas beaucoup plus grand pour les femmes que pour les hommes, mais dans tous les cas plus grand en condition bruyante qu'en condition silencieuse.

Ces résultats semblent indiquer que l'adaptation de la parole dans le bruit ne consiste pas seulement en une élévation moyenne de la fréquence fondamentale, comme nous avons pu le voir dans le chapitre précédent, mais également en un élargissement de la gamme des fréquences fondamentales produites. Celui-ci peut être interprété comme l'indicateur d'une stratégie d'augmentation de la dynamique vocale en fréquences, visant à émerger du bruit ambiant. Cependant, comme la fréquence fondamentale est liée à l'intonation, cet élargissement de la gamme de F0 peut également être la conséquence d'une autre stratégie phonétique de renforcement des indices intonatifs, comme par exemple des chutes mélodiques en fin d'énoncé ou des montées mélodiques caractéristiques de l'accent primaire en fin de syntagme accentuel (Jun et al. 1995^[170]). Nous examinerons cette autre hypothèse au chapitre 8.

Quelle que soit l'interprétation de cet élargissement de l'ambitus de F0 dans le bruit, celui-ci confirme les observations de Welby 2006^[375] pour 4 de ses 7 locutrices, mais va à l'encontre des observations de Rostolland 1982^[288] rapportant une « saturation » de la fréquence fondamentale de la voix lors de son élévation en parole criée, entraînant un manque de modulation de la voix en fréquence et affectant les indices intonatifs de la parole.

En revanche, nous observons comme Sodersten et al. 2005^[322] que les locuteurs et les locutrices utilisant régulièrement le mécanisme laryngé M0 (ou fry) en condition silencieuse, lors de pauses d'hésitation ou en fin de phrase, l'utilisent moins ou plus du tout dans le bruit.

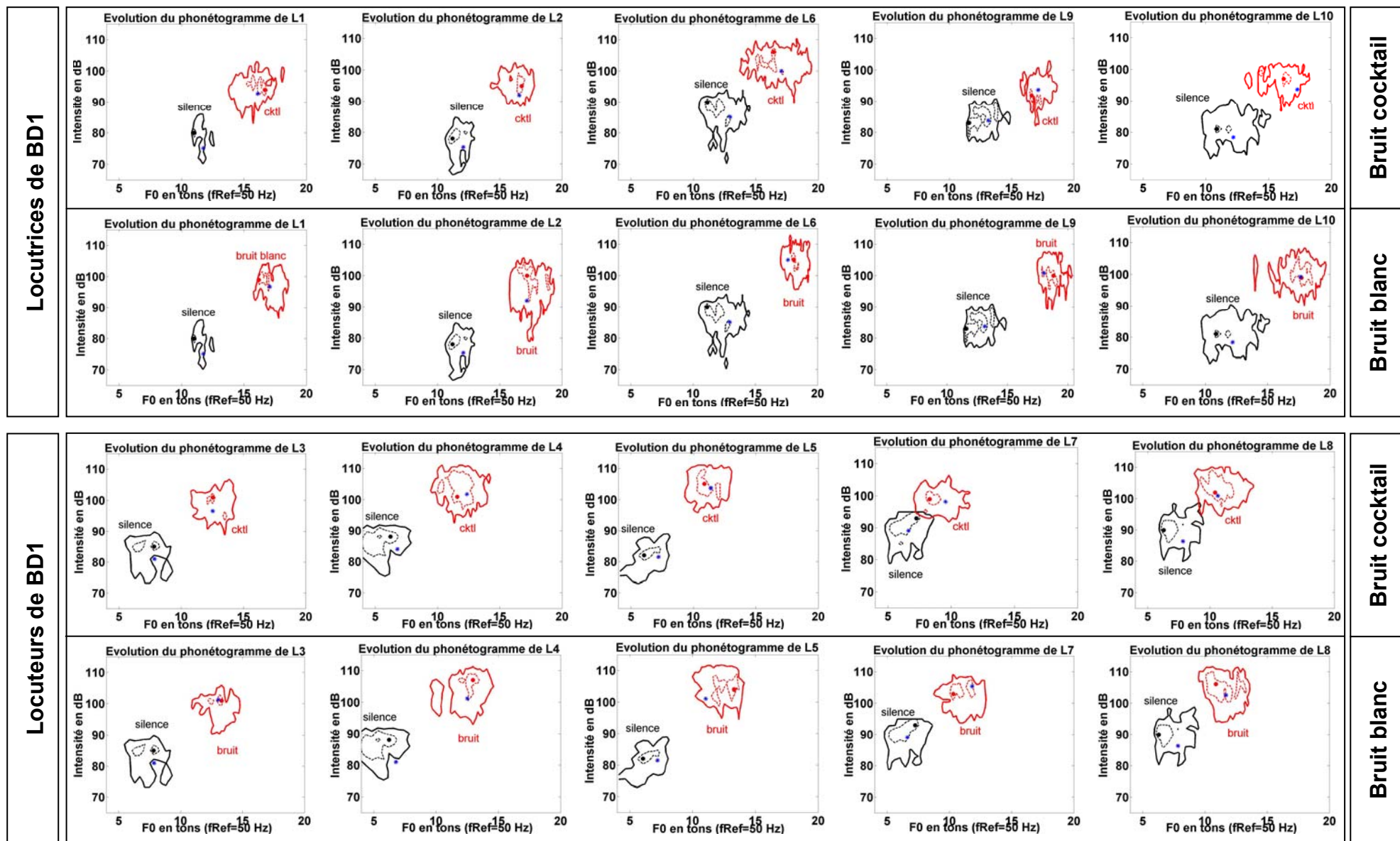


Figure 80. Représentation dans le plan F0*Intensité de l'évolution de l'étendue vocale (en trait plein) du silence (en noir) au bruit (en rouge) pour chaque locuteur et locutrice de la BD1. En traits pointillés sont représentées les zones cibles de ces étendues vocales. Le point à l'intérieur de cette zone cible représente la cible vocale (de F0 et d'intensité), tandis que l'étoile bleue représente la valeur moyenne (de F0 et d'intensité).

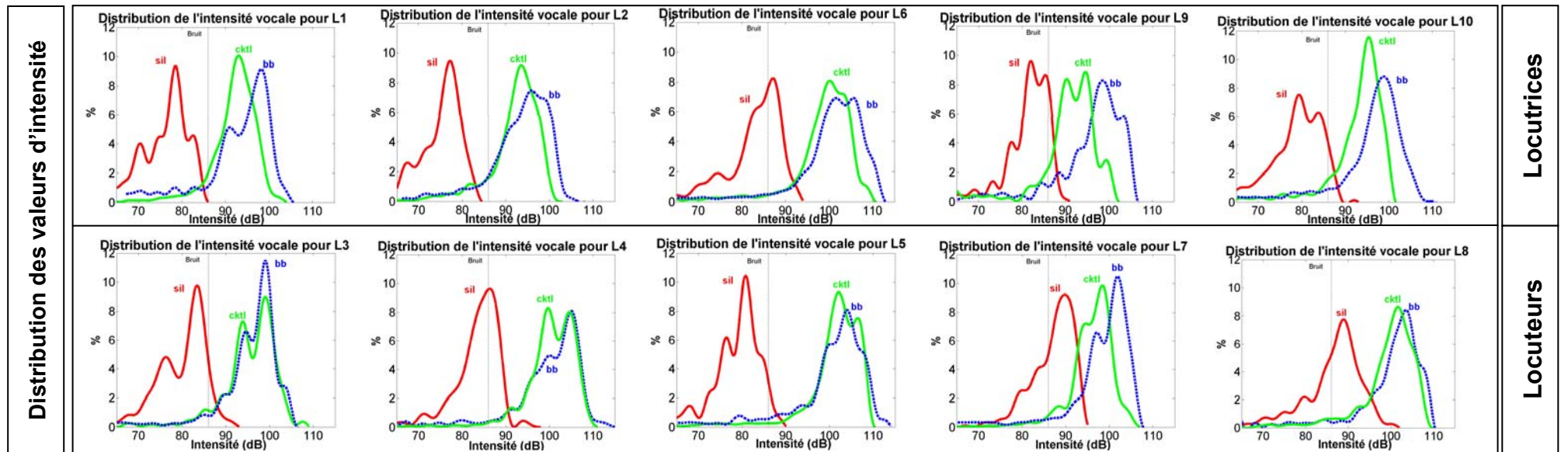


Figure 81. Représentation de la distribution des valeurs d'intensité dans le silence (en rouge), dans le bruit blanc à 86dB (en bleu pointillé) et dans le bruit cocktail à 86dB (en vert) pour chaque locuteur et locutrice de la BD1. Le niveau des deux bruits (86dB) est représenté par un trait vertical. Leur dynamique est effectivement quasi nulle.

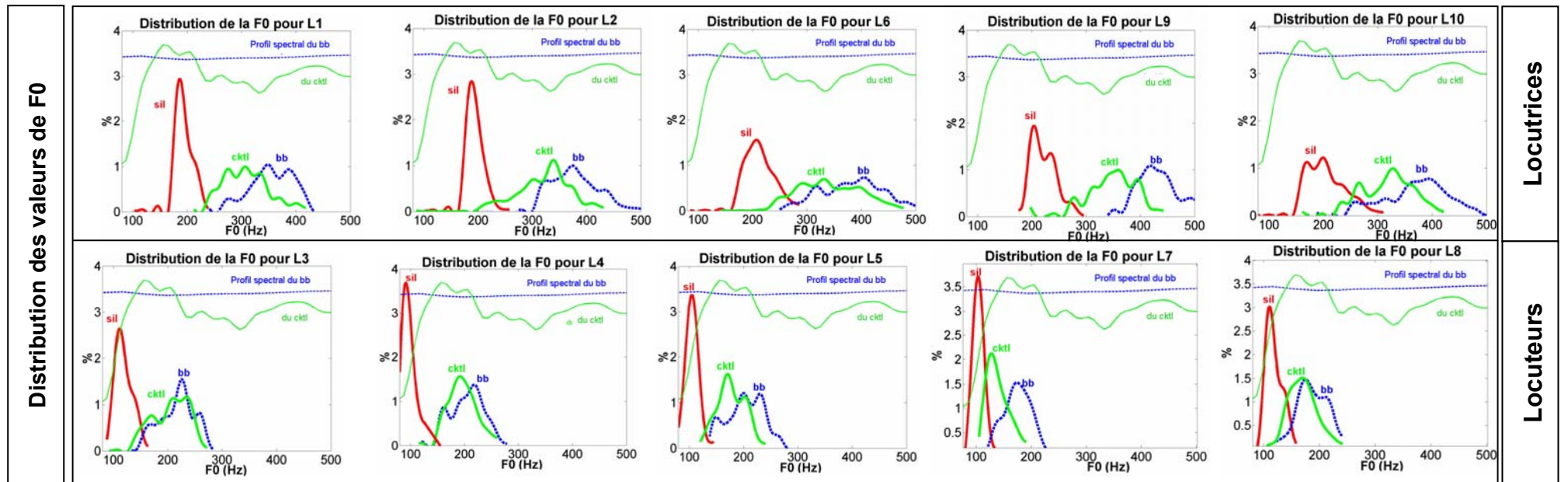


Figure 82. Représentation de la distribution des valeurs de fréquences fondamentales dans le silence (en rouge), dans le bruit blanc à 86dB (en bleu pointillé) et dans le bruit cocktail à 86dB (en vert) pour chaque locuteur et locutrice de la BD1. Les profils spectraux des deux types de bruits sont représentés dans le haut de chaque figure.

3. ADAPTATION EN FREQUENCES

L'intelligibilité de la parole est également fortement dépendante de sa fréquence fondamentale ou de son contenu spectral. Ainsi, Koopmans-van Beinum 1980^[184] a montré que les femmes sont moins intelligibles que les hommes, bien qu'elles montrent une tendance à articuler plus soigneusement. Par ailleurs, le système auditif est sujet à des effets de masquage fréquentiel, si bien que le bruit ambiant perturbe l'intelligibilité de la parole de façon différente selon la façon dont leur énergie spectrale rentre « en compétition » : Ainsi, le bruit blanc affecte davantage les consonnes que ne le fait le bruit de conversations, et au contraire le bruit de conversations affecte davantage les voyelles que le bruit blanc. Miller et al. 1955^[242] ont également montré que le bruit diminuait davantage l'intelligibilité de certaines classes de sons que d'autres et que le voisement ou la nasalité étaient les traits les moins affectés par le bruit blanc. De même, Junqua 1993^[171] émet l'hypothèse que l'intelligibilité des femmes est moins dégradée par l'ajout de bruit blanc plutôt que par du bruit de conversations (qui masque les basses fréquences davantage que les hautes fréquences), du fait que leur voix est plus soufflée (Henton et al. 1985^[143]) et donc que l'énergie de leur parole est davantage concentrée dans les basses fréquences.

Par conséquent, nous envisageons dans ce paragraphe l'existence possible de stratégies adoptées par les locuteurs en environnement bruyant pour compenser ces effets de masquage, en adaptant la fréquence fondamentale au contenu spectral du bruit ambiant ou en modifiant la répartition d'énergie en fréquences en fonction des zones où l'énergie du bruit est plus ou moins importante.

3.1. Adaptation tonale

Nous avons vu dans le chapitre précédent que l'augmentation de la fréquence fondamentale avait tendance à être légèrement plus importante du silence au bruit blanc que du silence au bruit cocktail. Cela peut être tout simplement lié au fait que l'intensité perçue du bruit blanc est plus importante que celle du bruit cocktail, à même niveau d'intensité physique, ce qui pousse les locuteurs à davantage s'adapter dans le bruit blanc en augmentant davantage leur intensité vocale, mais aussi leur fréquence fondamentale.

Néanmoins, l'observation des distributions de fréquences fondamentale de la Figure 82 laissent penser que la différence d'élévation de la fréquence fondamentale entre les deux types de bruit puisse être liée à une stratégie d'émergence « tonale ». Le profil spectral du bruit de conversations présente en effet un minimum vers les 340Hz, et au contraire deux maxima autour de 150Hz et de 400Hz. On observe justement chez les femmes que leurs valeurs de F_0 sont distribuées dans le bruit cocktail de façon inverse au profil spectral de ce bruit. La fréquence fondamentale « cible », correspondant au maximum de la distribution des valeurs de F_0 , coïncide avec le minimum spectral du bruit cocktail, ce qui n'est pas le cas pour la parole produite dans le bruit blanc. Cet ajustement est observé pour les 5 locutrices de la BD1. Les femmes semblent donc adopter une stratégie d'ajustement de leur fréquence fondamentale au contenu spectral du bruit ambiant afin d'émerger par rapport à lui.

Au contraire, un tel ajustement de la fréquence fondamentale par rapport au spectre du bruit n'est pas observé chez les locuteurs. Leur fréquence fondamentale est en effet plus basse que celle des femmes et se situe toujours dans la zone de fréquences où l'énergie du bruit cocktail est maximum. Il serait donc intéressant de renouveler cette expérience en immergeant des locuteurs dans des bruits de conversations réalisés avec des voix très graves, des voix exclusivement masculines, et des voix exclusivement féminines, afin d'observer si les hommes adoptent également de telles stratégies d'ajustement de leur fréquence fondamentale au contenu spectral du bruit ambiant, lorsque cela leur est physiologiquement possible.

3.2. Adaptation spectrale

De la même façon, nous avons vu au chapitre précédent que l'évolution de l'énergie spectrale en fonction des différentes bandes de fréquences était légèrement variable entre les deux types de bruit. Nous allons ici examiner si cette différence peut être liée à l'existence d'une stratégie d'adaptation du spectre de la voix aux caractéristiques spectrales de l'environnement bruyant. Mokbel 1992^[245] ou Egan 1972^[80] ont déjà rapporté des observations allant dans ce sens, en constatant que les locuteurs

augmentaient davantage leur énergie spectrale dans les zones où le spectre du bruit était maximum. Cette stratégie serait alors contraire à la stratégie d'adaptation « tonale » observée précédemment chez les femmes. Les locuteurs ne chercheraient alors pas à placer leur voix dans les « trous » du spectre de bruit mais au contraire à compenser le masquage induit par le bruit en renforçant l'énergie de la voix là où la perturbation est maximale. Il s'agirait en quelque sorte d'une préservation du rapport signal sur bruit non plus seulement de façon globale, mais pour chaque bande de fréquences.

3.2.1. Evolution spectrale des segments voisés et bruités de la parole dans le bruit

Dans le cas où une telle stratégie existerait, on pourrait s'attendre à ce que les locuteurs renforcent davantage l'énergie de leurs segments sonores voisés (en particulier de leurs voyelles) dans les basses fréquences pour le bruit cocktail que pour le bruit blanc, et au contraire qu'ils renforcent davantage l'énergie de leurs segments bruités non voisés (tels que les fricatives non voisées par exemple) au-delà de 2kHz dans le bruit blanc plutôt que dans le bruit cocktail.

C'est pourquoi nous avons comparé l'évolution de l'énergie spectrale dans le bruit entre les voyelles et les fricatives non voisées ([ʃ], [s] et [f]) produites par les différents locuteurs et locutrices de la 1^{ère} base de données (BD1).

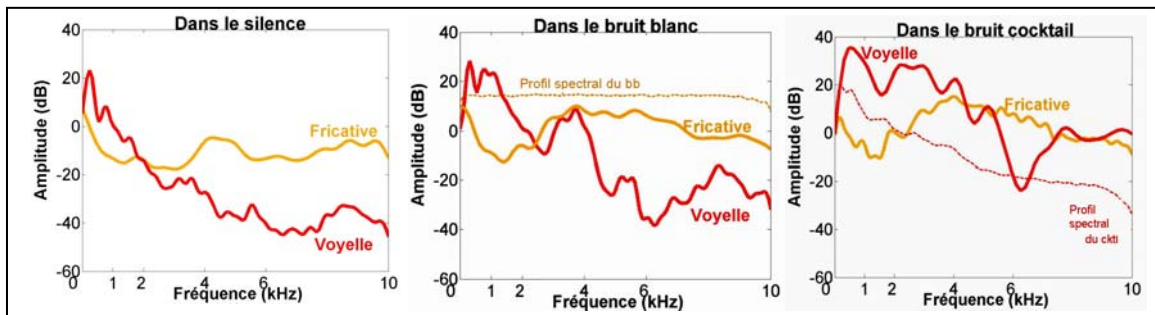


Figure 83. Exemples de spectres de voyelles ou de fricatives non voisées observés dans le silence, dans le bruit blanc et dans le bruit cocktail.

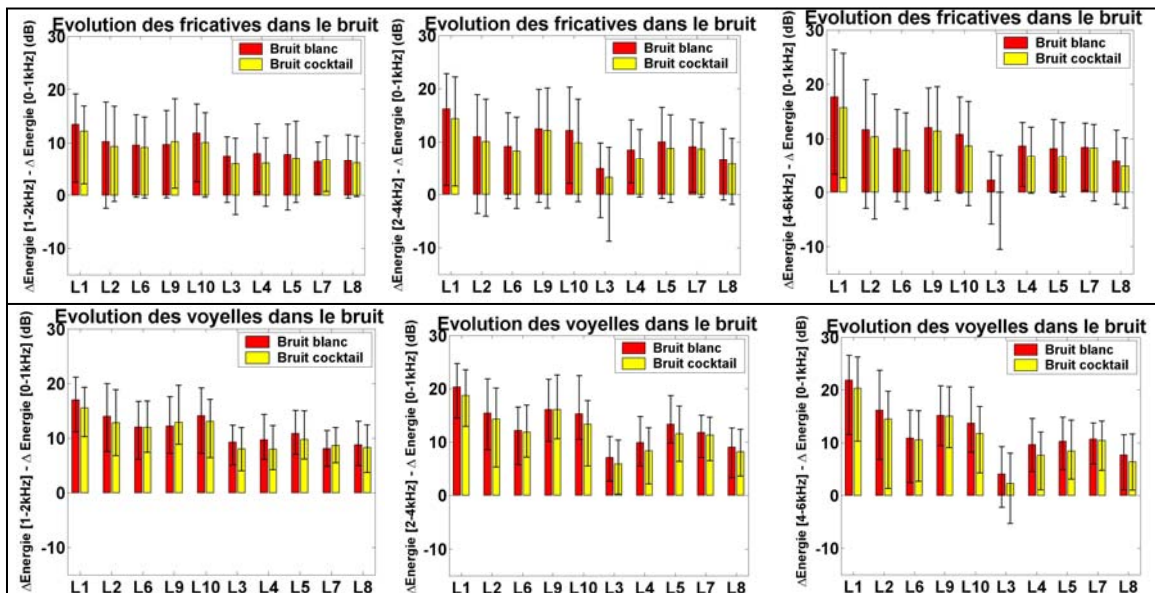


Figure 84. Evolution de l'énergie spectrale dans les bandes de fréquences [1-2kHz], [2-4kHz] et [4-6kHz] par rapport à celle de la zone [0-1kHz] pour les locuteurs et locutrices de la première base de données, pour des voyelles (en haut) ou des fricatives non voisées (en bas) produites dans 2 types de bruit (bruit blanc et bruit cocktail) à 86dB.

De façon générale, on observe que l'énergie spectrale augmente avec le bruit davantage dans les bandes de fréquences médium de la voix que dans les basses fréquences inférieures à 1kHz, que ce soit pour les voyelles ou pour les fricatives non voisées. Cela correspond bien à un

enrichissement spectral de la voix dans les fréquences médium, comme on peut l'observer sur la Figure 83. Au-delà de cette observation générale, les résultats obtenus infirment notre hypothèse de l'existence d'une stratégie d'adaptation timbrale de la voix aux caractéristiques spectrales de l'environnement bruyant, du moins pour les locuteurs et locutrices de notre 1^{ère} base de données. En effet, on observe bien, comme attendu, un renforcement du spectre des fricatives plus important dans le bruit blanc que dans le bruit cocktail pour les bandes fréquences supérieures à 2kHz. Mais ce renforcement plus important dans le bruit blanc est également observé pour les voyelles, et pour toutes les bandes de fréquences. Il semble plutôt que le renforcement spectral plus important dans le bruit blanc soit en fait relié à l'augmentation plus importante de l'effort vocal dans le bruit blanc (intensité, F0, etc.).

Lorsqu'on compare l'augmentation dans le bruit de l'intensité des voyelles et de celle des fricatives non voisées, on observe tout d'abord que l'intensité augmente pour ces deux types de segments, et non pas uniquement pour les segments voisés (Castellanos et al. 1996 ^[46]). Cependant, l'augmentation de l'intensité est en moyenne significativement plus importante pour les voyelles que pour les fricatives (cf. Figure 85). Cela confirme les résultats d'études antérieures rapportant un renforcement plus important des voyelles par rapport aux consonnes dans le bruit ou en situation d'effort vocal (Fairbanks et al. 1957 ^[87]). Ce phénomène est interprété par certains auteurs comme le fait que l'intelligibilité de la parole serait principalement portée par les voyelles, plus audibles à forte distance et dans le bruit (Fonagy et al. 1966 ^[94], Dohalska et al. 2000 ^[73]). Ainsi, Draegert 1951 ^[76] a montré que l'intensité et la durée des voyelles étaient les paramètres qui affectaient le plus l'intelligibilité dans un bruit intense. Mais cette hypothèse n'est pas acceptée par tous et Gordon-Salant (cité dans Nabelek 1990 ^[249]) a au contraire montré qu'un renforcement du rapport d'intensité Consonnes/Voyelles améliorerait significativement l'intelligibilité de 14%.

Au-delà de cette première observation, on remarque toujours une augmentation de l'intensité des voyelles plus importante dans le bruit blanc que dans le bruit cocktail, allant plutôt dans le sens d'une stratégie de préservation du rapport signal sur bruit par rapport à l'intensité perçue du bruit ambiant (cf. §1.1), et au contraire à l'encontre d'une stratégie de compensation du masquage fréquentiel induit par le bruit. Car dans ce cas, l'augmentation de l'intensité des voyelles devrait être plus importante dans le bruit cocktail que dans le bruit blanc. On remarque également de façon surprenante que si l'intensité des fricatives augmente pour la plupart des locuteurs davantage dans le bruit blanc que dans le bruit cocktail, il existe quelques locuteurs où cette intensité augmente de façon comparable (L6, L3 ou L7) entre les deux types de bruit, et même de façon plus importante dans le bruit cocktail pour L10. Cela va de nouveau à l'encontre d'une stratégie de compensation du masquage fréquentiel, qui correspondrait plutôt à une plus grande augmentation de l'intensité des fricatives dans le bruit blanc que dans le bruit cocktail.

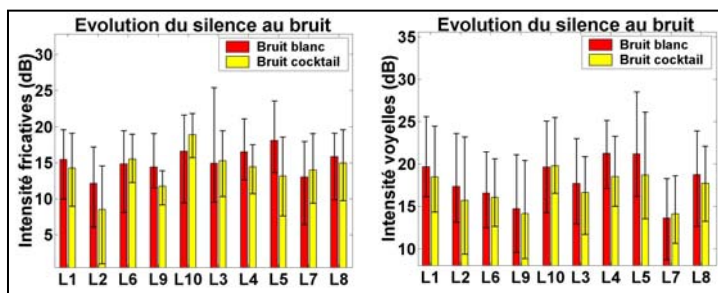


Figure 85. Evolution dans le bruit de l'intensité des fricatives non voisées (à gauche) et des voyelles (à droite) produites par les locuteurs et locutrices de la première base de données, pour 2 types de bruit (bruit blanc et bruit cocktail) à 86dB.

Enfin, nous avons comparé l'évolution de la durée des voyelles dans le bruit par rapport à celle des fricatives non voisées. Nous retrouvons encore les résultats d'études antérieures concernant l'allongement plus important des voyelles que des consonnes dans le bruit : Ainsi, Fairbanks et al. 1957 ^[87] note une augmentation avec l'effort vocal du rapport de durée des voyelles sur celle des consonnes, et Sodersten et al. 2005 ^[322] rapporte une augmentation du pourcentage de voisement dans le bruit. Dans notre cas, on observe même plutôt une absence d'évolution de la durée des fricatives non voisées pour les locuteurs L1, L2, L6, L9, L4, L5 et L8 et un raccourcissement de ces consonnes dans le bruit pour les locuteurs L10, L3 et L7 (cf. Figure 86). On n'observe pas d'effet significatif du type de bruit sur l'allongement ou le raccourcissement de différents segments dans le bruit. Ces résultats concordent encore avec ceux des études antérieures puisque tous les auteurs s'accordent sur l'augmentation dans le bruit ou en situation d'effort vocal de la durée des voyelles. En revanche, certains auteurs n'observent pas d'évolution significative de la durée des consonnes

(Castellanos et al. 1996^[46]) voire plutôt leur raccourcissement (Junqua 1993^[171] ; Kim 2005^[180] ; Rostolland 1982^[289]).

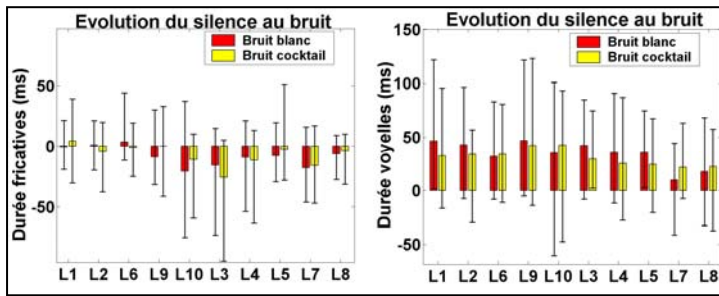


Figure 86. Evolution dans le bruit de la durée des fricatives non voisées (à gauche) et des voyelles (à droite) produites par les locuteurs et locutrices de la première base de données, pour 2 types de bruit (bruit blanc et bruit cocktail) à 86dB.

Pour conclure vis à vis de ces différentes explorations, la comparaison de l'évolution dans le bruit des fricatives non voisées (susceptibles d'être davantage masquées par le bruit blanc) et des voyelles (susceptibles d'être davantage masquées par le bruit cocktail) n'a pas montré de résultats allant dans le sens d'une stratégie de compensation de ce masquage ni au niveau spectral, ni au niveau de l'intensité globale, et pas non plus au niveau de la durée segmentale. Il ne semble pas exister pour les locuteurs de la BD1 de stratégies d'adaptation spectrale de la parole aux caractéristiques de l'environnement bruyant. En revanche, on observe globalement un renforcement de l'énergie de tous les segments dans les fréquences médium, que ces segments soient voisés ou bruités, et quel que soit le type de bruit. Nous avons également vu que ce renforcement de l'énergie ne montrait pas pour autant une très forte corrélation avec l'intensité vocale (cf. § 1.3.4). L'interprétation la plus plausible de ces différents éléments d'information est qu'il existe une stratégie générale de « timbrage » de la voix dans la zone sensible de l'oreille, indépendamment du type de bruit, de façon à renforcer l'énergie de la parole là où celle-ci est la mieux perçue par l'auditeur, de la même façon que les chanteurs lyriques renforcent l'énergie spectrale de leur voix autour de 3kHz (singing formant) pour émerger de l'orchestre (Sundberg 1972^[337]) ou que les acteurs renforcent l'énergie spectrale autour de 4kHz (Bele 2006^[23]).

3.2.2. Lien entre l'évolution formantique et le renforcement de l'énergie spectrale dans les médiums

Puisque nous avons évoqué la notion de Singing Formant, il se trouve que Sundberg 1972^[337] propose que ce renforcement spectral soit réalisé par les chanteurs en rapprochant les 3èmes et 4èmes formants vocaliques. Cela nous amène à effectuer l'hypothèse que l'évolution du 3^{ème} formant observée dans le bruit chez certains locuteurs puisse être lié à cette stratégie de timbrage de la voix dans la zone sensible de l'oreille (2-4kHz). En examinant qualitativement un grand nombre de spectres moyennés, on observe effectivement que

- dans le cas de la voyelle [a], F3 tend à se rapprocher de F4 dans le bruit pour former un renforcement spectral autour de 3.5-4kHz (cf. Figure 87). La fréquence de F3 a donc tendance à augmenter.
- dans le cas de la voyelle [u], l'amplitude de F3 tend à être renforcée dans le bruit ou F3 tend à se rapprocher de F4 et F5 pour former un renforcement spectral autour de 4kHz (cf. Figure 87). La fréquence de F3 a donc tendance à rester constante dans le bruit ou à légèrement augmenter.
- enfin dans le cas du [i], on observe deux stratégies différentes : Dans certains cas, F2 et F3 fusionnent dans le bruit autour de 2.5kHz et F4 et F5 autour de 4.5kHz, dans d'autres on remarque un rapprochement de F2, F3, F4 pour renforcer le spectre entre 2 et 4 kHz (cf. Figure 87). Dans tous les cas, la fréquence du F3 a alors tendance à diminuer.

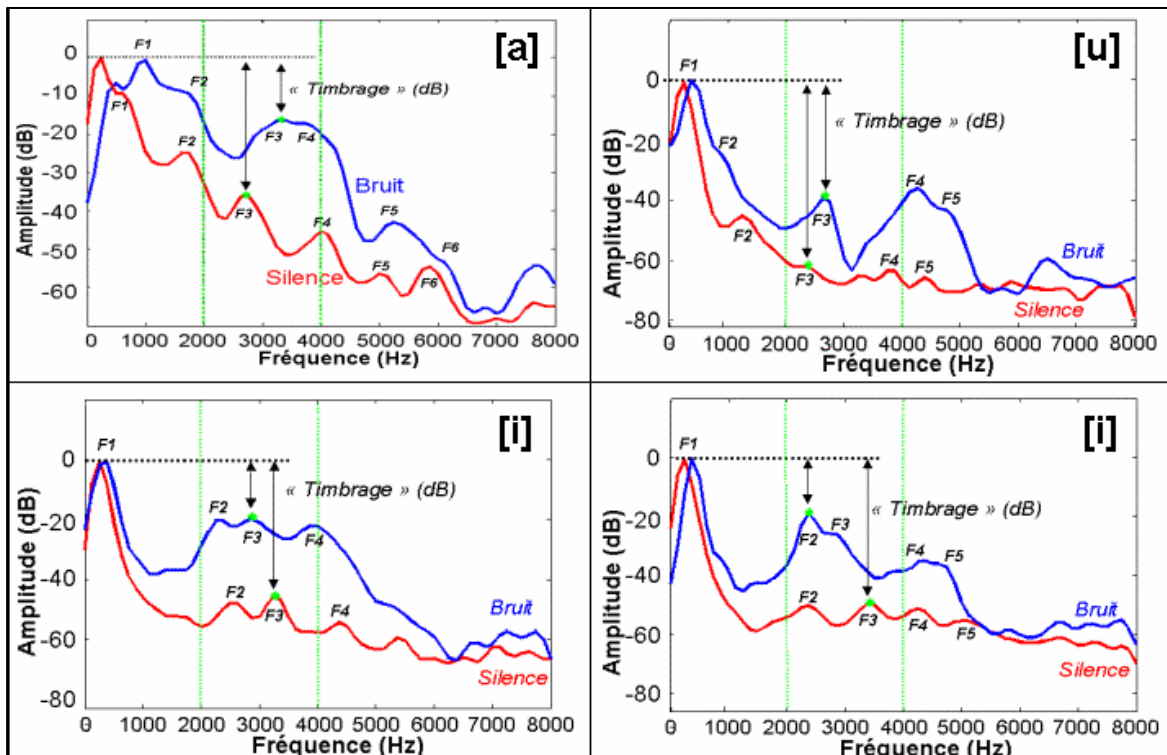


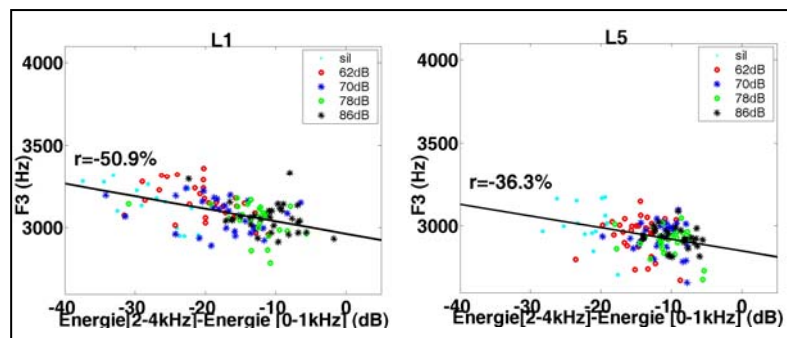
Figure 87. Exemple d'évolution dans le bruit du spectre moyenné sur les voyelles [a] (en haut à gauche), [u] (en haut à droite) et [i] (en bas).

Nous avons donc cherché à explorer la corrélation entre la fréquence centrale du 3^{ème} formant et le renforcement du spectre dans la zone sensible de l'oreille (entre 2 et 4kHz) par rapport à la zone du fondamental (entre 0 et 1kHz). Pour cela, nous avons tracé dans un même plan les valeurs du 3^{ème} formant vocalique en fonction de la différence d'énergie spectrale entre ces deux zones (cf. Figure 88). Ces deux paramètres ont été mesurés

- dans la BD1 sur les syllabes des 16 mots cibles contenant un [a], un [i] ou un [u], prononcés dans chaque condition de bruit, et cela pour les 9 conditions (silence et 8 conditions de bruit à différents niveaux d'intensité). Cela représente donc au total 108 mesures par locuteur et locutrice pour les voyelles [a], 135 pour les [i] et 54 pour les [u].
- dans la BD2 sur la syllabe du déterminant [la] et sur les 1^{ères} ou 2^{èmes} syllabes [la] des 17 logatomes cibles prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 255 mesures par locutrice pour les voyelles [a]. Nous n'avons pas examiné les corrélations pour les autres voyelles [i] et [u] faute d'effectif suffisamment important pour ces voyelles.

Les coefficients de régression linéaire de ces données sont rapportés dans la colonne « En Absolu » du Tableau 23. Nous avons également tracé dans un même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence. Les coefficients de régression linéaire de ces données sont également rapportés dans le Tableau 23, cette fois dans la colonne « En Evolution ».

Figure 88. Exemples pour une locutrice (L1) et un locuteur (L5) de la corrélation entre leur 3^{ème} formant et l'énergie spectrale de leur voix entre 2 et 4kHz par rapport à la zone [0-1kHz], pour les voyelles [i] produites en situation calme et bruyante. L'ensemble des figures pour tous les locuteurs et locutrices des deux bases de données sont consultables en Annexe Res6.



	[a]		[i]		[u]	
	En Absolu	En Evolution	En Absolu	En Evolution	En Absolu	En Evolution
L1	3.4 %	40.4 %	-50.9 %	-50.8 %	27.3 %	-10.3 %
L2	35.3 %	49.6 %	-48.1 %	-60.8 %	23.1 %	56.2 %
L6	27.6 %	35.8 %	-25.1 %	-8.4 %	-2.0 %	17.9 %
L9	-37.6 %	-10.1 %	-20.5 %	-16.4 %	14.4 %	-23.2 %
L10	50.2 %	13.6 %	-19.6 %	-16.1 %	54.4 %	36.3 %
L3	28.2 %	-19.7 %	-19.6 %	26.2 %	-22.1 %	7.4 %
L4	16.5 %	-29.3 %	-2.8 %	-3.3 %	28.2 %	10.5 %
L5	-29.5 %	-3.2 %	-36.3 %	-25.4 %	-47.5 %	-15.7 %
L7	11.2 %	-2.1 %	-21.3 %	-6.0 %	15.4 %	1.6 %
L8	-11.9 %	1.8 %	-34.2 %	-13.6 %	-12.8 %	12.3 %

	[a]	
	En Absolu	En Evolution
L2	40.2 %	-11.5 %
L6	21.9 %	-19.3 %
L11	43.7 %	28.7 %

BD2

BD1

Tableau 23. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs du 3ème formant vocalique des voyelles [a], [i] et [u] et celles de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz], toutes conditions confondues de la BD1 (à gauche) et de la BD2 (à droite). La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit du 3ème formant vocalique et la valeur d'évolution de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz].

L'examen de ces figures et des coefficients de régression linéaire montre peu de corrélation entre le 3ème formant vocalique et le renforcement de l'énergie spectrale dans la zone sensible de l'oreille, que ce soit dans l'absolu ou en évolution du silence au bruit. Cela se retrouve également au niveau des tendances globales d'évolution qui montrent autant de cas de diminution du 3^{ème} formant avec le renforcement de l'énergie spectrale que d'augmentation conjointe du 3^{ème} formant avec le renforcement de l'énergie spectrale (cf. Tableau 24). Ces résultats peuvent indiquer que le déplacement en fréquence du 3^{ème} formant ne participe pas au renforcement spectral de la voix entre 2 et 4kHz. Les résultats de Lienard et al. 1999^[215] iraient dans ce sens puisque ces auteurs observent en voix forte un renforcement de l'énergie spectrale de la voix dans les médiums bien que le 2^{ème} et le 3^{ème} formant vocalique ne se déplacent pas en fréquence. En revanche, ces auteurs constatent une augmentation de l'amplitude de ces formants dans le bruit, indépendamment de tout rapprochement et de toute fusion.

	[a]		[i]		[u]	
	F3 diminue	F3 augmente	F3 diminue	F3 augmente	F3 diminue	F3 augmente
L1	49.4 %	50.6 %	62.1 %	29.3 %	33.3 %	63.3 %
L2	21.8 %	75.9 %	59.3 %	25.0 %	44.7 %	29.8 %
L6	39.8 %	60.2 %	58.8 %	35.3 %	60.4 %	35.4 %
L9	73.8 %	26.2 %	58.1 %	40.0 %	18.9 %	51.4 %
L10	19.6 %	79.3 %	80.9 %	19.1 %	6.8 %	68.2 %
L3	12.7 %	87.3 %	72.5 %	23.1 %	48.9 %	12.8 %
L4	28.2 %	71.8 %	41.1 %	51.6 %	23.9 %	52.2 %
L5	85.1 %	13.8 %	79.7 %	19.5 %	79.2 %	16.7 %
L7	38.0 %	62.0 %	70.2 %	29.8 %	31.9 %	61.7 %
L8	69.5 %	25.3 %	61.3 %	30.3 %	75.0 %	22.9 %

	[a]	
	F3 diminue	F3 augmente
L2	26.0 %	74.0 %
L6	17.3 %	82.7 %
L11	25.0 %	75.0 %

BD2

BD1

Tableau 24. Pourcentage de cas où le 3ème formant vocalique augmente ou diminue conjointement à un renforcement de l'énergie spectrale dans la zone sensible de l'oreille, pour chaque locuteur et locutrice de la 1^{ère} base de données (à gauche) et de la 2^{ème} base de données (à droite). La somme des deux colonnes ne correspond pas toujours à 100%. Il existe en effet quelques rares cas où l'énergie spectrale est moins renforcée avec le bruit dans la zone [2-4kHz] que dans la zone [0-1kHz]. L'effectif de ces cas correspond au pourcentage restant pour atteindre 100%.

3.2.3. Lien entre l'évolution des mouvements articulatoires et le renforcement de l'énergie spectrale dans les médiums

Si l'augmentation de l'énergie de la voix dans les fréquences médium n'est pas uniquement liée à l'aplatissement de la pente spectrale de par l'augmentation de l'effort glottique, alors on peut raisonnablement émettre l'hypothèse que les articulateurs puissent jouer un rôle dans le renforcement du spectre observé entre 2 et 4kHz, ainsi que dans le renforcement spectral observé chez certains locuteurs entre 4 et 6kHz.

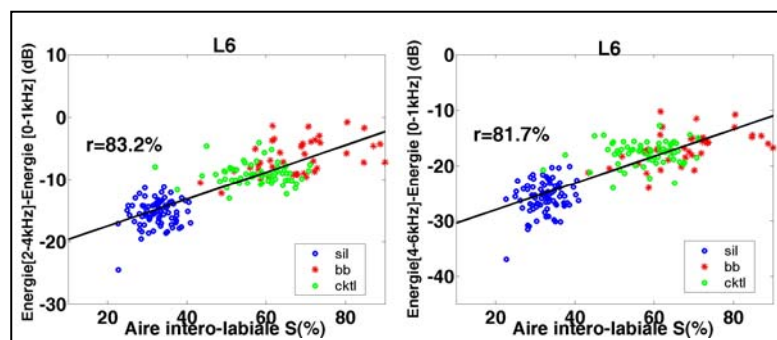
Une première idée consiste à envisager ce renforcement spectral du point de vue du rayonnement de la voix, et à considérer le conduit vocal comme un pavillon. On sait, d'après la théorie des pavillons, que l'augmentation de la surface d'ouverture du pavillon permet d'améliorer son rayonnement et s'accompagne d'une diminution de la fréquence critique du pavillon (cf. Annexe Met1). Nous explorerons donc dans un premier temps si l'augmentation de l'aire intéro-labiale observée dans le bruit peut être directement reliée aux renforcements du spectre existant dans le bruit entre 2 et 4kHz, ou entre 4 et 6 kHz (cf. Figure 87).

Une deuxième idée provient d'une autre interprétation du formant du chanteur également proposée par Sundberg 1972^[337] puis reprise par Dang et al. 1997^[61]. Ces auteurs proposent en effet que le renforcement du spectre autour de 3kHz soit lié à la fréquence de résonance des sinus piriformes (vers les 2-3kHz) ou des cavités pharyngées plus globales (vers les 3-3.5kHz), et contrôlable dans une certaine mesure par l'abaissement du larynx, provoquant un élargissement de cette cavité. Nous explorerons donc dans un deuxième temps si les mouvements verticaux du larynx sont corrélés au renforcement du spectre de la voix entre 2 et 4kHz.

3.2.3.1. L'aire intéro-labiale et le rayonnement de la voix

Pour explorer la potentielle corrélation entre le timbrage de la voix et l'aire intéro-labiale, nous avons tracé dans un même plan les valeurs d'aire intéro-labiale en fonction de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz]. Nous avons de la même manière tracé dans un même plan les valeurs d'aire intéro-labiale en fonction de la différence d'énergie spectrale entre la zone [4-6kHz] et la zone [0-1kHz] (cf. Figure 89). Ces trois paramètres ont été mesurés sur la syllabe du déterminant [la] et sur les 2 syllabes des 17 logatomes cibles de la BD2, prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 255 mesures par locutrice. Les coefficients de régression linéaire de ces données sont rapportés dans la colonne « En Absolu » du Tableau 25. Nous avons également tracé dans un même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence. Les coefficients de régression linéaire de ces données sont également rapportés dans le Tableau 25, cette fois dans la colonne « En Evolution ».

Figure 89. Exemples pour une locutrice de la corrélation entre ses valeurs d'aire intéro-labiale et le timbrage de sa voix dans les zones [2-4kHz] et [4-6kHz] par rapport à la zone [0-1kHz], en situation calme et bruyante. L'ensemble des figures pour toutes les locutrices de la BD2 sont consultables en Annexe Res6.



L'examen de ces figures et des coefficients de régression linéaire montre une forte corrélation chez les 3 locutrices de BD2 entre l'aire intéro-labiale et la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz]. Une corrélation très forte est également observée entre l'aire intéro-labiale et la différence d'énergie spectrale entre la zone [4-6kHz] et la zone [0-1kHz], mais seulement pour les locutrices L6 et L11. En revanche, les évolutions de ces différents paramètres dans le bruit sont peu corrélées, bien qu'au niveau des tendances globales d'évolution l'aire intéro-labiale augmente dans tous les cas de façon conjointe à l'augmentation de l'énergie spectrale entre 2 et 4 kHz.

En ce qui concerne l'énergie entre 4 et 6 kHz, on retrouve que celle-ci augmente dans le bruit de façon conjointe à l'augmentation de l'aire intéro-labiale pour les locutrices L6 et L11, tandis que la locutrice L2 montre 36.5% de cas où son aire intéro-labiale augmente conjointement à une atténuation de l'énergie entre 4 et 6kHz par rapport à la zone [0-1kHz].

	Renforcement [2-4 kHz]		Renforcement [4-6 kHz]	
	En Absolu	En Evolution	En Absolu	En Evolution
L2	68.2 %	11.8 %	-6.5 %	-7.5 %
L6	83.2 %	35.7 %	81.7 %	23.0 %
L11	87.7 %	50.5 %	80.3 %	31.1 %

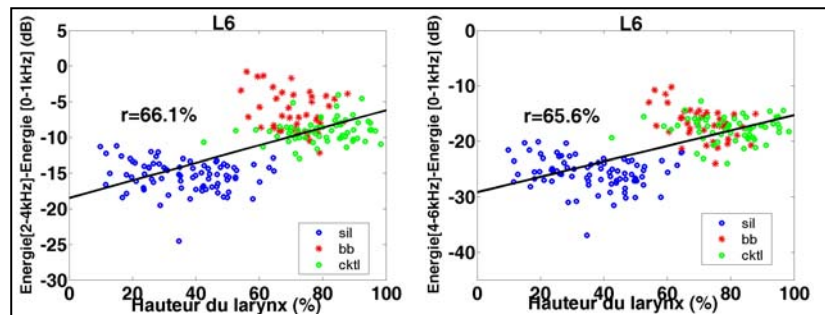
Tableau 25. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs de l'aire intéro-labiale et celles de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz], ou entre la zone [4-6kHz] et la zone [0-1kHz], toutes conditions confondues de la BD2. La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit de l'aire intéro-labiale et les valeurs d'évolution de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz] ou entre la zone [4-6kHz] et la zone [0-1kHz].

En conclusion, l'augmentation de l'aire intéro-labiale semble assez liée à la stratégie de renforcement du spectre dans les fréquences médiums adoptée par les locuteurs pour émerger du bruit ambiant. Cela est également le cas pour les fréquences entre 4 et 6 kHz, mais seulement pour les locutrices L6 et L11.

3.2.3.2. Les mouvements verticaux du larynx et la résonance des cavités pharyngées

Pour explorer la potentielle corrélation entre le timbrage de la voix et les mouvements verticaux du larynx, nous avons tracé dans un même plan les valeurs d'amplitude des mouvements verticaux du larynx en fonction de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz]. Nous avons procédé de la même manière pour la différence d'énergie spectrale entre la zone [4-6kHz] et la zone [0-1kHz] (cf. Figure 90). Ces trois paramètres ont été mesurés sur la syllabe du déterminant [la] et sur les 2 syllabes des 17 logatomes cibles de la BD2, prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 255 mesures par locutrice. Les coefficients de régression linéaire de ces données sont rapportés dans la colonne « En Absolu » du Tableau 26.

Figure 90. Exemples pour une locutrice de la corrélation entre la hauteur de son larynx et le timbrage de sa voix dans les zones [2-4kHz] et [4-6kHz] par rapport à la zone [0-1kHz], en situation calme et bruyante. L'ensemble des figures pour toutes les locutrices de la BD2 sont consultables en Annexe Res6.



	Renforcement [2-4 kHz]		Renforcement [4-6 kHz]	
	En Absolu	En Evolution	En Absolu	En Evolution
L2	12.8 %	-18.9 %	-10.9 %	-0.8 %
L6	66.1 %	-29.0 %	65.6 %	-5.4 %
L11	71.7 %	15.2 %	67.6 %	10.6 %

Tableau 26. La colonne « En Absolu » représente les coefficients de régression linéaire entre les valeurs de l'amplitude des mouvements verticaux du larynx et celles de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz], ou entre la zone [4-6kHz] et la zone [0-1kHz], toutes conditions confondues de la BD2. La colonne « En Evolution » représente les coefficients de régression linéaire entre les valeurs d'évolution du silence au bruit de l'amplitude des mouvements verticaux du larynx et les valeurs d'évolution de la différence d'énergie spectrale entre la zone [2-4kHz] et la zone [0-1kHz] ou entre la zone [4-6kHz] et la zone [0-1kHz].

4. CONCLUSION DU CHAPITRE 6

Nous avons examiné dans ce chapitre différentes hypothèses de stratégies pouvant être adoptées par les locuteurs pour émerger du bruit ambiant.

- De nombreuses observations argumentent en faveur de l'existence d'une stratégie de préservation d'un rapport signal sur bruit acceptable par l'augmentation de l'intensité vocale, non pas par rapport à l'intensité physique du bruit ambiant mais à son intensité perçue. Cette stratégie est intimement liée à d'autres modifications telles que l'élévation moyenne de la fréquence fondamentale ou l'augmentation de l'ouverture labiale. Un enrichissement spectral de la voix dans les médiums l'accompagne également, sans lui être pour autant fortement corrélé.
- Nous avons également observé une possible stratégie d'augmentation de la dynamique vocale dans le bruit chez tous les locuteurs par l'élargissement de la gamme de fréquences fondamentale produites. Celle-ci peut cependant être uniquement la conséquence d'une autre stratégie de renforcement dans le bruit de certains indices intonatifs (cf. chapitre 8).
- Enfin, nous avons remarqué chez les femmes l'existence d'une possible stratégie d'adaptation « tonale » consistant à un ajuster leur fréquence fondamentale dans les minima spectraux du bruit ambiant. Celle-ci nécessite cependant d'être confirmée par des expériences supplémentaires. Nous n'avons en revanche pas observé de tels comportements en ce qui concerne la répartition de l'énergie spectrale de la voix, contrairement à d'autres études antérieures (Mokbel 1992^[245], Egan 1972^[80]).
- Pour finir, nous avons constaté un renforcement récurrent de l'énergie spectrale de la voix dans les fréquences médiums chez tous les locuteurs et quel que soit le type de bruit. Celui-ci n'est pas très fortement corrélé à l'augmentation de l'intensité vocale. Il est davantage corrélé à l'augmentation de l'aire intéro-labiale sans que nous puissions pour autant déduire un quelconque lien de causalité entre ces deux paramètres.

Nous avons ici examiné des stratégies de renforcement global de certaines caractéristiques acoustiques dans le but d'émerger du bruit ambiant. Cependant, l'intelligibilité est loin d'être uniquement liée à des caractéristiques acoustiques globales de la parole, mais également à plusieurs indices plus locaux de l'énoncés, tels que les indices de reconnaissance des phonèmes, les indices intonatifs et articulatoires de segmentation de l'énoncé, ou encore les différences de prévisibilité entre les différents mots d'un énoncé. C'est pourquoi nous allons explorer dans les chapitres suivant d'autres hypothèses de stratégies pouvant être adoptées pour communiquer dans le bruit, visant cette fois à renforcer ces différents indices de « codage » et de compréhension de la parole.

Chapitre 7 :

Une adaptation visant à favoriser la reconnaissance des unités phonétiques ?

- 1. RENFORCEMENT DU CONTRASTE
ACOUSTIQUE ENTRE LES VOYELLES**
- 2. RENFORCEMENT DES INDICES VISUELS
D'IDENTIFICATION OU DE DISCRIMINATION
DES UNITES PHONETIQUES**
- 3. ALLONGEMENT DU TEMPS DE PRODUCTION
OU DE RECUPERATION DE CES INDICES**
- 4. CONCLUSION DU CHAPITRE 7**

Nous avons observé au chapitre 5 que l'adaptation de la parole dans le bruit est en partie caractérisée par des modifications articulatoires allant dans le sens d'une hyper-articulation, et par des modifications formantiques, en particulier du 1^{er} formant. Nous avons déjà envisagé certaines hypothèses de stratégies adoptées par les locuteurs en environnement bruyant et examiné comment ces modifications articulatoires et formantiques peuvent être liées à l'augmentation de l'intensité vocale ou au renforcement de l'énergie spectrale de la voix dans les fréquences médiums, visant toutes les deux à émerger acoustiquement du bruit ambiant.

Cependant, les formants vocaliques et les mouvements articulatoires labiaux ont un double statut car ils sont intimement liés au geste global de production de la voix, mais ils constituent également des indices de « codage » de la parole dans les domaines audio et visuels, permettant à l'interlocuteur de reconnaître les unités phonétiques, et par conséquent de « décoder » le message produit par le locuteur.

- Ainsi, la théorie de l'invariance acoustique (Blumstein et al. 1979 ^[28]) et la théorie « direct realist » (Fowler 1986 ^[96]) s'accordent sur l'idée que le système perceptif est sensible à des propriétés invariantes, à partir desquelles il lui est possible de catégoriser des unités phonétiques. C'est le cas en particulier des caractéristiques formantiques des voyelles ou des propriétés spectrales des bruits de plosion ou de friction dans le cas des consonnes occlusives ou fricatives. Les expériences sur la parole compensatoire montrent ainsi que les locuteurs s'adaptent sur le plan articulatoire pour maintenir des caractéristiques formantiques ou spectrales stables (Jones et al. 2003 ^[169], Savariaux et al. 1999 ^[302], Houde et al. 1998 ^[156]).
- Pour la théorie « direct realist » (Fowler 1986 ^[96]), la théorie motrice de la parole (Liberman et al. 1985 ^[212]), ou la théorie de la perception pour le contrôle de l'action (Schwartz et al. 2002 ^[308]), les catégories phonétiques sont définies au niveau du conduit vocal ou du geste de production et le signal acoustique n'est qu'un indicateur, au même titre que le signal visuel, pour remonter aux gestes articulatoires.
- Au contraire, la théorie de variabilité acoustique de la parole rejette l'hypothèse de la présence d'invariant physique. Les catégories phonétiques sont uniquement définies au niveau de la compréhension de l'auditeur et non pas au niveau du signal physique (Lindblom 1996 ^[218]). L'auditeur ne se baserait pas seulement sur les indices physiques du signal mais aussi sur l'information concernant des conditions dans lesquelles la parole a été produite. Cette théorie conçoit davantage la sensibilité du système perceptif en terme de rapports entre des indices acoustiques qu'en terme de valeur de ces indices dans l'absolu.

Une très bonne synthèse bibliographique et critique de ces différentes théories a été effectuée par Perrier 2005 ^[268].

Puisque les formants et les mouvements articulatoires labiaux participent à l'intelligibilité de la parole, on peut penser que leur évolution observée dans le bruit est liée à l'existence d'une stratégie de renforcement dans le bruit des indices de reconnaissance des unités phonétiques. C'est pourquoi nous examinerons au cours de ce chapitre si les locuteurs cherchent à augmenter leur intelligibilité dans le bruit en renforçant des indices acoustiques ou visuels de reconnaissance ou de discrimination des unités phonétiques. Nous examinerons également si le ralentissement du débit de parole observé dans le bruit peut contribuer à faciliter la récupération de ces indices par l'interlocuteur.

1. RENFORCEMENT DU CONTRASTE ACOUSTIQUE ENTRE LES VOYELLES

La question de la perception auditive des voyelles est double :

- Tout d'abord, quels sont les indices acoustiques permettant de « coder » les différentes voyelles et à l'inverse de les reconnaître ?

Cette question a suscité de nombreuses études depuis une quarantaine d'année. Le rôle des 3 premiers formants vocaliques dans l'identification des voyelles est maintenant communément accepté (Carlson et al. 1970 ^[42]). La Figure 91 montre ainsi comment les dispositions différentes des 3 premiers formants pour les voyelles [a], [i], [u], contribuent au codage et à la reconnaissance de ces 3 voyelles. De nombreuses recherches ont ensuite montré que les indices de reconnaissance des catégories vocaliques ne correspondaient pas tant aux valeurs de ces 3 premiers formants dans l'absolu qu'à leur distance tonotopique les uns par rapport aux autres, et à l'intégration de formants proches en un seul formant effectif perçu (effet du centre de gravité : Schwartz et al. 1989 ^[310], Chistovich et al. 1979 ^[51]).

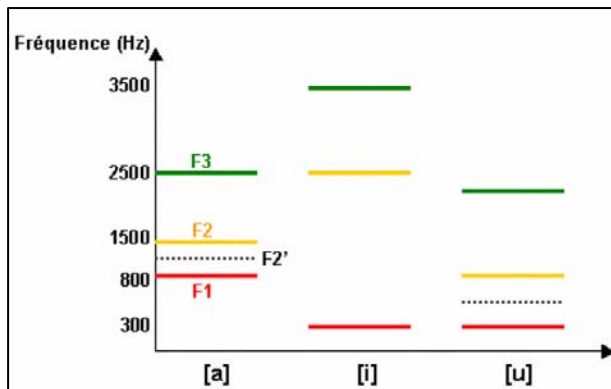


Figure 91. Représentation de la disposition typique des 3 premiers formants vocaliques des voyelles [a], [i] et [u] les uns par rapport aux autres. F2' représente l'intégration perceptive des deux premiers formants lorsque ceux-ci sont proches.

- Un autre problème se pose alors : ces indices peuvent évoluer graduellement dans l'espace physique tandis que la perception des voyelles est catégorielle. Aussi, comment cette catégorisation est-elle effectuée par l'auditeur ? Une voyelle est-elle d'autant plus reconnaissable que les indices acoustiques qui la « codent » se rapprochent d'un prototype ? Ou bien une catégorie de voyelles est-elle associée à un domaine de l'espace acoustique délimité par des « frontières » en deçà desquelles une voyelle est reconnue comme appartenant à la catégorie et au-delà desquelles la voyelle est reconnue comme appartenant à la catégorie voisine ?

Certaines études ont montré qu'il existe pour chaque catégorie de voyelles un stimulus acoustique associé à des scores d'identification supérieurs aux autres ou à une identification plus rapide (Grieser et al. 1989^[123], Sussman 1993^[339]). Par ailleurs, d'autres études ont montré que la production des voyelles est d'autant plus dispersée pour une même catégorie vocalique que la langue présente un nombre restreint de voyelles (Manuel et Krakow 1984, cités dans Meunier et al. 2006^[238]). De façon similaire, des travaux ont montré que les locuteurs d'une langue tendent à utiliser un « hyper-espace » pour rendre leurs catégories vocaliques les plus distinctes les unes des autres (Johnson et al. 1993^[167]), et qu'une expansion du système vocalique est observée dans la parole « claire », associée à un ralentissement du débit de parole (Lindblom et al. 1992^[219]). À l'inverse, une réduction du système vocalique est observée en parole spontanée par rapport à de la parole lue, associée à une accélération du débit de parole. Ce phénomène est interprété par Meunier et al. 2006^[238] comme le signe d'une adaptation de l'effort articulatoire aux exigences de la situation de communication : une réduction ne serait possible que si d'autres aspects contextuels ou communicationnels peuvent la compenser et ainsi maintenir la distinctivité du système à un autre niveau.

Dans ce paragraphe, nous allons donc explorer si, pour le français, l'adaptation formantique des locuteurs observée en environnement bruyant peut être l'indicateur de l'existence d'une stratégie de renforcement de la distinctivité du système vocalique. Nous allons donc en particulier examiner comment évoluent les 3 extrema [a], [i] et [u] de ce système vocalique Français dans le bruit. Puisque nous nous intéressons à leur discrimination et non à leur évolution isolée, nous examinerons comment évolue le contraste entre ces 3 voyelles, relativement à différents indices perceptifs de discrimination, et de façon globale dans l'espace de leurs 3 premiers formants.

1.1. Evolution de la discrimination des voyelles selon plusieurs indices perceptifs

Nous allons donc examiner tour à tour plusieurs indices acoustiques permettant de discriminer entre eux les extrema du système vocalique français ([a], [i] et [u]), et étant susceptibles d'être accentués dans le bruit.

Nous avons adopté la même méthode dans les 3 paragraphes suivants. Pour chacun des cas, nous avons pris une voyelle V1 ([a] dans le §1.1.1, [i] dans le §1.1.2 et [u] dans le §1.1.3), et un indice acoustique (IA) permettant de distinguer perceptivement cette voyelle V1 des deux autres (V2 et V3, respectivement [i] et [u] dans le §1.1.1, [u] et [a] dans le §1.1.2, et [i] et [a] dans le §1.1.3). Nous avons mesuré l'indice (IA) :

- dans BD1, sur les syllabes [la] de [lagon] et [va] de [vallée], sur les syllabes [Zi] de [gitans] et [zi] de fusil, sur les syllabes [Zu] de [bijou] et [tu] de [toupie]. Au total, nous disposons de 6 mesures de chaque voyelle V1, V2 et V3, par locutrice et par condition de silence ou de bruit cocktail à 86dB. Nous avons délibérément ignoré d'autres syllabes des mots-cibles contenant des [a] et des [i], de manière à avoir le même effectif pour les 3 groupes de voyelles [a], [i] et [u] et à pouvoir mener des comparaisons statistiques de ces groupes.

- dans BD2, sur les syllabes [la] des logatomes [lala] de la deuxième base de données (BD2), sur les syllabes [lu] des logatomes [lula] et [lalu] et sur les syllabes [li] des logatomes [lila] et [lali]. Au total, nous disposons de 8 mesures de chaque voyelle V1, V2 et V3, par locutrice et par condition de silence ou de bruit cocktail à 85dB.

A partir de ces mesures, nous avons réalisé des tests Anova à un facteur (TYPE VOYELLE) pour vérifier si l'indice IA était significativement différent:

- entre V1 et V2, que ce soit dans le silence ou dans le bruit.
- entre V1 et V3, que ce soit dans le silence ou dans le bruit.

Nous avons également réalisé un autre test Anova à un facteur (BRUIT) pour déterminer :

- si le contraste entre V1 et V2 selon l'indice IA évoluait significativement avec le bruit.
- si le contraste entre V1 et V3 selon l'indice IA évoluait significativement avec le bruit.

Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res7. Nous n'en présenterons ici qu'une synthèse.

1.1.1. Distinction des voyelles [a] par rapport aux voyelles [u] et [i]

Pour commencer, nous avons examiné la distinctivité des voyelles [a] par rapport aux autres voyelles [i] et [u]. Dans la littérature, le 1^{er} formant est classiquement associé à la dimension d'ouverture des voyelles de par sa très forte corrélation avec l'ouverture de la mâchoire (Lindblom et al. 1971^[220]). Ainsi, le 1^{er} formant est assez haut dans le cas du [a], et bas dans le cas du [i] et du [u] (cf. Figure 91). Le premier formant est cependant dépendant de la fréquence fondamentale. Traunmüller 1981^[355] ont alors montré que la différence $F1-F0$, exprimée en barks, est l'indice le plus pertinent pour rendre compte de la perception de « hauteur » des voyelles, tout en s'abstrayant de la variabilité interindividuelle et contextuelle. Nous avons donc choisi d'examiner comment évolue la distinctivité dans le bruit des voyelles [a], [i] et [u] vis à vis de cette dimension $(F1-F0)_{\text{barks}}$.

Les Figure 92 et Figure 93 synthétisent les résultats obtenus en représentant l'évolution du silence au bruit du contraste entre les voyelles [a] et [u] et entre les voyelles [a] et [i] selon la dimension $(F1-F0)_{\text{barks}}$. Ce contraste est toujours très significatif pour tous les locuteurs, aussi bien dans le silence que dans le bruit. Il tend à augmenter dans le bruit pour tous les locuteurs et locutrices de la BD1, mais n'augmente significativement que pour L6, L3, L4, L5 et L8. Au contraire, il tend à diminuer dans le bruit pour les locutrices L6 et L11 de la BD2, sans pour autant que cela soit significatif.

1.1.2. Distinction des voyelles [i] par rapport aux voyelles [a] et [u]

Nous avons ensuite examiné la distinctivité des voyelles [i] par rapport aux autres voyelles [a] et [u]. Dans la littérature, cette voyelle se caractérise surtout par un timbre très clair, correspondant à un centre de gravité spectrale très haut, vers les 3kHz. Cette voyelle peut être assez bien caractérisée par un formant dit « effectif » $F2'$, n'existant pas acoustiquement parlant, mais correspondant à l'intégration perceptive des 3^{ème} et 4^{ème} formants (cité par Vaissière 2006^[359]). Par ailleurs, Traunmüller 1985 (cité dans Savariaux et al. 1999^[302]) ont proposé la différence $(F2-F1)$, exprimée en barks, comme un bon indice perceptif de discrimination des voyelles antérieures (telles que [i]) et des voyelles postérieures arrondies (telles que [u]). La distance entre le 1^{er} et le 2^{ème} formant est en effet assez importante dans le cas des voyelles [i] tandis que ces deux formants sont plutôt rapprochés pour les voyelles [a] et [u] (cf. Figure 91). Par ailleurs, Lienard et al. 1999^[215] ont noté une augmentation de la différence $(F2-F1)$ avec l'effort vocal. Nous avons donc choisi d'examiner comment évolue la distinctivité dans le bruit des voyelles [a], [i] et [u] vis à vis de cette dimension $(F2-F1)_{\text{barks}}$.

Les Figure 94 et Figure 95 synthétisent les résultats obtenus en représentant l'évolution du silence au bruit du contraste entre les voyelles [i] et [u] et entre les voyelles [i] et [a] selon la dimension $(F2-F1)_{\text{barks}}$. Cette différence entre les voyelles [i] et les autres voyelles [a] et [u] est toujours très significative pour tous les locuteurs (excepté L1 dans le silence), aussi bien dans le silence que dans le bruit, et tend à légèrement augmenter dans le bruit. Cette augmentation est significative pour les locuteurs L6, L4, L5 et L7. Au contraire, le contraste entre les voyelles [i] et [u] tend à décroître dans le bruit, de façon significative pour L11 uniquement.

Il semble donc que l'évolution de $(F2-F1)_{\text{barks}}$ soit plutôt une conséquence indirecte de l'élévation de $F1$ qu'un réel renforcement de l'indice de discrimination des voyelles [i] par rapport aux autres voyelles.

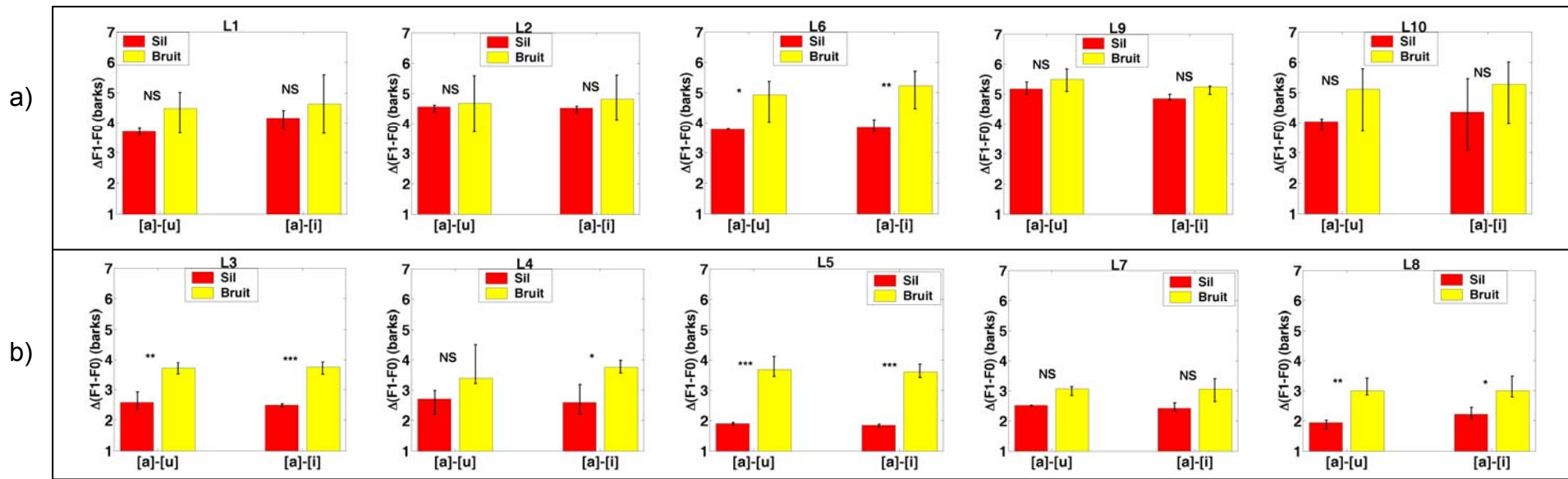


Figure 92. Evolution du silence au bruit du contraste entre [a] et [u] et entre [a] et [i] selon la dimension (F1-F0), exprimée en barks, pour les locutrices (a) et les locuteurs (b) de la BD1.

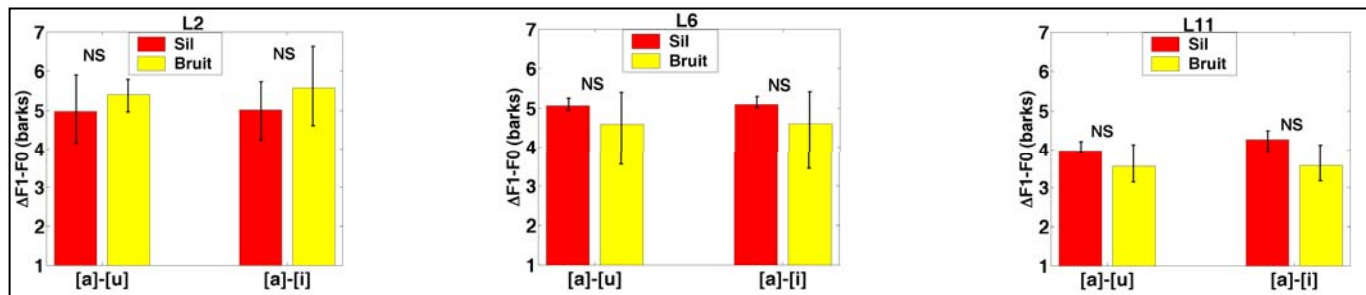


Figure 93. Evolution du silence au bruit du contraste entre les [a] et les [u] et entre les [a] et les [i] selon la dimension (F1-F0), exprimée en barks, pour les locutrices de la BD2

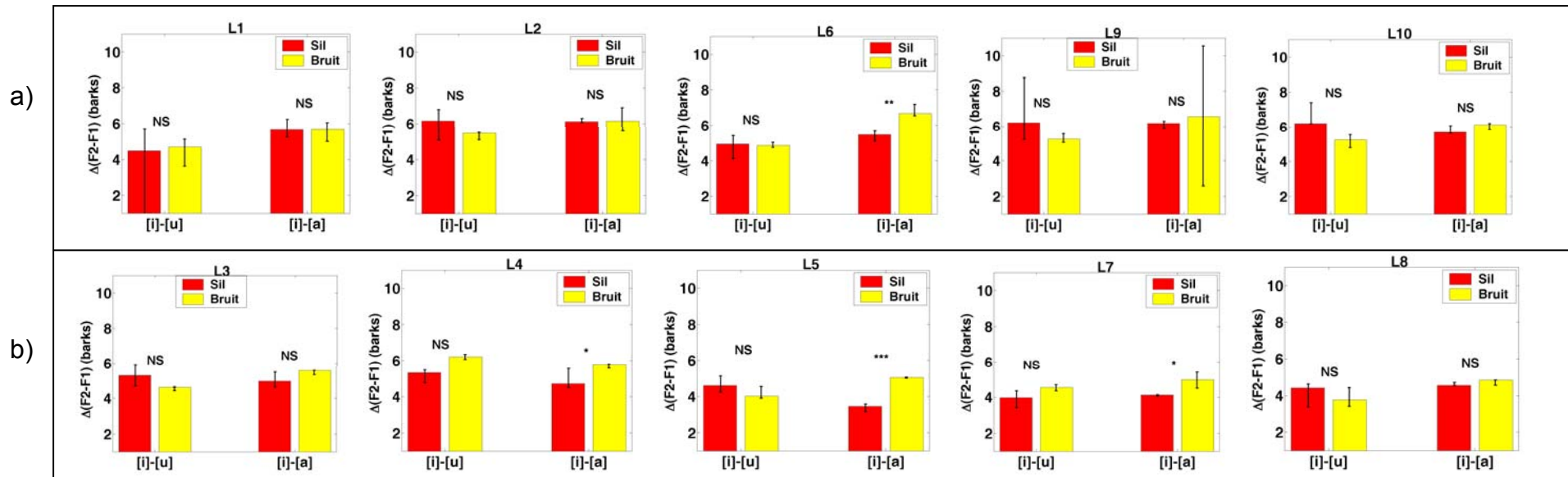


Figure 94. Evolution du silence au bruit du contraste entre [i] et [u] et entre [i] et [a] selon la dimension (F2-F1), exprimée en barks, pour les locutrices (a) et les locuteurs (b) de la BD1.

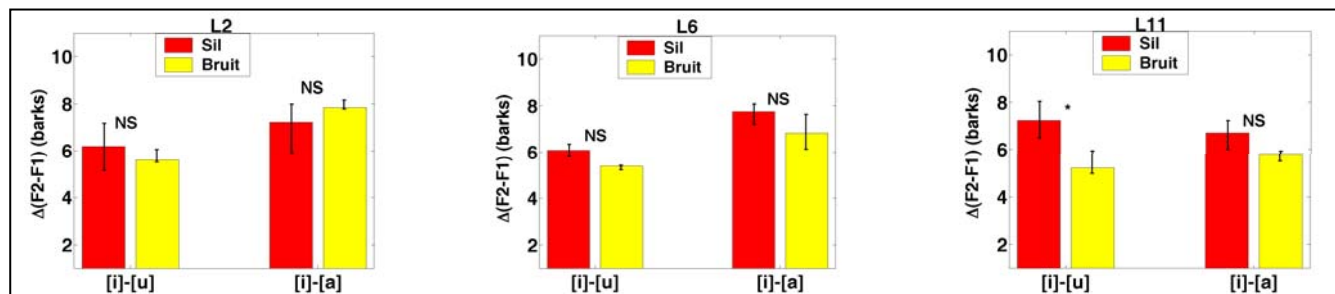


Figure 95. Evolution du silence au bruit du contraste entre les [i] et les [u] et entre les [i] et les [a] selon la dimension (F2-F1), exprimée en barks, pour les locutrices de la BD2

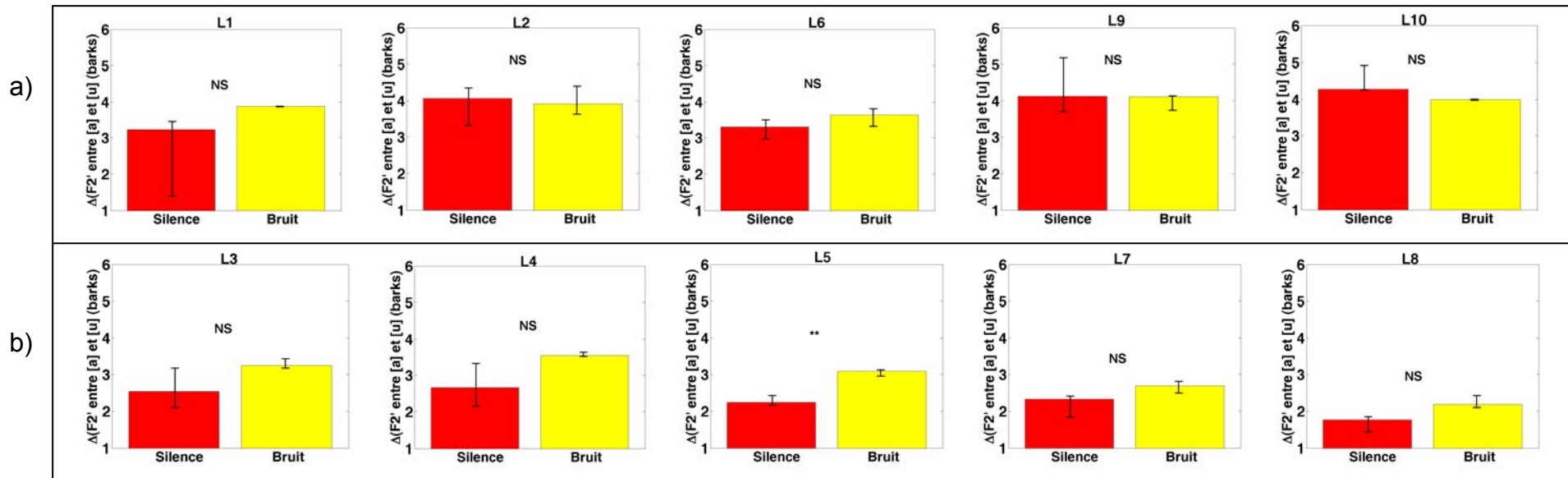


Figure 96. Evolution du silence au bruit de la différence entre [a] et [u] de $F2'=(F1+F2)/2$, exprimée en barks, pour les locutrices (a) et les locuteurs (b) de la BD1.

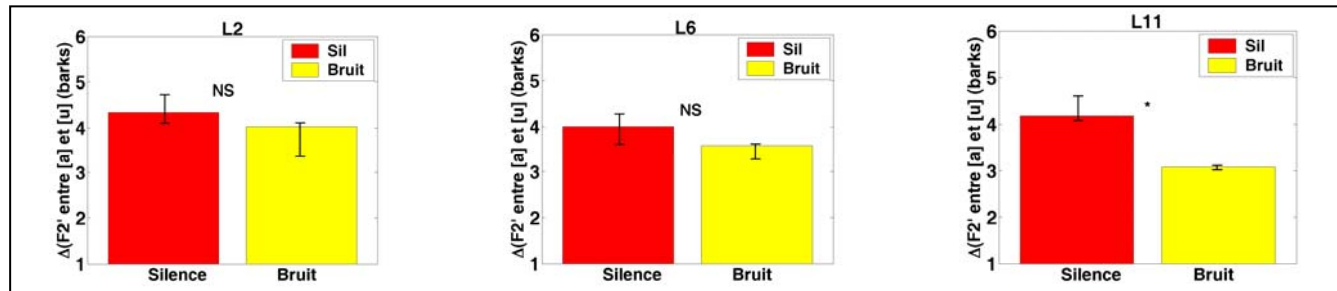


Figure 97. Evolution du silence au bruit de la différence entre [a] et [u] de $F2'=(F1+F2)/2$, exprimé en barks, pour les locutrices de la BD2

1.1.3. Distinction des voyelles [u] par rapport aux voyelles [a] et [i]

Enfin, nous avons examiné la distinctivité des voyelles [u] par rapport aux autres voyelles [a] et [i]. La voyelle [u] est la plus « sombre » du système vocalique et se caractérise principalement par un barycentre spectral très bas en fréquences. Ses deux premiers formants sont assez proches et intégrés perceptivement en un formant effectif $F2'$ assez bien approximé par la moyenne entre $F1$ et $F2$, en barks. Les voyelles [a] montrent également deux premiers formants assez rapprochés, mais dont la moyenne est beaucoup plus haute que celle du [u] (cf. Figure 91). Nous avons donc choisi d'examiner comment évolue la distinctivité dans le bruit des voyelles [a] et [u] vis à vis de ce formant effectif $F2'_{\text{barks}}$.

Les Figure 96 et Figure 97 synthétisent les résultats obtenus en représentant l'évolution du silence au bruit de la différence de $F2'_{\text{barks}}$ entre les voyelles [a] et [u]. Cette différence est toujours très significative pour tous les locuteurs aussi bien dans le silence que dans le bruit. Elle reste constante dans le bruit pour toutes les femmes, excepté L11 où elle diminue significativement. Au contraire, elle tend à légèrement augmenter dans le bruit chez les hommes, en particulier pour L5.

Ce résultat indique donc que si $F2'$ est un indice de « gravité » des voyelles [u] par rapport aux autres voyelles, celui-ci n'est pas renforcé dans le bruit. Ce résultat reste malgré tout extrêmement positif car il montre que cet indice est conservé dans le bruit, ce qui était loin d'être évident étant donné l'augmentation importante de $F1$ et de $F2$ (donc de $F2'$) dans le bruit pour les voyelles [u]. Cet indice n'est donc pas conservé par abaissement de $F2'$ sur les [u], mais au contraire par une plus grande augmentation sur les [a].

1.2. Evolution du système vocalique dans son ensemble

Jusqu'à maintenant, nous avons examiné la distinctivité des extrema [a], [i] et [u] du système vocalique selon des dimensions isolées ($F1-F0$, $F2-F1$, $F2'$). Nous allons maintenant examiner comment les extrema du système vocalique évoluent les uns par rapport aux autres dans l'espace plus global de leurs 3 premiers formants vocaliques, et non plus selon des dimensions isolées.

Les Figure 98 et Figure 99 représentent les voyelles [a], [i] et [u] produites par les locuteurs et locutrices de la BD1 et de la BD2 dans le silence et dans le bruit cocktail respectivement à 85dB et 86dB, dans l'espace $(F1-F0)*F2*F3$ en barks. Nous avons calculé pour chaque locuteur l'aire de ce triangle vocalique dans l'espace, pour la condition de silence et pour la condition de bruit. Nous avons également calculé pour chaque condition la dispersion des voyelles [a], [i], [u] autour de leur barycentre (inertie intra-classe) ainsi que la distance entre ces barycentres (inertie inter-classes). Le rapport de ces deux inerties nous a servi à estimer la distinctivité des différentes voyelles [a], [i] et [u] dans l'espace des indices acoustiques ($F1-F0$), $F2$ et $F3$.

Ces différentes analyses montrent que l'aire du triangle vocalique :

- augmente dans le bruit pour les locuteurs L1, L3, L4, L5 et L7.
- reste constante pour L6 (dans BD1), L9 et L8
- diminue pour L2, L6 (dans BD2), L10 et L11.

Nous n'observons donc pas de stratégie systématique d'expansion du système vocalique dans le bruit, ni d'ailleurs de réduction systématique comme dans plusieurs études antérieures (Junqua 1992^[173] ; Rostolland 1982^[289]). Par ailleurs, dans les cas où l'aire du triangle vocalique augmente dans le bruit, on observe que cela ne provient pas d'un élargissement du triangle au niveau de ses 3 sommets, mais principalement par un étirement au niveau du [a] et au contraire souvent un rétrécissement de la distance entre [i] et [u].

Si l'on prend en compte l'aspect de dispersion ou de concentration de chaque groupe de voyelles dans l'espace $(F1-F0)*F2*F3$, on observe que :

- la dispersion des voyelles [a] reste constante avec le bruit pour la majorité des locuteurs, exceptés L1, L2 et L6 pour lesquelles la dispersion augmente très légèrement, et L5 et L8 pour lesquels la dispersion diminue très légèrement.
- la dispersion des voyelles [i] reste constante avec le bruit pour la majorité des locuteurs, exceptés L2 (dans BD2), L3, L4 et L8 pour lesquels la dispersion diminue très légèrement, et L7 pour lequel la dispersion augmente très légèrement.
- la dispersion des voyelles [u] diminue de façon notable pour L1, L9, L3, L4 et L6 (dans BD2) et légèrement pour L2, L6 (dans BD1), L10 et L7. Elle augmente au contraire pour L8, et reste constante dans le reste des cas.

Par conséquent, la distinctivité auditive des voyelles [a], [i] et [u] tend à augmenter dans le bruit pour les locuteurs L1, L2, L9, L3, L4, L5, L8 conservée par les locuteurs L10 et L7, et diminuée par les locutrices L6 et L11

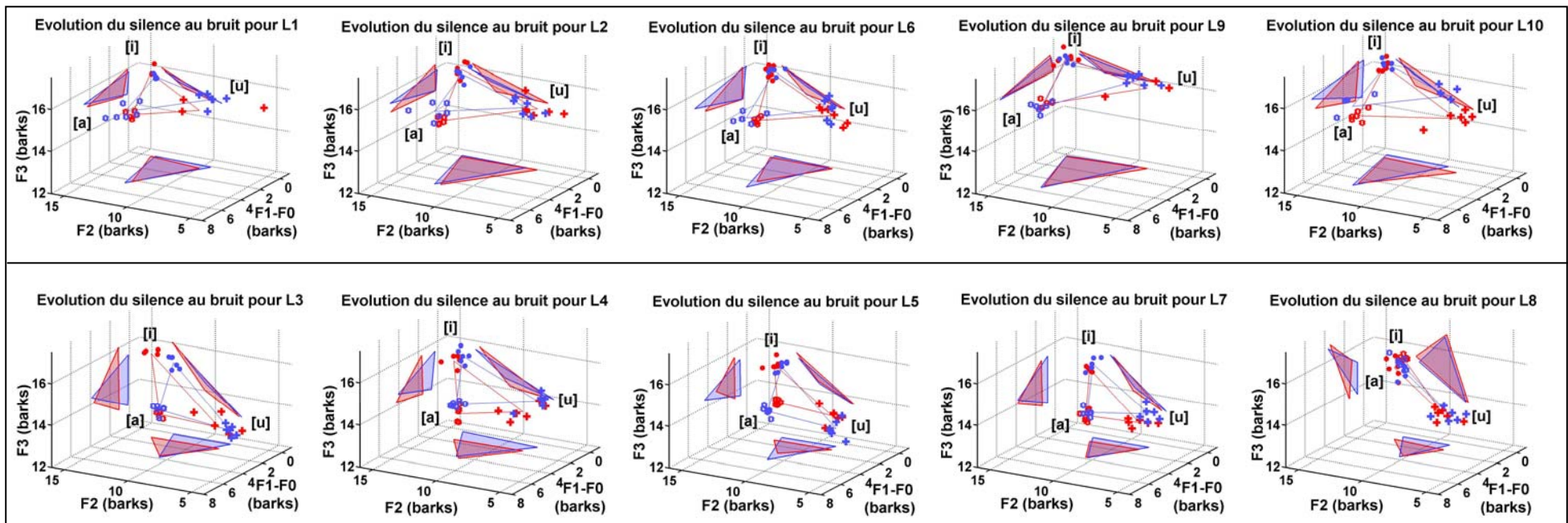


Figure 98. Représentation des voyelles [a], [i] et [u] de chaque locuteur et locutrice de la BD1 dans l'espace formantique $(F1-F0)*F2*F3$, dans le silence (en rouge) et dans le bruit (en bleu clair). Les fréquences formantiques sont considérées comme des indices perceptifs. C'est pourquoi elles sont exprimées en barks.

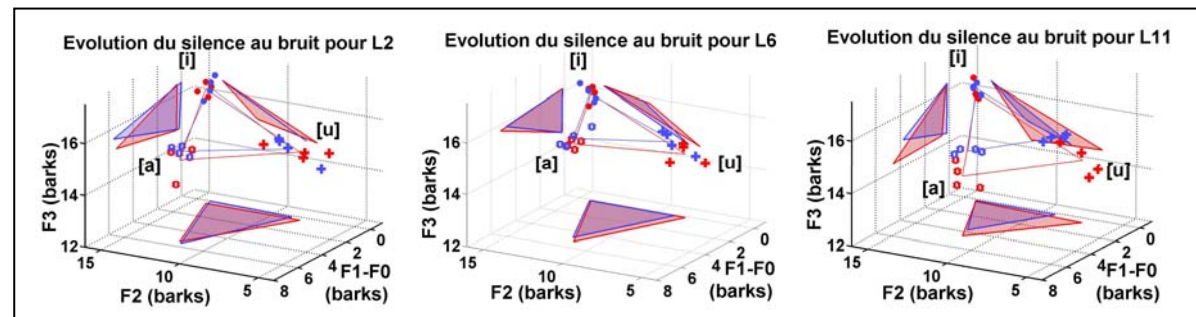


Figure 99. Représentation des voyelles [a], [i] et [u] de chaque locutrice de la BD2 dans l'espace formantique $(F1-F0)*F2*F3$, dans le silence (en rouge) et dans le bruit (en bleu clair). Les fréquences formantiques sont considérées comme des indices perceptifs. C'est pourquoi elles sont exprimées en barks.

1.3. Conclusion quant à la distinctivité acoustique des voyelles

De façon générale, le système vocalique ne montre pas d'expansion dans le bruit par un prototypage de chaque extrémité [a], [i], [u]. L'évolution du triangle vocalique consiste surtout en un étirement vis à vis de la dimension (F1-F0), accompagné chez certains locuteurs d'une compression selon la dimension F2 par l'abaissement du deuxième formant des [i], par l'augmentation du F2 des [u], ou les deux à la fois.

Cette observation générale se traduit par :

- la conservation ou le renforcement de la distinctivité des [a] par rapport aux [i] et [u], selon la dimension (F1-F0).
- la conservation ou le renforcement de la distinctivité des [i] par rapport aux [a] selon la dimension (F2-F1) et plutôt une diminution par rapport aux [u].
- la conservation de la distinctivité entre les [u] et les [a] selon la hauteur de leur F2', voire une légère augmentation de cette distinctivité chez les hommes.

Ces résultats complètent ceux de Rostolland 1982 ^[289] notant que le 2^{ème} formant évolue avec l'effort vocal de façon différente selon les voyelles, tout en conservant une distinctivité entre elles, tandis que le 3^{ème} formant évolue au contraire dans le sens d'une uniformisation de toutes les voyelles selon ce paramètre, à forte intensité vocale.

L'allongement du triangle vocalique dans le bruit se traduit par une réduction de son aire, lorsque la réduction selon la dimension F2 est supérieure à son étirement selon la dimension (F1-F0), par une conservation lorsque les deux influences se compensent, et par une augmentation lorsque l'augmentation du 1^{er} formant est très importante. Il n'y a donc pas de stratégie systématique d'expansion du système vocalique dans le bruit comme dans le cas de la parole « claire » (Lindblom et al. 1992 ^[219]). De même, on observe que la distinctivité globale du système vocalique augmente pour certains locuteurs, est conservée pour d'autres, et diminue pour d'autres encore. On observe néanmoins que quelle que soit son évolution du silence au bruit, la distinctivité du système vocalique reste toujours importante dans le bruit, excepté pour la locutrice L11 dont la distinctivité descend beaucoup plus bas que les autres. Et quelle que soit son évolution, cette distinctivité atteint des valeurs comparables dans le bruit pour la majorité des locuteurs. Les locutrices L2 et L9 et le locuteur L3 se démarquent cependant par une distinctivité beaucoup plus importante que les autres dans le bruit.

Pour revenir à notre hypothèse, plusieurs locuteurs et locutrices montrent une adaptation de la parole dans le bruit allant dans le sens d'un renforcement de la distinctivité auditive du système vocalique. Cet effet provient principalement de l'augmentation du 1^{er} formant et d'une meilleure distinctivité des voyelles [a] par rapport aux voyelles [i] et [u]. Cette augmentation de la distinctivité peut participer à l'intelligibilité accrue de la parole Lombard, observée en comparaison de la parole produite dans le silence (Junqua 1993 ^[171]). Cependant, elle semble davantage la conséquence indirecte de l'augmentation du 1^{er} formant, liée à l'augmentation de l'intensité vocale ou de la fréquence fondamentale, que le résultat d'une stratégie particulière.

Cependant, on peut interpréter ces résultats d'une autre manière, en envisageant que toutes les modifications acoustiques et articulatoires observées dans le bruit ne sont pas nécessairement les indicateurs de stratégies visant à augmenter l'intelligibilité. En effet, il est possible que la mise en place d'une stratégie de communication fasse gagner en intelligibilité d'un côté, mais perturbe la production de parole d'un autre côté. Ainsi, l'augmentation globale de l'intensité s'accompagne d'une augmentation de l'ouverture labiale et du 1^{er} formant vocalique. En soi, ces modifications peuvent perturber la production de certaines voyelles, telles que les [u]. De la même façon que dans l'expérience du « Lip Tube », où une ouverture des lèvres était imposée au locuteur pendant la réalisation de voyelles [u] (Savariaux et al. 1995 ^[301]), il est possible que les locuteurs adoptent dans le bruit des stratégies acoustiques et articulatoires, non pas pour renforcer l'intelligibilité, mais pour compenser la perturbation de la production induite par la nécessité de parler plus fort. Autrement dit, il est possible que l'adaptation du locuteur résulte d'un compromis entre le gain d'intelligibilité obtenu par l'augmentation de l'intensité vocale et la perte d'intelligibilité induite par l'augmentation de l'ouverture de la bouche. Des études perceptives ont montré que la parole forte ou la parole Lombard étaient plus intelligibles que la parole « normale », mais qu'elles devenaient au contraire moins intelligibles lorsque l'intensité vocale atteignait le niveau du cri (Rostolland et al. 1973 ^[290]). Ce phénomène pourrait être le signe que le locuteur n'arrive plus à réaliser de compromis entre l'intensité vocale et l'articulation des voyelles au-delà d'un certain seuil.

Vues sous cet angle, les évolutions du système vocalique que nous avons observées précédemment représentent un très bon compromis puisque malgré la réduction des voyelles [i] et [u] chez certains locuteurs, la distinctivité globale du système est conservée, voire même augmentée dans certains cas.

Ces différentes considérations motivent à examiner plus en détail si l'augmentation de l'intelligibilité de la parole Lombard provient réellement de l'étirement du triangle vocalique selon la dimension (F1-F0), où si l'intelligibilité des unités phonétiques est plutôt diminuée, mais compensée par d'autres aspects de timbre, de fréquence fondamentale ou de segmentation syllabique plus marquée (cf. chapitre 8). En effet, les tests perceptifs antérieurs sur la parole Lombard ont été réalisés à partir de listes de mots, et non à partir de logatomes ou de voyelles isolées. Aussi, rien ne prouve que l'augmentation de l'intelligibilité observée soit due à une meilleure distinctivité des unités phonétiques. Il serait donc très intéressant de mener des tests perceptifs à partir de voyelles isolées ou des logatomes de la 2^{ème} base de données (où seule une voyelle change entre chaque mot), de façon à examiner de façon ciblée si les voyelles produites dans le bruit sont effectivement plus distinctes, ou si l'augmentation de l'intelligibilité dans le bruit provient du renforcement d'autres indices ou d'autres aspects de la parole.

Par ailleurs, nous n'avons étudié ici que les extrema du système vocalique. La distinctivité entre ceux-ci est bien sûr beaucoup plus importante qu'entre deux catégories vocaliques voisines dans l'espace acoustique. Il serait donc nécessaire de compléter cette première exploration en considérant les autres voyelles intermédiaires du système vocalique.

Enfin, la parole « claire » ne montre pas seulement une expansion du système vocalique mais également un renforcement de l'intensité des sons obstruents, en particulier des plosives (Picheny et al. 1985^[269]). Pendant la thèse, nous avons également mené une étude pilote pour étudier l'évolution dans le bruit des caractéristiques acoustiques et de durée des occlusives bilabiales voisées et non voisées. Nous n'avons pas présenté cette étude dans le manuscrit, faute de résultat très probant. Cette piste reste néanmoins à explorer plus en détail.

2. RENFORCEMENT DES INDICES VISUELS D'IDENTIFICATION OU DE DISCRIMINATION DES UNITÉS PHONÉTIQUES

La parole ne consiste pas uniquement en la production et la perception d'indices acoustiques, mais également visuels (Schwartz 2004^[307]). Ainsi, la modalité visuelle accroît considérablement l'intelligibilité de la parole, pas seulement pour les personnes malentendantes mais pour tous les auditeurs. Sumbly et al. 1954^[335] ont ainsi montré que l'ajout de la modalité visuelle était équivalente à une augmentation du rapport signal sur bruit de 15dB. L'information visuelle est particulièrement intéressante car elle n'est pas redondante mais complémentaire à l'information acoustique : les caractéristiques acoustiques de la parole les plus facilement détériorées par du bruit ambiant sont justement celles qui sont les plus robustes sur le plan visuel. Inversement, l'intelligibilité auditive des voyelles ouvertes est assez bonne tandis que ces voyelles sont assez confondues visuellement avec les voyelles étirées (Benoit et al. 1994^[24]) Par ailleurs, les indices visuels de la parole apportent plutôt des informations sur le lieu d'articulation, tandis que les indices acoustiques renseignent davantage sur le mode d'articulation (voisé/non voisé, par exemple). La modalité visuelle ne permet pas seulement de percevoir des indices de reconnaissance des unités phonétiques, mais sert également à la détection d'événements (Schwartz et al. 2004^[309], Grant et al. 2000^[121]). Cette perception anticipée du mouvement articulatoire par rapport au son permet d'attirer l'attention sur certains indices acoustiques discriminants comme le voisement préalable à une occlusive par exemple (Schwartz et al. 2004^[309]).

Pour augmenter son intelligibilité, le locuteur peut renforcer certains indices acoustiques mais également certains indices visuels. Ainsi, des tests perceptifs audiovisuels ont montré que la parole Lombard est visuellement plus intelligible que la parole produite dans le silence (Davis et al. 2006^[62]). Parallèlement, l'adaptation de la parole dans le bruit se caractérise par une amplification de certains mouvements articulatoires. Aussi, il est assez naturel de formuler l'hypothèse que les locuteurs adoptent, dans le bruit, des stratégies articulatoires visant à renforcer les indices visuels de reconnaissance ou de discrimination des unités phonétiques. C'est ce que nous allons examiner dans les paragraphes suivants, en considérant tout d'abord l'évolution dans le bruit des indices visuels de discrimination des voyelles, puis celle des indices visuels de reconnaissance des consonnes bilabiales.

2.1. Renforcement du contraste visible entre les voyelles

Les voyelles du Français peuvent être discriminées perceptivement à partir de quatre traits articulatoires :

- le degré d'ouverture du conduit vocal, lié à l'abaissement de la mandibule et à l'ouverture labiale, permettant de distinguer les voyelles ouvertes, telles que [a] des voyelles fermées telles que [i].
- la position de la constriction du conduit vocal par la langue, placée vers l'avant ou l'arrière de la cavité buccale, permettant de distinguer les voyelles antérieures, telles que [y] des voyelles postérieures telles que [u].
- la protrusion et l'arrondissement des lèvres ou au contraire leur étirement, permettant de distinguer les voyelles étirées, telles que [i] des voyelles arrondies, telles que [y].
- la nasalité contrôlée par le degré de couplage des cavités nasale et buccale grâce au vélum.

Parmi ces différents traits, l'ouverture de la mâchoire, l'étirement des lèvres, leur arrondissement ou leur protrusion sont visibles extérieurement. La position de la langue peut être devinée par l'interlocuteur, mais plus difficilement, et les mouvements du voile du palais ne sont pas perceptibles visuellement. C'est pourquoi nous avons choisi d'examiner dans ce paragraphe si les adaptations articulatoires labiales observées dans le bruit allaient dans le sens d'un renforcement de la distinctivité visuelle des voyelles (cf. Figure 100). Nous allons donc en particulier examiner comment évoluent les 3 extréma [a], [i] et [u] de ce système vocalique dans le bruit. Puisque nous nous intéressons à leur discrimination et non à leur évolution isolée, nous examinerons comment évolue le contraste entre ces 3 voyelles selon les dimensions d'ouverture, d'étirement, d'arrondissement ou de protrusion des lèvres.

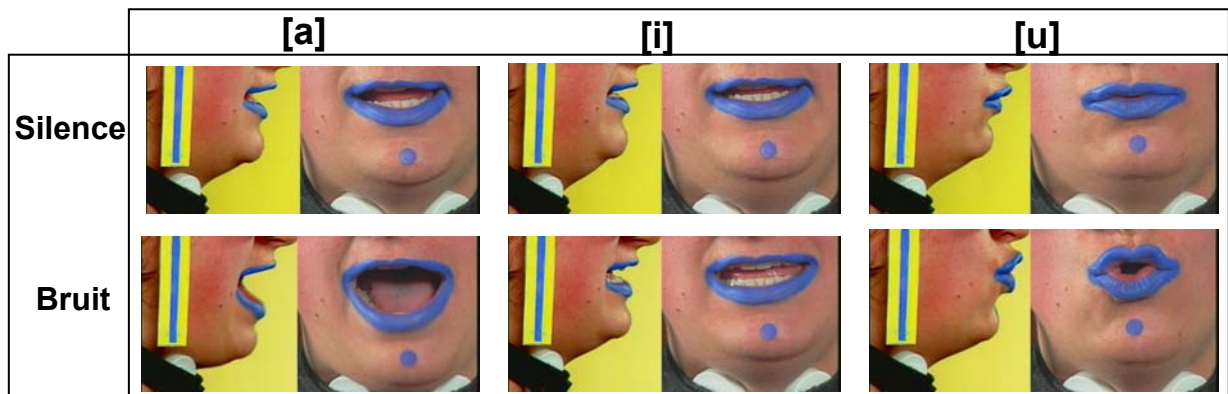


Figure 100. Distinctivité visuelle dans le silence et dans le bruit entre les voyelles ouvertes [a], étirées [i] et arrondies [u].

Nous avons adopté la même méthode dans les 3 paragraphes suivants. Pour chacun des cas, nous avons pris une voyelle V1 ([a] dans le §2.1.1, [i] dans le §2.1.2 et [u] dans le §2.1.3), et un indice articulaire IA permettant de distinguer perceptivement cette voyelle V1 des deux autres (V2 et V3, respectivement [i] et [u] dans le §2.1.1, [u] et [a] dans le §2.1.2, et [i] et [a] dans le §2.1.3). Nous avons mesuré l'indice IA sur les syllabes [la] des logatomes [lala] de la deuxième base de données (BD2), sur les syllabes [lu] des logatomes [lula] et [lalu] et sur les syllabes [li] des logatomes [lila] et [lali]. Au total, nous disposons de 8 mesures de chaque voyelle V1, V2 et V3, par locutrice et par condition de silence ou de bruit cocktail à 85dB.

A partir de ces mesures, nous avons réalisé des tests Anova à un facteur (TYPE VOYELLE) pour vérifier si l'indice IA était significativement différent:

- entre V1 et V2, que ce soit dans le silence ou dans le bruit.
- entre V1 et V3, que ce soit dans le silence ou dans le bruit.

Nous avons également réalisé un autre test Anova à un facteur (BRUIT) pour déterminer

- si le contraste entre V1 et V2 selon l'indice IA évoluait significativement avec le bruit.
- si le contraste entre V1 et V3 selon l'indice IA évoluait significativement avec le bruit.

Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res7. Nous n'en présenterons ici qu'une synthèse.

2.1.1. Distinction des voyelles [a] par rapport aux voyelles [i] et [u]

Nous avons commencé par examiner un indice articulaire labial susceptible de différencier la voyelle [a] des autres voyelles [i] et [u]. La voyelle [a] se caractérise en effet par une ouverture de mâchoire plus importante que les autres voyelles, considérées comme des voyelles fermées. Nous allons donc regarder si l'ouverture labiale B peut permettre de distinguer les voyelles [a] des voyelles [i] et [u] et si cet indice de discrimination est renforcé dans le bruit.

La Figure 101 synthétise les résultats obtenus en représentant l'évolution du silence au bruit de la différence d'ouverture labiale B entre les voyelles [a] et [u] et entre les voyelles [a] et [i].

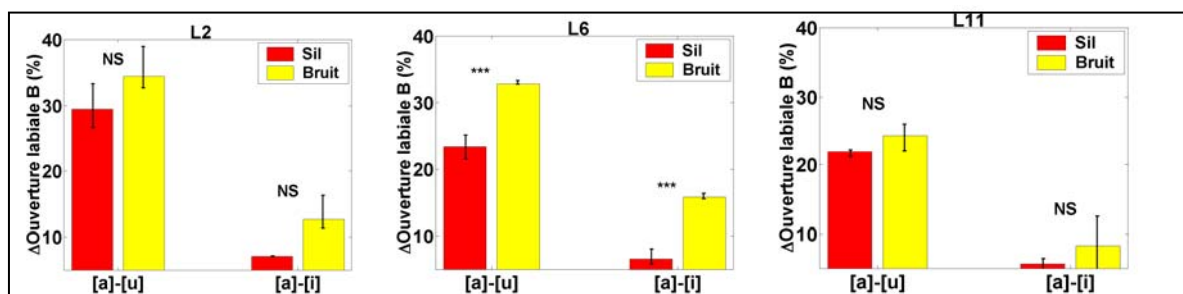


Figure 101. Représentation de la différence d'ouverture labiale B entre les voyelles ouvertes [a] et les autres voyelles protruses [u] et étirées [i], dans le silence et dans le bruit.

La différence d'ouverture labiale entre les voyelles [a] et [u] est toujours très significative pour les 3 locutrices de la BD2 aussi bien dans le silence que dans le bruit. Elle tend à augmenter dans le

bruit mais cela n'est significatif que pour la locutrice L6. En revanche, l'indice B permet moins de discriminer les voyelles [a] et [i], bien que la différence entre ces deux voyelles reste toujours significative pour les 3 locutrices, aussi bien dans le silence que dans le bruit. Ce contraste est renforcé dans le bruit, de façon significative uniquement pour L6.

Ces résultats confirment ceux de Schulman 1989 ^[304] observant que la distinction des voyelles selon la dimension articulatoire d'ouverture/fermeture labiale reste conservée à forte intensité vocale.

2.1.2. Distinction des voyelles [i] par rapport aux voyelles [a] et [u]

Nous avons ensuite examiné un indice articulatoire labial susceptible de différencier la voyelle [i] des autres voyelles [a] et [u]. La voyelle [i] se définit comme une voyelle étirée. Nous allons donc regarder si l'étirement labial A peut permettre de distinguer les voyelles [i] des voyelles [a] et [u] et si cet indice de discrimination est renforcé dans le bruit.

La Figure 102 synthétise les résultats obtenus en représentant l'évolution du silence au bruit de la différence d'étirement labial A entre les voyelles [i] et [u] et entre les voyelles [i] et [a].

La différence d'étirement labial entre les voyelles [i] et [u] est toujours très significative pour les 3 locutrices de la BD2 aussi bien dans le silence que dans le bruit. Elle tend à diminuer dans le bruit, de façon significative pour L2 et L11. Au contraire, l'indice A ne permet pas de discriminer les voyelles [i] et [a], que ce soit dans le silence ou dans le bruit. Leur contraste n'est jamais significatif, ou très peu pour L2 dans le bruit.

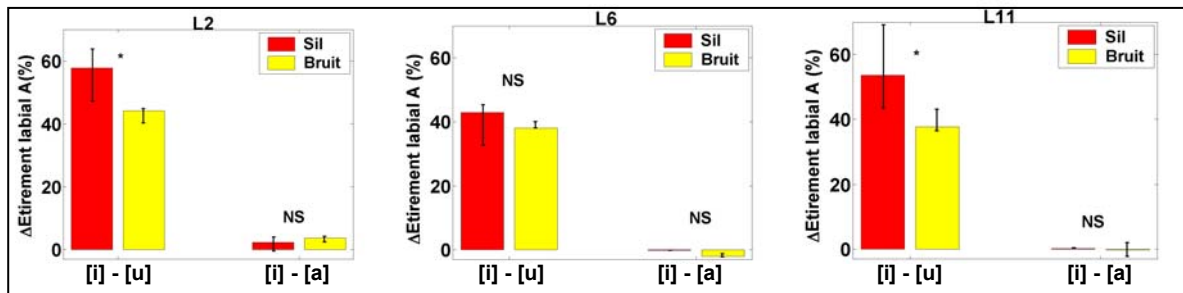


Figure 102. Représentation de la différence d'étirement labial A entre les voyelles étirées [i] et les autres voyelles protruses [u] et ouvertes [a], dans le silence et dans le bruit.

2.1.3. Distinction des voyelles [u] par rapport aux voyelles [a] et [i]

Pour finir, nous avons examiné des indices articulatoires labiaux susceptibles de différencier la voyelle [u] des autres voyelles [a] et [i]. La voyelle [u] est à la fois une voyelle protruse et arrondie, contrairement au [a] et au [i]. C'est pourquoi nous avons choisi d'explorer si la protrusion P1 de la lèvre supérieure et l'arrondi des lèvres AR peuvent permettre de distinguer les voyelles [u] des voyelles [a] et [i] et si ces indices de discrimination sont renforcés dans le bruit.

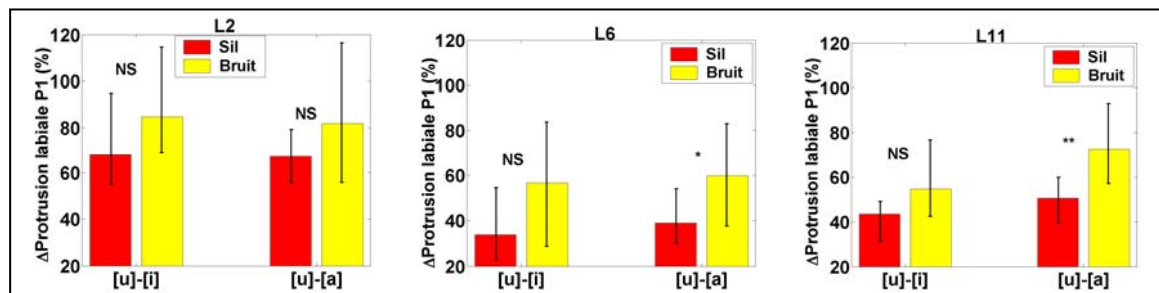


Figure 103. Représentation de la différence de protrusion labiale P1 entre les voyelles protruses [u] et les autres voyelles étirées [i] et ouvertes [a], dans le silence et dans le bruit.

La différence de protrusion labiale entre les voyelles [u] et [i] et entre les voyelles [u] et [a] est toujours très significative pour les 3 locutrices de la BD2 aussi bien dans le silence que dans le bruit. Elle tend à augmenter dans le bruit pour toutes les locutrices, mais de façon significative uniquement pour L6 et L11 entre les voyelles [u] et [a] (cf. Figure 103). Cependant, les effets sont en moyenne non négligeables, de 11 à 22% de la protrusion maximale. L'absence de significativité

semble donc provenir dans certains cas de l'effectif restreint de nos données (8 occurrences de chaque voyelle pour chaque condition). Par ailleurs, certaines augmentations ne sont pas significatives dans le bruit mais la valeur moyenne du contraste est déjà très importante dans le silence : supérieur à 60% pour L2, et supérieur à 40% pour L11. Enfin, nous avons vu au chapitre 5 que la protrusion a tendance à augmenter dans le bruit sur les voyelles [u], mais de façon non significative. Aussi, on aurait pu s'attendre à ce que les voyelles soient moins bien discriminées dans le bruit selon cet indice. Les résultats que nous obtenons ici sont donc très intéressants car ils montrent que la discrimination des voyelles selon leur protrusion labiale peut être renforcée dans le bruit, non pas par un renforcement de la protrusion des voyelles [u], mais surtout en réduisant significativement la protrusion des autres voyelles.

La différence d'arrondi labial entre les voyelles [u] et [i] et entre les voyelles [u] et [a] est toujours très significative pour les 3 locutrices de la BD2 aussi bien dans le silence que dans le bruit. Elle tend à augmenter dans le bruit pour toutes les locutrices, de façon significative pour L2 entre les voyelles [u] et [i], et entre les voyelles [u] et [a] pour L6 et L11 (cf. Figure 104). Ces résultats confirment ceux de Schulman 1989^[304] observant que la distinction des voyelles selon la dimension articulaire d'étirement/arrondi labial reste conservée à forte intensité vocale.

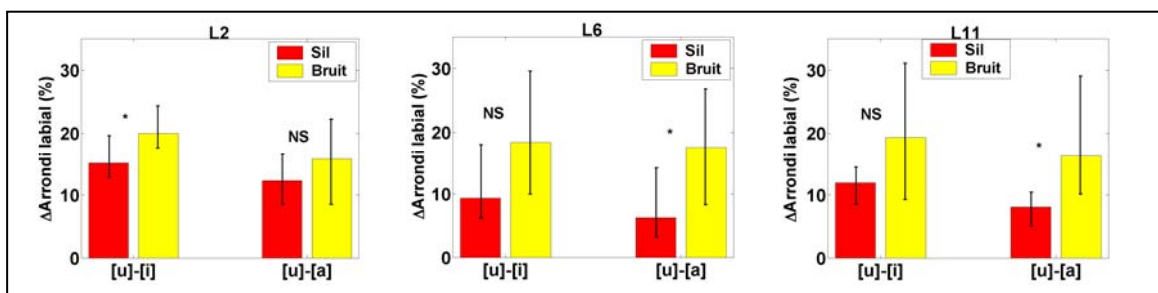


Figure 104. Représentation de la différence d'arrondi labial AR entre les voyelles arrondies [u] et les autres voyelles étirées [i] et ouvertes [a], dans le silence et dans le bruit.

2.2. Renforcement des indices visuels de reconnaissance des consonnes bilabiales

Lors de l'exploration de la deuxième base de données, nous avons également remarqué que les locutrices n'avaient pas seulement tendance à hyper-articuler dans le bruit en ouvrant davantage les lèvres, mais également en les compressant davantage lors de leur fermeture sur des consonnes bilabiales ([m], [p] et [b]). Nous avons donc voulu explorer plus en détail ce pincement des lèvres, pouvant être considéré comme un indice visible de reconnaissance des consonnes bilabiales.

Pour cela, nous avons mesuré le pincement labial (défini en Annexe Met1) sur les syllabes [pa] des logatomes [lapa] et [pala] de la deuxième base de données (BD2), sur les syllabes [ma] des logatomes [lama] et [mala] et sur les syllabes [ba] des logatomes [laba] et [bala]. Cela représente 24 mesures de pincement, toutes consonnes bilabiales confondues, par locutrice et par condition de silence ou de bruit cocktail à 85dB. Nous avons réalisé des tests Anova à un facteur (BRUIT) pour déterminer si le pincement des lèvres évoluait significativement avec le bruit. Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res7.

La Figure 105 synthétise les résultats obtenus en représentant l'évolution du silence au bruit du pincement des lèvres sur les consonnes bilabiales de la deuxième base de données. On observe effectivement que les toutes les locutrices renforcent significativement cet indice dans le bruit, de 7% (pour L2) jusqu'à 26% (pour L6).

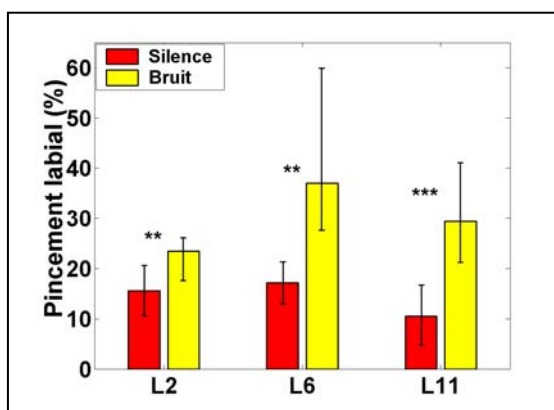


Figure 105. Amplitude maximale du pincement des lèvres sur les consonnes bilabiales [m], [b] et [p] produites dans le silence et dans un bruit de cocktail-party à 85dB, par les 3 locutrices de la deuxième base de données (BD2)

2.3. Conclusion quant à la distinctivité visuelle des segments

Les résultats obtenus au niveau des indices visuels de la parole sont beaucoup plus marqués que ceux observés au niveau formantique. On observe une tendance chez toutes les locutrices à renforcer la distinctivité visuelle dans le bruit entre les voyelles [a] et [i], entre les [a] et les [u] et entre les [i] et les [u] au niveau de la différence d'ouverture labiale, de la différence de protrusion labiale et de la différence d'arrondi labial entre ces voyelles. En revanche, la distinctivité visuelle des voyelles a tendance à diminuer dans le bruit selon la dimension d'étirement labial. On observe également un renforcement dans le bruit de l'amplitude de pincement des lèvres sur les consonnes bilabiales pour toutes les locutrices.

Ces résultats sont très intéressants pour plusieurs raisons :

- Le renforcement dans le bruit de certains mouvements articutoires tels que l'arrondissement labial peut difficilement être expliqué comme la conséquence d'une stratégie acoustique d'augmentation de l'intensité ou de renforcement de l'énergie spectrale dans les médiums. Il semble donc que l'augmentation de la distinctivité visuelle dans le bruit ne soit pas due au hasard mais témoigne de l'existence d'une stratégie d'adaptation dans le bruit visant à faciliter la reconnaissance visuelle des unités phonétiques pour l'interlocuteur.
- La distinctivité des voyelles n'est pas toujours augmentée par l'amplification du geste articutoire permettant de reconnaître une catégorie de voyelles, mais au contraire en renforçant le geste opposé sur les autres catégories. Ainsi, on observe un léger renforcement dans le bruit de la protrusion sur les voyelles [u], mais surtout une ouverture et un étirement plus important sur les [a] et les [i], induisant un recul important de la lèvre supérieure vers l'arrière. Cela peut être expliqué également par le fait que les voyelles arrondies/protruses sont les plus intelligibles visuellement (Benoit et al. 1994 ^[24]). Aussi, leur trait articutoire n'a pas besoin d'être beaucoup renforcé pour améliorer leur identification, contrairement aux voyelles ouvertes.
- Ces résultats peuvent également être interprétés du point de vue de la complémentarité entre les indices acoustiques et visuels de la parole. Les voyelles [u] sont en effet les plus perturbées sur le plan acoustique par l'augmentation de l'effort qui provoque une augmentation de leur ouverture labiale, de leurs deux premiers formants. Nous avons évoqué au paragraphe 1.3 l'idée que l'adaptation de la parole dans le bruit puisse consister en un compromis entre le gain d'intelligibilité apporté par l'augmentation de l'intensité vocale et la perte d'intelligibilité qu'elle peut induire par la déformation des indices formantiques. Le renforcement de la distinctivité visuelle pourrait participer à ce compromis en compensant au niveau visuel, grâce à la complémentarité des deux modalités, la dégradation des indices acoustiques de reconnaissance des voyelles. A l'inverse, l'augmentation importante du 1^{er} formant sur les voyelles [a] pourrait contribuer à améliorer la discrimination auditive entre les voyelles ouvertes et étirées, qui n'est en revanche pas très importante, ni très renforcée dans le bruit, dans le domaine visuel.
- Enfin, par rapport à notre problématique de caractérisation du forçage vocal, l'existence d'une telle stratégie audiovisuelle de communication dans le bruit est très intéressante car elle pourrait décharger le larynx d'une partie de l'effort dépensé pour être intelligible en le « délocalisant » sur les articulateurs. Cependant, nous devons être prudents quant aux bénéfices d'une telle stratégie pour la santé vocale. En effet, d'après son expérience de la rééducation vocale, Guerin 2004 ^[128] avance que l'enchaînement de mouvements articutoires très différents pourrait induire des problèmes d'adaptation de l'impédance ramenée sur le larynx, et finalement mettre le larynx en situation de faiblesse. De même, les chanteurs lyriques cherchent à articuler les voyelles sur un même « placement », sans que l'on sache si cela est effectué à des fins esthétiques ou de protection du larynx. Il serait donc très utile de clarifier l'influence d'un renforcement de l'articulation sur la charge supportée par le larynx, afin de déterminer si une telle stratégie de communication peut être la cause de troubles vocaux ou au contraire les prévenir.

Après avoir observé ces renforcements de la distinctivité visuelle des voyelles en production, il serait maintenant nécessaire de mener des tests perceptifs afin de déterminer s'ils sont effectivement utilisés en perception par les interlocuteurs. De plus, nous espérons pouvoir prochainement examiner si cette stratégie de renforcement de la distinctivité visuelle des voyelles est adoptée dans toutes les situations d'exposition au bruit ou si elle est particulièrement renforcée lorsque les interlocuteurs disposent de la modalité visuelle pour communiquer.

3. ALLONGEMENT DU TEMPS DE PRODUCTION OU DE RECUPERATION DE CES INDICES

Nous avons vu dans le chapitre 5 que l'adaptation de la parole dans le bruit s'accompagne pour la majorité des locuteurs d'un ralentissement du débit de parole. Celui-ci a souvent été relié dans la littérature à une expansion du système vocalique dans le plan $F1 \times F2$ et à une amplification des gestes articulatoires atteignant davantage leur cible (Lindblom 1963^[216]; Picheny et al. 1986^[270]). Il est donc envisageable dans le cas de la parole produite en environnement bruyant que la durée syllabique soit allongée de façon à donner le temps au locuteur de mieux réaliser les indices acoustiques et/ou visuels de reconnaissance des phonèmes. C'est pourquoi nous testerons dans le premier paragraphe si l'amplification des mouvements articulatoires dans le bruit est corrélée à l'allongement syllabique.

Cependant, on peut envisager une autre hypothèse où l'allongement de la durée ne servirait pas tant au locuteur pour produire ces indices qu'à l'interlocuteur pour augmenter ses chances de les percevoir. Aussi, nous examinerons dans le deuxième paragraphe comment le ralentissement global du débit de parole influence en particulier la durée des cibles vocaliques, durant lesquelles les caractéristiques articulatoires et spectrales des voyelles, servant d'indices pour leur reconnaissance, sont stables. De même, nous examinerons si la durée des transitions vocaliques ou si le temps de pincement des lèvres sur les bilabiales sont allongés dans le bruit.

3.1. Un ralentissement lié à l'amplification des gestes articulatoires

Nous avons donc examiné la potentielle corrélation entre la durée syllabique et l'amplitude des mouvements articulatoires labiaux. Pour cela, nous avons tracé dans un même plan les valeurs de durée des syllabes en fonction des valeurs de l'ouverture labiale B (cf. Figure 106). Ces valeurs ont été mesurées dans la BD2 sur la syllabe du déterminant [la] et sur les 1^{ères} ou 2^{èmes} syllabes [la] des 17 logatomes cibles prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 175 mesures par locutrice. Nous avons estimé la corrélation entre l'ouverture labiale et la durée syllabique à partir du coefficient de régression linéaire de ces données. Celui-ci est indiqué sur chaque figure et montre une assez faible corrélation entre l'ouverture labiale des syllabes [la] et leur durée, pour toutes les locutrices de la BD2.

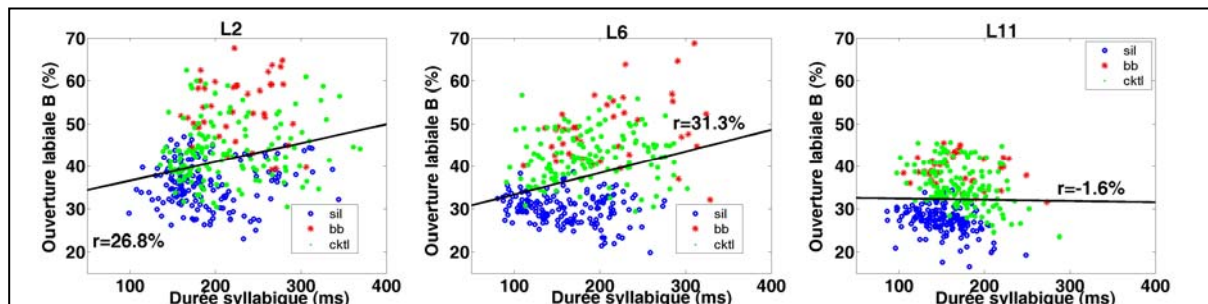


Figure 106. Représentation pour les 3 locutrices de la BD2 de leur ouverture labiale en fonction de la durée syllabique pour toutes les syllabes [la] prononcées dans le silence, dans le bruit blanc et dans le bruit cocktail.

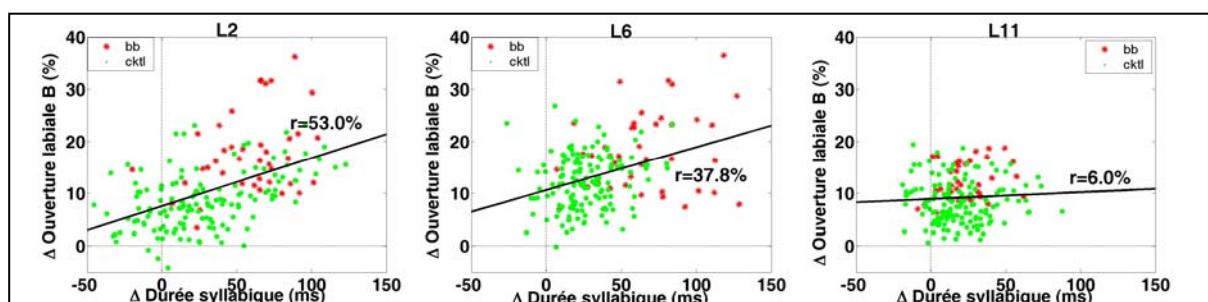


Figure 107. Représentation pour les 3 locutrices de la BD2 de l'évolution du silence au bruit blanc et au bruit cocktail de leur ouverture labiale en fonction de l'évolution de la durée syllabique pour toutes les syllabes [la].

Nous avons examiné si l'évolution de l'ouverture labiale du silence au bruit était davantage corrélée à l'évolution de la durée syllabique dans le bruit. Nous avons donc cette fois tracé dans un même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence (cf. Figure 107) et rapporté le coefficient de régression linéaire de ces données sur chaque figure.

Nous remarquons tout d'abord que l'ouverture labiale et la durée syllabique n'augmentent conjointement dans le bruit de façon systématique que pour la locutrice L6. Les locutrices L2 et L11 montrent au contraire de nombreux cas où la durée syllabique diminue bien que l'ouverture labiale augmente (cf. Tableau 28). Cela explique en partie les faibles corrélations de nouveau observées entre l'évolution de la durée syllabique dans le bruit et l'augmentation de l'ouverture labiale.

	durée diminuée	durée augmentée
L2	21.1 %	77.3 %
L6	5.9 %	93.5 %
L11	13.2 %	86.8 %

Tableau 28. Pourcentage de cas où la durée syllabique augmente ou diminue conjointement à une augmentation de l'ouverture labiale dans le bruit, pour chaque locutrice de la 2^{ème} base de données. La somme des deux colonnes ne correspond pas toujours à 100%. Il existe en effet quelques rares cas où l'ouverture labiale diminue avec le bruit. L'effectif de ces cas correspond au pourcentage restant pour atteindre 100%.

Nous avons vérifié si cela était également valable pour le 1^{er} formant vocalique, puisque l'ouverture labiale n'est qu'une approximation de l'ouverture de la mâchoire. Nous avons donc tracé dans un même plan les valeurs de durée des syllabes en fonction des valeurs du 1^{er} formant (cf. Figure 108). Ces valeurs ont été mesurées dans la BD2 sur la syllabe du déterminant [la] et sur les 1^{ères} ou 2^{èmes} syllabes [la] des 17 logatomes cibles prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail, et seulement en position finale de l'énoncé pour la condition de bruit blanc. Cela représente donc au total 175 mesures par locutrice. Nous avons également tracé dans le même plan la différence entre chaque valeur mesurée dans les différentes conditions de bruit et sa valeur correspondante dans le silence (cf. Figure 109). Les coefficients de régression linéaire de ces données sont rapportés sur chaque figure.

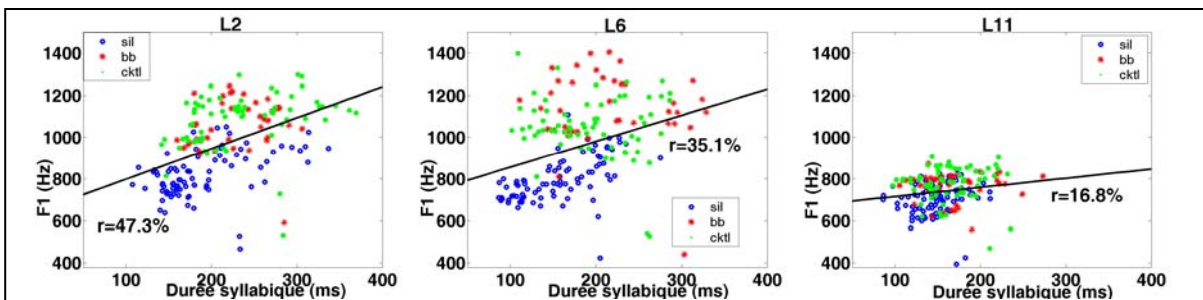


Figure 108. Représentation pour les 3 locutrices de la BD2 de leur ouverture labiale en fonction de la durée syllabique pour toutes les syllabes [la] prononcées dans le silence, dans le bruit blanc et dans le bruit cocktail.

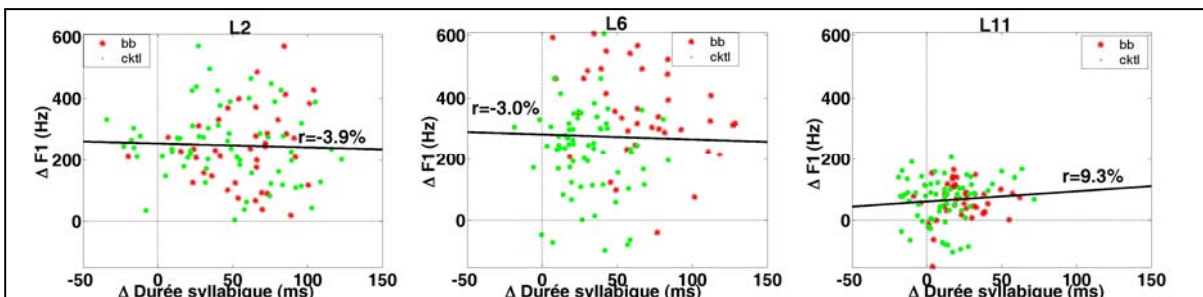


Figure 109. Représentation pour les 3 locutrices de la BD2 de leur ouverture labiale en fonction de la durée syllabique pour toutes les syllabes [la] prononcées dans le silence, dans le bruit blanc et dans le bruit cocktail.

Nous remarquons tout d'abord que le 1^{er} formant et la durée syllabique n'augmentent pas toujours conjointement dans le bruit. En particulier, la locutrice L11 montre de nombreux cas où la durée syllabique diminue bien que le 1^{er} formant augmente (cf. Tableau 29). Par conséquent, il n'est pas étonnant d'observer d'assez faibles corrélations avec le 1^{er} formant.

	durée diminuée	durée augmentée
L2	9.3 %	90.7 %
L6	3.7 %	89.7 %
L11	16.8 %	68.2 %

Tableau 29. Pourcentage de cas où la durée syllabique augmente ou diminue conjointement à une augmentation de l'ouverture labiale dans le bruit, pour chaque locutrice de la 2^{ème} base de données. La somme des deux colonnes ne correspond pas toujours à 100%. Il existe en effet quelques rares cas où l'ouverture labiale diminue avec le bruit. L'effectif de ces cas correspond au pourcentage restant pour atteindre 100%.

3.2. Un allongement de la durée de perception de ces indices

Nous avons alors examiné si l'allongement de la durée syllabique n'était pas l'artefact d'un allongement plus particulier du temps durant lequel les indices acoustiques et articulatoires sont stables et « récupérables » perceptivement par l'interlocuteur, ou du temps des transitions formantiques pouvant également jouer un rôle important dans l'identification des voyelles (Lindblom et al. 1967^[221]) ou des consonnes occlusives Ohde 1988^[259]).

Nous avons segmenté dans la 2^{ème} base de données les phases formantiques stables et les transitions formantiques des consonnes aux voyelles et des voyelles aux consonnes. Les critères de cette segmentation sont présentés en détail en Annexe Met3. Nous nous sommes alors intéressés à l'évolution dans le bruit de la durée des phases formantiques stables sur les voyelles (i.e. la durée des cibles vocaliques) et de la durée des transitions vocaliques. Nous avons mesuré la durée de la cible vocalique pour les 2 voyelles des 17 logatomes cibles de la BD2, prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail à 85dB. De même, nous avons mesuré la durée des transitions formantiques Voyelle-Consonne et Consonne-Voyelle pour les 2 syllabes de ces mêmes logatomes. Cela représente donc au total 68 mesures de ces 2 types de durée, par locutrice et par condition. Nous avons réalisé des tests Anova à un facteur (BRUIT) pour déterminer si la durée des cibles vocaliques évoluait significativement avec le bruit, de même que pour la durée des transitions formantiques. Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res7.

La Figure 110 synthétise les résultats obtenus en représentant l'évolution du silence au bruit de la durée des cibles vocaliques et de la durée des transitions formantiques. On observe effectivement un allongement significatif dans le bruit de la durée durant laquelle les formants d'une voyelle sont stables, pour les locutrices L2 et L11. Pour L2, le ralentissement du débit de parole semble allonger toutes les phases de la syllabes : autant les parties stables que les parties de transitions. Pour L11, le ralentissement du débit touche davantage les cibles vocaliques, mais pas particulièrement les transitions formantiques. Enfin, au contraire, l'allongement syllabique dans le bruit chez L6 sert davantage à renforcer la durée des transitions formantiques que celle des cibles vocaliques. Ces différences interindividuelles peuvent être mises en relation avec l'amplitude des gestes articulatoires, au contraire davantage renforcée dans le bruit pour la locutrice L6 que pour la locutrice L11. Il est donc envisageable que ces deux stratégies d'allongement des cibles ou de leur amplification soient complémentaires, bien que visant toutes les deux à faciliter l'identification des unités phonétiques par l'interlocuteur.

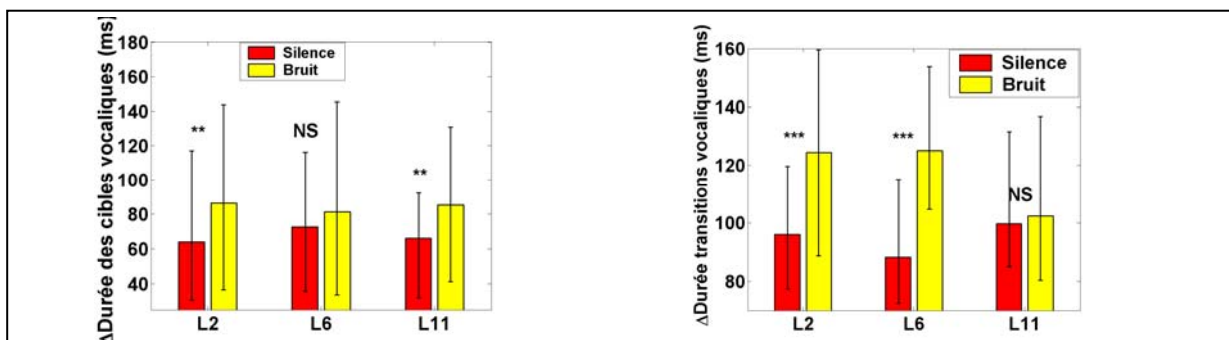


Figure 110. Durée des cibles vocaliques et des transitions formantiques (VC+CV) dans le silence et dans un bruit de cocktail-party à 85dB, pour les 3 locutrices de la deuxième base de données (BD2)

De même, nous avons examiné le temps durant lequel les lèvres sont fermées pour produire une consonne bilabiale. Nous avons mesuré cette durée sur les syllabes [pa] des logatomes [lapa] et [pala] de la deuxième base de données (BD2), sur les syllabes [ma] des logatomes [lama] et [mala] et sur les syllabes [ba] des logatomes [laba] et [bala]. Cela représente 24 mesures, toutes consonnes bilabiales confondues, par locutrice et par condition de silence ou de bruit cocktail à 85dB. Nous

avons réalisé des tests Anova à un facteur (BRUIT) pour déterminer si la durée de pincement des lèvres sur les bilabiales évoluait significativement avec le bruit. Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res7.

La Figure 111 synthétise les résultats obtenus en représentant l'évolution du silence au bruit de la durée de pincement des lèvres sur les consonnes bilabiales. On observe que celle-ci tend à augmenter dans le bruit pour les locutrices L2 et L6, mais cette augmentation reste dans tous les cas non significative, du fait d'un écart-type assez important. Il est possible que cette variabilité provienne du fait que nous avons examiné de façon indifférenciée des occlusives bilabiales [p] et [b] et des bilabiales nasales [m]. Aussi, le pincement des lèvres peut servir à produire un bruit de plosion dans le cas des occlusives, et non pour les [m]. Par ailleurs, le timing articulatoire des occlusives est important puisque la durée de certaines phases (en particulier le délai d'établissement du voisement) constitue un indice d'identification de ces consonnes. Pour toutes ces raisons, il est possible que la durée de pincement n'évolue pas dans le bruit de la même façon pour ces différentes catégories de consonnes bilabiales. Il serait donc intéressant de refaire cette expérience avec davantage d'effectif de chacune des consonnes bilabiales, afin de vérifier sur le plan statistique les tendances que nous avons pu observer ici.

Dans tous les cas, nos résultats vont à l'encontre de ceux de Schulman 1989^[304] observant que les bilabiales intervocaliques sont raccourcies à forte intensité vocale. Cela peut indiquer qu'il existe des différences entre la parole forte et la parole Lombard, ou bien une réorganisation des différentes phases d'une occlusive, avec un allongement de la phase d'occlusion (correspondant au temps de pincement des lèvres) et au contraire un raccourcissement des phases de plosion, et de transition. Pendant la thèse, nous avons mené une première étude exploratoire sur l'évolution de la durée de ces différentes phases d'une occlusive bilabiale, définies de façon acoustique. Nous n'avons pas observé d'évolution significative. De nouveau, il serait utile de mener une expérience dédiée à l'évolution des occlusives bilabiales en environnement bruyant, à la fois sur le plan acoustique et articulatoire, et sur un effectif de données plus important.

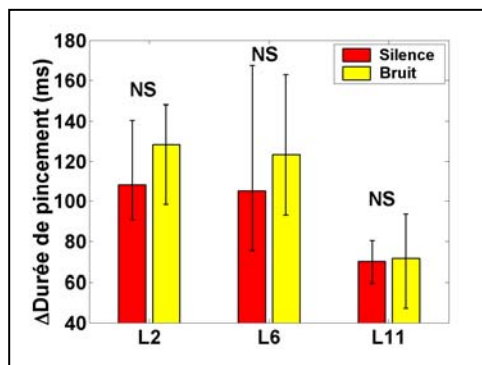


Figure 111. Durée du pincement des lèvres sur les consonnes bilabiales dans le silence et dans un bruit de cocktail-party à 85dB, pour les 3 locutrices de la deuxième base de données (BD2)

3.3. Conclusion quant à l'allongement de la durée des segments

Contrairement à ce qui est classiquement observé dans la littérature, le ralentissement du débit de parole dans le bruit n'est pas entièrement corrélé à l'amplification des gestes articulatoires même s'il l'accompagne. On observe effectivement que cet allongement des syllabes correspond pour 2 locutrices sur 3 à une augmentation de la durée durant laquelle les cibles articulatoires et formantiques sont atteintes, et peuvent être perçues par l'interlocuteur. Pour 2 autres locutrices sur 3, le ralentissement du débit est lié à une augmentation de la durée des transitions vocaliques. Certains auteurs soutiennent l'hypothèse que les indices de perception des voyelles soient davantage dynamiques que statiques, et que les transitions formantiques participent en cela de façon déterminante à l'identification des unités phonétiques (Browman et al. 1989^[35] ; Carre et al. 2000^[44]). De même, certaines études ont montré que les transitions formantiques contribuent à l'identification des consonnes occlusives, et peuvent même suffire en l'absence de plosion (Ohde 1988^[259]). Il est donc envisageable que ces transitions soient renforcées dans le bruit pour compenser le fait que le bruit de plosion est plus susceptible d'être masqué par le bruit.

Dans le cas de la parole claire comme dans celui de la parole Lombard, le ralentissement du débit n'est pas uniforme mais dépend des phonèmes (Kim 2005^[180], Castellanos et al. 1996^[46], Junqua 1993^[171], Picheny et al. 1985^[269]), ce qui argumente également en faveur de l'existence d'une

stratégie de renforcement des caractéristiques acoustiques ou visuelles d'identification des unités phonétiques et de leur temps de récupération par l'interlocuteur.

Enfin, le ralentissement du débit de parole ne consiste pas seulement en un allongement des phonèmes mais également en l'insertion de pauses. Nous examinerons dans le chapitre suivant comment le ralentissement de la parole peut également contribuer à un renforcement des indices prosodiques de structuration de l'énoncé.

4. CONCLUSION DU CHAPITRE 7

Nous avons vu au cours de ce chapitre que l'élévation du premier formant contribue chez certains locuteurs à une meilleure distinctivité auditive des voyelles ouvertes par rapport aux voyelles fermées selon la dimension (F1-F0), pouvant expliquer l'intelligibilité accrue de la parole Lombard rapportée dans des études antérieures (Junqua 1993^[171]). Cependant, la distinctivité acoustique des voyelles tend plutôt à être réduite par rapport au 2^{ème} et au 3^{ème} formant, si bien qu'il semble que les modifications formantiques observées dans le bruit soient davantage à l'augmentation de l'effort vocal qu'à une recherche de « prototypage » des voyelles.

L'adaptation articulatoire au niveau du mouvement des lèvres contribue également à une meilleure distinctivité visuelle des voyelles dans le bruit ainsi qu'à un renforcement des indices de reconnaissance des consonnes bilabiales. Contrairement à l'évolution des formants, toutes les adaptations articulatoires ne semblent pas pouvoir être totalement expliquées par une augmentation de l'effort vocal : c'est le cas en particulier de l'augmentation du contraste visuel selon la dimension de protrusion ou d'arrondi labial, ainsi que de l'augmentation du pincement des lèvres sur les consonnes bilabiales. Il semble donc exister une stratégie de renforcement des indices visuels de reconnaissance et de discrimination des unités phonétiques dans le bruit.

Nous avons proposé dans ce chapitre l'hypothèse que les adaptations formantiques et articulatoires ne soient pas seulement liées à des stratégies de renforcement de l'intelligibilité mais résultent d'un compromis entre le gain d'intelligibilité apporté par l'augmentation de l'intensité vocale et au contraire à la perte d'intelligibilité que cette stratégie principale peut engendrer au niveau de la réalisation des unités phonétiques. Les résultats que nous observons ici semblent montrer que ce compromis peut être trouvé grâce à la complémentarité des indices audio et visuels. Certaines voyelles, telles que le [u] sont davantage affectées que d'autres par l'adaptation acoustique de la parole dans le bruit, mais leur intelligibilité visuelle est au contraire conservée ou renforcée. A l'inverse, certaines voyelles telles que le [a] et le [i] sont autant, voire moins discriminables visuellement dans le bruit que dans le silence, tandis que leur distinctivité acoustique augmente dans le bruit. L'étude de la parole Lombard comme un type de parole compensatoire, non pas en réaction au bruit ambiant, mais aux perturbations qu'induisent la nécessité de parler plus fort, serait intéressante à mener plus en détail dans des travaux futurs. En particulier, il serait nécessaire de prolonger cette première exploration par l'analyse des mouvements linguaux pour déterminer s'il existe des adaptations au niveau de la langue, notamment pour les voyelles arrondies, afin de compenser l'augmentation de l'ouverture de la mâchoire liée à l'augmentation de l'intensité vocale.

Enfin, nous avons également montré dans ce chapitre que le ralentissement de la parole dans le bruit ne semble pas être corrélé à l'amplification des gestes articulatoires. Il contribue en revanche à un allongement du temps durant lequel les indices statiques ou dynamiques d'identification des voyelles ou des consonnes bilabiales peuvent être récupérés par l'interlocuteur. Dans ce travail, nous avons laissé de côté la vitesse des mouvements articulatoires car nous avons observé que les pics de vitesse des mouvements labiaux, reliés à la force des mouvements articulatoires, n'évoluent pas de façon significative avec le bruit (cf. chapitre 5). Il serait cependant intéressant d'examiner plus en détail les courbes de vitesse et d'accélération des mouvements labiaux. Selon Roy 2005^[294], ces paramètres seraient en effet utilisés par l'interlocuteur pour focaliser son attention sur les instants où le locuteur atteint des cibles articulatoires. On peut donc faire l'hypothèse que le locuteur renforce ces indices dans le bruit si le ralentissement du débit de parole est pour lui une stratégie visant à faciliter la récupération des indices phonétiques par l'interlocuteur.

Chapitre 8 :

Une adaptation visant à renforcer des indices de structuration de l'énoncé ?

- 1. INDICES DE FRONTIÈRE SYLLABIQUE**
- 2. INDICES DE FRONTIÈRE ENTRE UN DÉTERMINANT ET UNE UNITÉ LEXICALE**
- 3. INDICES DE DÉBUT DE L'ÉNONCÉ**
- 4. INDICES DE FIN DE L'ÉNONCÉ**
- 5. CONCLUSION DU CHAPITRE 8**

Comme nous l'avons vu dans le chapitre 5, l'adaptation de la parole dans le bruit se traduit entre autres par une augmentation de l'intensité vocale, par un ralentissement du débit de parole, par un élargissement de l'ambitus de F0 et par l'amplification des gestes articulatoires. Dans les chapitres 6 et 7, nous avons examiné en quoi ces différentes modifications de la parole pouvaient être les indicateurs de stratégies de communication visant à émerger acoustiquement du bruit ambiant ou cherchant à faciliter la reconnaissance des unités phonétiques pour l'interlocuteur. Or ces différents aspects acoustiques et articulatoires de la parole sont également reliés à la prosodie. C'est pourquoi nous allons examiner maintenant dans ce chapitre une autre hypothèse selon laquelle l'évolution dans le bruit de ces caractéristiques pourrait être l'indicateur de stratégies de communication visant à améliorer l'intelligibilité en renforçant certains indices prosodiques de la parole, en particulier les indices de structuration de l'énoncé.

La parole n'est en effet pas seulement une succession de sons ou de mouvements articulatoires, mais leur organisation pour former des mots, des syntagmes et des énoncés. Plusieurs théories linguistiques complémentaires tentent d'expliquer la façon dont l'auditeur segmente la parole continue en mots (cf. Wauquier-Gravelines 1999^[371] pour une revue) : d'un côté les théories post-lexicales de la segmentation soutiennent l'hypothèse que l'auditeur se sert de ses connaissances sur le lexique, les règles syntaxiques et phonotaxiques de la langue écoutée pour deviner l'emplacement le plus probable des frontières entre les mots. De l'autre côté, mais pas nécessairement de façon opposée, les théories pré-lexicales de la segmentation postulent l'existence dans le signal de parole d'indices produits par le locuteur pour marquer l'emplacement des frontières entre les mots. Nous nous intéresserons ici à l'hypothèse pré-lexicale de segmentation du flux de parole, afin de poursuivre notre exploration des stratégies d'adaptation de la parole en environnement bruyant, motivée par la recherche d'intelligibilité. C'est pourquoi nous allons dans ce chapitre nous intéresser plus particulièrement aux indices produits par le locuteur pour structurer son énoncé en Français, et examiner si ces indices sont conservés, atténués ou renforcés lorsque le locuteur s'exprime dans le bruit.

Parmi les différents indices de structuration de l'énoncé ayant été mis en évidence dans des études antérieures, l'allongement de la consonne initiale d'un mot de contenu a été montré par Christophe 1993^[52] comme étant un indice de frontière entre la fin du déterminant et le début d'une unité lexicale. De même, Vaissière 1983^[357] et Fougeron et al. 1997^[95] ont montré que la syllabe initiale d'un énoncé a tendance à être allongée et hyperarticulée, ce qui peut être considéré comme un indice de début d'énoncé. Tous les auteurs s'accordent également sur le fait de considérer l'allongement de la syllabe finale d'une unité prosodique comme un indice de fin en Français (Wenk et al. 1982^[378] ; Delattre 1966^[67]), avec un allongement d'autant plus marqué que l'unité en question est de haut niveau (Christophe 1993^[52] ; Bagou et al. 2002^[19]), donc maximal en fin d'énoncé. Cet allongement est également associé à une « chute finale majeure » de la fréquence fondamentale à la fin d'un énoncé déclaratif, correspondant à un ton bas de frontière (Vaissière 1983^[357]). Tabain, 2003 et Lœvenbruck, 1999 ont également observé une hyperarticulation de la syllabe finale d'un énoncé. Enfin, il arrive fréquemment que les frontières syntaxiques importantes de l'énoncé soient marquées par l'introduction de courtes pauses dans le flux de parole (Vaissière 1983^[357] ; Grosjean et al. 1975^[125]).

Nous examinerons donc ici s'il existe dans le bruit des stratégies de renforcement de ces indices acoustiques et articulatoires de frontière syllabique, de frontière entre un déterminant et une unité lexicale, de début et de fin d'énoncé.

1. INDICES DE FRONTIERE SYLLABIQUE

Nous avons vu dans le chapitre 6 que les voyelles sont davantage renforcées dans le bruit par rapport aux consonnes, conformément à d'autres études antérieures sur la parole Lombard (Fairbanks et al. 1957^[87] ; Dohalska et al. 2000^[73] ; Castellanos et al. 1996^[46]). Nous avons envisagé que ce phénomène soit lié à une stratégie de compensation du masquage fréquentiel induit par le bruit ambiant. Cela ne semble pas être le cas. D'autres auteurs interprètent ce renforcement des voyelles dans le bruit ou dans la parole à forte distance de l'interlocuteur, comme le signe que les voyelles portent davantage l'intelligibilité du message que les consonnes. Dohalska et al. 2000^[73] ont également proposé l'idée de « patrons vocaliques » à partir desquels l'interlocuteur pourrait reconstituer le message.

En observant l'allure de la courbe d'intensité vocale en fonction du temps, nous avons été frappés par son aspect relativement « lisse » dans le silence, suivant globalement les syntagmes accentuels de l'énoncé, avec une chute de l'intensité au niveau de leur frontière, et au contraire l'apparition de « lobes » dans le bruit, alignés sur chaque syllabe (cf. Figure 112), et justement liés au renforcement plus important des voyelles que des consonnes. Nous nous sommes alors interrogés sur le rôle de ces lobes sur l'intelligibilité de la parole, et si la chute d'intensité avant une syllabe ne pouvait pas constituer un indice de frontière syllabique qui serait renforcé dans le bruit. La parole produite dans le bruit nous a en effet paru à l'écoute plus « scandée » dans le bruit que dans le silence. Nous avons même observé que des locuteurs inséraient une courte pause entre les différentes syllabes d'un mot, dans certains niveaux de bruit très intenses ou lors de la répétition d'un mot incompris par l'interlocuteur.

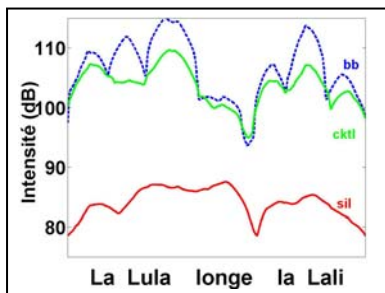


Figure 112. Représentation des variations de l'intensité vocale au cours d'un énoncé, dans le silence, dans le bruit blanc (bb) et dans le bruit cocktail (ckt).

Nous avons donc tout d'abord voulu vérifier si l'apparition de ces « lobes » d'intensité dans le bruit était un phénomène marginal ou observé dans de nombreux cas. Pour cela, nous avons mesuré la dynamique du lobe d'intensité de chaque syllabe (définie en Annexe Met1)

- dans la BD1, sur les 2 syllabes des 16 mots cibles prononcés dans le silence et dans la condition de bruit cocktail à 86dB. Cela représente donc 32 mesures par locuteur pour chaque condition.
- dans la BD2, sur la syllabe du déterminant [la] et sur les 2 syllabes des 17 logatomes cibles, prononcés en position initiale et finale de l'énoncé pour la condition de silence et de bruit cocktail. Cela représente donc au total 102 mesures par locutrice pour chaque condition.

Nous avons réalisé des tests Anova à un facteur (BRUIT) pour déterminer si la dynamique d'intensité de chaque syllabe évoluait significativement avec le bruit. Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res8.

Les Figure 113 et Figure 114 de la page suivante synthétisent ces résultats en représentant l'évolution du silence au bruit de la dynamique du lobe d'intensité de chaque syllabe, pour les locuteurs et les locutrices de la 1^{ère} et de la 2^{ème} base de données. On observe effectivement un renforcement significatif de cette dynamique chez tous les locuteurs, excepté chez L7. Ces résultats laissent penser qu'il puisse exister dans le bruit une stratégie de renforcement des indices de frontière syllabique. Cependant, avant de pouvoir conclure, il serait nécessaire de mener des tests perceptifs afin de déterminer si la chute d'intensité en fin de syllabe peut être un indice perçu et utilisé par l'auditeur pour segmenter un mot en syllabes.

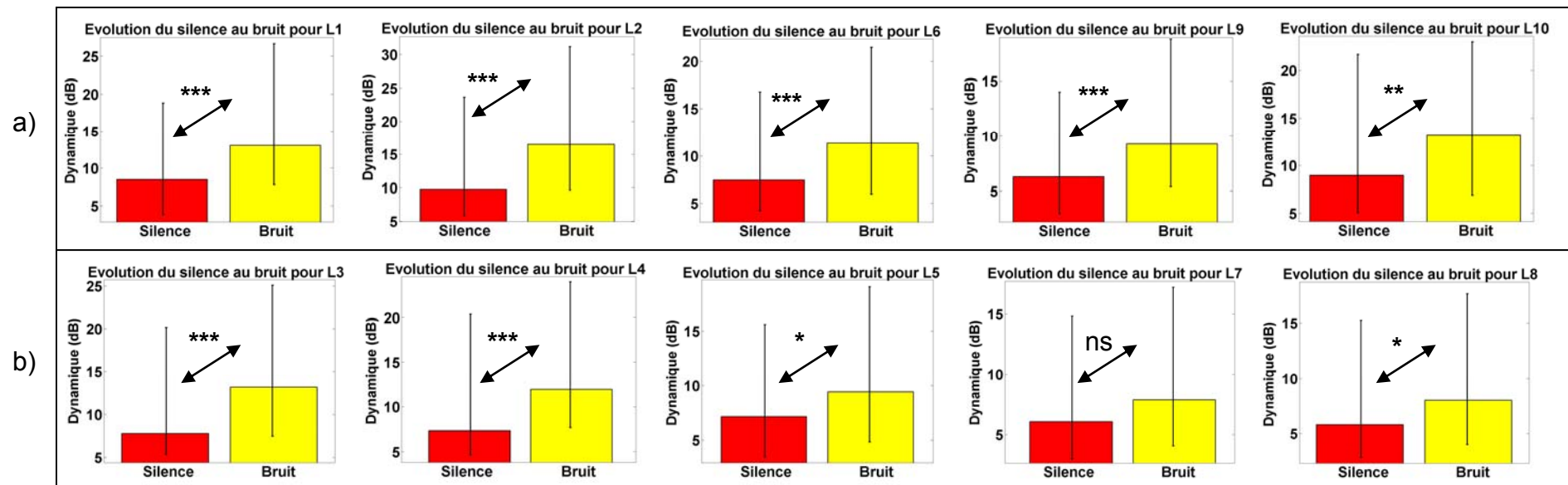


Figure 113. Représentation pour les locutrices (a) et locuteurs (b) de la BD1 de la dynamique d'intensité des syllabes dans le silence et dans le bruit.

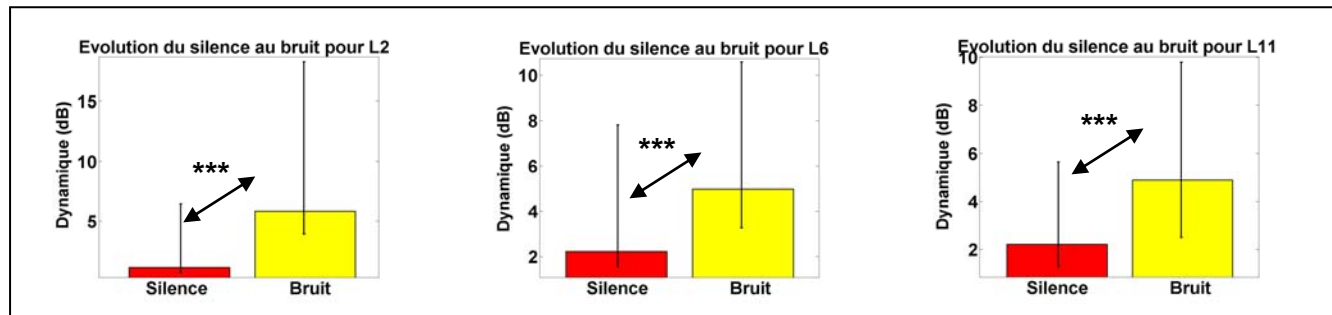


Figure 114. Représentation pour les locutrices de la BD2 de la dynamique d'intensité des syllabes dans le silence et dans le bruit.

2. INDICES DE FRONTIERE ENTRE UN DETERMINANT ET UNE UNITE LEXICALE

2.1. Matériel et méthode

Pour étudier le marquage de la frontière entre la fin du déterminant et le début d'une unité lexicale, nous avons examiné la durée, l'intensité et l'articulation des consonnes initiales de mots de contenu par comparaison aux autres consonnes non initiales de ces mots.

L'intensité et la durée ont été mesurées pour chaque locuteur et locutrice de la 1^{ère} base de données, sur les consonnes initiales des 16 mots-cibles produits dans le silence et dans le bruit cocktail à 86dB, ainsi que sur les 16 consonnes non initiales de ces 16 mots-cibles. Ces mots-cibles ont été constitués sur une structure CVCV et de façon à représenter la quasi-totalité des consonnes du français en position C1 et C2 (cf. § 3.1 du chapitre 3 pour plus de détails sur la BD1). Cela nous permet de comparer chaque consonne lorsqu'elle se trouve en début de mot de contenu ou au milieu du mot. Les premières analyses montraient une grande variabilité. En observant chaque type de consonne de façon distincte, nous avons observé que cette variabilité était due à un comportement très différent entre les consonnes voisées et non voisées. C'est pourquoi nous présentons ici les résultats de nos analyses pour ces deux catégories de consonnes.

L'amplitude et les pics de vitesse du pincement labial sur les consonnes bilabiales ont été mesurés, pour chaque locutrice de la 2^{ème} base de données, sur les consonnes initiales des logatomes [pala], [bala] et [mala], précédés du déterminant [la], produits en position initiale et finale des énoncés pour les conditions de silence et de bruit cocktail à 85dB. Nous avons fait de même pour les consonnes non initiales des logatomes [lapa], [laba] et [lama]. Cela représentait au total 12 mesures de pincement par condition et locutrice, pour les consonnes initiales comme pour les consonnes non initiales d'un mot de contenu.

Nous avons alors réalisé des tests Anova à un facteur (TYPE CONSONNE) à deux niveaux : initiale et non initiale, afin de déterminer si la consonne initiale des mots de contenu est produite de façon significativement différente des autres (ici de la deuxième) dans le silence et dans le bruit. Autrement dit, nous avons cherché à vérifier si la différence de valeur des différents paramètres examinés entre la consonne initiale et non initiale d'un mot de contenu, joue le rôle d'indice de marquage du début d'un mot de contenu dans le silence, et si cela est toujours le cas dans le bruit.

Nous avons également mené un autre test Anova à un facteur (BRUIT) pour déterminer si cet indice, autrement dit la différence d'un paramètre entre la consonne initiale et non initiale d'un mot de contenu, évolue de façon significative du silence au bruit. Nous insistons sur le fait que ce qui nous intéresse n'est pas tant l'évolution du silence au bruit de chaque type de consonne, mais l'évolution de leur différence, pouvant constituer un indice perceptif de contraste pour l'interlocuteur.

Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res8. Nous n'en présenterons ici qu'une synthèse

2.2. Résultats

Les Figure 115 et Figure 116 représentent la différence de durée et d'intensité entre la consonne initiale des mots de contenu et une autre consonne non initiale, dans le silence et dans le bruit, pour les 10 locuteurs et locutrices de la BD1, et en distinguant le cas des consonnes voisées de celui des consonnes non voisées.

On observe en premier lieu deux tendances différentes pour les consonnes voisées et les consonnes non voisées, valable chez tous les locuteurs :

- pour les consonnes voisées, l'intensité de la consonne initiale d'un mot de contenu est renforcée par rapport à celle d'une consonne non initiale. Au contraire, la consonne initiale tend à être raccourcie par rapport à une consonne non initiale.
- pour les consonnes non voisées, la consonne initiale d'un mot de contenu est allongée par rapport à celle d'une consonne non initiale, tandis que son intensité est moins importante que celle d'une consonne non initiale.

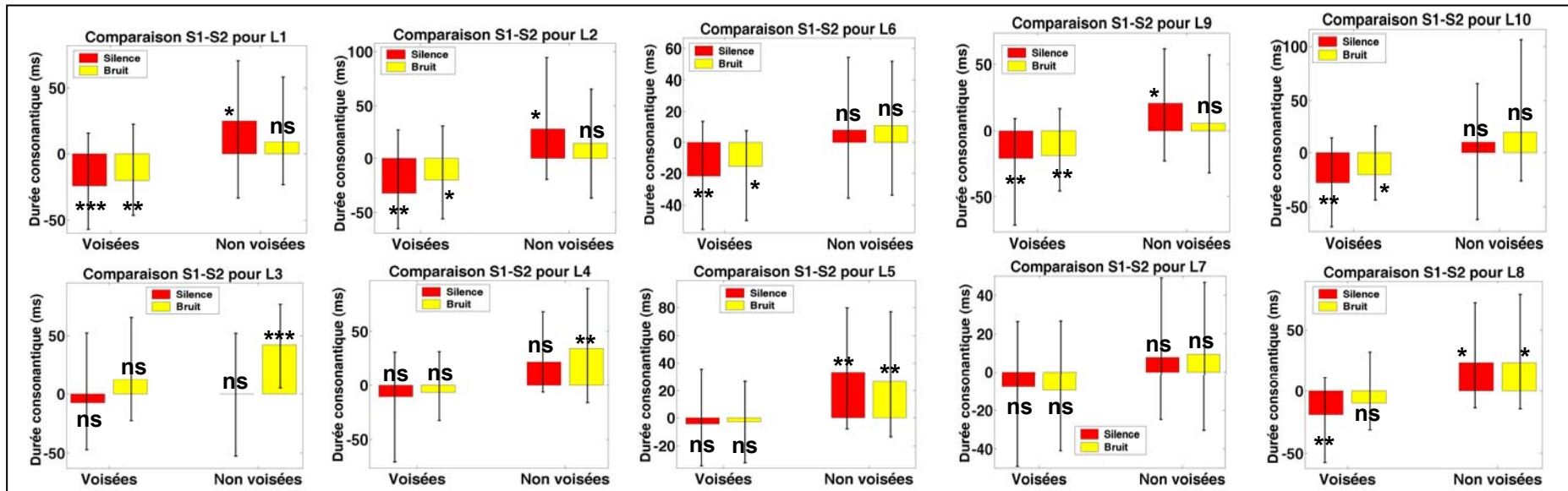


Figure 115. Représentation pour 10 locuteurs de la différence de durée entre la consonne initiale des mots de contenu (S1) et une autre consonne non initiale (S2), dans le silence (en rouge) et dans le bruit (en jaune). Le cas des consonnes voisées et celui des non voisées sont présentés séparément.

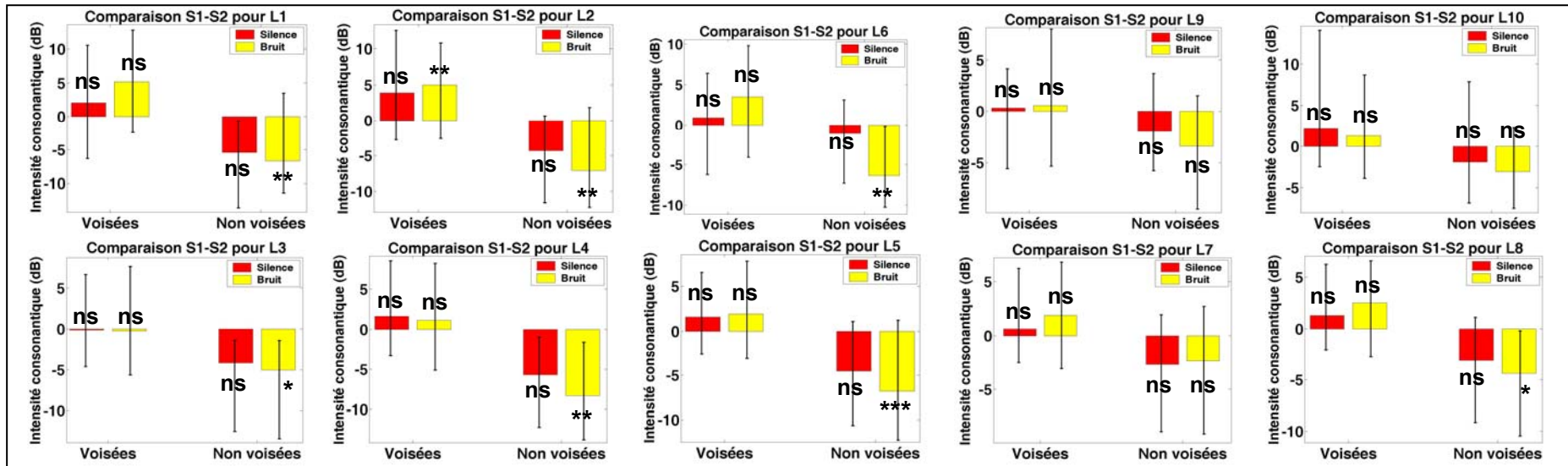


Figure 116. Représentation pour 10 locuteurs de la différence d'intensité entre la consonne initiale des mots de contenu (S1) et une autre consonne non initiale (S2), dans le silence (en rouge) et dans le bruit (en jaune). Le cas des consonnes voisées et celui des non voisées sont présentés séparément.

Pour les consonnes voisées, le raccourcissement consonantique en début de mot n'est significatif dans le silence et dans le bruit que pour les femmes et pour le locuteur L8 dans le silence. Celui-ci a tendance à être moindre dans le bruit, mais pas de façon significative. De même, le renforcement de leur intensité en début de mot n'est quasiment jamais significatif, et n'évolue pas non plus par conséquent de façon notable du silence au bruit.

Pour les consonnes non voisées, l'allongement consonantique en début de mot est significatif pour plusieurs locuteurs dans le silence et le bruit cocktail (L5 et L8), seulement dans le silence (L1, L2, L9) ou seulement dans le bruit (L3). Dans tous les cas, excepté pour L3, cet indice n'évolue pas de façon notable du silence au bruit. Enfin, l'atténuation de l'intensité des consonnes non voisées en début de mot est significative chez certains locuteur, uniquement dans le bruit (L1, L2, L3, L4, L5 et L8) et a tendance à être davantage marquée dans le bruit que dans le silence.

Les Figure 117 et Figure 118 représentent la différence d'amplitude de vitesse au niveau du pincement labial entre la consonne initiale des mots de contenu et une autre consonne non initiale, dans le silence et dans le bruit, pour les 3 locutrices de la BD2.

Aucune locutrice ne montre de renforcement significatif de l'amplitude du pincement sur la consonne initiale d'un mot de contenu dans le silence, tandis que ce renforcement existe chez toutes les locutrices dans le bruit. Par ailleurs, les locutrices L2 et L11 montrent des pics de vitesse du pincement labial plus importants sur les consonnes initiales des mots de contenu déjà dans le silence. Ce renforcement est conservé dans le bruit par L2, amplifié par L11 et créé par L6.

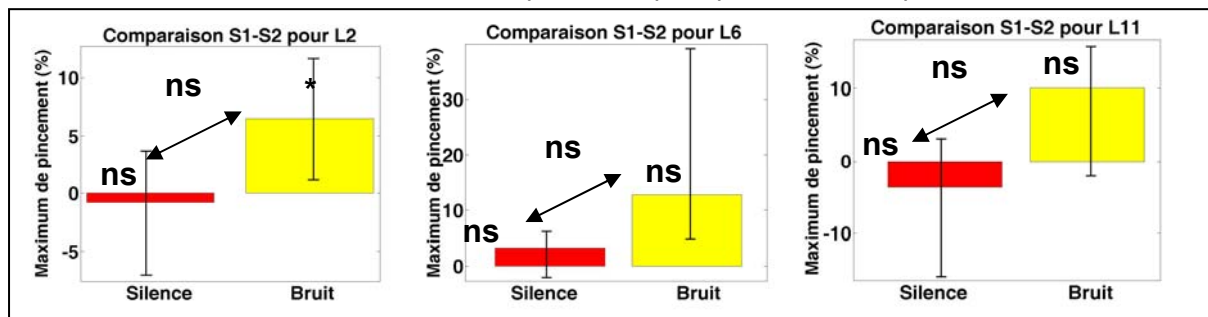


Figure 117. Représentation pour 3 locutrices de la différence d'amplitude du pincement labial entre la consonne initiale des mots de contenu (S1) et une autre consonne non initiale (S2), dans le silence (en rouge) et dans le bruit (en jaune). Seules les consonnes bilabiales [m], [b] et [p] ont été considérées.

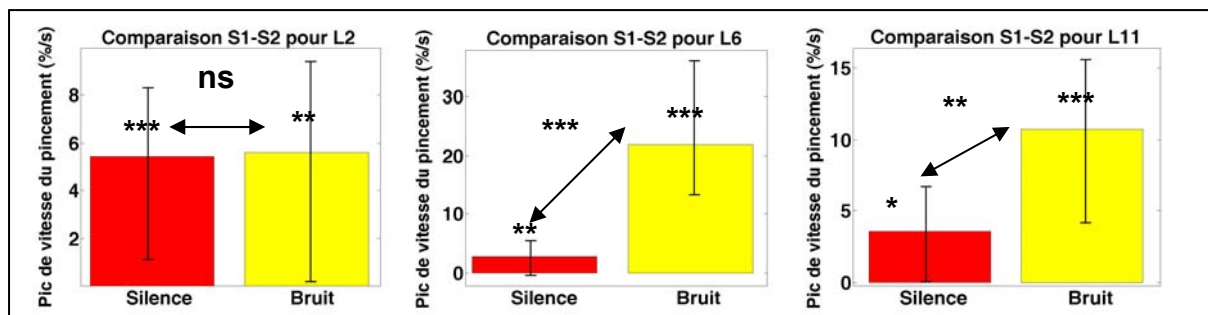


Figure 118. Représentation pour 3 locutrices de la différence de pic de vitesse du pincement labial entre la consonne initiale des mots de contenu (S1) et une autre consonne non initiale (S2), dans le silence (en rouge) et dans le bruit (en jaune). Seules les consonnes bilabiales [m], [b] et [p] ont été considérées.

2.3. Conclusion sur les indices de début d'une unité lexicale

Nous n'observons pas dans cette étude un allongement systématique de la consonne initiale d'un mot de contenu, mais uniquement lorsque la consonne initiale est non voisée. En revanche, nous avons observé que de façon complémentaire, l'intensité de la consonne initiale d'un mot de contenu est renforcée par rapport aux autres consonnes lorsqu'elle est voisée. Il est donc envisageable que ces deux indices se complètent selon le type de consonne.

Dans les deux cas, nous n'avons pas observé de renforcement dans le bruit de ces indices. Il ne semble donc pas exister de stratégie de renforcement dans le bruit des indices de durée et d'intensité marquant le début d'un mot de contenu. Cependant, notre corpus n'était pas idéal pour cette

exploration, puisque nous avons examiné de façon confondue plusieurs consonnes, suivies de voyelles différentes. Il est possible que les grands écart-types observés sur nos données proviennent de ces effets de coarticulation et nous empêchent d'examiner précisément le phénomène qui nous intéresse.

Au niveau articulatoire, nous avons observé que le pincement labial n'est pas non plus significativement renforcé sur les consonnes bilabiales en début de mot. En revanche sa force est significativement plus importante en début de mot que ce soit dans le silence ou dans le bruit, et est également significativement renforcée dans le bruit par 2 locutrices sur 3. Il serait donc nécessaire de mener un test perceptif afin de déterminer si l'augmentation de la force du pincement labial contribue à une meilleure perception visuelle ou auditive (par le bruit de plosion qui pourrait y être associé) de la frontière entre un déterminant et un mot de contenu.

Dans ce paragraphe, nous n'avons pas examiné de façon exhaustive tous les indices prosodiques permettant de marquer le début d'un mot de contenu. D'autres auteurs se sont également intéressés au renforcement de tels indices en conditions d'intelligibilité perturbée. En particulier, Welby 2007^[377] a porté son attention sur le « coude » de fréquence fondamentale en début de mot de contenu pouvant être un indice intonatif de marquage de la frontière avec le déterminant qui précède. Dans une première étude effectuée en parallèle de notre étude articulatoire, sur le même corpus pilote (cf. Annexe Pil2, Garnier et al. 2006^[102] ; Garnier et al. 2006^[104]), cet auteur a observé que certains locuteurs montraient dans le bruit des cas de renforcement de ce « coude », sans que cette tendance soit significative (Welby 2005^[374]). Celle-ci n'a pas pu être observée dans une seconde étude (Welby 2006^[375]).

3. INDICES DE DEBUT DE L'ENONCE

Pour étudier le marquage de début d'énoncé, nous avons exploré la deuxième base de données, où chaque énoncé est de la forme « *la Bala longe la Lanla* » et commence donc par le déterminant [la]. Pour avoir une référence de comparaison, nous avons dû trouver dans l'énoncé une autre occurrence de cette même syllabe, toujours portée par un mot outil, mais cette fois en position neutre de l'énoncé. Le deuxième déterminant introduisant l'objet de la phrase satisfait à cette condition.

Nous avons ainsi pu comparer la différence de durée entre le premier et le deuxième déterminant, représentant la différence de durée entre la syllabe initiale de l'énoncé et une autre syllabe non initiale. Par contre, nous ne pouvons pas examiner la différence d'intensité ou d'amplitude des mouvements articulatoires entre ces deux syllabes à cause de l'effet de déclinaison. Il existe en effet une diminution progressive de la fréquence fondamentale, de l'intensité et de l'amplitude de l'articulation au cours de l'énoncé. Ainsi, la différence que l'on observerait entre la syllabe initiale et non initiale de l'énoncé risquerait d'être davantage imputable à ce phénomène de déclinaison qu'au marquage du début de l'énoncé.

Nous avons pour cela mesuré la durée des premières syllabes [la] de chaque énoncé, en condition de silence et de bruit cocktail à 85dB. Nous avons fait de même pour la durée du deuxième déterminant [la] situé au milieu de l'énoncé. Cela représentait au total 34 mesures pour chaque type de syllabe et dans chaque condition.

Nous avons alors réalisé des tests Anova à un facteur (TYPE SYLLABE) à deux niveaux : initiale et non initiale, afin de déterminer si la syllabe initiale des énoncés est produite de façon significativement différente des autres dans le silence et dans le bruit. Autrement dit, nous avons cherché à vérifier si la différence de valeur des différents paramètres examinés entre la syllabe initiale et non initiale d'un énoncé, joue le rôle d'indice de marquage du commencement de l'énoncé dans le silence, et si cela est toujours le cas dans le bruit.

Nous avons également mené un autre test Anova à un facteur (BRUIT) pour déterminer si cet indice, autrement dit la différence d'un paramètre entre la syllabe initiale et non initiale d'un énoncé, évolue de façon significative du silence au bruit. Nous insistons sur le fait que ce qui nous intéresse n'est pas tant l'évolution du silence au bruit de chaque syllabe, mais l'évolution de leur différence, pouvant constituer un indice perceptif de contraste pour l'interlocuteur.

Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res8. Nous n'en présenterons ici qu'une synthèse

La Figure 119 représente la différence de durée entre la syllabe initiale des énoncés et une syllabe comparable mais neutre de l'énoncé, dans le silence et dans le bruit. On observe que l'allongement de la syllabe initiale de l'énoncé est bien effectué dans le silence par les locutrices L6 et L11, et par toutes les locutrices dans le bruit. Conformément à notre hypothèse et aux études antérieures, les syllabes initiales des énoncés tendent à être plus allongées dans le bruit que dans le silence. Cependant, cette tendance n'est pas significative et inférieure à 10ms. Il ne semble donc pas y avoir, chez les 3 locutrices examinées, de stratégie de renforcement de l'indice de durée marquant le début de l'énoncé. Il serait intéressant d'examiner si d'autres indices (d'intensité, d'amplitude des mouvements articulatoires) sont plus significativement renforcés dans le bruit.

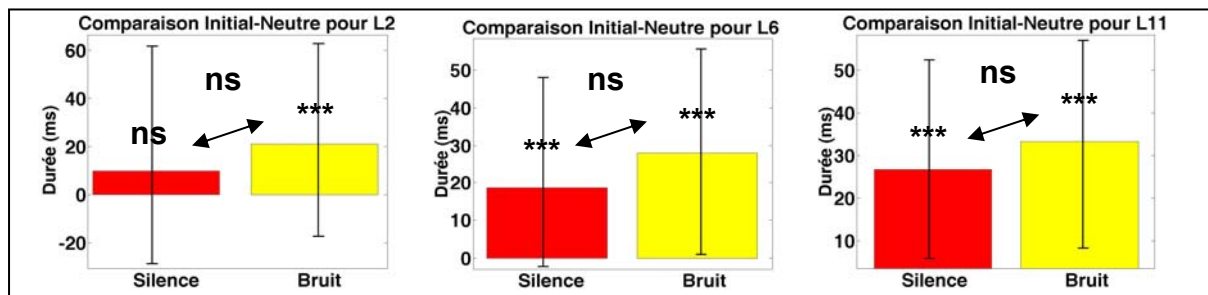


Figure 119. Représentation pour 3 locutrices de la différence de durée entre la syllabe initiale de l'énoncé (Initial) et une autre syllabe comparable mais non initiale (Neutre), dans le silence (en rouge) et dans le bruit (en jaune).

4. INDICES DE FIN DE L'ÉNONCÉ

4.1. Matériel et méthode

Pour étudier le marquage de fin d'énoncé, nous avons exploré la deuxième base de données, dans laquelle les énoncés se terminent tous par un logatome cible de deux syllabes. Or ceux-ci sont de structure CVCV et ont été constitués de façon à représenter un ensemble de 17 syllabes en position initiale ou finale du mot (cf. § 2.1 du chapitre 4 pour plus de détails). Cela nous a permis de comparer une même syllabe en position finale et non finale (il s'agissait alors de l'avant dernière syllabe) de l'énoncé.

Nous avons commencé par mesurer la durée de la syllabe finale de chaque énoncé, son intensité et son aire intéro-labiale, en condition de silence et de bruit cocktail à 85dB. Nous avons fait de même pour la durée des mêmes syllabes en avant dernière position de l'énoncé. Cela représentait au total 34 mesures pour chaque type de syllabe et dans chaque condition.

Nous avons alors réalisé des tests Anova à un facteur (TYPE SYLLABE) à deux niveaux : finale et non finale, afin de déterminer si la syllabe finale des énoncés est produite de façon significativement différente des autres dans le silence et dans le bruit. Autrement dit, nous avons cherché à vérifier si la différence de valeur des différents paramètres examinés entre la syllabe finale et non finale d'un énoncé, joue le rôle d'indice de marquage de fin de l'énoncé dans le silence, et si cela est toujours le cas dans le bruit. Nous avons également mené un autre test Anova à un facteur (BRUIT) pour déterminer si cet indice, autrement dit la différence d'un paramètre entre la syllabe finale et non finale d'un énoncé, évolue de façon significative du silence au bruit. Nous insistons sur le fait que ce qui nous intéresse n'est pas tant l'évolution du silence au bruit de chaque syllabe, mais l'évolution de leur différence, pouvant constituer un indice perceptif de contraste pour l'interlocuteur.

Nous avons également examiné l'ambitus de la chute mélodique finale, en repérant sous Praat le ton bas final de frontière (L% dans le modèle de l'intonation du français de Jun et al. 1995^[170]) et le ton haut précédent (correspondant au Hi de l'accent secondaire dans ce même modèle). Cette chute mélodique n'est pas nécessairement alignée par rapport aux syllabes. Au cours de cette annotation, nous avons observé dans quelques rares cas (5 au total) que les locutrices réalisaient une continuation majeure en fin d'énoncé (c'est à dire un ton haut, H%) à la place de la chute finale attendue. Cela peut arriver lors de la lecture d'une liste de mots ou de phrase (Delattre 1966^[67]). Par contre, il est assez étrange que ce phénomène se soit produit ici, alors que les locutrices étaient enregistrées en parole semi-spontanée, et marquaient des pauses relativement longues entre chaque énoncé. Cette continuation majeure correspond à un phénomène prosodique différent de la chute mélodique finale. Aussi, il ne serait pas rigoureux de les étudier de façon confondue. Nous nous sommes ici intéressés uniquement à l'indice de marquage de fin d'énoncé que peut constituer la chute mélodique finale, et avons par conséquent mis de côté les cas de continuation majeure. Mais nous insistons sur le fait qu'il ne s'agit en aucun cas d'une sélection des données par rapport à nos hypothèses : nous ne cherchons pas ici à montrer qu'il existe des chutes mélodiques finales en fin d'énoncé (cela est une observation plus que classique), mais à déterminer si ces chutes mélodiques finales, quand elles sont produites, sont renforcées dans le bruit.

Nous avons alors réalisé sur ces mesures un test Anova à un facteur (BRUIT) pour examiner si l'intervalle de fréquences fondamentales (en tons) couvert par les chutes mélodiques en fin d'énoncé évolue de façon significative du silence au bruit. Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res8. Nous n'en présenterons ici qu'une synthèse.

4.2. Résultats

La Figure 120 représente la différence de durée entre la syllabe finale des énoncés et une autre syllabe comparable, mais neutre de l'énoncé, dans le silence et dans le bruit. La Figure 121 représente l'ambitus tonal de la chute mélodique finale en fin d'énoncé, dans le silence et dans le bruit.

On observe que la syllabe finale des énoncés est bien allongée dans le silence par toutes les locutrices, et que cet allongement est renforcé dans le bruit également par toutes les locutrices. De

même, l'ambitus tonal de la chute mélodique finale en fin d'énoncé est renforcé par toutes les locutrices dans le bruit.

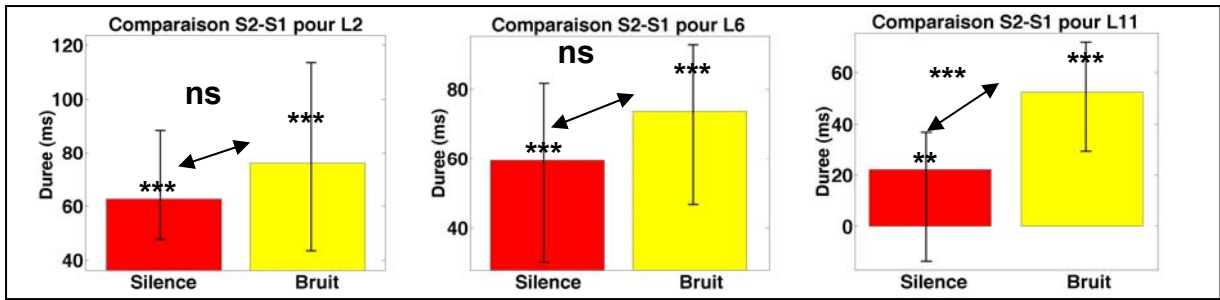


Figure 120. Représentation pour 3 locutrices de la différence de durée entre la syllabe finale des énoncés (S2) et une autre syllabe comparable mais non finale (S1), dans le silence (en rouge) et dans le bruit (en jaune).



Figure 121. Représentation pour 3 locutrices de l'ambitus tonal de la chute mélodique finale en fin d'énoncé, dans le silence (en rouge) et dans le bruit (en jaune).

La Figure 122 représente la différence d'intensité et d'aire intéro-labiale entre la syllabe finale des énoncés et une autre syllabe comparable, mais neutre de l'énoncé, dans le silence et dans le bruit. On observe bien une diminution de l'intensité en fin de phrase chez toutes les locutrices dans le silence, i.e. une différence d'intensité négative entre la syllabe finale de l'énoncé et la syllabe qui la précède. Cette diminution est moins importante dans le bruit que dans le silence pour toutes les locutrices. L'amplitude des mouvements articulaires d'ouverture des lèvres suit la même tendance que l'intensité. Autrement dit, nous observons dans ce corpus aucune différence articulatoire sur la syllabe finale de l'énoncé, voire plutôt une tendance d'hypo-articulation.

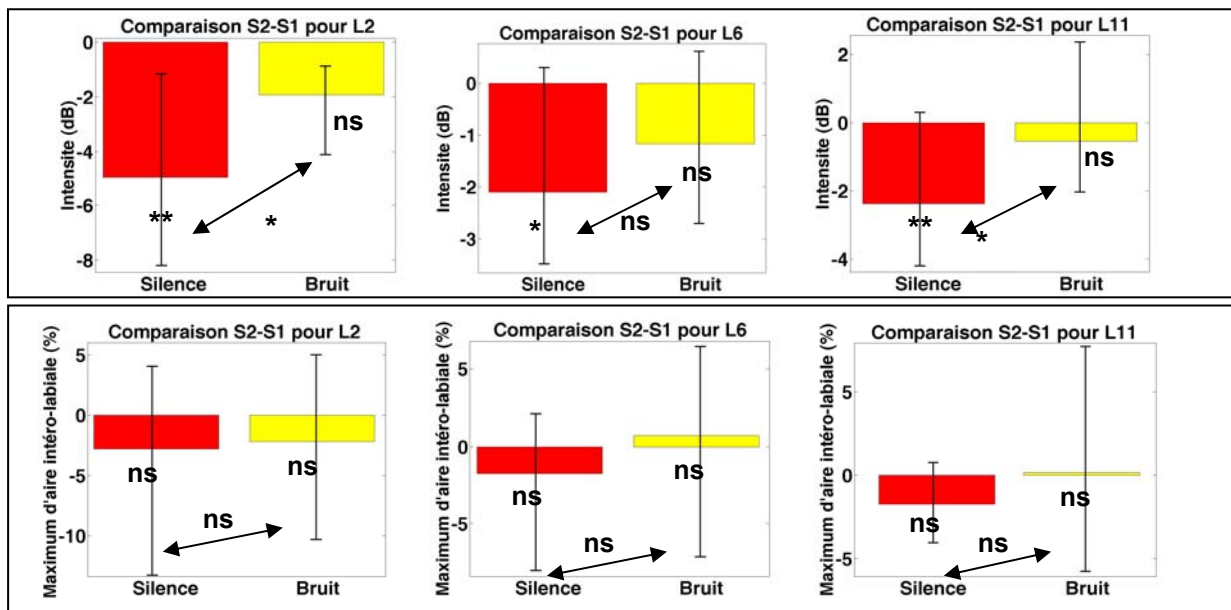


Figure 122. Représentation pour 3 locutrices de la différence d'intensité entre la syllabe finale des énoncés (S2) et une autre syllabe comparable mais non finale (S1), dans le silence (en rouge) et dans le bruit (en jaune).

4.3. Conclusion quant aux indices de fin d'énoncé

Conformément à notre hypothèse et aux études antérieures, les syllabes finales des énoncés sont encore plus allongées dans le bruit que dans le silence. Ces résultats vont dans le sens de ceux de Rostolland 1982^[289] observant un allongement syllabique deux fois plus important pour les syllabes finales que pour les syllabes initiales des mots. De même, la chute mélodique en fin d'énoncé est amplifiée dans le bruit. Il semble donc que les locutrices cherchent à renforcer ces indices pour augmenter leur intelligibilité dans le bruit. Ce résultat va à l'encontre de Rostolland 1982^[289] rapportant que la marge de modulation de la voix en fréquences se restreint à forte intensité vocale, affectant en cela l'intonation de la parole. En revanche, ce renforcement de la chute mélodique en fin d'énoncé est parfaitement cohérent avec notre observation d'une augmentation de l'ambitus de F0 dans le bruit (cf. Chapitre 6). Aussi, il est tout à fait possible que cette augmentation de l'ambitus de F0 soit davantage lié au renforcement de certains indices intonatifs qu'à l'élargissement de l'entendue vocale dans une recherche d'émergence par rapport au bruit ambiant. Il serait maintenant intéressant de poursuivre cette première observation en examinant si les montées mélodiques en fin de syntagme accentuel ou en fin de syntagme intonatif interrogatif sont également renforcées dans le bruit.

En revanche, nous n'observons pas d'hyperarticulation des syllabes finales de l'énoncé, même dans le silence, ce qui est contraire aux résultats d'études antérieures (Loevenbruck 2000^[223] ; Rolland et al. 2002^[287] ; Tabain 2003^[341]). La fin de l'énoncé s'accompagne plutôt dans notre corpus d'une hypo-articulation au niveau de l'aire intéro-labiale, conjointement à la chute d'intensité en fin de phrase. Cette hypo-articulation n'est cependant pas significative ni dans le silence, ni dans le bruit.

Nous n'avons ici examiné que les indices de fin d'énoncé. Au cours de l'étude pilote préalable à la constitution de la deuxième base de données (cf. Annexe Pil2, Garnier et al. 2006^[104]), et menée sur une seule locutrice (L11), nous avons également observé que l'allongement de la durée en fin de syntagme accentuel était également renforcé dans le bruit, mais de façon moins importante qu'en fin de syntagme intonatif. Par ailleurs, nous avons observé dans la deuxième base de données que 9.8% des énoncés présentent une ou plusieurs pauses dans le silence contre 25.5% dans le bruit blanc et 21.5% dans le bruit cocktail. Ces pauses sont presque toujours situées entre le Sujet et le Verbe de la phrase porteuse (« La _ longue la _ »), très rarement entre le Verbe et l'Objet de la phrase ou les deux à la fois. Par ailleurs, ces pauses ne sont pas seulement plus fréquentes mais également plus longues : de 47 ms en moyenne dans le silence et atteignant 194 ms en moyenne dans le bruit. Même ramené au débit de parole, ces pauses augmentent de 3.5% de la durée de l'énoncé dans le silence à 11.5% dans le bruit. Ces différentes observations motivent donc à explorer par la suite plus en détail ce qui se passe à l'intérieur de l'énoncé, et pas uniquement son début ou sa fin.

5. CONCLUSION DU CHAPITRE 8

Dans ce chapitre, nous avons pu montrer l'existence de stratégies visant à renforcer des indices de structuration du discours au niveau de la syllabe, du mot et de l'énoncé.

Ainsi, le renforcement dans le bruit de la dynamique d'intensité par syllabe pourrait contribuer à une meilleure délimitation des syllabes d'un mot. L'augmentation de la force de pincement des lèvres sur les consonnes bilabiales en début de mot de contenu peut être liée à une stratégie visuelle ou acoustique de renforcement de l'indice de début du mot. Nous avons également observé un allongement plus important dans le bruit de la syllabe initiale et finale d'un énoncé, de même qu'une augmentation de l'ambitus tonal de la chute mélodique finale.

De nombreux autres indices restent encore à explorer. Il est également nécessaire de valider ces observations d'un point de vue perceptif, afin de vérifier que ces renforcements en production sont bien perçus par l'interlocuteur et l'aident dans sa compréhension.

Chapitre 9 :
**Une adaptation visant à
mettre en valeur l'information
importante ?**

Nous avons évoqué dans le chapitre 2 que l'intelligibilité est liée à la prévisibilité proche et lointaine du message par rapport au contexte dans lequel il est émis et perçu (Leipp 1977^[208]). La notion de prévisibilité est en réalité liée aux mécanismes cognitifs d'anticipation, d'activation lexicale et de reconstitution. Elle suggère que l'auditeur n'est pas passif et ne perçoit pas nécessairement l'intégralité des indices du message mais construit une signification à partir de la reconstitution des différents indices perçus et de la probabilité des différentes reconstitutions possibles.

Cette prévisibilité ne conditionne pas seulement la perception de l'auditeur mais peut être également exploitée en production de parole. Ainsi, par rapport à la théorie H&H (Lindblom 1990^[217]) on peut envisager que le locuteur fournisse moins d'effort sur les indices du message pouvant être facilement reconstitué par l'auditeur dans un contexte donné. Ainsi, Meunier et al. 2006^[238] interprètent la réduction vocalique observée en parole spontanée comme le fait que les mots prononcés dans un contexte sont plus facilement reconstituables que des mots lus hors contexte, et par conséquent ne nécessitent pas une articulation des voyelles aussi « précise ».

Mais les indices du message ne sont pas seulement des indices de reconnaissance de chaque phonème constituant les mots du message, mais également les mots constituant le message, qui n'ont pas tous le même poids informationnel : Certains mots par exemple sont repris d'un énoncé précédent et n'apportent pas d'information nouvelle. Il peuvent donc être facilement reconstitués par l'auditeur (Gravano et al. 2006^[122]). D'autres mots de l'énoncé ont davantage un rôle syntaxique que sémantique. On les appelle les *mots-outil* ou *mots de fonction* (déterminants, pronoms, mots de liaison), par opposition aux *mots de contenu*, dont le sens est aussi important que le rôle syntaxique (nom, adjectif, verbe et adverbe). A noter que certains auteurs parlent également de mots grammaticaux vs. de mots lexicaux. Une minimalisation des mots de fonction par rapport aux mots de contenu a été montrée en français et pour d'autres langues (Cutler 1993^[60]). Il existe en particulier des différences prosodiques entre ces deux catégories de mots (Selkirk 1996^[312] ; Welby 2006^[376]) rendant les mots de contenu plus saillant perceptivement (Dodane et al. 2006^[72]). Par ailleurs, il a été montré que ces indices sont perçus par les auditeurs : en particulier ils sont utilisés par les bébés lors de l'acquisition du lexique (Shi et al. 1999^[314]).

On peut donc faire l'hypothèse que cette différence entre les mots de fonction et de contenu soit renforcée lorsque le locuteur cherche à augmenter son intelligibilité. Cette hypothèse a été vérifiée par Dodane et al. 2006^[72] en ce qui concerne le langage produit par des mères s'adressant à leur bébé. Nous nous intéressons ici à examiner cette hypothèse dans le cas de la parole produite en environnement bruyant.

1. MATERIEL ET METHODE

Pour cela, nous avons analysé la deuxième base de données (BD2), dans laquelle 3 locutrices (L2, L6 et L11) ont été enregistrées en parole semi-spontanée dans le silence et dans un bruit de cocktail-party diffusé à 85 dB SPL par le biais de haut-parleurs (cf. § 2.1 du chapitre 4 pour plus de détails). Ce corpus permet d'examiner notre hypothèse car il présente :

- des mots de contenu. Il est en effet constitué de 17 logatomes-cibles comportant la syllabe [la] en position initiale (pour 9 logatomes sur 17) ou finale (pour également 9 logatomes).
- des mots outils comparables aux mots de contenu. Tous les logatomes sont en effet introduits par le déterminant [la].
- un contexte de communication et de prévisibilité du message. Les locutrices avaient pour tâche de commander à l'expérimentateur de tracer des flèches au tableau, reliant des rivières portant le nom des 17 logatomes. Le vocabulaire possible était donc connu à l'avance à la fois par le locuteur et l'auditeur, et était limité à 17 mots hautement confondables puisqu'ils sont tous dérivés du logatome [lala] et ne s'en différencient que par une syllabe.
- une différence de poids informationnel entre les mots de fonction et les mots de contenu. Le déterminant était en effet toujours le même, une phrase porteuse était imposée (« la *rivière1* longe la *rivière2* »), avec toujours le même verbe. Seuls les noms de rivière apportaient donc une information nouvelle et utile vis à vis de la tâche à accomplir.

Pour chacun des 17 logatomes-cibles prononcés en position initiale et finale de l'énoncé, dans le silence et dans le bruit de cocktail à 85dB, nous avons comparé la syllabe [la] du déterminant (mot de fonction) avec la 1^{ère} ou la 2^{ème} syllabe [la] du logatome-cible (mot de contenu). Cela représente au

total 68 mesures par condition et par locutrice. Ces comparaisons ont été menées pour plusieurs descripteurs de la parole : la durée syllabique, l'intensité moyenne, la F0 moyenne et le timbrage moyen, correspondant ici au pourcentage d'énergie sonore entre 2 et 4kHz relativement à l'ensemble de l'énergie sonore produite. Nous avons également considéré l'amplitude des mouvements articulatoires d'étirement labial (A), d'ouverture labiale (B) et d'aire intéro-labiale (S).

Nous avons alors réalisé des tests Anova à un facteur (TYPE SYLLABE) à deux niveaux : fonction et outil, afin de déterminer si les syllabes des mots de contenu sont produites de façon significativement différente des syllabes des mots de fonction, dans le silence et dans le bruit. Autrement dit, nous avons cherché à vérifier si la différence de valeur des différents paramètres examinés entre les syllabes de mots de fonction et de mots de contenu, permet de rendre les mots de contenu plus saillants perceptivement, et si cela est toujours le cas dans le bruit.

Nous avons également mené un autre test Anova à un facteur (BRUIT) pour déterminer si cet indice, autrement dit la différence d'un paramètre entre les syllabes de mots de fonction et de mots de contenu, évolue de façon significative du silence au bruit. Nous insistons à nouveau sur le fait que ce qui nous intéresse n'est pas tant l'évolution du silence au bruit de chaque type de syllabe, mais l'évolution de leur différence, pouvant constituer un indice perceptif de contraste pour l'interlocuteur.

Nous avons ensuite mené des analyses complémentaires plus détaillées en distinguant les syllabes initiales (S1) et finales (S2) des logatomes cibles, afin de vérifier si les résultats des analyses précédentes sont valables pour toutes les syllabes des logatomes-cibles, ou bien si l'écart entre les mots de contenu et les mots de fonction est différent selon la syllabe considérée du mot de contenu. Un comportement similaire des deux syllabes du mot de contenu, comparativement à la syllabe du déterminant, argumenterait plutôt en faveur de l'existence d'un indice de discrimination des mots de fonction et des mots de contenu. Au contraire, si seulement une des deux syllabes du mot de contenu se distingue de la syllabe du déterminant, les premiers résultats seraient en fait l'artefact d'un autre effet, par exemple d'un indice de frontière lexicale, et non le signe d'une distinction des mots par rapport à leur poids informationnel.

Les résultats détaillés de ces tests statistiques sont présentés en Annexe Res9. Nous n'en présenterons ici qu'une synthèse

2. RESULTATS

La Figure 123 représente pour plusieurs paramètres la différence moyenne entre les syllabes des logatomes cibles et la syllabe du déterminant, dans le silence et dans le bruit cocktail à 85dB.

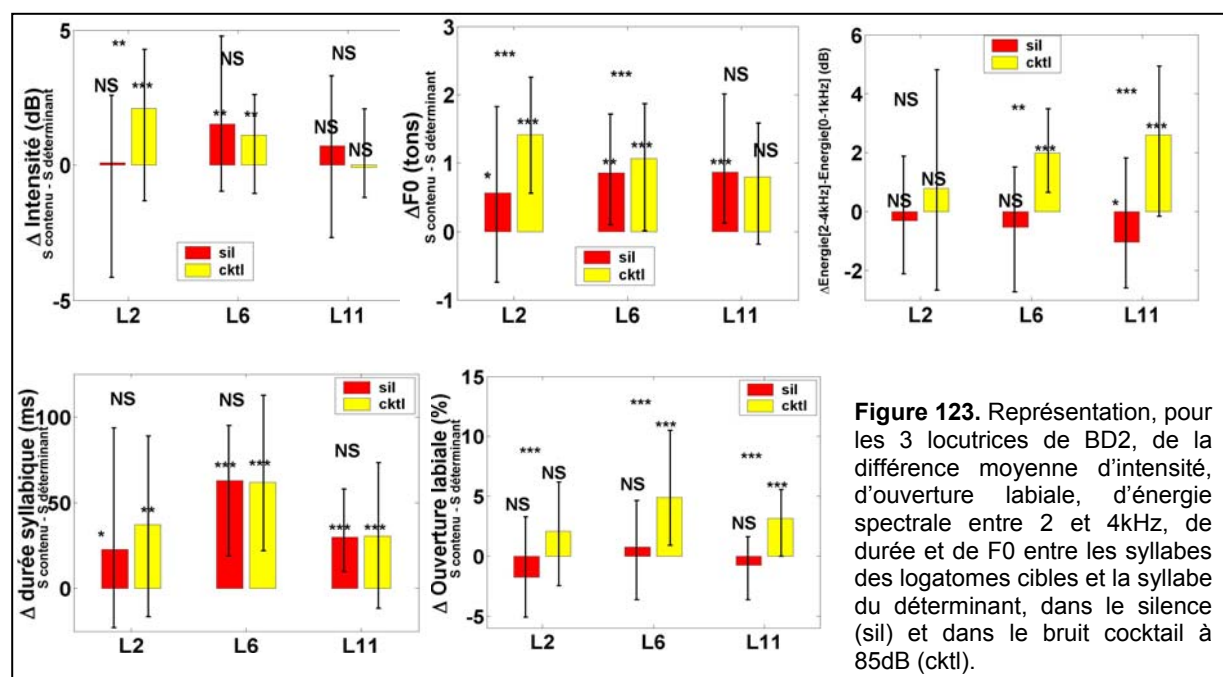


Figure 123. Représentation, pour les 3 locutrices de BD2, de la différence moyenne d'intensité, d'ouverture labiale, d'énergie spectrale entre 2 et 4kHz, de durée et de F0 entre les syllabes des logatomes cibles et la syllabe du déterminant, dans le silence (sil) et dans le bruit cocktail à 85dB (cktl).

On peut distinguer 4 cas en ce qui concerne la différence entre les mots de fonction et de contenu :

- (1) Tout d'abord, certains paramètres ne semblent pas être utilisés par des locutrices pour marquer la différence entre les mots de contenu des mots de fonction : c'est le cas de l'intensité pour la locutrice L11, du timbrage et de l'ouverture labiale pour L2. Ces paramètres ne montrent pour ces locutrices aucune différence entre les mots de fonction et de contenu, que ce soit dans le silence ou dans le bruit, et pas non plus d'évolution significative de cette différence du silence au bruit. Nous avons vérifié cette observation pour les deux syllabes du mot de contenu (cf. Figure 124(a)), excepté dans le cas de l'amplitude des mouvements d'ouverture labiale chez L2, où la 2^{ème} syllabe du mot de contenu est légèrement hyper-articulée par rapport à la syllabe du déterminant et à la 1^{ère} syllabe du mot de contenu (cf. Figure 124 (b)). Dans ce dernier cas, il s'agit donc davantage d'un indice de fin de syntagme que d'un indice de différenciation des mots de fonction et de contenu.

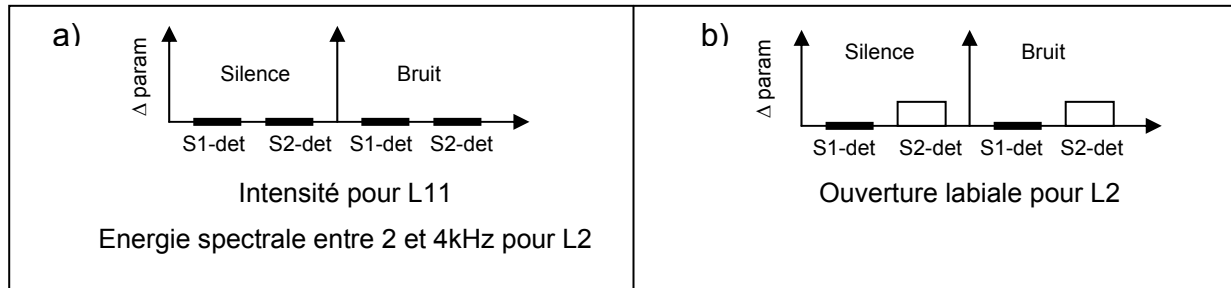


Figure 124. Schématisation de deux cas où les syllabes des mots de contenus ne sont en moyenne pas significativement différentes des syllabes de déterminant (det), aussi bien dans le silence que dans le bruit. Dans le 1^{er} cas (a), cela est valable pour les deux syllabes initiales (S1) et finales (S2) du mot de contenu. Dans le 2^{ème} cas (b), cela est seulement valable pour la 1^{ère} syllabe, la deuxième montrant une légère différence avec la syllabe du déterminant et avec la 1^{ère} syllabe.

- (2) D'autres paramètres sont renforcés sur les mots de contenu par rapport aux mots de fonction, mais cet indice de distinction n'est pas renforcé dans le bruit : c'est le cas de la durée pour toutes les locutrices et de la F0 moyenne pour L11, ainsi que de l'intensité pour L6. Pour ces paramètres, toutes les syllabes des mots de contenu comme des mots de fonction sont renforcées de façon équivalente avec le bruit. Dans certains cas, la deuxième syllabe du mot de contenu est légèrement renforcée par rapport à la première à la fois dans le silence et le bruit (cf. Figure 125 (a)) : c'est le cas de la durée pour toutes les locutrices. Il semble donc y avoir deux effets : non seulement un allongement des syllabes d'un mot de contenu par rapport à celle du déterminant, mais aussi un allongement final en fin de syntagme. Ce dernier est parfaitement connu comme étant un indice de fin d'une unité prosodique en français (Wenk et al. 1982^[378]; Delattre 1966^[67]). Dans d'autres cas, les deux syllabes du mot de contenu sont produites de façon semblable (cf. Figure 125 (b)) : c'est le cas de la F0 pour L11 et de l'intensité pour L6. Mais dans tous les cas, les deux syllabes du mot de contenu restent toujours significativement renforcées par rapport à la syllabe du déterminant.

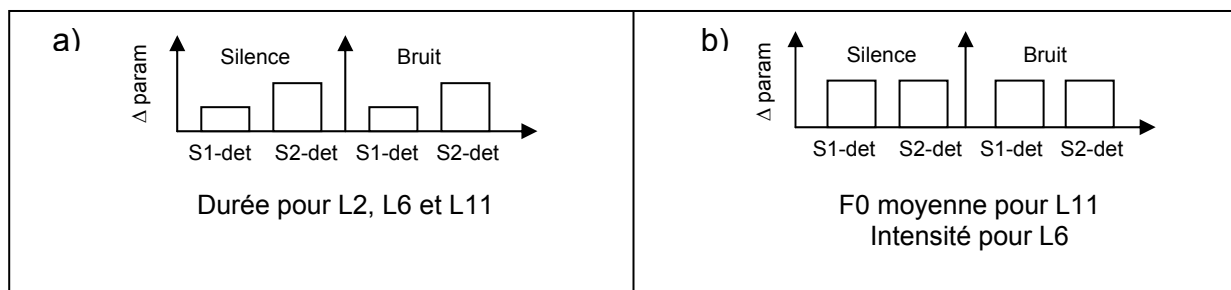


Figure 125. Schématisation de deux cas où les syllabes des mots de contenus sont en moyenne significativement différentes des syllabes de déterminant (det), aussi bien dans le silence que dans le bruit, sans que cette différence n'augmente avec l'exposition au bruit. Dans le 1^{er} cas (a), la syllabe finale montre une accentuation supplémentaire par rapport à la syllabe initiale (S2) du mot de contenu. Dans le 2^{ème} cas (b), cela est seulement valable pour la 1^{ère} syllabe, la deuxième montrant une légère différence avec la syllabe du déterminant et avec la 1^{ère} syllabe.

Enfin, il existe bien des paramètres qui vérifient notre hypothèse, autrement dit pour lesquels la différence entre les mots de fonction et de contenu est renforcée dans le bruit. On peut distinguer deux sous-cas :

- (3) Ce renforcement peut en fait correspondre à la création dans le bruit d'un indice de distinction entre ces deux types de mots, qui n'existe pas dans le silence : c'est le cas de l'intensité pour la locutrice L2, du timbrage pour L6 et L11, et de l'amplitude des mouvements d'ouverture labiale pour toutes les locutrices.

Pour tous ces paramètres, nous avons vérifié qu'il s'agit bien de la création d'une distinction pour les deux syllabes des mots de contenu (cf. Figure 126 (a)) ou de la création d'une distinction pour la syllabe initiale et d'une amplification de la distinction existant dans le silence pour la deuxième syllabe (cf. Figure 126 (b)). Dans tous les cas, l'amplification de l'écart avec le déterminant concerne bien les deux syllabes du mot de contenu, avec le maintien d'un léger renforcement de la deuxième syllabe par rapport à la première.

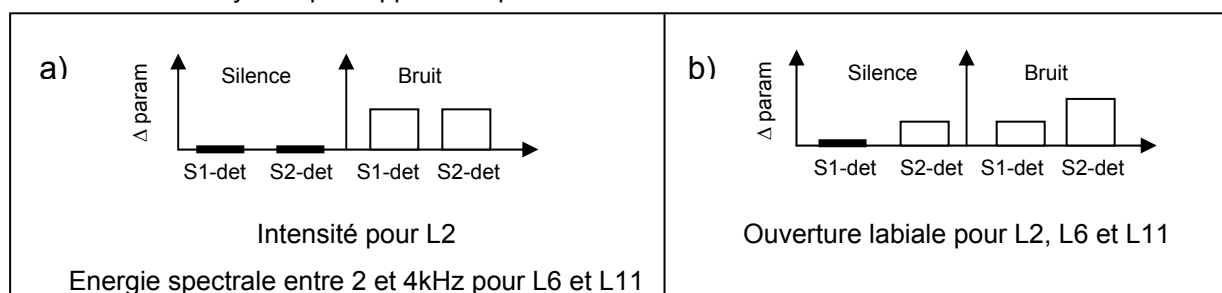


Figure 126. Schématisation de deux cas où les syllabes des mots de contenus ne sont pas en moyenne significativement différentes des syllabes de déterminant (det) dans le silence, mais le deviennent dans le bruit. Dans le 1^{er} cas (a), cela est valable pour les deux syllabes initiales (S1) et finales (S2) du mot de contenu. Dans le 2^{ème} cas (b), il y a création dans le bruit d'une différence entre la syllabe du déterminant et la 1^{ère} syllabe du mot de contenu seulement, tandis qu'une différence existe déjà dans le silence entre la syllabe du déterminant et la 2^{ème} syllabe du mot de contenu, amplifiée dans le bruit.

- (4) Il existe également des cas où cet indice existe déjà dans le silence et est encore renforcé dans le bruit : c'est le cas de la F0 moyenne pour les locutrices L2 et L6.

En réalité, lorsqu'on observe séparément les syllabes initiales et finales des mots de contenu, on remarque que seule la 2^{ème} syllabe du mot de contenu est significativement différente de la syllabe du déterminant dans le silence. En revanche, il se crée dans le bruit une différence entre la syllabe du déterminant et la 1^{ère} syllabe du mot de contenu. Cette différence créée avec la 1^{ère} syllabe devient aussi importante que le renforcement de la différence existant avec la 2^{ème} syllabe, au point que le léger renforcement de la 2^{ème} syllabe du mot de contenu par rapport à la 1^{ère}, observé dans le silence, n'existe alors plus dans le bruit (cf. Figure 127)

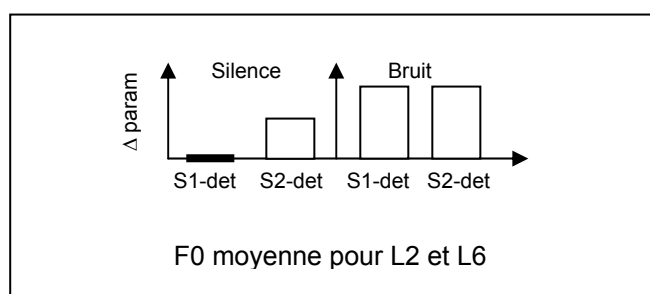


Figure 127. Schématisation du cas où les syllabes des mots de contenus sont en moyenne significativement différentes des syllabes de déterminant (det) dans le silence comme dans le bruit, et où cette différence est amplifiée avec l'exposition au bruit. Il s'agit plus précisément de la création dans le bruit d'une différence entre le déterminant et la syllabe initiale du mot de contenu (S1), et d'une amplification de la différence entre le déterminant et la syllabe finale (S2) du mot de contenu.

3. CONCLUSION

La durée est bien un indice de distinction des mots de contenu par rapport aux mots de fonction pour toutes les locutrices, et ce pour toutes les syllabes du mot de contenu, même si la syllabe finale du mot de contenu est toujours plus allongée que la syllabe initiale. En revanche, cet indice est seulement conservé avec le bruit par toutes les locutrices et non renforcé.

L'intensité sert également à marquer la distinction entre les mots de fonction et de contenu, uniquement pour les locutrices L2 et L6. Cet indice est déjà utilisé par L6 dans le silence et est simplement conservé dans le bruit. Tandis que L2 ne produit pas cet indice dans le silence mais le marque très nettement dans le bruit. Que ce soit pour L2, L6 ou L11, la différence d'intensité entre déterminant et mot de contenu est semblable pour les deux syllabes du mot de contenu.

La F0 moyenne n'est utilisée dans silence que par la locutrice L11 pour différencier les mots de contenu des mots de fonction. En effet, cette locutrice est la seule à montrer une différence significative pour les deux syllabes du mot de contenu tandis que L2 et L6 ne présentent une différence qu'entre le déterminant et la 2^{ème} syllabe, ce qui semble davantage correspondre à un indice de fin d'unité prosodique qu'à un indice de mise en valeur de l'information importante dans l'énoncé. En revanche, toutes les locutrices marquent une distinction significative de F0 dans le bruit entre les mots de contenu et de fonction, cette fois pour les deux syllabes du mot de contenu. Pour résumer, il semble donc que L11 utilise la F0 moyenne comme indice de distinction des mots de contenu des mots outil dans le silence et que cet indice soit conservé dans le bruit, mais pas non plus renforcé, tandis que les locutrices L2 et L6 ne produisent cet indice que dans le bruit.

Seules L6 et L11 renforcent davantage l'énergie spectrale entre 2 et 4kHz par rapport à la zone [0-1kHz] sur les mots de contenu par rapport aux mots de fonction, et uniquement en condition bruyante. Ce renforcement concerne bien les deux syllabes du mot de contenu par rapport à la syllabe du déterminant.

Enfin, l'amplitude des mouvements d'ouverture labiale n'est utilisée par aucune locutrice de la BD2 dans le silence pour renforcer les mots de contenu par rapport aux mots de fonction dans le silence. En revanche, la différence d'ouverture labiale entre les mots de contenu et de fonction augmente significativement avec l'exposition au bruit pour toutes les locutrices, et pour les deux syllabes des mots de contenu, même si la syllabe finale du mot de contenu reste en moyenne toujours plus articulée que la 1^{ère} syllabe.

Une première conclusion de ces différents résultats est la confirmation d'études antérieures sur l'existence d'indices de distinction entre les mots de contenu et les mots de fonction pour plusieurs paramètres, tels que la durée, la F0 ou l'intensité. Nous avons également montré ici que cette distinction entre les mots de fonction et de contenu pouvait être également articulatoire ou spectrale pour certaines locutrices.

Nous avons observé que si certains paramètres sont utilisés par toutes les locutrices (durée, F0, ouverture labiale), il peut également exister des stratégies individuelles de mise en valeur de l'information par le biais d'autres paramètres tels que l'intensité ou l'énergie spectrale de la voix dans les médiums.

De même, nous avons remarqué que si certaines stratégies de renforcement des mots de contenu par rapport aux mots de fonction utilisées à fois dans le silence et dans le bruit (F0, durée) de la même façon ou de façon amplifiée, il existe des paramètres pour lesquels ce renforcement n'est effectué qu'en situation bruyante (ouverture labiale, timbrage). Cela attire notre attention sur le fait que l'adaptation du silence au bruit ne semble pas consister simplement en une amplification graduelle de la parole produite en situation calme, mais également en une réorganisation. Il n'est donc pas du tout évident de modéliser l'effet Lombard sous la forme d'un phénomène graduel et linéaire en fonction du niveau de bruit ambiant. Nous examinerons ce point au chapitre suivant.

Sur le plan individuel, les résultats de cette étude montrent que toutes les locutrices sans exception renforcent davantage les mots de contenu par rapport aux mots de fonction dans le bruit, en créant ou en renforçant avec le bruit des indices de distinction entre ces deux types de mots. En revanche, chaque locutrice semble adopter des stratégies différentes pour cela : L2 semble plutôt renforcer les

mots de contenu par des indices prosodiques acoustiques (F0, intensité, durée). L6 utilise tous les paramètres examinés pour marquer une différence entre des mots de contenu et de fonction, mais renforce cette distinction dans le bruit essentiellement au niveau du timbre de sa voix et de son ouverture labiale. De même, L11 se sert de la F0 et de la durée pour distinguer ces deux types de mots mais se sert surtout du timbre de sa voix et de son ouverture labiale pour amplifier ce contraste dans le bruit.

Pour plusieurs paramètres (Durée, F0 moyenne et Ouverture labiale), nous avons également observé un autre effet prosodique de renforcement de la syllabe finale d'un mot de contenu par rapport à sa syllabe initiale, auquel nous avons pris garde, de façon à ne pas le confondre avec l'effet sémantique que nous cherchions à mettre en évidence dans ce chapitre. Ce renforcement observé va tout à fait dans le sens des résultats d'études antérieures en ce qui concerne l'allongement de la syllabe finale d'une unité prosodique (Delattre 1966^[67] ; Wenk et al. 1982^[378]), la montée mélodique caractéristique de l'accent primaire en fin de syntagme accentuel (Jun et al. 1995^[170] ; Vaissière 2006^[359]) et l'hyper-articulation de la syllabe finale d'un syntagme accentuel (Loevenbruck 2000^[223] ; Tabain 2003^[341]). Ce dernier résultat peut paraître contradictoire avec l'hypo-articulation de la syllabe finale d'un énoncé mise en évidence au chapitre précédent sur le même corpus. Cependant, nous avons considéré dans le chapitre précédant uniquement les mots de contenu en fin d'énoncé, donc à la fin d'un syntagme intonatif, tandis que nous avons ici considéré à la fois ces mêmes mots de contenu en fin d'énoncé mais également les mots de contenu constituant le Sujet des énoncés de la BD2 (« La Lula longe la Lalé », par exemple) et donc plutôt susceptibles d'être placés à la fin d'un syntagme accentuel. La différence de résultats articulatoires avec le chapitre précédent suggère donc qu'il puisse exister, du moins dans notre corpus, une différence de marquage articulatoire entre la fin d'un syntagme accentuel (\Leftrightarrow hyper-articulation de la syllabe finale) et celle d'un syntagme intonatif (\Leftrightarrow hypo-articulation de la syllabe finale), de la même façon qu'au niveau intonatif, la fin d'un syntagme accentuel est marquée par un ton haut de frontière H* et celle d'un syntagme intonatif par un ton bas L* dans le cas d'un énoncé affirmatif. Cependant, Loevenbruck 2000^[223] a plutôt observé que l'articulation suivait la même tendance que la durée : renforcée en fin de syntagme accentuel et encore plus en fin de syntagme intonatif. Pour clarifier ce point, il serait donc maintenant utile de relever les Sujets de nos énoncés constituant un syntagme accentuel (ce qui n'est pas toujours le cas), et d'examiner précisément comment évolue l'articulation de leur syllabe finale, et s'il existe une différence entre le marquage articulatoire de la fin d'un syntagme accentuel et celui de la fin de l'énoncé. Au-delà de cette nouvelle expérience à mener, ces résultats indiquent aussi qu'il aurait peut être été plus prudent de ne pas mélanger ici des mots de contenu situés en milieu et en fin d'énoncé et qu'il serait utile de vérifier si les résultats que nous obtenons en terme de renforcement des mots de contenu par rapport aux déterminants sont bien valables pour ces deux positions des mots de contenu dans l'énoncé.

Chapitre 10 :

**Synthèse quant à l'utilisation
des différentes stratégies
d'adaptation en fonction des
individus et des contextes**

A ce stade de l'étude, nous pouvons faire un bilan des différentes hypothèses explorées et proposer une synthèse de la sélection et de la combinaison de ces différentes stratégies d'adaptation en fonction des individus et des contextes.

Nous commencerons par examiner en quoi les différentes informations apportées tout au long de la thèse permettent de déterminer si les locuteurs subissent l'effort vocal du fait qu'ils adaptent leur parole de façon uniquement réflexe, ou peuvent le gérer dans une certaine mesure, dans le cas où cette adaptation est motivée par la recherche d'intelligibilité vis à vis de l'interlocuteur.

Nous évoquerons ensuite en quoi les différentes modifications de la parole observées dans le bruit, et pouvant être les indicateurs de stratégies de communication, sont susceptibles d'augmenter la charge portée par le larynx, et de quelles façons différentes.

Nous dresserons le bilan des différences interindividuelles rencontrées dans les différents chapitres, et tenterons de les mettre en relation avec des descripteurs de la vibration des cordes vocales ainsi qu'avec la façon dont le locuteur évalue la situation de communication.

Enfin, nous examinerons comment les stratégies et les différences interindividuelles mises en évidence dans les chapitres précédents uniquement à fort niveau de bruit et la plupart du temps dans un bruit de cocktail party, évoluent en fonction des caractéristiques de l'environnement bruyant, en particulier ses caractéristiques spectrales et son niveau d'intensité.

1. PREMIERS BILANS

1.1. Bilan concernant la possibilité de gestion de l'effort vocal par les individus communicant en environnement bruyant

Nous avons vu tout au long de cette 3^{ème} partie que la parole produite en environnement bruyant se caractérise par un ensemble de modifications descriptibles de façon physique (acoustiques, articulatoires, glottiques, posturaux, etc.) entre une condition de silence et une condition de bruit. Certaines de ces modifications globales de la parole, telles que l'élévation de la F0, le renforcement de l'énergie spectrale dans les médiums, l'augmentation de l'ouverture labiale et du premier formant vocalique peuvent être tout simplement liées à l'augmentation de l'intensité vocale dans le bruit. Celle-ci constitue la particularité majeure et la plus classique de la parole Lombard, et peut être à la fois provoquée par un réflexe physiologique de régulation de la voix à partir du retour auditif atténué (Lombard 1911 ^[224]; Pick et al. 1989 ^[271]) ou être motivée par l'intention (pas nécessairement consciente) de préserver l'intelligibilité pour l'interlocuteur en maintenant un rapport signal sur bruit acceptable (Lane et al. 1971 ^[192]).

Au cours des Chapitres 6 et 7, nous avons envisagé une autre interprétation et montré que certaines de ces modifications peuvent également très bien constituer des stratégies de communication en soi visant à préserver l'intelligibilité dans le bruit en renforçant certaines caractéristiques dynamiques, tonales et spectrales de la voix pour lui permettre d'émerger par rapport au bruit ambiant. C'est le cas en particulier de l'augmentation de l'ambitus de F0, de l'ajustement chez les femmes de la F0 au trou spectral du bruit dans lequel elles sont immergées, ou encore du renforcement énergétique de la voix dans la zone sensible de l'oreille. De même, certaines de ces modifications peuvent tout à fait correspondre à des stratégies de communication visant à renforcer le contraste audiovisuel entre les voyelles. C'est le cas de l'évolution dans le bruit de l'ouverture labiale ainsi que de la différence (F1-F0) pour les différentes voyelles, allant dans le sens d'une augmentation de leur distinctivité. Enfin, le ralentissement du débit de parole peut très bien être lié à une stratégie de communication visant à favoriser la récupération des indices phonétiques audiovisuels par l'interlocuteur.

Nos analyses prouvent donc l'existence possible de ces stratégies de communication et le bien-fondé de ces hypothèses, mais ne permettent pas pour autant d'exclure la possibilité que ces modifications soient simplement liées à l'augmentation réflexe ou intentionnelle de l'intensité vocale. Il est d'ailleurs fort probable que la réalité se situe entre les deux. D'autres expériences complémentaires seraient donc nécessaires pour vérifier si l'on retrouve systématiquement ou non ces stratégies de communications dans d'autres conditions de bruit très variées, afin de conclure avec davantage de confiance que ces stratégies de communications existent et que leur observation dans cette étude ne relève pas seulement de la coïncidence.

Cependant, nous disposons dans cette étude d'observations supplémentaires argumentant en faveur de l'interprétation « communicationnelle » et pas uniquement réflexe de ces modifications :

- la plupart de ces modifications évoluent de façon conjointe à l'intensité vocale, sans pour autant que l'on observe toujours une grande corrélation statistique.
- dans le chapitre 4, nous avons mis en évidence que ces modifications de la parole dans le bruit sont plus importantes lorsque le locuteur interagit avec un interlocuteur, et même que certaines d'entre elles n'existent que dans cette condition d'élocution.
- enfin, nous avons également observé que ces modifications ne sont pas semblables pour tous les individus mais montrent des différences interindividuelles significatives. Une fois encore, on peut interpréter ces modifications comme d'origine réflexe et alors les différences interindividuelles comme les indicateurs de différences physiologiques d'audition entre les individus. Cependant, nous avons mesuré l'audiogramme de chacun de nos locuteurs. Ceux-ci ne montrent pas de différence significative qui puisse être reliée à une adaptation plus ou moins importante. Il semble donc plus probable que ces différences interindividuelles d'adaptation soient d'origine cognitive, et par conséquent que l'adaptation de la parole dans le bruit ne soit pas uniquement régie par des mécanismes physiologiques réflexes.

Nous avons jusqu'à maintenant discuté de l'interprétation de certaines modifications comme les indicateurs de stratégies de communication ou être simplement comme étant liées à l'augmentation de l'intensité vocale. Cependant, nous avons également pu mettre en évidence au cours de cette thèse tout un ensemble d'autres modifications de la parole dans le bruit qui elles, n'ont pas de lien a priori avec l'augmentation de l'intensité vocale et argumentent très fortement en faveur d'une recherche d'intelligibilité de la part du locuteur.

Ainsi, nous avons observé plusieurs modifications articulatoires dans le bruit, telles que l'amplification du pincement des lèvres sur les consonnes bilabiales ou le renforcement du contraste entre les voyelles selon leur dimension de protrusion et d'arrondi labial, pouvant difficilement être expliquées par une augmentation de l'intensité vocale.

Par ailleurs, nous avons mis en évidence dans les chapitres 7, 8 et 9 qu'il n'existe pas uniquement des modifications globales de la parole dans le bruit mais également des renforcements acoustiques ou articulatoires spécifiques à certaines unités phonétiques, à certains indices prosodiques de structuration de l'énoncé ou à certains mots portant davantage d'information que d'autres. Ainsi, nous avons observé un renforcement du pincement des lèvres sur les consonnes bilabiales initiales des unités lexicales dans le bruit, pouvant servir d'indice de frontière avec le déterminant qui précède. Nous avons également montré un renforcement dans le bruit de l'allongement des syllabes finales d'un énoncé ainsi que de l'ambitus tonal de la chute intonative finale. Enfin, nous avons mis en évidence une augmentation dans le bruit de la fréquence d'insertion de pauses au sein de l'énoncé, ainsi qu'un renforcement acoustique ou articulatoire, selon les locutrices, des mots de l'énoncé apportant davantage d'information à l'interlocuteur.

La synthèse de ces différents éléments laisse donc fortement penser que, conformément à la proposition initiale de Lane et al. 1971 ^[192], reprise et argumentée par d'autres auteurs (Junqua 1993 ^[171]; Amazi et al. 1982 ^[8]), l'adaptation de la parole dans le bruit ne soit pas uniquement réflexe mais en grande partie motivée par la recherche d'intelligibilité vis à vis de l'interlocuteur.

L'interprétation réflexe ou cognitive des mécanismes sous-tendant l'adaptation de la parole dans le bruit a des conséquences importantes en ce qui concerne la façon d'envisager le phénomène de forçage vocal, puisqu'elle implique que l'effort vocal d'adaptation à un environnement bruyant n'est pas uniquement subi par le locuteur mais peut être en partie géré par le locuteur.

Par ailleurs, puisque le locuteur ne s'adapte pas uniquement de façon réflexe en parlant plus fort, mais aussi de façon cognitive en cherchant à préserver son intelligibilité par différentes stratégies émergentielles et linguistiques, alors son effort d'adaptation n'est pas nécessairement semblable à celui d'un autre locuteur, et peut être lié à la façon dont il évalue la situation de communication et à la façon dont il pense (non consciemment) devoir modifier sa parole pour être intelligible.

Enfin, si l'on considère que le locuteur s'adapte dans le bruit en partie dans le but d'être intelligible pour son interlocuteur, alors dans ce cas, le forçage vocal ne se caractérise pas uniquement par le comportement vocal du locuteur mais aussi par ses stratégies de communication, qui dans les deux cas peuvent avoir des conséquences pour le larynx.

1.2. Bilan concernant la différence de charge que ces modifications de la parole sont susceptibles de faire porter sur le larynx

Les différentes stratégies d'adaptation de la parole dans le bruit ne sont a priori pas équivalentes du point de vue de la charge qu'elles font porter sur le larynx, et donc du point de vue des troubles vocaux qu'elles sont susceptibles d'induire lorsqu'elles sont mises en œuvre de façon répétée.

L'objet de ce paragraphe n'est pas de déterminer quelles stratégies sont saines et lesquelles sont malsaines car le problème est beaucoup plus complexe que cela, et surtout, nous n'avons pas mené dans cette thèse d'expérience nous permettant d'affirmer quoi que ce soit sur ce sujet. Après avoir consacré la majeure partie du manuscrit au recensement et à la caractérisation des différentes stratégies d'adaptation de la parole dans le bruit, notre but est maintenant de montrer en quoi ce travail sur les stratégies d'adaptation est pertinent du point du phénomène de forçage vocal et ouvre de nouvelles perspectives à explorer plus en détails dans des travaux futurs.

Nous allons donc ici envisager, en nous appuyant sur les observations et les résultats d'études antérieures, en quoi les différentes modifications de la parole dans le bruit recensées et caractérisées dans cette thèse sont susceptibles d'augmenter ou d'alléger la charge portée par le larynx, et par quelles différentes façons.

Pour commencer l'augmentation de l'intensité peut faire porter différentes charges au larynx selon la technique vocale avec laquelle elle produite :

- L'intensité vocale peut tout d'abord être accrue en augmentant la pression sous-glottique, entraînant une augmentation de l'amplitude de voisement Av et à une augmentation de la vitesse de fermeture E des cordes vocales lors de leur vibration. La force de collision des cordes vocales est alors plus importante et davantage susceptible de provoquer des inflammations ou des lésions des cordes vocales. Par ailleurs, la contraction des muscles laryngés est plus importante pour résister à cette pression sous-glottique.
- Une autre façon d'augmenter l'intensité vocale consiste à restreindre la surface glottique en compressant davantage les aryténoïdes tout en se servant très peu des muscles expirateurs pour moduler la pression sous-glottique. Dans ce cas, le serrage laryngé est important. La phase ouverte de la vibration glottique est alors assez courte et le son enrichi en hautes-fréquences, contribuant à l'augmentation globale de l'intensité vocale.
- Il existe aussi des techniques résonnantes pour renforcer l'énergie acoustique de la parole dans certaines zones du spectre et ainsi augmenter l'intensité vocale de manière indirecte. C'est le cas du singing formant ou du speaker's formant correspondant à un renforcement du spectre autour de 3 ou 4kHz par le conduit vocal. De même, l'ajustement grâce aux mouvements articulatoires du premier formant vocalique sur un des premiers harmoniques de la voix, aussi appelé Formant Tuning, permet de renforcer l'énergie de la voix dans la zone du fondamental, et ainsi d'augmenter l'intensité vocale globale. Ces deux techniques ne feraient donc a priori pas porter de charge sur la source glottique, et allégeraient même au contraire.

On observe ensuite dans le bruit une élévation de la fréquence fondamentale pouvant simplement être liée à l'augmentation de l'intensité vocale mais constituer une stratégie en soi d'émergence tonale par rapport au bruit ambiant. Dans tous les cas, cette augmentation correspond à un nombre croissant de collisions des cordes vocales par unité de temps, et augmenterait les risques d'inflammations ou de lésions. Par ailleurs, l'augmentation de la F0 en mécanisme 1 est liée à une contraction croissante du muscle vocal, et à son étirement croissant en mécanisme 2. Quel que soit le mécanisme laryngé, l'augmentation de la F0 s'accompagne donc d'une augmentation de la raideur des cordes vocales.

Comme nous l'avons vu aux chapitres 5 et 6, la voix produite en environnement bruyant se caractérise par un déplacement du spectre vers les médiums par rapport à la voix produite en situation calme. Celui-ci peut toujours être lié à l'augmentation de l'effort vocal ou être considéré comme l'indicateur d'une stratégie d'émergence spectrale par rapport au bruit ambiant en renforçant l'énergie dans la zone sensible de l'oreille. Dans les deux cas, cet enrichissement du spectre dans les fréquences médiums peut provenir d'un serrage laryngé plus important s'accompagnant de l'aplatissement de la pente spectrale. Il peut aussi correspondre à un renforcement du spectre par résonance du conduit vocal, sans charge supplémentaire pour le larynx

Nous avons noté chez plusieurs locuteurs et locutrices un ralentissement du débit de parole dans le bruit, pouvant simplement être lié à l'amplification des gestes articulatoires d'ouverture du conduit vocal pour produire des sons de plus forte intensité. L'allongement syllabique peut aussi contribuer à augmenter le temps durant lequel l'interlocuteur perçoit et interprète les indices phonétiques, et ainsi faciliter sa compréhension dans un environnement perturbé. Quoi qu'il en soit, l'augmentation du temps et du pourcentage de voisement accroît bien évidemment le nombre de collisions des cordes vocales pour transmettre une même information à l'interlocuteur, et par la même occasion les risques d'inflammation ou de lésions.

Au niveau articulatoire, nous avons mis en évidence dans le chapitre 7 une augmentation dans le bruit de l'amplitude de pincement des lèvres sur les consonnes bilabiales ainsi que de la durée de ce

pincement. Nous ne disposons pas d'information sur l'impact d'une telle modification sur le larynx. Si celle-ci est purement articulatoire, elle n'entraîne a priori aucune charge sur la source glottique. En revanche, si celle-ci correspond à une très forte augmentation de la pression buccale pendant la phase d'occlusion du conduit vocal, elle peut alors être associée à un serrage laryngé important et à une élévation du larynx pour les occlusives voisées. Ces remarques ne restent que des considérations intuitives et mériteraient d'être explorées scientifiquement.

De même, plusieurs modifications de la parole dans le bruit ne consistent pas tant en l'augmentation globale d'un paramètre sur tout l'énoncé, qu'en son évolution différente selon les catégories vocaliques, selon les syllabes des syntagmes, selon les constituants de l'énoncé ou selon le poids informatif des mots, allant dans le sens d'un renforcement des contrastes au sein de l'énoncé.

Nous avons très peu, voire aucune information sur l'impact de ces forts contrastes sur la charge que cela fait porter au larynx.

- Ainsi, nous ne savons pas du tout si l'augmentation de l'étendue vocale en fréquence et le renforcement des indices intonatifs sont susceptibles de davantage fatiguer le larynx, s'ils n'ont aucune influence par rapport au fait d'utiliser une F0 quasi constante, ou encore s'ils pourraient permettre au contraire d'« assouplir » la production vocale. On constate seulement que les locuteurs en état de fatigue vocale avancée ont tendance à très peu moduler leur fréquence fondamentale (Gelfer et al. 1991 ^[111]), ce qui pourrait être interprété comme une stratégie d'économie vocale. Par ailleurs, Ternström et al. 2006 ^[345] rapportent que les femmes ont une étendue vocale plus importante que les hommes, que ce soit dans le silence comme dans le bruit. Cela pourrait éventuellement expliquer leur sensibilité accrue aux troubles de la voix (Vilkman 2004 ^[365] ; Autesserre et al. 2006 ^[16])
- De même, nous n'avons aucune information quant aux conséquences laryngées de l'augmentation du contraste d'intensité entre les constituants de l'énoncé, ou de la scansion syllabique plus marquée dans le bruit. On peut néanmoins comparer ces stratégies à la conduite d'un automobiliste qui appuierait par intermittente sur l'accélérateur. Cela consommerait non seulement plus de carburant mais userait aussi plus rapidement le moteur ou les amortisseurs. Encore une fois, cette analogie reste de l'ordre de l'intuition et mériterait de faire l'objet d'une étude approfondie.
- Enfin, on pourrait penser intuitivement que l'augmentation de la distinctivité acoustique et articulatoire des voyelles peut constituer une stratégie efficace pour renforcer l'intelligibilité sans demander d'effort supplémentaire à la source glottique. Cependant, ce point de vue se base implicitement sur l'idée d'indépendance entre la source et le filtre (Fant 1960 ^[88]) ce qui n'est pas tout à fait vrai. La configuration du conduit vocal exerce en effet une influence sur la vibration laryngée. C'est ce que nous avons exposé au § 3.2.3 du 1^{er} Chapitre, à travers la théorie d'impédance ramenée sur le larynx. Aussi, une ouverture plus importante de la mâchoire, si elle n'est pas associée à d'autres ajustements (tels que le soulèvement du voile du palais ou le recul de la langue), peut augmenter la charge portée par le larynx. Les chanteurs utilisent ainsi la technique de couverture pour faciliter leur émission vocale dans les aigus et protéger leur larynx. Au contraire, l'augmentation de la protrusion labiale peut contribuer à allonger le conduit vocal et ainsi à faire porter moins de charge sur le larynx. D'après son expérience de rééducation de la voix, Guerin 2004 ^[128] soutient ainsi que l'hyper-articulation des voyelles et le renforcement de leur contraste pourrait être néfaste à la santé vocale, du fait que l'équilibre pneumo-phonique (cf. §3.2.2. du Chapitre 1) entre le timbre de la voix, la pression sous-glottique et la tension laryngée serait dans ce cas constamment perturbé et que le larynx en ferait alors les frais. Cette hypothèse n'a cependant jamais été démontrée scientifiquement. Mais certaines techniques de chant lyrique vont effectivement dans de même sens, en cherchant un placement de voix homogène, le plus indépendant possible des voyelles, au point d'ailleurs de déformer légèrement leur prononciation. Nous n'avons pas connaissance de techniques semblables pour les acteurs de théâtre où l'intelligibilité est essentiel contrairement au domaine du chant lyrique.

Pour résumer, les modifications de la parole que nous avons observées dans le bruit, qu'elles soient induites par des mécanismes physiologiques réflexes ou par des mécanismes cognitifs, sont susceptibles d'augmenter la charge portée par le larynx de façon différente : en augmentant par exemple la tension musculaire du vocalis, celle nécessaire à la compression des aryténoïdes ou à la bascule du cartilage cricoïdien, en augmentant la force ou le nombre de collisions des cordes vocales, etc. Il n'est donc pas du tout évident de considérer ces différentes stratégies sur le même plan, tant au niveau du gain d'intelligibilité qu'elles peuvent apporter que de la charge qu'elles font porter sur le larynx ou des troubles vocaux qu'elles peuvent entraîner (fatigue neuromusculaire, échauffement des cordes vocales, œdème localisé).

Par ailleurs, nous remarquons que certaines stratégies ne peuvent pas être réalisées de plusieurs façons. Leur efficacité ou leur nocivité pour la voix peut donc être évaluée en soi. Au contraire, d'autres stratégies, telles que l'augmentation de l'intensité vocale, peuvent avoir plusieurs réalisations motrices, et ainsi ne pas être en soi bonnes ou mauvaises, mais le devenir selon la façon dont elles sont gérées par le locuteur.

Enfin, il serait bien sûr simpliste et inutile de vouloir évaluer l'efficacité ou la nocivité de ces diverses stratégies séparée, puisque celles-ci ne sont pas adoptées de façon isolée mais combinée. La charge portée par le larynx, de même que l'intelligibilité globale, résulte de l'addition et de la compensation de ces différents ajustements. Il est néanmoins intéressant d'essayer de clarifier de façon théorique l'impact différent que peuvent avoir ces différentes stratégies sur le larynx.

En l'absence d'information supplémentaire, nous nous garderons donc d'exprimer des jugements de valeurs sur l'efficacité de communication ou la nocivité de ces différentes stratégies, qu'il conviendra d'explorer dans des travaux futurs. Nous nous contentons ici de constater que toutes ces stratégies ne sont pas équivalentes pour le larynx et donc que l'adoption de ces stratégies avec un « dosage » différent selon les locuteurs peut contribuer à leur sensibilité variable face aux troubles de la voix.

1.3. Premier bilan sur les différences interindividuelles d'adaptation dans le bruit

Conformément à notre hypothèse de départ, nous avons bien observé que les individus ne s'adaptent pas tous de la même façon à un environnement bruyant. Ces différences sont de plusieurs ordres :

- En premier lieu, il existe des stratégies adoptées seulement par certains locuteurs, et non par d'autres. C'est le cas de l'élargissement de l'étendue tonale dans le bruit, non effectué par le locuteur L7 de la BD1 ni par L2 dans la BD2 (cf. S2 sur la Figure 128 et la Figure 130), ainsi que de l'allongement de la durée syllabique, observé chez tous les individus exceptée L1 dans la BD1 et L11 dans la BD2 (cf. S9 sur la Figure 128 et la Figure 129). Enfin, L6 et L11 n'augmentent pas dans le bruit le contraste d'intensité entre les mots de contenu et les mots de fonction (cf. S11 sur la Figure 128), de même que L11 n'augmente pas non plus ce contraste par la F0 (cf. S16 sur la Figure 128). En dehors de ces exceptions, on observe surtout que chaque locuteur n'adopte pas une unique stratégie mais quasiment toutes les stratégies examinées. Les différences interindividuelles d'adaptation au bruit ne résident donc pas dans le fait que chaque locuteur adopte une stratégie de communication différente mais plutôt dans le fait que chaque locuteur ne privilégie pas la même stratégie de communication ou n'adopte pas la même combinaison de stratégies.
- Par ailleurs, on observe que tous les locuteurs ne s'adaptent pas avec la même magnitude du silence au bruit : Des locuteurs montrent une marge d'évolution plus importante que d'autres sur certaines stratégies et au contraire une marge moins importante sur d'autres stratégies (cf. Tableau 30, Tableau 31, Figure 128, Figure 129 et Figure 130). Ainsi, on voit par exemple dans la BD1 que la locutrice L1 et le locuteur L8 augmentent davantage leur intensité vocale que les autres locuteurs dans le bruit, tandis qu'au contraire les locuteurs L3, L4 et L5 jouent davantage sur le ralentissement du débit de parole que les autres. Cela rappelle les résultats obtenus par Van Summers et al. 1988^[361] où un de leurs locuteurs augmentait principalement sa F0 et son premier formant vocalique dans le bruit, tandis qu'un autre locuteur augmente plutôt la durée. Dans notre première base de données, nous observons également une certaine ressemblance entre les locutrices L2, L9 et L10 augmentant davantage leur ambitus de F0 que les autres dans le bruit.

De façon globale, on voit que pour certains locuteurs tels que L8 ou L9, « l'hyperadaptation » dans le bruit en ce qui concerne une stratégie de communication est compensée par une « hypo-adaptation » sur une autre. En revanche, on observe également des locuteurs, comme L4, ayant tendance à s'adapter davantage que les autres locuteurs sur un grand nombre de stratégies, et à l'inverse des locuteurs tels que L6 ou L7 ayant tendance à moins s'adapter que d'autres de manière globale. Cette constatation est intéressante vis à vis du forçage vocal puisqu'elle suggère que, face à une même situation de communication, certains individus fournissent une adaptation supérieure à d'autres, correspondant à un effort vocal potentiellement plus important.

	BD1										BD2							
	Femmes					Hommes					L2	L6	L11					
	L1	L2	L6	L9	L10	L3	L4	L5	L7	L8				S10 Dynamique syllabique	L2	L6	L11	
S1 Intensité	21.0 dB	17.3 dB	17.0 dB	12.3 dB	15.9 dB	17.5 dB	16.4 dB	20.0 dB	19.0 dB	22.2 dB	S1 Intensité	14.6 dB	21.9 dB	12.8 dB	S10 Dynamique syllabique	6.1 dB	6.6 dB	2.3 dB
S2 Ambitus F0	3.6 tons	5.6 tons	3.7 tons	4.6 tons	5.1 tons	3.1 tons	2.6 tons	2.8 tons	2.6 tons	3.0 tons	S2 Ambitus F0	1.5 tons	1.0 tons	0.6 tons	S11 Indice début de mot par le pincement labial	7.6 %	12.4 %	13.8%
S3 F0 cible	5.6 tons	5.1 tons	5.6 tons	5.3 tons	5.8 tons	5.1 tons	6.4 tons	5.5 tons	4.8 tons	3.7 tons	S3 F0 cible	129 Hz	191 Hz	101 Hz	S12 Indice fin d'énoncé par durée syllabique	15 ms	22 ms	31 ms
S4 Timbrage	19.1 dB	14.4 dB	9.5 dB	11.8 dB	10.1 dB	9.7 dB	18.6 dB	13.0 dB	12.3 dB	14.5 dB	S4 Timbrage	11.3 dB	9 dB	10.1 dB	S13 Indice fin d'énoncé par chute mélodique	2.4 tons	1 ton	1 ton
S9 Durée syllabique	42 ms	55 ms	59 ms	30 ms	54 ms	88 ms	78 ms	87 ms	42 ms	57 ms	S5 Contraste B	14.5 %	14.2 %	12.0 %	S14 Indice de mot de contenu par intensité	2.2 dB	0.4 dB	0.6 dB
S10 Dynamique syllabique	6.2 dB	8.1 dB	3.3 dB	3.9 dB	4.4 dB	8.0 dB	6.2 dB	5.9 dB	4.4 dB	7.0 dB	S6 Contraste P1	55.9 %	52.3 %	46.5 %	S15 Indice de mot de contenu par durée syllabique	18.5 ms	13.4 ms	19.8 s
											S7 Contraste AR	18.1%	26.5 %	8.7 %	S16 Indice de mot de contenu par F0	0.8 tons	0.3 tons	0.3 tons
											S8 Pincement labial	10.7 %	28.1 %	21.2 %	S17 Indice de mot de contenu par timbrage	3.9 dB	3 dB	3.6 dB
											S9 Durée syllabique	46 ms	55 ms	25 ms	S18 Indice de mot de contenu par ouv. labiale	5.7 %	9 %	4 %

Tableau 30. Marges d'évolution de chaque locuteur et locutrice de BD1 et BD2 pour différentes stratégies de communication dans le bruit. Ces marges correspondent aux différences entre les valeurs maximales et minimales observées de ces stratégies sur l'ensemble des conditions de chaque base de données (silence, bruit de type et de niveau d'intensité variable, cf. Figure 128, Figure 129 et Figure 130).

	BD1										BD2							
	Femmes					Hommes					L2	L6	L11					
	L1	L2	L6	L9	L10	L3	L4	L5	L7	L8				S10 Dynamique syllabique	L2	L6	L11	
S1 Intensité	75.1 dB	75.3 dB	81.0 dB	89.1 dB	86.3 dB	85.1 dB	83.7 dB	78.3 dB	84.0 dB	81.5 dB	S1 Intensité	87.3 dB	85.1 dB	91.5 dB	S10 Dynamique syllabique	3.4 dB	3.9 dB	3.7 dB
S2 Ambitus F0	4.1 tons	4.1 tons	6.6 tons	3.1 tons	7.8 tons	4.9 tons	3.5 tons	3.8 tons	4.2 tons	4.8 tons	S2 Ambitus F0	7.1 tons	5.2 tons	5.7 tons	S11 Indice début de mot par le pincement labial	-0.8%	3.2 %	-3.7 %
S3 F0 cible	186 Hz	188 Hz	207 Hz	204 Hz	199 Hz	111 Hz	91 Hz	105 Hz	104 Hz	111 Hz	S3 F0 cible	221 Hz	233 Hz	197 Hz	S12 Indice fin d'énoncé par durée syllabique	63 ms	59 ms	22 ms
S4 Timbrage	-30.3 dB	-29.9 dB	-28.0 dB	-27.4 dB	-29.6 dB	-22.4 dB	-26.7 dB	-26.0 dB	-23.6 dB	-27.7 dB	S4 Timbrage	-22.5 dB	-16.2 dB	-16.3 dB	S13 Indice fin d'énoncé par chute mélodique	2,8 tons	2,7 tons	1,8 tons
S9 Durée syllabique	212 ms	219 ms	196 ms	188 ms	199 ms	223 ms	169 ms	247 ms	226 ms	177 ms	S5 Contraste B	17.7 %	15.0 %	12.5 %	S14 Indice de mot de contenu par intensité	-0,1 dB	1,5 dB	0,7 dB
S10 Dynamique syllabique	9.9 dB	8.7 dB	9.4 dB	5.8 dB	6.9 dB	9.2 dB	6.1 dB	10.5 dB	6.9 dB	6.7 dB	S6 Contraste P1	67.6 %	36.4 %	47.0 %	S15 Indice de mot de contenu par durée syllabique	23 ms	63 ms	30 ms
											S7 Contraste AR	13.7 %	7.8 %	9.9 %	S16 Indice de mot de contenu par F0	0,6 tons	0,9 tons	0,9 tons
											S8 Pincement labial	15.6 %	17.2 %	10.5 %	S17 Indice de mot de contenu par timbrage	-0,3 dB	-0,5 dB	-1,0 dB
											S9 Durée syllabique	191 ms	163 ms	151ms	S18 Indice de mot de contenu par ouv. labiale	-1.8 %	0.8 %	-0.8 %

Tableau 31. Différences inter-individuelles quant à l'utilisation de plusieurs stratégies de communication dans le silence (cf. Figure 128, Figure 129 et Figure 130). Les cases rouges foncées mettent en valeur les locuteurs montrant une valeur supérieure aux autres (> moyenne + écart type sur l'ensemble des locuteurs d'une même base de données) et celles en vert clair, les locuteurs montrant une valeur plus basse que les autres (< moyenne - écart-type).

- L'observation de cette marge d'évolution est encore plus intéressante en regard de la variabilité interindividuelle déjà observée dans le silence.

Ainsi, on remarque que dans de nombreux cas, l'adaptation d'un locuteur moins importante que d'autres sur une stratégie peut s'expliquer par le fait qu'il utilise déjà cette stratégie de façon importante dans le silence et doit donc difficilement pouvoir surenchérir dans le bruit. C'est le cas par exemple des locutrices L9 et L11 montrant une augmentation d'intensité moins importante que les autres dans le bruit, mais parlant déjà avec une intensité vocale élevée dans le silence (cf. Tableau 30, Tableau 31, Figure 128, Figure 129 et Figure 130). De même L3 enrichit moins que d'autre l'énergie spectrale de sa voix dans les médiums mais sa voix est déjà fortement timbrée dans le silence.

A l'inverse, l'hyper-adaptation d'un locuteur sur une stratégie de communication correspond dans plusieurs cas au fait que ce locuteur utilise peu ou pas cette stratégie dans le silence. C'est le cas par exemple pour l'intensité vocale ou le timbrage chez la locutrice L1, pour la F0 et la durée syllabique chez le locuteur L4 ou le timbrage chez la locutrice L2 dans la 2^{ème} base de données (cf. Tableau 30, Tableau 31, Figure 128, Figure 129 et Figure 130).

Ces observations sont d'un grand intérêt vis à vis du forçage vocal car elles suggèrent que le forçage vocal puisse ne pas consister uniquement en une dépense d'énergie plus importante chez certains locuteurs pour s'adapter à une situation de communication perturbée, mais également en une forte dépense d'énergie constante chez certains individus pour communiquer, quelle que soit la situation dans laquelle ils se trouvent, et donc sans adaptabilité de l'effort vocal aux exigences d'intelligibilité de la situation de communication.

Il existe néanmoins quelques cas où le locuteur s'adapte davantage que d'autres sur une stratégie mais utilise également davantage cette stratégie que les autres dans le silence. C'est le cas de l'ambitus de F0 chez L10 et L2 dans BD2, ou de la durée syllabique chez L5. On peut penser que l'augmentation de ces paramètres n'est pas limitée par des contraintes physiologiques ou que d'après leurs valeurs, ils se situent bien en dessous des limites physiologiques du locuteur, contrairement au timbrage ou à l'intensité vocale. Cela attire encore une fois notre attention sur le fait que l'adaptation des locuteurs dans le bruit est susceptible d'être restreinte par leur physiologie, leur technique vocale ou par des contraintes d'ordre phonologique. Il est donc fort possible que certaines stratégies ne puissent pas être augmentées indéfiniment en fonction d'un bruit ambiant d'intensité croissante. Nous examinerons dans le paragraphe 2 comment évolue chaque stratégie en fonction du niveau de bruit, s'il existe des « saturations » de certaines stratégies et des réorganisations conséquentes de l'adaptation.

L'examen des valeurs moyennes des stratégies par chaque locuteur de BD1 permet également de distinguer plusieurs profils individuels (cf. Figure 129 et Figure 130). Ainsi, les locutrices L9 et L10 ainsi que les locuteurs L3, L4, L7 et L8 parlent à plus forte intensité que les autres, et cela déjà dans le silence, tandis qu'ils ne jouent pas beaucoup sur la durée syllabique ni sur la dynamique syllabique, que ce soit dans le silence ou dans le bruit.

Au contraire, les locutrices L1 et L2, ainsi que le locuteur L5 sont davantage économes au niveau de l'intensité vocal. En revanche, ils montrent un débit de parole plus lent et une dynamique d'intensité par syllabes plus importante que les autres.

Enfin, la locutrice L6 se distingue par une intensité vocale moyenne, combinée à un allongement syllabique et à une dynamique syllabique plutôt importantes.

Il semble donc exister une certaine complémentarité entre l'utilisation de stratégies acoustiques telles que l'augmentation de l'intensité vocale, et l'utilisation de stratégies plus linguistiques comme l'allongement syllabique ou la scansion par syllabe. On retrouve en effet ce type de complémentarité dans la BD2, où la locutrice L11 est celle montrant l'intensité vocale et le timbrage de la voix les plus importants par rapport aux autres locutrices. En contrepartie, cette locutrice présente également les plus faibles contrastes articulatoires entre les voyelles (S5, S6 et S7 sur la Figure 128), un débit de parole rapide ne ralentissant quasiment pas dans le bruit (S9), une faible dynamique d'intensité des syllabes (S10) ainsi qu'un marquage plus faible que les autres de la fin de l'énoncé, aussi bien au niveau de l'allongement syllabique (S12) que de la chute mélodique finale (S13). A l'inverse, la locutrice L2 parle un peu moins fort que les deux autres locutrices, et timbre moins sa voix. En revanche, elle est celle qui présente le plus de distinctivité visuelle entre les voyelles, qui montre le débit de parole le plus lent, et elle marque nettement la fin de l'énoncé au niveau de la durée syllabique (S12) et de l'ambitus tonal de la chute intonative (S13). De façon similaire à la BD1, L6 a un profil intermédiaire entre ces deux extrêmes, et utilise toutes les stratégies de façon moyenne. Elle se distingue également des autres locutrices par un contraste de durée très marqué entre les mots de contenu et de fonction.

On remarque que ces différents profils individuels correspondent également à des comportements laryngés différents. Les locuteurs tels que L11, L7, L8, L9 et L10 utilisant leur voix à forte intensité et beaucoup moins des stratégies « linguistiques » montrent un quotient ouvert plus faible que les autres, et tendant à diminuer avec l'effort vocal. Perceptivement, il semble que les locutrices L9 et L10 utilisent très probablement leur 1^{er} mécanisme laryngé et les valeurs de leur quotient ouvert inférieures à 0.5 argumentent également dans ce sens. Au contraire, les locutrices L1 et L2, jouant davantage sur le débit de parole que sur l'intensité vocale, présentent un quotient ouvert plus haut que les autres, restant quasiment constant avec l'effort vocal pour L1 et augmentant même pour L2. Ce dernier résultat est assez surprenant, étant donné que ces deux locutrices semblent, à l'écoute, utiliser leur mécanisme 2. Cependant, d'autres études antérieures observent également une corrélation positive entre la F0 et le quotient ouvert chez certaines femmes (des mezzo-sopranoes sans grand entraînement vocal, chez Henrich 2001^[139]), aussi bien en mécanisme 1 qu'en mécanisme 2. De même, nous avons observé que ces profils individuels étaient reliés à différents mouvements verticaux du larynx : L11 parlant à forte intensité montre un larynx assez haut que ce soit dans le silence ou dans le bruit. L2, qui utilise davantage l'allongement syllabique à un larynx s'élevant de façon moyenne par rapport à sa position au repos, sans modification significative dans le bruit. Enfin, L6 montre la plus forte élévation du larynx en situation d'effort, passant d'une position assez basse dans le silence à une position haute comparable à L11 dans le bruit.

Une dernière conclusion générale quant aux différences interindividuelles concerne l'absence de regroupement de locuteurs aux profils semblables par genre. Dans cette étude, nous avons rarement observé de comportements spécifiques à l'un des deux genres, excepté un ajustement de la F0 dans le trou spectral du bruit cocktail chez les femmes seulement (cf. Chapitre 6) ainsi qu'un ajustement plus important chez les femmes du 1^{er} formant des voyelles sur un harmonique de la voix.

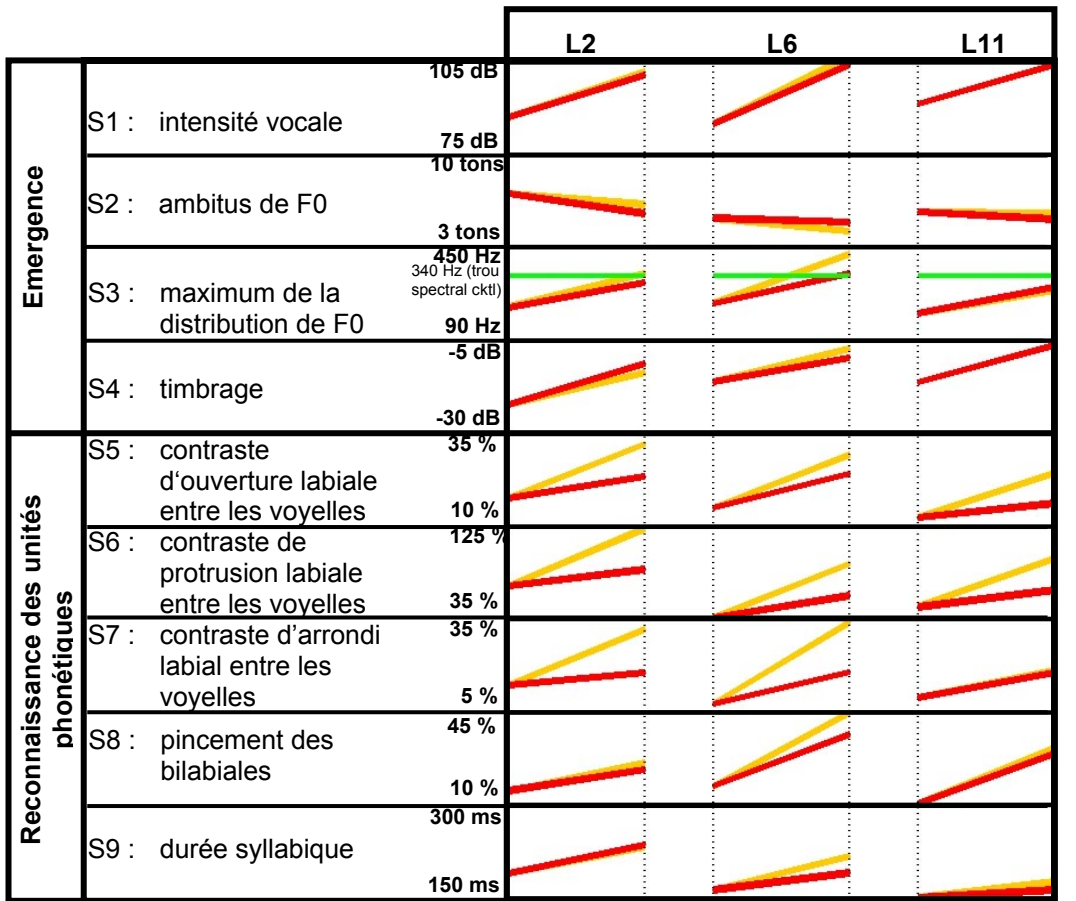
En revanche, la locutrice L2 se distingue des autres locuteurs par un profil laryngé atypique. Il se trouve que cette locutrice est la seule à avoir une utilisation semi-professionnelle de sa voix en tant qu'actrice et lectrice. Il est donc envisageable que sa particularité comportementale soit liée à son expertise.

2. REORGANISATION DE L'ADAPTATION EN FONCTION DU CONTEXTE

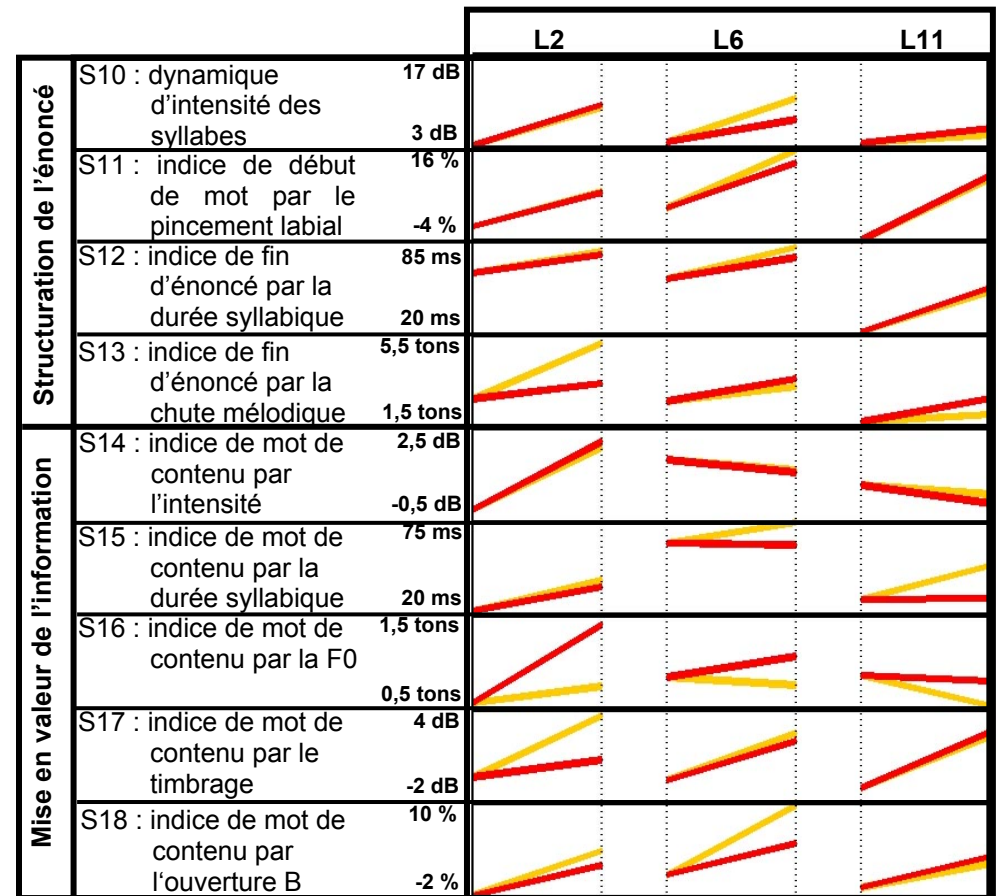
Dès le chapitre 6, nous avons abandonné une description paramétrique de l'adaptation de la parole dans le bruit selon chaque descripteur de la parole, pour privilégier la caractérisation de stratégies d'adaptation, correspondant à l'évolution simultanée de plusieurs paramètres acoustiques et articulatoires. De la même façon, nous venons de voir que l'adaptation globale des locuteurs dans le bruit ne consiste pas en l'adoption exclusive d'une seule stratégie de communication, ni en l'adoption de plusieurs stratégies en parallèle, mais en leur sélection et leur combinaison selon les locuteurs. De la même façon nous faisons l'hypothèse que ces sélections et ces combinaisons de stratégies puissent être différentes selon le contexte, en particulier selon le type et le niveau de bruit. Dans les chapitres précédents, nous avons quasiment toujours considéré l'adaptation de la parole dans un bruit cocktail de 85 ou 86dB (pour la BD1 et la BD2) pour examiner l'existence ou non de plusieurs stratégies potentielles de communication dans le bruit. Il est donc maintenant nécessaire d'examiner comment ces sélections et combinaisons de stratégies évoluent pour chaque locuteur en fonction du type de bruit et du niveau de bruit, et s'il existe une réorganisation de l'adaptation en fonction du contexte ou une évolution graduelle de celle-ci.

Pour cela, nous avons représenté sur la Figure 128, la Figure 129 et la Figure 130 l'évolution des principales stratégies de communications dans le bruit mises en évidence au cours de cette thèse, pour les locuteurs et locutrices des deux bases de données, et cela pour les deux types de bruit (blanc et de conversations) à un seul niveau de bruit (85dB) pour la BD2 et pour 4 niveaux de bruit croissants (62, 70, 78 et 86 dB) dans le cas de la BD1.

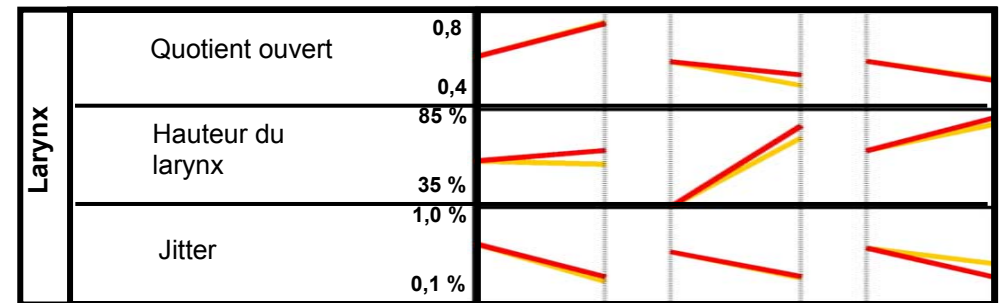
Locutrices de BD2



Silence Bruit Silence Bruit Silence Bruit



Silence Bruit Silence Bruit Silence Bruit



Silence Bruit Silence Bruit Silence Bruit

Figure 128. Représentation, pour les 3 locutrices de la BD2, de l'évolution de plusieurs stratégies de communication et de plusieurs descripteurs laryngés, du silence à un bruit blanc de 85dB (en orange clair) et du silence à un bruit de conversations à 85dB (en rouge foncé). Pour la stratégie S3, le trait vertical vert clair à 340Hz représente le trou spectral du bruit cocktail vers lequel tend la F0 lorsque les locutrices s'expriment dans ce type de bruit.

Locutrices de BD1

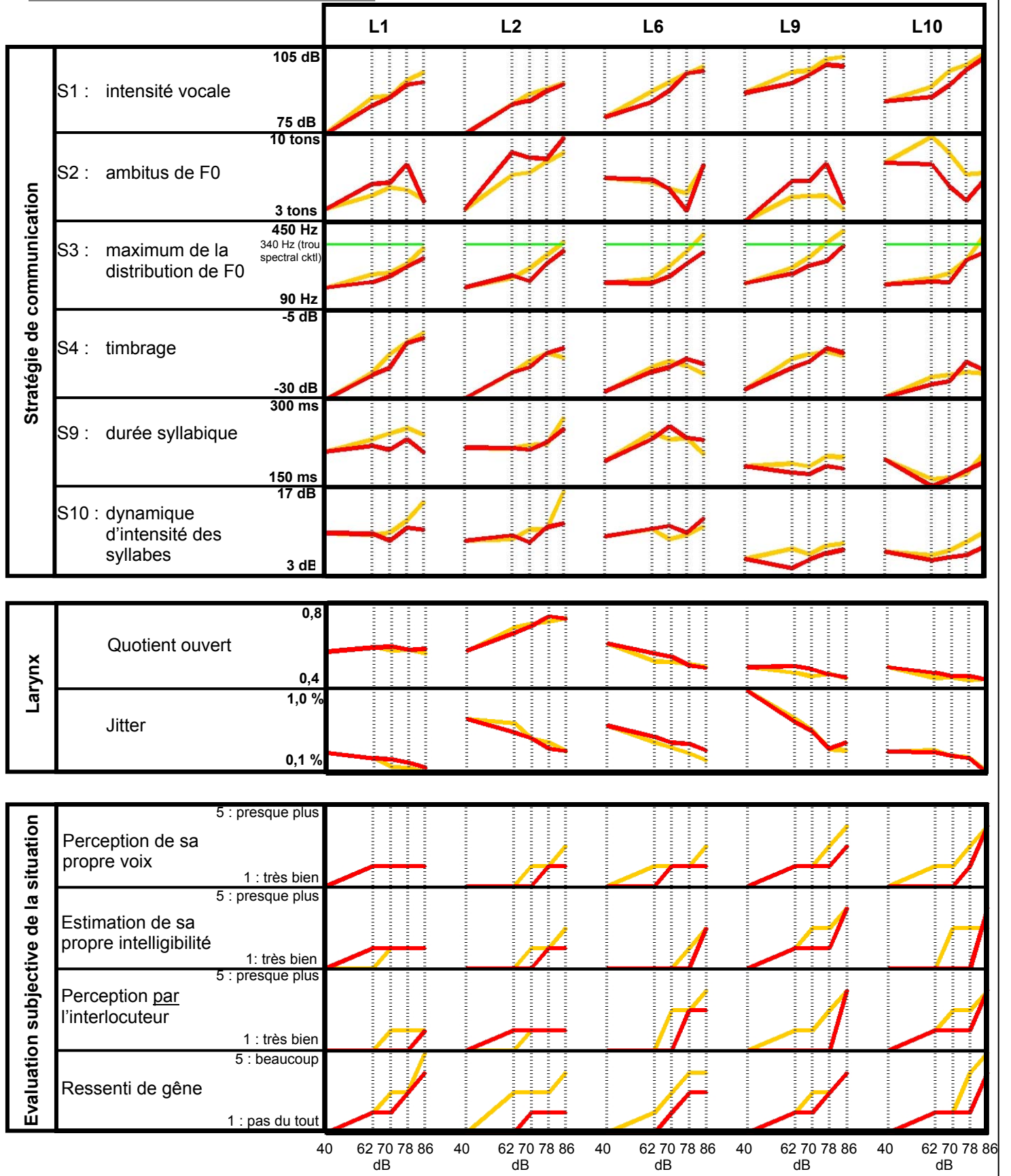


Figure 129. Représentation, pour les 5 locutrices de la BD1, de l'évolution de plusieurs stratégies de communication, de plusieurs descripteurs laryngés, et de plusieurs paramètres subjectifs en fonction du niveau de bruit ambiant (40dB ⇔ calme, puis 62, 70, 78 et 86dB de bruit), pour un bruit blanc (en orange clair) et un bruit de conversations (en rouge foncé). Pour la stratégie S3, le trait vertical vert clair à 340Hz représente le trou spectral du bruit cocktail vers lequel tend la F0 lorsque les locutrices s'expriment dans ce type de bruit.

Locuteurs de BD1

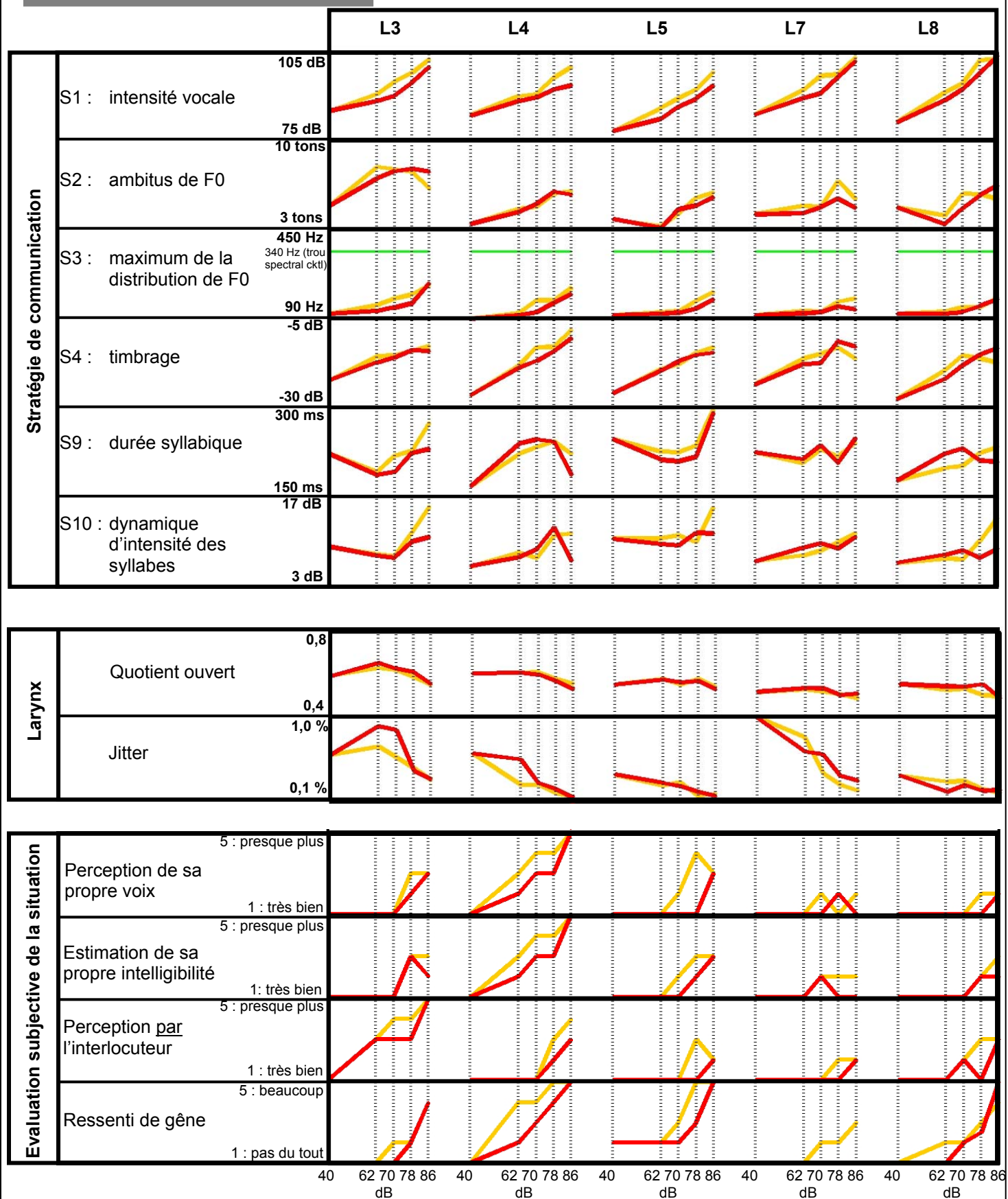


Figure 130. Représentation, pour les 5 locuteurs de la BD1, de l'évolution de plusieurs stratégies de communication, de plusieurs descripteurs laryngés, et de plusieurs paramètres subjectifs en fonction du niveau de bruit ambiant (40dB ⇔ calme, puis 62, 70, 78 et 86dB de bruit), pour un bruit blanc (en orange clair) et un bruit de conversations (en rouge foncé). Pour la stratégie S3, le trait vertical vert clair à 340Hz représente le trou spectral du bruit cocktail vers lequel tend la F0 lorsque les locutrices s'expriment dans ce type de bruit.

2.1. Influence du type de bruit

On observe sur les Figures précédentes une différence entre le bruit blanc et le bruit cocktail. Cette différence ne concerne pas l'adoption d'une stratégie exclusivement dans un type de bruit, ni tellement l'allure de l'évolution des stratégies du silence au bruit en fonction d'un niveau de bruit, qui reste assez semblable pour chaque locuteur entre les deux types de bruit. En revanche, l'adaptation de la parole dans le bruit montre très souvent une différence de magnitude entre le bruit blanc et le bruit de conversations.

Dans la 1^{ère} base de données, l'augmentation de l'intensité vocale (S1), de la F0 cible (S3), du timbrage (S4) et de la dynamique d'intensité des syllabes (S10) est plus importante dans le bruit blanc que dans le bruit de cocktail-party chez tous les locuteurs et locutrices. En revanche, l'augmentation de l'ambitus de F0 (S2) est plus importante dans le bruit cocktail pour les locutrices L1, L2 et L6, au contraire plus importante dans le bruit blanc pour la locutrice L10 et les locuteurs L7 et L8, et semblable entre les deux types de bruit pour la locutrice L9 et les locuteurs L3, L4 et L5. De même, la durée syllabique (S9) est plus importante dans le bruit blanc pour les locutrices L1 et L9, ainsi que pour les locuteurs L3 et L5, et relativement semblable entre les deux types de bruit pour les autres locuteurs et locutrices.

Dans la 2^{ème} base de données, l'intensité vocale (S1) n'est pas significativement différente entre les deux types de bruit chez toutes les locutrices, de même que le timbrage (S4), l'ambitus de F0 (S2), l'allongement syllabique en fin d'énoncé (S12) ou le contraste d'intensité entre les mots de fonction et de contenu (S14). Similairement, la F0 cible (S3), le pincement des lèvres sur les occlusives bilabiales (S8), la durée syllabique (S9), la dynamique d'intensité des syllabes (S10) et le renforcement du pincement labial en début de mot (S11) ne sont pas différents entre les deux types de bruit pour les locutrices L2 et L11, mais sont en revanche tous plus importants dans le bruit blanc pour la locutrice L6. Les autres stratégies présentent globalement une évolution plus importante dans le bruit blanc que dans le bruit de conversations, à l'exception de quelques cas individuels : ainsi, L11 ne renforce pas son contraste d'arrondi labial entre les voyelles (S7) de façon différente selon le type de bruit, contrairement aux autres, l'augmentation de l'ambitus tonal de la chute mélodique en fin d'énoncé (S13) et le contraste de timbrage entre les mots de contenu et de fonction (S17) ne sont pas sensibles au type de bruit chez les locutrices L6 et L11. De même pour le contraste d'ouverture labiale entre les mots de contenu et de fonction (S18) pour la locutrice L11.

Dans les Chapitres 5 et 6, nous avons déjà observé pour certains paramètres cette tendance à évoluer de façon plus importante dans le bruit blanc que dans le bruit cocktail. Le fait que cette différence ne soit pas observée uniquement pour des stratégies de type « émergentielles » (Augmentation globale de l'intensité, du timbrage, de l'ambitus de F0, etc.) mais également pour des stratégies plutôt « linguistiques » (contraste audiovisuel entre les voyelles, contraste entre les mots de fonction et de contenu, indices prosodiques de structuration de l'énoncé) laisse penser que le type de bruit n'influence pas seulement l'adaptation du locuteur par un masquage différent des propriétés spectro-temporelles de sa parole, mais de façon plus générale par une atténuation différente du retour auditif que le locuteur a de sa propre voix, ou par une gêne différente ressentie par le locuteur. Plusieurs arguments étayaient cette possibilité : on voit effectivement sur les Figure 129 et Figure 130 que la majorité des locuteurs et locutrices ressentent une gêne plus importante dans le bruit blanc que dans le bruit cocktail. De même, la plupart d'entre eux perçoivent moins bien leur propre voix dans le bruit blanc que dans le bruit cocktail. Nous rappelons que nous avons calibré le niveau d'intensité des deux types de bruit pour que celui-ci soit équivalent en dB SPL au niveau des oreilles des locuteurs (cf. Annexe Met2). En revanche, leur contenu spectral n'est pas semblable : le bruit blanc présente davantage d'énergie dans la zone sensible de l'oreille, ce qui peut expliquer qu'il soit perçu comme plus fort, et provoque par conséquent une adaptation plus importante que le bruit de cocktail, dont l'énergie se concentre principalement dans les basses-fréquences.

Le fait que le contenu spectral du bruit ambiant puisse exercer une influence globale sur l'intensité perçue et sur la gêne du locuteur n'exclue pas la possibilité qu'il masque également de façon différente la parole produite par le locuteur et que celui-ci cherche alors à émerger du bruit de façon différente. Plusieurs études ont en effet montré que le contenu spectral du bruit influençait les modifications du spectre de la voix et de certaines unités phonétiques dans le bruit (Mokbel 1992^[245] ; Egan 1972^[80]). Nous n'avons pas observé de tels comportements dans nos bases de données. En revanche, nous avons observé que les femmes montrent dans le bruit cocktail à fort niveau une distribution de F0 ajustée au niveau du trou spectral de ce bruit ambiant (cf. Figure 129). En l'absence d'expériences complémentaires, nous resterons prudents quant à ce résultat : Bien que celui-ci

semble reproductible puisqu'il se retrouve chez toutes les locutrices, il reste contrebalancé par le fait que la F0 cible augmente également dans le bruit blanc et que l'on retrouve une même augmentation chez les hommes bien que ceux-ci soient beaucoup trop bas en fréquence pour pouvoir atteindre le trou spectral du bruit cocktail. Il serait donc intéressant de mener une expérience sur ce phénomène, en examinant un niveau de bruit supérieur à 86dB pour voir si les locutrices continuent à augmenter leur F0 et dépassent donc le trou spectral du bruit cocktail, ou bien si elles restent ajustées au niveau de ce trou à très fort niveau de bruit. De même, il serait intéressant de faire varier la position fréquentielle de ce trou spectral de façon à examiner si les hommes peuvent également utiliser une telle stratégie lorsqu'elle leur est accessible, et si la position de ce trou influence l'évolution de la fréquence fondamentale dans le bruit chez les locuteurs.

2.2. Influence du niveau de bruit

2.2.1. Profil d'évolution de chaque stratégie en fonction du niveau de bruit

De façon générale, on observe que la plupart des stratégies sont utilisées avec une magnitude croissante en fonction du niveau de bruit (cf. Figure 129 et Figure 130). Pour l'intensité vocale, la F0 ou le timbre de la voix cette croissance est strictement monotone, et peut être en première approximation considérée comme quasi-linéaire en fonction de niveau de bruit. Elle montre cependant chez de nombreux locuteurs une saturation à fort niveau de bruit, ou au contraire un pallier sans grande évolution à faible niveau de bruit puis une augmentation à partir d'un niveau de bruit moyen.

On recense également de nombreux cas où la stratégie n'évolue pas de façon monotone : c'est le cas par exemple de l'ambitus de F0 chez L1, L9 et L3, ou de la durée syllabique chez L4 qui commencent par augmenter avec un niveau de bruit croissant puis diminuent à fort niveau de bruit (cf. Figure 129 et Figure 130). A l'inverse, l'ambitus de F0 commence par diminuer chez L6, L5 et L8 pour ensuite augmenter à fort niveau de bruit, de même façon que la durée syllabique chez L10, L3 et L5.

Ces observations montrent que si certains descripteurs de la parole évoluent de façon linéaire en fonction du bruit, cela n'est pas le cas pour tous. Par conséquent, l'adaptation globale de la parole dans le bruit n'est pas un phénomène pouvant se modéliser par une simple transformation linéaire définie par une matrice A et un vecteur de translation B, tel que le propose Mokbel 1992^[245]. De nombreuses études ont de la même manière cherché à modéliser l'effet Lombard par une relation linéaire entre la variation de l'intensité vocale et la variation du niveau de bruit, et ont même proposé de quantifier cet effet par le coefficient de cette droite (Egan 1972^[80] ; Lane et al. 1971^[192] ; Korn 1954^[185] ; Kryter 1946^[188] ; Dejonckere 1979^[63] ; Van Heusden et al. 1979^[360]). Mais il faut comprendre que cette modélisation linéaire de l'adaptation de la parole dans le bruit a été principalement recherchée par le domaine de la reconnaissance automatique de la parole pour lequel il était nécessaire de caractériser la « déformation » de la parole dans le bruit afin d'améliorer la robustesse des algorithmes basés sur un modèle acoustique de parole conversationnelle. Cette approximation linéaire trouve donc sa justification par rapport au but poursuivi par ces études. Elle n'est en revanche pas appropriée à une démarche de compréhension des mécanismes impliqués dans l'adaptation de la parole en environnement bruyant, puisqu'elle risque en particulier par contre de passer sous silence certains aspects intéressants tels que d'éventuelles réorganisations de l'action en fonction des caractéristiques du bruit ambiant ou certaines particularités individuelles.

En effet, lors d'une étude préliminaire (Garnier et al. 2006^[107]), nous avons exploré la linéarité d'évolution de plusieurs descripteurs acoustiques de la parole en fonction du niveau de bruit, ainsi que du ressenti du locuteur. Nous avons observé que l'évolution des paramètres acoustiques en fonction du niveau de bruit pouvait être modélisée par une relation linéaire ou polynomiale d'ordre 2, à condition d'effectuer cette régression à partir des données de tous les locuteurs confondus. En revanche, l'examen des évolutions par individus, comme elles sont représentées Figure 129 et Figure 130, fait apparaître de nettes ruptures dans l'évolution, des plateaux à faible niveau ou des saturations à fort niveau. Etant donné la consistance de l'effectif des données par individus et la reproductibilité de ces non linéarités d'un type de bruit à un autre, il semble bien que la modélisation linéaire observée en moyenne sur plusieurs individus ne corresponde pas à seulement à un « lissage » de la variabilité par une augmentation du nombre de données, mais en réalité à un « moyennage » de profils individuels nettement différents, ne rendant finalement compte du comportement d'aucune catégorie d'individus.

Certains résultats d'études antérieures vont également dans ce sens : Hanley et al. 1949^[133] ont par exemple montré que le ralentissement du débit de parole évoluait plutôt de façon logarithmique en fonction du niveau de bruit ambiant. Ternström et al. 2006^[345] ont également observé que la balance spectrale de la voix augmente avec un niveau de bruit croissant puis sature à très fort niveau de bruit pour des [a] et diminue même pour des [i]. De la même façon, Sodersten et al. 2005^[322] ont rapporté une augmentation de l'aire du phonétogramme à faible niveau de bruit puis une diminution à fort niveau de bruit. Egan 1972^[80] a observé que l'intensité vocale évolue de façon aléatoire en dessous de 50dB de bruit ambiant puis évolue de façon linéaire jusqu'à 100dB de bruit ambiant environ où ce paramètre sature. De telles non-linéarités ont également été notées pour des stratégies linguistiques : Traunmüller et al. 2000^[356] ont ainsi observé que les hommes ont tendance à diminuer le nombre de pauses dans l'énoncé avec un éloignement croissant de l'interlocuteur, puis l'augmente à très forte distance. Enfin, ces non-linéarités dans la production semblent se retrouver au niveau perceptif puisque des études antérieures ont montré que l'intelligibilité augmente avec une intensité croissante de la voix puis diminue à très forte intensité vocale (Rostolland et al. 1973^[290]).

L'interprétation de ces non-linéarités reste une question ouverte. Pour certaines stratégies telles que l'augmentation de l'intensité vocale, de la fréquence fondamentale, du timbre de la voix ou de l'ouverture de la mâchoire, il existe sans aucun doute des limitations physiologiques et techniques à l'adaptation du locuteur, qui peuvent expliquer les paliers observés dans l'évolution de ces paramètres à très fort niveau de bruit. Les locuteurs de nos deux bases de données atteignent en effet des intensités vocales de 92 à 105dB dans le bruit, ce qui correspond effectivement à l'ordre de grandeur des intensités maximales pouvant être produites en parole.

Il existe des non-linéarités dans le comportement de la source glottique elle-même qui peuvent également expliquer ces changements d'évolution de certains paramètres en fonction du niveau de bruit croissant. Vintturi 2001^[368] a montré par exemple que l'intensité vocale n'augmente pas de façon linéaire en fonction de la vitesse de fermeture des cordes vocales et qu'il existe deux comportements distincts à faible intensité vocale puis à intensité de parole conversationnelle. Alku et al. 1999^[6] a constaté une autre linéarité dans cette relation à forte intensité vocale, où la vitesse de fermeture n'augmente plus conjointement à l'intensité vocale. De même, la fréquence fondamentale n'augmente pas du grave à l'aigu de la tessiture d'un individu de façon linéaire en fonction des paramètres laryngés, du fait de l'utilisation de plusieurs mécanismes laryngés (cf. Annexe Met1) et de la non-linéarité biomécanique du larynx.

A ces limitations et non-linéarités physiologiques contraignant l'adaptation du locuteur dans le bruit peuvent également s'ajouter des limitations d'ordre phonologique. L'augmentation de l'intensité vocale permet par exemple d'améliorer l'intelligibilité mais s'accompagne d'une augmentation de la F0, de l'ouverture de la mâchoire, du premier formant vocalique. Il arrive donc un moment où le gain d'intelligibilité apporté par l'augmentation de l'intensité vocale est compensé par une détérioration de l'intelligibilité au niveau de la reconnaissance des voyelles par exemple et ce seuil semble être inférieur à l'intensité maximale productible par un locuteur (Rostolland et al. 1973^[290]).

Il est également tout à fait envisageable que l'adaptation du locuteur soit en partie limitée par des contraintes sociologiques. Dans certains contextes de communication, un individu préférera parfois ne pas fournir davantage d'effort d'adaptation au point de produire une voix trop aiguë ou trop criarde par rapport à l'effet qu'il veut produire sur son interlocuteur.

Enfin, nous avons jusque là principalement parlé de limitations à l'adaptation du locuteur dans le bruit. Or les non-linéarités observées correspondent également à des paliers à faible niveau de bruit dans l'évolution de certaines stratégies. Nous avons ainsi vu dans les chapitres 7, 8 et 9 que l'adaptation de la parole dans le bruit ne consiste pas seulement en un renforcement d'indices ou de caractéristiques existant déjà dans le silence, mais dans certains cas en l'adoption de nouvelles stratégies de communication dans le bruit n'existant pas dans le silence. De même, nous avons observé lors de la constitution de la 1^{ère} base de données que les interlocuteurs ne se servaient pas particulièrement de la modalité visuelle pour communiquer à faible niveau de bruit tandis qu'ils se regardaient attentivement à partir de 78 ou 86dB de bruit, ce qui suggère que des stratégies audiovisuelles soient mises en oeuvre à fort niveau de bruit. Ces différentes observations argumentent en faveur du fait que les non-linéarités de l'adaptation de la parole dans le bruit ne traduisent pas seulement une limitation de l'adaptation mais plus globalement une réorganisation de l'action en fonction de la perturbation imposée par l'environnement bruyant.

2.2.2. Réorganisation de l'adaptation en fonction du niveau de bruit

Les réflexions précédentes nous incitent donc à explorer la façon dont chaque locuteur réorganise son adaptation en fonction du type et du niveau de bruit. Pour cela, nous avons synthétisé dans le tableau la façon dont chaque stratégie évolue, pour chaque locuteur et locutrice de la BD1, à faible niveau de bruit et à fort niveau de bruit.

Locuteurs		Niveau de bruit	Stratégies communes						Stratégies individuelles						
			S4	S1		S3		S10		S2			S9		
				Tous niveaux	Fort niveau seulement	Tous niveaux	Fort niveau seulement	Tous niveaux	Fort niveau seulement	Tous niveaux	Fort niveau seulement	Aucun niveau	Tous niveaux	Fort niveau seulement	Aucun niveau
Groupe 1	L10	faible	aug		const		const		const					dim	
		fort	sat		aug		aug		aug					aug	
	L1	faible	aug	aug			const		const	aug					const
		fort	sat	sat			aug		aug	dim					const
	L9	faible	aug	aug		aug			dim	aug					dim
		fort	sat	sat		aug			aug	dim					aug
	L2	faible	aug	aug			const		const	aug					const
		fort	aug	aug			aug		aug	aug					aug
	L3	faible	aug	aug			const		dim	aug					dim
		fort	sat	aug			aug		aug	sat					aug
L5	faible	aug	aug			const		const		const				dim	
	fort	sat	aug			aug		aug		aug				aug	
Groupe 2	L8	faible	aug	aug			const		const		dim			aug	
		fort	sat	sat			aug		aug		aug			aug	
	L6	faible	aug	aug			const		aug		dim			aug	
		fort	sat	sat			aug		aug		aug			sat	
	L4	faible	aug	aug			const		aug		aug			aug	
		fort	aug	aug			aug		dim		sat			dim	
	L7	faible	aug	aug			const		aug			const			const
		fort	sat	aug			sat		aug			dim			aug

Tableau 32. Synthèse des différences individuelles de réorganisation de l'adaptation de la parole dans le bruit pour les différents locuteurs (L3, L4, L5, L7, L8) et locutrices (L1, L2, L6, L9, L10) de la BD1, en considérant les changements d'évolution entre un faible et un fort niveau de bruit pour plusieurs stratégies de communication : la préservation du rapport signal sur bruit par l'augmentation de l'intensité vocale (S1), l'élargissement de l'ambitus de F0 (S2), l'augmentation de la F0 cible (S3), l'augmentation de l'énergie dans la zone sensible de l'oreille (S4), le ralentissement du débit de parole (S9) et le renforcement de la dynamique d'intensité de chaque syllabe (S10).

Le label « aug » rend compte d'une augmentation de la stratégie considérée pour un faible ou fort niveau de bruit, « dim » d'une diminution, « const » d'une absence d'évolution à faible niveau de bruit et « sat » d'une saturation à fort niveau de bruit.

Pour chaque stratégie, nous avons séparé dans des colonnes différentes les locuteurs pour lesquels la stratégie est utilisée à tous niveaux de bruit avec ou sans saturation à fort niveau de bruit, ceux qui utilisent cette stratégie uniquement à fort niveau de bruit et ceux pour lesquels aucune évolution notable de cette stratégie n'est observée, que ce soit à faible ou fort niveau de bruit.

Deux principaux groupes de locuteurs se dégagent par rapport à la réorganisation de ces stratégies entre un faible et un fort niveau de bruit : un premier groupe se forme autour des locuteurs L2 et L3 dont ils diffèrent uniquement par la réorganisation d'une de leur stratégie (entourée par un cercle noir) d'un faible à un fort niveau de bruit. Un deuxième groupe rassemble les autres locuteurs autour de L6, dont chacun diffère également par la réorganisation d'une stratégie d'un faible à un fort niveau de bruit.

On observe que certaines stratégies, comme l'augmentation de l'intensité (S1), le timbrage de la voix (S4) et l'augmentation de la F0 (S3) sont utilisées de façon semblable par tous les locuteurs. S1 et S4 sont utilisées à faible comme à fort niveau de bruit, avec une éventuelle saturation à fort niveau de bruit chez certains locuteurs, tandis que S3 au contraire est plutôt utilisée à fort niveau de bruit

seulement. Parmi ces grandes tendances partagées par tous les locuteurs, seule L10 n'augmente son intensité qu'à fort niveau de bruit L9 augmente sa F0 déjà à faible niveau de bruit.

Au contraire, on observe davantage de différences interindividuelles dans l'évolution de l'ambitus de F0 (S2), dans le ralentissement du débit de parole (S9) et dans le renforcement de la dynamique d'intensité par syllabe (S10). On peut dégager 2 principaux profils individuels de réorganisation de l'adaptation dans le bruit, autours desquels peuvent être rattachés plusieurs locuteurs et locutrices :

- d'un côté, le locuteur L3 et la locutrice L2 présentent une augmentation de l'ambitus de F0 (S2) à faible comme à fort niveau de bruit, et au contraire un allongement syllabique (S9) et un renforcement de la dynamique syllabique (S10) uniquement à fort niveau de bruit. La locutrice L10 peut être rattachée à ce premier groupe de locuteurs, de même que L1 qui diffère uniquement par sa non-utilisation de la stratégie S9, et L5 qui diffère en augmentant la stratégie S2 uniquement à fort niveau de bruit. On remarque que ce groupe de locuteurs aux comportements semblables est mixte.
- de l'autre côté, la locutrice L6 n'augmente au contraire son ambitus de F0 (S2) qu'à fort niveau de bruit, mais ralentit son débit de parole (S9) et renforce la dynamique syllabique (S10) à faible comme à fort niveau de bruit. L8 diffère de L6 uniquement par l'évolution de la stratégie S10, L4 par la stratégie S2 et L7 par la stratégie S9 ainsi que par sa non-utilisation de la stratégie S2.

Il est intéressant de noter que tous les locuteurs, sans exception, montrent une réorganisation de l'adaptation dans le bruit, même si celle-ci est différente pour chacun. Il semble donc que l'adaptation de la parole dans le bruit ne soit pas un phénomène graduel, mais davantage catégoriel en fonction du niveau de perturbation imposée par le bruit ambiant. De telles réorganisations peuvent être observées au niveau psychomoteur, comme l'illustre classiquement l'exemple de la transition marche-course d'un individu se déplaçant sur un tapis roulant à vitesse croissante (Diedrich et al. 1995^[69]). Dans le domaine de la parole, de telles réorganisations ont également déjà été mises en évidence, par exemple au niveau articulatoire entre un débit de parole lent et rapide (Vaxelaire et al. 1998^[362]). Ces bifurcations comportementales d'un patron de coordination vers un autre sous l'effet d'une variation d'un paramètre de contrôle sont expliquées par l'auto-organisation du système en perte de stabilité vers un autre état stable.

Dans notre cas, il ne semble pas y avoir de bifurcation brutale d'un patron vers un autre mais une réorganisation plus progressive en fonction du niveau de bruit. Par ailleurs, cette réorganisation n'est pas identique pour tous les individus comme cela peut être le cas dans la transition marche-course.

Une première observation peut permettre de proposer une explication à cette réorganisation de l'adaptation : on remarque en effet que tous les locuteurs, à l'exception de L2,aturent à fort niveau de bruit sur une ou plusieurs de leurs stratégies utilisées à faible niveau de bruit. Cela pourrait donc justifier le fait que les locuteurs recourent alors à de nouvelles stratégies supplémentaires pour continuer à augmenter leur intelligibilité à fort niveau de bruit. L2 ne présente certes pas de saturation sur les 6 stratégies observables à partir des données de la BD1, mais il est fort probable qu'il existe des saturations sur d'autres stratégies « linguistiques ». Comme nous l'avons déjà envisagé au paragraphe précédent, ces saturations pourraient correspondre à des limitations physiologiques du locuteur. Ces saturations pourraient également être des indicateurs d'un mécanisme de protection de la voix chez certains locuteurs, évitant de dépasser un certain seuil d'effort sur une stratégie et privilégiant alors une diversification des stratégies. Ces saturations pourraient également correspondre à un seuil de compromis entre le gain d'intelligibilité apporté par l'augmentation d'une stratégie de communication et la détérioration de l'intelligibilité qu'elle engendre sur d'autres plans. Par exemple, la saturation de l'intensité vocale à fort niveau de bruit ne correspond pas nécessairement au fait qu'un locuteur a atteint sa limite vocale mais peut être seulement qu'au-delà de cette intensité, son effort devient contre-productif pour l'intelligibilité.

Enfin, il reste une autre explication possible que nous n'avons pas encore abordée : la perception d'intensité et de gêne liées à un environnement bruyant n'est pas linéaire en fonction du niveau de bruit « physique » en dB, tant du point de vue physiologique – il existe un réflexe, dit stapédien, permettant d'atténuer un son à très forte intensité au niveau de l'oreille moyenne (Margolis et al. 1975^[233]) – que du point de vue cognitif (Dubois et al. 2001^[78]). Sur la Figure 129 et la Figure 130, on observe effectivement que la perception de sa propre voix ou la perception de gêne n'augmente pas de façon linéaire en fonction du niveau de bruit, et on retrouve chez certains locuteurs un pallier à faible niveau de bruit, avec une évolution du ressenti uniquement à fort niveau de bruit. Ces observations pourraient donc tout à fait expliquer la réorganisation de l'adaptation observée dans le bruit, à partir d'une analyse de la situation par le locuteur d'ordre davantage catégorielle que graduelle.

3. CONCLUSION

Tout au long de cette thèse, nous avons adopté une démarche progressive pour caractériser les mécanismes d'adaptation de la parole dans le bruit (cf. Figure 131).

- (1) Au chapitre 5, nous avons commencé par recenser les différentes modifications acoustiques, articulatoires et glottiques observables entre une condition de silence et une condition de bruit.
- (2) Dans les chapitres 6 à 9, nous avons remonté d'un cran dans le système d'adaptation en cherchant à relier ces différentes observations dans le domaine physique à des stratégies de communication pouvant être adoptées dans le bruit pour préserver l'intelligibilité, à une régulation réflexe de la voix par l'audition, ou encore à des stratégies de protection du larynx ou de compensation de la détérioration de l'intelligibilité pouvant être induite par certaines modifications articulatoires liées à l'augmentation de l'intensité vocale. La mise en œuvre d'une stratégie de communication peut de ce point de vue induire d'autres modifications collatérales que celle directement visée par la stratégie. Quand au contraire la modification d'un paramètre dans le domaine physique peut correspondre à plusieurs stratégies de communication différentes. Il reste donc maintenant à évaluer l'efficacité de ces différentes stratégies de communication, non seulement au niveau de leur influence sur l'intelligibilité de la parole dans le bruit, mais également au niveau de la différence de charge qu'elles sont susceptibles de faire porter au larynx. Ce chapitre nous a permis de synthétiser et de discuter ces différents résultats et de conclure que l'adaptation de la parole dans le bruit est en partie motivée par la recherche d'intelligibilité et qu'elle peut donc être en partie gérée par le locuteur, et pas uniquement subie. Cela ouvre donc une perspective intéressante de prévention et de rééducation possible de cette adaptation.
- (3) Enfin, nous avons examiné dans ce chapitre comment ces stratégies sont adoptées de façon combinée et complémentaire différemment selon les interlocuteurs et selon le contexte de communication (type de bruit, niveau de bruit, etc.). Il reste maintenant à explorer plus en détail le fonctionnement de ce système de sélection et d'organisation des stratégies en fonction des individus et des contextes, afin de déterminer l'influence de l'analyse de la situation de communication par le locuteur (perception de gêne, évaluation de l'attention qu'on lui porte, de sa propre intelligibilité, évaluation de la distance de l'interlocuteur, de l'intensité du bruit ambiant) sur sa prise de décision (non consciente) et sur sa mise en œuvre de stratégies d'adaptation. Plusieurs études laissent en effet penser que la personnalité du locuteur joue un rôle dans le phénomène de forçage vocal (Roy et al. 2000 ^[295] ; Giovanni 2007 ^[114]), de même que son manque d'attention à la gêne, à la fatigue ou à la douleur qu'il ressent (Aronsson et al. 2005 ^[12]) ou encore sa sous-estimation de sa propre intelligibilité (observée chez les femmes par Sodersten et al. 2005 ^[322]).

Figure 131. Modélisation de l'adaptation de la parole dans le bruit (page suivante->).

Au niveau physique (1), on peut observer des modifications de la parole sur les plans acoustique et articulatoire. Ces modifications ont un double statut puisqu'elles sont à la fois perceptibles par l'interlocuteur et d'ailleurs en partie produites à son attention. Mais elles constituent également des indices pour remonter à la façon dont elles sont été produites (sur le plan respiratoire, phonatoire ou articulatoire) ainsi qu'à l'intention poursuivie par le locuteur en terme de stratégie de communication, de protection ou de compensation (2). Enfin, la mise en œuvre et l'organisation de ces différentes stratégies est décidée par chaque locuteur (3) de façon différente en s'appuyant sur sa perception et son analyse du contexte ((3) faible ou fort niveau de bruit, type de bruit, distance de l'interlocuteur), sur son estimation de sa propre intelligibilité, sur son ressenti de gêne ou de fatigue, etc.

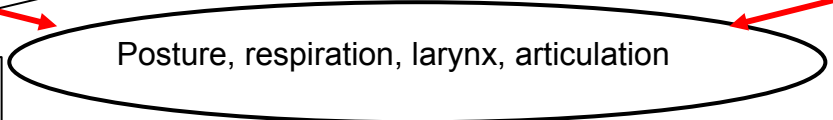
Nous avons indiqué en rouge les endroits de ce système d'adaptation où un dysfonctionnement peut faire passer le locuteur d'une simple adaptation à un forçage vocal : ainsi, les facteurs du forçage vocal classiquement étudiés dans la littérature peuvent influencer la constitution physiologique du locuteur et dégrader les conditions de production. Tout comme l'acquisition d'une technique vocale peut influencer la réalisation motrice des stratégies dans le domaine physique. Mais le forçage vocal peut également résider dans le choix et l'organisation combinée de certaines stratégies moins efficaces que d'autres. Il peut également consister en une mauvaise prise de décision dans la mise en œuvre de l'adaptation par surestimation des exigences du contexte, ou par sous-estimation de la gêne ou de sa propre intelligibilité.

(1) Modifications mesurables par le chercheur, perceptibles par l'interlocuteur



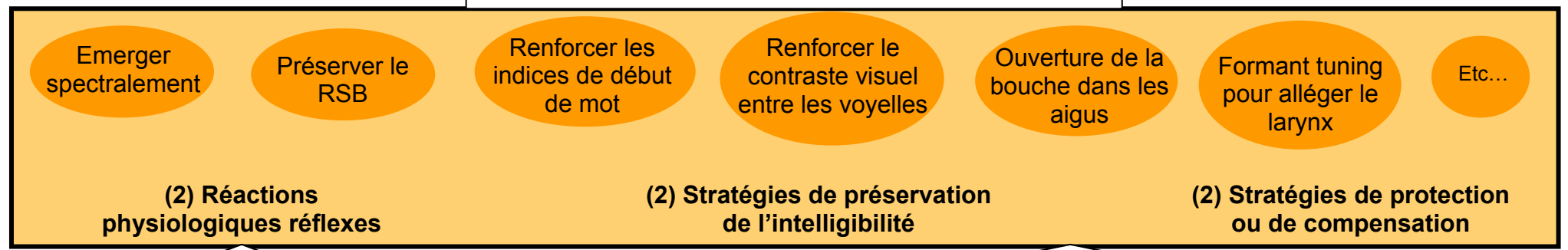
Humidité, poussière, posture inconfortable, maladie ORL

Constitution anatomique



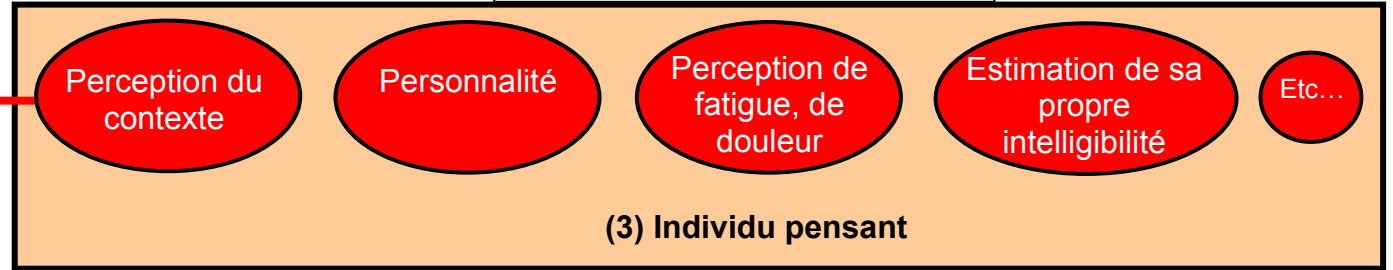
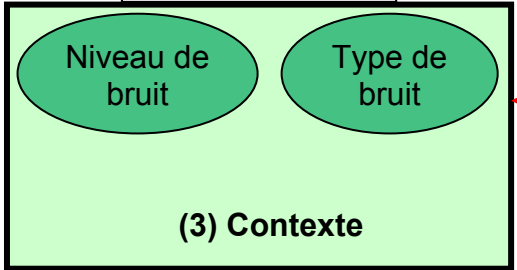
Technique vocale

Contrôle moteur



Régulation automatique

Prise de décision, choix d'une adaptation



Conclusion générale de l'étude

Apports de la thèse

Une nouvelle façon de concevoir le problème du forçage vocal

De façon à mieux comprendre pourquoi tous les locuteurs ne sont pas équivalents devant les troubles de la voix et devant le problème général du forçage vocal, nous nous sommes attachés dans cette étude à définir le plus précisément possible notre objet d'étude et à le situer dans l'espace conceptuel et méthodologique des études antérieures. Notre intérêt et notre champ disciplinaire d'origine (acoustique, psychologie) nous ont conduits à envisager la problématique avec un nouveau point de vue, en cherchant à expliquer les différences interindividuelles face aux troubles de la voix par des différences comportementales de gestion de l'effort vocal, plutôt que par des différences d'origine physiologiques ou par des différences d'exposition à certains facteurs. Nous avons par conséquent envisagé le problème du forçage vocal à un niveau global, à la fois corporel et cognitif, où le locuteur joue un rôle actif dans ce phénomène, et non pas à un niveau uniquement laryngé. Cela nous a amené à considérer l'effort vocal non pas comme quelque chose de subi par le locuteur, mais régulé par le locuteur en fonction des exigences d'intelligibilité de la situation de communication. Par conséquent, nous n'avons plus seulement envisagé le forçage vocal du point de vue des symptômes qu'il peut générer, mais du point de vue de l'efficacité de l'adaptation de l'effort d'un locuteur à une situation de communication. Cette efficacité est liée à la fois à l'intelligibilité du locuteur, mais également à la façon dont il réalise cette stratégie de communication en répartissant plus ou moins l'effort de production sur les différents organes de la parole. C'est pourquoi nous avons cherché durant cette thèse à caractériser les différences interindividuelles d'adaptation de la parole à une situation de communication perturbée, en l'occurrence un environnement bruyant, afin de mieux comprendre en quoi certaines personnes s'adaptent différemment que d'autres et peuvent éventuellement développer des troubles de la voix.

Proposition de méthodologies expérimentales appropriées à l'étude de l'adaptation de la parole en environnement bruyant

Cette restriction de la problématique du forçage vocal à la caractérisation, dans un premier temps, des différences interindividuelles d'adaptation de la parole dans le bruit, nous a amenés à prendre en compte dans nos protocoles expérimentaux plusieurs facteurs ayant une influence sur les mécanismes physiologiques et cognitifs de l'adaptation de la parole dans le bruit.

Nous avons d'abord comparé plusieurs moyens de reconstituer en laboratoire un environnement bruyant. Si certains aspects de la parole ne sont pas sensibles à la méthode d'immersion dans le bruit (telles que la fréquence fondamentale ou la durée syllabique), l'usage d'un casque influence en revanche significativement l'adaptation de la parole pour la plupart des autres paramètres acoustiques (intensité vocale, répartition spectrale de l'énergie, etc.). Cette observation nous a conduits dans nos protocoles expérimentaux à faire en sorte que le locuteur conserve des retours auditifs naturels de l'environnement et de sa propre voix. Pour cela, nous avons implémenté une technique numérique de débruitage permettant d'immerger le locuteur dans le bruit par le biais de haut-parleurs, tout en pouvant réaliser des mesures acoustiques fiables.

Nous avons ensuite montré que l'interaction avec un interlocuteur conditionne de façon significative l'adaptation acoustique et articulatoire des locuteurs dans le bruit, principalement en l'amplifiant. Nous avons conçu plusieurs jeux interactifs permettant d'enregistrer les locuteurs en situation d'interaction semi-spontanée avec un interlocuteur, tout en contrôlant le matériel linguistique.

Constitution de 2 grandes bases de données acoustiques, électroglottographiques et articulatoires pour l'étude de la parole Lombard

Nous avons au cours de cette thèse constitué deux grandes bases de données d'étude de la parole Lombard, que nous sommes prêts à mettre à disposition d'autres chercheurs.

Le premier intérêt de ces bases de données est d'être composées de parole semi-spontanée, avec 16 mots-cibles pour la 1^{ère} base de données et 17 logatomes-cibles pour la deuxième. Les données de ces corpus sont donc beaucoup plus proches de la parole naturelle et peuvent être plus intéressantes pour des études sur la reconnaissance automatique de la parole Lombard qui

chercheraient à améliorer la robustesse de leur algorithme en situation réelle, par exemple pour des applications de billetteries automatiques dans des gares, ou pour améliorer la reconnaissance de commandes vocales dans un habitacle automobile.

Ces données sont également très riches de par le nombre de locuteurs examinés (10 dans le 1^{ère} et 3 dans la 2^{ème}), dans 2 types de bruit (bruit blanc et bruit cocktail) et dans 4 niveaux de bruit pour la 1^{ère} base de données (62, 70, 78 et 86dB). Elles offrent également l'avantage de pouvoir mettre en relation des observations acoustiques, articulatoires et électroglottographiques simultanées.

L'exploration de ces deux bases de données n'est pas achevée. En particulier, la deuxième base de données permettra d'observer plus en détail certains aspects phonétiques d'articulation des voyelles, du fait qu'elle présente 6 voyelles différentes dans un même contexte [I_]. Elle permettra également d'examiner plus en détail l'adaptation de la parole dans le bruit au niveau prosodique, grâce à la structure contrôlée des énoncés. Nous projetons enfin de mener très prochainement des tests perceptifs à partir de ses extraits audio et vidéo.

Apports théoriques sur l'effet Lombard et sur l'adaptation de la parole dans le bruit

Dans la 3^{ème} partie de cette thèse, nous avons vérifié les observations principales de la littérature en ce qui concerne les aspects acoustiques et formantiques de la parole Lombard, que nous avons complétées par de nouvelles observations électroglottographiques et articulatoires en ce qui concerne l'ouverture, l'étirement, la protrusion et le pincement des lèvres.

Nous ne nous sommes pas contentés de décrire l'adaptation de la parole dans le bruit de façon paramétrique et globale, mais nous avons ensuite exploré différentes hypothèses de stratégies de communication pouvant être adoptées par les locuteurs dans le bruit de façon à préserver leur intelligibilité pour l'interlocuteur. Nous avons examiné en quoi l'évolution des paramètres acoustiques et articulatoires dans le bruit peut contribuer à ces stratégies.

Parmi ces différentes hypothèses, nous avons pu mettre en évidence plusieurs stratégies de communication adoptées par certains locuteurs, parfois par tous, et avec des magnitudes différentes. En premier lieu, nous avons observé comment les locuteurs peuvent émerger du bruit ambiant en préservant un rapport signal sur bruit acceptable, en élargissant la distribution de leur fréquence fondamentale, en adaptant cette distribution au minimum spectral du bruit ambiant, en compressant leur dynamique vers les fortes intensités vocales, ou en renforçant l'énergie de leur voix dans la zone de sensibilité maximale de l'oreille. Nous avons ensuite mis en évidence chez certains locuteurs l'existence de stratégies visant à renforcer le contraste acoustique et visuel entre les différentes voyelles, à renforcer le pincement des lèvres sur les consonnes bilabiales, et à allonger la durée des cibles vocaliques et des transitions formantiques pouvant servir à faciliter la récupération des indices de reconnaissance des unités phonétiques par l'interlocuteur. Nous avons également montré l'existence, chez certains locuteurs, de stratégies de renforcement d'indices de structuration du discours, tels que l'allongement des syllabes initiales et finales d'un énoncé, le renforcement du pincement labial sur la consonne initiale d'un mot de contenu, ou encore le renforcement de la chute mélodique en fin d'énoncé. Pour finir, nous avons vu qu'il existe des stratégies de renforcement dans le bruit des mots portant plus d'information utile pour l'interlocuteur, mais marqués par des indices acoustiques ou prosodiques différents selon les individus.

Au-delà de la mise en évidence de ces différentes stratégies de communication en environnement bruyant, cette étude a apporté de nouveaux arguments au débat sur l'interprétation réflexe ou communicationnelle de l'effet Lombard : Tout d'abord, la comparaison de la parole Lombard produite avec et sans interaction avec un interlocuteur plaide en faveur d'une recherche d'intelligibilité même en l'absence d'interlocuteur, qui s'ajouterait à la régulation réflexe de l'intensité vocale face à une atténuation du retour auditif. L'adaptation plus importante observée lorsqu'un interlocuteur est présent et donne des retours de sa compréhension argumente bien en faveur de l'intervention de mécanismes cognitifs haut-niveau dans l'adaptation de la parole en environnement bruyant. Au cours des chapitres 7, 8 et 9, nous avons également montré que l'adaptation de la parole n'est pas globale sur tout l'énoncé, mais plus importante sur certains constituants : Ainsi, le renforcement de la protrusion et de l'arrondissement des voyelles [u], ou le renforcement du pincement labial des consonnes [m] ne peuvent pas être expliqués uniquement comme des conséquences de l'augmentation de l'intensité vocale. De même, nous ne voyons pas

d'autre explication au renforcement sélectif de certains indices prosodiques de structuration du discours, qu'une recherche d'intelligibilité de la part du locuteur. Le renforcement de certains mots de l'énoncé portant davantage d'information utile pour l'interlocuteur que les autres va également dans ce sens.

Afin de synthétiser ces différents résultats, nous avons mené à la fin de cette étude une première exploration de l'évolution de l'adaptation d'un locuteur en fonction du type de bruit et du niveau de bruit. Nous avons ouvert la perspective d'une adaptation davantage catégorielle que graduelle, consistant en une sélection et une combinaison des différentes stratégies de communication en fonction du contexte. Nous avons ainsi observé que certaines adaptations globales étaient privilégiées à faible niveau de bruit, et d'autres au contraire à fort niveau de bruit. Ces premiers résultats nécessiteront d'être réexaminés plus en détail lors de travaux futurs.

Éléments de réponse au problème du forçage vocal

A travers l'étude des mécanismes physiologiques ou communicationnels sous-tendant l'adaptation de la parole dans le bruit, nous avons pu apporter de nombreux arguments en faveur du fait que le locuteur a la capacité, dans une certaine mesure, de gérer son adaptation dans le bruit et ainsi de réguler son effort à la situation de communication et non de la subir. Ce résultat ouvre de nouvelles perspectives quant à la prévention du forçage vocal dans de tels environnements, en permettant d'envisager des solutions au problème du forçage vocal au niveau du comportement du locuteur lui-même, et non plus uniquement de façon extérieure (hydratation, microphone, acoustique de la salle, poussière, ...).

Nous avons ensuite mis en évidence des différences interindividuelles d'adaptation et donc de gestion de l'effort vocal face à une même situation de communication. Ces différences peuvent être de deux principaux ordres :

- tout d'abord, nous avons observé que certains locuteurs s'adaptent davantage que d'autres de façon globale, et aussi que d'autres s'adaptent peu du fait qu'ils produisent toujours un effort de communication important, aussi bien dans le silence que dans le bruit. De ce point de vue, les individus sujets au forçage vocal pourraient ne pas être seulement ceux dont l'adaptation du silence au bruit est la plus importante, mais également ceux dont l'effort est toujours maximal et non régulé aux exigences de la situation de communication.
- nous avons également observé que les locuteurs n'adoptent pas une unique stratégie sur laquelle ils s'adaptent plus ou moins que d'autres, mais sélectionnent chacun des combinaisons différentes de stratégies de communication, avec des dosages différents et des réorganisations différentes de ces combinaisons en fonction du niveau de bruit. Ces résultats sont intéressants du point de vue du forçage vocal car ces différentes stratégies ne sont a priori pas équivalentes au niveau de l'effort qu'elles nécessitent pour produire un même niveau d'intelligibilité, ainsi qu'au niveau de la charge qu'elles sont susceptibles de faire porter sur le larynx. Ces différences interindividuelles dans l'utilisation et l'organisation des stratégies de communication pourrait donc être un facteur important permettant d'expliquer la plus grande sensibilité de certains individus aux troubles de la voix. Il reste maintenant à encore remonter d'un cran dans le fonctionnement cognitif des locuteurs afin de comprendre les causes de ces prises de décision différentes face à une même situation de communication. Au vu de plus observations effectuées au cours de ce travail, il nous semble important de mieux cerner le lien entre l'analyse de la situation par le locuteur et sa prise de décision d'une stratégie appropriée, en particulier au niveau de sa perception de la perturbation imposée par le bruit, au niveau de sa perception de sa propre voix, au niveau de la façon dont il pense être intelligible pour l'interlocuteur, mais également écouté par ce dernier, au niveau de l'image qu'il veut lui donner, au niveau de sa connaissance de ses propres limites vocales ou encore au niveau de sa perception de gêne, de fatigue et de douleur.
- enfin, nous avons également observé que si certaines stratégies de communication semblent être associées en soi à un comportement laryngé, il existe quelques cas où plusieurs individus produisent une même stratégie de communication avec des comportements laryngés différents. Aussi, indépendamment de l'analyse de la situation ou du choix de mettre en place certains stratégies de communication de façon privilégiée, il semble bien exister une gestion individuelle au niveau de la réalisation motrice (phonatoire, posturale, articulatoire, respiratoire) de ces stratégies. Cela va bien entendu dans le sens de l'intuition qu'un apprentissage de la technique vocale peut être bénéfique aux orateurs ou aux personnes amenées à se servir fréquemment de leur voix.

Ces différents résultats ouvrent donc de nouvelles perspectives quant à la prévention du forçage vocal. On pourrait dès lors enseigner aux personnes utilisant leur voix de façon importante, des stratégies de communication efficaces et appropriées par rapport à la situation à laquelle elles sont confrontées dans leur vie professionnelle (bruit, forte réverbération, prise de parole à distance, en plein air, etc.), en association avec des techniques vocales pour améliorer la gestion de leur production. Pour certaines personnes, il serait également utile de les faire travailler sur leur perception de la situation, de leur propre voix, sur leur sentiment d'être entendu et écouté, de façon à les amener à trouver un niveau d'effort approprié par rapport aux exigences de la situation de communication.

Limites

Le principal regret vis à vis de cette étude concerne le choix des locuteurs de la première base de données, que nous aurions pu choisir avec des profils plus différents, afin de mettre en évidence des contrastes plus marqués.

Par ailleurs, la limite principale de cette étude consiste à n'être restée que du point de vue de la production de la parole en environnement bruyant. Nous ne pouvions bien entendu pas tout traiter. Il n'en reste pas moins indispensable maintenant de valider ou d'infirmer l'existence des différentes stratégies d'adaptation de la parole dans le bruit, en examinant si celles-ci améliorent effectivement ou non l'intelligibilité de la parole pour l'interlocuteur.

La problématique du forçage vocal était très ambitieuse et bien sûr impossible à aborder dans son ensemble en une thèse. Une première approche aurait pu consister à n'en considérer qu'un aspect bien précis à le creuser en profondeur. Nous avons au contraire préféré une approche plus globale du phénomène, nous ayant par conséquent empêchés de rentrer davantage dans les détails de chaque point abordé. Ce choix peut nous être reproché et être interprété en terme de dispersion. Cependant, comme nous l'avons vu tout au long de ce manuscrit, le forçage vocal n'est pas un problème simple, mais pluridisciplinaire et faisant intervenir des aspects complémentaires de la parole : acoustiques, mécaniques, phonétiques, psychologiques et médicaux. C'est pourquoi il nous a paru indispensable d'adopter une approche globale et transversale à différents champs disciplinaires pour pouvoir mieux comprendre ce phénomène.

Perspectives

Approfondissement de l'étude des stratégies d'adaptation et de compensation

Il serait par la suite nécessaire d'approfondir l'étude des différentes stratégies d'adaptation, en explorant plus en détail comment l'évolution des résonances du conduit vocal et du mouvement des lèvres peut participer à un renforcement de l'énergie spectrale au-dessus de 2kHz. Il serait également intéressant d'examiner davantage de types de bruit, afin de déterminer comment le contenu spectral du bruit ambiant peut influencer l'adaptation des locuteurs dans le bruit, en particulier au niveau du timbre de leur voix.

Comme nous l'avons déjà évoqué, il reste encore de nombreuses hypothèses à explorer en ce qui concerne l'existence de certaines stratégies de renforcement d'indices prosodiques dans le bruit. Nous devons également valider d'un point de vue perceptif les bénéfices de ces différentes stratégies pour l'intelligibilité. Par ailleurs, certaines adaptations posturales observées dans le bruit, telles que l'avancée du buste et du menton, pourraient également être prises en compte en tant qu'indicateurs d'un désir de persuasion ou d'une recherche de réduction de la distance par rapport à l'interlocuteur. Il serait donc intéressant d'étudier également les stratégies de communication des locuteurs non plus seulement à un niveau acoustique ou phonétique, mais également au niveau psychologique, afin d'essayer de relier certains comportements acoustiques, articulatoires ou posturaux à la façon dont le locuteur évalue la situation de communication, sa propre intelligibilité, ou encore l'attention que lui porte son interlocuteur.

Nous avons soulevé la possibilité que certaines modifications de la parole dans le bruit ne soient pas seulement liées à des stratégies de préservation de l'intelligibilité, mais également des compensations de ces stratégies soit pour la santé du larynx, soit pour l'intelligibilité. Aussi, il serait très intéressant d'examiner les mouvements de la langue et non plus seulement ceux des lèvres, afin de voir s'il existe des mouvements de recul lingual compensatoires à l'augmentation globale de l'ouverture de la bouche, en particulier pour les voyelles arrondies telles que [u]. De même, il serait très intéressant d'examiner si l'augmentation de l'ouverture de la mâchoire et de la

protrusion peuvent servir à faciliter la production de sons de forte intensité et de fréquence fondamentale élevée, comme cela est le cas dans le chant lyrique.

La multimodalité de l'adaptation de la parole dans le bruit reste également un point à examiner plus en détail, en particulier pour déterminer si les indices visuels renforcés dans le bruit sont uniquement des conséquences de stratégies acoustiques, ou si le locuteur cherche effectivement à les renforcer pour eux-mêmes. Dans ce but, nous espérons pouvoir prochainement examiner si le renforcement de ces indices visuels est davantage marqué lorsque les interlocuteurs disposent de la modalité visuelle pour communiquer, par rapport à la modalité audio seule.

Enfin, nous avons examiné dans cette étude l'adaptation de la parole dans le bruit pour la langue française uniquement. Il serait intéressant d'observer si les stratégies linguistiques que nous avons pu mettre en évidence pour le français se retrouve dans d'autres langues et si au contraire il existe des stratégies linguo-spécifiques d'adaptation à un environnement bruyant.

Estimation de l'efficacité des différentes stratégies de communication dans le bruit

La caractérisation des différences interindividuelles d'adaptation de la parole à un environnement bruyant consistait en la première étape dans la démarche générale de caractérisation du forçage vocal que nous avons exposée en introduction. La deuxième étape est maintenant d'évaluer l'efficacité des différentes stratégies observées, du point de l'effort de production qu'elles exigent pour une même intelligibilité perçue par l'interlocuteur, et du point de vue de la répartition des efforts sur les différents organes de la parole.

Pour cela, on peut envisager de mettre en relation ces différentes stratégies de communication avec des mesures aérodynamiques ou électromyographiques, ou de quantifier l'effort dépensé à partir des pics de vitesse des mouvements articulaires ou posturaux en général.

Des tests perceptifs, à la fois d'intelligibilité et d'évaluation de la qualité vocale, pourraient également permettre d'estimer l'efficacité de communication de ces différentes stratégies.

Au niveau laryngé, l'observation vidéostroboscopique de la vibration des cordes vocales au cours de la réalisation de ces différentes stratégies pourrait également apporter des informations complémentaires aux données électroglottographiques afin de déterminer en particulier si l'abaissement du quotient ouvert est seulement lié à un meilleur contact des cordes vocales ou à une tension laryngée pouvant provoquer des lésions à moyen terme. Cela permettrait également d'avoir plus d'information sur les mécanismes laryngés utilisés dans le silence, à faible ou à fort niveau de bruit.

Elargissement de la problématique à d'autres situations d'adaptation de l'effort vocal

Nous avons restreint dans un premier temps l'étude de la gestion de l'effort vocal à la communication en environnement bruyant. Cependant, le forçage vocal peut advenir dans d'autres conditions telles qu'un usage prolongé de la voix, ou une prise de parole dans une salle de mauvaise acoustique. Il serait donc utile, pour mieux comprendre le forçage vocal, d'élargir par la suite cette étude à d'autres conditions de communication, et d'examiner si certaines stratégies de communication sont retrouvées. Dans le cas d'un usage prolongé de la voix, il serait particulièrement intéressant d'examiner comment le locuteur s'adapte non pas à la situation, qui n'a pas nécessairement de particularité, mais à la fatigue qui s'installe progressivement. Autrement dit, existe-t-il des stratégies compensatoires à cette fatigue et en quoi peuvent elles accélérer ou ralentir l'apparition de symptômes pour la voix ?

Améliorer la segmentation automatique de la parole Lombard

Pour finir, les résultats de cette étude peuvent tout à fait être intégrés à un modèle de reconnaissance automatique de la parole, afin d'en améliorer la robustesse dans le bruit. En retour, cette amélioration pourrait aider de futures recherches acoustiques et phonétiques sur la parole Lombard en permettant peut être de segmenter automatiquement ce type de parole de façon plus fiable, et ainsi de gagner beaucoup de temps. Etant donné que certains indices de la parole sont particulièrement renforcés dans le bruit, cela devrait être envisageable. Nous avons de notre côté co-encadré un stage sur le développement d'un outil logiciel qui permettrait, à partir de la transcription phonétique des mots ou énoncés, d'en placer automatiquement les frontières de segmentation. Des résultats plutôt positifs nous incitent à poursuivre ce travail.

Table des matières

PARTIE I : ACQUIS ET HYPOTHESES SUR LE FORÇAGE VOCAL**Chapitre 1. Le forçage vocal**

1. Définitions du forçage vocal.....	8
1.1. Définitions du forçage vocal dans la littérature.....	8
1.2. Enquête auprès de différentes personnes concernées par le forçage vocal.....	10
2. Etat des connaissances sur le forçage vocal.....	14
2.1. Un forçage défini selon les études d'un niveau physiologique à un niveau intentionnel.....	14
2.2. Un forçage observé à différents niveaux et par différents champs disciplinaires.....	16
2.3. Un forçage défini par ses conséquences, mais observé à plusieurs niveaux de temporalité et de causalité.....	18
3. Positionnement et hypothèses de notre travail.....	25
3.1. Des stratégies individuelles de gestion de l'effort.....	26
3.2. L'évaluation de ces stratégies de gestion à partir des symptômes qu'elles provoquent.....	27
3.3. L'évaluation de ces stratégies de gestion à partir des individus qui les adoptent.....	27
3.4. L'évaluation de ces stratégies de gestion à partir du degré de répartition des efforts.....	28
3.5. L'évaluation de ces stratégies de gestion par l'adéquation de l'effort aux exigences de la situation de communication.....	31
4. Conclusion du chapitre 1.....	34

PARTIE II : METHODOLOGIES ET BASES DE DONNEES POUR L'ETUDE DE L'ADAPTATION DE LA PAROLE EN ENVIRONNEMENT BRUYANT**Chapitre 2. Mécanismes et facteurs de l'adaptation de la parole en environnement bruyant**

1. Les acquis des travaux antérieurs.....	40
1.1. Des objectifs différents selon les champs disciplinaires.....	40
1.2. A la recherche d'une modélisation unique des modifications de la parole dans le bruit.....	41
1.3. Une « variabilité » individuelle importante.....	43
2. Les mécanismes de l'adaptation de la parole dans le bruit.....	45
2.1. Une régulation basée sur le retour auditif de sa propre voix.....	45
2.2. Une régulation physiologique réflexe.....	48
2.3. Une régulation psychologique et communicationnelle.....	48
2.4. Une adaptation plus complexe qu'une simple régulation.....	50
3. Les méthodologies d'étude de l'adaptation de la parole dans le bruit.....	51
3.1. L'environnement bruyant.....	51
3.2. La situation de communication.....	51
3.3. La tâche de parole.....	52
3.4. Le traitement des données expérimentales.....	52
4. Positionnement de notre étude par rapport à ces différentes théories et méthodologies.....	53
4.1. Bilan sur les mécanismes gouvernant l'adaptation de la parole dans le bruit et sur les facteurs qui les influencent.....	53
4.2. Hypothèses sur les différences interindividuelles d'adaptation de la parole dans le bruit... ..	54

Chapitre 3. Influence de la méthode d'immersion dans le bruit sur l'adaptation de la parole

1. Les éventuels problèmes liés au port d'un casque	58
1.1. Perception de l'environnement sonore	58
1.2. Perception de l'interlocuteur	59
1.3. Retour auditif de sa propre voix.....	59
1.4. Modifications conséquentes de la parole	59
2. Recherche d'une méthode de débruitage.....	60
2.1. Revue de différentes méthodes d'atténuation du bruit.....	60
2.2. Comparaison de 3 méthodes numériques de débruitage	66
3. Comparaison de trois méthodes d'immersion du locuteur dans le bruit	72
3.1. Matériel et méthode	73
3.2. Résultats concernant l'influence du port du casque.....	78
3.3. Résultats concernant l'influence de l'ajout d'un retour auditif dans le casque.....	81
3.4. Conclusion sur les méthodes d'immersion du locuteur dans le bruit	84

Chapitre 4. Influence du dialogue avec un interlocuteur sur l'adaptation de la parole dans le bruit

1. Recherche d'alternatives à la lecture par un locuteur isolé	88
1.1. Les premières initiatives dans la littérature	90
1.2. Nos propositions. Intérêts et limites.....	90
2. Comparaison de deux tâches de parole avec et sans interaction avec un interlocuteur.....	95
2.1. Matériel et méthode	95
2.2. Résultats	98
2.3. Conclusion sur la comparaison des deux tâches de parole	100
3. Conclusion du chapitre 4	101

PARTIE III : CARACTERISATION DES STRATEGIES D'ADAPTATION DE LA PAROLE EN ENVIRONNEMENT BRUYANT

Chapitre 5. Principales observations

1. Quels sont les aspects de la parole significativement modifiés dans le bruit ?	108
1.1. Au niveau acoustique	109
1.2. Au niveau articulatoire	110
1.3. Au niveau glottique	113
2. Quelles sont les tendances partagées et les différences individuelles ?	115
2.1. Tendance partagée sans grande variabilité interindividuelle	115
2.2. Tendance unique avec différentes magnitudes.....	116
2.3. Tendances opposées	120
3. Quels sont les paramètres dont l'évolution dans le bruit est dépendante du type de bruit ?	123
3.1. Paramètres dont l'évolution dans le bruit ne dépend pas du type de bruit	123
3.2. Paramètres dont l'évolution dans le bruit montre la même dépendance au type de bruit pour tous les locuteurs	123
3.3. Paramètres dont l'évolution dans le bruit montre une dépendance au type de bruit différente selon les locuteurs.....	125
4. Conclusion du chapitre 5	127

Chapitre 6. Une adaptation visant à émerger du bruit ambiant ?

1. Préservation d'un rapport signal sur bruit acceptable	130
1.1. Relations entre l'intensité vocale et l'intensité du bruit ambiant	130
1.2. Des évolutions directement liées à l'augmentation de l'intensité vocale	132
2. Expansion ou compression de la dynamique vocale	141
2.1. Evolution de l'étendue vocale dans le bruit	141
2.2. Evolution de la dynamique d'intensité vocale dans le bruit	141
2.3. Evolution de l'ambitus de fréquence fondamentale dans le bruit	142
3. Adaptation en fréquences	145
3.1. Adaptation tonale	145
3.2. Adaptation spectrale	145
4. Conclusion du chapitre 6	154

Chapitre 7. Une adaptation visant à favoriser la reconnaissance des unités phonétiques ?

1. Renforcement du contraste acoustique entre les voyelles	156
1.1. Evolution de la discrimination des voyelles selon plusieurs indices perceptifs	157
1.2. Evolution du système vocalique dans son ensemble	162
1.3. Conclusion quant à la distinctivité acoustique des voyelles	164
2. Renforcement des indices visuels d'identification ou de discrimination des unités phonétiques	166
2.1. Renforcement du contraste visible entre les voyelles	166
2.2. Renforcement des indices visuels de reconnaissance des consonnes bilabiales	169
2.3. Conclusion quant à la distinctivité visuelle des segments	170
3. Allongement du temps de production ou de récupération de ces indices	171
3.1. Un ralentissement lié à l'amplification des gestes articulatoires	171
3.2. Un allongement de la durée de perception de ces indices	173
3.3. Conclusion quant à l'allongement de la durée des segments	174
4. Conclusion du chapitre 7	175

Chapitre 8. Une adaptation visant à renforcer les indices de structuration de l'énoncé ?

1. Indices de frontière syllabique	179
2. Indices de frontière entre un déterminant et une unité lexicale	181
2.1. Matériel et méthode	181
2.2. Résultats	181
2.3. Conclusion sur les indices de début d'une unité lexicale	183
3. Indices de début de l'énoncé	185
4. Indices de fin de l'énoncé	186
4.1. Matériel et méthode	186
4.2. Résultats	186
4.3. Conclusion quant aux indices de fin d'énoncé	188
5. Conclusion du chapitre 8	188

Chapitre 9. Une adaptation visant à mettre en valeur l'information importante ?

1. Matériel et méthode	190
2. Résultats	191
3. Conclusion	194

Chapitre 10. Synthèse quant à l'utilisation des différentes stratégies d'adaptation en fonction des individus et des contextes

1. Premiers bilans.....	198
1.1. Bilan concernant la possibilité de gestion de l'effort vocal par les individus communicant en environnement bruyant	198
1.2. Bilan concernant la différence de charge que ces modifications de la parole sont susceptibles de faire porter sur le larynx.....	199
1.3. Premier bilan sur les différences interindividuelles d'adaptation dans le bruit.....	202
2. Réorganisation de l'adaptation en fonction du contexte	205
2.1. Influence du type de bruit	209
2.2. Influence du niveau de bruit	210
3. Conclusion.....	214

CONCLUSION GENERALE DE L'ETUDE

TABLE DES MATIERES

REFERENCES BIBLIOGRAPHIQUES

THESE de DOCTORAT de l'UNIVERSITE PARIS 6
Ecole doctorale SMAE
Laboratoire d'Acoustique Musicale (UPMC, CNRS, Ministère de la Culture)

Spécialité :
Acoustique-Phonétique

Sujet de la thèse :

**Communiquer en environnement bruyant :
de l'adaptation jusqu'au forçage vocal**

Présentée par **Maëva Garnier**

Pour obtenir le grade de DOCTEUR de l'UNIVERSITE PARIS 6

ANNEXES DU MANUSCRIT

- Sommaire des Annexes -

ANNEXES BIBLIOGRAPHIQUES

Annexes Bib1.	Synthèse bibliographique des méthodologies d'étude des conséquences d'un effort vocal.....	237
Annexes Bib2.	Synthèse bibliographique des méthodologies d'étude de l'effet Lombard.....	239

ANNEXES METHODOLOGIQUES

Annexes Met1.	Principes et descripteurs de la production de parole.....	251
Annexes Met2.	Détails techniques liés à la constitution de bases de données.....	269
Annexes Met3.	Critères de segmentation.....	275

ANNEXES D'ETUDES PILOTES

Annexes Pil1.	Etude pilote de l'influence des retours de compréhension de l'interlocuteur.....	285
Annexes Pil2.	Corpus et étude articulatoire pilote.....	293

ANNEXES DE RESULTATS

Annexe Res1.	Questionnaire sur le forçage vocal.....	313
Annexe Res2.	Questionnaire sur la perception des environnements bruyants.....	339
Annexe Res3.	Résultats statistiques détaillés du chapitre 3.....	345
Annexe Res4.	Résultats statistiques détaillés du chapitre 4.....	355
Annexe Res5.	Résultats statistiques détaillés du chapitre 5.....	359
Annexe Res6.	Résultats statistiques détaillés du chapitre 6.....	379
Annexe Res7.	Résultats statistiques détaillés du chapitre 7.....	395
Annexe Res8.	Résultats statistiques détaillés du chapitre 8.....	403
Annexe Res9.	Résultats statistiques détaillés du chapitre 9.....	407

Annexes bibliographiques

Annexe Bib1 :

**Synthèse bibliographique
des méthodologies d'étude
des conséquences d'un
effort vocal**

Etude	Tâche d'effort	Durée de la phase de warm-up	Durée de la phase d'effort
Vinturi 2001 ^[368]	Lecture prolongée	Déterminée perceptivement	45 min
Sihvo et al. 1996 ^[317]	Lecture prolongée		45 min
Stone et al. 1973 ^[332]	Lecture de voyelles à différentes hauteurs et intensités		20 min
Elliot et al. 1995 ^[62]	Exercices vocaux chantés	30 min	X
Verstraete et al. 1993 ^[364]	Répétition de voyelles à différentes hauteurs	X	25 min
Laukkanen et al. 1996 ^[199]	Exercices vocaux du type bilabiale + voyelle	1 min	X
Lauri et al. 1997 ^[201]	Parole spontanée dans une salle de classe		45 min
Vilkman et al. 1999 ^[367]	Lecture prolongée		45 min
Rantala et al. 2002 ^[281]	Parole spontanée dans une salle de classe pendant une journée		1 journée
Neils et al. 1987 ^[252]	Lecture prolongée dans du bruit à 50, 70 et 90 dB		45 min
Gelfer et al. 1991 ^[111]	Lecture prolongée en parole forte	X	1h
Stemple et al. 1995 ^[329]	Lecture prolongée en parole forte		2h, avec une pause de 5min au bout d'1h
Scherer et al. 1991 (dans Stemple et al. 1995 ^[329])	Lecture prolongée à forte intensité et une octave au dessus du pitch habituel		Alternance 15min effort / 15 min pause jusqu'à épuisement
Linville 1995 ^[222]	Lecture prolongée à forte intensité	X	15 min
Buekers 1998 ^[38]	Lecture en parole forte, chant dans les deux mécanismes laryngés, imitation de voix et d'animaux, toux	X	30 min
Kitch et al. 1996 ^[182]	chant		La durée d'un spectacle

Tableau 33. Récapitulatif des différentes tâches et temps de parole des protocoles de plusieurs études récentes sur l'usage prolongé de la parole.

Annexe Bib2:

**Synthèse bibliographique
des méthodologies d'étude
de l'effet Lombard**

Etude	Type(s) de bruit	Niveau(x) de bruit	Immersion du locuteur dans le bruit
Amazi et al. 1982 ^[8]	Bruit de conversations	90 dB SPL	Casque avec retour de sa propre voix
Bond et al. 1989 ^[31]	Bruit rose	95 dB	Casque avec
Boril et al. 2005 ^[32]	25 bruits de voiture (issus de CAR2E) , 4 bruit blancs entre 62-125Hz, 75-300Hz, 220-1120Hz et 840-2500Hz	90 dB SPL	Casque
Castellanos et al. 1996 ^[46]	Bruit blanc	85dB SPL	Casque
Davis et al. 2006 ^[62]	Bruit blanc, Bruit de conversations	80 dB SPL	Casque, Haut-parleurs
Dejonckere et al. 1983 ^[65]	Bruit blanc	30, 50, 70, 90 dB	Casque
Dieroff et al. 1966 ^[71]	Bruit d'enfants dans une maternelle	85-90 dB	In situ
Dohalska et al. 2000 ^[73]	Bruit de trafic ferroviaire	Non précisé	In situ
Egan 1972 ^[80]	Bruit blanc large bande, basse fréquences, hautes fréquences et médium	20 à 120 dB par pas de 10 dB	Casque
Hofler 1984 ^[149]	Bruit blanc	100-105 dB (A)	Casque
Junqua 1992 ^[173]	Bruit blanc et bruit de conversations	85dB SPL	Casque
Junqua 1993 ^[171]	Bruit blanc	85dB SPL	Casque
Kadiri 1998 ^[176]	Bruit blanc et bruit de conversations	85dB SPL	Casque avec et sans retour de sa propre voix à 72 dB
Kim et al. 2005 ^[179]	Bruit blanc et bruit de conversations	80dB SPL	Casque, Haut-parleurs
Kim 2005 ^[180]	Distance par rapport à l'interlocuteur	0.5, 1, 2 et 3 m	-
Korn 1954 ^[185]	Bruit blanc de même enveloppe spectrale qu'un bruit de conversations	40, 50, 60, 70, 80, 90 dB	Casque
Lamprecht 1988 ^[191]	Bruit blanc	95 dB (A)	Casque
Lane et al. 1970 ^[194]	Bruit blanc	71, 75, 79 et 83 dB	Casque avec retour de sa propre voix
Lee et al. 2004 ^[204]	Bruit de voiture tournant au ralenti, roulant à 35 et 55 miles /h avec la vitre ouverte ou fermée	Non précisé	In situ
Mixdorff et al. 2006 ^[244]	Bruit de conversations	60, 70 et 80 dB SPL	Haut-parleurs
Mokbel 1992 ^[245]	Bruit de voiture roulant à 90 et 130 km/h, 4 bruits blanc : large bande, filtré par un passe bas à 1kHz, à 1.5 kHz ou par un passe haut à 1.7 kHz	66 dB (A), 79 dB (A)	Casque
Nonaka et al. 1997 ^[256]	Son pur, 1kHz	105dB SPL	Haut-parleurs
Papon 2006 ^[265]	Bruit de conversations	80 dB SPL	Casque
Pick et al. 1989 ^[271]	Bruit blanc	90 dB SPL	Casque
Pisoni et al. 1985 ^[272]	Bruit blanc filtré par un passe-bas à 3.5 kHz	80, 90, 100 dB	Casque
Schultz-Coulon et al. 1976 ^[305]	Bruit blanc	60, 80, 100, 110 dB	Casque
Siegel et al. 1974 ^[315]	Bruit de même enveloppe spectrale qu'un bruit de conversations	60, 70, 80 dB SPL	Casque avec retour de sa propre voix
Sinnott et al. 1975 ^[318]	Bruit dans la bande 200-500Hz, Bruit dans la bande 8-16kHz	70, 80, 90 dB SPL	Casque
Sodersten et al. 2005 ^[322]	Bruit d'une classe calme, bruit de ventilation, bruit de classe bruyante, bar avec de la musique forte	30, 70-78, 74, 78-85, 87 dB	Haut-parleurs
Stanton 1988 ^[327]	Bruit rose	90 dB	Casque
Ternström et al. 2002 ^[346]	bruit blanc stationnaire, maternelle, bar avec de la musique forte, bruit de ventilation	68, 70, 82, 85 dB SPL	Haut-parleurs

Ternström et al. 2006 ^[345]	Bruit d'une classe calme, bruit de ventilation, bruit de classe bruyante, bar avec de la musique forte	30, 70-78, 74, 78-85, 87 dB	Haut-parleurs
Tonkinson 1994 ^[353]	Chorale	65 à 95 dB SPL	Casque
Frank et al. 2003 ^[97]	Bruit rose	60, 70, 80, 90, 100 SPL	Casque + protections auditives
Van Heusden et al. 1979 ^[360]	Bruit blanc de même enveloppe spectrale qu'un bruit de conversations	35, 45, 55, 65 dB(A)	Casque spécial, sans atténuation
	Salle réverbérante, cantine, salle informatique, voiture, chaufferie	31, 39, 61, 65, 75 dB(A)	In situ
Van Summers et al. 1988 ^[361]	Bruit blanc	80, 90, 100 dB	Casque
Webster 1962 ^[372]	Bruit de conversations	65, 75, 85 dB (C)	Bruit « in situ », mais condition de laboratoire
Welby 2006 ^[375]	Bruit blanc	80 dB	Casque
Zeiliger et al. 1994 ^[382]	Bruit blanc et bruit de conversations	85dB SPL	Casque avec et sans retour de sa propre voix
Zeine et al. 1988 ^[383]	Bruit de même enveloppe spectrale qu'un bruit de conversations	100 dB SPL	Casque

Tableau 34. Bilan des types de bruit, niveaux de bruit et moyens de restitutions utilisés par plusieurs études de l'effet Lombard

Etude	Tâche de parole	Matériel linguistique	Interaction
Amazi et al. 1982 ^[8]	désignation d'images, puis contage d'une histoire à propos de ces images	25 cartes représentant une image	oui + système de gratification
Bond et al. 1989 ^[31]	lecture	10 spondées issus de Hirsh et al. 1952 ^[148]	non
Boril et al. 2005 ^[32]	lecture	Phrases équilibrées phonétiquement, chiffres et nombres, commandes, dates, heures, etc.	non
Castellanos et al. 1996 ^[46]	lecture	13 phrases équilibrées phonétiquement	non
Davis et al. 2006 ^[62]	lecture	10 phrases équilibrées issus des phrases de 1969 ^[2]	Auditeur passif à 2.5 m
Dejonckere et al. 1983 ^[65]	Lecture	Texte d'environ 30s	non
Dieroff et al. 1966 ^[71]	Parole spontanée	-	oui
Dohalska et al. 2000 ^[73]	Parole spontanée	-	Non précisé
Egan 1972 ^[80]	Lecture	44 phrases de 7 à 12 syllabes	non
Hofler 1984 ^[149]	lecture	Passage d'un texte	non
Junqua 1992 ^[173]	Lecture	9 voyelles en contexte h-d, puis en contexte d-d et s-s. Logatomes incorporés dans des phrases types.	non
Junqua 1993 ^[171]	Lecture	49 mots (chiffres, lettres, commandes)	non
Kadiri 1998 ^[176]	Lecture	50 phrases équilibrées, nombres, 82 logatomes de type C(C)VC	non
Kim et al. 2005 ^[179]	Lecture	50 mots (commandes) dont 4 monosyllabiques, 36 bisyllabiques, 6 trisyllabiques et 4 quadrisyllabiques.	Ordinateur passif à 1.5, 1, 2 et 3 m
Kim 2005 ^[180]	Lecture	10 phrases équilibrées issus des phrases de 1969 ^[2]	Auditeur passif à 2.5 m
Korn 1954 ^[185]	Dialogue spontané	-	oui
Lamprecht 1988 ^[191]	-	Voyelle /a/ tenue	-
Lane et al. 1970 ^[194]	Enonciation de phrases types	Phrase porteuse avec un mot cible changeant (commande aéronautique)	Retour de l'auditeur par le biais d'une ardoise
Lee et al. 2004 ^[204]	lecture	13 chiffres, 26 lettres, 10 numéros de téléphone, 20 phrases	non
Mixdorff et al. 2006 ^[244]	lecture	Phrases phonétiquement équilibrées	non
Mokbel 1992 ^[245]	Lecture	12 mots (chiffres et commandes)	non
Nonaka et al. 1997 ^[256]	Phonation induite par stimulation électrique	-	-
Papon 2006 ^[265]	lecture	Mots du jeu Pictionary, mots construits et non existants	Oui, système d'ardoise
Pick et al. 1989 ^[271]	Parole spontanée	-	non
Pisoni et al. 1985 ^[272]	Lecture	15 mots aéronautiques (chiffres et commandes)	non
Schultz-Coulon et al. 1976 ^[305]	Lecture	Texte	non
Siegel et al. 1974 ^[315]	Parole spontanée	-	non
Sinnott et al. 1975 ^[318]	Induit par conditionnement	Cri d'appel	non
Sodersten et al. 2005 ^[322]	lecture	Texte	Retour de la compréhension d'auditeur (invisible) via un vumètre
Stanton 1988	lecture	56 phrases	non

[327]			
Ternström et al. 2002 [346]	lecture	5 petits textes différents pour chaque condition	Retour de la compréhension d'auditeur (invisible) via un vumètre
Ternström et al. 2006 [345]	lecture	6 textes de 90 s chacun	Retour de la compréhension d'auditeur (invisible) via un vumètre
Tonkinson 1994 [353]	Chant	-	non
Frank et al. 2003 [97]	Lecture	12 passages	non
Van Heusden et al. 1979 [360]	Dialogue spontané	-	oui
Van Summers et al. 1988 [361]	Lecture	15 mots aéronautiques (chiffres et commandes)	non
Webster 1962 [372]	Lecture	Liste de mots	oui
Welby 2006 [375]	Lecture	Court texte	non
Zeiliger et al. 1994 [382]	lecture	30 phrases phonétiquement équilibrées, 82 logatomes de type C(C)VC, 30 nombres	non
Zeine et al. 1988 [383]	Lecture	Voyelle /a/ tenue, liste de chiffres, 50 mots	non

Tableau 35. Bilan des types de bruit, niveaux de bruit et moyens de restitutions utilisés par plusieurs études de l'effet Lombard

Etude		Principaux résultats
Bond et al. 1989 ^[31]	Lombard	Élévation de la F0, de l'intensité vocale, de la durée des voyelles uniquement sur la deuxième syllabe, augmentation de F1, surtout pour /u/, augmentation de F2 pour /u/ et diminution pour /i/, pas de modification pour les autres
Boril et al. 2005 ^[32]	Lombard	Déplacement de la distribution de F0 vers l'aigu. Augmentation de F1, surtout pour /a/. Augmentation plus légère de F2, sauf pour /i/. Tableau récapitulatif des augmentations de F1 et F2 et de leur largeur de bande pour les voyelles /a/, /e/, /i/, /o/, et /u/
Borsuk and Klajman 1967	Voix forte, Lombard	Pas de différence significative entre la voix forte et la voix lombard au niveau spectral
Castellanos et al. 1996 ^[46]	Lombard	Pas d'augmentation significative de la durée des consonnes, mais augmentation de la durée des voyelles, sauf pour /u/. L'intensité augmente sur les voyelles et les consonnes voisées (de 9 à 16dB pour un bruit ambiant de 95dB), mais a tendance à diminuer sur les non voisées. Augmentation de la F0, davantage pour les hommes que les femmes. (+40% en moyenne sur les voyelles) Augmentation de la pente spectrale BF pour les plosives, les nasales, les liquides et les voyelles, mais pas pour les fricatives non voisées. Diminution de la pente spectrale HF pour segments voisés (plosives voisées, nasales, liquides et voyelles). Augmentation au contraire pour les fricatives non voisées. Augmentation de F1 et de F2 pour toutes les voyelles Différences observées entre les hommes et les femmes en particulier sur les nasales, les fricatives et les liquides, mais pas vraiment sur les voyelles. L'augmentation du spectre est centré autour de 5kHz chez les femmes sur les voyelles, les nasales et les liquides, tandis qu'il est entre 2 et 4kHz chez les hommes, avec au contraire un trou spectral à 5kHz. Par contre, comportement assez similaire pour les plosives et les fricatives : les fricatives sont atténuées, sauf dans les médium (entre 1 et 2,5kHz), et les plosives sont légèrement renforcées dans les médiums (2 à 4 kHz).
Davis et al. 2006 ^[62]	Lombard	Augmentation de l'intensité (+11dB pour un bruit de 80dB), de la durée des phrases (+290 ms), augmentation de F1, F2 et surtout F3. Plus grande augmentation de F3 dans le bruit blanc que dans le babble. Augmentation de F0 et de l'ambitus de F0. L'ambitus est plus important dans le bruit blanc que dans le babble. Au niveau visuel, les mouvements du visage peuvent être décrits à 90% par une augmentation de la mâchoire, de l'ouverture de la bouche, de la protrusion, de la fermeture de la bouche. Pas d'avancée significative de la tête vers l'interlocuteur Les mouvements articulatoires sont plus amples dans le bruit cocktail que dans le bruit blanc, et plus amples par restitution au casque que par restitution haut-parleurs.
Dejonckere et al. 1983 ^[65]	Lombard	Augmentation de l'intensité moyenne avec une pente de 0.083 pour un bruit ambiant de 30 à 90 dB. Augmentation de l'intensité maximale avec le bruit ambiant (r=0.39)
Dohalska et al. 2000 ^[73]	Lombard	La reconstitution d'un message peut se faire, dans un contexte et un vocabulaire restreint à partir de patrons vocaliques. Les voyelles /i/, /a/, /u/ sont moins confondues que les voyelles /e/ et /o/
Egan 1972 ^[80]	Lombard	Plus grande augmentation de l'intensité vocale chez les femmes que chez les hommes (3 à 4 dB). Plus grande augmentation dans le bruit blanc large bande et médium que dans le bruit blanc basses fréquences et hautes fréquences. La variabilité individuelle est plus marquée à forte intensité de bruit ambiant, et davantage chez les femmes. L'intensité vocale augmente de façon « random » en dessous d'un certain niveau de bruit ambiant, variable selon les types de bruit (en moyenne 50dB), puis augmente de façon linéaire. Il existe un pallier dans l'augmentation vers 100dB de bruit pour le bruit blanc large bande, quel que soit le genre du locuteur.
Fairbanks et al. 1957 ^[87]	Voix forte	Le rapport entre l'intensité des voyelles et des consonnes est d'environ -7dB en parole conversationnelle et de -14dB en effort important, mais non extrême.
Hanley et al. 1949 ^[133]	Lombard	Diminution du débit de parole: l'augmentation de la durée d'une syllabe peut être modélisée par $y = 0.426 \cdot \log(x) + 0.109$ (en s) L'augmentation de l'intensité vocale peut être modélisée soit par une droite $y = 2.35x + 2.13$ ou par $y = 3.72 \cdot \log(x) + 1.64$
Huang et al.	Voix forte	Augmentation du quotient fermé avec l'effort vocal. Diminution du Jitter et du

1995 ^[160]		Shimmer
Junqua 1992 ^[173]	Lombard	<p>Augmentation de l'intensité, de la F0 (davantage pour les hommes, et moins pour les enfants) et de la durée.</p> <p>Augmentation de F1</p> <p>L'espace vocalique dans le plan F2-F1/F1 est plus important pour les enfants mais comparable entre les hommes et les femmes, et tend à diminuer avec le bruit ambiant. Les valeurs formantiques sont plus élevées chez les femmes (parole plus ouverte).</p> <p>Globalement, le triangle vocalique pivote par rapport au /i/ lorsque le niveau de bruit augmente, et se déplace vers des valeurs formantiques plus élevées de F1 et F2.</p> <p>Au niveau articulatoire, une bonne répétabilité des mouvements pour chaque condition est observée. Le type de bruit, le contexte consonantique et le locuteur influencent l'articulation des voyelles dans le bruit : La parole produite dans le babble se rapproche de la parole claire, et est différente de la parole dans le bruit blanc.</p>
Junqua 1993 ^[171]	Lombard	<p>Augmentation de la durée des voyelles, légère diminution de la durée des consonnes, augmentation globale de la durée des syllabes.</p> <p>Augmentation de la F0 davantage chez les hommes que chez les femmes</p> <p>Le centre de gravité spectral augmente pour tous les phonèmes, en particulier les voyelles.</p> <p>Diminution de l'énergie des voyelles en dessous de 500Hz, également mais moindre entre 500 et 1000Hz. Augmentation de l'énergie chez les femmes entre 4 et 5 kHz. Diminution de l'énergie spectrale dans toutes les bandes pour les nasales, fricatives, plosives.</p> <p>La pente spectrale BF augmente pour les voyelles, liquides et nasales tandis que la pente spectrale HF diminue pour la plupart des voyelles et des liquides.</p> <p>Augmentation de F1 pour les voyelles, liquides et nasales, davantage pour les femmes, et indépendamment du contexte. Augmentation de F2 seulement pour les femmes.</p>
Kadiri 1998 ^[176]	Lombard	<p>Augmentation de la durée des voyelles (+23%), des mots (+14%) et des phrases (+16%). Pas de différence sur la durée entre restitution avec ou sans retour de la voix dans le casque, ni entre le bruit blanc et cocktail.</p> <p>Augmentation de F0, plus importante sans retour de sa propre voix (+23%) qu'avec (+18%) et légèrement plus grande dans le bruit blanc que dans le cocktail (+2%)</p> <p>Augmentation de l'intensité vocale, plus importante avec retour (+112%) que sans (65%) (!)</p> <p>Augmentation de F1, indépendant du retour de la voix dans le casque. Pas de variation significative des autres formants.</p> <p>L'augmentation du couple F0/ Intensité est plus importante chez les femmes que chez les hommes.</p>
Kim et al. 2005 ^[179]	Lombard	Cf. Davis et al. 2006 ^[62] .
Kim 2005 ^[180]	Voix à distance	<p>Augmentation de la durée des mots de 9% en moyenne, moins importante plus le mot a de syllabes, au point que l'augmentation de la durée n'est pas significative pour les mots de 4 syllabes.</p> <p>Plus grande augmentation sur la 1ère syllabe, plus grande augmentation chez les femmes. Diminution de la durée des consonnes (-9% en moyenne, plus importante chez les hommes) et augmentation de celle des voyelles (+34%, plus importante chez les femmes). Ces résultats sont vrais pour les voyelles et les consonnes non voisées, mais pas de modification significative de la durée des consonnes voisées, nasales et liquides. En particulier les plosives sont plus courtes de 15.6%.</p> <p>Il existe des stratégies individuelles différentes sur les nasales et les liquides, allongées ou raccourcies selon les locuteurs.</p>
Lane et al. 1970 ^[194]	Lombard	Augmentation de l'intensité vocale de 5dB pour une augmentation du bruit ambiant de 10 dB
Lienard et al. 1999 ^[215]	Voix à distance	<p>Augmentation de F0 à 5Hz/dB d'intensité vocale, et augmentation de F1 à 3.5 Hz/dB d'intensité vocale. Forte corrélation entre F0, F1 et l'effort vocal (r=0.43). F1-F0 varie moins que F1 avec l'effort vocal, mais toujours significativement.</p> <p>Variation non significative de F2 et F3. Par contre, augmentation de l'amplitude des trois premiers formants (+11, +12 et +13dB pour une augmentation de 10dB de l'intensité vocale). Augmentation de l'énergie dans les hautes fréquences du spectre.</p> <p>F2-F1 varie avec l'effort vocal, mais pas F3-F2.</p>
Mixdorff et al.	Lombard	Augmentation de F1, de F2 sur le /u/ pour l'allemand. Pas de réduction

2006 ^[244]		significative de la largeur des formants
Mokbel 1992 ^[245]	Lombard	Augmentation de la pente spectrale, amplification de l'énergie dans les zones fréquentielles où sont situés des pics de bruit. La transformation de la parole normale en parole Lombard peut être modélisée par une homothétie et une translation de l'espace des paramètres spectraux.
Pisoni et al. 1985 ^[272]	Lombard	Augmentation de l'intensité vocale (environ 7dB pour 90 dB de bruit ambiant), de la durée des mots (+65ms) et des voyelles (+50ms). Augmentation de la F0, de la pente spectrale (+1.5dB/octave). Diminution de l'énergie basse fréquences (-1.3 dB) et augmentation de l'énergie haute fréquences (+2.4dB) Augmentation de F1, et diminution de F2.
Rostolland 1982 ^[288]	Voix criée	Plus grande augmentation de l'intensité chez les hommes (+28 dB) que chez les femmes (+20 dB). Augmentation importante de l'énergie au dessus de 500 Hz, augmentation de la F0, La durée des voyelles augmente (+67% pour la voyelle finale, +33% pour la voyelle de la 1ère syllabe) et celle des consonnes diminue (-20% pour la consonne intervocalique d'un CVCV)
Rostolland 1982 ^[289]	Voix criée	Augmentation de F1, sauf pour /a/ où une légère diminution est observée. Réduction de l'espace vocalique. F3 est comparable pour toutes les voyelles sauf pour /i/ où il est plus important et /u/ où il est plus faible. F2 reste assez distinctif entre les voyelles. Augmentation de l'énergie des voyelles vers 4-5kHz, et parfois au delà de 10kHz. L'intonation est affectée par l'augmentation de la F0 moyenne. Rostolland parle de "frequency clipping", comme si le locuteur ne pouvait plus moduler sa voix. La fin des mots en particulier montre une forme de plateau pour la courbe intonative.
Schulman 1989 ^[304]	Voix forte	Augmentation de l'ouverture de la mâchoire, et de l'ouverture labiale. La dimension vocalique ouvert/fermé montre encore une distinction entre les différentes voyelles, tandis que l'articulation labiale maintient la distinction entre voyelles étirées et arrondies. L'augmentation de l'ouverture de la mâchoire peut être modélisée de façon linéaire. L'augmentation de l'ouverture labiale peut être également modélisée de façon linéaire, mais par des fonctions différentes selon la catégorie vocalique (spread, inrounded, outrounded). Une compensation de la lèvre supérieure est observée à la plus grande ouverture de mâchoire, sur les bilabiales Augmentation de la vitesse des mouvements. Les bilabiales intervocaliques sont raccourcies, tandis que les voyelles sont allongées
Sodersten et al. 2005 ^[322]	Lombard	Augmentation de l'intensité vocale, de la F0, du temps de lecture et du pourcentage de voisement. L'aire du phonétogramme commence par augmenter avec un niveau de bruit faible ou moyen, puis diminue à fort niveau de bruit. La perception de voix pressée, forcée, instable et rauque augmente avec le niveau de bruit. Au contraire, le vocal fry diminue. Les femmes montrent un pourcentage de voisement plus élevé. Leur phonétogramme est plus large que celui des hommes. Les femmes ont l'impression de moins bien se faire entendre que les hommes, et de dépenser plus d'effort.
Stanton 1988 ^[327]	Lombard	Parole forte et Lombard semblables pour : Augmentation de F0, de la durée des voyelles et des liquides, diminution de la durée des plosives. Augmentation de la pente spectre BF et diminution de la pente spectre HF, augmentation du centre de gravité spectral. Diminution de l'énergie entre 0 et 500Hz, sauf les nasales dont l'énergie augmente entre 250 et 500Hz. Augmentation de l'énergie spectrale entre 500Hz et 3kHz pour les voyelles, liquides et nasales et diminution pour les fricatives. Augmentation de l'énergie entre 3 et 5kHz pour les fricatives et diminution pour les nasales. Diminution de l'énergie entre 5 et 8kHz pour les voyelles, liquides et nasales, et augmentation pour les fricatives et plosives. Augmentation de F1. Parole forte et Lombard différentes pour : Augmentation de l'énergie des voyelles, des liquides et des plosives entre 3 et 5kHz en parole Lombard tandis qu'elle diminue en parole criée. Augmentation de la pente spectrale HF sur /p/ et /f/ en parole forte, et diminution en parole Lombard.
Ternström et al. 2002 ^[346]	Lombard	Augmentation de l'intensité (+12-15dB pour 82-85dB de bruit ambiant), de la F0, et diminution de l'aire du phonétogramme. Les femmes ont globalement toujours une aire de phonétogramme plus importante. Le temps de phonation est plus important chez les femmes que chez les hommes dans le bruit. Le centre de gravité spectral est relativement stable chez les hommes, autour de 600Hz, tandis qu'il augmente proportionnellement à la F0 chez les femmes.

Ternström et al. 2006 ^[345]	Lombard	Augmentation du premier formant à très forte intensité vocale. La balance spectrale, définie comme la différence de l'intensité entre 100-1000 Hz et entre 2-6kHz, augmente avec l'intensité vocale puis « sature » vers 93B pour les /a/ des femmes et 97.5dB chez les hommes. Ce niveau d'intensité de saturation est variable selon les individus et selon les voyelles. Pour les /i/, la balance spectrale diminue à très forte intensité vocale. Cette saturation est interprétée comme un changement de stratégie de la parole forte au cri.
Traunmüller et al. 2000 ^[356]	Parole à distance	Augmentation de l'intensité vocale avec un doublement de la distance, de façon comparable pour les hommes et les femmes. Augmentation de F0 et F1 avec l'effort vocal, par contre pas de variation de F3. Augmentation de la durée des segments voisés avec la distance. Diminution du nombre de pauses avec la distance chez les hommes et les enfants, et augmentation chez les femmes. Cependant, à très grande distance, les hommes augmentent aussi brutalement le nombre de pauses
Frank et al. 2003 ^[97]	Lombard avec protection auditives	Les locuteurs parlent moins fort avec des protections auditives, que ce soit dans le silence ou dans le bruit, mais cela est encore plus marqué dans le bruit. De même leur centre de gravité spectral est plus bas avec des protections auditives que sans.
Van Summers et al. 1988 ^[361]	Lombard	Augmentation de l'intensité, de la F0, de la durée, diminution de la pente spectrale. Différence entre les deux sujets de l'études : le premier augmente très peu sa F0, ni F1, mais augmente par contre plus la durée. A l'inverse, le deuxième augmente davantage sa F0, son F1, indépendamment de F0, mais n'augmente pas beaucoup la durée. Pas d'effet significatif sur F2.
Webster 1962 ^[372]	Lombard	Augmentation de l'intensité vocale de 5dB pour une augmentation du bruit ambiant de 10 dB
Welby 2006 ^[375]	Lombard	Elévation de la F0 moyenne. 4 locutrices sur 7 augmentent leur étendue de F0. Coude de F0 pas davantage renforcé dans le bruit que dans le silence

Tableau 36. Bilan des types de bruit, niveaux de bruit et moyens de restitutions utilisés par plusieurs études de l'effet Lombard

Annexes méthodologiques

Annexe Met1 :

Principes et descripteurs de la production de parole

- 1. RAPPELS SUCCINCTS DE LA PHYSIOLOGIE DE LA PAROLE**
- 2. LA MODULATION DE LA VOIX EN FREQUENCE ET EN INTENSITE**
- 3. LE TIMBRE DE LA VOIX**
- 4. LE DOUBLE ROLE DE L'ARTICULATION**

1. RAPPELS SUCCINCTS DE LA PHYSIOLOGIE DE LA PAROLE

1.1. Description synthétique du larynx

Les cordes vocales sont attachées d'un côté au cartilage thyroïdien, de l'autre aux 2 cartilages aryténoïdes (cf. Figure 132). Lors de l'inspiration, les muscles aryténoïdiens sont relâchés. Les cordes vocales sont alors distantes et forment un « V » ouvert, laissant passer l'air sans obstacle dans la trachée. En attitude de phonation, les muscles aryténoïdiens compressent les aryténoïdes et rapprochent les cordes vocales (cf. Figure 132). Le passage de l'air au travers de la glotte permet alors de mettre en vibration les cordes vocales dans un mouvement passif de collision périodique. Les cordes vocales se composent d'un muscle : le vocalis, recouvert d'une muqueuse. La longueur, la masse et la raideur des cordes vocales peuvent se contrôler par des actions musculaires, engendrant des modes de vibration différents. En particulier, la contraction du vocalis entraîne un raccourcissement et un raidissement des cordes vocales. La contraction du muscle crico-thyroïdien, quant à lui, fait basculer le cartilage thyroïde vers le bas, provoquant un étirement, un affinement et un raidissement des cordes vocales (cf. Figure 132).

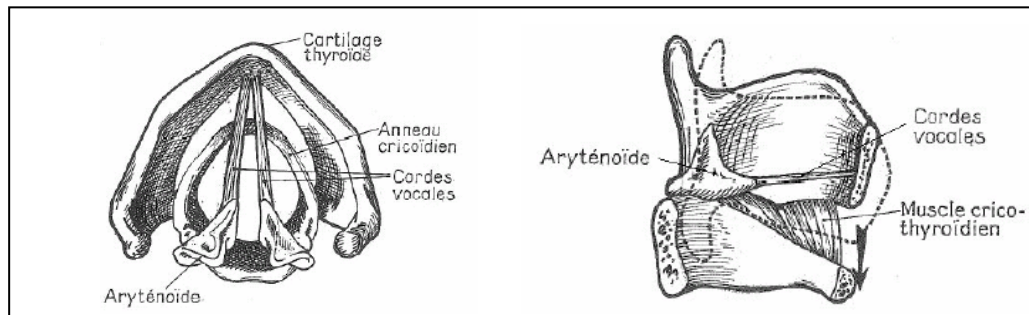


Figure 132. Représentation du larynx vu de dessus et de profil. D'après Legent et al. 1975^[205]

1.2. Action des différents articulateurs

Le mouvement des différents articulateurs permet de contrôler dans une certaine mesure la forme du conduit vocal et la taille des différentes cavités de résonance par lesquelles le son glottique est filtré.

Ainsi, la production des voyelles peut être principalement décrite à partir des mouvements de la pointe, du dos et du corps de la langue (Calliope 1989^[41]). L'abaissement de la mandibule permet d'élargir la cavité buccale. Le mouvement des lèvres permet également de modifier la forme de la cavité buccale en coordination avec la langue et la mandibule. Les principaux mouvements des lèvres, à la base de la catégorisation phonologique des voyelles, consistent en l'arrondissement, la protrusion (avancée des lèvres vers l'avant) et l'étirement. Les lèvres peuvent également s'ouvrir « passivement » en suivant les mouvements de la mandibule. Elles peuvent également se fermer en se pinçant pour produire des consonnes bilabiales telles que [m], [p] ou [b]. Les mouvements du voile du palais permettent d'ajuster le degré de couplage entre le conduit vocal et le conduit nasal. Enfin, le larynx peut être également considéré comme un articulateur, dans le sens où son abaissement vertical permet d'allonger le conduit vocal et d'agrandir la cavité pharyngée.

2. LA MODULATION DE LA VOIX EN FREQUENCE ET EN INTENSITE

2.1. L'intensité de la voix

Rappels

L'augmentation de la pression sous-glottique est le premier moyen pour accroître l'intensité vocale (Sundberg et al. 1993^[338] ; Alku et al. 2002^[7] ; Fant 1960^[88]). Le contrôle de la pression sous-

glottique est assuré par les muscles expirateurs (abdominaux, diaphragmatiques ou thoraciques). L'intensité vocale est également liée au débit glottique, inversement proportionnel à la surface de l'orifice d'écoulement. Ainsi, l'intensité vocale croît avec l'adduction des cordes vocales et le rétrécissement de la glotte, contrôlé par l'action des muscles aryténoïdiens (Leothaud 2005 ^[209]). Les résonateurs peuvent également permettre de renforcer l'énergie spectrale dans certaines bandes de fréquences, et par conséquent avoir une influence sur l'intensité vocale globale. De même, l'ajustement de la fréquence fondamentale de façon à ce qu'un harmonique de la voix coïncide sur une résonance du conduit vocal (en général le premier formant), aussi appelé « formant tuning », peut permettre d'augmenter l'intensité vocale autrement que par l'augmentation de la pression sous-glottique seulement (Miller et al. 1990 ^[240] ; Carlsson et al. 1992 ^[43] ; Henrich et al. 2006 ^[142]).

Mesure

L'intensité acoustique est une mesure de la puissance du signal acoustique par unité de temps, et permet d'estimer la force de voix. L'intensité en dB se calcule très classiquement par la formule :

$$Intensité_{dB} = 20 \cdot \log\left(\frac{P}{P_0}\right), \text{ où } P \text{ est le signal de Pression acoustique et } P_0 = 20\mu\text{Pa, le seuil d'audibilité servant de niveau de référence.}$$

Le signal audio en Volts, acquis à l'aide d'un microphone, est proportionnel au signal de pression d'un facteur α de transduction (à déterminer en effectuant une calibration du niveau sonore). On peut donc exprimer l'intensité acoustique par la formule :

$$\begin{aligned} Intensité_{dB} &= 10 \cdot \log\left(\frac{V^2}{P_0^2}\right) + 20 \cdot \log(\alpha), \text{ où } V \text{ représente l'amplitude du signal audio.} \\ &= 10 \cdot \log\left(\frac{V^2}{P_0^2}\right) + CalibDiff \end{aligned}$$

Le facteur correctif CalibDiff correspond à la différence entre l'intensité d'un son de calibration (un [a] tenu produit par le locuteur), mesurée au sonomètre au niveau des lèvres du locuteur, et l'intensité calculée à partir de l'enregistrement audio de ce même son de calibration (par la formule $10 \cdot \log(V^2/P_0^2)$).

Les mesures d'intensité sont donc présentées en dB SPL (pour « sound pressure level »). Dans les 2 bases de données constituées ainsi que dans le corpus pilote, nous avons choisi d'utiliser un microphone cardioïde pour sa directivité permettant d'obtenir de meilleurs résultats de débruitage du bruit ambiant (cf. Chapitre 4). Nous avons choisi des microphones à la réponse la plus plate possible en fréquence. Nous sommes néanmoins conscients que ce type de microphone induit un effet de proximité avantageant l'énergie des basses fréquences et que cela peut avoir un effet sur nos mesures d'intensité sonore.

2.2. La hauteur de la voix

Rappels

La fréquence de vibration des cordes vocales dépend de trois facteurs principaux : la longueur des cordes vocales, leur masse mise en vibration et leur raideur. L'ajustement des muscles intrinsèques du larynx, permettant de contrôler ces paramètres, peut se faire de différentes façons. Les travaux de Roubeau et al. 2004 ^[292] ou Henrich 2006 ^[140] ont contribué à préciser la notion de mécanisme laryngé comme une « configuration glottique particulière, caractérisée par la forme des cordes vocales (longueur, épaisseur) ainsi que par la tension musculaire mise en jeu ». Il existe 4 mécanismes laryngés que nous exposons ici par ordre croissant de fréquence fondamentale correspondante :

- le mécanisme 0 (M0) ou « fry » correspond aux fréquences les plus graves émises par un individu. La perception de hauteur est mal définie et l'on entend plutôt une succession d'impulsions régulières. Le mécanisme est associé à une tension minimale des cordes vocales (Hollien et al. 1968 ^[152])
- le mécanisme 1 (M1, défini de façon physiologique et objective) est souvent assimilé au registre de « voix de poitrine » (défini perceptivement) et est utilisé pour produire les fréquences graves et médiums de la tessiture. Dans ce mécanisme, la raideur des cordes vocales augmente avec la fréquence fondamentale, du fait de la contraction du vocalis. Les cordes vocales vibrent sur la globalité de leur masse et sur toute la longueur des cordes vocales. La phase ouverte de la glotte au cours de la vibration est plutôt courte, correspondant à un quotient ouvert laryngé assez faible (Henrich et al. 2005 ^[141]).

- le mécanisme2 (M2, défini de façon physiologique et objective) est de la même façon souvent assimilé au registre de « voix de tête » et il est utilisé pour produire les fréquences médiums et aiguës de la tessiture. Dans ce mécanisme, seule une petite partie de la masse des cordes vocales participe encore à la vibration avec la couche muqueuse, et les cordes vocales ne vibrent plus que sur 2/3 de leur longueur. L'élévation de la fréquence fondamentale n'est plus contrôlée par la contraction du vocalis, mais par la tension des muscles crico-thyroïdiens, provoquant une nouvelle tension croissante des cordes vocales, cette fois par leur élongation. Le quotient ouvert est plus important en M2, puis diminue avec l'augmentation de la F0 (Henrich et al. 2005^[141]).
- le mécanisme 3 (M3) ou « sifflet » est utilisé pour produire les fréquences extrêmement aiguës de la tessiture. Les cordes vocales sont alors très fines et très tendues.

Le mécanisme 3 n'est pas utilisé en parole et le mécanisme 0 très peu, parfois sur des hésitations, des pauses, des attaques ou des fins de phrases. La voix parlée utilise essentiellement les mécanismes 1 et 2 pour les femmes, et presque exclusivement le mécanisme 1 pour les hommes. Il existe une zone de fréquences et d'intensités vocales appelée « zone de passage » sur laquelle le locuteur peut choisir d'utiliser le mécanisme 1 ou le mécanisme 2. Cette zone se situe dans le milieu de la tessiture pour les hommes et plutôt dans le cas de la tessiture pour les femmes (Roubeau et al. 2004^[292]).

Dans la parole, la hauteur de la voix varie au cours de l'énoncé. Cette variation, aussi appelée intonation, n'est pas réalisée au hasard et ne constitue pas uniquement un aspect de la qualité vocale. Elle possède une fonction phonologique propre de structuration du discours. Ainsi, une montée mélodique en fin de phrase permet de marquer l'interrogation tandis qu'une chute mélodique indique une affirmation (Delattre 1966^[67]). Les ruptures de pente ou les proéminences de la courbe mélodiques sont également des indices de frontières entre des constituants de l'énoncé : l'accent primaire s'accompagnant entre autres d'une montée mélodique permet ainsi de marquer la fin d'un syntagme accentuel (Jun et al. 1995^[170]). Il faut également mentionner que dans les langues à tons, ou les langues sifflées, la hauteur avec laquelle certaines syllabes sont réalisées permet de discriminer des mots aux sens différents (Rialland 2005^[285] ; Rialland 2003^[284]).

Mesure

La fréquence fondamentale (F0) correspond au premier harmonique d'un signal périodique. Elle peut être uniquement calculée sur les segments voisés de parole. La fréquence fondamentale a été ici calculée par une méthode d'autocorrélation à partir du signal électroglottographique.

Afin de mieux rendre compte de la perception relative de la hauteur, nous avons choisi d'exprimer la F0 en tons. Sachant qu'un doublement de la fréquence fondamentale correspond à une octave, i.e. à 6 tons, la fréquence fondamentale en tons se déduit de sa valeur en Hz par la formule :

$$f_{0_{\text{tons}}} = \frac{6}{\log(2)} \cdot \log\left(\frac{f_{0_{\text{Hz}}}}{f_{\text{ref}}}\right), \text{ où } f_{\text{ref}} \text{ est une fréquence de référence arbitrairement choisie.}$$

Nous avons choisi ici $f_{\text{ref}} = 50$ Hz, de façon à ce que toutes les f_0 en tons soient positives.

2.3. La marge de modulation de la voix

Rappels

La mesure de l'intensité vocale moyenne et de la fréquence fondamentale moyenne permet de suivre le déplacement global moyen de la parole, mais n'apporte pas d'information sur la capacité de modulation de la voix. La modulation de la voix correspond aux variations de la parole autour de ces valeurs moyennes en fréquence et en intensité. La marge de modulation de la voix peut être assez intuitivement décrite par l'écart-type des distributions de fréquence fondamentale et d'intensité vocale, ainsi que par l'ambitus de F0 et la dynamique d'intensité, correspondant à la différence entre le maximum et le minimum de ces distributions. Cependant, la marge de modulation en intensité n'est pas la même pour toutes les fréquences de la voix, et à l'inverse la marge de modulation en fréquence est différente selon l'intensité vocale. Le désir de représenter la marge de modulation de la voix d'un sujet à la fois en fréquence et en intensité a conduit les chercheurs à introduire la notion de phonétogramme (cf. Figure 133), aussi appelé « Voice Range Profile » (VRP) dans la littérature anglophone (Schutte et al. 1983^[306] ; Sulter et al. 1994^[334] ; Gramming et al. 1988^[119] ; Titze 1992^[351]). Le phonétogramme est utilisé par le milieu médical pour diagnostiquer des pathologies ou dysfonctionnements vocaux. Les professeurs de chant ou les orthophonistes s'en servent également

afin de suivre les progrès qu'apportent une rééducation vocale ou l'apprentissage de la technique vocale à un sujet (Pabon et al. 1988 ^[263]). Le phonétogramme a également été utilisé pour mettre en évidence l'étendue vocale couverte par les mécanismes laryngés, ainsi que leur plage de recouvrement (Roubeau et al. 2004 ^[292]). L'aire du phonétogramme est alors un bon descripteur de l'étendue vocale en fréquence et en intensité.

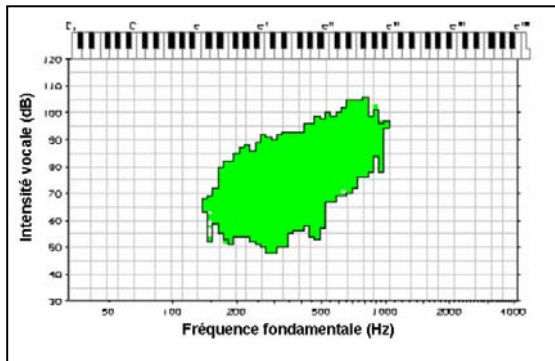


Figure 133. Représentation du contour du phonétogramme d'un chanteur, obtenue grâce au logiciel Voice Profiler (Pabon et al. 1988 ^[263])

Mesures

Dans le cas d'un phonétogramme en voix parlée, toutes les fréquences et intensités situées à l'intérieur du contour du phonétogramme ne sont pas produites avec la même probabilité, et le contour risque alors de s'élargir lorsque le temps sur lequel le phonétogramme est calculé augmente, puisque certaines valeurs très ponctuelles, voire aberrantes seront prises en compte. Il est donc plutôt intéressant de savoir quelles sont les zones du domaine F0/intensité vocale où la parole est produite avec le plus de probabilité, s'il s'agit d'ailleurs d'une ou de plusieurs zones, et si ces zones sont très localisées ou plus diffuses. C'est pourquoi nous avons introduit la notion de « cible vocale », correspondant au maximum de ces zones de grande densité de probabilité à la fois en fréquence et en intensité.

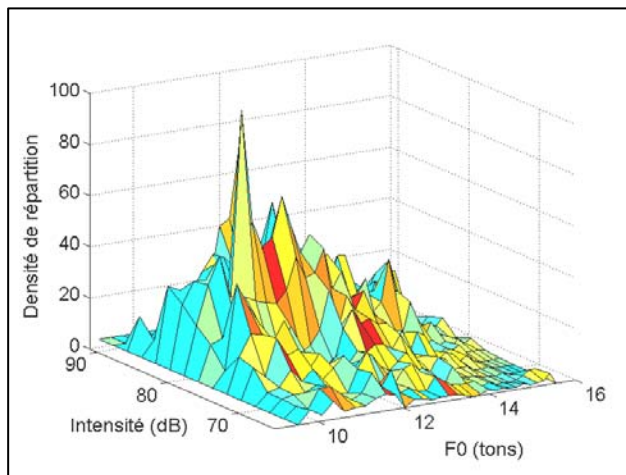


Figure 134 : Distribution des valeurs de F0 et d'intensité vocale d'une locutrice sur 1.5 mn de parole semi-spontanée.

La Figure 134 représente la distribution 3D des valeurs de F0 et d'intensité vocale mesurées à partir d'1min30 de signal audio de parole. La détermination de la fréquence et de l'intensité de la « cible vocale » est effectuée en repérant le maximum de cette distribution (cf. Figure 135). Les distributions 2D des valeurs de F0 et d'intensité sont également calculées à partir de cette distribution 3D (cf. courbes en bas et à gauche de la Figure 135). La courbe de niveau à 50% de cette distribution 3D permet de repérer la zone cible de F0 et d'intensité dans laquelle se situe la parole du locuteur la plupart du temps (cf. courbe en pointillés de la Figure 135). Une zone de faible superficie traduit un manque de modulation de la voix. Enfin, l'étendue vocale est estimée à partir de la courbe de niveau à 10% de cette distribution (cf. courbe en trait plein de la Figure 135). Ce seuil permet en effet d'éliminer les valeurs marginales de F0 et d'intensité. Enfin, cette étendue vocale permet de définir l'ambitus de F0, correspondant à la différence entre les valeurs maximales et minimales de F0 de ce contour, ainsi que la dynamique d'intensité, correspondant à la différence entre les valeurs maximales et minimales d'intensité de ce contour.

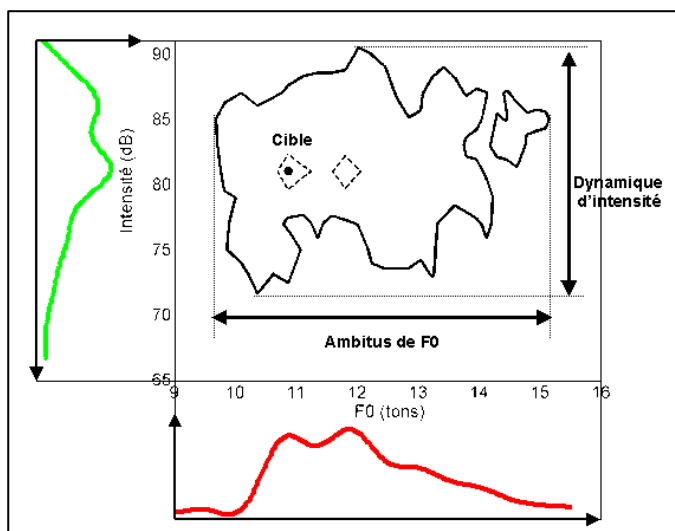


Figure 135. A partir de la distribution 3D représentée Figure 134, on définit l'étendue vocale, par la courbe niveau à 10% (en trait plein), la zone « cible » de la parole par la courbe de niveau à 50%, et la cible elle-même, par le maximum de cette distribution. On définit également les distributions des valeurs de F0 et d'intensité, respectivement en bas et à gauche du phonétogramme. Enfin, on définit à partir de l'étendue vocale l'ambitus de F0 ainsi que la dynamique d'intensité.

3. LE TIMBRE DE LA VOIX

Au-delà de sa hauteur et de son intensité, la voix peut présenter des timbres très variables d'une personne à l'autre ou pour une même personne dans des situations différentes. La qualité vocale est liée à la fois aux caractéristiques vibratoires de la source glottique, mais également aux aspects résonantiels du conduit vocal, ainsi qu'à la façon dont le son est rayonné par le pavillon labial (cf. Figure 136). Nous allons rappeler brièvement le principe de ces différentes contributions au timbre de la voix, et définir les différents descripteurs physiques qui permettent d'en rendre compte.

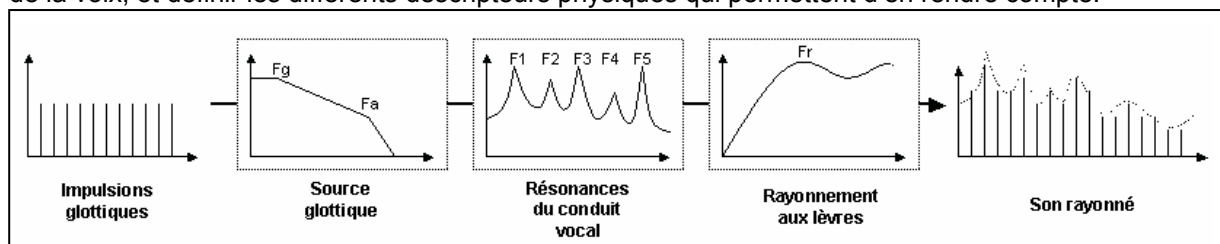


Figure 136. Schématisation des différentes contributions de la source glottiques, du conduit vocal, et du pavillon labial, au contenu spectral du son rayonné.

3.1. La source glottique

3.1.1. Rappels sur les caractéristiques de la source glottique

Quelques rappels sur la source glottiques sont ici nécessaires pour comprendre son effet sur la qualité vocale.

Comme nous l'avons rappelé dans le 1^{er} paragraphe, les cordes vocales se mettent à vibrer de façon passive lorsqu'elles sont suffisamment proches l'une de l'autre (grâce à l'action des aryténoïdes) et qu'un débit d'air est expiré avec une pression sous-glottique suffisante. La vibration des cordes vocales consiste alors en une succession d'accolements/décollements. Ces cycles réguliers d'ouverture et de fermeture de la glotte vont moduler le débit d'air qui la traverse avec une rapidité directement liée à la hauteur du son produit. La façon particulière avec laquelle les cordes vocales vibrent va moduler le débit d'air expiré de façon différente et en cela avoir des conséquences sur la qualité vocale. Les mécanismes laryngés (Henrich 2001 ^[139]) ou la compression des aryténoïdes (Hess et al. 2000 ^[144]) semblent en particulier jouer un rôle sur le contrôle de ces caractéristiques de vibration des cordes vocales.

Ainsi, il est classique de décrire la vibration des cordes vocales à partir de l'onde de débit glottique (ODG), correspondant aux variations en fonction du temps du débit d'air traversant la glotte. L'ODG est nulle ou constante lorsque la glotte est fermée (il peut en effet exister des fuites

glottiques). Ensuite, l'instant d'ouverture se traduit alors par une augmentation progressive du débit, du fait que l'ouverture/fermeture de la glotte n'est pas binaire mais graduelle. De même, les cordes vocales n'entrent pas en contact sur toute leur surface au moment où la glotte se referme, mais de façon progressive. L'instant de fermeture correspondant au moment où les cordes vocales rentrent en contact : l'excitation glottique est alors maximum ce qui se traduit par un pic minimum de la dérivée du débit glottique (ODGD, pour « onde de débit glottique dérivée »). Le temps écoulé entre l'instant de fermeture et la fermeture elle-même est appelé durée effective de la phase de retour (T_a). Celle-ci est donc d'autant plus faible que la fermeture des cordes vocales est abrupte. A partir de ces instants, on peut définir la phase ouverte de la glotte (T_e), correspondant à la durée entre l'instant de fermeture et l'instant d'ouverture, ainsi que la phase fermée de la glotte (T_p), i.e. le complément de temps sur une période de vibration des cordes vocales (T_0). Plusieurs paramètres existent pour décrire la forme des impulsions de la source glottique (cf. Henrich 2001^[139] pour une revue détaillée). En particulier :

- Le quotient ouvert (O_q , pour « open quotient ») se définit de façon simple comme le rapport entre la durée de la phase ouverte sur la période fondamentale ($O_q = T_e/T_0$). Il varie en pratique entre 0.3 et 0.98 (Henrich 2001^[139]). On trouve également dans la littérature la notion de quotient fermé (C_q pour « closed quotient ») qui en est l'équivalent et se déduit simplement par la relation $C_q = 1 - O_q$. Ce paramètre varie avec le temps d'accolement des cordes vocales et est par conséquent relié à l'effort et à la tension vocale (Huang et al. 1995^[160] ; Hillman et al. 1989^[146]).
- L'amplitude de voisement (A_v) correspond à la différence entre le maximum de l'onde de débit glottique et sa valeur constante ou nulle lors de la fermeture des cordes vocales. Elle dépend de la pression sous-glottique et de l'adduction des cordes vocales (Isshiki 1964^[163]) et est corrélée très fortement à l'intensité vocale.
- La vitesse de fermeture de la glotte correspond à l'amplitude du pic négatif du signal DODG à l'instant de fermeture et rend compte de la vitesse du débit glottique au moment de la fermeture des cordes vocales. Ce paramètre est désigné par l'abréviation E dans certaines études (Henrich 2001^[139] ; Doval..) et par d_{peak} dans d'autres (Vintturi 2001^[368], Alku et al. 1996^[5], Holmberg et al. 1988^[153]).
- Le quotient d'amplitude (A_q , pour « amplitude quotient ») se déduit des deux derniers paramètres puisqu'il correspond au rapport de l'amplitude de voisement sur la vitesse de fermeture. Le paramètre R_d , défini par Fant 1995^[89] tient compte en plus de la fréquence de vibration et se définit comme le rapport du quotient d'amplitude sur la période fondamentale multipliée par un facteur 110.

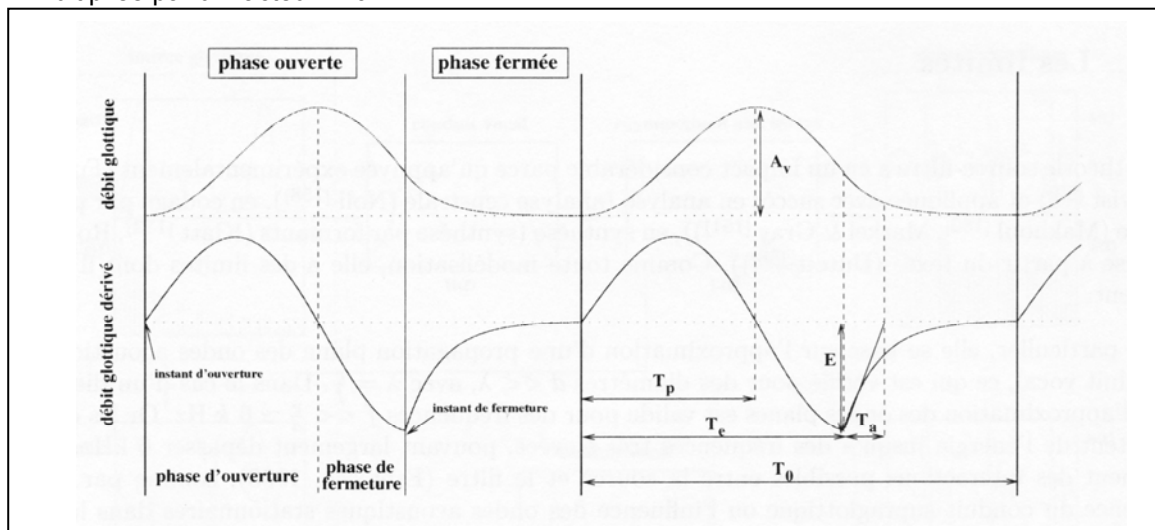


Figure 137. Rappel des différentes phases du cycle de vibration des cordes vocales définies à partir de l'onde de débit glottique (ODG) : phase ouverte de la glotte (T_e), phase fermée (T_p), phase de retour (T_a). Les différents paramètres glottiques examinés dans cette étude sont également indiqués : la période fondamentale (T_0), le quotient ouvert (O_q), l'amplitude de voisement (A_v) et la vitesse de fermeture (E). D'après Henrich 2001^[139].

3.1.2. Méthodes d'exploration de la source glottique

Nous avons décrit dans le paragraphe précédent les principales caractéristiques de la source glottique, par référence à l'onde de débit glottique. Reste maintenant à savoir comment mesurer

cette dernière. Une première méthode assez précise consiste à mesurer le débit d'air au niveau de la bouche et du nez en plaçant un masque, dit de Rothenberg devant le visage du locuteur. Ce dispositif ne permet cependant pas d'effectuer une acquisition simultanée du signal acoustique puisqu'il filtre considérablement la voix. Evidemment, il est encore moins envisageable de réaliser des mesures des mouvements labiaux en parallèle. C'est pourquoi nous avons choisi une autre méthode un peu moins précise, mais cependant couramment utilisée, qui consiste, par filtrage inverse, à estimer l'onde de débit glottique à partir du signal rayonné acquis à l'aide d'un microphone.

Le filtrage inverse est basé sur la théorie source-filtre (Fant 1960 ^[88]) qui conçoit le son rayonné aux lèvres comme la résultante du filtrage du son glottique par le conduit vocal et par l'effet de rayonnement aux lèvres, et s'appuie sur une hypothèse forte d'indépendance entre la source (glottique) et le filtre (conduit vocal). Faisant cette hypothèse, on peut alors modéliser l'effet de la source et du filtre comme un système de filtrage par blocs (cf. Figure 136). Une fois que les résonances du conduit vocal ont été estimées (cf. paragraphe 4.1), et en faisant l'hypothèse que le rayonnement aux lèvres peut être modélisé par un filtre passe-haut de pente +6dB/octave, le spectre du son rayonné peut alors s'exprimer comme la multiplication du spectre glottique par une fonction de transfert $H(f)$ prenant en compte les effets de filtrage par le conduit vocal et le rayonnement aux lèvres. Il est alors possible d'estimer l'onde de débit glottique comme la convolution du son rayonné par la réponse impulsionnelle inverse $h^{-1}(t)$.

L'amplitude de voisement et la vitesse de fermeture sont mesurées à partir du signal ODG de façon assez fiable. Par contre, le quotient ouvert est plus délicat à mesurer directement sur l'ODG du fait que l'instant d'ouverture est difficilement repérable sur un signal réel légèrement bruité.

L'usage de l'électroglottographie s'avère au contraire très intéressante pour la mesure du quotient ouvert (Henrich 2001 ^[139]). Il s'agit d'une méthode d'exploration du contact entre les cordes vocales. Elle n'est donc pas équivalente mais complémentaire à l'observation du débit glottique : le signal ODG renseigne sur ce qui se passe pendant la phase ouverte des cordes vocales, tandis que le signal électroglottographique (EGG) renseigne davantage sur ce qui se passe pendant la phase fermée des cordes vocales.

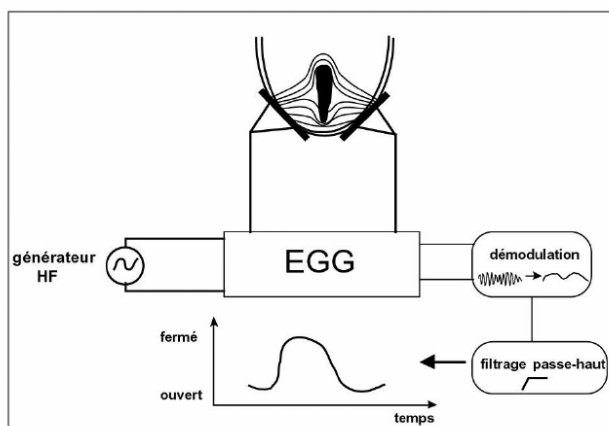


Figure 138 : Principe de fonctionnement d'un électroglottographe. D'après Henrich 2001 ^[139].

L'électroglottographie est une méthode non-invasive très simple d'utilisation, pouvant être utilisée en parallèle d'un enregistrement du signal acoustique. L'acquisition du signal EGG s'effectue à l'aide de deux électrodes que l'on place sur le cou du locuteur, au niveau du cartilage thyroïdien, et reliées à un générateur de courant constant de très faible intensité (cf. Figure 138). La mesure de la différence de potentiel aux bornes des deux électrodes revient à déterminer la résistance que le cou oppose au courant. Sachant que les tissus conduisent davantage le courant que l'air, le signal EGG est alors plus important lorsque les cordes vocales sont en contact que lorsqu'elles sont ouvertes (cf. Figure 139). Les phases ouvertes et fermées des cordes vocales se repèrent au niveau de la dérivée du signal EGG. L'instant de fermeture des cordes vocales n'est toujours pas problématique à estimer car il correspond à un pic positif bien marqué et d'amplitude importante sur le signal DEGG. Par contre, l'instant d'ouverture des cordes vocales est moins évident à repérer. Il correspond à un pic négatif sur le signal DEGG, mais n'est malheureusement pas toujours très marqué. Nous avons donc choisi ici d'estimer l'instant d'ouverture par une méthode de seuil à 3:7, proposée par Howard 1995 ^[158]. Celle-ci consiste à détecter sur chaque période de vibration des cordes vocales le maximum et le minimum de la forme d'onde du signal EGG, et à

déterminer l'instant d'ouverture comme le moment où la valeur du signal EGG est égale à $\text{Min} + \frac{3}{7} * (\text{Max} - \text{Min})$.

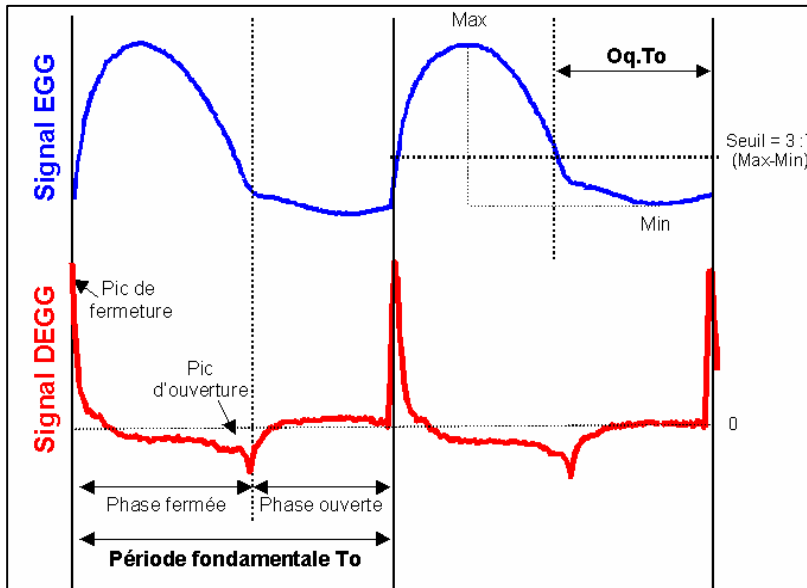


Figure 139 : Représentation de la forme du signal électroglottographique (EGG) et de sa dérivée (DEGG) en fonction des différentes phases du cycle de vibration des cordes vocales.

3.1.2. Descripteurs glottiques de l'effort vocal

Les caractéristiques vibratoires des cordes vocales ont des conséquences sur la forme d'onde du son glottique et par conséquent sur son contenu spectral. L'influence spectrale de la source glottique se modélise mathématiquement sous la forme d'un filtre passe bas du 2^{ème} ordre de pente -12dB/octave et de fréquence de coupure F_g . Celle-ci est appelée « formant glottique » bien qu'elle n'ait pas d'origine résonnante, et varie en fonction des caractéristiques de vibration des cordes vocales d'une valeur proche du fondamental dans le cas de phonations relâchées ou soufflées à quelques multiples de la F_0 dans le cas de phonations pressées. Ce déplacement du formant glottique a pour conséquence un enrichissement spectral plus ou moins important du son glottique en hautes fréquences. Par ailleurs, dans le cas où la fermeture des cordes vocales n'est pas abrupte, nous avons vu que l'on peut alors définir au niveau de l'onde de débit glottique une phase de retour T_a . Celle-ci se traduit spectralement par un deuxième filtrage passe bas dans les hautes fréquences, à une fréquence de coupure F_a d'autant plus haute que la fermeture des cordes vocales est abrupte (cf. Figure 136). Le son glottique est alors d'autant plus riche en harmoniques aiguës que la fermeture des cordes vocales est brutale.

La notion de pente spectrale cherche à rendre compte directement de l'atténuation plus ou moins importante du spectre glottique en fonction des fréquences croissantes. Celle-ci est globalement très marquée pour les voix soufflées et au contraire peu marquée lorsque la tension vocale est importante (Childers et al. 1991^[50], Gauffin et al. 1989^[109], Novak et al. 1995^[258]). Bien que la pente spectrale soit sensée rendre compte de l'influence de la source glottique sur la qualité vocale, la plupart des méthodes proposent de l'estimer à partir du spectre rayonné : Hammarberg a par exemple introduit un index, le Hamml, correspondant à la différence entre le maximum d'énergie dans la zone 0-2kHz et celui dans la zone 2-5kHz. D'autres études se basent sur le calcul de l'énergie par bandes de fréquences (sur 33 bandes en dessous de 3kHz chez Castellanos et al. 1996^[46]), pour ensuite estimer la pente spectrale « basses fréquences » par régression logarithmique de la décroissance de ces différentes énergies. De la même manière, une autre pente spectrale « hautes-fréquences » est définie entre 3 et 8 kHz (Castellanos et al. 1996^[46]) pour rendre compte du caractère abrupt ou non de la fermeture glottique. Dans ce travail, nous avons repris ces définitions et méthodes de mesure des pentes glottiques BF et HF.

La vitesse de fermeture E , le quotient ouvert Oq et l'amplitude de voisement Av permettent également de caractériser la contribution de la source glottique à l'enrichissement du spectre de la voix en harmoniques aiguës. Comme nous l'avons vu précédemment, ces paramètres varient en fonction du mode de vibration des cordes vocales et sont par conséquent reliés à un enrichissement spectral plus ou moins important du son glottique en hautes fréquences (Henrich 2001^[139]). Une vitesse de fermeture élevée, un quotient ouvert faible ou une amplitude de

voisement importante correspondent à un déplacement du formant glottique vers les fréquences aiguës et par conséquent à un son glottique plus riche en harmoniques aigus. Dans ce travail, nous estimerons donc la vitesse de fermeture des cordes vocales E et l'amplitude de voisement Av à partir de l'ODG obtenue par filtrage inverse, et le quotient ouvert à partir du signal électroglottographique. Nous les envisagerons comme des indicateurs potentiels d'une phonation pressée ou relâchée.

Dans les deux bases de données, le signal audio a été normalisé par rapport à la calibration d'intensité effectuée pour chaque locuteur. En revanche, nous ne connaissons pas la transduction Pression→Volts des enregistrements. C'est pourquoi l'amplitude de voisement Av et la vitesse de fermeture E sont exprimées en unité arbitraire (u.a.) dans ce travail. Grâce à la calibration en intensité, nous pouvons néanmoins comparer les valeurs des différents locuteurs au sein de chaque base de données. Nous ne pouvons en revanche pas comparer les valeurs entre les deux bases de données.

3.1.3. Descripteurs de l'apériodicité de la vibration laryngée

Un autre aspect de la qualité vocale réside dans l'apériodicité de la vibration des cordes vocales en fréquence (Jitter) ou en intensité (Shimmer), tout à fait normale et contribuant au naturel de la voix lorsqu'elle reste faible, mais devenant une caractéristique pathologique associée à la perception d'instabilité ou de raucité dans la voix lorsqu'elle dépasse un certain seuil (Dejonckere et al. 1996 [66]).

Le Jitter est défini initialement par Liebermann 1961 [214] comme un pourcentage de perturbation de la période fondamentale. Ce paramètre est associé à la perception d'instabilité et de raucité de la voix (Coleman 1971 [57]), et est beaucoup utilisé dans les études concernant la dysphonie ou la caractérisation des voix pathologiques (Teston 2001 [348]). Le Jitter moyen est défini comme la moyenne de la variation de F0 entre deux cycles de vibration consécutifs, et s'exprime en Hz.

$$Jitter\ moyen(Hz) = \frac{1}{N-1} \sum_{i=1}^{N-1} |F0_i - F0_{i+1}|$$

Le Jitter moyen est intéressant, mais dépend beaucoup de la hauteur de la voix. Cela est gênant pour nous puisque nous souhaitons comparer des voix d'hommes et de femmes. C'est pourquoi dans cette étude, nous avons choisi de mesurer plutôt le Jitter factor, défini comme la normalisation du Jitter moyen sur la F0 moyenne.

$$Jitter\ factor(\%) = 100 \cdot \frac{Jitter\ moyen}{F0\ moyenne}$$

L'instabilité à court terme de l'amplitude de la vibration se traduit par des variations d'amplitude entre chaque cycle d'oscillation. Le Shimmer moyen se définit ainsi comme la moyenne des rapports d'amplitudes entre deux périodes de vibration consécutives. Il s'exprime alors en dB.

$$shimmer_{dB} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \cdot \log\left(\frac{A_i}{A_{i+1}}\right) \right|, \text{ où } A_i \text{ représente l'amplitude de la vibration au cycle } i.$$

Nous avons ici choisi de calculer plutôt le Shimmer factor, qui consiste à normaliser le Shimmer moyen par l'amplitude moyenne de vibration. Ce paramètre est alors exprimé en %.

$$shimmer\ factor = 100 \cdot \frac{shimmer}{20 \cdot \log(A_{moyen})}$$

L'instabilité de la vibration laryngée se manifeste comme un bruit qui lui est superposé. Le HNR (harmonic noise ratio) proposée par Yumoto et al. 1982 [381], le SNR (signal to noise ratio) de Laukkanen et al. 1995 [198] et la NNE (normalized noise energy) proposée par Kasuya et al. 1986 [177] sont trois méthodes permettant de quantifier le rapport entre la composante harmonique et bruitée du signal vocal.

3.1.3. Descripteurs de l'aspect soufflé d'une voix

La source glottique peut également avoir une influence sur la qualité vocale dans le cas où la fermeture glottique est incomplète et que les « fuites d'air » se traduisent par l'apparition de composantes inharmoniques dans le spectre (Yanagihara 1967 [379]). Une phonation soufflée se

caractérise également par la différence H1-H2 (en dB) entre les deux premiers harmoniques du spectre de l'ODGD. En effet, puisque le spectre du signal ODG est modélisé par un passe bas de pente -12dB/octave à partir du formant glottique F_g , celui du signal ODGD se présente donc sous la forme d'une asymptote de pente $+6\text{dB/octave}$ jusqu'à F_g , et de pente -6dB/octave au-dessus de F_g . Dans le cas d'une phonation relâchée, le formant glottique est proche du fondamental, tandis qu'il se déplace vers les fréquences aiguës dans le cas d'une phonation pressée (Gauffin et al. 1989^[109]; Klatt et al. 1990^[183], Henrich 2001^[139]). Dans le premier cas, les deux premiers harmoniques sont d'amplitude semblable, voire décroissante ($H1 > H2$). C'est le cas en particulier des phonations soufflées où le fondamental prédomine (Hillenbrand et al. 1994^[145]). Au contraire, le déplacement du formant glottique vers les aigus entraîne une diminution de H1 par rapport aux harmoniques de rang supérieur (en particulier H2). Une forte corrélation a été montrée entre le quotient ouvert et cette mesure de H1-H2 (Holmberg et al. 1995^[154], Henrich 2001^[139]; Hanson 1997^[135]) mais Henrich 2001^[139] précise que H1-H2 dépend également des autres paramètres de source et du mécanisme laryngé.

3.2. Le conduit vocal et le rayonnement aux lèvres

La qualité vocale ne dépend pas seulement de la source glottique mais également de la façon dont le conduit vocal et le rayonnement aux lèvres filtrent le son glottique. Les différentes cavités pharyngée, buccale, nasale (cf. Figure 140) peuvent être assimilées à des résonateurs de Helmholtz qui renforcent l'énergie spectrale du son glottique au niveau de certaines zones fréquentielles. Ces renforcements spectraux sont aussi appelés formants, et permettent d'estimer dans une certaine mesure les résonances physiques du conduit vocal. Le conduit vocal présente 5 principaux formants en dessous de 5kHz, dont les fréquences centrales sont liées aux formes des différentes cavités et peuvent être modifiées par l'ouverture de la mâchoire, la forme des lèvres, la position de la langue, le soulèvement ou l'abaissement du voile du palais et l'élévation ou l'abaissement du larynx. La position fréquentielle de ces formants et leur amplitude va déterminer la répartition en fréquences de l'énergie acoustique de la voix, et en cela modifier le timbre de la voix.

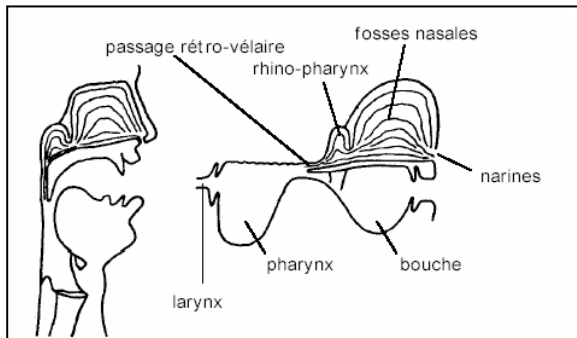


Figure 140. Représentation en coupe et redressée du conduit vocal, avec les différentes cavités pharyngée, buccale, nasale. D'après Husson 1962^[162].

A cet effet du conduit vocal se rajoute, l'effet de rayonnement du son au niveau du pavillon labial, pouvant être assimilé à celui d'un piston circulaire dans un champ infini. L'impédance acoustique de rayonnement correspondante est représentée Figure 141. Lorsque $ka \ll 1$ (où k est le nombre d'onde, égal à $2\pi f/c$, avec c , la vitesse de propagation du son dans l'air et f , la fréquence du son), le rayonnement aux lèvres peut être approximé par un filtre passe haut du premier ordre, de pente $+6\text{dB/octave}$ (Flanagan 1972^[93]). C'est ce qui est classiquement effectué en parole, où le rayon a de l'aire intero-labiale n'excède en général pas 5mm. La condition $ka \ll 1$ est alors satisfaite pour toutes les fréquences inférieures à 10kHz. Cette approximation doit être reconsidérer dans le cas de la parole criée où l'ouverture de la bouche est plus importante.

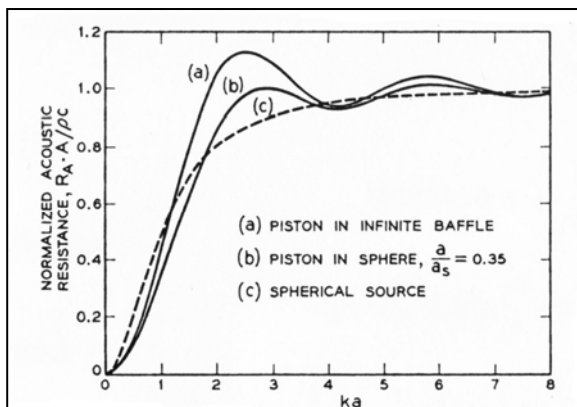


Figure 141. Impédance de rayonnement (a) d'un piston circulaire dans un champ infini, (b) dans un champ sphérique et (c) d'une source sphérique. Extrait de Flanagan 1972^[93].

3.2.1. Emergence et portée de la voix

Différentes études se sont intéressées au timbre particulier de la voix des chanteurs lyriques ainsi qu'à la voix projetée des acteurs, en cherchant à comprendre quelles sont les caractéristiques spectrales de ces qualités vocales qui leur permettent d'émerger par rapport un orchestre et d'avoir une voix « qui porte » sans l'aide de sonorisation.

Il a en particulier été remarqué chez les chanteurs un renforcement important du spectre autour de 3kHz, c'est à dire dans la zone sensible de l'oreille, aussi appelé « formant du chanteur » (Sundberg 1972^[337] ; Castellengo 1973^[47], Bloothoof et al. 1986^[26]). Plusieurs interprétations ont été proposées quant à son origine et son contrôle : Sundberg 1972^[337] et Gauffin et al. 1989^[109] proposent que le chanteur crée ce renforcement de façon articuloire en rapprochant les 3^{ème} et 4^{ème} formants vocaliques (voire le 5^{ème} aussi). L'amplitude des formants dépend en effet de leur proximité. Lorsque leur fréquence centrale est très rapprochée, ils peuvent fusionner en un seul « pic » spectral d'amplitude conséquente. Ce renforcement pourrait également provenir d'une résonance des cavités pharyngées et hypopharyngée pouvant être dilatées par l'abaissement du larynx (Sundberg 1972^[337] ; Dang et al. 1997^[61]).

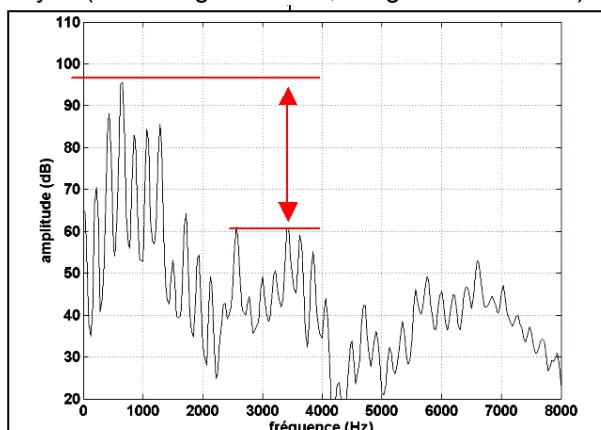


Figure 142. Estimation du « Singing formant », comme la différence entre le maximum du spectre dans la zone [50-1800] et le maximum dans la zone [1800-3500]. (extrait de Garnier et al. 2005^[106])

Plusieurs méthodes de mesure ont été proposées du formant du chanteur. En tout premier lieu, certaines études calculent le spectre moyenné sur la voyelle produite et mesurent la différence entre le maximum du spectre dans la zone [0-2000Hz] et le maximum dans la zone [2000-4000Hz] (cf. Figure 142). Cette mesure présente l'avantage de ne pas dépendre de l'intensité globale de la voix puisqu'il s'agit d'une différence en dB. Une autre possibilité consiste à estimer le pourcentage d'énergie acoustique dans la bande de fréquence 2-4kHz par rapport à l'énergie acoustique totale audible (inférieure à 20kHz). Enfin, une solution intermédiaire consiste à calculer la différence (en dB) d'énergie acoustique globale (et non pas maximale) entre la zone du fondamental (0-2kHz) et la zone sensible de l'oreille (2-4kHz).

Plusieurs auteurs ont également cherché à caractériser un formant du locuteur (« speaker's formant » ou « speaker's ring ») chez les acteurs de théâtre ou chez certains chanteurs en voix parlée (Bele 2006^[23] ; Cleveland et al. 2001^[56] ; Oliveira Barrichelo et al. 2001^[260] ; Nawka et al. 1997^[250], Thunberg 2003^[349]). Leino 1993^[207] observe toujours chez les acteurs un renforcement énergétique comparable au singing formant mais d'amplitude plus faible et décalé d'environ 1kHz vers les fréquences aiguës. Il rassemblerait alors plutôt F4 et F5 selon lui. Bele 2006^[23] par contre observe surtout un abaissement de F4 chez les acteurs professionnels, ayant pour conséquence le rapprochement de F3 et de F4 en un même cluster d'amplitude spectrale importante. Nawka et al. 1997^[250] enfin constate chez les acteurs sur scène une diminution de l'énergie dans la région de F2, ce qui tend à accentuer l'acuité du pic d'énergie autour de F1, et une augmentation de l'énergie dans la zone 2000-2500Hz.

3.2.2. Timbre vocalique et extravocalique

La répartition d'énergie en fréquences participe également à d'autres aspects de la qualité vocale que l'émergence ou la portée de la voix. En particulier, Husson 1962^[162] distingue le timbre vocalique, en dessous de 3kHz environ, du timbre extra-vocalique. Le timbre vocalique comprend les 2 ou 3 premiers formants, déterminants pour l'intelligibilité des voyelles. Cependant, ce timbre vocalique peut également contribuer à la perception d'un placement de voix en avant ou en arrière, baillé, ouvert ou couvert, d'un timbre clair ou sombre (Garnier et al. 2005^[106]). Au contraire, le

caractère « brillant » ou « sourd », « timbré » ou « détimbré » de la voix est davantage relié à l'énergie acoustique dans la zone sensible de l'oreille par comparaison à l'énergie acoustique des premiers harmoniques de la voix. Enfin, la « dureté », la « stridence » d'une voix serait, d'après Husson 1962 ^[162], déterminée par l'énergie acoustique au-dessus de 4kHz, toujours par comparaison à l'énergie des premiers harmoniques de la voix.

C'est pourquoi dans ce travail nous chercherons à décrire les modifications de répartition spectrale de l'énergie de la voix en mesurant l'intensité de chaque segment (voyelles, fricative, etc.) dans les bandes de fréquences [1-2kHz], [2-4kHz] et [4-6kHz]. Pour cela, nous filtrons le signal audio du segment considéré par un filtre passe bande (Butterworth, 3^{ème} ordre, de fréquences de coupure $0.95 \cdot f_1$ et $1.05 \cdot f_2$, où $f_1=1\text{kHz}$ et $f_2=2\text{kHz}$ par exemple dans le cas de la bande [1-2kHz]). Nous avons ramené l'intensité de la voix dans chacune de ces bandes de fréquences à l'intensité de la voix dans la zone du fondamental [0-1kHz].

4. LE DOUBLE ROLE DE L'ARTICULATION

Comme nous l'avons vu précédemment, le conduit vocal joue un rôle sur la qualité vocale par la façon dont il renforce l'énergie acoustique du son glottique au niveau des fréquences de résonance des différentes cavités (pharyngée, buccale, nasale, etc.) qui le composent et dont la forme peut être contrôlée à l'aide des articulateurs de la parole (langue, voile du palais, mandibule, lèvres, larynx). Hors ces renforcements spectraux, de même que les mouvements articulatoires qui leur sont associés, ne contribuent pas seulement au timbre de la voix et à la répartition globale de l'énergie spectrale, mais sont des indices de codage et de reconnaissance des voyelles. En cela, l'articulation et ses conséquences formantiques joue un double rôle dans la production de parole de contrôle du timbre de la voix ainsi que de l'intelligibilité de la parole.

4.1. Les indices formantiques de l'articulation

Rappel

Les renforcements énergétiques du spectre rayonné, ou formants, donnent une approximation des résonances du conduit vocal. Leur fréquence centrale est donc liée aux mouvements articulatoires, même si cette relation n'est pas linéaire (Perrier 2005 ^[268]). Les trois premiers formants F1, F2 et F3 jouent un rôle déterminant dans la reconnaissance statique des voyelles, et ont été reliés à plusieurs mouvements articulatoires. Ainsi, l'augmentation de F1 est corrélée à une augmentation de l'ouverture de la mâchoire (Lindblom et al. 1971 ^[220]), tandis que l'augmentation de F2 est reliée à l'antériorisation des voyelles (Calliope 1989 ^[41]). Il est ainsi très classique de représenter l'ensemble des voyelles dans le plan F1/F2 selon un triangle ou un quadrilatère, en attribuant l'axe de F1 à la dimension phonologique d' « ouverture/fermeture » vocalique et l'axe de F2 à la dimension d' « antériorisation / postériorisation » des voyelles (cf. Figure 143).

La diminution du troisième formant serait reliée à l'augmentation de la protrusion (Sundberg 1970 ^[336]) et la diminution de F4, à un abaissement du larynx (Sundberg 1970 ^[336]). Par ailleurs, il a été montré qu'un allongement du pharynx pour les voyelles antérieures (/i/, /e/, /a/) était relié à une diminution de F2 (Fant 1960 ^[88]), tandis qu'une augmentation de F3 était induite par l'avancée de la pointe de la langue (Sundberg 1970 ^[336]).

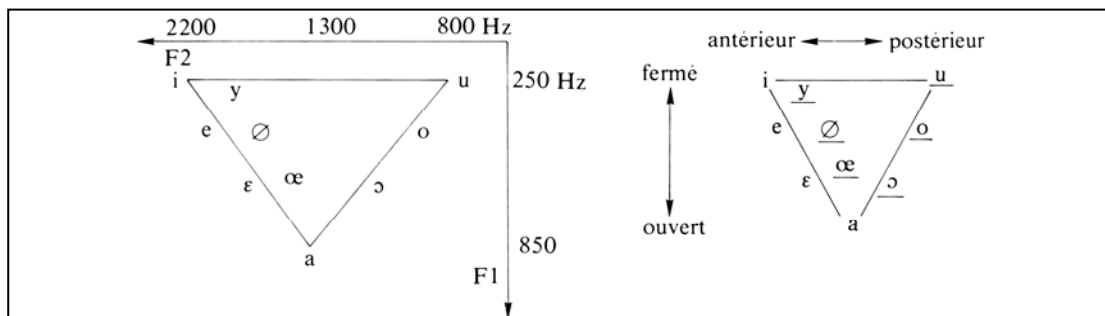


Figure 143. Représentation du système vocalique français dans le plan F1/ F2. D'après Calliope 1989 ^[41]

Mesure

La fréquence centrale, l'amplitude et la largeur de bande des formants vocaliques peuvent être obtenus par filtrage inverse du signal acoustique rayonné. Nous avons déjà évoqué cette méthode au paragraphe 3.12. Elle consiste à considérer le son rayonné comme la sortie d'un système de filtrage par blocs prenant en compte l'influence spectrale de la source glottique, l'influence du conduit vocal et celle du rayonnement aux lèvres (cf. Figure 136). L'estimation des formants revient à estimer la fonction de transfert du conduit vocal à partir du son rayonné. Cela suggère de connaître les caractéristiques des autres blocs du système. Cela n'est malheureusement pas le cas. Ce qui oblige à réaliser des hypothèses fortes, en modélisant l'effet du rayonnement aux lèvres par un filtre passe-haut de pente +6dB/octave et l'effet de la source glottique par un filtre basse bas du deuxième ordre de pente -12dB/octave. Numériquement, cette hypothèse se traduit par la convolution du signal rayonné par un filtre tout-zéros de la forme $F(z) = 1 - az^{-1}$, où a , le paramètre d'accentuation est choisi proche de 1 (ici $a=0.976$).

L'estimation de la fonction de transfert du conduit vocal peut alors être réalisée à l'aide d'une méthode de prédiction linéaire (ou LPC pour Linear Predictive Coding). Celle-ci consiste à modéliser le spectre redressé au cours de l'étape précédente sous forme d'un filtre linéaire causal. Autrement dit, cette méthode suppose que le signal à un instant t peut être prédit à partir d'une combinaison linéaire d'échantillons du signal à des instants précédents. Il existe plusieurs méthodes de prédiction linéaire (cf. Henrich 2001 ^[139] pour une revue de ces méthodes). Nous avons ici utilisé une méthode de prédiction linéaire par autocorrélation de Matlab, en fenêtrant le signal par une fenêtre de Hanning de 20 ms. Le principe très général de cette méthode consiste à déterminer les coefficients $[1 \ c(2) \dots \ c(N+1)]$ d'un estimateur X_p d'ordre N du spectre X ($X_p(n) = -c(2)*X(n-1) - c(3)*X(n-2) - \dots - c(N+1)*X(n-N)$) de façon à ce que la somme quadratique des erreurs $err(n) = X(n) - X_p(n)$ soit minimisée.

Cette estimation formantique n'est cependant pas fiable à 100%. Il arrive par exemple que l'estimation soit bonne mais « décalée » d'un formant. Il arrive également qu'un formant ne soit pas estimé ou avec une valeur aberrante. C'est le cas en particulier sur les voyelles nasales ou sur les [u], posant également des problèmes dans d'autres études (Gendrot 2005 ^[112]). Pour réduire ces erreurs d'estimation, nous avons tout d'abord pensé à appliquer un suivi de trajectoires formantiques, puis à écarter les estimations s'écartant des valeurs typiques de la littérature au-delà d'une certaine marge, comme ce que propose Gendrot 2005 ^[112]. Cependant, cela pose un problème de fond dans notre cas puisque notre objectif dans cette thèse est justement d'examiner comment la parole produite dans le bruit est modifiée par rapport à la parole conversationnelle. Ce serait donc un contre-sens de définir des valeurs formantiques a-priori alors que nous recherchons justement à examiner comment elles évoluent en parole Lombard. Nous avons donc opté pour une autre méthode : pour chaque locuteur et locutrice de chaque base de données, et pour chaque voyelle [a], [i] et [u], nous avons tracé de façon juxtaposée les trajectoires des 4 premiers formants estimés par filtrage inverse (cf. Figure 144), toutes conditions confondues (silence, types de bruit, niveaux de bruit). Nous avons ensuite regroupé manuellement les valeurs correspondant à chacun des 4 premiers formants et par cette sélection manuelle éliminé simultanément certaines valeurs marginales. Cette méthode n'est donc pas parfaite et est plus coûteuse en temps qu'une méthode complètement automatique. C'est pourquoi nous n'avons examiné dans la thèse que les voyelles [a], [i] et [u] de nos bases de données. Cependant cette méthode est celle qui nous a permis de réduire au mieux les erreurs d'estimation tout en minimisant l'a priori vis à vis des résultats à obtenir.

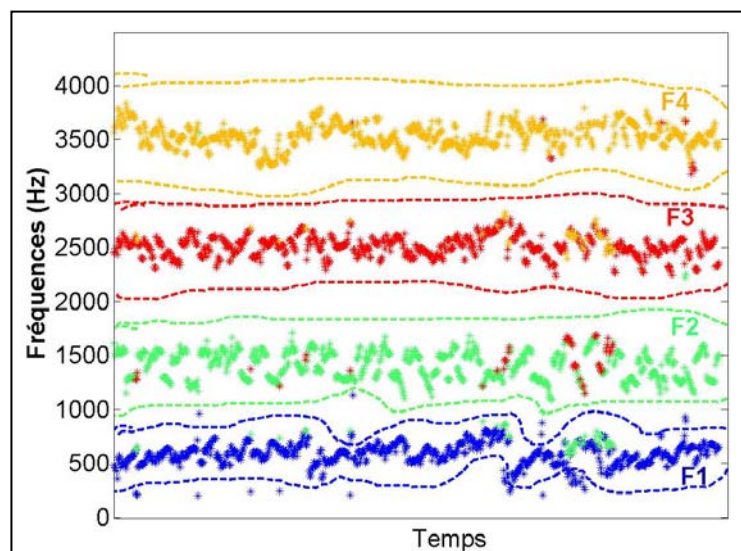


Figure 144. Juxtaposition des trajectoires formantiques estimées par filtrage inverse pour les différents [a] d'un locuteur aussi bien dans le silence que dans le bruit.

Pour la première base de données, nous avons ainsi mesuré les valeurs moyennes des 3 premiers formants sur la durée totale des voyelles [a], [i] et [u]. Dans la deuxième base de données, nous avons mesuré ces valeurs moyennes seulement sur les cibles vocaliques, c'est à dire pour la partie stable de la structure formantique. Nous avons également mesuré la durée et des transitions de chaque voyelle avec les [l] précédant et suivant cette voyelle (cf. Annexe Met3).

4.2. Les indices labiaux de l'articulation

4.2.1. Paramètres considérés

Dans cette étude, nous nous sommes intéressés principalement aux mouvements articulatoires labiaux, décrit à partir de plusieurs paramètres, représentés sur la Figure 145. Tout d'abord l'aire intéro-labiale (S) peut être calculée à partir de l'ouverture (B) et de l'étirement (A), mesurés de face à partir du contour intérieur des lèvres. Une vue de profil du locuteur permet de définir la protrusion de la lèvre supérieure (P1). Une mesure absolue de la protrusion peut être effectuée de profil par référence à une règle fixée sur le profil du sujet (cf. Figure 145). Cependant, cette mesure doit être ramenée à une position neutre des lèvres (P1neutre), c'est à dire au repos, afin de refléter l'amplitude réelle des mouvements. La protrusion peut alors être positive dans le cas des voyelles protrusées, mais aussi négative dans le cas des voyelles très ouvertes ou très étirées.

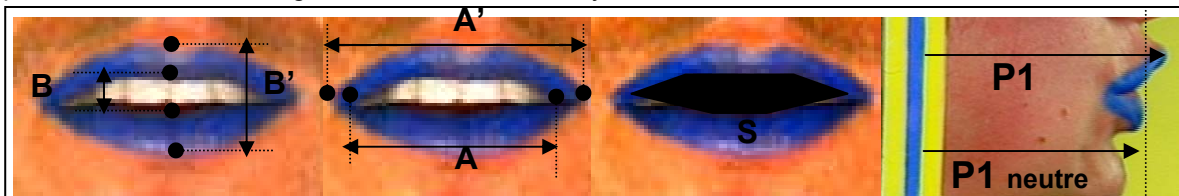


Figure 145. Représentation des paramètres labiaux observables de face et de profil : A et A', l'étirement labial mesuré au niveau du contour intérieur ou extérieur des lèvres, B et B', l'ouverture labiale mesurée au niveau du contour intérieur ou extérieur des lèvres, S, l'aire intéro-labiale et P1, la protrusion de la lèvre supérieure. Inspiré de Calliope 1989^[41].

La mesure de ces différents paramètres peut être exprimée de façon absolue en unité de longueur (mm, cm, pixels). Hors chaque locuteur présente des capacités articulatoires différentes liées à sa morphologie. Puisque nous nous intéressons dans cette étude à l'évolution en relatif des mouvements articulatoires d'une condition silencieuse à une condition bruyante, nous avons choisi de normaliser chaque paramètre articulatoire par rapport au maximum que chaque sujet peut effectuer (cf. Figure 146). Cela nous permet ainsi de savoir comment le sujet exploite l'étendue de ses possibilités articulatoires, et pourquoi éventuellement il ne peut pas augmenter davantage l'amplitude de ses mouvements. Pour résumer, nous avons donc examiné

- le pourcentage d'ouverture labiale par rapport à l'ouverture maximale : B / B_{max} .
- le pourcentage d'étirement labial par rapport à l'étirement maximal = A / A_{max} .
- le pourcentage d'aire intéro-labiale par rapport à l'aire intéro-labiale maximale = S / S_{max} .
- le pourcentage de protrusion labiale = $(P1 - P1_{neutre}) / (P1_{max} - P1_{neutre})$.

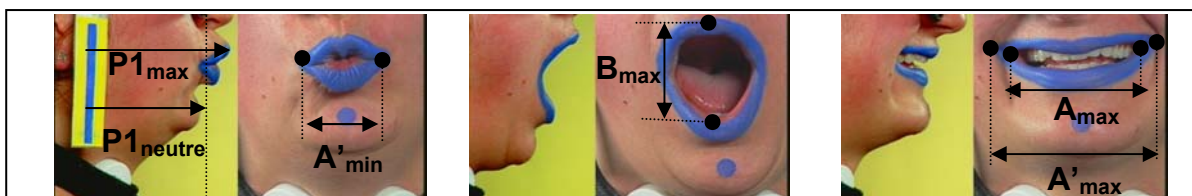


Figure 146. Exemple pour une locutrice des mouvements maximaux de protrusion, d'ouverture et d'étirement qu'elle peut produire et par rapport auxquels nous avons normalisé ses mesures articulatoires.

Nous avons également cherché à estimer l'arrondissement des lèvres sur les voyelles arrondies telles que [u] et [y]. Pour cela, nous avons considéré l'étirement des lèvres mesuré à partir du contour extérieur des lèvres. Celui-ci est en effet minimum lorsque le mouvement d'arrondissement des lèvres est le plus marqué, et au contraire maximum lorsque les lèvres sont les moins arrondies.

Aussi, nous avons introduit le paramètre $AR = (A'_{max} - A') / (A'_{max} - A'_{min})$ correspondant au pourcentage d'arrondissement des lèvres.

Nous avons mesuré chacun de ces paramètres au niveau de la cible articulatoire vocalique de chaque syllabe CV. Cette cible a été détectée automatiquement pour les syllabes contenant un [a], un [e], ou un [i] grâce à un maximum de l'aire intéro-labiale, et pour les syllabes contenant un [y], un [u] ou un [ä] grâce à un maximum de la protrusion de la lèvre supérieure.

Nous avons également mesuré, lorsqu'ils existaient, les pics de vitesse de chacun des ces paramètres (correspondant à un pic de la dérivée des paramètres articulatoires en fonction du temps), pouvant être reliés à la force articulatoire (Nelson 1983 ^[253]).

Nous avons également voulu rendre compte de la compression des lèvres plus ou moins importante lors de leur fermeture sur des consonnes bilabiales. Nous avons observé deux types de pincement différents :

- la plupart du temps les lèvres se compressent en rentrant à l'intérieur de la bouche. La surface visible des lèvres diminue avec le pincement (cf. Figure 147).
- dans quelques cas, les lèvres se compressent au contraire en sortant vers l'extérieur de la bouche, comme pour effectuer un baiser. La surface labiale visible augmente avec le pincement (cf. Figure 147).

Le 1^{er} type de pincement est beaucoup plus fréquent que le second. Nous n'avons pas pu mettre en évidence une utilisation privilégiée de l'un ou l'autre type de pincement selon les locutrices ou selon les syllabes de l'énoncé. Nous avons donc décidé dans cette étude de nous focaliser uniquement sur le premier type de pincement. Lorsque nous parlons de pincement labial, il s'agit donc du premier type. Pour l'objectiver, nous avons considéré la hauteur visible des lèvres, correspondant à la différence entre l'ouverture labiale mesurée à l'extérieur (B') et à l'intérieur (B) du contour des lèvres. Par ailleurs, le pincement n'est défini qu'à partir du moment où les lèvres se ferment, autrement dit, lorsque B est nul. La différence B'-B correspond alors tout simplement à B'. Nous considérons que le pincement des lèvres est minimal et nul lorsque la bouche est fermée, mais au repos, sans crispation. La hauteur B' prend alors une valeur de référence à B'_{neutre}. Le pincement des maximal est maximal lorsque les lèvres entrent complètement à l'intérieur de la bouche et que la surface labiale n'est plus visible, c'est à dire lorsque B' est nul. Nous avons par conséquent introduit le paramètre $PI = (B'_{neutre} - B') / B'_{neutre}$, défini uniquement lorsque B=0, correspondant au pourcentage de pincement des lèvres lors de leur fermeture.



Figure 147. Représentation des deux types de pincement labial : avalé (à gauche) et protrus (à droite)

4.2.2. Méthodes de mesure

Il existe plusieurs systèmes d'acquisition des mouvements articulatoires labiaux (optotrack, EMA, etc.). Nous avons choisi pour une première étude d'utiliser le système de suivi labiométrique (ou Lip Tracking) de l'ICP, développé par Lallouache 1991 ^[190], basé sur une détection des contours labiaux à partir d'enregistrements vidéo de face et de profil du visage du locuteur dont les lèvres ont été maquillées en bleu (cf. Figure 146). Nous tenons à ce propos à renouveler nos remerciements à l'ICP et à Christophe Savariaux en particulier qui nous a beaucoup aidé dans l'acquisition des vidéos et l'extraction des paramètres articulatoires.

La couleur bleue avec laquelle sont maquillées les lèvres du locuteur a été choisie car elle n'est pas présente naturellement dans le visage du locuteur. Cela permet ensuite, lors du traitement d'image, d'appliquer un chroma-key¹ sur l'image, de manière à convertir en noir les zones

¹ « Incrustation en chrominance » en français. Il s'agit d'un code de chromaticité utilisé pour modifier les couleurs d'une image, et permettant l'incrustation d'une autre image. Ce procédé est utilisé en particulier dans le cinéma où un objet est filmé sur un fond uni de couleur pure. Après numérisation, ce fond peut être supprimé par sélection de cette couleur spécifique, et remplacé par une autre image de fond, par exemple un décor.

initialement colorées de ce bleu spécifique, et à filtrer les images par un filtre médian. Le contraste entre les lèvres et le reste de l'image (peau, dents, intérieur de la bouche) se trouve considérablement amélioré par rapport à une image de lèvres non maquillées. La détection des contours est alors plus fiable et précise. Ce système est très robuste et permet de détecter des aires intero-labiales extrêmement faibles pouvant atteindre 0.5mm². Par contre, il nécessite pour être performant de remaquiller régulièrement le locuteur au cours de l'expérience pour que la couleur bleue soit bien uniforme sur l'ensemble des lèvres. La tête du locuteur est maintenue immobile pendant l'enregistrement, grâce à un casque dans lequel il met sa tête, fixé au mur derrière lui. Une règle est attachée au casque, sur le profil gauche filmé par la caméra (cf. Figure 148). Elle permet de définir une position de référence pour le calcul des variations de la protrusion des lèvres.

Au début de l'enregistrement, une trame quadrillée (chaque carré faisant 1cmx1cm) est placée contre le visage du locuteur, de face et de profil (cf. Figure 148). Les images de ces instants sont enregistrés par les caméras. Il est ainsi possible par la suite de convertir les distances en pixel mesurées à partir des images vidéo, en distances en cm.

Les images vidéo correspondant à chaque énoncé sont ensuite numérisées. A partir de ces séquences, l'application «TACLE» (Traitement Automatique du Contours des LEvres), développée à l'ICP (Audouy 2000 ^[15]), permet d'extraire des paramètres descripteurs des lèvres.

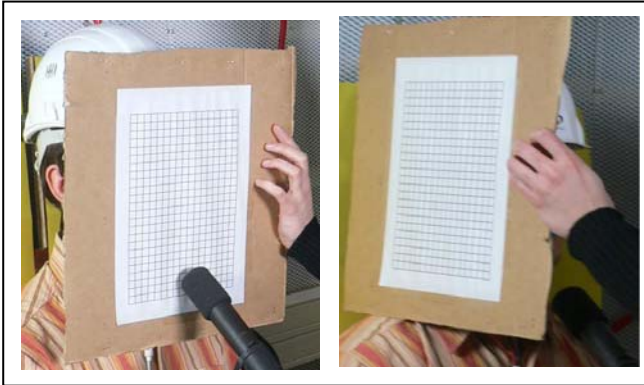


Figure 148. Trames de face et de profil utilisées au début de l'expérience pour assurer la conversion en centimètres des distances en pixels.

4.3. Les indices laryngés de l'articulation

Définition

En plus des mouvements articulatoires labiaux, nous nous sommes également intéressés dans ce travail aux mouvements verticaux du larynx (VLP, pour vertical larynx position). Les mouvements articulatoires labiaux peuvent relier aussi bien à des variations de qualité vocale qu'à l'intelligibilité acoustique et visuelle des unités phonétiques. Par contre, une élévation du larynx est plutôt associée à une tension vocale lors de la phonation et un abaissement peut être relié à l'allongement du conduit vocal ainsi qu'à la dilatation de la cavité pharyngée pour renforcer le timbre de la voix autour de 3kHz.

De la même façon que pour les mouvements labiaux, nous avons choisi d'exprimer le déplacement vertical du larynx non pas en unités de longueur mais de façon relative par rapport à son maximum et à sa position au repos. Nous avons ainsi introduit le paramètre $(VLP - VLP_{\text{neutre}}) / \max(|VLP - VLP_{\text{neutre}}|)$ rendant compte du pourcentage de déplacement vertical du larynx.

Mesure

Nous avons mesuré les mouvements verticaux du larynx à l'aide d'un électroglottographe (EGG) à 2 paires d'électrodes, placées sur le cou du locuteur, départ et d'autre du cartilage thyroïdien. Le signal acquis à l'EGG correspond à la différence de potentiel entre les électrodes hautes et basses, et donc à un signal électrique continu dont la variation de voltage est linéairement reliée au mouvement vertical du larynx. La bande passante de notre carte son ne permettant pas d'acquérir un signal de fréquence inférieure à 50Hz, nous avons du moduler en amplitude le signal issu de l'EGG à l'aide d'une porteuse de fréquence 1kHz, afin d'enregistrer sur la carte son un signal sinusoïdal dont l'enveloppe représente l'amplitude des mouvements du larynx.

Annexe Met2 :

Détails techniques liés à la constitution des bases de données

- 1. CALIBRATION DU NIVEAU DE BRUIT ET DES
RETOURS AUDITIFS**
- 2. ACQUISITION SYNCHRONE, RESTITUTION ET
MIXAGE DES SIGNAUX**
- 3. TRAITEMENT DES DONNEES**

1. CALIBRATION DU NIVEAU DE BRUIT ET DES RETOURS AUDITIFS

1.1. Calibration des niveaux de bruit sur haut-parleurs et par casque, et pour les deux types de bruit.

Pour pouvoir comparer rigoureusement les deux types de bruit, et les trois méthodes d'immersion du locuteur dans le bruit, nous avons dû calibrer les niveaux de sortie du signal audio de bruit de façon à obtenir exactement les mêmes niveaux au niveau du locuteur pour une restitution du bruit par casque ou par haut-parleurs.

Nous avons effectué un premier choix de réaliser la calibration en dB SPL.

Pour la calibration des niveaux de bruit à restituer sur haut-parleurs (Tanoy System 600), nous avons utilisé un microphone de pression 1/2" (B&K 4165) placé au niveau de l'oreille d'une tête artificielle en polystyrène, à l'endroit où devait être placé le locuteur dans la pièce (à 1.5 m à équidistance des deux enceintes).

Pour la calibration des niveaux de bruit à restituer dans le casque (Sennheiser HD250 Linear II), nous avons utilisé une oreille artificielle (B&K 4153), représentée Figure 149.

Dans les deux cas, la capsule de pression a ensuite été montée sur le même préamplificateur (B&K 2669) et le même amplificateur de mesure (B&K Nexus 2690). Ce dernier délivre un signal de calibration interne (1V RMS à 1kHz) servant de référence pour pouvoir ensuite mesurer avec précision le niveau de pression au niveau de la capsule du microphone.



Figure 149. Oreille artificielle B&K 4153 permettant de calibrer avec précision le niveau d'intensité à l'intérieur d'un casque.

Nous avons ainsi déterminé le niveau du bruit blanc et du bruit cocktail à envoyer en entrée des haut-parleurs de façon à mesurer 62dB SPL, 70 dB SPL, 78 dB SPL et 86 dB SPL au niveau de l'oreille de la tête artificielle. De façon similaire, nous avons déterminé le niveau du bruit blanc et du bruit cocktail à envoyer en entrée du casque de façon à mesurer 62dB SPL, 70 dB SPL, 78 dB SPL et 86 dB SPL au niveau de l'oreille artificielle.

1.2. Calibration des retours auditifs dans le casque

L'atténuation auditive provoquée par le port du casque a été estimée à 20 dB.

Le port du casque atténue à la fois le retour auditif que le locuteur a de sa propre voix, mais il va diminuer également l'intelligibilité du partenaire. Cela est gênant car le partenaire risque alors de parler encore plus fort, non plus pour s'adapter au bruit mais pour s'adapter à la « surdité » artificielle du locuteur. Cela constituerait un biais important à l'expérience. Par ailleurs, nous souhaitons que la perception que le locuteur a de son partenaire pour les sessions 1 et 2 (port du casque) soit comparable à la session 3 (haut-parleurs). Ainsi, ce qui change entre les trois sessions est principalement la perception que le locuteur a de sa propre voix.

Pour pallier ce problème, nous avons réinjecté dans le casque du locuteur le signal acquis au microphone de l'interlocuteur, en tenant compte de l'atténuation naturelle du son liée à la distance entre les sujets. Après avoir réalisé des essais de calibration très précise mais peu convaincante, nous avons finalement opté pour une calibration perceptive. Nous avons demandé à deux personnes de prendre la place des locuteurs de l'expérience dans la pièce et de converser librement, équipés du microphone serre-nuque. L'une de ces personnes portait le casque sur les oreilles pendant que nous

calibrions le retour de son interlocuteur dans le casque. Nous lui avons demandé de comparer ce qu'elle percevait avec et sans port de casque et de nous faire signe quand le niveau de retour de l'interlocuteur lui paraissait le plus naturel possible par rapport à sa distance, et le plus comparable sa perception de l'interlocuteur sans casque.

De la même manière, pour la condition de restitution du bruit par casque avec ajout du retour de sa propre voix, nous avons calibré la ré-injection de la voix acquise au microphone avant l'expérience, dans le silence et avec le concours d'une autre personne. Nous lui avons demandé de comparer la façon dont elle percevait sa voix avec et sans casque et de nous faire signe quand le niveau de retour de sa voix lui paraissait le plus naturel possible et le plus comparable à la perception de sa propre voix sans casque.

Nous avons gardé ces mêmes calibrations pour tous les sujets de l'expérience. Nous n'avons pas appliqué d'effet de salle à la voix de l'interlocuteur ni tenu compte de la fonction de transfert bouche-oreille mais seulement envisagé l'atténuation de niveau sonore globale.

2. ACQUISITION SYNCHRONE, RESTITUTION ET MIXAGE DES SIGNAUX

2.1. Pour la première base de donnée BD1

La synchronisation des signaux Audio enregistrés avec le fichier de bruit restitué est une condition nécessaire au « débruitage » des enregistrements où le bruit est restitué sur haut-parleurs. Par ailleurs, il est également nécessaire d'acquérir les signaux Audio et EGG de façon simultanée pour pouvoir les mettre en relation. Pour cela nous avons utilisé un logiciel d'acquisition multitrack (Adobe Audition 1.0) permettant de diffuser une piste Audio (l'environnement bruyant) simultanément à l'acquisition d'autres pistes (Audio et EGG, pour les deux locuteurs).

Lors des situations de restitution du bruit par casque, nous avons dû mixer le signal de bruit avec le signal audio de l'interlocuteur afin de diffuser cet ensemble sur le casque du locuteur. Dans le cas de la session 2 où le locuteur dispose d'un retour de sa propre voix, nous avons dû encore rajouter le signal audio du locuteur à ce mixage. Nous avons dû opter pour une solution de mixage analogique, afin de mixer les signaux audio analogiques en cours d'acquisition avec le signal audio de bruit en sortie du convertisseur numérique-analogique. La Figure 150 récapitule le câblage de l'expérience.

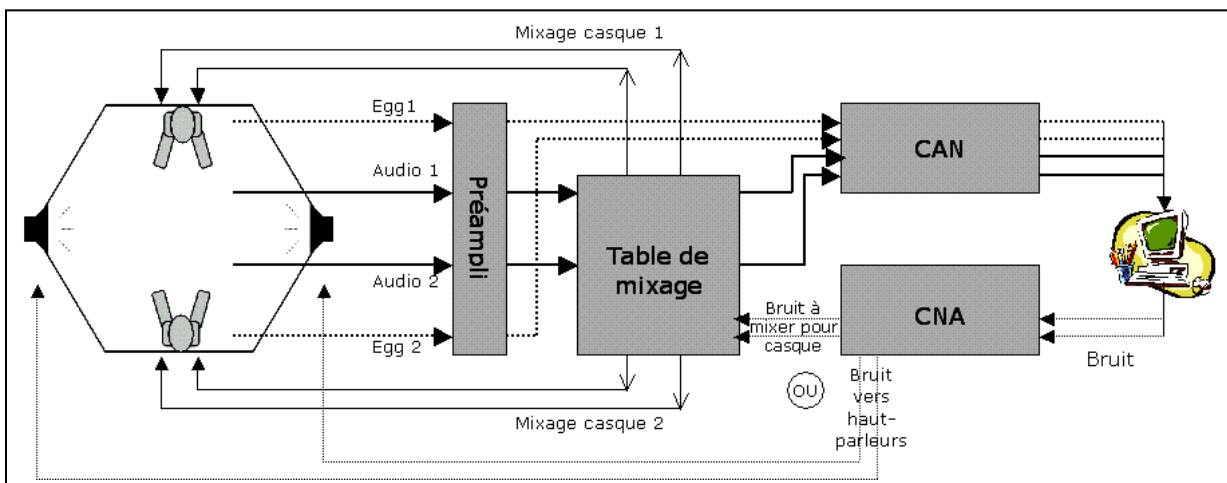


Figure 150 : Récapitulatif du dispositif expérimental d'acquisition, restitution et mixage des signaux. Les signaux audio de parole sont représentés en gras, les signaux EGG en gras pointillé.

Le signal audio de l'environnement bruyant à reconstituer est représenté en gris foncé pointillé.

Le signal audio en sortie de la table de mixage à destination du casque est représenté en gris foncé.

La table de mixage est intercalée entre le Convertisseur analogique/numérique (CAN) (RME ADI 8 Pro) et le préamplificateur (RME Octamic) pour les signaux Audio seulement. Les signaux EGG sont directement numérisés à la sortie du préamplificateur.

Pour la session 1 (restitution par casque sans retour de sa propre voix), le signal diffusé sur le casque du locuteur correspond au mixage du bruit (à la sortie du CNA) avec le signal audio de l'interlocuteur.

Pour la session 2 (restitution par casque avec retour de sa propre voix), le signal diffusé sur le casque du locuteur correspond au mixage du bruit (à la sortie du CNA), du signal audio de l'interlocuteur et du signal audio du locuteur.

Pour la session 3 (restitution par haut-parleurs), la table de mixage ne sert à rien puisque le bruit est diffusé directement sur les hauts-parleurs et qu'aucun son n'est diffusé dans les casques. Les signaux Audio continuent de transiter par la table de mixage mais celle-ci est complètement neutre vis à vis de l'expérience et de l'acquisition des signaux.

L'acquisition des signaux et le contrôle des appareils de mesure se passait à l'extérieur de la pièce assourdie où se trouvaient les locuteurs. Nous avons donc installé une WebCam dans la pièce ainsi qu'un microphone d'ambiance pour pouvoir communiquer avec les sujets et contrôler le déroulement du protocole depuis l'extérieur.

2.2. Pour le corpus articulatoire pilote (CP) et la deuxième base de donnée BD2

Nous avons toujours utilisé le logiciel d'acquisition multitrack (Adobe Audition 1.0) pour enregistrer les signaux Audio et Egg de façon synchrone avec le fichier de bruit restitué sur haut-parleurs.

Dans cette expérience, il était également nécessaire de synchroniser les signaux audio et EGG avec les données articulatoires, autrement dit avec le signal vidéo. Il se trouve que le signal audio est également mixé à la vidéo et enregistré sur les bandes Betacam. Nous avons mis en correspondance les données articulatoires avec les données acoustiques en calculant l'intercorrélation entre le signal audio issu de la vidéo et le signal audio acquis à l'ordinateur.

Pour chaque énoncé, nous avons ensuite synchronisé de façon plus précise les signaux articulatoires avec les enregistrements audio en synchronisant le signal $S+|P1|$ (où S représente l'aire intéro-labiale et P1, la protrusion labiale) avec le signal d'intensité de la voix, mesurée à partir de l'enregistrement acoustique. Ces deux courbes présentent en effet la même allure, avec des lobes sur chaque voyelle.

3. TRAITEMENT DES DONNEES

3.1. Débruitage des fichiers audio de la session 3

Afin de réaliser des analyses acoustiques et phonétiques à partir des fichiers Audio comprenant à la fois la voix du locuteur et le bruit environnant, nous avons utilisé une méthode de traitement du signal exposée dans le chapitre 5 et permettant de séparer la voix du bruit ambiant tout en conservant une très bonne qualité du signal de parole.

3. 2. Segmentation et étiquetage des fichiers

Nous avons segmenté et labellisé chaque énoncé, mot-cible, syllabe, voyelle et consonne de ces mots-cibles à la main sous Praat. Pour cela, nous avons établi, et suivi les critères de segmentation détaillés en annexe Met3.

1.4.3. Système d'indexation du corpus

Le corpus étant constitué de parole semi-spontanée, il nous a fallu indexer certaines particularités de l'énonciation des mots-cibles pour pouvoir réaliser par la suite des analyses automatiques du corpus.

En particulier, nous avons indexé la **position des mots-cibles dans l'énoncé** : **Initiale** lorsque le mot-cible est le sujet de l'énoncé (par exemple « Le requin est associé à... »), **finale** lorsque le mot-cible est en fin d'énoncé (par exemple « le chemin va vers le requin »), et **intermédiaire** lorsque le mot-cible est situé au milieu de l'énoncé (par exemple « je pars du requin pour ensuite me diriger vers... »). La durée des mots ou leur intonation étant dépendante de leur position dans

l'énoncé, cette indexation nous a permis de trier automatiquement les mots-cibles en position comparable pour effectuer les analyses.

Nous avons également indexé le **mode d'énonciation de chaque mot cible**, selon s'il était **prononcé pour la première fois** ou bien **répété** suite à une incompréhension de l'interlocuteur, ou bien encore **prononcé par le locuteur pour lui-même** (lorsqu'il le cherchait sur sa carte, par exemple).

Enfin, nous avons également indexé l'**ordre d'énonciation des mots-cibles**, dans le but d'examiner un éventuel effet d'ordre ou encore si une demande de répétition par l'interlocuteur affectait la prononciation des mots suivants.

Annexe Met3 :

Critères de segmentation

- 1. REPERAGE DU DEBUT ET DE LA FIN DES ENONCES**
- 2. SEGMENTATION DES SYLLABES**
- 3. SEGMENTATION DES CONSONNES ET DES VOYELLES**
- 4. SEGMENTATION DES CIBLES ET DES TRANSITIONS**

Avant de commencer à segmenter la première base de données, nous avons effectué des recherches bibliographiques sur les conventions et critères de segmentation, et questionné plusieurs chercheurs en phonétique à ce propos. Il semble que chacun adopte ses propres critères, cohérents au sein de chaque étude, mais sans qu'il existe de consensus partagé par tous. Nous avons cependant trouvé un article proposant quelques critères de segmentation de la parole spontanée en coréen (Keating et al. 1994 ^[178]), sur lequel nous nous sommes basés pour établir nos propres critères de segmentation, en complément de la description acoustique des phonèmes du français donnée par Calliope 1989 ^[41].

Par ailleurs, puisque dans cette étude nous travaillons en relatif, en comparant la parole produite dans le bruit à la parole produite dans le calme, l'essentiel n'est pas tant d'adopter un critère de segmentation donné, mais surtout de conserver le même critère pour l'étiquetage de l'ensemble d'un corpus.

1. REPERAGE DU DEBUT ET DE LA FIN DES ENONCES

Dans le corpus pilote et dans la deuxième base de données, nous avons étudié des énoncés entiers. Ceux-ci commençaient toujours par des segments voisés. Nous avons donc indexé le début de l'énoncé à l'apparition du voisement, et nous pas uniquement à l'apparition de la structure formantique. Nous n'avons pas non plus tenu compte d'éventuels coups de glotte ou de bruits de bouche ou de déglutition précédant l'énoncé (cf. Figure 151).

De même, nous avons commencé par indexer la fin des énoncés au niveau de la disparition du son, qu'il soit voisé ou bruité. Cependant, nous avons observé que certaines locutrices finissaient leur énoncé par du fry ou par la continuation des voyelles finales sans voisement mais uniquement avec du souffle. Ces phénomènes semblent intéressants et mériteraient d'être segmentés en soi. Mais ils introduisaient trop de variabilité intra et inter-locutrice. Nous avons obtenu des résultats moins variables en indexant la fin des énoncés au niveau de la disparition du voisement ET de la structure formantique des voyelles finales (donc en ignorant les épisodes de fry ou de chuchotement).

Nous avons adopté les mêmes critères dans le cas où les locutrices inséraient des pauses à l'intérieur de l'énoncé : le début de la pause a été indexé au niveau de la disparition de la structure formantique de la voyelle précédant la pause, et la fin de la pause a été repérée par l'apparition du voisement de la consonne suivant la pause.

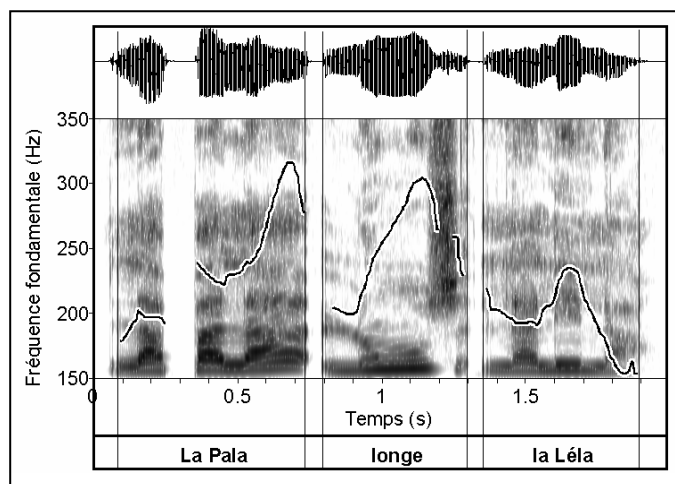


Figure 151. Exemple d'indexation du début et de la fin d'un énoncé, ainsi que de deux pauses insérées entre le sujet et le verbe de l'énoncé, et entre le verbe et l'objet. Le début de l'énoncé est ici précédé d'un bruit d'accolement des cordes vocales non voisé, que nous ne prenons pas en compte dans la segmentation. De même, la première pause entre le sujet et le verbe commence sur une fin de la voyelle précédente en fry.

2. SEGMENTATION DES SYLLABES

La segmentation en syllabes des syntagmes-cibles des différentes bases de données a été facilitée par le fait que tous ces syntagmes ont été construits à partir de syllabes de type CV. La fin de chaque syllabe a donc été indexée au niveau de la fin de chaque voyelle, assez facilement repérable sur le spectrogramme (cf. Figure 152).

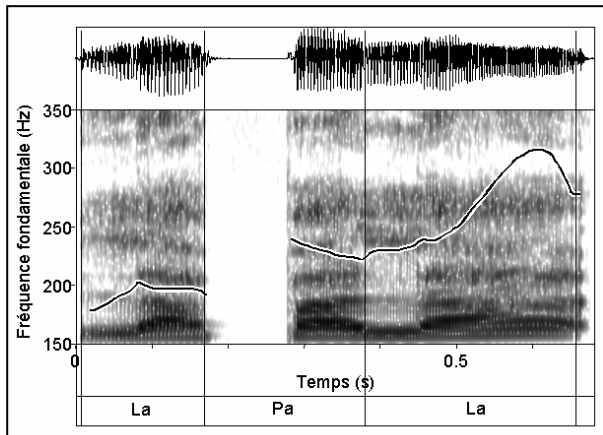


Figure 152. Exemple d'indexation des différentes syllabes du syntagme « La Pala », correspondant au sujet de l'énoncé représenté sur la Figure précédente. La dernière syllabe [la] se termine par un *fry*, que nous ne prenons pas en compte dans la segmentation du syntagme.

Toutefois, dans la 1^{ère} base de données, il arrivait que les mots-cibles soient suivis d'un verbe ne commençant pas par une consonne (par exemple : « Le fusil est associé... », cf. Figure 153). Dans ce cas, la frontière entre la fin du syntagme-cible et la voyelle suivante était trop ambiguë pour être déterminée seulement à partir du spectrogramme et de la visualisation de la structure formantique. Nous nous sommes alors aidés de la courbe intonative pour indexer la fin de la dernière syllabe du mot-cible. C'est pourquoi nous avons ensuite construit le corpus de l'étude pilote et celui de la 2^{ème} base de données en imposant des verbes commençant par une consonne, de façon à faciliter la segmentation.

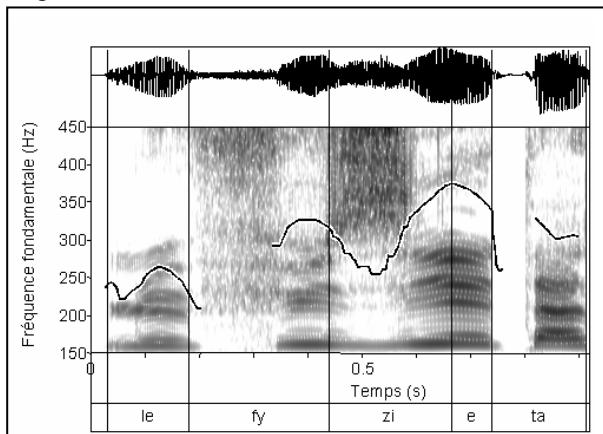


Figure 153. Exemple d'indexation des différentes syllabes du syntagme «Le fusil», suivi du verbe « est associé ». La frontière entre le syntagme cible et le verbe a été dans ce cas indexée grâce au maximum de la courbe intonative.

3. SEGMENTATION DES CONSONNES ET DES VOYELLES

La grande majorité des études sur la parole Lombard ont étudié les caractéristiques formantiques, et spectrales ainsi que la durée des différents phonèmes en segmentant des mots-cibles ou des énoncés en consonnes et voyelles. Au tout début de la thèse, nous avons donc suivi cet exemple pour les mots-cibles de la 1^{ère} base de données. En comptant toutes les occurrences des mots, tous locuteurs et conditions confondues, cela a représenté 5773 mots de forme CVCV à segmenter en consonnes et voyelles, autrement dit 23092 segments !

Devant l'ampleur de la tâche, nous avons tout d'abord cherché à automatiser ou semi-automatiser cette segmentation. La segmentation automatique possède en outre l'avantage de conserver exactement les mêmes critères pour une large base de données. Pour cela, nous avons cherché à utiliser un outil d'étiquetage automatique développé au LIMSI (Gauvain et al. 2002^[110]). Il nous a fallu pour cela transcrire textuellement tous les énoncés semi-spontanés produits par les locuteurs, de façon à ce que le programme n'ait pas à reconnaître les segments mais seulement à en délimiter les frontières. Il nous a donc également fallu recenser toutes les prononciations possibles des différents mots produits semi-spontanément par les locuteurs, afin d'établir un dictionnaire de prononciation pour le logiciel. Enfin, nous avons dû supprimer tous les silences trop longs entre deux énoncés qui entraînaient de grosses erreurs d'étiquetage.

Malgré tout ce travail, les résultats de l'étiquetage automatique n'ont pas pu être utilisés tels quels car l'erreur d'indexation par rapport à ce que nous aurions étiqueté à la main dépassait très souvent les 20ms et parfois les 70 ms ! Ces résultats décevants étaient en fait assez prévisibles puisque ce type de programme se base sur un modèle « normal » de parole, et non sur de la parole Lombard. Par conséquent, il rencontre les mêmes difficultés que les autres algorithmes de reconnaissance de la parole face à la parole produite dans le bruit, non à cause du bruit ambiant, mais à cause des modifications de la parole elle-même qui ne sont plus en correspondance avec le modèle d'apprentissage. D'ailleurs, les principales études acoustiques-phonétiques menées sur la parole Lombard ont cherché à modéliser ces modifications de la parole dans le but d'améliorer la robustesse des algorithmes de reconnaissance vocale (Junqua et al. 1990^[174], Hanson et al. 1990^[134]).

Nous avons donc tout repris à la main en réajustant sous Praat (Boersma et al.^[30]) les frontières entre les voyelles et les consonnes des mot-cibles. Nous avons adopté des critères différents selon les catégories de consonnes.

3.1. Frontières entre une consonne occlusive et des voyelles adjacentes

Nous avons fixé le début de l'occlusive à la fin de la structure formantique de la voyelle précédente. Ainsi, l'occlusive peut commencer sur un résidu de voisement (cf. [p] de « panda » sur la Figure 154) ou sur la nasalisation du voisement due à la voyelle nasale précédente (cf. [d] de « panda » sur la Figure 154). Plusieurs conventions existent pour l'indexation de la fin d'une consonne occlusive : certains la fixent au niveau de la phase de plosion, sans ambiguïté visuelle sur le spectrogramme, et correspondant au relâchement de l'occlusion. D'autres déterminent la fin d'une occlusive au niveau de l'apparition du voisement après la plosion, d'autres encore au niveau de l'apparition de la structure formantique de la voyelle suivante. Hors cette structure formantique n'est pas stable dès son apparition. Ces transitions formantiques contribuent autant que la plosion à la reconnaissance perceptive des différents types de consonnes occlusives (Ohde 1988^[259]). Par ailleurs, le but de notre segmentation n'est pas seulement de réaliser des mesures de durée des segments mais également d'analyser la structure formantique des voyelles. Nous avons donc choisi de fixer la fin des consonnes occlusives au moment où la structure formantique de la voyelle suivante devient stable, autrement dit après la phase de transition. Nous ne nous sommes donc pas seulement servi du spectrogramme pour cette segmentation mais nous nous sommes également aidés des trajectoires formantiques estimées par Praat (Boersma et al.^[30]).

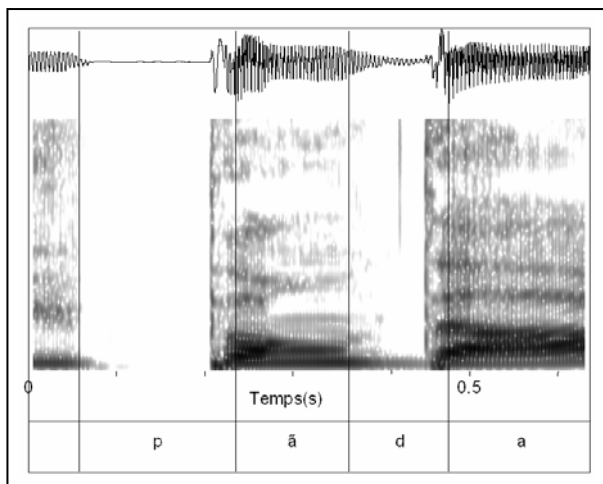


Figure 154. Exemples d'indexation de la frontière entre des consonnes occlusives (ici [p] et [d]) et des voyelles pour le syntagme « Le panda ». Le début de l'occlusive coïncide avec la disparition de la structure formantique de la voyelle précédente, même s'il reste un voisement pendant la phase d'occlusion d'une plosive non voisée telle que [p], ou une nasalisation pendant la phase d'occlusion d'une plosive voisée telle que [d].

3.2. Frontières entre une consonne fricative et des voyelles adjacentes

Nous avons défini des critères de segmentation différents pour les consonnes fricatives voisées et non voisées.

L'indexation du début et de la fin d'une fricative non voisée ([s], [f] et [ʃ]) est ce qui nous a paru le moins ambigu, car cela est assez bien marqué au niveau du spectrogramme : Nous avons donc fait débuter une fricative non voisée à l'instant où disparaît la structure formantique de la voyelle

précédente. Au contraire, nous avons indexé sa fin à l'instant où la structure formantique de la voyelle suivante réapparaît (cf. Figure 155).

Au contraire, nous avons eu beaucoup de difficulté à trouver un critère objectif et peu variable entre les conditions et les locuteurs pour segmenter des fricatives voisées. Puisqu'un de nos objectifs était de réaliser des analyses acoustiques sur les voyelles, nous avons choisi de segmenter la voyelle « propre » et d'inclure les transitions formantiques fricative-voyelle et voyelle-fricative dans la consonne (cf. Figure 155). Nous avons donc fixé le début de la fricative au moment où la structure formantique de la voyelle précédente amorce une transition montante et où l'énergie spectrale au-dessus de 6kHz apparaît de façon importante sur le spectrogramme par des barres « verticale » (i.e sur toutes les fréquences) et masque l'organisation formantique horizontale de la voyelle. Au contraire, nous avons indexé la fin de la fricative au moment où l'énergie spectrale hautes-fréquences diminue et laisse réapparaître la structure formantique de la voyelle suivante, et au moment où les 3 premiers formants redeviennent stables. Pour effectuer cette segmentation, nous avons donc dû visualiser le spectrogramme non plus seulement en dessous 5kHz, mais plus globalement de 0 à 20kHz, et nous nous sommes également aidés des trajectoires formantiques estimées par Praat (Boersma et al. ^[30]).

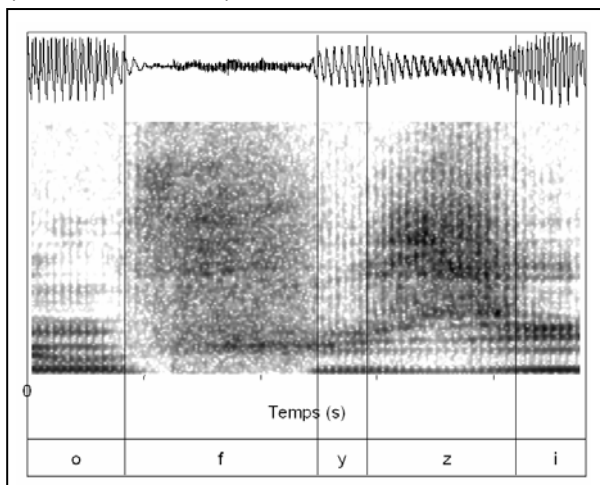


Figure 155. Exemples d'indexation de la frontière entre des consonnes fricatives (ici [f] et [z]) et des voyelles pour le syntagme « au fusil ».

3.3. Frontières entre une consonne nasale ou liquide et des voyelles adjacentes

L'indexation des frontières entre des voyelles et des consonnes liquides [l] ou des consonnes nasales [n] ou [m] n'est également pas toujours évidente du fait que ces consonnes présentent également une structure formantique comme les voyelles. Cependant, l'énergie acoustique est beaucoup plus faible en hautes fréquences pour ses consonnes que pour les voyelles. Ainsi, nous avons indexé le début des consonnes [l] et des consonnes nasales grâce à l'atténuation de l'énergie hautes fréquences sur le spectrogramme en complément de la variation rapide de la structure formantique. Au contraire, la fin de ces consonnes a été indexée à partir des variations formantiques et de l'énergie spectrale hautes-fréquences, mais également grâce à un petit « claquement » très rapide, correspondant au relâchement de l'occlusion bilabiale sur les [m] ou au relâchement de l'occlusion palatale dans le cas des [l] et des [n].

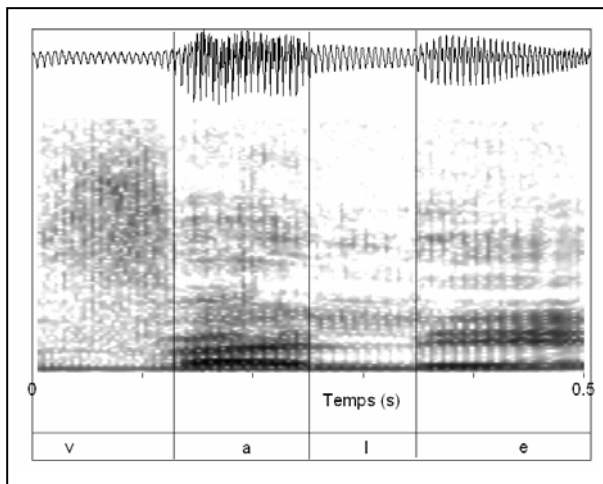


Figure 156. Exemple d'indexation de la frontière entre un [v] et ses voyelles adjacentes dans le cas du mot « vallée ».

Les consonnes [ʁ] se démarquent un peu mieux des voyelles par leur prononciation souvent bruitée. Leur segmentation est alors comparable à celle des fricatives voisées : nous avons choisi de segmenter la voyelle « propre » et d'inclure les transitions formantiques [ʁ]-voyelle et voyelle-[ʁ] dans la consonne (cf. Figure 157).

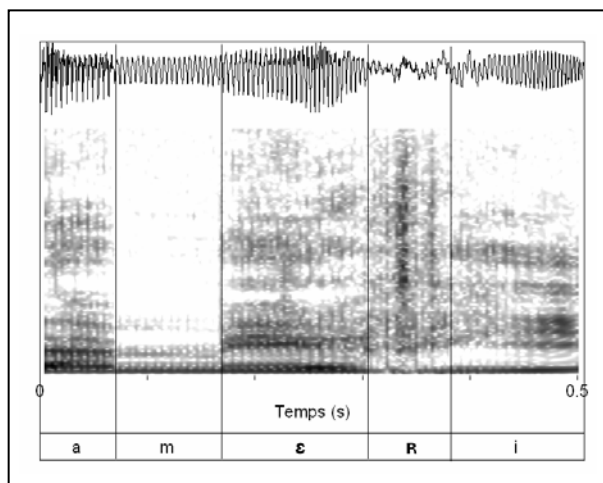


Figure 157. Exemple d'indexation de la frontière entre une consonne nasale (ici [m]) et ses voyelles adjacentes dans le cas du syntagme « la mairie ». De même pour la frontière entre un [ʁ] bruité ses voyelles adjacentes.

3.4. Problèmes liés à la segmentation des voyelles et des consonnes

La définition de critères de segmentation et leur conservation sur une quantité importante de données est un premier problème pour étudier de façon rigoureuse les caractéristiques acoustiques et de durée des différents segments phonétiques. De plus, ces critères sont avant tout « visuels » par référence au spectrogramme ou aux trajectoires formantiques. Ils sont donc très influencés par le nombre de points sur lequel est calculée la FFT, sur la taille de la fenêtre d'analyse, sur le contraste de niveaux de gris choisi pour afficher le spectrogramme, etc. Nous avons par conséquent toujours gardé les mêmes paramètres de réglages du spectrogramme sous Praat (Boersma et al. ^[30]) pendant la segmentation, en fixant une taille de fenêtre d'analyse (gaussienne) à 5ms, une résolution temporelle de 3 ms, et une résolution fréquentielle de 50Hz, et une dynamique d'intensité de 50dB pour l'affichage du spectrogramme.

De même, le niveau de « zoom » du spectrogramme a une grande influence sur la visualisation des frontières entre les phonèmes, et donc sur la segmentation. Ainsi, la segmentation des occlusives nécessite la visualisation précise des transitions des trois premiers formants, autrement dit une focalisation sur les fréquences inférieures à 3.5-4kHz. Tandis que la segmentation des fricatives ou des [ʁ] nécessite au contraire une visualisation de l'énergie hautes-fréquences jusqu'à 20kHz.

Enfin, il existe d'autres problèmes à la segmentation des voyelles et des consonnes, cette fois davantage de fond : la notion de phonème est en fait davantage définie au niveau phonologique que d'un point de vue acoustique ou articuloire, étant donné que la parole ne consiste pas en une juxtaposition de sons ou de mouvements articuloires, mais en leur continuum et leur recouvrement. Il est donc tout d'abord assez problématique de délimiter exactement la frontière entre deux phonèmes. Un choix doit toujours être fait d'attribuer la phase de transition entre deux phonèmes consécutifs à l'un d'entre eux, ou de mettre de côté cette phase de transition comme n'appartenant ni à l'un ni à l'autre. Ce dernier choix serait dommage puisque ces transitions ont une grande importance perceptive.

Par ailleurs, au-delà du problème de délimiter la frontière temporelle entre deux phonèmes, il est bon de ne pas perdre de vue à quel questionnement nous souhaitons répondre et si la segmentation en voyelles et consonnes est réellement la plus appropriée. En effet, si notre questionnement concerne essentiellement la façon dont les paramètres de production sont modifiés dans le bruit, il est peut être plus pertinent de segmenter les phases d'occlusion de la parole, les phases d'aspiration, de voisement, de friction, de plosion ou encore les phases présentant une structure formantique, plutôt que de segmenter le flux acoustique en consonnes et voyelles. De même, si notre questionnement concerne les mouvements articuloires, il serait peut être plus pertinent de définir des critères de segmentation de ces mouvements à partir des signaux articuloires eux-mêmes plutôt que du signal acoustique. Au contraire, si notre questionnement est plutôt d'ordre phonologique et concerne la façon dont les indices de reconnaissance des phonèmes sont modifiés dans le bruit, il serait peut être plus approprié de segmenter les phases formantiques stables, transitoires, la plosion des occlusives, leur phase d'occlusion, la phase de friction « propre » des fricatives, de façon distincte plutôt que de regrouper dans un même phonème des phases contenant des indices différents.

4. SEGMENTATION DES CIBLES ET DES TRANSITIONS

Ces réflexions nous ont amenés pour la 2^{ème} base de données à reconsidérer nos objectifs avant de réaliser la segmentation de ce nouveau corpus :

- Nous souhaitons tout d'abord réaliser des mesures formantiques et articuloires des voyelles. Dans ce but, il était surtout important d'indexer les « cibles » formantiques et articuloires ou du moins les phases pendant lesquelles la structure formantique est relativement stable.
- Les voyelles nous intéressaient également dans le cadre de l'étude de l'effort vocal puisqu'elles correspondent à des segments voisés (la plupart du temps), et par conséquent à des phases où le larynx fonctionne et est susceptible de se fatiguer. Or finalement, il existe également d'autres segments voisés que les voyelles : c'est le cas des consonnes nasales et liquides, des fricatives et des occlusives voisées. Aussi, il est plus pertinent dans cet objectif de segmenter les phases voisées et non voisées de la parole plutôt que les voyelles et les consonnes. Cela peut être assez simplement effectué à partir de l'estimation de la fréquence fondamentale.

Pour ces différentes raisons, il nous a semblé plus pertinent dans la 2^{ème} base de données d'abandonner cette distinction consonne/voyelle et de segmenter plutôt les phases formantiques stables et les phases de transitions formantiques.

Les différents logatomes constituant la deuxième base de donnée présentent des transitions entre la consonne [l] et les voyelles [a], [e], [i], [y], [u], [ã], ainsi que des transitions entre la voyelle [a] et les consonnes bilabiales [p], [b] et [m]. Pour déterminer le début et la fin de chacune de ces transitions formantiques, nous avons examiné le 1^{er} formant des transitions consonne-[a] et des transitions entre un [l] et un [e]. Nous avons examiné le 2^{ème} formant pour les transitions entre un [l] et une voyelle [y], [u], [ã] ou [i]. Au moment de la transition, la trajectoire du formant pris comme référence a globalement l'allure d'un S ou d'un S inversé de transition à l'allure d'un S ou d'un S inversé (cf. Figure 158). Nous avons donc adopté comme convention d'indexer le début de la transition formantique au début de ce « S », c'est à dire à l'instant où la trajectoire du formant pris comme référence quitte le « plateau » du premier phonème. Nous avons repéré la fin de la transition à l'instant où le « S » se termine, c'est à dire quand la trajectoire formantique atteint le plateau du phonème suivant.

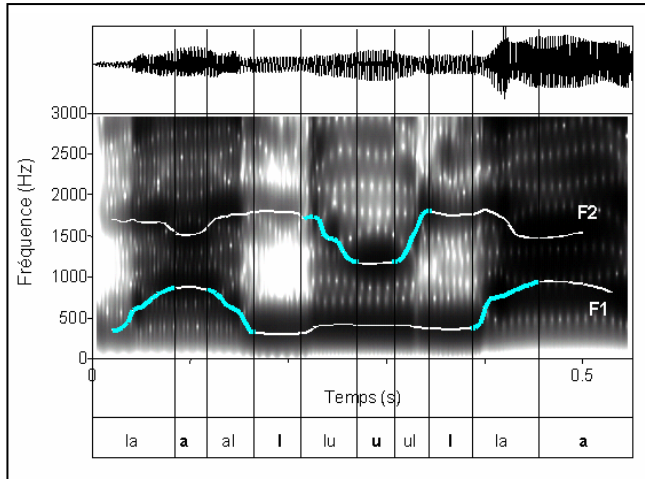


Figure 158. Exemple d'indexation des transitions formantiques et des phases formantiques stables pour le syntagme «La Lula ». Dans ce cas, la trajectoire du 1^{er} formant est utilisée pour indexer les transitions [l]-[a] et [a]-[l], tandis que la trajectoire du 2^{ème} formant est utilisée pour indexer les transitions [l]-[u] et [u]-[l].

Il arrive que la transition prenne l'allure d'un double S (cf. transition [l]-[u] de la Figure 158), ou qu'il n'y ait pas de « plateau » formantique mais uniquement un maximum. Dans tous les cas, nous avons conservé les critères de segmentation des transitions formantiques exposés ci-dessus.

De la même façon que pour la segmentation des voyelles et des consonnes, l'indexation des transitions formantiques est fortement dépendante des paramètres d'estimation formantique. Nous avons donc toujours conservé une estimation de 5 formants en dessous 5.5kHz, sur une fenêtre d'analyse de 20ms, avec une résolution temporelle de 5ms (1/4 de la fenêtre d'analyse). De même, la gamme de fréquences pour laquelle sont affichés le spectrogramme et les trajectoires formantiques est déterminante sur la visualisation des transitions formantiques et sur la précision de leur annotation. Puisque nous nous intéressons exclusivement aux deux premiers formants, nous avons conservé pendant toute la segmentation des transitions formantiques un affichage du spectrogramme et des formants en dessous de 3kHz.

Annexes d'études pilotes

Annexe Pil1 :

**Etude pilote de l'influence
des retours de
compréhension de
l'interlocuteur**

- 1. INFLUENCE DE CETTE INCOMPREHENSION A
COURT TERME**
- 2. INFLUENCE DE CETTE INCOMPREHENSION A
MOYEN TERME**

Dans la première base de données, où les 10 locuteurs dialoguaient par binômes de façon semi-spontanée, nous avons observé plusieurs manifestations verbales et non verbales d'incompréhension de l'interlocuteur poussant le locuteur à répéter son message. Nous insistons sur le fait que nous nous avons considéré ces répétitions comme un « bonus » de la première base de données, et que nous les avons analysées à part et jamais de façon mélangée au reste du corpus principal.

Bien que nous n'ayons pas conçu la 1^{ère} base de données pour examiner de telles répétitions, celles-ci nous ont semblées riches d'information sur la boucle de rétroaction dans la communication, de l'interlocuteur vers le locuteur. Ce corpus de répétitions n'étant pas totalement bien « calibré » : nous avons choisi de ne pas présenter son exploration dans le manuscrit sur le même plan que les autres analyses plus rigoureuses. Cependant, nous avons quand même souhaité partager dans ces annexes les premières observations que nous avons réalisées de ces répétitions qui apportent des arguments concernant les hypothèses que nous effectuons sur les mécanismes d'adaptation de la parole dans le bruit, et qui mériteraient maintenant d'être approfondies par un corpus et une étude dédiée à ce questionnement. Ce travail a fait l'objet d'une communication (Garnier et al. 2006^[105]).

Pour constituer ce corpus de répétitions, nous avons commencé par recenser les différents cas d'incompréhension entre le locuteur et l'interlocuteur :

- l'interlocuteur peut marquer son incompréhension de façon « ouverte », par des formes interrogatives telles que « hein ? », « quoi ? », une élévation des sourcils, des formes affirmatives telles que « je ne comprends pas » ou encore des demandes explicites de répétition telles que « tu peux répéter ? ». Ces différents signes ont entraîné une répétition de la phrase entière par le locuteur.
- l'interlocuteur peut également répéter ce qu'il pense avoir compris du message mais sur lequel il a un doute. La plupart du temps, l'interlocuteur ne répète alors que le constituant « utile » (le mot-cible), en forme interrogative (« le chausson ? »), et non pas la phrase entière. Cette question peut alors entraîner trois réactions différentes du locuteur :
 - si l'interlocuteur avait effectivement bien compris, le locuteur peut se contenter de confirmer par un hochement de tête, ou de répondre par l'affirmative.
 - il peut également dans certains cas confirmer en répétant la phrase entière ou seulement le constituant « utile ».
 - enfin, si au contraire l'interlocuteur a mal compris le message, le locuteur le corrige en répétant la phrase entière initiale ou seulement son constituant « utile ».

Nous nous intéresserons ici aux répétitions du locuteur dans des buts de confirmation (focalisation informative) ou de correction (focalisation contrastive). Nous sommes bien conscients que ces deux types de répétition ne représentent pas les mêmes processus linguistiques. Il ont par contre en commun d'être motivés par le doute ou l'incompréhension de l'interlocuteur. Ce qui nous intéresse n'est pas de caractériser précisément ces répétitions d'un point de vue acoustique ou phonétique, mais d'observer si le retour d'incompréhension de l'interlocuteur va influencer ou non le locuteur à réévaluer la situation de communication et à revoir sa stratégie pour faire passer au mieux son message, à court et moyen terme.

1. INFLUENCE DE CETTE INCOMPREHENSION A COURT TERME

Dans un premier temps, nous avons donc examiné l'influence à court terme des marques d'incompréhension de l'interlocuteur et les procédés par lesquels le locuteur cherche à mieux faire passer son message lors de sa répétition. Nous insistons donc encore sur le fait que nous ne cherchons pas à caractériser la focalisation mais plutôt, à travers elle, à remonter aux stratégies du locuteur.

1.1. Matériel et méthode

Pour cela, nous avons donc examiné les différentes répétitions extraites de la première base de données. Nous avons regroupé toutes ces répétitions, produites par 10 locuteurs, pour des niveaux de bruit et des types de bruit différents. Les forts niveaux de bruit ont logiquement suscité plus de répétitions (cf. Figure 159). De même, certains mots-cibles ont moins bien compris que d'autres, et par conséquent davantage répétés (cf. Figure 159).

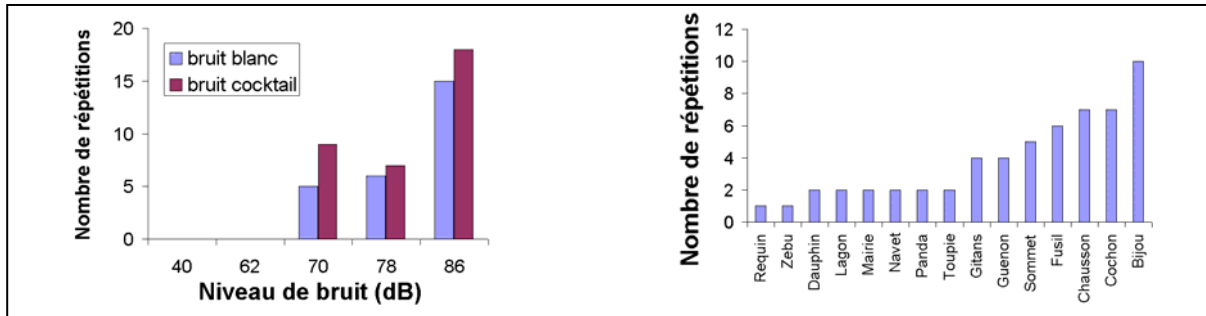


Figure 159. Composition du corpus de répétitions selon les niveaux et types de bruit dans lesquels elles ont été produites, et selon le mot-cible concerné.

Nous avons examiné pour chacune de ses 2 syllabes, la différence entre la répétition d'un mot cible et son énonciation première pour l'Intensité, la F0 moyenne, la durée syllabique, l'ambitus de F0, la dynamique d'intensité, la différence d'énergie des voyelles entre la zone [0-2kHz] et la zone [2-4kHz], (aussi appelé timbrage par la suite), et le déplacement formantique de chaque voyelle dans le plan F1/F2. Nous avons fait le choix de considérer la tendance globale d'évolution de chaque paramètre pour chaque répétition, plutôt que la magnitude de cette évolution, en vue de dégager différents profils-type d'évolutions. Pour cela nous avons repéré pour chaque paramètre la répétition dont l'évolution était la plus grande dans l'absolu. Puis nous avons fixé un seuil de 5% par rapport à cette évolution maximum, en dessous duquel nous avons considéré que le paramètre n'évoluait pas (codage 0), et au-dessus duquel au contraire nous avons considéré que le paramètre augmentait (codage 1) ou diminuait (codage -1) (cf. Figure 160). Le déplacement des voyelles dans le plan F1/F2 est par nature toujours positif. Cependant, nous avons considéré chaque voyelle séparément et codé par 1 un déplacement allant dans le sens d'une expansion du triangle vocalique dans le plan F1/F2 (que nous appellerons par la suite « Prototypage » de la voyelle), et par -1 un déplacement allant dans le sens de d'une réduction. Le code 0 a toujours été attribué aux déplacements formantiques inférieurs au seuil de 5%.

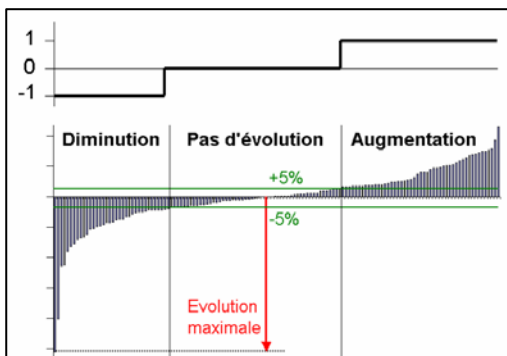


Figure 160. Schématisation de la méthode utilisée pour quantifier les différences observées pour un paramètre entre chaque répétition et son énonciation initiale, afin d'observer les tendances globales d'évolution de chaque paramètre pour chaque répétition.

Cette méthode nous a permis d'établir un profil « codé » (par des -1, des 0 ou des 1) de chaque répétition, dans un espace multiparamétrique, permettant une meilleure lisibilité des stratégies de répétition.

1.2. Résultats

Une classification ascendante hiérarchique à partir de ces profils a permis de dégager 5 stratégies principales de répétition dans le bruit (cf. Figure 161):

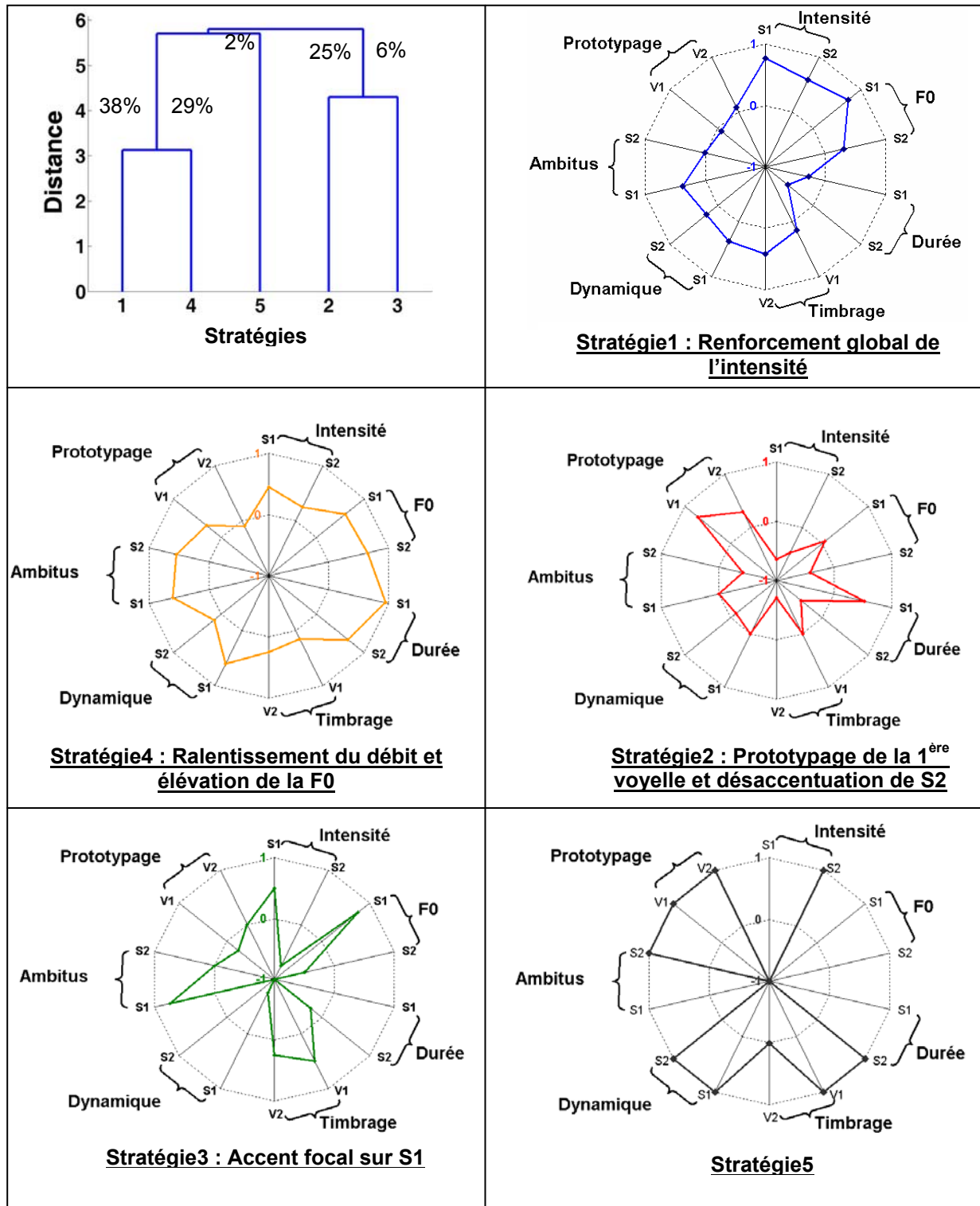


Figure 161. Illustration des 5 différentes stratégies observées de répétition d'un mot dans le bruit. Une valeur moyenne supérieure à 0.5 correspond à une augmentation du paramètre pour la majorité des répétitions de cette stratégie, une valeur inférieure à -0.5 correspond à une diminution du paramètre pour la majorité des répétitions de cette stratégie. S1 et S2 sont des abréviations pour 1^{ère} et 2^{ème} syllabe du mot-cible, tout comme V1 et V2 pour 1^{ère} et 2^{ème} voyelle du mot-cible.

- La première stratégie est adoptée dans 38% des cas. Elle consiste à parler plus fort de façon globale aussi bien sur la 1^{ère} que sur la 2^{ème} syllabe. L'augmentation de l'intensité s'accompagne de l'augmentation conjointe de la F0 moyenne, surtout sur la première syllabe et d'une accélération

- du débit sur la 2^{ème} syllabe. Le timbrage, l'ambitus de F0 ou la dynamique d'intensité n'évoluent pas particulièrement, les formants vocaliques non plus.
- La quatrième stratégie est adoptée dans 29% des cas. Elle consiste avant tout à ralentir le débit global de la parole, et à élever la F0 moyenne globale. Une augmentation de l'ambitus de F0 accompagne cette stratégie, de même qu'un renforcement de la dynamique d'intensité sur la 1^{ère} syllabe. L'intensité et le timbrage global n'augmentent pas notablement. Les formants vocaliques n'évoluent pas particulièrement non plus.
 - La deuxième stratégie est adoptée dans 25% des cas. Elle consiste surtout à prototyper la 1^{ère} voyelle et à allonger la première syllabe. Pour autant, cette stratégie ne s'accompagne pas d'un accent focal sur la 1^{ère} syllabe : la F0 moyenne n'évolue pas particulièrement et l'intensité au contraire diminue notablement sur la première syllabe. Par contre, la deuxième syllabe est désaccentuée : l'intensité, la F0 moyenne, le timbrage, l'ambitus de F0 diminuent sur la deuxième syllabe.
 - La troisième stratégie est adoptée dans 6% des cas. Elle consiste au contraire à accentuer la 1^{ère} syllabe : augmentation de l'intensité, de la F0, du timbrage et de l'ambitus de F0, diminution de la dynamique d'intensité et raccourcissement de la 1^{ère} syllabe. En contraste, la 2^{ème} syllabe tend à être désaccentuée (diminution de l'intensité, de la F0 et de la dynamique d'intensité).
 - Enfin, la cinquième stratégie est très marginale et ne correspond qu'à un cas isolé. Il est donc difficile de tirer des généralités. Mais elle semble consister à renforcer la 2^{ème} syllabe (augmentation de l'intensité, de la durée, de l'ambitus de F0 et de la dynamique d'intensité, mais diminution de la F0). Au contraire, la 1^{ère} syllabe tend à être désaccentuée en intensité, F0, durée et ambitus de F0. Son timbrage et sa dynamique d'intensité sont par contre renforcés.

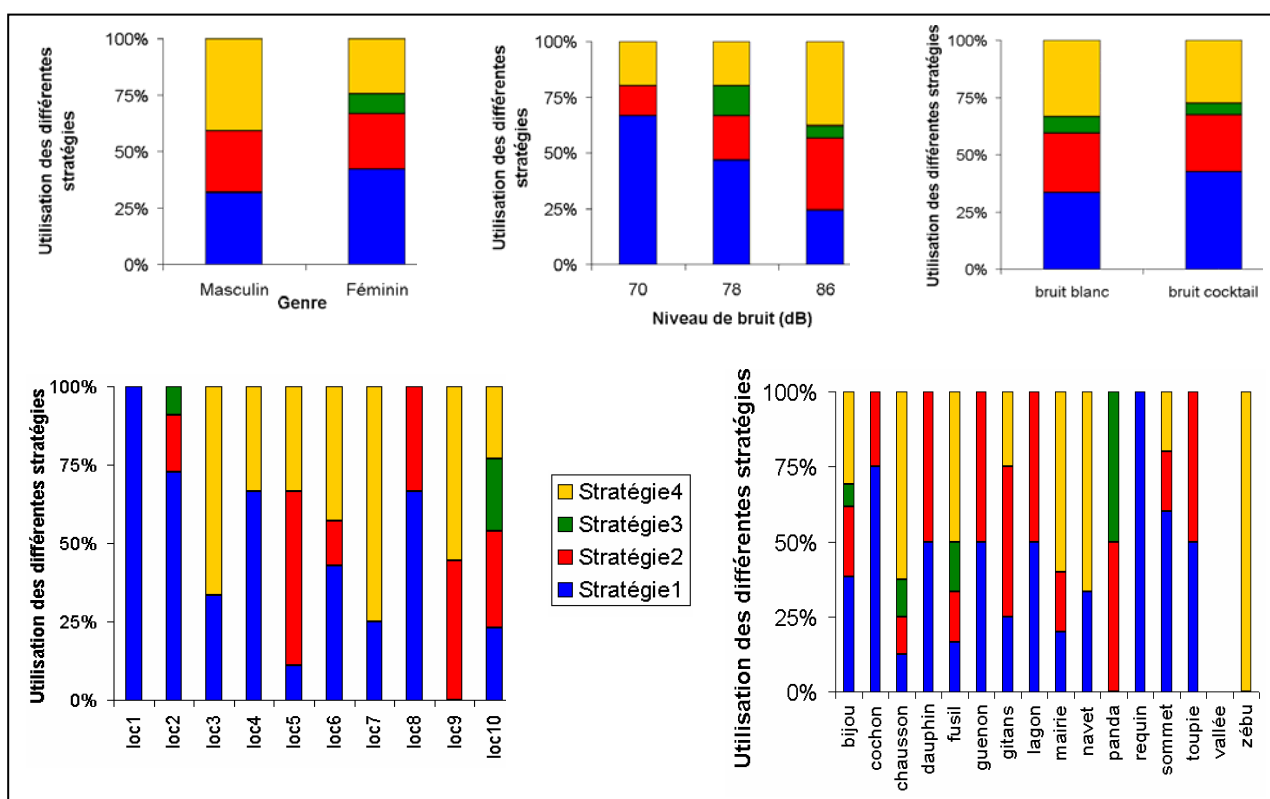


Figure 162. Utilisation des différentes stratégies de répétition d'un mot dans le bruit, selon le genre du locuteur, le niveau de bruit ambiant, le type de bruit, le locuteur ou encore le mot répété. La 5^{ème} stratégie, marginale, n'a pas été représentée.

Nous avons également essayé de comprendre si l'utilisation d'une stratégie plutôt qu'une autre était influencée par des caractéristiques individuelles (locuteur, genre) ou contextuelles (niveau de bruit, type de bruit, mot prononcé). Ainsi, nous avons observé que la 3^{ème} stratégie est uniquement utilisée dans ce corpus par les locutrices, et que celles-ci se servent majoritairement de la 1^{ère} stratégie, tandis que les locuteurs masculins utilisent davantage la 4^{ème} stratégie.

La 3^{ème} stratégie n'est utilisée qu'à partir de 78 dB de bruit ambiant. La première stratégie est la plus utilisée à moyen niveau de bruit et laisse la place à fort niveau de bruit à la 2^{ème} et la 4^{ème} stratégie. On n'observe pas de différence d'utilisation des différentes stratégies dans le bruit blanc et le bruit cocktail. À part la locutrice 1, utilisant exclusivement la 1^{ère} stratégie, tous les locuteurs utilisent au moins 2, voire jusqu'à 4 stratégies, dans des proportions différentes. Les locuteurs 3 (H), 4(H), 6(F) et 7(H) présentent une certaine similarité dans leur utilisation des différentes stratégies. Les locuteurs 5(H) et 9(F) aussi, de même que les locuteurs 2(F) et 8(H). Les locutrices 1 et 10 se distinguent à part. On remarque que ces similarités n'ont pas de relations avec le genre des locuteurs. Enfin, on observe que l'utilisation des différentes stratégies n'est a priori pas dépendante du mot répété.

Pour finir, nous avons examiné avec attention les cas de répétition successive, c'est à dire lorsque la première répétition n'est toujours pas comprise par l'interlocuteur et donne lieu à une deuxième voire une troisième répétition du mot. Ces cas ne sont pas nombreux et nos remarques resteront ici prudentes et qualitatives. Nous avons observé que les locuteurs ont plutôt tendance à surenchérir sur leur première stratégie lors de la deuxième répétition, tandis qu'ils ont tendance à changer de stratégie à partir de la troisième répétition, témoignant d'une réévaluation de la situation et d'une réadaptation.

1.3. Conclusion

Bien qu'imparfait, ce corpus représente un exemple de ce que peut apporter l'utilisation d'une tâche semi-spontanée. De telles interactions ne peuvent en effet avoir lieu que dans un contexte où un interlocuteur est présent et où le discours du locuteur n'est pas prévisible à l'avance. Il ne pourrait être obtenu de façon réaliste en tâche de lecture. Notre questionnaire ne portant pas initialement sur ces répétitions, ce corpus a été acquis ici de façon imprévue. Maintenant que nous avons vu l'intérêt de ces répétitions, il serait intéressant de réadapter le principe de nos jeux interactifs afin de mieux contrôler l'acquisition d'un tel corpus de répétitions.

Malgré la taille restreinte du corpus, nous obligeant à rester à des considérations qualitatives, les premières observations permettent de soutenir notre propos dans cette thèse car elles montrent l'influence du dialogue avec un interlocuteur dans le bruit. L'adaptation du locuteur n'est pas seulement influencée par la présence de l'interlocuteur (cf. Chapitre 4), mais également par ses retours de compréhension ou d'incompréhension. Lorsque l'interlocuteur ne comprend pas le message, nous avons ainsi observé que le locuteur ne se contente pas de le lui répéter de la même façon mais modifie sa façon de parler, autrement dit il semble réévaluer la situation de communication et s'y réadapter. Cette réadaptation ne concerne pas seulement l'intensité ou la F0 globale du message mais peut privilégier une syllabe du mot-cible (souvent la première), ou concerner des paramètres considérés comme des indices de reconnaissance des voyelles (déplacement des formants vocaliques).

Enfin, ce corpus est très riche d'informations. Son observation nous a donné la plupart des pistes concernant les stratégies d'adaptation de la parole en environnement bruyant que nous avons explorées dans les chapitres suivants :

- en premier lieu, les analyses de ce corpus semblent aller dans le sens de notre hypothèse d'existence de plusieurs stratégies de gestion de l'effort vocal en environnement bruyant, effort motivé par le désir d'être compris par l'interlocuteur.
- au contraire, ce corpus semble infirmer notre hypothèse initiale d'existence de stratégies d'adaptation propres à chaque locuteur ou du moins à des groupes de locuteurs. Il semble plutôt exister un « panel » de stratégies possibles dans lequel chaque locuteur en utilise certaines, voire toutes, mais rarement une seule.
- le recours à certaines stratégies de ce panel plutôt que d'autres ne semble pas dépendre du mot prononcé mais semble au contraire influencé par le contexte, en particulier le niveau de bruit ambiant.
- de la même façon, une recherche d'intelligibilité croissante ne semble pas être réalisée uniquement par le renforcement d'une même stratégie, mais parfois par un changement de stratégie. Nous nous sommes d'ailleurs ici limités à l'analyse des mots-cibles, mais nous avons observé que la stratégie de répétition peut parfois consister en une réorganisation de l'énoncé (répétition du mot de contenu uniquement, ou reformulation du message avec des mots plus simples). Ces observations nous amènent à douter que l'adaptation de la parole dans le bruit puisse être décrite de façon graduelle sur une unique dimension. Ce phénomène semble en partie graduel mais également catégoriel. Par conséquent, il n'est pas évident qu'il puisse être non plus décrit de façon linéaire en fonction du niveau de bruit ambiant.

2. INFLUENCE DE CETTE INCOMPREHENSION A MOYEN TERME

Nous avons vu que l'incompréhension de l'interlocuteur influence l'adaptation immédiate du locuteur sur le mot qui n'a pas été compris. Nous avons ensuite cherché à examiner si cette influence perdure au-delà de la réadaptation immédiate, ou bien si le locuteur revient à sa stratégie initiale après s'être fait comprendre.

2.1. Matériel et méthode

Pour cela, nous avons toujours considéré la première base de données, constituée de 10 locuteurs dialoguant par binômes dans différents niveaux et types de bruit (cf. chapitre 3). Nous avons examiné la différence entre les mots-cibles prononcés avant et après un signe d'incompréhension de l'interlocuteur, pour plusieurs paramètres de description de la parole : l'intensité et la F0 moyenne, la durée des mots-cibles, l'ambitus de F0, la dynamique d'intensité, la différence d'énergie des voyelles entre la zone [0-2kHz] et la zone [2-5kHz], aussi appelé timbrage par la suite, et le quotient ouvert).

Lorsque aucune incompréhension n'a eu lieu lors d'une condition expérimentale (c'est le cas de toutes les conditions de silence), aucun mot n'a donc été considéré. Au contraire, lorsque plusieurs incompréhensions ont eu lieu au cours d'une condition expérimentale, nous n'avons considéré que l'ensemble des mots-cibles avant le signe d'incompréhension et l'ensemble des mots cibles prononcés entre cette incompréhension et la suivante (en excluant la répétition immédiate induite par cette incompréhension).

La valeur moyenne de chaque paramètre a été ainsi calculée sur l'ensemble des mots avant et après chaque incompréhension. Nous avons regroupé dans un même vecteur les différentes conditions expérimentales et les différents locuteurs. Nous avons examiné l'influence du signe d'incompréhension de l'interlocuteur à l'aide d'un test statistique Anova.

2.2. Résultats

Aucun paramètre ne s'est montré significativement influencé à moyen terme par les marques d'incompréhension de l'interlocuteur. Ainsi, l'intensité moyenne diminue de -0.38 dB ($F(1)=0.12$, ns) après une demande de répétition, la F0 moyenne augmente de 0.1 ton ($F(1)=0.03$, ns), l'ambitus de F0 diminue de 0.1 ton ($F(1)=0.45$, ns), la dynamique d'intensité diminue de 0.3 dB ($F(1)=0.15$, ns), l'énergie entre 2 et 5 kHz augmente de 0.3 dB ($F(1)=0.03$, ns) et le quotient ouvert reste stable ($F(1)=0.03$, ns).

2.3. Conclusion

Contrairement à nos hypothèses et à ce que les résultats du paragraphe précédent auraient pu laisser penser, l'adaptation à moyen terme du locuteur ne semble pas influencée par le retour de l'interlocuteur, au-delà de la répétition immédiate du mot incompris.

En premier lieu, ces résultats peuvent être interprétés comme le signe que l'effet Lombard n'est pas influencé par le retour de l'interlocuteur. L'effet à court terme observé suite à une incompréhension du locuteur pourrait alors être « simplement » un effet de focalisation, indépendant de la présence de bruit, similaire à la focalisation classiquement observée dans le silence. Le corpus de répétitions examiné au paragraphe précédent ne comporte pas de répétitions dans le silence. Il ne nous est donc pas possible ici de conclure sur ce point. Cependant, nous avons observé que l'utilisation des stratégies de répétitions était différente selon le niveau de bruit. Il semble donc quand même que le retour de l'interlocuteur n'influence pas l'adaptation du locuteur de la même manière selon le niveau de bruit, donc certainement aussi entre une condition silencieuse et bruyante. Cela reste en tout cas à vérifier sur un corpus dédié à ce questionnement. Cependant, l'idée de focalisation en condition silencieuse nous semble non naturelle, excepté si d'autres éléments de la situation perturbent la communication (distance des interlocuteurs, interlocuteur de langue étrangère, etc). Ce n'est pas un hasard si notre corpus spontané ne comporte aucune répétitions dans le silence

Nous avons également envisagé que le forçage vocal soit justement un dysfonctionnement de l'estimation de l'intelligibilité à fournir. Il est donc possible que les différences interindividuelles de stratégies de répétition observées proviennent en partie de différences individuelles d'évaluation de l'intelligibilité à fournir, dues au fait que tous les locuteurs ne se baseraient pas sur les mêmes indices

pour estimer leur intelligibilité. Nous avons voulu vérifier ce point rapidement. Pour cela, nous avons calculé pour chaque locuteur les corrélations statistiques entre le nombre de répétitions par condition et la façon dont chaque locuteur percevait sa propre voix (question1), la façon dont ils estimaient leur intelligibilité (question2), la façon dont leur interlocuteur les a effectivement perçus (question3 de l'interlocuteur) et la façon dont ils percevaient l'interlocuteur (question3). Nous avons également calculé les corrélations entre la façon dont les locuteurs ont estimé leur intelligibilité (question2) et la question1, la question 3 et la question 3 de l'interlocuteur. Les résultats de ces corrélations sont présentés dans le Tableau 37.

	Estimation de son intelligibilité * Nombre de répétitions	Estimation de son intelligibilité * Perception de sa propre voix	Estimation de son intelligibilité * Perception de la voix de l'interlocuteur	Estimation de son intelligibilité * Perception <u>par</u> l'interlocuteur	Nombre de répétitions * Perception <u>par</u> l'interlocuteur
L1	0.36	0.26	0.55	0.60	0.52
L2	0.36	0.65	0.79	0.71	0.45
L6	0.55	0.72	0.91	0.68	0.59
L9	0.51	0.92	0.99	0.70	0.49
L10	0.60	0.68	0.73	0.73	0.49
L3	0.59	0.68	0.61	0.71	0.41
L4	0.40	0.56	0.63	0.51	0.40
L5	0.65	0.72	0.83	0.73	0.19
L7	0.38	0.79	0.94	0.55	0.63
L8	0.38	0.80	0.83	0.53	0.13

Tableau 37. Corrélations statistiques pour chaque locuteur entre le nombre de mots non compris par l'interlocuteur et demandés d'être répétés, et les différentes évaluations subjectives du locuteur sur la perception de sa propre voix, de l'interlocuteur, sur l'estimation de son intelligibilité. Les corrélations supérieures à 50% sont indiquées en orange. Celles supérieures à 65% sont indiquées en rouge.

L'estimation du locuteur de sa propre intelligibilité n'est que partiellement corrélée au nombre de mots n'ayant pas été compris par l'interlocuteur. L'interlocuteur lui-même ne semble pas non plus tellement juger l'intelligibilité du locuteur sur cette unique base. La majorité des locuteurs, excepté L1, montrent une estimation de leur propre intelligibilité davantage corrélée à la perception de leur propre voix et surtout à leur perception de l'interlocuteur. L'existence d'une forte corrélation ne prouve en aucun cas un lien de causalité entre deux facteurs. Enfin, il existe effectivement des locuteurs qui estiment mieux que d'autres la façon dont leur interlocuteur les perçoit effectivement : L1 et L6 par exemple arrivent relativement bien à estimer la façon dont l'interlocuteur les a perçues, alors que L5 ou L8 surévaluent ou sous-évaluent leur intelligibilité. Pour conclure, plusieurs indices peuvent servir au locuteur pour estimer sa propre intelligibilité. Manifestement, certains locuteurs se basent moins que d'autres sur les retours de compréhension de l'interlocuteur pour cela. Cependant, si on ré-effectue des tests statistiques pour évaluer l'influence de l'incompréhension de l'interlocuteur sur l'adaptation à moyen terme de chaque locuteur, les résultats restent non significatifs pour tous les locuteurs.

Enfin, une dernière interprétation peut être proposée de ces résultats : il est envisageable que le locuteur ait besoin de plusieurs expériences par essais-erreurs pour pouvoir associer une conséquence à son action. Une incompréhension isolée de l'interlocuteur ne suffirait pas à ce que le locuteur assimile cette expérience de réadaptation. Il continue alors sur sa stratégie initiale.

Quoi qu'il en soit, ces résultats vont plutôt à l'encontre de nos hypothèses d'adaptation du locuteur aux retours d'incompréhension de l'interlocuteur. Par contre, ils garantissent la validité des analyses réalisées sur le corpus principal de la 1^{ère} base de données, puisque la prononciation de tous les mots d'une même condition peut être considérée comme homogène.

Annexe Pil2 :

**Corpus et étude articulatoire
pilote**

- 1. CONSTITUTION D'UN CORPUS ARTICULATOIRE
PILOTE**
- 2. COMMUNICATION A ICSLP 2006,
PITTSBURGH**
- 3. COMMUNICATION A ISSP 2006, UBATUBA**

La conception, l'enregistrement et la segmentation du corpus pilote a fait l'objet du stage de DEA de Lucie Bailly 2005 ^[21]. L'étude articulatoire a été faite en collaboration avec Marion Dohen, Hélène Loevenbruck et Pauline Welby de l'ICP à Grenoble.

Cette étude exploratoire avait pour but d'investiguer plusieurs points :

- Tout d'abord, nous avons cherché à examiner si l'adaptation de la parole en environnement bruyant s'accompagnait d'une modification des mouvements articulatoires. Si tel est le cas, s'agit-il d'une hyper-articulation ou d'une réorganisation articulatoire? Les éventuelles modification articulatoires concernent-elles l'amplitude et/ou la vitesse des mouvements? Enfin, tous les types de mouvements articulatoires sont-ils modifiés de façon comparable ou différente ?
- De plus, toujours dans le but d'explorer l'aspect adaptatif et non réflexe de la parole produite en environnement bruyant, nous avons voulu examiner si le type de bruit avait une influence sur les mouvements articulatoires.
- Enfin, nous nous sommes demandées si les modifications articulatoires étaient globales sur tout l'énoncé ou bien si elles étaient plus spécifiques à certains constituants de l'énoncé (syntaxiques, syllabiques...)
- Pour finir, cette base de données a été également constituée pour permettre de répondre à une étude intonative parallèle menée par Welby 2005 ^[374] cherchant à voir si les indices intonatifs de segmentation lexicale étaient renforcés dans le bruit.

Ce corpus, comprenant seulement une locutrice, a été préalable à la constitution d'une base de données plus vaste sur plusieurs locuteurs et ciblée sur des points intéressants émergeant de cette étude pilote (cf. Deuxième base de données BD2).

Ce travail exploratoire a donné lieu à plusieurs publications (Garnier et al. 2006 ^[103] ; Garnier et al. 2006 ^[102] ; Garnier et al. 2006 ^[104]), dont deux sont présentées en paragraphe 2 et 3 de cette annexe.

1. CONSTITUTION D'UN CORPUS ARTICULATOIRE PILOTE

1.1. Matériel linguistique

Choix des 16 syntagmes cibles

Afin de garder un matériel linguistique comparable à la première base de donnée, nous avons de nouveau choisis des syntagmes cibles construits à partir de **syllabes de structure CV** (Consonne-Voyelle).

Par ailleurs, l'étude intonative menée en parallèle sur ce corpus a introduit un certain nombre de contraintes sur le choix des syntagmes cibles :

- Tout d'abord, seules des **consonnes sonores** ont été retenues afin de minimiser les ruptures de la courbe de fréquence fondamentale. Les segments /m/, /n/, /l/, /R/ et plus occasionnellement /v/ ont été utilisés.
- Ensuite, nous souhaitons examiner si les « Early rise » pouvaient être renforcés dans le bruit. Cet indice intonatif de segmentation lexicale (Welby 2003^[373]) est davantage présent dans le silence sur des syntagmes de longueur égale ou supérieure à 3 syllabes. Nous nous demandions s'il pouvait exister dans le bruit sur des syntagmes plus courts de 2 syllabes. De plus, l'alignement tonal du early rise dépend du fait que le syntagme est uniquement composé d'un mot de contenu ou bien également d'un déterminant. Pour ces différentes raisons, nous avons retenus des **syntagmes cibles de 2 et 3 syllabes, composés ou non d'un déterminant**.

La résolution de ces différentes contraintes a été effectuée grâce au logiciel Lexique (New 2006^[254]). Nous avons finalement retenu **16 syntagmes-cibles** représentant 4 catégories d'analyse :

Syntagmes à 2 syllabes de contenu

- avec déterminant : **La mamie, le minet, les moulins, les lamas** ;
- sans déterminant : **Loulou, Maman, Manon, Ninon** ;

Syntagmes à 3 syllabes de contenu

- avec déterminant : **Les monuments, les moulinets, la nominée, le minaret** ;
- sans déterminant : **Mélanie, Marilou, Mélina, Marina**.

Choix d'une structure syntaxique fixe de type Sujet-Verbe-Objet

Afin de comparer les syntagmes cibles dans différentes positions de l'énoncé, nous avons choisi une même structure syntaxique pour toutes les phrases du corpus.

Le choix de cette structure s'est arrêté sur une syntaxe de type **Sujet-Verbe-Objet (SVO)** pour deux raisons :

Une étude antérieure [Dohen et al., 2003] a montré une variabilité plus faible de la réalisation des gestes articulatoires sur une telle structure.

De plus, cette structure a l'avantage de ne présenter aucune ambiguïté, ce qui garantit l'unicité de l'interprétation de la phrase par le locuteur et par conséquent, la comparabilité des formes intonatives produites.

Positions initiale, intermédiaire et finale des syntagmes dans l'énoncé

Dans le but d'examiner si les modifications articulatoires dans le bruit sont globales sur tout l'énoncé ou différentes selon la position du syntagme dans l'énoncé (cf. But 3), nous avons construit 33 phrases permettant, par combinaison, de représenter les 16 syntagmes cibles en 3 positions de l'énoncé :

Position initiale : par exemple « *Loulou nourrit le minet* »

Position intermédiaire : par exemple « *Nina et Loulou mimaient les lamas* »

Position finale : par exemple « *Maman ramena Loulou* »

Construction des 33 phrases du corpus

Les phrases ont été entièrement construites à partir d'enchaînements de syllabes de structure CV. De même, le choix de consonnes sonnantes a été étendu à l'ensemble des phrases et non pas seulement aux syntagmes cibles.

Bien que pour cette étude pilote, la priorité ait été accordée au contrôle du matériel linguistique au détriment de la spontanéité du discours, nous avons essayé autant que possible de construire des phrases ayant du sens pour le locuteur. De plus, nous avons fait varier le verbe de la phrase afin d'éviter que la locutrice tende à focaliser sur les syntagmes cibles changeant.

Les **33 phrases** du corpus étaient donc les suivantes :

1. Maman ramena Loulou.
2. Lili et Maman louent les monuments.
3. Mélina lavera Manon.
4. Le minaret rallie la mairie.
5. Les lamas ramenaient Maman.
6. Mélanie ramena les moulinets.
7. Nina et Loulou mimaient les lamas.
8. Ninon lira le roman.
9. Le loup et le minet levaient le nez.
10. Marilou mima Mélanie.
11. Mélanie et Ninon loueront le minaret.
12. La mamie lavait le muret.
13. Loulou nourrit le minet.
14. Les mulets et les lamas rallieront la vallée.
15. Momo et Marilou lavaient les moulins.
16. Marina nourrit les mulets.
17. Manu et la nominée menaient la mamie.
18. Le monument et le minaret rallient la vallée.
19. Loulou et Marina liront le roman.
20. Nina et Mélanie ramenaient Ninon.
21. Manon leva le nez.
22. Le monument et les moulins rallient la vallée.
23. Les monuments rallient le muret.
24. Ninon et la mamie mimaient Marilou.
25. Lili et Mélina ramenaient Marina.
26. Le minaret et les monuments rallient la vallée.
27. La nominée mena Mélina.
28. Lola lavera la mamie.
29. Mélanie et Manon menaient la nominée.
30. Les moulinets ramenaient les morues.
31. Le minet leva le nez.
32. Le minaret et les moulinets rallient la vallée.
33. Les moulins rallient la mairie.

1.2. Tâche

L'expérimentateur était assis à 1,5 m en face de la locutrice et lui présentait sur une feuille les phrases du corpus à lire dans un ordre aléatoire, différent selon les conditions expérimentales.

1.3. Sujet

La locutrice, L11, âgée de 24 ans, était naïve concernant la problématique, et ne présentait aucun troubles de la parole ni de l'audition.

1.4. Conditions expérimentales variées

Les enregistrements ont eu lieu dans une pièce insonorisée (*grâce à l'aimable accueil de l'ICP*). Nous avons tout d'abord enregistré la locutrice en **condition calme de référence**. Puis, toujours dans le but d'examiner l'influence potentielle des caractéristiques spectrales de l'environnement sur l'adaptation de la parole, nous l'avons enregistré dans deux types de bruit différents

- un **bruit blanc**, plat en fréquence jusqu'à 10kHz puis atténué au delà de 10kHz.
- un **bruit de conversation**, ou bruit de « **cocktail-party** » mixte inintelligible, d'énergie plutôt basse-fréquences, située en dessous de 800Hz.

Ces deux types de bruit sont identiques à ceux utilisés pour la première base de données, et sont issus de la base de données BD_Bruit (Zeiliger et al. 1994^[382]). Ils ont été diffusés à un niveau de 85dB SPL par le biais de deux hauts-parleurs (A2t, de puissance maximale 200 Watts) situés à 2 m de la locutrice (cf. Figure 163), après avoir été amplifiés par un amplificateur (JVC Ax40). Le niveau d'intensité des bruit a été calibré à l'aide d'un sonomètre au niveau des oreilles de la locutrice.

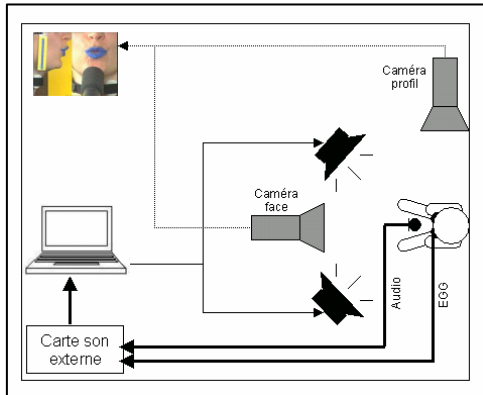


Figure 163. Schéma du dispositif expérimental

1.5. Mesures effectuées

Afin de réaliser des analyses acoustiques et phonétiques de la parole des locutrices, nous avons enregistré leur signal audio de parole à l'aide d'un microphone à électret (AKG C1000S), situé à 50cm de la bouche de la locutrice. Pour calibrer le niveau d'intensité acquis au microphone, nous avons enregistré le locuteur en début d'enregistrement sur un [a] tenu stable et mesuré cette intensité vocale de référence au sonomètre au niveau des lèvres du locuteur. Ce signal a ensuite été numérisé sur 16 bits à 44.1 kHz sur la première voie d'une carte son externe (Sound Blaster Extigy)

Afin d'étudier la vibration des cordes vocales, nous avons acquis à l'aide d'un électroglottographe à deux paires d'électrodes (Glottal Enterprise EG2) le signal d'accolement des cordes vocales, que nous désignerons par la suite par l'abréviation EGG a été directement numérisé sur 16 bits à 44.1 kHz sur la deuxième voie de la carte son externe (Sound Blaster Extigy).

Afin d'explorer les mouvements articulatoires labiaux, nous avons effectué des enregistrements vidéo de face et de profil du visage des locutrices et extrait leurs mouvements articulatoires au niveau des lèvres grâce à un système de « Lip Tracking » explicité en détail en Annexe Met1.

2. COMMUNICATION A ICSLP 2006, PITTSBURGH

An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance

Maëva Garnier¹, Lucie Bailly^{1,2}, Marion Dohen^{2,3}, Pauline Welby², H el ene L evenbruck²

¹Laboratoire d'Acoustique Musicale, CNRS UMR 7604, UMPC, Minist ere de la culture, Paris, France

²Institut de la Communication Parl ee, CNRS UMR 5009, INPG, Univ. Stendhal, Grenoble, France

³ATR – Cognitive Information Science Laboratories, Kyoto, Japan

garnier@lam.jussieu.fr, {lucie.bailly, pauline.welby, helene.loevenbruck}@icp.inpg.fr, marion.dohen@gmail.com

Abstract

This study aims at characterizing the acoustic and articulatory modifications that occur in speech in noisy environments, and at examining them as compensatory strategies. Audio, EGG and video signals were recorded for a female native speaker of French. The corpus consisted of short sentences with a subject-verb-object (SVO) structure. The sentences were recorded in three conditions: silence, 85dB white noise, and 85dB cocktail party noise. Labial parameters were extracted from the video data. The analyses enabled us to examine the effect of the type of noise and to show that hyper-articulation concerns lip aperture and spreading rather than lip pinching. The analysis of the relationship between acoustic and articulatory parameters shows that this speaker especially adapts to noise not only by talking louder or increasing vowel recognition cues but also by increasing spectral emergence.

Index Terms: Lombard speech, noise, acoustic, articulation.

1. Introduction

Communication in noisy environments is disturbed for two main reasons: speakers get attenuated feedback of their own voices, and their intelligibility is decreased for the listener. The speech compensatory adaptation to these perturbations is called the Lombard effect, and has been mainly described from acoustic and phonetic points of view [1-4]. In this study, we aim at characterizing the articulatory gestures that accompany these acoustic modifications in noise and at comparing them to speech produced in silence. Is there a hyper-articulation? Which articulatory parameters are concerned by this speech adaptation? Are the articulatory and acoustic modifications correlated? Lastly, does the articulatory adaptation depend on the type of noise? Some studies have already described articulation of "hyper" types of speech [5-6] but few specifically examine Lombard speech [7-8].

2. Methods

2.1. Corpus

The corpus consisted of 33 short sentences with a subject-verb-object (SVO) structure. They consisted of CV syllables (see (1)) and voiced consonants were chosen in order to minimize the segmental perturbations of the intonation contour.

(1) *Le monument et les moulins rallient la vall ee.*
'The monument and the mills are rallying the valley.'

2.2. Audiovisual recordings

We simultaneously recorded audio, electroglottographic (EGG) and articulatory signals from a female native speaker of French. She read sentences to a person standing two meters in front of her. Articulatory data were extracted from video recordings (25 images/s) of the speaker's lips, using a labiometric device developed at the Institut de la Communication Parl ee [9]. In this study, we focused on the analysis of lip **spreading (A)**, lip **aperture (B)**, and **inter-lip area (S)** (see Figure 1). For these three parameters, we examined the amplitude maxima of the articulatory movements (**max**), as well as their more global evolution (**glob**), which corresponds to the integral over a normalized time period [9, pp. 113- 114].

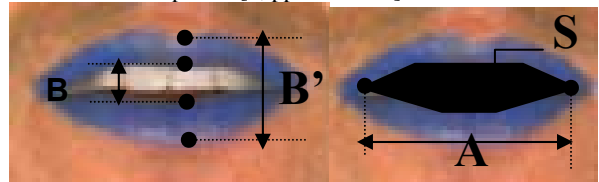


Figure 1. Articulatory parameters.

We also analyzed **lip pinching**, defined as lip compression when the mouth is closed for [m] segments. It is equivalent to B' when B is equal to zero (see Figure 1). Two types of lip pinching could be observed: **protruded lip pinching**, i.e. when the visible labial surface increases with lip compression, and **swallowed lip pinching**, i.e. when the lips flatten and move towards the inside of the mouth, resulting in a decrease of the visible labial surface.

The audio signal was recorded with an AKG microphone placed 20cm away from the lips. The EGG signal was recorded with an electroglottograph EG2. Both signals were digitized at a rate of 44.1kHz, over 16bits. Two noisy environments were used, both extracted from the BD_Bruit database [11]: white noise (**wn**) and cocktail party noise (**cktl**). They were played over two loudspeakers located 2m away from the speaker and 2m away from each other. The noise level was calibrated to 85dB at the participant's ears. The speaker was first recorded in a silent

reference condition, and then in both noisy environments. Noise was removed from the acoustic signal using the method designed by [12]. The utterance boundaries were then labelled using Praat [13], as well as word and segment boundaries. Fundamental frequency (F0) and open quotient (Oq) were extracted from the EGG signal using an autocorrelation method.

3. Results

The following notation was adopted for indicating statistical significance: * $p < .05$, ** $p < .01$, *** $p < .001$, and *ns* (not significant) $p > .05$.

3.1. Articulation of the utterances in noise

3.1.1. Amplitude of the articulatory movements

A significant increase of A, B and S was observed in noisy conditions for both types of noises. Protruded lip pinching decreased significantly in white noise (see Table 1). The type of noise had a significant effect on the increase of all articulatory parameters, except swallowed lip pinching. This increase of the articulatory parameters (henceforth called "hyper-articulation") was globally greater in cocktail party noise than in white noise.

Table 1. Evolution of the amplitude of the articulatory movements in the two noisy environments.

		Δ in wn	Δ in cktl	Effect of the type of noise. Δ cktl- Δ wn
A	max	+3.4% ***	+6.5% ***	+3% **
	glob	+8.9% ***	+14.2% ***	+5.3% ***
B	max	+19.5% ***	+27.5% ***	+7.9% ***
	glob	+16.0% ***	+27.1% ***	+11.2% ***
S	max	+28.4% ***	+42.3% ***	+13.9% ***
	glob	+25.3% ***	+43.6% ***	+18.3% ***
Protruded pinching	max	-16.3% ***	-11.0% *	+27.3% ***
	glob	-36.7% ***	-10.2% *	+26.5% ***
Swallowed pinching	max	+83.0% **	+95.5% **	+12.5% ns
	glob	+149.2% **	+141.1% **	-8.1% ns

3.1.2. Velocity peaks of the articulatory movements

We observed a significant increase in the velocity peaks (Vmax) of A, B and S for both types of noise. The protruded lip pinching velocity peaks decreased significantly in white noise (see Table 2). The velocity peaks of B, S and protruded lip pinching increased significantly more in cocktail party noise than in white noise. However, the type of noise had no significant effect on the increase in the velocity peaks of A and swallowed lip pinching. Peak velocity has been related to the notion of articulatory effort [14].

Table 2. Evolution of the articulatory velocity peaks in the two noisy environments.

		Δ in wn	Δ in cktl	Effect of the type of noise. Δ cktl- Δ wn
Vmax of A		+12.7% **	+13.9% ***	+1.2% ns
Vmax of B		+14.9% ***	+29.9% ***	+15.0% ***
Vmax of S		+17.8% ***	+38.5% ***	+20.7% ***
Vmax of	protruded pinching	-16.5% ***	+3.3% ns	+19.8% **
	swallowed pinching	+120.1% **	+127.9% **	+7.8% ns

3.2. Evolution of acoustic parameters in noise

3.2.1. Over the whole utterance

Table 3 presents the mean values of the acoustic parameters over the whole utterance for all utterances. The acoustic results confirmed those obtained in previous studies of Lombard speech [1-3]: voice intensity, F0 and word duration increased significantly for both types of noise (see Table 3). We also observed a large decrease of Oq in noise, that can be related to a more pressed voice quality, and an increase in the difference between the spectrum maximum in the 80-1500Hz frequency band (first harmonics) and the spectrum maximum in the 1500-3500Hz frequency band. A comparable result was obtained for the energy difference between the 80-1500Hz frequency band and the 3500-5500Hz frequency band.

Table 3. Evolution of the acoustic parameters in the two noisy environments.

		Δ in wn	Δ in cktl	Effect of the type of noise. Δ cktl- Δ wn
Intensity		+12.9dB ***	+8.6dB ***	-4.3dB ***
F0		+55.5Hz ***	+65Hz ***	+9.5Hz ***
Spectral energy	1500-3500Hz	+23dB ***	+19dB ***	-4dB ***
	3500-5500Hz	+27dB ***	+19dB ***	-8dB ***
Oq		-0.123 ***	-0.118 ***	+0.005 ns
Word duration		+11% ***	+8.2% ***	-2.8% *

The type of noise had a significant effect on the evolution of the acoustic parameters in noisy conditions, except for Oq. The increase of voice intensity, spectral energy, and word duration was greater in white noise than in cocktail party noise. On the contrary, mean F0 increased more in cocktail party noise than in white noise.

3.2.2. For vowels and consonants

Vowel intensity increased more in noisy conditions than did consonant intensity (see Table 4). Vowel duration increased significantly in noise whereas consonant duration tended to decrease, as it has already been observed ([1-3]). On the contrary, spectral energy did not significantly increase more in noise for vowels than for consonants.

Vowel duration increased more in white noise than in cocktail party noise. The type of noise had a significant effect on the increase of spectral energy in the 3500-5500Hz frequency band, and for consonants, on the increase of spectral energy in the 1500-3500Hz frequency band. The reinforcement is greater in white noise than in cocktail party noise.

Table 4. Evolution of the acoustic parameters in noisy conditions for vowels and consonants.

	Intensity	Spectral energy		Word duration
		1500-3500Hz	3500-5500Hz	
Δ Vowels in wn	+12.9dB ***	+17.7dB ***	+24.1dB ***	25.2% ***
Δ Vowels in cktl	+8.4dB ***	+17.4dB ***	+13.3dB ***	16% ***
Difference cktl-wn for Δ Vowels	-4.5dB ***	-0.3dB ns	-10.8dB ***	-9.2% ***
Δ Consonants in wn	+8.5dB ***	+21.6dB ***	+28.2dB ***	-5.7% ns
Δ Consonants in cktl	+3.5dB ***	+19.6dB ***	+11.8dB ***	-3.6% ns
Difference cktl-wn for Δ Consonants	-5.0dB ***	-2.0% ***	-16.4dB ***	+2.1% ns
Difference Δ Vowels - Δ Consonants	+4.6dB ***	+2.9dB ns	+1.3dB ns	+25.2% ***

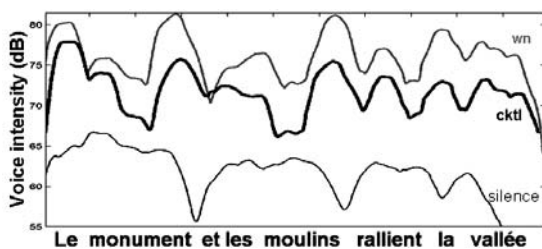


Figure 2. Voice intensity of an utterance in silence, white noise (wn) and cocktail party noise (cktl).

As expected, vowels were louder than consonants in silence (with an average increase of 4.5dB, $p < .001$). We especially observed that voice intensity followed the prosodic phrasing of the utterance, falling at prosodic phrase boundaries. In noisy conditions, we noticed that the dynamic of the voice intensity

"lobes" per syllable increased significantly (+4dB, $p < .001$). This could contribute to the sometimes "choppy" nature of Lombard speech (see Figure 2).

3.2.3. For the vowels /u/, /i/ and /a/

Table 5. Evolution of the acoustic parameters in noisy conditions for the vowels /a/, /i/ and /u/.

		Δ / u / in noise	Δ / i / in noise	Δ / a / in noise	Effect of the vowel type
Intensity		+9.6dB ***	+8.9dB ***	+14.6dB ***	4.5dB ***
Spectral energy	1500-3500Hz	+18.4dB ***	+20.5dB ***	+11.4dB ***	7.0dB ***
	3500-5500Hz	+21.5dB ***	+19.2dB ***	+13.8dB ***	5.7dB **
Oq		-0.111 ***	-0.115 ***	-0.092 ***	0.017 ns
Word duration		26.7% ***	23.2% ***	14.3% **	9.2% ns

Table 5 presents the mean values of the acoustic parameters over all the /a/, /i/ and /u/ segments of the corpus. These vowels correspond to the extrema of the vowel triangle in the F1/F2 plane. Voice intensity and spectral energy varied differently for these three vowels (see Table 5). On the contrary, the evolution of word duration and Oq did not significantly depend on the vowel considered.

Figure 3 shows that F1 tended to increase in noisy conditions, particularly for the open vowel /a/ and the rounded vowel /u/. This result confirms the observations made by Van Summers et al., 1988 for one of their speakers [3], and by Rostolland for shouted speech [16]. The second formant is slightly modified for /i/ and /a/ and increases considerably in noisy conditions for /u/.

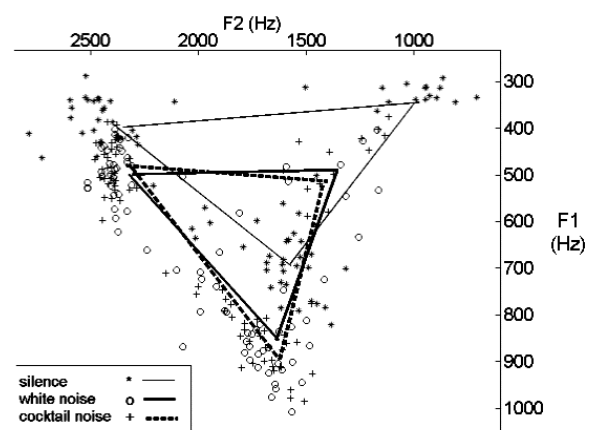


Figure 3. Modification of the vowel system in white noise and in cocktail party noise. (The corners of the triangles correspond to the barycenters of /a/, /i/ and /u/ for each condition, in the F1/F2 plane).

3.3. Correlations between acoustic and articulatory parameters

We observed a strong correlation between the maximum amplitudes of A and B over the whole utterance ($r = .91$), and also a strong correlation between their velocity peaks ($r = .70$). F0 and the spectral energy are correlated with maximal amplitude of B ($r = .66$ and $r = .70$ respectively). This correlation is stronger than the one between voice intensity and maximum amplitude of B ($r = .59$). Word duration is weakly correlated with other acoustic and articulatory parameters ($|r| < .15$).

4. Discussion and Conclusions

Hyper-articulation was observed when the speaker spoke in noisy environments. This amplification of the articulatory movements concerned lip aperture and spreading rather than lip pinching, not only for the movement amplitude, but also for the velocity peaks, which can be related to articulatory effort. We observed an associated evolution of the acoustic parameters that confirmed the observations of earlier studies [1-3]. Type of noise had a significant effect on the evolution of the acoustic parameters (the increase was greater in white noise than in cocktail party noise, except for F0). An influence of the spectral characteristics of noise has already been noticed in previous studies [2,12]. We observed here that it also had a significant effect on the evolution of the articulatory parameters (the increase was more important in cocktail party noise than in white noise). In addition, spectral energy was the parameter which increased the most. Lip aperture was more correlated with spectral energy than with voice intensity. Therefore it seems that acoustic and articulatory adaptations in noise include not only talking louder but also specifically increasing spectral emergence.

As it has already been noticed in previous studies, acoustic adaptation in noisy conditions favors vowels rather than consonants in general. This result is understandable for voiceless consonants, which are expected to be masked by the ambient noise. However, it is also observed for voiced consonants. This could be in line with the hypothesis of Dohalska [15], who suggested that vowels are used as "patterns" to help the listener reconstitute the intelligibility of the message. However, we note that the opposite observation has been made for clear speech, for which consonants are reinforced in comparison to vowels [5].

The observation of both articulatory and acoustic (formant values) data show that hyper-articulation does not seem to reinforce vowel recognition cues, except for the open vowel /a/. This could explain the decrease in intelligibility noticed for shouted speech [16]. A new corpus controlling for coarticulation has recently been recorded in order to rigorously explore the hyper-articulation corresponding to different vowels. This corpus will also enable us to additionally analyze lip protrusion. Several speakers have been recorded in order to analyze whether the results of the present study can be generalized. It would also be interesting to go further and explore whether there exists compensatory tongue movements for the increased aperture of the lips, especially for rounded vowels. Compensation strategies have indeed been observed for the vowel [u] produced with a lip-tube (increased lip aperture) [17].

5. Acknowledgments

We would like to thank N. Henrich, C. Savariaux, A. Arnal, A. Noiray, J. Poitevineau, C. Lalevée, and C. Vilain for their contribution to the project. Pauline Welby's participation was supported by a Marie Curie International Fellowship within the 6th European Community Framework Programme.

6. References

- [1] Castellanos, A., Benedi, J. M., and Casacuberta, F. "An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect", *Speech Comm.*, 20: 23–35, 1996.
- [2] Junqua, J. C. "The Lombard reflex and its role on human listener and automatic speech recognisers", *J. Acoust. Soc. Amer.*, 93: 510–524, 1993.
- [3] Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. "Effect of noise on speech production: acoustic and perceptual analyses", *J. Acoust. Soc. Amer.*, 84: 912–928, 1988.
- [4] Garnier, M., Henrich, N., Dubois, D., and J. D. Polack. "Est-il valide de considérer l'effet Lombard comme un phénomène linéaire en fonction du niveau de bruit ?", *Proc. Congrès Français d'Acoustique*, 2006.
- [5] Lindblom, B. "Speech transforms", *Speech Comm.*, 11: 357–368, 1992.
- [6] Schulman, R. "Articulatory dynamics of loud and normal speech", *J. Acoust. Soc. Amer.*, 85: 295–312, 1988.
- [7] Junqua, J. C. "Acoustic and production pilot studies of speech vowels produced in noise", *Proc. ICSLP*: 811–814, 1992.
- [8] Kim, J., Davis, C., Vignali, G., and Hill, H. "A visual concomitant of the Lombard reflex", *Proc. Listenery-Visual Speech Processing 2005*: 17–21, 2005.
- [9] Lallouache, M. T. "Un poste "Visage-Parole" couleur. Acquisition et traitement automatique des contours des lèvres", *Doctoral dissertation*, Institut National Polytechnique de Grenoble, 1991.
- [10] Dohen, M. "Deixis prosodique multisensorielle: production et perception audiovisuelle de la focalisation contrastive en français", *Doctoral dissertation*, Institut National Polytechnique de Grenoble, 2005.
- [11] Zeiliger, J. "BD_Bruit, une base de données de parole de locuteurs soumis à du bruit", *Proc. Xèmes Journées d'Étude sur la Parole*: 287–290, 1994.
- [12] Ternström, S., Sodersten, M. and Bohman, M. "Cancellation of simulated environmental noise as a tool", *J. Voice*, 16: 195–206, 2002.
- [13] Boersma, P. and Weenink, D. "Praat: doing phonetics by computer (Version 4.2.28) [Computer program]". Retrieved from <http://www.praat.org/>, 2004.
- [14] Nelson, W. L. "Physical principles for economies of skilled movements", *Biol. Cybern.*, 46: 135–147, 1983.
- [15] Dohalska, M. and Mejvaldova, J. "Rôle de la prosodie dans la communication en milieu bruite", *Proc. Xèmes Journées d'Étude sur la Parole*: 265–268, 2000.
- [16] Rostolland, D. "Phonetic structure of shouted voice", *Acustica*, 51: 80–89, 1982.
- [17] Savariaux, C., Perrier, P. and Orliaguet, J.P., "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production", *J. Acoust. Soc. Amer.*, 98 (5): 2428–2442, 1995.

3. COMMUNICATION A ISSP 2006, UBATUBA

The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement?

Maëva Garnier¹, Marion Dohen², Hélène Lœvenbruck², Pauline Welby², Lucie Bailly^{1,2}

¹Laboratoire d'Acoustique Musicale (CNRS UMR 7604, UMPC, Ministère de la culture)

²Institut de la Communication Parlée (CNRS UMR 5009, INPG, Université Stendhal)

garnier@lam.jussieu.fr, (Marion.Dohen, Helene.Loevenbruck,welby, Lucie.Bailly)@icp.inpg.fr

***Abstract.** The purpose of this study was to examine whether speech in noisy environments consists of global acoustic and articulatory modifications or if there are some changes specific to units within the utterance. Changes on a more local level could be interpreted as a controlled intelligibility enhancement of specific speech cues such as cues to word segmentation or prosodic phrasing. Audio and video signals were recorded for a female native speaker of French in three conditions: silence, 85dB white noise, and 85dB “cocktail party” noise. The corpus consisted of 33 short sentences with a subject-verb-object (SVO) structure. Labial parameters were extracted from the video data. A controlled intelligibility enhancement in noise was observed for some cues to word segmentation and utterance structure.*

1. Introduction

1.1. Background

Speech production is influenced by the immersion of the speaker in a noisy environment. Speech changes in noise are collectively called the Lombard effect and include an increase in vocal intensity, fundamental frequency (F_0), and word duration (Lombard, 1911; Junqua, 1993), as well as an increase in amplitude of articulatory movements (Garnier et al., 2006; Davis, 2006). A number of perception studies have found that speech produced in noisy conditions is more intelligible than speech produced in silence (see, for example, Junqua, 1993; there are however, limits: past a certain point, shouting decreases intelligibility, Rostolland, 1982).

This increased intelligibility raises the question of whether speech changes in noisy environments are motivated by increased intelligibility. This question is part of a larger question on the physiological and cognitive mechanisms underlying the Lombard effect. Communication in noisy environments might be disturbed for at least two reasons: speakers get attenuated feedback of their own voices, and their intelligibility is decreased for listeners. Two main interpretations of the Lombard effect have been proposed. The first argues that the effect is a physiological audio-phonatory reflex (Lombard, 1911), the second that Lombard changes are motivated by compensation on the part of the speaker for decreased intelligibility (Lane and Tranel, 1971). Some authors have also argued that both mechanisms may contribute to the changes made by the speaker in noisy environments (Junqua, 1993).

A number of arguments have been put forth in favor of these different interpretations: On the one hand, speakers cannot totally inhibit speech changes in noise (Pick et al., 1989); and similar changes occur in animals (for example in monkeys, Sinnott et al., 1975). On the other hand, Lombard changes are greater in adults than in children and in spontaneous speech than in reading tasks (Lane and Tranel, 1971; Amazzi and Garber, 1982).

1.2. Aims

One aspect has not yet been examined that could bring arguments to this debate. If the Lombard effect is motivated by a search for intelligibility, Lombard changes might enhance some cues rather than others, and thus might vary over the utterance as a whole or within a smaller unit within the utterance. In this

study we therefore examine some specific aspects of speech production in French, across the utterance as a whole and within smaller units, in both quiet and noisy conditions (Condition 1: NOISE).

First, in French, as in other languages, there are prosodic differences between function words (determiners, conjunctions, etc.) and content words (nouns, verbs, etc.). (Delais-Roussarie, 1995; Welby, 2006; *inter alia*). For example, content word syllables tend to be longer, *ceteris paribus*. The evidence suggests that these differences between function words and content words help listeners in the task of word segmentation (for example, Christophe, 1993). A greater difference between the duration of function word syllables and content word syllables in noisy conditions could suggest an effort to enhance this linguistic category distinction. We therefore examined differences between syllables like [le] in the article *les* and the first and the last syllables of content words like [mu], and [nE] in *les moulinets* 'the reels', in quiet and in noise. (Condition 2: WORD TYPE).

In addition, utterance-initial syllables have been shown to be longer and over-articulated compared to other syllables (Fougeron and Keating, 1997). Similarly, syllables at the end of prosodic units are lengthened (Beckman and Edwards, 1994) and over-articulated (Tabain, 2003; Løevenbruck, 1999), with greater lengthening associated with higher prosodic levels. This relative lengthening is considered to be a cue to word and higher-level boundaries (e.g. Rietveld, 1980; Christophe, 1993; Bagou et al., 2002). We therefore compared initial, intermediate, and final position in the utterance (Condition 3: POSITION). Duration differences at unit boundaries may be enhanced in noise, that is, the difference between the duration of initial syllables in phrase-initial position (like [mi] in *Le minet leva le nez* 'The kitty raised his nose') and the other positions and the difference between the duration of final syllables in phrase-final position (like [nE] in *Loulou nourrit le minet* 'Loulou is feeding the kitty') and the other positions.

Finally, in French, as in other languages, an F_0 declination is observed across the utterance, as is a final lowering at the very end of an utterance (in declaratives, for example). If we consider this declination as a cue to an utterance boundary, we might expect it to be enhanced in noise.

2. Methods

2.1. Background

The corpus consisted of 33 short sentences with a subject-verb-object (SVO) structure. Only CV syllables were used in order to simplify the acoustic labeling of the corpus. Four groups of 4 targets were selected, each consisting of a 2- or 3- syllable content word, with or without a preceding determiner. The 33 sentences combined these 16 targets, so that each target appeared in three positions in the utterance²:

Initial (1) Loulou nourrit le minet. 'Loulou is feeding the kitty.'

Intermediate (2) Nina et Loulou mimaient les lamas. 'Nina and Loulou were imitating the llamas.'

Final (3) Maman ramena Loulou. 'Mom brought back Loulou.'

2.2. Audiovisual recordings

Audio and articulatory signals were simultaneously recorded for a female native speaker of Hexagonal French. The speaker read the sentences to a person standing two meters in front of her. Articulatory data were extracted from video recordings (25 images/s) of the speaker's lips, using a labiometric device developed at the Institut de la Communication Parlée (Lallouache, 1991). In this study, we focused on the analysis of lip spreading (A), lip aperture (B), and inter-lip area (S) (see Figure 1). We then examined the mean amplitude of the articulatory movements, corresponding to the integral over a normalized time period (Dohen, 2005). We also analyzed lip pinching, defined as lip compression when the mouth is closed for [m] segments (i.e. B' when B=0, see Figure 1). The audio signal was recorded with an AKG microphone placed 20cm away from the lips and digitized at a rate of 44.1kHz, over 16bits. Two noisy environments (white noise and cocktail party noise) were used, both extracted from the BD_Bruit database (Zeiliger, 1994). They were played over two loudspeakers located 2m away from the speaker and 2m away from each other. The noise level was calibrated to 85dB at the participant's ears. The speaker was first recorded in a silent reference condition, and then in both noisy environments. Noise was removed from the acoustic signal using a method especially designed for that problem (Ternström et al., 2002). The utterance, target, and syllable boundaries were then labeled using Praat (Boersma and Weenink, 2004). The results presented here report the two types of noise considered together.

² To reduce recording time, some sentences contain more than one target word

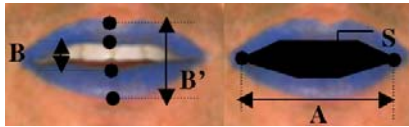


Figure 1. Articulatory parameters.

The parameter values are highly dependent on the segmental composition of the syllable studied (e.g. [mu] vs. [li]; smaller lip area (S) for rounded [u] compared to spread [i]). The following normalization procedure was applied to allow inter-syllable comparisons: for each item: parameter values in the noisy conditions were divided by the corresponding values in the silent condition. After normalization, a value of 1 corresponds to no variation between the noisy and silent conditions; a value above 1 corresponds to an increase in noise compared to the silent condition; and a value below 1 to a decrease in noise compared to silence.

3. Results

Global articulatory and acoustic effects across the utterance as a whole from the quiet to the noisy condition are described in an earlier study (Garnier et al. 2006). We focus here on effects within smaller units. Figure 2 gives an example of the variation across several parameters from (a) silence to (b) noise. Note, for example, that F_0 , intensity, as well as articulatory parameters show a general increase in noise.

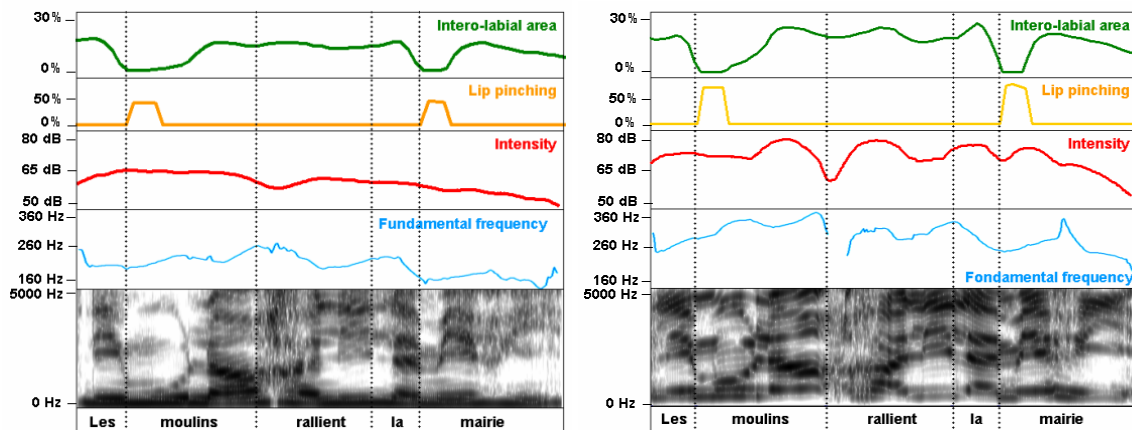


Figure 2. Several parameters for the same item produced in (a) silence and in (b) noise. From top to bottom: inter-lip area, lip pinching, intensity, fundamental frequency, time aligned with the spectrogram and segmentation into words.

As detailed in the following, analyses of variance (ANOVA) were carried out with Matlab. The following notation is adopted in reporting statistical significance:

* $p < .05$, ** $p < .01$, *** $p < .001$, and *ns* (not significant) $p > .05$.

3.1. Comparison of function and content words

The first research question we examined was the following: Is there a difference in enhancement between content word syllables and function word syllables? In order to address this question, we considered the eight targets containing a determiner (*le*, *la* or *les*).³ Of these, four consisted of 2-syllable content words and the other four of 3-syllable content words. We measured the values for the function word syllable, and the two syllables at the edges of the content word, i.e. the first syllable and the last syllable. Figure 3 presents the normalized values for duration and several articulatory parameters in noise relative to silence. For each normalized parameter, we conducted a one-way ANOVA (factor: SYLLABLE TYPE, three levels: function word syllable, content word initial syllable, content word final syllable).

³ Note that throughout the paper, the values reported for function words were measured for these determiners; they do not include, for example, conjunctions.

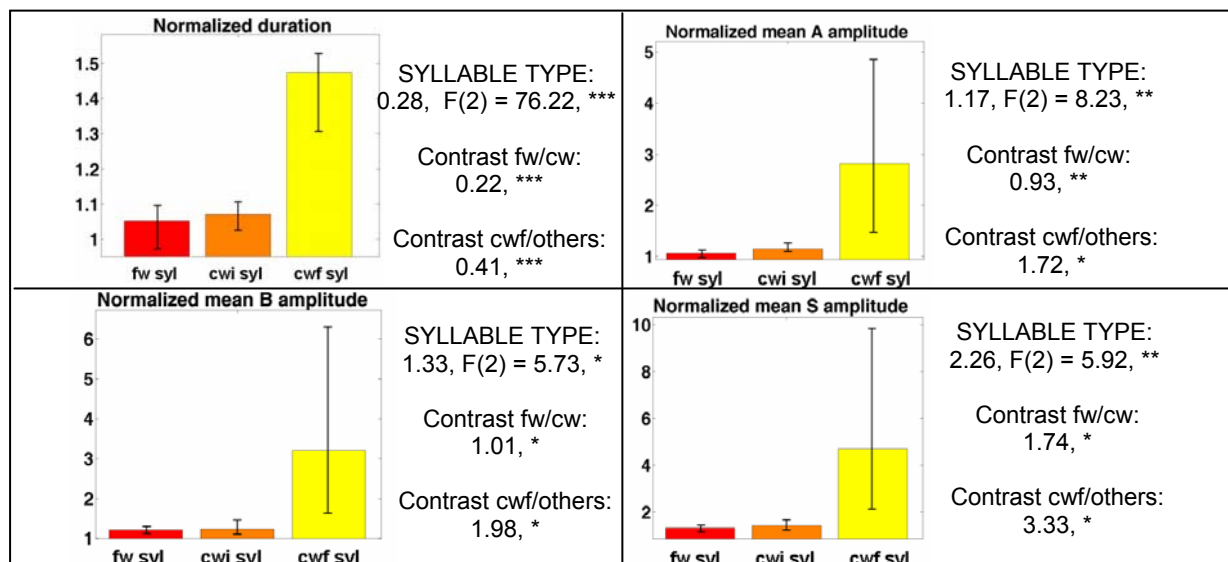


Figure 3. Normalized values for duration and several articulatory parameters in noise relative to silence. Bars represent standard deviation. *fw syl* represents the function word syllable, *cwi syl* the content word-initial syllable and *cwf syl* the content word-final syllable.

First, it is interesting to note that all normalized values are greater than 1, which means that all parameters increased in noise relative to silence, regardless of the word type (function word or content word).

Secondly, a significant effect of word type on the variation from silence to noise was found for duration and mean A, B, and S amplitude. Specifically, the durational and articulatory amplitude parameters are more enhanced in noise for content word-final syllables than for other syllables. It therefore seems, for this speaker, that content word-final syllables are more enhanced in noise than are function word syllables or content word-initial syllables. This could suggest that the speaker tried to reinforce cues to the ends of content words.

3.2. Analysis of the initial syllable of a content word depending on word position within the utterance

The second research question we examined was the following: Is there a difference in initial syllable enhancement when the content word to which the syllable belongs appears in initial, intermediate or final position? For example, is the first syllable in *Loulou* more enhanced in noise in (1), (2) or (3)? (see 2.1). In order to address this question, we considered the initial syllable of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). Figure 4 presents the normalized values for duration and several articulatory parameters in noise relative to silence for the word-initial syllables in the 3 positions. The results of the one-way ANOVA (factor: position, three levels: initial, intermediate, final) are reported beside each graph (label POSITION). The results labeled *contrast init/others* correspond to a comparison between the initial level and the other two levels. These results were obtained using multiple comparison tests derived from the ANOVA.

First, it is interesting to note that all normalized values, except lip pinching in non-initial positions, are greater than 1, which means that on initial syllables, all parameters increased in noise relative to silence, regardless of the position of the word to which the initial syllables belong (initial, intermediate or final).

Secondly, a significant effect of position on the variation from silence to noise was found only for the mean B amplitude. More specifically, for the content word initial syllables, only the B mean amplitude showed significantly greater increases in noise in the initial position than in the other positions. However, syllable duration, mean A amplitude, as well as mean amplitude of lip pinching, all tended to show a greater increase in noise for initial syllables of content words in utterance-initial position relative to content words in other positions.

It therefore seems that for this speaker, content word-initial syllables tend to be longer, produced with a more open articulation in noise than in silence, and especially so when the content word is in initial position in the utterance.

As mentioned in section 1.2, it has been suggested in the literature that initial syllable lengthening could be a cue to word and higher-level boundaries. The fact that the initial position in the utterance is accompanied by greater articulatory and durational increases in noise (relative to silence) suggests that the speaker might have been trying to enhance articulatory and durational cues to word segmentation and prosodic hierarchy.

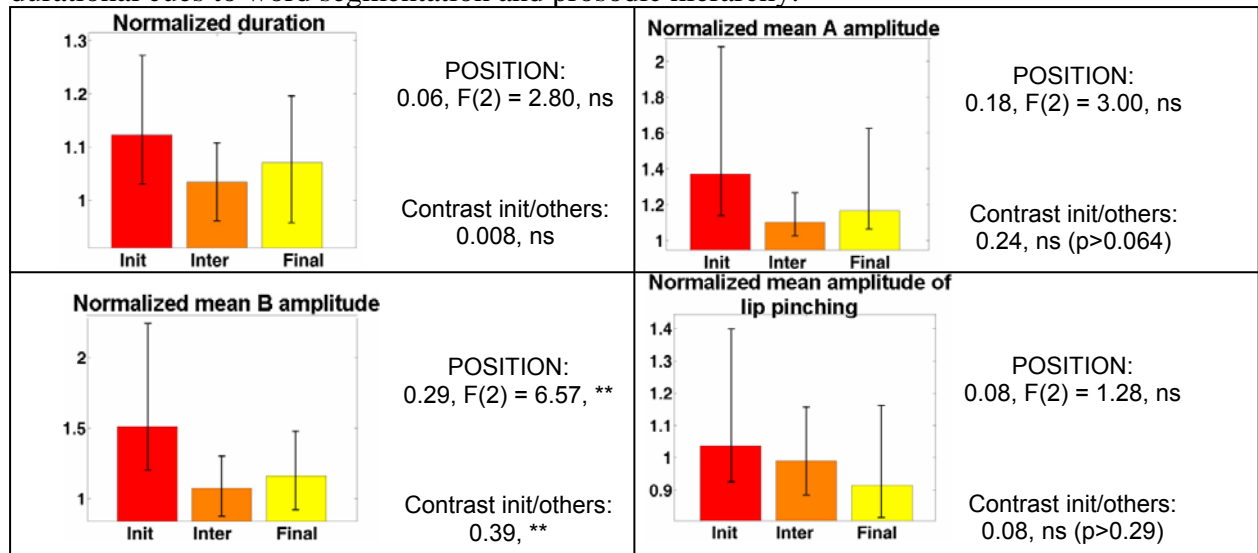


Figure 4. Normalized values for duration and several articulatory parameters in noise relative to silence for initial syllables in content words in initial, intermediate and final position in the utterance. Bars represent standard deviation.

3.3. Analysis of the final syllable of a content word depending on word position within the utterance

Next we examined the question: Is there a difference in final syllable enhancement when the content word to which the syllable belongs appears in initial, intermediate or final position? For example, is the last syllable in *Loulou* more enhanced in noise in (1), (2) or (3) (see 2.1)?

In order to address this question, we considered the initial syllable of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). For eight of the content words, the final syllable was the second one; for the other eight, it was the third one. In order to address this question, we considered the initial syllable of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). Figure 5 presents the normalized values for duration and several articulatory parameters in noise relative to silence for the word-final syllables in the three positions. The results of the one-way ANOVA (factor: position, three levels: initial, intermediate, final) are reported beside each graph (label POSITION). The results labeled *contrast init/others* correspond to a comparison between the initial level and the other two levels. These results were obtained using multiple comparison tests derived from the ANOVA.

First, it is interesting to note that all normalized values are greater than 1, which means that on final syllables, all parameters increased in noise relative to silence, regardless of the position of the word to which the final syllables belong (initial, intermediate or final).

Secondly, a significant effect of position on the variation from silence to noise was found for duration. More specifically, for the content word-final syllables, the syllable duration showed a significantly higher increase in noise in the utterance-final position than in the other positions. In addition, all articulatory parameters tend to show a greater increase in noise for final syllables of content words in utterance-final position relative to content words in other positions.

It therefore seems that for this speaker, final content word syllables tend to be longer, produced with a more open articulation in noise than in silence, particularly when the content word is in final position in the utterance. As mentioned in section 1.2, it has been suggested in the literature that final syllable lengthening could be a cue to word and higher-level boundaries. The fact that the final position in the utterance is accompanied by articulatory and durational increases in noise (relative to silence) suggests that the speaker might have been trying to enhance articulatory and durational cues to word segmentation and prosodic hierarchy.

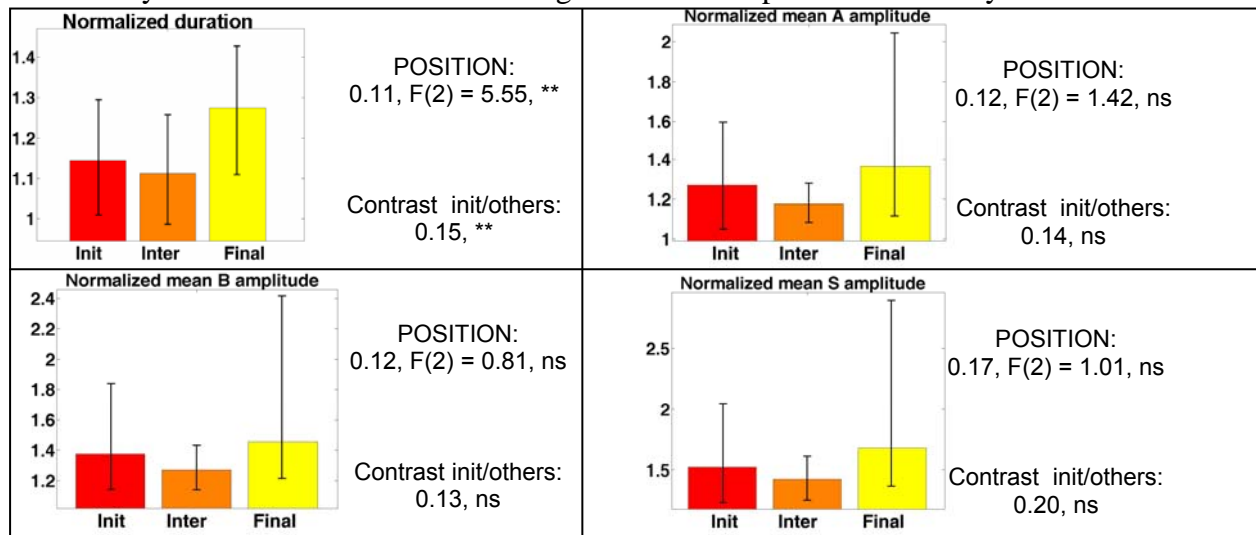


Figure 5: Normalized values for duration and several articulatory parameters in noise relative to silence for final syllables in content words in initial, intermediate and final position in the utterance. Bars represent standard deviation.

3.4. Analysis of F_0 and intensity declination over the utterance

Finally, we examined the question: Are F_0 and intensity declination across the utterance enhanced in noise? To answer this question, we considered mean F_0 and intensity of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). Figure 6 shows the results of the F_0 and intensity analyses.

In silent conditions, as expected, an F_0 declination utterance is observed across the (POSITION effect: 23 Hz, $F(2) = 64.14$, ***). This declination is still observed in noisy conditions (POSITION effect: 30 Hz, $F(2) = 96.64$, ***) but is not significantly enhanced. It has been previously demonstrated (Lombard, 1911; Junqua, 1993) that speech in noise is produced with a globally higher F_0 than in silence. We might suspect that the lesser declination observed here in noise could be due to the fact that F_0 values for initial and intermediate content words have reached a ceiling (cf. Rostolland, 1992). However, the standard deviation of mean F_0 does not decrease in noise, even for the initial and intermediate content words, which are produced with higher mean F_0 . Thus the declination attenuation does not seem to be due to a ceiling effect. Similar observations can be made about the intensity declination across the utterance (POSITION effect: 3.3 dB, $F(2) = 90.78$, ***, *in silence*; 2.3 dB, $F(2) = 21.66$, ***, *in noise*). As mentioned in 1.2, F_0 declination can be considered to be a cue to an utterance boundary. An enhancement of F_0 declination or intensity declination would thus reinforce cues to utterance boundary. It seems that this speaker, however, did not particularly reinforce these F_0 and intensity cues to utterance boundaries in noise. It may be that other cues playing the same role (such as final syllable lengthening and over-articulating) were sufficiently enhanced. F_0 value could thus be maintained to a sufficiently high level, even at the utterance end.

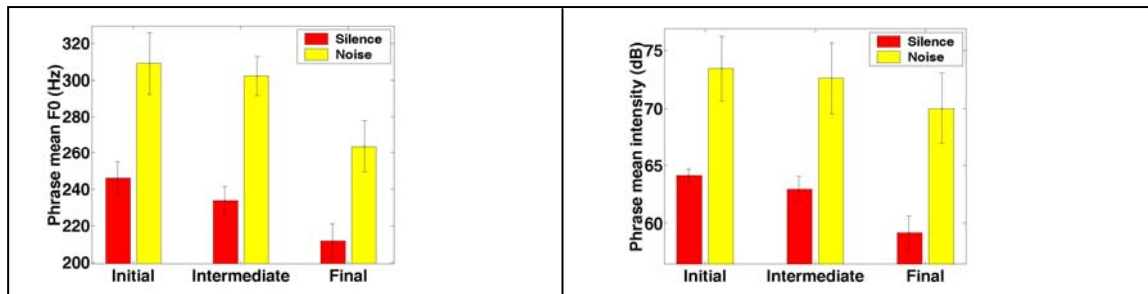


Figure 6. Declination of F_0 and intensity over the utterance in silent and noisy conditions. Bars represent standard deviation.

4. Discussion and Conclusions

In this study, we examined the variation from silence to noise of specific aspects of speech that could enhance intelligibility in noise. We conducted these analyses in order to explore if the Lombard effect affects the whole utterance in the same way or if cues to intelligibility are more enhanced for certain units of the utterance. This could shed light on the question of whether the Lombard effect is simply a physiological reflex or a controlled intelligibility enhancement.

Our previous research has shown global lengthening and over-articulation in noise (Garnier et al. 2006). We sought here to examine whether more localized effects would be found, in particular ones which might contribute to easing word segmentation or phrasing decisions. The results show that content word-final syllables are more enhanced in noise than are other syllables (increased lengthening and increased over-articulation). These changes are therefore not purely physiological in nature, since the changes from silence to noise depend on linguistic category (function word vs. content word). The fact that our speaker enhances content word-final syllables more than content word-initial syllables is consistent with a reinforced marking of the end of content words. We note that in French, the ends of content words (or of accentual phrases) tend to be more consistently and redundantly marked than the beginnings (see Welby 2006 and references therein). We also observed a general tendency for initial content word syllables in utterance-initial position to be accompanied by greater articulatory and durational increases in noise. This is consistent with the enhancement of articulatory and durational cues to word segmentation and prosodic phrasing. We observe similar results for absolute utterance-final syllables.

In addition, F_0 and intensity declinations across the utterance were attenuated rather than enhanced in noise. Rostolland, 1992 observed a plateau for shouted voice at the top of the intonational curve, which he interpreted as being a limitation to voice frequency modulation. Our data, however, do not show a significantly lower standard deviation corresponding to the increase of the strongest units and the weakest ones. Thus, it does not seem that the speaker reached her limit.

This study therefore shows some variability in acoustic and articulatory changes from silence to noise depending on unit within an utterance or a word, changes which seem to be consistent with enhanced cues to word segmentation and prosodic phrasing. These results are also in line with those obtained in previous studies on clear speech (Cutler and Butterfield, 1990) and emphatic speech (Løevenbruck, 2000). The lack of statistical significance for some measures may (or may not) be due the fact that this was a single speaker study with a small corpus.

Our results call for follow-up studies in a number of areas, including: the extent to which the results found are dependent on the speaker, on the language studied, and on the corpus used. In addition, if the changes observed do indeed enhance cues to word segmentation and prosodic phrasing, perception studies should demonstrate this. Finally, other intonational parameters are clearly worth examining. For example, Welby, 2006 suggested that the intonational cues to content word beginning in French might be enhanced in Lombard speech.

5. Acknowledgments

The authors would like to thank Christophe Savariaux, Alain Arnal, Aude Noiray, Claire Lalevée and Coriandre Vilain for their contribution to this study. Pauline Welby's participation in this project was supported by a Marie Curie International Fellowship of the 6th European Community Framework Programme.

References

- Amazi, D. K. and Garber, S. R. The Lombard sign as a function of age and task. *The Journal of Speech and Hearing Research*, 25(4): 581–585, 1982.
- Bagou, O. and Frauenfelder, U. H. Stratégie de segmentation prosodique: rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle. In *Proceedings of the XXVIèmes Journées d'Etude sur la Parole*, pages 571–574, 2006.
- Beckman, M.E. and Edwards, J. Articulatory evidence for differentiating stress categories. In Keating, P.A, editor, *Phonological structure and phonetic form: Papers in laboratory phonology III*, pages 7–33, Cambridge University Press, 1994.
- Boersma, P. and Weenink, D. Praat: doing phonetics by computer (Version 4.2.28) [Computer program]. Retrieved from <http://www.praat.org/>, 2004.
- Christophe, A. Role de la prosodie dans la segmentation en mots. Doctoral thesis, École des Hautes Études en Sciences Sociales, Paris, 1993.
- Cutler, A. and Butterfield, S. Durational cues to word boundaries cues in clear speech. *Speech Communication*, 9: 485–495, 1990.
- Davis, C., Kim, J., Grauwinkel, K. and Mixdorff, H. Lombard speech: Auditory(A), Visual(V) and AV effects. In *Proceedings of Speech prosody*, pages 361–365, 2006.
- Delais-Roussarie, E. Pour une approche parallèle de la structure prosodique : étude de l'organisation prosodique et rythmique de la phrase française. *Doctoral dissertation, University of Toulouse*, 1995.
- Dohen, M. Deixis prosodique multisensorielle : production et perception audiovisuelle de la focalisation contrastive en français. *Doctoral dissertation, Institut National Polytechnique de Grenoble*, 2005.
- Fougeron, C. and Keating, P. A. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6): 3728–3740, 1997.
- Garnier, M., Bailly, L., Dohen, M., Welby, P. and Løevenbruck, H. An acoustic and articulatory study of Lombard speech : global effects on the utterance. In *Proceedings of ICSLP*, pages 2246–2249, 2006.
- Junqua, J. C. The Lombard reflex and its role on human listener and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93: 510–524, 1993.
- Lallouache, M. T. Un poste “Visage-Parole” couleur. Acquisition et traitement automatique des contours des lèvres. *Doctoral dissertation, Institut National Polytechnique de Grenoble*, 1991.
- Lane, H. and Tranel, B. The Lombard sign and the role of hearing in speech. *The Journal of Speech and Hearing Research*, 14: 677–709, 1971.
- Løevenbruck, H. An investigation of articulatory correlates of the accentual phrase in French. In *Proceeding of ICPHS 99*, vol. 1, pages 667–670, 1999.
- Lombard, E. Le signe de l'élévation de la voix. *Annales des maladies de l'oreille et du larynx*, 37: 101–119, 1911.
- Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R. and Kearney, J. K. Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America*, 85(2): 894–900, 1989.
- Rietveld, A. C. M. French word boundaries. *Language and Speech*, 23(3): 289–296, 1980.
- Rostolland, D. Phonetic structure of shouted voice, *Acustica*, 51: 80–89, 1982.
- Sinnott, J. M., Stebbins, W. C. and Moody, D. B. Regulation of voice amplitude by the monkey. *The Journal of the Acoustical Society of America*, 58: 412–414, 1975.
- Tabain, M. Effects of prosodic boundary on /aC/ sequences: articulatory results. *The Journal of the Acoustical Society of America*, 113(5): 2834–2849, 2003.
- Ternström, S., Sodersten, M. and Bohman, M. Cancellation of simulated environmental noise as a tool. *The Journal of Voice*, 16: 195–206, 2002.
- Welby, P. Intonational differences in Lombard speech: looking beyond F_0 range. In *Proceedings of Speech Prosody*, pages 763–766, 2006.
- Welby, P. French intonational structure: Evidence from tonal alignment. *The Journal of Phonetics*, 34(3): 343–371, 2006.
- Zeiliger, J. BD_Bruit, une base de données de parole de locuteurs soumis à du bruit. In *Proceedings of the Xèmes Journées d'Étude sur la Parole*, pages 287–290, 1994.

Annexes de résultats

Annexe Res1 :
Questionnaire sur le
forçage vocal

Error! No table of contents entries found.

1. QUESTIONS POSEES

Vous êtes :

- Chercheur en _____
- Orthophoniste
- Professeur de chant
- Médecin

- Acteur
- Chanteur
- Enseignant
- Autre : _____

- 1) Qu'est ce que le forçage vocal pour vous? Pourriez-vous en donner une définition ?
- 2) Pourriez vous citer des situations de forçage vocal ?
- 3) Quelles sont les caractéristiques de ces situations qui vous font penser qu'il s'agit de forçage vocal ?
- 4) Qu'est ce qu'une voix forcée pour vous ? Pourriez-vous en donner une définition ?
- 5) Sur quelles caractéristiques vous basez vous pour déterminer qu'une voix est forcée ?
- 6) Personnellement, vous arrive-t-il de forcer sur votre voix ? Si oui, dans quelle(s) situation(s) ?

Page réservée aux orthophonistes, médecins ou professeurs de chant (sinon, vous pouvez passer à la question 12)

- 7) Quels sont les profils des personnes qui viennent vous consulter et qui ont un problème de forçage vocal ? (professions, milieux, comportements, personnalités...)
- 8) Pouvez vous décrire les voix de ces personnes ?
- 9) De quelles gênes, fatigues ou douleurs se plaignent-elles en venant consulter?
- 10) Remarquez vous d'autres caractéristiques de forçage vocal chez elles, dont elles ne parlent pas ?
- 11) Quelles type de rééducation, ou de traitement mettez vous en œuvre pour aider ces personnes ?

- 12) Qu'est ce que l'efficacité vocale pour vous ? Pourriez-vous en donner une définition ?
- 13) Qu'est ce que l'effort vocal pour vous ? Pourriez-vous en donner une définition ?
- 14) Qu'est ce que la fatigue vocale pour vous ? Pourriez-vous en donner une définition ?
- 15) Qu'est ce que le malmenage vocal pour vous ? Pourriez-vous en donner une définition ?
- 16) En quoi consiste un échauffement vocal pour vous ?

2. REPONSES INDIVIDUELLES

2.1. Première question : définition du forçage vocal

Médecins	M1	C'est un cercle vicieux caractérisé par l'utilisation du mécanisme de la voix d'insistance (ou de détresse), là où normalement devrait intervenir le mécanisme de la voix projetée.
	M2	c'est un comportement phonatoire utilisant de façon inadéquate ou trop importante, mal adaptée, la musculature pharyngo-laryngée et cervicale, voie thoraco-abdominale. La fermeture glottique est réalisée soit sous forme d'une attaque dure (« coup de glotte »), serrée, avec relâchement discret pour la mise en vibration cordale, soit par serrage aryténoïdien forcé (glotte ovalaire à l'attaque du son). Dans cette définition, une aphonie psychogène correspond à un forçage vocal extrême.
	M3	sollicitation supra-physiologique de l'appareil phonatoire
	M4	C'est un terme que j'emploie peu. Mais la définition serait celle que C'est un geste vocal qui nécessite une énergie musculaire qui n'est pas en rapport avec l'efficacité acoustique du son produit. Du fait de la mobilisation contrariée de groupes musculaires aussi bien au niveau articulaire que respiratoire, laryngé ou postural

Orthophonistes	O1	Perturbation de l'accord pneumophonique et perturbation de la vibration des cordes vocales. -> Trop de pression sous-glottique + trop grande constriction du larynx.
	O2	Altération du bon fonctionnement vocal, c'est-à-dire une mauvaise utilisation de la coordination pneumophonatoire, avec un comportement demandant beaucoup d'effort pour peu d'efficacité
	O3	C'est l'augmentation de la pression expiratoire visant à "forcer" la résistance laryngée
	O4	C'est l'abus d'une utilisation inadéquate de l'appareil phonatoire
	O5	C'est un comportement d'effort vocal utilisé de façon systématique et inadéquate dans respect du repos que l'organisme adopte de façon physiologique lorsque les signes de fatigue vocale de font sentir.
	O6	Le forçage vocal est le résultat perceptible et/ou subjectif d'efforts excessifs lors de l'adduction des cordes vocales lors du chant ou de la parole. Le plus souvent le forçage résulte d'un excès de tension dans les muscles intrinsèques ou extrinsèques du larynx (la région du cou, de la nuque, du maxillaire, ou du larynx lui-même) Le forçage, pratiqué de manière systématique/continue, sont à la base du malmenage.

Chercheurs en phonétique	CP1	Sollicitation trop importante des cordes vocales
	CP2	Si c'est l'aspect le plus pathologique, je dirais qu'il s'agit des conséquences (accidentelles, ponctuelles ou chroniques) d'une mauvaise utilisation du larynx / des cordes vocales
	CP3	Action de pousser/forcer la voix à dépasser un "seuil" lors d'un chant et de l'émission de la parole, le seuil étant le niveau d'intensité minimale de la voix audible en situation conversationnelle
	CP4	Faire porter l'effort sur les cordes vocales au lieu de soutenir au niveau du diaphragme

Enseignants	E1	Devoir hausser la voix pour se faire entendre, ou pour mieux faire passer un message.
	E2	utiliser sa voix en faisant des efforts. Essayer de lui faire faire plus qu'elle ne peut normalement donner.
	E3	Il y a forçage quand il y a dysfonctionnement de l'appareil phonatoire dû à une mauvaise émission vocale (forcée).

Professeurs de chant	PC1	Le terme est impropre et amène à l'erreur. Le forçage n'est pas vocal mais laryngé. Le phénomène de la voix étant pneumatique, ondulatoire et acoustique, obéissant aux lois de la physique, il est immuable. Dans le phénomène de phonation, le larynx est fermé et « passif ». Dès que cette passivité est transgressée, il y a un risque de forçage. Mécaniquement, le forçage est un surcroît de pression sous glottique provoquant une tétanisation des muscles du larynx
	PC2	C'est une "mauvaise" solution pour obtenir un (plusieurs) son(s) qu'on n'arrive pas à faire. On s'oppose alors volontairement à la production du son telle que le corps l'aurait fait sans forçage, mais qui ne serait pas adéquat en rapport avec l'effet recherché.
	PC3	Une rupture de l'équilibre énergétique entre pression sous-glottique, résistance de la glotte, et impédance des résonateurs. Schématiquement, une pression sous-glottique trop faible conduit à une compensation musculaire laryngée pour assurer la fréquence fondamentale. De même une pression sous glottique trop forte conduit à une compensation musculaire laryngée forte pour éviter la fuite d'air. Dans les deux cas, le larynx est brutalisé, ses muscles hyper sollicités.
	PC4	sensation de gêne physique au niveau de la gorge
	PC5	Lors d'un effort en recherche d'efficacité, aller au-delà de la limite garantissant la préservation du bon état de la voix.

2.2. Deuxième question : exemples de situations de forçage vocal

Médecins	M1	Toute situation où l'on se sent dépassé ou impuissant du fait de l'insuffisance réelle ou supposée de sa propre voix au sens large du terme .Exemple : « Je n'arrive pas à me faire entendre » ou « Je n'arrive pas à me faire comprendre »
	M2	Enseignement en milieu bruyant (sport), Aphonie psychogène, Enfant exubérant jouant en groupe ou prenant la parole à table, en milieu bruyant, Ouvrier en usine, devant donner des consignes devant une chaîne de montage bruyante
	M3	communication en milieu bruyant ou entre deux interlocuteurs distants ou pour faire rentrer quelque chose dans la tête de son partenaire qui, en dépit du bon sens, refuse de comprendre
	M4	La situation de forçage est un état dans lequel se trouve un sujet en difficulté avec sa voix. Dans le bruit par exemple où il n'arrive pas faire émerger sa voix qualitativement ou quantitativement du milieu (assemblée, rue, restaurant). Ou bien lors d'une prise de parole prolongée, lorsque la qualité du son s'altère et que le sujet cherche à compenser cette fatigue.

Orthophonistes	O1	Présence d'une anomalie organique (nodule, polype), phonation en milieux bruyants, phonation prolongée (profs, instituteurs...), chorale, voix "de détresse" dans des situations d'urgence, situations "émotionnelles".
	O2	Conversation en milieu bruyant, utilisation importante de la voix (enseignement, chant), situations émotionnelles, présence de lésions organiques, parfois troubles psychologiques
	O3	Je vois surtout des situations pathologiques : certaines formes de compensation chez des patients avec pathologies neurodégénératives.
	O4	en environnement bruyant (enseignement, théâtre de rue, manifestations, débats politiques et syndicaux), dans les utilisations extrêmes de l'appareil vocal (chant, théâtre)
	O5	Je pourrais citer des situations d'effort vocal plus ou moins prolongé dans le temps (faire un cours sans micro, parler dans une boîte de nuit, jouer une pièce de théâtre en plein air....) mais pas de situation de forçage
	O6	Prise de parole sur air 'résidu' (manque de soutien respiratoire à la parole), Chant dans un registre peu confortable pour le sujet (trop haut ou trop bas, ou tenu trop longue) - ceci arrive aux chanteurs amateurs Prise de parole prolongée dans des postures difficiles, p.ex. au téléphone (flexion asymétrique de la tête), autre exemple : les enseignants qui parlent aux élèves en tordant le cou alors qu'ils font face au tableau.= mauvaise posture, peu propice à la voix portée et qui engendre des tensions. Gros fumeurs et personnes grippées qui s'efforcent à parler, alors qu'ils souffrent d'une inflammation aigüe ou chronique des cordes vocales

Chercheurs en phonétique	CP1	Dans un environnement trop bruyé. Le supporter de sport qui crie dans les tribunes pour encourager, les groupies dans un concert, les gens qui utilisent mal leurs cordes vocales pour chanter
	CP2	J'imagine, à la suite d'une voix forcée en environnement bruyant, ou sous l'effet d'une forte émotion, ou d'une voix criée.
	CP3	La voix criée, la voix chantée plus aigüe que sa voix normale ou plus grave que sa voix normale.
	CP4	chanteur inexpérimenté s'attaquant au registre aigu

Enseignants	E1	Conversation dans un bar (musique trop forte). Enseignement quand les élèves chahutent. Gronder son enfant quand il ne veut pas aller se coucher ! Le marchand de légumes de la rue de Lourmel ...
	E2	en situation de chant ou de cri
	E3	Par exemple dans une salle de classe avec des élèves bruyants. Le professeur va forcer s'il essaye de passer au dessus du bruit de fond. De même en discothèque, ou dans une salle de concert. Encore, dans un transport en commun telle que métro, bus ...Ou lors d'une discussion croisée dans un repas à table avec beaucoup de personnes. Mais aussi dans le cadre d'une mauvaise place de voix dans le chant (lyrique, rock...) ou pour un acteur de théâtre.

Professeurs de chant	PC1	Dès qu'il y a recherche d'intensité sur un larynx insuffisamment tonique, dès qu'il y a recherche de hauteur de note non maîtrisée (chanteurs), il y a forçage
	PC2	Etat de trac, de crainte ou d'angoisse, retour du son produit trop faible (écoute intérieure), déficiences techniques, fatigue occasionnelle.
	PC3	Chanter hors tessiture (trop grave ou trop aigu). Chanter trop longtemps, devoir passer un orchestre trop bruyant, chanter des partitions mal écrites pour la voix, parler en appui sur le larynx (sans « soutenir »), dans des situations de stress. Parler dans une boîte de nuit avec la sono à fond. Chanter du répertoire exigeant pour le contrôle du vibrato (XVIIe par exemple) et opérer ce contrôle par un appui laryngé plutôt que par un contrôle diaphragmatique.
	PC4	hurler d'une façon prolongée, parler non stop avec tension en donnant de cours ou de conférences, pendant des heures, chanter en forçant de la gorge, chanter en prétendant du volume sans emploi du souffle parler ou chanter ayant une laryngite ou autre maladie ORL
	PC5	Enseignement (bruit ambiant, grande salle), chant (répertoire inapproprié, grande salle, plein air), conversation (lieu bruyant), en cas de pathologie,...

2.3. Troisième question : Caractéristiques des situations de forçage vocal

Médecins	M1	Les situations qui déclenchent le cercle vicieux du forçage vocal sont toutes celles qui entraînent le sentiment – plus ou moins conscient et plus ou moins fondé – de l'insuffisance vocale.
	M2	Tensions et instabilité vocale ressenties par le patient - Sensation de fatigue vocale progressive - Désodorisations intempestives, de quelques minutes à quelques heures - Usage vocal récemment modifié, avec gêne hypo pharyngée (augmentation des heures de cours / tensions, émotions inhabituelles / conflits / personne « débordée »)
	M3	volonté du locuteur de se faire entendre dans un environnement jugé hostile
	M4	La sensation d'effort, d'altération du timbre (appauvrissement ou érailement), l'inversement du rapport entre l'effort fourni et le résultat sonore

Orthophonistes	O1	Intensité demandée plus forte: dans le bruit. Fatigabilité vocale dans la durée. Importance des émotions : la voix traduit le monde intérieur.
	O2	Utilisation de la voix importante, impact des émotions et de l'état psychologique sur la voix, caractéristiques perceptives de voix pathologiques.
	O3	Augmentation de la raucité, du souffle (conséquences ?)
	O4	milieu bruyant, utilisation poussée de l'appareil vocal (Augmentation de la fréquence ou de l'intensité dans les domaines artistiques), volonté d'imposer ses idées par la voix (politique, manifestations)
	O5	la nécessité d'utiliser une voix projetée avec une intention de communiquer dans un environnement plus ou moins favorable à la propagation du son ou rendu difficile par l'état physique des organes nécessaires à la production vocale
	O6	Fatigue vocale, en fin de journée surtout Raucité intermittente, instabilité vocale (dysphonie intermittente, pire après efforts) Manque d'intensité de la voix Plaintes de douleurs au niveau du larynx Efforts visibles à la phonation au niveau des muscles du cou et posture caractéristique (extension de la tête) Mauvais patterns respiratoires (respiration claviculaire superficielle p.ex.)

Chercheurs en phonétique	CP1	Du bruit autour du locuteur qui l'oblige à parler plus fort pour s'entendre et se faire entendre
	CP2	Je pense que la qualité de l'émission laryngée va se trouver dégradée. Handicap au niveau des muscles vocaux, mauvaise vibration, etc...
	CP3	changement de ton, niveau d'intensité élevé
	CP4	la voix est dépourvue d'harmoniques et elle est désagréable de dureté

Enseignants	E1	soit c'est du à un fort bruit extérieur ; soit c'est dans une situation où l'autre n'a pas envie d'entendre (l'enfant, on cherche à «en imposer ») ou n'écoute pas vraiment (l'étudiant, la ménagère de la r de Lourmel).
	E2	le résultat est désagréable
	E3	Dans toutes ces situations il y a un niveau de bruit ambiant assez élevé (par rapport au niveau sonore de la voix parlée à l'état normal). Seule exception : dans le cadre d'un chanteur ou d'un acteur qui place mal sa voix le paramètre du bruit ambiant n'intervient pas forcément. On en arrive à la seconde caractéristique : le niveau sonore de l'émission vocale est élevé car il est aussi fonction d'une certaine distance avec l'auditoire. Une troisième caractéristique mais qui n'implique pas forcément la deuxième caractéristique est le fait que l'on ne s'entende pas bien (on s'en rend compte quand on parle longtemps lors d'un voyage en voiture sur l'autoroute ... on est aphone après). Une dernière caractéristique, je pense, est la pose de voix : suivant les muscles que l'on contracte, ou si l'on « force » sur le larynx ...

Professeurs de chant	PC1	Le son peut être momentanément de qualité parce que le forçage amène à un bon accolement. C'est de courte durée, proportionnelle à la tonicité du larynx. La congestion des cordes amène vite souffrance et fatigue, et ensuite lésion et détérioration du son
	PC2	Désagrément à l'écoute extérieure, faible volume sonore alors qu'on voit le chanteur faire des efforts, signes croissants de fatigue (difficultés d'émission, de justesse, souffle court) suivant une exécution.
	PC3	Situations qui exigent trop de la voix, ou qui demandent une technique que le locuteur/chanteur n'a pas.
	PC4	le sens de peine physique, de stress de peur et de malaise
	PC5	La nécessité de trouver le maximum d'efficacité (être entendu et surtout compris) dans des situations difficiles pour la voix.

2.4. Quatrième question : Définition d'une voix forcée

Médecins	M1	Une voix forcée est une voix qui a perdu sa maniabilité et sa qualité acoustique du fait de l'altération de la muqueuse des plis vocaux due au forçage
	M2	C'est une voix provoquant subjectivement une sensation d'effort, de tension, pour l'interlocuteur ; son timbre est appauvri, ses vibrations diminuées, avec éventuel enrouement. Elle peut être émise à trop forte intensité, mais, le plus souvent (adultes), à trop faible intensité, retenue, trop grave (« économie » de l'air expiré).
	M3	voix dont les caractéristiques de timbre, et/ou de fréquence et/ou de volume sonore sont poussés à leurs limites optimales
	M4	Une voix dont les caractéristiques acoustiques peuvent être très différentes mais qui donnent l'impression que le résultat sonore (timbre et/ou intensité et/ou variations de timbre d'intensité et de hauteur) n'est pas en rapport avec les mobilisations corporelles du sujet.

Orthophonistes	O1	Une voix forcée donne l'impression que la phonation devient laborieuse pour le locuteur, qu'il utilise plus d'énergie pour mettre en vibration ses cordes vocales. C'est une voix qui semble pathologique. Cependant la voix forcée peut être culturelle (femmes méditerranéennes) et non perçue comme pathologique. Elle peut même être "cultivée" = effet de style dans le chant (rock-variété).
	O2	Voix soufflée, rauque, serrée. Peu d'utilisation optimum de la voix.
	O3	C'est une voix plus sonore, avec une insuffisance du contrôle laryngé.
	O4	C'est une voix qui fonctionne à l'envers. Plus on fait d'efforts pour qu'elle soit efficace et moins elle est efficace
	O5	une voix qui développe une dépense d'énergie importante pour le sujet sans efficacité
	O6	Voir point 1

Chercheurs en phonétique	CP1	La conséquence d'un forçage vocal. C'est après une utilisation abusive des cordes vocales, lorsqu'elles vibrent moins et que par conséquent, la voix est moins audible ou "rocailleuse"
	CP2	Si c'est l'aspect le moins pathologique, je dirais que c'est une voix qui atteint ses limites en voulant effectuer une certaine tâche (intensité, contrôle mélodique, etc.) et qui n'y parvient pas.
	CP3	Action de pousser sa voix à dépasser un seuil lors de l'émission de la parole. Par seuil, j'entends le niveau d'intensité minimal de la voix audible en situation conversationnelle.
	CP4	à la longue, les cordes vocales s'abiment, avec formation de nodules

Enseignants	E1	Une voix forcée : par définition c'est une voix qui doit être différente de la voix « normale » : on veut faire passer un message de manière plus claire.
	E2	une voix qui n'est pas dans un registre normal -ou suraigüe ou trop criarde
	E3	C'est une voix « agressive ». Saturée, dénaturée, enraillée, enrouée. Elle est souvent caractérisée par un grésillement dans le son. Elle est abîmée par le mauvais usage qui en a été fait et des conséquences physiologique et acoustique sont apparues. La voix est fatiguée avec des irrégularités (bruits parasites), la personne est souvent aphone, et il peut y avoir des risques de formations de nodules ou polypes sur les cordes vocales.

Professeurs de chant	PC1	Une voix forcée est un déséquilibre entre la pression sous glottique trop forte et l'impédance acoustique du timbre. Jusqu'à plus ample recherche, la pression sous glottique se mesure mal (elle se déduit, mais ne se vérifie pas). L'impédance ne se mesure pas mais s'observe. La pratique pragmatique des chanteurs et orateurs à forte intensité m'impose cette conclusion.
	PC2	Idem que 1 : pas d'adéquation entre le son désirable et le son produit compensé par trop d'effort plutôt que par des défaillances incontrôlées et imprévisibles.
	PC3	Une voix qui donne à entendre que la fermeture glottique n'est pas complète ou est irrégulière. (apériodicités, fuite d'air, mauvaise distribution de l'énergie dans les harmoniques, effort phonatoire disproportionné avec le résultat acoustique). En d'autres termes, voix éraillée, rauque, émission laryngée, voix sourde...
	PC4	une sensation de dérèglement physique et psychique concernant la personne qui parle ou chante, un son désagréable à la perception globale, non seulement à l'audition, une sensation de mal être.
	PC5	Voix insuffisamment maîtrisé (soutien, richesse, placement) pour être efficace en certaines circonstances et conduisant à une fatigue vocale.

2.5. Cinquième question : Caractéristiques d'une voix forcée

Médecins	M1	Acoustiquement sur son timbre altéré mais plus encore sur le manque de souplesse des attaques et des arrêts vocaux
	M2	Sensation subjective de forçage par l'examineur - Voix trop grave, tendue / air expiré « économisé » - Voix tendue et aiguë, trop forte.
	M3	sur la perception qu'en donne le locuteur et/ou l'auditeur des caractéristiques de timbre, et/ou de fréquence et/ou de volume sonore utilisées à la phonation
	M4	C'est un terme que j'emploie peu. Ce n'est pas la voix qui est forcée mais le geste pour la produire qui est peu adapté. C'est l'importance de l'énergie apparemment dépensée rapportée au résultat. La fatigabilité est fréquente.
Orthophonistes	O1	Caractéristiques acoustiques : phonation "serrée", voix bitonale, intensité trop forte ou hypophonie par serrage trop important, raucité, modulation réduite. Caractéristiques visuelles : Crispation des muscles sterno-cléido-mastoïdiens, jugulaires saillantes, respiration plutôt thoracique supérieure lors de la phonation, posture du corps.
	O2	Caractéristiques perceptives, indices de posture, de comportement du patient, si possible examen phoniatrique, examen voix parlée, chantée, projetée.
	O3	Augmentation de la raucité, du souffle
	O4	Une voix serrée, rauque, avec une composante soufflée. En regardant la personne on peut parfois voir des crispations au niveau du cou, des épaules, de la mâchoire.
	O5	écoute de la voix, observation du geste vocal
	O6	Je laisse ceci aux spécialistes, mais je pense que sur un spectro on verrait – comme dans les cas de dysphonies, des moments d'interruption du voisement et une instabilité des paramètres de la voix F0, ainsi qu'une limite d'intensité (selon situations de production). A juger selon l'âge et le sexe, s'il s'agit d'un enfant ou d'un adulte. Aussi un aminuement des variations prosodiques qui, tout en variant selon la langue, traduisent l'appauvrissement des possibilités prosodiques du sujet.
Chercheurs en phonétique	CP1	voix moins audible, rocailleuse
	CP2	Fo : mauvais contrôle, ruptures, creaky voice... Timbre rocailleux, etc (conséquences de l'effort laryngé)
	CP3	Niveau d'intensité élevé
	CP4	la voix est dépourvue d'harmoniques et elle est désagréable de dureté
Enseignants	E1	Plus forte, probablement avec un effort pour mieux articuler, plus posée, peut-être moins de variations d'intensité.
	E2	l'esthétique- si elle donne envie de la fuir.
	E3	Sur le son et la posture vocale.
Professeurs de chant	PC1	Visuellement, la sollicitation des muscles extrinsèques se voit, la tension des muscles de la face aussi. Auditivement, la détérioration du timbre se mesure sur les spectres.
	PC2	Sensation d'effort trop grand pendant le chant
	PC3	Acoustiques (voir ci-dessus) mais parfois aussi posturales
	PC4	sur son aspect non harmonieux
	PC5	Sur la fatigue perceptible (érailement, dysphonie, trous dans la voix,...).

2.6. Sixième question : Expérience personnelle du forçage vocal

Médecins	M1	Je ne crois pas, même s'il m'arrive de donner de la voix. Mais crier ou s'emporter à bon escient dans le contexte d'une voix d'insistance bien assumée n'a rien à voir avec le forçage vocal.
	M2	En cas d'énerverment extrême, avec enjeu important, envie de convaincre à tout prix, le plus rapidement possible (voix tendue, trop aiguë, « à bout de souffle »... Qui ne convainc personne !) Lors de cours (enseignement) prolongés (auditoire de 20 à 50 personnes), plus de 1h30 consécutive, en utilisant une voix projetée forte, un peu trop grave (fatigue !)
	M3	communication en milieu bruyant ou entre deux interlocuteurs distants ou pour faire rentrer quelque chose dans la tête de son partenaire qui, en dépit du bon sens, refuse de comprendre
	M4	Fatiguer, oui. Forcer, j'évite mais cela peut arriver lors de prises de parole prolongée. Du type 6 heures de cours dans une même journée

Orthophonistes	O1	Oui, conversations téléphoniques longues, enseignement pendant 6 heures de suite, lors de rééducation de patientes dysphoniques (trop d'empathie ?), chorale, lar de conversations contrariantes.
	O2	Parfois en situation à composante émotionnelle
	O3	Oui, en public.
	O4	Rarement, plutôt dans des situations émotionnellement fortes, quand la voix a tendance à s'en aller et qu'il faut absolument la "rattraper"
	O5	oui, après une longue journée avec des rééducations difficiles ou après avoir trop chanté dans un état de fatigue générale
	O6	Comme je suis bavarde, j'ai tendance à forcer si je suis enrouée ou malade. Parce que j'ai de la peine à tenir un repos vocal... J'ai chanté dans un chœur amateur : je sais qu'au moment du concert public j'étais plus tendue et efforçais de chanter mieux, je forçais ma voix (douleurs après etc.) Mais je ne souffre pas de forçage en temps normal de voix parlée.

Chercheurs en phonétique	CP1	rarement
	CP2	Professionnellement non, même en amphi. Peut être parfois en voix de colère, il m'arrive de sentir l'effort, la tension exagérée et il s'ensuit une toux laryngée...(top secret)
	CP3	des cris, des chants
	CP4	en m'attaquant à des airs du répertoire lyrique alors que je ne suis pas assez entraîné, ou encore lorsque j'essaye de me faire entendre dans une discothèque malgré le potin ambiant

Enseignants	E1	Toutes les situations du 3) sauf la dernière
	E2	exceptionnellement : le chant
	E3	Rarement. Le plus souvent, c'est juste après avoir chanté longtemps, si je parle beaucoup j'ai tendance à forcer car ma voix est plus haute (mes cordes vocales plus tendues) et j'essaye de la baisser involontairement, d'où une fatigue. Ou encore dans les transports en commun. Je m'en rend compte car la gorge me gratte ou me pique.

Professeurs de chant	PC1	A partir d'une erreur de classement de mon professeur de chant (je suis baryton et non basse profonde), j'ai vécu le forçage permanent, constant, régulier, de 20 ans à 40 ans. Enrouements, aphonies, kissing nodules, impossibilité de chanter jusqu'au bout un rôle lyrique, mauvaise voix parlée. Après rééducation et bon enseignement de technique vocale, j'ai chanté après 40 ans des rôles de baryton lyrique sans difficulté, et n'ai plus jamais été en situation de forçage. MAIS, faisant partie d'une population dite « à risque », je gère la tonicité de mon larynx pour pouvoir, par exemple, enseigner régulièrement aux orthophonistes 9 heures de suite la rééducation de la voix (deux stages de 4h plus 1 heure de discussion). La moindre erreur de gestion de mon larynx m'amènerait au forçage.
	PC2	Cf 3 : Pour mieux s'entendre dans un environnement trop bruyant (musical ou autre) et sans assez de réverbération, compensation de la fatigue et de la mollesse qu'elle impliquerait par un surcroît d'effort, trop grande difficulté non surmontée (encore ou définitivement) par rapport à ses moyens vocaux (trop aigu en tessiture, trop lourd en puissance, trop grave et fort)
	PC3	Le moins possible, mais cela arrive : face à un orchestre qui n'écoute pas le chanteur (et on apprend justement à ne pas lutter...) en cours pour donner à entendre des contre-exemples aux élèves, ou dans une situation musicale extrême où il faut réaliser quelque chose qui dépasse nos moyens ou notre technique.
	PC4	oui, à cause de mon métier de chanteuse, surtout en fin de concert, pour les bis et en tout cas pour parler avec le publique après le concert
	PC5	Quand je chante un répertoire qui est trop « lourd » pour moi ou quand j'ai à chanter ou parler longtemps alors que je suis malade (rhume, laryngite, etc).

2.7. Septième question : Population concernée par le forçage vocal

Médecins	M1	Enseignants surtout. Mais encore tous les professionnels de la voix qui par définition sont mis dans l'obligation de parler ou de chanter même les jours où ils ne sentent pas en voix Aptitudes aux réactions immédiates (primarité au sens de la caractérologie de Le Senne) Perfectionnisme Suractivité professionnelle
	M2	Enseignants stressés (élèves agressifs, inattentifs, bruyants) - Travailleurs en milieu bruyant (usine, chantier), plutôt féminins. - Mères de famille stressées (nombre d'enfant indifférent !) - Coiffeuses (produits irritants, parlent beaucoup, bruit ambiant) - Jeunes chanteur(s)(pop, lyrique), ou jeunes enseignants, après 2 ans environ de formation ou d'activité professionnelle - Personnes au comportement très réservé, autoritaire, avec travail à responsabilité
	M3	X
	M4	Je ne voit que des chanteurs ou des comédiens en consultation. Leur forçage est lié à des surmenages vocaux le plus souvent. Ces patients sont dans des utilisations « performancielles » de leur voix et ils sont amenés à ne pas respecter les signes de fatigues que leur voix leur donne. Les raisons de la « non écoute » de ces signes est complexe et constitue l'aspect le plus important de la prise en charge.
Orthophonistes	O1	Personnalités plutôt impulsives, voire hystériques, ou au contraire personnalités inhibées-nerveuses. Chez les enfants, instituteurs, professeurs, chanteurs, comédiens, mère de grandes familles, plutôt des femmes dans mon expérience. Contexte allergique (craie, poussière).
	O2	X
	O3	Pathologies neuro-dégénératives (je suis neurologue, spécialisé dans la maladie de Parkinson)Le forçage peut être observé plutôt dans les syndromes parkinsoniens.
	O4	enseignement, avocats, chanteurs, comédiens, commerciaux... personnalités plutôt volontaires, ayant beaucoup investi leur voix, comme instrument de séduction, persuasion ou autorité personnalités anxieuses
	O5	enseignants, chanteurs, personnalités anxieuses
	O6	Je ne vois pas cette population. Beaucoup de chanteurs amateurs et d'enseignants je crois Alcooliques et gros fumeurs viennent aussi, ils sont peu demandeurs....
Professeurs de chant	PC1	Il n'existe pas, à mon avis, de professions à risque concernant la voix. Une profession à risque fait prendre le même risque à tous les membres de la profession (dresseur de tigres, peintre de la tour Eiffel). Certaines professions sollicitent la voix plus que d'autres, à la criée des marées, forains, enseignant, maître nageur. Mais la profession la plus risquée pour la voix est chanteur. En proportion, on rencontre moins de dysfonctionnement chez les chanteurs professionnels que dans l'enseignement par exemple. Il existe par contre des circonstances à risque internes ou externes au sujet, et des physiologies à risque (caisse de résonance, port de tête, forme de la bouche, position des pommettes, état endocrinien, hérédité etc) L'état affectif des patients (inhibition, stress) n'a pas d'influence sur la voix, mais sur le geste qui détermine la voix (pressions sous glottique, accollement du larynx, forme de la caisse de résonance).
	PC2	X
	PC3	Chanteurs, mais c'est rare, en général ils ont plutôt besoin de construire leur instrument, plus rarement de le réparer
	PC4	parmi mes élèves de chant j'ai eu des personnes qui avaient des problèmes de forçage vocal. Cela était dû à la non connaissance de l'emploi des muscles liés à la phonation et a disparu. Néanmoins quelques uns ont continué à l'avoir à cause: - de l'envie de vouloir chanter en pensant que les vocalises étaient inutiles et en sous estimant le travail vocal progressif et contrôlé - de leur personnalité coincée - de leur "éducation" ou façon d'être qui n'accepte pas un "lâcher prise" primordial qui est la base de tout travail sur la voix
	PC5	X

2.8. Huitième question : Qualité vocale des voix forcées

Médecins	M1	Altération variable du timbre en rapport avec l'altération de la muqueuse des plis vocaux : Rauque ou éraillée ou irrégulière ou voilée ou sourde ou voilée... Donnant l'impression d'effort ; ou encore, étouffante.
	M2	Non, pas de voix « type »
	M3	X
	M4	Non. La voix parlée d'un chanteur qui force peut être « normale » ou altérée (appauvrissement du timbre). La voix parlée d'un comédien, elle sera altérée même en situation conversationnelle.

Orthophonistes	O1	Coup de glotte fréquent aux attaques. Modulation réduite; intensité mal adaptée (trop ou pas assez), voix "soufflée", Fo plus grave, parfois plus aigue, raucité, parfois bitonalité, phrases respiratoires longues avec reprise d'air en urgence, débit rapide assez souvent.
	O2	Serrée, soufflée, souvent rauque, parfois aigue.
	O3	Voix rauque, soufflée, difficulté du contrôle respiratoire et laryngé.
	O4	serrée, rauque, un peu soufflée, légèrement détimbrée
	O5	de nombreuses caractéristiques peuvent s'intriquer et il est quasiment impossible de rapprocher un type de voix d'un type de patient
	O6	X

Professeurs de chant	PC1	En règle générale, timbres trop vocaliques, absence d'harmoniques extra vocaliques. L'état de la muqueuse du larynx et la forme prise par la caisse de résonance déterminent la qualité harmonique du son.
	PC2	X
	PC3	Fuite d'air, vibrato trop large, ou/et irrégulier, ou au contraire vibrato inexistant (empêché par un serrage laryngé permanent, timbre mat)
	PC4	Les voix de ces personnes sont désagréables à l'entente et transmettent une sensation de gêne, malaise, fatigue et tension
	PC5	X

2.9. Neuvième question : Symptômes du forçage dont se plaint le patient

Médecins	M1	Fatigue, lassitude à la phonation prolongée, voire étourdissement (en cas de surventilation dû à la désorganisation du souffle). Douleurs cervicales, voire abdominales.
	M2	<p>dysesthésies pharyngées</p> <ul style="list-style-type: none"> - sensation de serrement cervical, d'étouffement, tensions sous-mandibulaires - « boule » dans la gorge, corps étranger hypopharyngé, mucosités adhérentes - hémhage (raclement de gorge) fréquent - fatigue vocale (enrouement, voix voilée)
	M3	
	M4	Fatigue, perte d'efficacité, douleurs, réduction de l'étendue ou de la portée, perte de la dynamique voire de la justesse pour certains.
Orthophonistes	O1	"J'ai des aphonies régulières, j'ai mal dans la gorge, je sens un corps étranger dans la gorge, on m'appelle Monsieur au téléphone, je ne peux plus chanter, ça me fatigue de parler, je racle souvent, j'ai une gêne à la déglutition..."
	O2	Fatigue vocale en situation de stress, en fin de journée, douleurs dans les épaules et le cou, voix inefficace, patients n'aiment pas leur voix.
	O3	Mes patients ne consultent pas pour le forçage vocal, c'est l'observation clinique qui note ce problème.
	O4	fatigue vocale et essoufflement, épisodes d'aphonie, sensations de brûlures laryngées lors de la phonation, contractions cervico scapulaires (parfois)
	O5	Serrage, fatigabilité, manque d'intensité et d'efficacité, brûlures et picotements dans la gorge, de tête et sans
	O6	X
Professeurs de chant	PC1	La gêne est vocale ; inaudibilité, absence de puissance, de résistance, la souffrance est localisée au larynx et aux muscles du cou. On peut aller jusqu'au traumatisme affectif.
	PC2	X
	PC3	Gêne : déphasage entre ce qu'ils veulent produire musicalement et ce qu'ils parviennent à réaliser. Fatigue rapide en situation de chant. Parfois douleurs laryngées, mais aussi cervicales, dorsales...
	PC4	X
	PC5	X

2.10. Dixième question : Autres symptômes du forçage dont le patient n'est pas toujours conscient

Médecins	M1	Attitude caractéristique au moment de la phonation : posture de la voix de détresse avec flexion du haut du dos, projection du visage vers l'avant et crispations faciales
	M2	discoordination pneumophonique - comportement laryngé : attaque des sons avec participation marquée des fausses cordes vocales ; compression laryngée antéro-postérieure en phonation / serrage aryténoïdien marqué en phonation - réflexes pharyngés exacerbés - langue hypertonique, bombée lors de l'examen laryngé épipharyngoscopique
	M3	X
	M4	Pour les chanteurs la perte des minima est le signe d'une fatigue vocale dont le forçage est l'origine ou la conséquence.
Orthophonistes	O1	Crispation des muscles sterno-cléido-mastoïdiens, jugulaires saillantes, respiration plutôt thoracique supérieure lors de la phonation, posture du corps.
	O2	Mobilisation du visage souvent les sourcils, serrage au niveau du cou.
	O3	X
	O4	La contraction des muscles de la mâchoire, langue, cou, épaules n'est souvent pas consciente oui elles sont nombreuses les patients n'ayant pas toujours conscience de leur comportement de forçage surtout en ce qui concerne la maîtrise du geste elle même mais on rencontre aussi des gens qui ont des signes de forçage avec une voix altérée et qui n'ont aucune sensation associée
	O5	
	O6	X
Professeurs de chant	PC1	le traumatisme ou blocage psychique occulte le mauvais geste. Le patient ne sait plus comment obtenir un bon son
	PC2	X
	PC3	Grimaces, asymétries, torsions pendant la phonation et juste avant.
	PC4	X
	PC5	X

2.11. Onzième question : Rééducation du forçage vocal

Médecins	M1	Information théorique sur le mécanisme du cercle vicieux du forçage vocal Amélioration de la gestion du niveau de tension psychomotrice grâce à la pratique personnelle de la relaxation (5 à 10 minutes par jour) Entraînement à la maîtrise du souffle Découverte et maîtrise du comportement de projection vocale avec ses quatre caractéristiques :1) Certitude d'être efficace 2) redressement du corps 3) Regard en face Mise en route automatique du souffle abdominal. Exercices vocaux parlés et chantés
	M2	Rééducation vocale (placement vocal, coordination pneumo-phonique, détente cervico-laryngée, hygiène vocale, amélioration de la voix projetée) - Fréquemment, traitement antacide (inhibiteur pompe à protons) si patient symptomatique ou si signes laryngés de reflux gastro-oesophagien - Fréquemment, physiothérapie (massages nuque, dos, ceinture scapulaire)
	M3	
	M4	Rien de systématique, sauf les prises de conscience des efforts et de leur localisation et de leur possibilité de réduction. Le travail passe souvent par l'abandon de la qualité du son au premier chef au profit de la « fluidité » du geste vocal.

Orthophonistes	O1	PRISE DE CONSCIENCE des attaques dures qu'elles font souvent inconsciemment, sur l'importance de faire des silences dans la parole, sur le feedback auditif et visuel du forçage, travail du souffle + RELAXATION
	O2	Rééducation de la respiration, du bon geste vocal (type Amy de la Brétèque, exercice de paille, de vocalisations douces), travail de l'intonation, relaxation.
	O3	Rééducation orthophonique centrée sur la relaxation, la reprise des contrôles respiratoires et laryngés. NB : dans la maladie de Parkinson, la méthode de rééducation LSVT (Lee Silverman voice therapy) se base sur l'effort vocal. Elle peut avoir des effets délétères et doit donc être contrôlée.
	O4	Parfois relaxation, détente laryngée (massages), respiration, travail de coordination pneumophonique, exercices vocaux sur la vibration des cordes vocales, exercices spécifiques en fin de rééducation(voix projetée, voix d'appel)
	O5	prise de conscience du forçage, respiration, relaxation, pose de voix (mais on pourrait développer ça sur de nombreuses pages)
	O6	X

Professeurs de chant	PC1	Après le rétablissement d'un larynx sain (médecin), il convient de parvenir à quatre gestes : 1/ une maîtrise par les obliques et transverses, d'un flux d'air de 15 à 30 secondes au travers d'une paille sans aucune effort 2/ Sur ce souffle, un accolement des cordes vocales sans tension extrême permettant, en utilisant les mécanismes I ET II, d'obtenir un son de qualité constante, sur toute la tessiture de la personne, sans aucun effort ni tétanisation 3/ une recherche de la bonne position des résonateurs pour obtenir un timbre extra vocalique sans aucun effort, sur toute la tessiture, et quelle que soit la voyelle. 4/ une tonification du larynx par la couverture des sons afin de lui permettre de résister à une augmentation de pression sous glottique légère, nécessaire pour soutenir certaines notes hautes, ou obtenir une augmentation d'intensité sans effort.
	PC2	X
	PC3	Travail technique vocal, affinement de la boucle audio-phonatoire (écoute et analyse se sa propre production vocale, puis ajustements). Quelque conseil et exercices posturaux.
	PC4	X
	PC5	X

2.12. Douzième question : Définition de l'efficacité vocale

Médecins	M1	La juste adaptation des trois qualités de la voix (tonalité, intensité, timbre) aux objectifs de l'acte vocal entrepris. (Cela peut être le silence !) Cela nécessite 1) l'absence de forçage et 2) l'attention portée à l'effet de cette voix dans l'autre (interlocuteur ou auditoire) plus qu'à cette voix elle-même.
	M2	Un équilibre harmonieux entre degré de fermeture glottique, tensions intra-cordales, pression sous-glottique et ajustement des cavités de résonance permettant l'émission d'une voix timbrée, stable, confortable, d'intensité adaptée au lieu, au type de communication et aux interlocuteurs ciblés.
	M3	optimisation de l'intelligibilité d'une séquence vocale
	M4	C'est la possibilité de moduler les différents paramètres acoustiques de la voix sans effort.
Orthophonistes	O1	La possibilité d'utiliser et d'adapter sa voix en toutes circonstances, avec facilité et aisance, sans la maltraiter + une "bonne" perception de l'interlocuteur.
	O2	Pouvoir utiliser sa voix à bon escient avec un minimum d'effort, et dans un geste non brutal pour les cordes vocales.
	O3	C'est un indice représentant le niveau de coordination pneumophonique (l'efficacité augmente si le "produit" vocal est réalisé à moindre coût. Expiration moins forte (débit) et "résistance" laryngée plus faible.
	O4	Pouvoir être entendu et compris par un auditoire situé à quelques mètres, en milieu calme.
	O5	Pouvoir appeler à plusieurs mètres une bonne résonance qui permet aussi d'avoir une voix audible qui porte sans effort
	O6	Pour un professionnel du chant ou pour un acteur, il s'agit de la capacité à produire dans les meilleures conditions les effets vocaux et articulatoires souhaités par lui et par son public au cours de la prestation. Cela englobe pour moi son intelligibilité, la portée de sa voix, l'optimisation des effets de timbre, d'intensité et de qualité vocale, de variations prosodiques. Cela englobe la voix, mais aussi la résonance et l'articulation, donc. Production + perception. En voix parlée et en consultation orthophonique, les personnes qui consultent de sont pas des spécialistes de la voix, ils cherchent juste à diminuer la portée de leurs symptômes pathologiques... je parle pour les problèmes de voix rencontrés dans le cadre de maladies neurogènes dont je m'occupe (p.ex. Sclérose en plaques, Maladie de Parkinson etc). Ils le font pour eux-mêmes ou, s'ils sont peu conscients du trouble ou peu demandeurs, pour faire plaisir à la personne demandeuse du traitement (médecin ou proche)
Chercheurs en phonétique	CP1	c'est une voix intelligible qui est la résultante d'une activation des cordes vocales demandant un effort minimum. La fatigue vocale mettra plus de temps à venir
	CP2	Etre capable de contrôler sa voix en intensité et en hauteur avec l'effort minimum
	CP3	Le fait de pouvoir varier la hauteur de la voix ou de pouvoir passer aisément de la plus élevée à la plus aiguë
	CP4	Sans qu'il y ait sensation d'effort, le locuteur est entendu à plusieurs dizaines de mètres et la voix est nette
Enseignants	E1	Ben je ne sais pas trop ce que c'est, intuitivement c'est le rapport entre le résultat (la capacité à faire passer le message) et l'effort (musculaire + respiratoire).
	E2	savoir ce que l'on peut obtenir et en fonction de cela l'utiliser au mieux
	E3	C'est la façon d'émettre la voix qui sera la plus efficace pour être intelligible pour l'auditoire. Cela implique de trouver une hauteur de son adéquate, avec le minimum d'énergie nécessaire et donc d'adapter en fonction des situations l'articulation, la projection, la force...
Professeurs de chant	PC1	Audibilité (émergence du singing formant) Modulation en amplitude et en intensité (résistance du larynx à la pression sous glottique) Tonicité du larynx (résistance à la pratique des heures, des jours, des années)
	PC2	Mise en œuvre harmonieuse du corps et de l'imagination (des facultés spirituelles en général) pour obtenir le résultat souhaité dans un effort minimal et souple (mais sans mollesse pour autant)
	PC3	Basiquement, c'est le rapport intensité/effort musculaire. Plus ce rapport est grand, plus l'efficacité est grande. Mais le critère intensité en décibels n'est pas le seul. C'est plutôt la perception de l'auditeur qui est le critère important : donc singing-formant, rapport consonnes/voyelles ...
	PC4	connaissance de ses possibilités de voix parlée où chantée contrôle constant de la situation et perception du degré de débit à fournir pour une durée déterminée, sens d'économie vocale
	PC5	C'est l'utilisation de la voix parler ou chanter <u>sans fatigue</u> pour l'obtention du résultat souhaité.

2.13. Treizième question : Définition de l'effort vocal

Médecins		Au sens positif cela peut-être l'effort de mobilisation psychomotrice pour mener à bien une entreprise vocale difficile. Au sens péjoratif c'est le forçage vocal avec ce porte à faux résultant de l'usage du mécanisme de la voix de détresse (flexion du haut du dos, souffle vertébral) là où le comportement de projection vocale était indiqué.
	M1	
	M2	Une situation ponctuelle, « de crise », durant laquelle la musculature cervicale, glottique et thoracique (voire plus) réalise une tension importante, exagérée, pour émettre un son
	M3	l'ensemble des coûts énergétiques exigés à l'élaboration d'une séquence phonatoire
	M4	C'est l' « énergie » globale mise en jeu pour la production de la voix.

Orthophonistes	O1	Le fait "d'emmener" sa voix au-delà de ses possibilités acoustiques et anatomiques
	O2	L'effort vocal non pathologique est la capacité à projeter sa voix, à chanter en utilisant un geste vocal approprié. L'effort vocal pathologique est l'utilisation d'un geste brutal, tendu avec un serrage vocal, pour peu d'efficacité vocale.
	O3	L'effort vocal consiste à augmenter l'expiration (P expiratoire) et/ou à augmenter la résistance laryngée.
	O4	L'effort vocal est ce qui accompagne une mise en situation un peu extrême de la voix (voix projetée ou chantée par exemple). Une préparation respiratoire, une position du corps adaptée, la volonté de toucher l'auditoire. Pour moi ce n'est pas pathologique, ça le devient si l'utilisation de l'appareil vocal est inadaptée
	O5	des situations où le sujets doit mobiliser une dynamique du geste vocal plus importante pour que la voix soit efficace
	O6	Effort : pour moi au sens 'apprentissage' : je fais un effort vocal quand je m'adapte à une nouvelle situation de phonation, peu habituelle, et qui me demande de me surpasser, quand je chante je quand je donne un cours p.ex. J'y parviens ou non, dans quel cas je risque le malmenage/le forçage....

Chercheurs en phonétique	CP1	Une activation plus importante des cordes vocales pour se faire comprendre. Elles sont alors plus tendues
	CP2	C'est pousser sa voix (tension, contrôle de Fo) à ses limites, sans atteindre le forçage
	CP3	X
	CP4	le fait de forcer sur les cordes vocales au lieu de soutenir au niveau du diaphragme

Enseignants	E1	Idem je ne peux pas en donner une def précise, mais on imagine un effort musculaire (les cordes vocales sont plus « tendues ») et aussi respiratoire (plus de souffle).
	E2	quand on a l'impression que les cordes tirent. Une impression légèrement douloureuse
	E3	C'est l'énergie vocale dépensée. Issue de la mise en action des muscles de l'appareil phonatoire

Professeurs de chant		Chez les professionnels de la voix, il peut y a voir effort qui éprouve la résistance du larynx (chanteurs lyriques ou de variété travaillant beaucoup, intensément, sur toute leur tessiture, à forte intensité, et dans l'adversité) Idem chez des forains, marchands ambulants, etc. L'effort n'est donc pas synonyme de forçage quand le larynx PEUT résister à la pression sous glottique. Dès qu'il y a dépassement de cet équilibre, il y a forçage
	PC1	
	PC2	C'est l'effort contrôlé qui permet d'obtenir l'efficacité vocale définie au 12)
	PC3	Effort n'est pas forçage. Le forçage, c'est l'effort mal géré. L'effort, c'est l'ensemble des actions musculaires nécessaires à la production optimale du son.
	PC4	donner plus de ce que nous avons
	PC5	Pression sous glottique trop importante, insuffisance de la protection.

2.14. Quatorzième question : Définition de la fatigue vocale

Médecins	M1	C'est le résultat du forçage vocal dont le déclenchement peut être aussi bien le fait de facteurs purement dysfonctionnels que de lésions organiques primaires de l'appareil vocal
	M2	Impression du patient d'une émission vocale difficile, pénible, demandant un effort de volonté, avec ou sans altération vocale audible. Le patient a envie de se taire à ces moments-là.
	M3	épuisement des ressources énergétiques disponibles à l'élaboration d'une séquence phonatoire
	M4	C'est l'altération des modulations des différents paramètres acoustiques sans effort.
Orthophonistes	O1	C'est la prise de conscience physique+ acoustique de l'effort vocal que l'on a fourni. (sensation de gêne, douleur, essoufflement, aphonie, manque d'appétence à la communication par fatigabilité.
	O2	
	O3	Elle se traduit plutôt par le "produit" : énergie acoustique plus basse.
	O4	La conséquence du forçage vocal qui se caractérise par une perte du rendement vocal et parfois des douleurs
	O5	lorsqu'après un effort vocal le corps "signale" au sujet par une altération de la voix ou des sensations kinesthésiques que le repos est nécessaire. En pathologie, la fatigue vocale correspond à un sujet qui a perdu l'efficacité de sa voix
	O6	C'est peut être faux mais pour moi c'est synonyme de forçage....
Chercheurs en phonétique	CP1	La conséquence d'un effort vocal trop important ou trop long
	CP2	Atteindre un état où l'on a du mal à parler (ou chanter) où l'on sent une fatigue / gêne laryngée
	CP3	X
	CP4	à force de vices d'émission, toutes sortes de pathologies s'installent dont celles où les cordes vocales ne sont plus jointives, d'où voix éraillée et presque inaudible.
Enseignants	E1	Personnellement après 3h de TD avec des groupes nombreux et pas très motivés, j'ai la gorge sèche, j'ai du mal à forcer la voix précisément, j'ai souvent carrément mal à la gorge. La fatigue vocale est donc un désagrément passager dû à une pratique trop intense ou prolongée ou mal maîtrisée du forçage vocal.
	E2	Une utilisation anormale. quand on a plus de voix pour l'avoir trop utilisée
	E3	C'est la conséquence d'un trop grand effort vocal. Cela peut se manifester par une perte de puissance vocale, une aphonie, une irritation ou pire, des kystes ou lésions.
Professeurs de chant	PC1	Terme impropre : il faut dire Fatigue du larynx (ce ne sont ni les zygomatiques ni les transverses qui sont fatigués, et encore moins la voix) qui déclenche une voix de mauvaise qualité
	PC2	Un surcroit d'effort des cordes vocales
	PC3	C'est l'altération des qualités vocales
	PC4	rentrer dans une situation désagréable et non contrôlée de phonation
	PC5	X

2.15. Quinzième question : Définition du malmenage vocal

Médecins	M1	C'est pour moi un équivalent du forçage, mais avec une nuance de traficotage volontaire surajouté des caractères acoustiques de la voix
	M2	Utilisation extrême (aigu, grave, trop fort) ou en force de ses organes phonatoires, plusieurs fois par jour ou par semaine, entraînant des altérations vocales, y compris tissulaires, répétées (enrouement, voile, aphonie).
	M3	production vocale sur un mode anti-physiologiques en terme de timbre, et/ou de fréquence, et/ou de volume sonore émis
	M4	C'est le non respect des indices de souffrance que peut fournir la voix.

Orthophonistes	O1	C'est le fait de ne pas faire attention aux limites vocales. C'est continuer à forcer sur sa voix alors qu'on est déjà fatigué vocalement: commerçants sur les marchés, instituteurs,... Ils peuvent consciemment ou inconsciemment "ignorer" leur fatigue vocale.
	O2	C'est une utilisation inappropriée de l'outil vocal
	O3	C'est une mauvaise coordination pneumophonique. Cela peut être une stratégie de compensation mais à potentiel délétère
	O4	Le malmenage serait peut être un stade avant le forçage vocal, une sorte d'utilisation inadaptée de la voix au quotidien mais qui n'est pas encore accompagnée de signes vocaux entravant la vie professionnelle et sociale.
	O5	le sujet utilise mal sa voix, pas de notion de durée
	O6	Malmenage= forçage continu ou répété de la voix lors d la prise de parole.

Chercheurs en phonétique	CP1	c'est lorsqu'on fait régulièrement ou trop d'effort pour se faire entendre
	CP2	je vois mal la différence avec la voix forcée ou alors ce serait une mauvaise utilisation régulière de la part d'un individu, avec d'éventuelles conséquences pathologiques.
	CP3	le fait de solliciter la voix trop longtemps
	CP4	forcer trop fréquemment sur ses cordes jusqu'à ce qu'elles ne soient plus jointives lors de l'émission

Enseignants	E1	Jamais entendu parler. J'imagine ca surtout pour des chanteurs qui se « cassent » la voix à force de travailler de manière trop intense, ou trop jeune, etc ...
	E2	une utilisation mauvaise mais quotidienne.
	E3	C'est le fait de trop solliciter sa voix et de mauvaise manière. Le forçage vocal est un malmenage de la voix. Il y a malmenage quand il y a fatigue vocale de manière répétée et trop souvent (voire continuellement).

Professeurs de chant	PC1	Quand une personne sollicite son larynx avec trop de pression, il y a forçage. Dès qu'elle continue, insiste (pour poursuivre son exposé, sa conversation etc), elle malmène son larynx et l'amène à un état pathologique où il pourra y avoir surmenage et lésion
	PC2	Surcroit d'effort systématique des cordes vocales voire du corps (souvent aussi de l'intellect)
	PC3	Mauvaise gestion de l'effort vocal. Sous tonicité, sur-tonicité, ou mauvaise répartition de l'effort.
	PC4	l'emploi inconscient et brutale de la voix ou les conséquences d'un comportement psychique troublée ou malade
	PC5	Utilisation d'une pratique vocale inappropriée conduisant à une fatigue vocale.

2.16. Seizième question : Définition de l'échauffement vocal

Médecins	M1	Se mettre en voix progressivement, en commençant par des productions moyennes pour aboutir en quelques minutes <i>avec des pauses</i> , à des productions plus contrastées. Mais attention à ne pas trop s'amuser avec le timbre. S'arrêter avant que ça picote !!!
	M2	Cinq à dix minutes de fredonnement et de glissades bien maîtrisées, sans précipitation, vers l'aigu et vers le grave. Vocalises avant chorale ou chant solo - Boire (eau) avant un effort vocal soutenu - Maintenir sa musculature de la nuque et du cou au chaud (foulard, linge éponge, écharpe) jusqu'à la prestation ou vocale ou entre deux prestations.
	M3	préparation douce et progressive de l'appareil phonatoire à un exercice d'effort vocal
	M4	Pas grand chose d'intéressant. Soit c'est réquisitionner l'énergie nécessaire pour pouvoir fournir l'effort nécessaire à la production vocale, c'est la vision négative. Soit c'est chercher les productions vocales nécessitant les mobilisations minimales. Dans ce cas ça ne porte plus le nom d'échauffement.

Orthophonistes	O1	L'échauffement vocal est surtout utilisé dans la pratique de la projection vocale (chant, théâtre...). Tout comme l'athlète qui "chauffe" ses muscles avant l'épreuve. L'échauffement est une préparation progressive qui consiste à assouplir les muscles de la phonation + respiration en vue d'une situation de projection vocale ultérieure. C'est tenter d'emmener petit à petit sa voix vers des extrêmes à moindres coûts (= avec efficacité donc)
	O2	Mise en route musculaire (tête + épaules), mise en vibration douce des cordes vocales (exercice de paille, exercice de la mouche)
	O3	Améliorer la coordination pneumophonique (exercices sur l'étendue tonale et vocale).
	O4	Reprendre les bases d'un bon fonctionnement vocal dans une utilisation "a minima" pour les appliquer ensuite en situation "extrême" (chant, voix projetée) sans risque de forçage vocal. Pour moi il n'y a pas de correspondance entre un échauffement sportif et un échauffement vocal car l'échauffement n'agit pas directement sur le muscle vocal.
	O5	permettre au sujet d'amener progressivement le sujet vers un travail vocal de plus en plus important.
	O6	Chanteurs : mise en tension très progressive des cordes vocales, associé à une pratique de la relaxation. Dissociation effort-tension.

Chercheurs en phonétique	CP1	Ca dépend de l'utilisation de la voix : Pour le chant : vocalises Pour la parole : peut être au réveil, ne pas faire d'efforts brusques, parler doucement, se racler la gorge avant un discours et boire un peu d'eau
	CP2	Je ne pratique pas le chant mais j'imagine qu'il s'agit comme en sport de préliminaires permettant de roder les cordes vocales (et les articulateurs). Ma voix du matin, lors de mes premiers cours est effectivement un peu laborieuse....racllements de gorge, toux sèche par exemple
	CP3	l'exercice qui consiste à s'entraîner en faisant travailler la voix avant de chanter
	CP4	le terme me semble approprié en ce sens qu'aucun effort ne doit porter sur les cordes vocales. La seule préparation consisterait, selon moi, à assurer une coordination entre diaphragme et lèvres via la respiration. Je pense à des exercices faisant intervenir des consonnes dont le M, par exemple émettre des séries de Momomomo et Mamamama pour "placer la voix dans le masque"

Enseignants	E1	Pour le chanteur -que je ne suis pas-, j'imagine bien vocalises et arpèges, ... Personnellement, c'est quelque chose que je ne fais jamais
	E2	faire des exercices vocaux (vocalises)
	E3	Dans le cadre de la voix parlée il suffit de commencer la journée sans trop forcer ni trop monter la hauteur, le temps que les cordes vocales se chauffent, se tendent (comme tous les muscles). Dans le cadre du chant, il convient aussi de préparer l'action des cordes vocales par des exercices progressifs pour éviter un « claquage ». Ceci dit il y a bien des manières de s'échauffer la voix. En Russie, un chanteur ne fait pas d'exercice de vocalise avant de chanter son morceau mais il se chauffe directement sur un air large peu rapide, bien connu, pas trop aigu, où l'on peut « étaler sa voix ». En France on fait plutôt des vocalises qui permettent de monter progressivement dans l'aigu et d'aller plus vite. Ailleurs c'est encore différent ...

Professeurs de chant	PC1	<p>Pour un orateur, la première heure de la journée est déterminante, du réveil à la sortie de la maison. Il ne s'agit pas d'échauffer la voix, mais de retrouver la mobilité des muscles intrinsèques et extrinsèques du larynx, du voile du palais et de la bouche pour une voix de qualité sans effort. Ensuite, la recherche de l'intensité doit se faire progressivement. On peut se rendre aphone dans la première heure sans avoir crié. On peut a contrario se rendre résistant dans la même heure, et pouvoir hurler ensuite toute la journée. Il suffit de ne JAMAIS demander au larynx un effort qu'il ne peut pas faire.</p> <p style="text-align: center;">Idem pour les chanteurs.</p> <p style="text-align: center;">Le larynx étant lié au système endocrinologique, il convient de « réveiller » le corps en même temps que le larynx</p>
	PC2	Retrouver quasiment instantanément la voix que l'on maîtrise et s'en servir jusqu'à atteindre assez vite (5 minutes) le maximum de ce que l'on sait faire, ou simplement ce qui permet de maîtriser bien ce que l'on a à chanter, le reste est du perfectionnement à mon avis. En somme : retrouver des réflexes et vérifier le bon fonctionnement corporel et musculaire (les muscles eux mêmes sont quasiment immédiatement échauffés). C'est de toute façon très rapide (moins de 5 minutes voire immédiat) si l'on est bien portant, sinon il faut trouver des trucs occasionnels pour surmonter ses maux.
	PC3	Eveiller le corps et mettre ses potentialités au service de la production vocale. Faire fonctionner progressivement les muscles de la phonation (qui sont très divers) afin que dans l'effort il n'y ait pas de forçage. Il y a aussi un aspect turgescence de la muqueuse vocale qui favorise son accolement.
	PC4	dans l'éveil progressif de tous les muscles liés à la phonation et dans l'adressement de sa pensée dans le sens du travail vocal à venir
	PC5	<p>L'échauffement est avant tout un moyen de se connecter avec l'activité que l'on va aborder : le chant ; C'est une mise en condition mentale et corporelle. Points abordés lors d'un échauffement :</p> <ul style="list-style-type: none"> Détente, relaxation, concentration sur le corps Ouverture sur le lieu Travail sur le souffle, la sensation diaphragmatique Travail sur la voix, le legato, la couleur, les voyelles Travail sur l'articulation (prononciation), le piqué, les consonnes Travail sur l'oreille, l'audition, la justesse

3. ANALYSES NON PRESENTEES DANS LE MANUSCRIT

3.1. Le concept d'efficacité vocale

3.1.1. Un concept sans jugement de valeur et objectif par la physique

Pour certains, l'efficacité vocale est un objet en soi sans lien direct avec le forçage vocal ni aucune pathologie vocale, et donc sans quelconque jugement de valeur. Bien que cette conception neutre soit majoritaire dans la littérature, elle n'est représentée que par deux personnes de notre questionnaire (*indice représentant le niveau de coordination pneumophonique (pouvant être grand ou petit) ; rapport entre l'effort (musculaire + respiratoire) et son résultat*).

3.1.2. Un concept positif et subjectif

L'efficacité vocale est bien plus souvent conçue de façon subjective par les personnes ayant répondu à notre questionnaire (*harmonieux ; confortable ; stable ; adaptée ; optimisation ; bonne perception ; juste adaptation ; bon escient ; bonne résonance ; dans les meilleurs conditions ; optimisation ; au mieux ; harmonieuse*).

3.1.3. Un concept à l'opposé du forçage vocal

L'efficacité vocale est enfin considérée par certaines personnes comme un concept opposé au forçage vocal. Une personne exprime explicitement cette opposition (*absence de forçage*), quand d'autres avaient évoqué l'inefficacité dans les définitions du forçage vocal (*l'énergie musculaire n'est pas en rapport avec l'efficacité acoustique du son produit*). Mais surtout, l'efficacité vocale est définie par des attributs opposés de ceux du forçage :

- A la notion d'effort et d'excès caractérisant le forçage vocal répond au contraire l'idée d'absence ou de minimisation de l'effort dans la définition de l'efficacité vocale (*confortable ; sans effort ; avec facilité et aisance ; un minimum d'effort ; à moindre coût ; effort minimum ; aisément ; sans qu'il y ait sensation d'effort ; minimum d'énergie nécessaire ; effort minimal et souple ; économie vocale*). Dans le même esprit, la tension qui accompagne le forçage est remplacée par l'harmonie et la souplesse pour l'efficacité vocale (*un équilibre harmonieux ; confortable ; facilité ; aisance ; harmonieuse ; aisément ; effort souple*).
- A l'aspect nocif du forçage, générateur de fatigue ou de symptômes, s'opposent les notions d'endurance, de préservation et d'absence de symptômes pour l'efficacité vocale (*sans la maltraiter [la voix] ; geste non brutal ; la fatigue vocale mettra plus de temps à venir ; tonicité du larynx (résistance à la pratique des heures, des jours, des années) ; sans fatigue*).
- A l'inadéquation/inadaptation du mode phonatoire associée au forçage vocal répliquent l'équilibre et l'adaptation des caractéristiques vocales dans la définition de l'efficacité. Il s'agit d'une adaptation générale (*un équilibre harmonieux entre degré de fermeture glottique, tensions intra-cordales, pression sous-glottique et ajustement des cavités de résonance ; utiliser sa voix à bon escient ; optimisation des effets de timbre, d'intensité et de qualité vocale, de variations prosodiques, mais aussi la résonance et l'articulation ; hauteur de son adéquate*), à la situation (*intensité adaptée au lieu, au type de communication et aux interlocuteurs ciblés ; adapter sa voix en toutes circonstances ; adapter l'articulation, la projection, la force en fonction des situations*) ou aux objectifs de la phonation (*adaptation des trois qualités de la voix (tonalité, intensité, timbre) aux objectifs de l'acte vocal entrepris ; optimisation de l'intelligibilité*).
- A une voix forcée, désagréable, rauque ou instable, se substitue une qualité vocale « *timbrée, stable, nette* » avec une « *bonne résonance* » ou un « *singing formant* » dans le cas d'une voix efficace.
- Enfin l'impuissance physique, technique ou psychologique évoquée dans les situations de forçage (mais pas spécialement dans la définition du forçage) trouve son pendant dans la notion de capacité et de possibilité, en ce qui concerne l'efficacité vocale. Celles-ci sont de 3 ordres : Elles concernent dans certains cas la capacité à communiquer et à transmettre un message ou des émotions (*pouvoir être entendu et compris ; capacité à faire passer*

le message). Dans d'autres cas, il s'agit de la capacité à contrôler les caractéristiques de sa propre voix (*possibilité de moduler les différents paramètres acoustiques de la voix ; possibilité d'utiliser sa voix en toutes circonstances ; capacité de produire [...] les effets vocaux et articulatoires ; être capable de contrôler sa voix en intensité et en hauteur ; pouvoir varier la hauteur de la voix, passer de la plus élevée à la plus aiguë ; modulation en amplitude et en intensité*). Enfin, l'efficacité vocale est parfois conçue comme une conscience de ses capacités et limites, permettant de ne pas les dépasser et ainsi d'éviter le forçage vocal (*savoir ce que l'on peut obtenir ; connaissance de ses possibilités, contrôle constant de la situation*).

3.2. Le concept de « Warm up » ou Echauffement vocal

Pour finir, nous avons cherché à mieux comprendre la notion d'échauffement vocal (ou Warm-up). Les personnes interrogées dans notre questionnaire conçoivent l'échauffement vocal de plusieurs façons.

Il peut s'agir de prendre soin de sa voix, au sens large, avant un effort, sans pour autant utiliser son larynx (*boire ; maintenir sa musculature de la nuque et du cou au chaud (foulard, linge éponge, écharpe) jusqu'à la prestation ; mise en route musculaire (tête + épaules) ; relaxation ; respiration ; boire un peu d'eau*), voire même justement en essayant de minimiser l'effort vocal (*utilisation "a minima" ; ne pas faire d'efforts brusques ; parler doucement ; le terme me semble approprié en ce sens qu'aucun effort ne doit porter sur les cordes vocales ; dans le cadre de la voix parlée il suffit de commencer la journée sans trop forcer ni trop monter la hauteur, le temps que les cordes vocales se chauffent ; ne JAMAIS demander au larynx un effort qu'il ne peut pas faire ; chercher les productions vocales nécessitant les mobilisations minimales*).

Mais la très grande majorité des définitions conçoivent l'échauffement vocal comme une série d'exercices vocaux. Pour autant, ces exercices sont associés à des buts différents, et certains de ces buts eux-mêmes sont sujet à polémique entre les différentes personnes interrogées. Les « vocalises » en sont un exemple typique : certaines personnes, en particulier les enseignants, pensent qu'il est bon de se chauffer la voix de cette manière, tandis que d'autres, plutôt phoniatres ou orthophonistes, préconisent une mise en route plus douce et sans effort du larynx, au risque sinon de se rendre aphone pour le reste de la journée.

- Pour commencer, l'échauffement vocal est très fréquemment conçu comme une phase de « rodage » du corps et du larynx, dans le but de se préparer à un effort important, par analogie à un échauffement sportif (*vocalises avant chorale ou chant solo ; préparation douce et progressive de l'appareil phonatoire à un exercice d'effort vocal ; tout comme l'athlète qui "chauffe" ses muscles avant l'épreuve, l'échauffement est une préparation progressive qui consiste à assouplir les muscles de la phonation + respiration en vue d'une situation de projection vocale ultérieure ; il s'agit comme en sport de préliminaires permettant de roder les cordes vocales (et les articulateurs) ; travailler la voix avant de chanter ; préparer l'action des cordes vocales par des exercices progressifs pour éviter un « claquage » ; faire fonctionner progressivement les muscles de la phonation (qui sont très divers) afin que dans l'effort il n'y ait pas de forçage*). Le champ lexical du muscle est d'ailleurs très présent dans les définitions (*maintenir sa musculature de la nuque et du cou au chaud ; assouplir les muscles de la phonation + respiration ; mise en route musculaire (tête + épaules) ; éviter un « claquage » ; l'éveil progressif de tous les muscles liés à la phonation*). Au contraire, une personne interrogée exprime le fait qu'un échauffement vocal n'est pas comparable à un échauffement sportif (*pour moi il n'y a pas de correspondance entre un échauffement sportif et un échauffement vocal car l'échauffement n'agit pas directement sur le muscle vocal*). Cette préparation n'est en tout cas pas uniquement physique mais consiste également à « se connecter avec l'activité que l'on va aborder » (*C'est une mise en condition mentale et corporelle ; [un] adressement de sa pensée dans le sens du travail vocal à venir ; concentration sur le corps ; ouverture sur le lieu*). Dans certains cas, aucune performance n'est envisagée derrière cet échauffement vocal et celui-ci est envisagé pour soi, comme une étape de transition non brutale vers un effort important (*c'est tenter d'emmener petit à petit sa voix vers des extrêmes à moindres coûts ; permettre au sujet d'amener progressivement le sujet vers un travail vocal de plus en plus important ; mise en tension très progressive des cordes vocales, associée à une pratique de la relaxation*). L'idée de progressivité de l'échauffement vocal est très présente dans toutes les réponses (*se mettre en voix progressivement, sans précipitation, préparation progressive, mise en tension très progressive des cordes vocales, préparer l'action des cordes vocales par des exercices progressifs, la recherche de l'intensité doit se faire progressivement, faire fonctionner progressivement les muscles de la phonation*), de même que les notions de douceur, souplesse ou

détente (*préparation douce, assouplir les muscles de la phonation, mise en vibration douce des cordes vocales ; associé à une pratique de la relaxation. Dissociation effort-tension ; ne pas faire d'efforts brusques, sans trop forcer*)

Ce rodage peut prendre la nuance de « réveil » du corps et de la voix. L'échauffement n'est alors plus tourné vers un effort à venir mais plutôt sur le fait de quitter un état d'engourdissement (*se mettre en voix ; au réveil ; voix du matin ; commencer la journée sans trop forcer ; pour un orateur, la première heure de la journée est déterminante, du réveil à la sortie de la maison. On peut se rendre aphone dans la première heure sans avoir crié. On peut a contrario se rendre résistant dans la même heure, et pouvoir hurler ensuite toute la journée ; Mise en route musculaire ; le larynx étant lié au système endocrinologique, il convient de « réveiller » le corps en même temps que le larynx ; éveiller le corps et mettre ses potentialités au service de la production vocale ; éveil progressif de tous les muscles liés à la phonation*). Pour certains, ce réveil peut s'accompagner d'une élimination des mucosités, point polémique puisque le hémage (ou raclement de gorge) est également cité de nombreuses fois par les professions rééducatrices de la voix comme un facteur d'irritation pouvant ensuite déclencher un malmenage vocal.

- Pour d'autres, l'échauffement vocal ne consiste pas en une transition vers l'effort mais simplement en une recupération des capacités de contrôle de la phonation (*retrouver quasiment instantanément la voix que l'on maîtrise et s'en servir jusqu'à atteindre assez vite (5 minutes) le maximum de ce que l'on sait faire ; retrouver des réflexes et vérifier le bon fonctionnement corporel et musculaire (les muscles eux mêmes sont quasiment immédiatement échauffés) ; retrouver la mobilité des muscles intrinsèques et extrinsèques du larynx, du voile du palais et de la bouche pour une voix de qualité sans effort ; reprendre les bases d'un bon fonctionnement vocal dans une utilisation "a minima" pour les appliquer ensuite en situation "extrême" sans risque de forçage vocal*), en particulier par la mise en relation de la respiration et du son (*améliorer la coordination pneumophonique ; la seule préparation consisterait, selon moi, à assurer une coordination entre diaphragme et lèvres via la respiration ; travail sur le souffle, la sensation diaphragmatique*) et par l'exploration de son étendue vocale (*glissades bien maîtrisées, vers l'aigu et vers le grave ; vocalises qui permettent de monter progressivement dans l'aigu*) .

- Au contraire, d'autres personnes conçoivent l'échauffement vocal comme l'équivalent d'un entraînement vocal. Il consiste alors à tonifier le larynx et à améliorer sa technique vocale (*travail sur la voix, le legato, la couleur, les voyelles ; travail sur le souffle, la sensation diaphragmatique ; travail sur l'articulation (prononciation), le piqué, les consonnes ; travail sur l'oreille, l'audition, la justesse ; faire des exercices vocaux ; vocalises et arpèges*).

Quelques personnes évoquent la durée de l'échauffement vocal. Pour certains, celui-ci n'existe presque pas (*rapide ; quasiment instantanément*) ou se réalise en un temps très court (*En quelques minutes ; Cinq à dix minutes ; 5 minutes*). Seule une personne parle de « la première heure de la journée ».

Enfin, certaines définitions de l'échauffement vocal suggèrent que celui-ci joue un rôle préventif vis-à-vis du forçage vocal (*c'est tenter d'emmener petit à petit sa voix vers des extrêmes à moindres coûts (= avec efficacité donc) ; reprendre les bases d'un bon fonctionnement vocal [...] pour les appliquer ensuite en situation "extrême" (chant, voix projetée) sans risque de forçage vocal ; dissociation effort-tension ; on peut [...] se rendre résistant [...], et pouvoir hurler ensuite toute la journée ; faire fonctionner progressivement les muscles de la phonation [...] afin que dans l'effort il n'y ait pas de forçage,...*).

4. ANALYSE DES QUESTIONS 2 ET 6 : RECENSEMENT DES SITUATIONS DE FORÇAGE VOCAL

Nous avons repris les réponses aux questions 2 et 6 des 22 personnes interrogées, et avons tenté de proposer une classification des différentes situations de forçage vocal.

4.1. Usage « a-normal » de la voix (84)

4.1.1. Situation d'effort vocal (74)

Utilisation importante de la voix (enseignement, chant) / dans les utilisations extrêmes de l'appareil vocal (chant, théâtre) / chorale / chant / chorale / chant

4.1.1.1. Usage prolongé (14)

Phonation prolongée (profs, instituteurs) / lors d'une prise de parole prolongée, lorsque la qualité du son s'altère et que le sujet cherche à compenser cette fatigue / situations d'effort vocal plus ou moins prolongé dans le temps / Prise de parole prolongée / Chanter trop longtemps / hurler d'une façon prolongée / parler non stop avec tension en donnant des cours ou des conférences, pendant des heures / cours prolongés, plus de 1h30 consécutive / prises de parole prolongée. Du type 6 heures de cours dans une même journée / conversations téléphoniques longues / enseignement pendant 6 heures de suite / après avoir chanté longtemps/ enseigner 9 heures de suite / en fin de concert, pour les bis

4.1.1.2. Usage à forte intensité (39)

voix criée / la voix criée / Le marchand de légumes de la rue de Lourmel / hurler / cri / voix projetée forte / trop lourd en puissance

- *Communication dans le bruit (25)*

Conversation en milieu bruyant / phonation en milieux bruyants / Enseignement en milieu bruyant (sport) / en milieu bruyant / Ouvrier en usine, devant donner des consignes devant une chaîne de montage bruyante / Dans le bruit par exemple où il n'arrive pas faire émerger sa voix qualitativement ou quantitativement du milieu (assemblée, rue, restaurant) / en environnement bruyant (enseignement, théâtre de rue, manifestations, débats politiques et syndicaux / parler dans une boîte de nuit / Dans un environnement trop bruyant / en environnement bruyant / Conversation dans un bar (musique trop forte) / Enseignement quand les élèves chahotent / Par exemple dans une salle de classe avec des élèves bruyants. Le professeur va forcer s'il essaye de passer au dessus du bruit de fond / en discothèque, ou dans une salle de concert / dans un transport en commun telle que métro, bus / lors d'une discussion croisée dans un repas à table avec beaucoup de personnes / devoir passer un orchestre trop bruyant / Parler dans une boîte de nuit avec la sono à fond / bruit ambiant / lieu bruyant / milieu bruyant / lorsque j'essaye de me faire entendre dans une discothèque malgré le bruit ambiant/ dans les transports en commun/ Pour mieux s'entendre dans un environnement trop bruyant (musical ou autre) / face à un orchestre qui n'écoute pas le chanteur

- *Parole émotionnelle de forte intensité (5)*

Le supporter de sport qui crie dans les tribunes pour encourager / les groupies dans un concert / Gronder son enfant quand il ne veut pas aller se coucher / énervement extrême/ voix de colère

- *Situation d'urgence (1)*

voix "de détresse" dans des situations d'urgence

- *Distance (1)*

Interlocuteurs distants

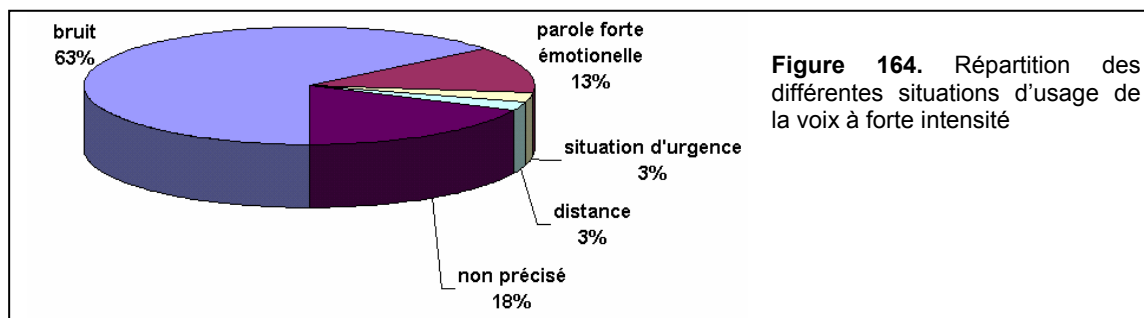


Figure 164. Répartition des différentes situations d'usage de la voix à forte intensité

4.1.1.3. Pitch inhabituel (9)

Chant dans un registre peu confortable pour le sujet (trop haut ou trop bas, ou tenu trop longue)/ ceci arrive aux chanteurs amateurs / Voix chantée plus aiguë que sa voix normale ou plus grave que sa voix normale / chanteur inexpérimenté s'attaquant au registre aigu / recherche de hauteur de note non maîtrisée (chanteurs) / Chanter hors tessiture (trop grave ou trop aigu) / répertoire inapproprié / voix un peu trop grave / trop aigu en tessiture / Quand je chante un répertoire qui est trop « lourd » pour moi

4.1.1.4. Acoustique inadéquate (6)

faire un cours sans micro / jouer une pièce de théâtre en plein air / retour du son produit trop faible / grande salle / plein air / Pour mieux s'entendre dans un environnement sans assez de réverbération

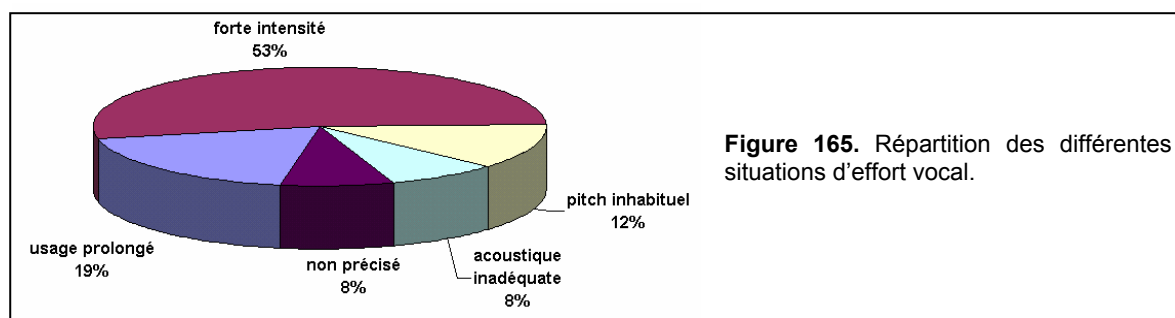


Figure 165. Répartition des différentes situations d'effort vocal.

4.1.2. Posture inconfortable (1)

Prise de parole dans des postures difficiles, p.ex. au téléphone (flexion asymétrique de la tête), autre exemple : les enseignants qui parlent aux élèves en tordant le cou alors qu'ils font face au tableau.= mauvaise posture, peu propice à la voix portée et qui engendre des tensions.

4.1.3. Situation de trouble émotionnel (9)

Situations émotionnelles / situations « émotionnelles » / sous l'effet d'une forte émotion / Etat de trac, de crainte ou d'angoisse / dans des situations de stress /conversations contrariantes / situation à composante émotionnelle / en public / dans des situations émotionnellement fortes, quand la voix a tendance à s'en aller et qu'il faut absolument la "rattraper".

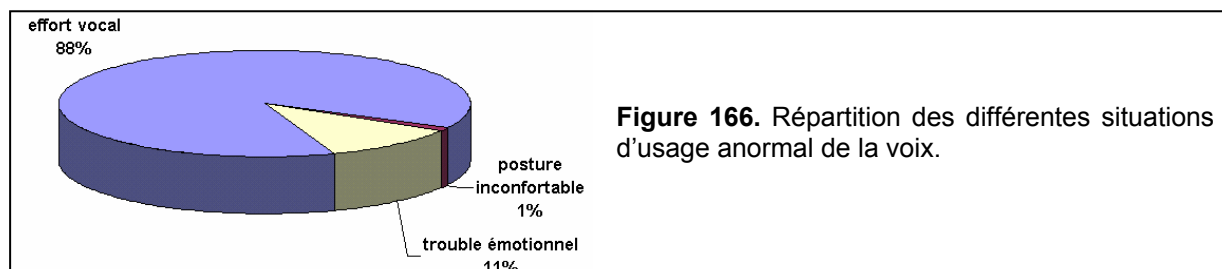


Figure 166. Répartition des différentes situations d'usage anormal de la voix.

4.2. Incapacité dans la situation même (13)

4.2.1. Impuissance psychologique (2)

Toute situation où l'on se sent dépassé ou impuissant du fait de l'insuffisance réelle ou supposée de sa propre voix au sens large du terme / La situation de forçage est un état dans lequel se trouve un sujet en difficulté avec sa voix.

4.2.2. Technique (12)

Prise de parole sur air 'résidu' (manque de soutien respiratoire à la parole) / mauvaise place de voix dans le chant (lyrique, rock...) ou pour un acteur de théâtre / déficiences technique / les gens qui utilisent mal leurs cordes vocales pour chanter / parler en appui sur le larynx (sans « soutenir ») / Chanter du répertoire exigeant pour le contrôle du vibrato (XVIIe par exemple) et opérer ce contrôle par un appui laryngé plutôt que par un contrôle diaphragmatique / chanter en forçant de la gorge / chanter en prétendant du volume sans emploi du souffle / recherche d'intensité sur un larynx insuffisamment tonique / en m'attaquant à des airs du répertoire lyrique alors que je ne suis pas assez entraîné / trop grande difficulté non surmontée (encore ou définitivement) par rapport à ses moyens vocaux / dans une situation musicale extrême où il faut réaliser quelque chose qui dépasse nos moyens ou notre technique

4.2.3. Impuissance ou échec de la communication (3)

Pour faire rentrer quelque chose dans la tête de son partenaire qui, en dépit du bon sens, refuse de comprendre/ Exemple : « Je n'arrive pas à me faire entendre » ou « Je n'arrive pas à me faire comprendre » / enjeu important, envie de convaincre à tout prix, le plus rapidement possible (voix tendue, trop aiguë, « à bout de souffle »... Qui ne convainc personne !)

4.3. Facteurs (10)

4.3.1. Troubles physiologiques préexistants (9)

Présence de lésions organiques / présence d'une anomalie organique(nodule polype) / Gros fumeurs et personnes grippées qui s'efforcent à parler, alors qu'ils souffrent d'une inflammation aiguë ou chronique des cordes vocales / fatigue occasionnelle / parler ou chanter ayant une laryngite ou autre maladie ORL / compensation chez des patients avec pathologies neurodégénérative / en cas de pathologie / après avoir trop chanté dans un état de fatigue générale / alors que je suis malade (rhume, laryngite, etc).

4.3.2. Troubles psychologiques (2)

Troubles psychologiques / Aphonie psychogène

4.3.3. Personnalités (1)

Enfant exubérant jouant en groupe ou prenant la parole à table.

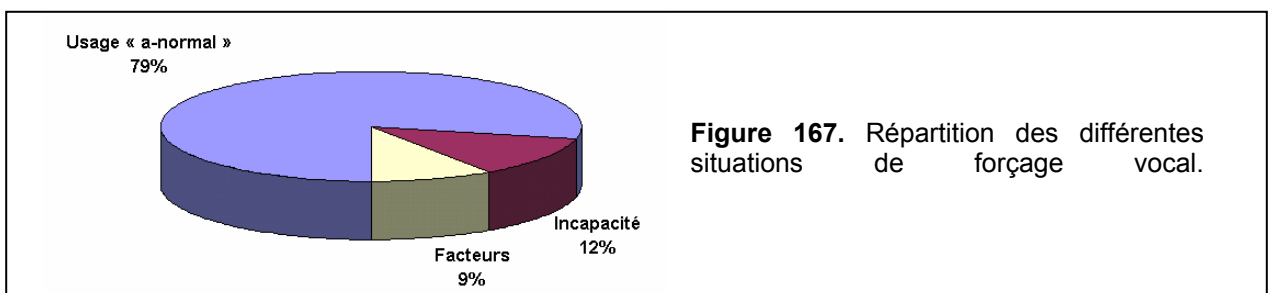


Figure 167. Répartition des différentes situations de forçage vocal.

Annexe Res2 :

**Questionnaire sur la
perception des
environnements bruyants**

Error! No table of contents entries found.

1. QUESTIONS POSEES

Pour commencer, je vais vous poser quelques questions relatives à des situations de communication.

1-Dans quelles situations vous est-il arrivé de forcer sur votre voix pour être entendu ou compris ?

2- Dans quelles situations vous est-il arrivé d'avoir du mal à entendre ou comprendre ce que dit une personne ?

Les prochaines questions vont porter sur la façon dont vous percevez les différents moments de la journée

3-Quel moment de la journée vous paraît le plus calme? En quoi?

4- Quel moment de la journée vous paraît le plus silencieux? En quoi ?

5- Quel moment de la journée vous paraît le plus animé ? En quoi ?

6- Quel moment de la journée vous paraît le plus bruyant ? En quoi ?

J'aimerais maintenant que vous m'expliquiez ce que représente pour vous la notion d'"environnement bruyant"

7-Qu'est ce qu'un environnement bruyant pour vous? Pourriez-vous en donner une définition ?

8-En quoi un environnement vous paraît-il bruyant ? Quelles caractéristiques vous le font percevoir comme tel?

9-Pouvez vous citer quelques exemples d'environnements qui vous paraissent typiquement bruyants ?

Pour finir, je vais vous demander de vous "replonger" mentalement dans certaines situations de votre vie quotidienne

10-Durant les activités particulières évoquées ci-dessous, vous souvenez-vous d'avoir été confronté à des situations bruyantes?

Si oui, pouvez vous préciser LESQUELLES et surtout EN QUOI elles vous paraissaient bruyantes. Si rien ne vous a paru spécialement bruyant dans cette situation, n'hésitez pas à écrire "NON"

- Vous êtes en train de travailler

- Vous marchez dans la rue, vous vous déplacez en métro, bus ou tramway, en voiture, en avion...etc

- Vous êtes chez vous, seul ou avec d'autres personnes.

- Vous êtes en vacances, en ville ou à la campagne

- Vous faites vos courses

- Vous effectuez une démarche administrative

- Vous sortez au cinéma, au théâtre, à un concert, dans un bar, une discothèque...etc

- Vous allez faire du sport ou votre loisir hebdomadaire favori

2. SYNTHÈSE DES RÉPONSES OBTENUES

31 personnes ont répondu à ce questionnaire. 22 hommes et 9 femmes, de 22 à 57 ans, moyenne d'âge 29 ans. 4 vivent à la campagne et 27 sont citadins.

Parmi les situations de forçage vocal, 58 % concernent une communication en environnement bruyant. (et 18% concernent une communication dans des grands espaces, 9.7% de la voix colérique, 42% de la communication téléphonique). Spontanément, presque la totalité des exemples d'environnements bruyants cités, induisant une sensation de forçage, sont des lieux de vie sociale (bar, discothèque, restaurant, foule)

Les définitions données d'un environnement bruyant font émerger certaines caractéristiques d'un environnement bruyant pertinentes vis à vis de la gêne qu'elle peuvent induire sur la communication : la constance ou continuité du bruit, la perturbation de la concentration ou de l'attention du locuteur, le niveau sonore, la difficulté de localisation du bruit, la distorsion du bruit ou encore le caractère non informatif du bruit.

D'autres caractéristiques sont citées comme le sentiment de soumission face à ce bruit, la perception de mouvement ou d'agitation dans l'espace, mais elles peuvent difficilement être prise en compte dans la mise au point d'un protocole expérimental.

Une autre particularité d'un environnement bruyant concerne la distinction des sources : 54.5% des définitions font référence à des bruits distincts, 45.5% à un ensemble de bruits non différenciés. Lorsque les bruits sont distincts, il s'agit à 72 % de sources multiples contre 28 % de sources isolées. Lorsque le bruit est considéré comme un ensemble non différencié, il s'agit à 47% d'un bruit de fond/brouhaha et 53% d'un bruit/son massif ou non mais toujours indéfini donc non caractérisé.

Lorsqu'on interroge des personnes sur les caractéristiques d'un environnement qui le leur font percevoir comme bruyant, on retrouve certaines des caractéristiques déjà citées (Niveau sonore, difficulté de localisation de la (des) source(s), la continuité ou le caractère répétitif du bruit). S'y ajoutent la non intelligibilité du bruit (proche du caractère « non informatif »), le caractère strident d'un son de l'environnement ou la nature motorisée des sources bruyantes.

On retrouve à occurrence relativement égale des conceptions d'environnements bruyants où se trouvent des sources multiples et distinctes, et des conceptions du bruit comme un magma sonore indéfini, non localisé, non compréhensible.

Les caractéristiques d'un environnement perçu comme bruyant sont à majorité décrites par des termes subjectifs, sans apporter d'information supplémentaire sur les caractéristiques du bruit mais plus sur les effets qu'il provoque du point de vue de l'auditeur (désagrément, gêne, agressivité, stress ou fatigue qu'il engendre). Il est néanmoins intéressant de noter que la gêne est principalement liée à des difficultés de communication (écoute et parole) ainsi qu'à un détournement de l'attention. De plus, la nuisance va pour 14.5% des réponses jusqu'au sentiment d'agression (douleur ou malaise physique, bruit lié à un danger, à la pollution par des gaz d'échappement ou dénotant d'un manque de respect de la part d'une personne). Quant au désagrément(25% des réponses) , il est de nouveau fréquemment associé au sentiment de soumission face au bruit (imposition d'un bruit non désiré, absence de contrôle sur le son, surprise du son dans le contexte d'écoute, obligation d'adopter un comportement non choisi mais contraint face à ce bruit) (14.5 % de l'ensemble des réponses)

Lorsqu'on demande à des personnes de citer spontanément des exemples d'environnements bruyants, 9.8% concernent des situations vécues sur le lieu de travail, 44.6% des bruits liés aux transports, 30.4% des environnements publics avec présence humaine (bruits de conversation et/ou musique d'ambiance),7.6% des bruits de travaux dans des lieux publics, 1% des signaux d'avertissement, et 4.3% seulement des bruits liés à l'habitat (chez soi ou bien chez le voisin).

Lorsqu'on essaye de former des groupes de ces exemples, il se dégage 3 principaux types de bruits : -23.9% concernent des bruits liés à des conversations ou à une présence humaine (14.1% cafés/restaurants, 3.3% magasins, 2.2% cour de récréation, 3.3% voisin)

-59.8% sont des bruits mécaniques ou de machines (dont 44.6% de bruits de transports, 8.7% de travaux publics, 3.3% de bruits d'usine, 2.2% de bruits de ventilation et 1% de signaux d'avertissement)

-13% citent explicitement la musique, recherchée ou subie, comme caractéristique bruyante d'un environnement.

Ces chiffres ne sont donnés qu'à titre indicatif car si certaines réponses ne citent pas explicitement la musique, notre propre expérience pourrait l'inférer, par exemple « dans les cafés », « dans les bars ». En fait, la plupart du temps, musique et bruits de conversation sont présents dans les mêmes environnements mais un locuteur peut prêter plus attention à la gêne engendrée par l'une des ces sources.

Le reste des réponses concernent des bruits issus de la nature (2.2%) et des déformations du son dues à l'acoustique des salles (1%).

Un environnement bruyant est avant tout un lieu pour la majorité des définitions recueillies (environ 75 %). 27% environ des exemples donnés citent une source de bruit identifiée de cet environnement bruyant (des véhicules, le voisin, un marteau piqueur, le métro, le système de ventilation, la télé, des chanteurs). Parmi ces sources, un peu moins des 2/3 sont directement considérées comme la définition même de l'environnement bruyant, les autres sont associées (« Avec ») à un lieu. Seules 1/5 des sources effectuent une action (déplacement, émission vocale...etc) et celle ci est très rarement située dans le temps (1 seule occurrence). Les 13% de définitions qui rendent compte d'une perception de l'environnement bruyant comme un événement temporel renvoient très peu à une action de la source sonore mais en très grande majorité à une activité de l'auditeur au moment où il perçoit l'environnement comme bruyant (voyage, dîner, ...etc) où à l'activité souhaitée par l'auditeur mais empêchée par le bruit (dormir, téléphoner...etc). Enfin, environ 10% des définitions assimilent un environnement bruyant directement à un phénomène acoustique (un fond sonore, de la musique, un effet de traitement ou de sonorisation du son...etc).

On remarque que les acteurs sont très peu cités, sauf dans la sphère de l'habitation (le voisin/la voisine, le colocataire). On a des « tout le monde » dans des environnements comme les bars et « des chanteurs » qui crient, « des oiseaux » qui gazouillent. Sinon, ce façon générale, les bruits semblent être produits « tout seuls » (la nature, la rue) ou par des objets (les marteaux piqueurs, les voitures, la rame de métro). Cela est d'autant plus étonnant lorsque les humains sont directement la source du bruit (bars, restaurants). On observe également que l'auditeur qui subit le bruit est également complètement absent (seulement 1 on, 2 je, 1 me, 1 tu) de son discours.

Le bruit sur le lieu de travail est plus souvent mis en relation avec la gêne qu'il occasionne sur la concentration que sur la difficulté de communication. Cependant, la plupart des personnes interrogées effectuant des emplois de bureau (biais introduit par le mode de réponse au questionnaire sur Internet), cela peut expliquer que peu de ces personnes soient soumises à des niveaux sonores intenses sur leur lieu de travail. On note cependant deux situations intéressantes : Celle de réunions de travail où tout le monde parle simultanément et qu'une personne tente de prendre la parole et de se faire entendre, ainsi que les bruits de ventilation ou de pompes (labos de chimie).

Le bruit au domicile ne concerne pratiquement pas la gêne vocale mais plutôt le besoin de repos. Les bruits cités sont donc de très faible niveau. La nuisance est plus relative à la personne qui produit le bruit qu'au bruit lui-même. On peut tout de même noter les travaux de bricolage (perceuse, tondeuse...etc), les soirées organisées à domicile (similaire à une discothèque ou un restaurant) ou encore les cris d'enfants (qui jouent ou pleurent).

Les bruits dans les transports et dans la rue sont plus pertinents pour notre étude sur la communication en environnement bruyant :

- 61 % des bruits cités proviennent des transports eux mêmes. 11.8% concernent en particulier le métro (grincement, roulement, « explosion » des compresseurs), 17,6% la circulation de véhicules (plutôt liés au bruit de moteur ou à l'absence de pot d'échappement qu'au bruit de roulement), 9.8% le bruit d'avions(toujours reliés à leur moteur), 5.9% des bruits de freinage (surtout de trains), et 15.7% les bruits de klaxon (relié à la notion d'embouteillage => nervosité des conducteurs, stationnarité des véhicules, pollution).
- 25.5% des exemples cités sont d'origines mécaniques ou électrique, non produits directement par les moyens de transports eux-mêmes (Travaux publics, sonneries de portables ou de métro, musique subie dans le métro assimilée à du bruit)
- Enfin 9.8% se rapportent à des bruits de conversations ou à des cris (d'enfants).

Le bruit sur les lieux de loisirs semble quasi inexistant, à part dans les piscines où la résonance amplifie les cris des baigneurs.

En ce qui concerne les lieux d'achat, les supermarchés sont plus souvent cités que les marchés, et pas exactement pour les mêmes raisons : Les marchés sont perçus comme bruyants à cause des

vendeurs qui crient. Mais ce qui dérange les acheteurs est plutôt la pression commerciale et la proximité du vendeur que sa voix. Par contre, les supermarchés sont considérés comme bruyants à cause des bandes son et des animations marketing qui créent un brouhaha ambiant. On remarque que la simple présence de « gens », l'animation du magasin ou la fatigue visuelle devant la multitude des produits, sont considérées comme des caractéristiques bruyantes d'un environnement. L'acoustique des grandes surfaces et des marchés couverts est citée à plusieurs reprises comme source/contribution au caractère bruyant de ce type d'environnement.

Enfin, les interrogés notent eux mêmes le paradoxe du bruit dans des lieux de divertissement où le bruit leur paraît intense et parfois gênant, tout en étant pourtant recherché. Pour la musique forte, le seuil de gêne semble être dans ce cas la douleur, le malaise physique ou bien la difficulté de communiquer avec une autre personne, et non pas seulement le niveau sonore. C'est le type d'environnement où l'effet du choix/de la soumission au bruit est la plus flagrante. Ainsi, dans d'autres types de lieux de divertissement où le niveau sonore n'est pas élevé, une très grande proportion des réponses font référence aux bruits intempestifs du public par rapport aux attentes du spectateur. (bonbons ou chuchotements au cinéma...etc).

3. CONCLUSION

Ce questionnaire a été mené afin de recenser les environnements bruyants auxquels les personnes sont confrontées dans leur vie quotidienne est dans lesquelles elles sont susceptibles de communiquer, voire de forcer sur leur voix pour y parvenir. Ce recensement nous a permis de dégager quelles pouvaient être les situations bruyantes les plus pertinentes à étudier dans le cadre de la thèse :

- les lieux tels que des bars-discothèques où le bruit ambiant est en grande partie basses-fréquences, de fort niveau, composé de musique et de conversations. Il s'agit d'un bruit continu pouvant aller jusqu'à induire un malaise physique pour l'auditeur.
- les lieux publics comme un restaurant dans lequel le bruit ambiant est essentiellement composé de conversations inintelligibles et non localisées, avec un contenu spectral coïncident avec celui de la parole.
- les situations au contraire où une personne parle de façon intelligible et à forte intensité à côté de soi.
- les lieux où l'acoustique est fortement réverbérante et déforme les indices spectro-temporels des sons.
- les lieux où une machine ou une soufflerie est en marche, et où le bruit ambiant est assez comparable à un bruit blanc.
- les transports en commun où les abords d'une voie de circulation où le bruit ambiant consiste essentiellement en bruits de moteurs et de roulement.
- les situations de travaux où le bruit est plutôt répétitif, d'origine mécanique.
- les situations dans lesquelles une ou plusieurs sources produisent des sons de fréquence aiguë : par exemple, un bébé qui hurle, une perceuse, une sirène, un freinage de train, un standard téléphonique où plusieurs téléphones sonnent simultanément, ou encore un concert de klaxons à un carrefour.

Dans un premier temps, nous avons choisi parmi ces situations de nous concentrer sur les bruits de conversations inintelligibles et les bruits continus de type soufflerie s'apparentant à du bruit blanc. Il sera ensuite intéressant par la suite d'examiner les autres types de situations.

Annexe Res3 :
Résultats statistiques
détaillés du chapitre 3

EFFET DU PORT D'UN CASQUE PAR COMPARAISON A UNE RESTITUTION DU BRUIT PAR HAUT-PARLEURS

Durée syllabique

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	2ms F(1) = 0.1 NS 0.74	
	62dB de bruit	13ms F(1) = 5.0 *	7ms F(1) = 1.4 NS (p=0.24)
	70dB de bruit	4ms F(1) = 0.6 NS 0.45	7ms F(1) = 1.6 NS (p=0.21)
	78dB de bruit	-4ms F(1) = 0.6 NS (p=0.46)	8ms F(1) = 2.4 NS (p=0.12)
	86dB de bruit	-13ms F(1) = 5.0 *	21ms F(1) = 13.5 ***
Effet moyen, silence et tous niveaux de bruit confondus		D=0ms F(1) = 0.01 NS (p=0.93)	D=9ms F(1) = 12.9 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit		13ms F(4) = 5.8 ***	9ms F(4) = 2.8 *
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant			a=0.34 r=8.0%

Identique	Légère amplification
-----------	----------------------

Intensité vocale

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	6.2 dB F(1) = 127.7 ***	
	62dB de bruit	4.9 dB F(1) = 114.2 ***	9.1 dB F(1) = 406.5 ***
	70dB de bruit	4.7 dB F(1) = 124.9 ***	8.8 dB F(1) = 557.8 ***
	78dB de bruit	4.4 dB F(1) = 104.4 ***	7.0 dB F(1) = 298.9 ***
	86dB de bruit	3.5 dB F(1) = 64.9 ***	7.8 dB F(1) = 294.7 ***
Effet moyen, silence et tous niveaux de bruit confondus		D=4.7 dB F(1) = 257.8 ***	D=7.8 dB F(1) = 740.6 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit		1.0 dB F(4) = 7.0 ***	1.4 dB F(4) = 12.8 ***
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=-0.05 r=-18.1%	

Légère réduction + décalage	Décalage
-----------------------------	----------

Dynamique d'intensité vocale

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	1.7 dB F(1) = 1.8 NS (p=0.20)	
	62dB de bruit	-1.1 dB F(1) = 0.4 NS (p=0.52)	-2.5 dB F(1) = 5.87 *
	70dB de bruit	-0.3 dB F(1) = 0.1 NS (p=0.79)	-1.3 dB F(1) = 0.6 NS (p=0.47)
	78dB de bruit	-1.8 dB F(1) = 4.6 *	1.0 dB F(1) = 0.6 NS (p=0.46)
	86dB de bruit	-0.1 dB F(1) = 0.004 NS (p=0.95)	1.7 dB F(1) = 2.2 NS (p=0.15)
Effet moyen, silence et tous niveaux de bruit confondus		D=-0.3 dB F(1) = 0.3 NS (p=0.59)	D=0.1 dB F(1) = 0.04 NS (p=0.85)
Variabilité moyenne de cet effet en fonction du niveau de bruit		1.6 dB F(4) = 1.0 NS (p=0.42)	2.3 dB F(4) = 2.3 NS (p=0.07)

Identique	Identique
-----------	-----------

Fréquence fondamentale

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	0.6 tons F(1) = 5.4275 *	
	62dB de bruit	0.3 tons F(1) = 1.6 NS (p=0.20)	1.7 tons F(1) = 40.6 ***
	70dB de bruit	0.5 tons F(1) = 3.5 NS (p=0.06)	1.5 tons F(1) = 32.2 ***
	78dB de bruit	-0.1 tons F(1) = 0.4 NS (p=0.54)	1.1 tons F(1) = 20.8 ***
	86dB de bruit	-0.4 tons F(1) = 2.9 NS (p=0.09)	1.4 tons F(1) = 26.2 ***
Effet moyen, silence et tous niveaux de bruit confondus		D=0.2 tons F(1) = 1.5 NS (p=0.22)	D=1.2 tons F(1) = 97.9869 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit		0.5 tons F(4) = 15.2 ***	0.5 tons F(4) = 16.5 ***
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant			a=0.01 r=11.8%

Identique	Légère amplification + décalage
-----------	---------------------------------

Ambitus de fréquence fondamentale

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPS	silence	0.4 tons F(1) = 0.1 NS (p=0.80)	
	62dB de bruit	0.6 tons F(1) = 0.18 NS (p=0.68)	-0.3 tons F(1) = 0.1 NS (p=0.81)
	70dB de bruit	tons F(1) = 1.0 NS (p=0.34)	-0.3 tons F(1) = 0.05 NS (p=0.82)
	78dB de bruit	-0.1 tons F(1) = 0.004 NS (p=0.95)	0.2 tons F(1) = 0.04 NS (p=0.85)
	86dB de bruit	0.8 tons F(1) = 1.3 NS (p=0.26)	0.4 tons F(1) = 0.3 NS (p=0.57)

Effet moyen, silence et tous niveaux de bruit confondus	D=0.6 tons F(1) = 1.1 NS (p=0.29)	D=0.1 tons F(1) = 0.03 NS (p=0.86)
Variabilité moyenne de cet effet en fonction du niveau de bruit	0.6 tons F(4) = 0.6 NS (p=0.69)	0.4 tons F(4) = 0.4 NS (p=0.77)

Identique	Identique
-----------	-----------

Différence d'énergie acoustique entre [2-4kHz] et [0-1kHz]

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPS	silence	1.5 dB F(1) = 7.5 **	
	62dB de bruit	-1.2 dB F(1) = 4.9 *	3.6 dB F(1) = 48.6 ***
	70dB de bruit	0.3 dB F(1) = 0.3 NS (p=0.60)	2.7 dB F(1) = 26.7 ***
	78dB de bruit	1.1 dB F(1) = 3.9 *	1.3 dB F(1) = 7.1 **
	86dB de bruit	2.1 dB F(1) = 11.9 ***	2.0 dB F(1) = 14.3 ***

Effet moyen, silence et tous niveaux de bruit confondus	D=0.7 dB F(1) = 6.3 *	D=2.2 dB F(1) = 62.4 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit	1.7 dB F(4) = 16.0 ***	1.1 dB F(4) = 9.1 ***

Très léger décalage	Décalage
---------------------	----------

Différence d'énergie acoustique entre [1-2kHz] et [0-1kHz]

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPS	silence	1.5 dB F(1) = 7.0 **	
	62dB de bruit	-0.9 dB F(1) = 2.6 NS (p=0.11)	2.5 dB F(1) = 23.3 ***
	70dB de bruit	0.1 dB F(1) = 0.02 NS 0.89	2.6 dB F(1) = 21.4 ***
	78dB de bruit	0.5 dB F(1) = 0.9 NS (p=0.35)	2.1 dB F(1) = 15.1 ***
	86dB de bruit	0.1 dB F(1) = 0.03 NS (p=0.87)	2.2 dB F(1) = 18.5 ***

Effet moyen, silence et tous niveaux de bruit confondus	D=0.2 dB F(1) = 0.7 NS (p=0.42)	D=2.2 dB F(1) = 62.8 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit	1.1 dB F(4) = 9.4 ***	0.4 dB F(4) = 1.9 NS (p=0.11)

Identique	Décalage
-----------	----------

Différence d'énergie acoustique entre [4-6kHz] et [0-1kHz]

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPS	silence	0.2 dB F(1) = 0.1 NS (p=0.72)	
	62dB de bruit	-0.3 dB F(1) = 0.3 NS (p=0.59)	4.5 dB F(1) = 71.8 ***
	70dB de bruit	1.3 dB F(1) = 5.0 *	3.4 dB F(1) = 34.7 ***
	78dB de bruit	2.3 dB F(1) = 16.9 ***	1.3 dB F(1) = 6.4 *
	86dB de bruit	3.3 dB F(1) = 28.0 ***	2.3 dB F(1) = 15.6 ***

Effet moyen, silence et tous niveaux de bruit confondus	D=1.4 dB F(1) = 22.4 ***	D=2.4 dB F(1) = 68.5 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit	2.2 dB F(4) = 20.3 ***	2.1 dB F(4) = 20.8 ***

Facteur d'amplification de cet effet en fonction du niveau de bruit croissant	a=0.07 r=20.4%	a=0.03 r=7.6%
---	-------------------	------------------

Amplification	Amplification
---------------	---------------

Quotient ouvert laryngé

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	-0.00 F(1) = 0.49 NS (p=0.48)	
	62dB de bruit	-0.02 F(1) = 13.5 ***	-0.03 F(1) = 14.2 ***
	70dB de bruit	-0.00 F(1) = 0.4 NS (p=0.51)	-0.01 F(1) = 4.3 *
	78dB de bruit	-0.01 F(1) = 1.8 NS (p=0.18)	-0.02 F(1) = 7.1 **
	86dB de bruit	-0 F(1) = 0.04 NS (p=0.85)	-0.01 F(1) = 1.4 NS (p=0.23)

Effet moyen, silence et tous niveaux de bruit confondus	D=-0.01 F(1) = 8.2 **	D=-0.02 F(1) = 22.2 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit	0.01 F(4) = 5.9 ***	0.01 F(4) = 4.4 **

Léger décalage	Léger décalage
----------------	----------------

H1-H2

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	-1.5 dB F(1) = 11.8 ***	
	62dB de bruit	-2.4 dB F(1) = 11.9 ***	-3.8 dB F(1) = 45.6 ***
	70dB de bruit	-1.0 dB F(1) = 3.3 NS (p=0.07)	-3.9 dB F(1) = 32.6 ***
	78dB de bruit	0.0 dB F(1) = 0.003 NS (p=0.99)	-3.3 dB F(1) = 22.1 ***
	86dB de bruit	1.4 dB F(1) = 2.2 NS (p=0.14)	-6.7 dB F(1) = 64.1 ***

Effet moyen, silence et tous niveaux de bruit confondus	D=-0.7 dB F(1) = 4.2 *	D=-3.8 dB F(1) = 147.6 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit	1.3 dB F(4) = 3.6 **	2.3 dB F(4) = 12.2 ***
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=-0.09 r=-15.5

Très léger décalage	réduction
---------------------	-----------

Vitesse de fermeture laryngée

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	0.8 u.a. F(1) = 97.3 ***	
	62dB de bruit	1.4 u.a. F(1) = 111.6 ***	2.11 u.a. F(1) = 206.6 ***
	70dB de bruit	1.5 u.a. F(1) = 91.0 ***	2.31 u.a. F(1) = 276.2 ***
	78dB de bruit	2.4 u.a. F(1) = 119.3 ***	2.56 u.a. F(1) = 173.0 ***
	86dB de bruit	2.6 u.a. F(1) = 129.8 ***	3.68 u.a. F(1) = 151.8 ***

Effet moyen, silence et tous niveaux de bruit confondus	D=1.8 u.a. F(1) = 397.1 ***	D=2.31 u.a. F(1) = 582.2 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit	1.0 u.a. F(4) = 26.7 ***	1.20 u.a. F(4) = 35.9 ***

Amplification + Décalage	Amplification + Décalage
--------------------------	--------------------------

Premier formant vocalique

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	-1Hz F(1) = 0.003 NS (p=0.95)	
	62dB de bruit	-5 Hz F(1) = 0.1 NS (p=0.81)	-9 Hz F(1) = 0.2 NS 0.65147
	70dB de bruit	-2 Hz F(1) = 0.01 NS (p=0.92)	12 Hz F(1) = 0.4 NS (p=0.52)
	78dB de bruit	-2 Hz F(1) = 0.01 NS (p=0.92)	12 Hz F(1) = 0.39854 NS (p=0.53)
	86dB de bruit	-8 Hz F(1) = 0.2 NS (p=0.67)	18 Hz F(1) = 0.9 NS (p=0.34)

Effet moyen, silence et tous niveaux de bruit confondus	D=4 Hz F(1) = 0.2 NS (p=0.68)	D=7 Hz F(1) = 0.6 NS (p=0.45)
Variabilité moyenne de cet effet en fonction du niveau de bruit	4 Hz F(4) = 0.2 NS (p=0.92)	9 Hz F(4) = 1.0 NS (p=0.43)

Identique	Identique
-----------	-----------

Question 1 : perception de sa propre voix

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	0.3 F(1) = 3.8 NS (p=0.07)	
	62dB de bruit	0.4 F(1) = 1.8 NS (p=0.20)	0.4 F(1) = 2.3 NS (p=0.14)
	70dB de bruit	0.7 F(1) = 3.6 NS (p=0.07)	-0.3 F(1) = 0.4 NS (p=0.52)
	78dB de bruit	1.1 F(1) = 12.8 **	-0.4 F(1) = 1.8 NS (p=0.20)
	86dB de bruit	0.4 F(1) = 0.5 NS (p=0.49)	0.3 F(1) = 0.5 NS (p=0.50)

Effet moyen, silence et tous niveaux de bruit confondus	D=0.6 F(1) = 8.4 **	D=0.0 F(1) = 0 NS (p=1)
Variabilité moyenne de cet effet en fonction du niveau de bruit	0.4 F(4) = 1.1 NS (p=0.37)	0.4 F(4) = 2.7 *
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=0.01 r=14.4%

Identique	Légère amplification
-----------	----------------------

Question 3 : perception de l'interlocuteur

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	0.0 F(1) = 0 NS (p=1)	
	62dB de bruit	0.4 F(1) = 1.8 NS (p=0.20)	0.2 F(1) = 0.4 NS (p=0.53)
	70dB de bruit	-0.4 F(1) = 0.8 NS (p=0.40)	0.9 F(1) = 5.6 *
	78dB de bruit	-0.6 F(1) = 2.9 NS (p=0.11)	1.1 F(1) = 6.9 *
	86dB de bruit	-0.3 F(1) = 0.4 NS (p=0.54)	0.7 F(1) = 2.0 NS (p=0.17)

Effet moyen, silence et tous niveaux de bruit confondus	D=-0.2 F(1) = 0.7 NS (p=0.40)	D=0.6 F(1) = 6.5 *
Variabilité moyenne de cet effet en fonction du niveau de bruit	0.5 F(4) = 1.9 NS (p=0.12)	0.6 F(4) = 2.4 NS (p=0.07)
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=0.02 r=33.6%

Identique	Légère amplification
-----------	----------------------

Question 2 : estimation de sa propre intelligibilité

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	0.0 F(1) = 0 NS (p=1)	
	62dB de bruit	0.4 F(1) = 2.3 NS (p=0.14)	0.1 F(1) = 0.1 NS (p=0.72)
	70dB de bruit	-0.3 F(1) = 0.4 NS (p=0.52)	0.7 F(1) = 4.4 NS (p=0.05)
	78dB de bruit	-0.4 F(1) = 1.8 NS (p=0.20)	0.9 F(1) = 5.8 *
	86dB de bruit	0.3 F(1) = 0.5 NS (p=0.50)	0.7 F(1) = 2.0 NS (p=0.17)

Effet moyen, silence et tous niveaux de bruit confondus	D=0.0 F(1) = 0 NS (p=1)	D=0.5 F(1) = 5.1 *
Variabilité moyenne de cet effet en fonction du niveau de bruit	0.4 F(4) = 2.7 *	0.5 F(4) = 2.0 NS (p=0.11)
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=0.02 r=33.4%

Identique	Légère amplification
-----------	----------------------

Question 4 : Ressenti de gêne

		Bruit blanc	Bruit cocktail
Effet du port d'un casque par rapport à une restitution du bruit par HPs	silence	0.4 F(1) = 2.7 NS (p=0.12)	
	62dB de bruit	0.0 F(1) = 0 NS (p=1)	0.5 F(1) = 1.6 NS (p=0.22)
	70dB de bruit	0.1 F(1) = 0.04 NS (p=0.84)	1.0 F(1) = 3.8 NS (p=0.07)
	78dB de bruit	-0.3 F(1) = 0.4 NS (p=0.55)	1.6 F(1) = 12.6 **
	86dB de bruit	-0.4 F(1) = 1.0 NS (p=0.34)	0.6 F(1) = 1.5 NS (p=0.24)

Effet moyen, silence et tous niveaux de bruit confondus	D=-0.0 F(1) = 0.02 NS (p=0.88)	D=0.8 F(1) = 9.4 **
Variabilité moyenne de cet effet en fonction du niveau de bruit	0.4 F(4) = 1.2 NS (p=0.34)	0.6 F(4) = 2.1 NS (p=0.10)
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=0.01 r=20.5%

Identique	Légère amplification
-----------	----------------------

EFFET DU RETOUR DE SA PROPRE VOIX DANS LE CASQUE

Durée syllabique

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	8ms F(1) = 2.2 NS (p=0.14)	
	62dB de bruit		10ms F(1) = 3.4 NS (p=0.07)
	70dB de bruit		-3ms F(1) = 0.3 NS (p=0.58)
	78dB de bruit		2ms F(1) = 0.1 NS (p=0.78)
	86dB de bruit		-1ms F(1) = 0.01 NS (p=0.92)
Effet moyen, silence et tous niveaux de bruit confondus			D=3ms F(1) = 1.5 NS (p=0.23)
Variabilité moyenne de cet effet en fonction du niveau de bruit			7ms F(4) = 1.8 NS (p=0.13)
			Identique

Fréquence fondamentale

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	0.5 tons F(1) = 4.1 *	
	62dB de bruit		0.5 tons F(1) = 3.3 NS 0.07
	70dB de bruit		-0.7 tons F(1) = 7.7 **
	78dB de bruit		-0.1 tons F(1) = 0.1 NS (p=0.73)
	86dB de bruit		-0.3 tons F(1) = 1.9 NS (p=0.17)
Effet moyen, silence et tous niveaux de bruit confondus			D=-0.0 tons F(1) = 0.1 NS (p=0.81)
Variabilité moyenne de cet effet en fonction du niveau de bruit			0.6 tons F(4) = 25.3 ***
			Identique

Intensité vocale

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	4.5 dB F(1) = 67.1 ***	
	62dB de bruit	1.9 dB F(1) = 15.3 ***	3.2 dB F(1) = 42.7 ***
	70dB de bruit	2.4 dB F(1) = 24.7 ***	1.7 dB F(1) = 14.3 ***
	78dB de bruit	3.2 dB F(1) = 41.8 ***	2.1 dB F(1) = 21.1 ***
	86dB de bruit	3.6 dB F(1) = 58.2 ***	2.3 dB F(1) = 25.7 ***
Effet moyen, silence et tous niveaux de bruit confondus		D=3.1 dB F(1) = 109.0 ***	D=2.7 dB F(1) = 79.3 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit		1.2 dB F(4) = 6.0 ***	1.1 dB F(4) = 6.3 ***
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant			a=-0.06 r=-14.6%
		Décalage Compensation partielle de l'effet du port du casque	Réduction Compensation partielle de l'effet du port du casque

Différence d'énergie acoustique entre [1-2kHz] et [0-1kHz]

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	1.1 dB F(1) = 3.7 NS (p=0.05)	
	62dB de bruit		0.3 dB F(1) = 0.3 NS (p=0.61)
	70dB de bruit		-1.6 dB F(1) = 8.9 **
	78dB de bruit		0.1 dB F(1) = 0.02 NS (p=0.90)
	86dB de bruit		-0.5 dB F(1) = 0.8 NS (p=0.36)
Effet moyen, silence et tous niveaux de bruit confondus			D=-0.1 dB F(1) = 0.2 NS (p=0.62)
Variabilité moyenne de cet effet en fonction du niveau de bruit			1.2 dB F(4) = 17.7 ***
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant			a=-0.04 r=-13.9%
			Identique

Différence d'énergie acoustique entre [2-4kHz] et [0-1kHz]

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	1.0 dB F(1) = 4.2 *	
	62dB de bruit	-3.5 dB F(1) = 48.4 ***	-0.0 dB F(1) = 0.001 NS (p=0.98)
	70dB de bruit	-1.0 dB F(1) = 3.9 *	-1.6 dB F(1) = 9.8 **
	78dB de bruit	-0.2 dB F(1) = 0.2 NS (p=0.64)	0.1 dB F(1) = 0.01 NS (p=0.91)
	86dB de bruit	-0.0 dB F(1) = 0.001 NS(p=0.98)	-0.7 dB F(1) = 1.9 NS (p=0.17)
Effet moyen, silence et tous niveaux de bruit confondus		D=-0.8 dB F(1) = 7.2 **	D=-0.3 dB F(1) = 0.9 NS(p=0.34)
Variabilité moyenne de cet effet en fonction du niveau de bruit		2.0 dB F(4) = 31.0 ***	1.2 dB F(4) = 15.0 ***
		Identique	Identique

Quotient ouvert laryngé

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	-0.01 F(1) = 3.0 NS (p=0.08)	
	62dB de bruit	-0.02 F(1) = 8.4 **	-0.01 F(1) = 0.8 NS (p=0.38)
	70dB de bruit	0.00 F(1) = 0.01 NS (p=0.93)	0.01 F(1) = 2.4 NS (p=0.12)
	78dB de bruit	-0.00 F(1) = 0.0005 NS (p=0.98)	-0.00 F(1) = 0.1 NS (p=0.73)
	86dB de bruit	-0.01 F(1) = 0.5 NS (p=0.46)	-0.01 F(1) = 1.2 NS (p=0.27)
Effet moyen, silence et tous niveaux de bruit confondus		D=-0.01 F(1) = 4.6 *	D=-0.00 F(1) = 0.8 NS (p=0.38)
Variabilité moyenne de cet effet en fonction du niveau de bruit		0.01 F(4) = 6.3 ***	0.01 F(4) = 6.4 ***
		Identique	Identique

Différence d'énergie acoustique entre [4-6kHz] et [0-1kHz]

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	-0.0 dB F(1) = 0.0004 NS (p=0.98)	
	62dB de bruit	-3.7 dB F(1) = 51.5 ***	0.4 dB F(1) = 0.5 NS (p=0.49)
	70dB de bruit	-1.6 dB F(1) = 8.6 **	-2.1 dB F(1) = 14.6 ***
	78dB de bruit	-0.2 dB F(1) = 0.2 NS (p=0.68)	-0.5 dB F(1) = 1.0 NS (p=0.32)
	86dB de bruit	0.0 dB F(1) = 0.001 NS (p=0.98)	-0.6 dB F(1) = 1.3 NS (p=0.26)
Effet moyen, silence et tous niveaux de bruit confondus		D=-1.1 dB F(1) = 14.5481 ***	D=-0.6 dB F(1) = 4.1 *
Variabilité moyenne de cet effet en fonction du niveau de bruit		1.9 dB F(4) = 21.9934 ***	1.2 dB F(4) = 10.9 ***
		Identique	Identique

Vitesse de fermeture laryngée

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	0.6 u.a. F(1) = 48.4 ***	
	62dB de bruit	0.7 u.a. F(1) = 24.4 ***	0.9 u.a. F(1) = 31.2 ***
	70dB de bruit	0.8 u.a. F(1) = 19.1 ***	0.7 u.a. F(1) = 12.5 ***
	78dB de bruit	1.3 u.a. F(1) = 31.3 ***	0.9 u.a. F(1) = 14.2 ***
	86dB de bruit	1.6 u.a. F(1) = 35.7 ***	1.5 u.a. F(1) = 21.3 ***
Effet moyen, silence et tous niveaux de bruit confondus		D=1.0 u.a. F(1) = 108.0 ***	D=0.9 u.a. F(1) = 69.7 ***
Variabilité moyenne de cet effet en fonction du niveau de bruit		0.8 u.a. F(4) = 14.1 ***	0.6 u.a. F(4) = 7.5 ***
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant		a=0.02 r=13.5%	a=0.02 r=8.3%
		Amplification Compensation partielle de l'effet du port du casque	Amplification Compensation partielle de l'effet du port du casque

H1-H2

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	-1.7 dB F(1) = 16.3 ***	
	62dB de bruit	-0.6 dB F(1) = 0.6 NS (p=0.43)	-2.1 dB F(1) = 13.0 ***
	70dB de bruit	-0.3 dB F(1) = 0.18 NS (p=0.67)	0.5 dB F(1) = 0.6 NS (p=0.46)
	78dB de bruit	-0.9 dB F(1) = 1.4 NS (p=0.24)	-0.1 dB F(1) = 0.01 NS (p=0.94)
	86dB de bruit	-0.8 dB F(1) = 0.9 NS (p=0.33)	-0.1 dB F(1) = 0.004 NS (p=0.95)
Effet moyen, silence et tous niveaux de bruit confondus		D=-0.8 dB F(1) = 6.5229 *	D=-0.7 dB F(1) = 3.3 NS (p=0.07)
Variabilité moyenne de cet effet en fonction du niveau de bruit		0.7 dB F(4) = 0.9 NS (p=0.48)	1.3 dB F(4) = 3.3 *
		Identique	Identique

Question 2 : estimation de sa propre intelligibilité

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	-0.1 F(1) = 1 NS (p=0.33)	
	62dB de bruit		-0.3 F(1) = 1 NS (p=0.33)
	70dB de bruit		0.1 F(1) = 0.1 NS (p=0.75)
	78dB de bruit		-0.3 F(1) = 0.4 NS (p=0.52)
	86dB de bruit		-0.3 F(1) = 0.4 NS (p=0.53)
Effet moyen, silence et tous niveaux de bruit confondus			D=-0.2 F(1) = 0.6 NS (p=0.46)
Variabilité moyenne de cet effet en fonction du niveau de bruit			0.2 F(4) = 0.4 NS (p=0.77)
			Identique

Question 1 : perception de sa propre voix

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	0.0 F(1) = 0 NS (p=1)	
	62dB de bruit		0.1 F(1) = 0.03 NS (p=0.86)
	70dB de bruit		-0.1 F(1) = 0.1 NS (p=0.76)
	78dB de bruit		-0.2 F(1) = 0.4 NS (p=0.55)
	86dB de bruit		-0.8 F(1) = 1.9 NS (p=0.18)
Effet moyen, silence et tous niveaux de bruit confondus			D=-0.2 F(1) = 0.9 NS (p=0.36)
Variabilité moyenne de cet effet en fonction du niveau de bruit			0.4 F(4) = 1.1 NS (p=0.36)
Facteur d'amplification de cet effet en fonction du niveau de bruit croissant			a=-0.01 r=-22.0%
			Réduction

Question 3 : perception de l'interlocuteur

		Bruit blanc	Bruit cocktail
Effet du rajout dans le casque d'un retour de sa propre voix	silence	0.0 F(1) = 0 NS (p=1)	
	62dB de bruit		0.0 F(1) = 0 NS (p=1)
	70dB de bruit		0.3 F(1) = 0.9 NS (p=0.36)
	78dB de bruit		-0.4 F(1) = 0.8 NS (p=0.37)
	86dB de bruit		-0.4 F(1) = 0.7 NS (p=0.42)
Effet moyen, silence et tous niveaux de bruit confondus			D=-0.1 F(1) = 0.2 NS (p=0.70)
Variabilité moyenne de cet effet en fonction du niveau de bruit			0.4 F(4) = 0.9 NS (p=0.49)
			Identique

Question 4 : ressenti de gêne

		Bruit blanc	Bruit cocktail
Effet du rejet dans le casque d'un retour de sa propre voix	silence	0.2 F(1) = 0.5 NS (p=0.47)	
	62dB de bruit		-0.1 F(1) = 0.1 NS (p=0.81)
	70dB de bruit		0.1 F(1) = 0.04 NS (p=0.85)
	78dB de bruit		0.1 F(1) = 0.05 NS (p=0.82)
	86dB de bruit		-0.3 F(1) = 0.8 NS (p=0.39)

Effet moyen, silence et tous niveaux de bruit confondus		D=0.0 F(1) = 0 NS (p=1)
Variabilité moyenne de cet effet en fonction du niveau de bruit		0.2 F(4) = 0.5 NS (p=0.74)

	Identique
--	-----------

Annexe Res4 :
Résultats statistiques
détaillés du chapitre 4

Intensité vocale

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	2.3 dB F(1) = 51.0 ***
	Sur la parole produite dans le silence	12.8 dB F(1) = 1902.0 ***
	Sur la parole produite dans le bruit	15.2 dB F(1) = 2424.5 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	2.3 dB F(1) = 107.6 ***
	Sur la parole produite hors situation de dialogue	0.1 dB F(1) = 0.2 NS (p=0.66)
	Sur la parole produite en situation de dialogue	2.5 dB F(1) = 87.9 ***

Différence d'énergie acoustique entre [2-4kHz] et [0-1kHz]

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	0.3 dB F(1) = 1.1 NS (p=0.29)
	Sur la parole produite dans le silence	9.0 dB F(1) = 522.8 ***
	Sur la parole produite dans le bruit	9.3 dB F(1) = 676.2 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	0.3 dB F(1) = 1.6367 NS (p=0.20)
	Sur la parole produite hors situation de dialogue	0.5 dB F(1) = 1.7 NS (p=0.19)
	Sur la parole produite en situation de dialogue	0.8 dB F(1) = 4.7 *

Fréquence fondamentale

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	0.7 tons F(1) = 67.1 ***
	Sur la parole produite dans le silence	2.5 tons F(1) = 485.7 ***
	Sur la parole produite dans le bruit	3.1 tons F(1) = 1333.9 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	0.7 tons F(1) = 94.1 ***
	Sur la parole produite hors situation de dialogue	tons F(1) = 0.001 NS (p=0.97)
	Sur la parole produite en situation de dialogue	0.7 tons F(1) = 46.8 ***

Différence d'énergie acoustique entre [4-6kHz] et [0-1kHz]

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	1.3 dB F(1) = 6.1 *
	Sur la parole produite dans le silence	7.9 dB F(1) = 189.8 ***
	Sur la parole produite dans le bruit	9.2 dB F(1) = 409.5 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	1.3 dB F(1) = 16.0 ***
	Sur la parole produite hors situation de dialogue	0.6 dB F(1) = 2.4 NS (p=0.12)
	Sur la parole produite en situation de dialogue	1.8 dB F(1) = 8.3 **

Durée syllabique

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	10.8 ms F(1) = 19.4 ***
	Sur la parole produite dans le silence	18.4 ms F(1) = 27.8 ***
	Sur la parole produite dans le bruit	29.1 ms F(1) = 52.2 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	10.8 ms F(1) = 17.3 ***
	Sur la parole produite hors situation de dialogue	-4.1 ms F(1) = 1.4 NS (p=0.24)
	Sur la parole produite en situation de dialogue	6.6 ms F(1) = 2.7 NS (p=0.10)

Quotient ouvert laryngé

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	-0.01 F(1) = 4.4 *
	Sur la parole produite dans le silence	-0.02 F(1) = 60.3 ***
	Sur la parole produite dans le bruit	-0.03 F(1) = 112.3 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	-0.01 F(1) = 11.8 ***
	Sur la parole produite hors situation de dialogue	-0.00 F(1) = 0.3 NS (p=0.56)
	Sur la parole produite en situation de dialogue	-0.01 F(1) = 27.3 ***

Pincement labial

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	12.4 % F(1) = 15.0 ***
	Sur la parole produite dans le silence	2.7 % F(1) = 1.8 NS (p=0.19)
	Sur la parole produite dans le bruit	16.5 % F(1) = 32.9 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	12.4 % F(1) = 18.7 ***
	Sur la parole produite hors situation de dialogue	-2.7 % F(1) = 2.2 NS (p=0.14)
	Sur la parole produite en situation de dialogue	9.4 % F(1) = 10.7 **

Premier formant vocalique

Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	49Hz F(1) = 1.6 NS (p=0.22)
	Sur la parole produite dans le silence	72Hz F(1) = 0.6 NS (p=0.45)
	Sur la parole produite dans le bruit	172Hz F(1) = 179.6 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	49Hz F(1) = 1.2 NS(p=0.29)
	Sur la parole produite hors situation de dialogue	67Hz F(1) = 0.6 NS (p=0.46)
	Sur la parole produite en situation de dialogue	104Hz F(1) = 1.1 NS (p=0.32)

Aire intéro-labiale

		Voyelles [a]	Voyelles [i]	Voyelles [u]
Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	6.5% F(1) = 92.7 ***	2.9 % F(1) = 1.2 NS (p=0.28)	1.88 % F(1) = 8.2 *
	Sur la parole produite dans le silence	10.7% F(1) = 213.8 ***	8.4 % F(1) = 9.4 **	2.90 % F(1) = 14.9 ***
	Sur la parole produite dans le bruit	17.3% F(1) = 425.5 ***	11.3 % F(1) = 11.7 **	5.06 % F(1) = 37.7 ***
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	6.5% F(1) = 183.8 ***	2.9 % F(1) = 0.9 NS (p=0.36)	1.88 % F(1) = 8.7 **
	Sur la parole produite hors situation de dialogue	2.5% F(1) = 14.9 ***	4.7 % F(1) = 2.6 NS (p=0.12)	0.55 % F(1) = 0.6 NS (p=0.46)
	Sur la parole produite en situation de dialogue	9.0% F(1) = 97.9 ***	7.7 % F(1) = 6.0 *	2.76 % F(1) = 10.6 **

Protrusion labiale

		Voyelles [a]	Voyelles [i]	Voyelles [u]
Effet du dialogue avec l'expérimentateur	Sur l'évolution de la parole du silence au bruit	-6.1 % F(1) = 12.6 ***	5.1 % F(1) = 1.1 NS 0.29547	8.4 % F(1) = 1.3 NS (p=0.27)
	Sur la parole produite dans le silence	-7.4 % F(1) = 38.7 ***	-11.8 % F(1) = 6.4 *	3.3 % F(1) = 0.5 NS (p=0.48)
	Sur la parole produite dans le bruit	-13.2 % F(1) = 69.4 ***	-6.7 % F(1) = 1.5 NS (p=0.23)	11.8 % F(1) = 2.4 NS (p=0.14)
Effet de l'exposition au bruit (⇔ Effet Lombard)	Sur la différence entre la parole produite avec ou sans dialogue avec l'expérimentateur	-6.0% F(1) = 14.7 ***	5.1 % F(1) = 1.6 NS (p=0.21)	8.5 % F(1) = 1.7 NS (p=0.21)
	Sur la parole produite hors situation de dialogue	3.0 % F(1) = 5.8 *	-4.8 % F(1) = 1.4 NS (p=0.25)	8.4 % F(1) = 2.1 NS (p=0.17)
	Sur la parole produite en situation de dialogue	-3.0 % F(1) = 3.7 NS (p=0.05)	0.3 % F(1) = 0.003 NS (p=0.96)	16.8 % F(1) = 6.2 *

Annexe Res5 :
Résultats statistiques
détaillés du chapitre 5

Intensité vocale

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	8.9 dB (soit 10.1%)	16.4 dB (soit 20.4 %) F(1) = 2356.5, ***	22.2 dB (soit 28.1%)	Tous les cas	Tous les cas
BD2	12.7 dB (soit 12.6%)	16.8 dB (soit 17.5 %) F(1) = 4086.2, ***	23.1 dB (soit 24.6%)	Tous les cas	Tous les cas

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0.2 dB	3.0 dB 1/6ème de l' E.L F(9) = 11.7, ***	9.2 dB	36%	31%	16%
BD2	1.4 dB	4.6 dB ¼ de l E.L F(2) = 182.8, ***	7.0 dB	Tous les cas	2 cas / 3	1 cas / 3

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-6.9 dB	-2.5 dB 1/11ème de l' EL F(1) = 25.7959, ***	1.5 dB	5 cas / 10	2 cas / 10	1 cas / 10
BD2	-4.0 dB	-3.3 dB 1/5 ^{ème} de l'E.L F(1) = 19.8, ***	-2.2 dB	2 cas / 3	2 cas / 3	Aucun cas

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	20.6 dB F(1) = 254.4 ***	17.3 dB F(1) = 287.3 ***	17.5 dB F(1) = 419.6 ***	16.5 dB F(1) = 253.3 ***	20.0 dB F(1) = 522.1 ***
Effet du bruit cocktail	17.2 dB F(1) = 155.1 ***	16.8 dB F(1) = 332.8 ***	15.0 dB F(1) = 312.3 ***	9.5 dB F(1) = 46.5 ***	14.9 dB F(1) = 272.4 ***
Effet du type de bruit	-3.4dB F(1) = 4.6 *	-0.5 dB F(1) = 0.3 NS (p=0.58)	-2.6 dB F(1) = 11.3 **	-6.9 dB F(1) = 28.1 ***	-5.1 dB F(1) = 25.0 ***

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	17.0 dB F(1) = 176.9 ***	18.8 dB F(1) = 396.1 ***	20.7 dB F(1) = 322.3 ***	12.2 dB F(1) = 154.3 ***	15.7 dB F(1) = 174.7 ***
Effet du bruit cocktail	15.2 dB F(1) = 127 ***	17.6 dB F(1) = 330 ***	22.2 dB F(1) = 449.9 ***	19.0 dB F(1) = 84.3 ***	14.4 dB F(1) = 147.3 ***
Effet du type de bruit	-1.7 dB F(1) = 2.8 NS (p=0.10)	-1.2 dB F(1) = 1.2 NS (p=0.27)	1.5 dB F(1) = 1.2 NS (p=0.27)	-3.2 dB F(1) = 14.4 ***	-1.3 dB F(1) = 1.7 NS (p=0.19)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	17.3 dB F(1) = 681.3 ***	23.1 dB F(1) = 2105.7 ***	14.9 dB F(1) = 665 ***
Effet du bruit cocktail	13.7 dB F(1) = 774.3 ***	19.1 dB F(1) = 2972.6 ***	12.7 dB F(1) = 742.7 ***
Effet du type de bruit	-3.7 dB F(1) = 3.0 NS (p=0.08)	-4.0 dB F(1) = 72.9 ***	-2.2 dB F(1) = 8.1 **

Fréquence fondamentale

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	2.8 tons (soit 40.5%)	4.7 tons (soit 76.4%) F(1) = 701.3, ***	6.4 tons (soit 110.8%)	Tous les cas	Tous les cas
BD2	2.4 tons (soit 32.3%)	3.7 tons (soit 53.8%) F(1) = 2383.0, ***	4.7 tons (soit 74.1%)	Tous les cas	Tous les cas

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0 ton	0.6 ton 1/8ème de l'E.L F(9) = 3.7, ***	2.3 tons	9%	7%	2%
BD2	0.1 ton	0.8 tons 1/5ème de l'E.L F(2) = 76.7, ***	1.2 tons	2 cas / 3	2 cas / 3	Aucun cas

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-1.9 ton	-1.1 ton 1/5ème de l'E.L F(1) = 56.4, ***	-0.5 ton	8 cas / 10	3 cas / 10	1 cas / 10
BD2	-2.0 tons	-1.0 ton 1/5ème de l'E.L F(1) = 42.6, ***	-0.4 ton	2 cas / 3	1 cas / 3	1 cas / 3

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	5.6 tons F(1) = 624.8 ***	5.1 tons F(1) = 408.5605, ***	5.6 tons F(1) = 998.0082, ***	5.3 tons F(1) = 396.4534, ***	5.8 tons F(1) = 254.1756, ***
Effet du bruit cocktail	4.4 tons F(1) = 345.6 ***	4.5 tons F(1) = 470.9 ***	4.1 tons F(1) = 270.6 ***	3.8 tons F(1) = 158.6 ***	5.1 tons F(1) = 66.8 ***
Effet du type de bruit	-1.2 tons F(1) = 12.0 **	-0.62 tons F(1) = 5.1 *	-1.5 tons F(1) = 22.3 ***	-1.5 tons F(1) = 17.1 ***	-0.7 tons F(1) = 0.006 NS (p= 0.94)

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	5.1 tons F(1) = 272.0 ***	6.4 tons F(1) = 428.2 ***	5.5 tons F(1) = 200.1 ***	4.8 tons F(1) = 342.4 ***	3.7 tons F(1) = 199.1 ***
Effet du bruit cocktail	4.6 tons F(1) = 231.5 ***	5.5 tons F(1) = 272.2 ***	4.2 tons F(1) = 119.9 ***	2.9 tons F(1) = 98.1 ***	2.8 tons F(1) = 83.4 ***
Effet du type de bruit	-0.5 tons F(1) = 1.6 NS (p=0.22)	-0.8 tons F(1) = 4.2 *	-1.3 tons F(1) = 6.2 *	-1.9 tons F(1) = 21.3 ***	-0.9 tons F(1) = 5.8 *

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	4.7 tons F(1) = 1004.9 ***	4.4 tons F(1) = 1192.3 ***	3.4 tons F(1) = 2004.9 ***
Effet du bruit cocktail	4.0 tons F(1) = 818.5 ***	2. tons F(1) = 346.3 ***	3.0 tons F(1) = 592.3 ***
Effet du type de bruit	-0.7 tons F(1) = 4.2 *	-2.0 tons F(1) = 185.3 ***	-0.4 tons F(1) = 0.004 NS (p=0.95)

Energie acoustique entre 0 et 1 kHz

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	9.4 dB (soit 10.4%)	16.2 dB (soit 19.5%) F(1) = 3059.4, ***	22.4 dB (soit 28.2%)	Tous les cas	Tous les cas
BD2	10.7 dB (soit 11.9%)	15.3 dB (soit 17.9%) F(1) = 4317.4, ***	21.3 dB (soit 25.5%)	Tous les cas	Tous les cas

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0 dB	3.4 dB 1/5ème de l E.L F(9) = 12.9651, ***	10.6 dB	47%	33%	22%
BD2	2.4 dB	5.1 dB 1/3 de l E.L F(2) = 221.6, ***	7.7 dB	Tous les cas	2 cas / 3	2 cas / 3

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-5.6 dB	-2.7 dB 1/15ème de l'E.L F(1) = 29.7, ***	-0.3 dB	5 cas / 10	2 cas / 10	1 cas / 10
BD2	-3.8 dB	-3.1dB 1/5ème de l E.L F(1) = 15.7, ***	-2.2 dB	Tous les cas	2 cas / 3	Aucun cas

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	20.2 dB F(1) = 485.6 ***	18.4 dB F(1) = 429.7 ***	16.9 dB F(1) = 464.0 ***	15.1 dB F(1) = 340.4 ***	18.5 dB F(1) = 644.1 ***
Effet du bruit cocktail	15.9 dB F(1) = 267.2 ***	16.7 dB F(1) = 501.3 ***	13.7 dB F(1) = 259.0 ***	9.5 dB F(1) = 72.1 ***	13.3 dB F(1) = 250.1 ***
Effet du type de bruit	-4.4 dB F(1) = 11.3 **	-1.7 dB F(1) = 1.6 NS (p= 0.20)	-3.2 dB F(1) = 6.9 *	-5.6 dB F(1) = 17.3 ***	-5.2 dB F(1) = 19.7 ***

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	17.5 dB F(1) = 203.0 ***	19.7 dB F(1) = 489.9 ***	22.4 dB F(1) = 195.0 ***	12.0 dB F(1) = 173.9 ***	15.1 dB F(1) = 164.4 ***
Effet du bruit cocktail	16.6 dB F(1) = 175.8 ***	17.7 dB F(1) = 282.2 ***	22.1 dB F(1) = 486.3 ***	9.4 dB F(1) = 104.1 ***	13.9 dB F(1) = 140.9 ***
Effet du type de bruit	-0.9 dB F(1) = 0.5 NS (p=0.47)	-1.9 dB F(1) = 3.3 NS (p=0.07)	-0.3 dB F(1) = 0.05 NS (p=0.83)	-2.6 dB F(1) = 9.7 **	-1.2 dB F(1) = 1.1 NS (p=0.31)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	16.3 dB F(1) = 946.1 ***	21.3 dB F(1) = 3663.1 ***	12.9 dB F(1) = 698.8 ***
Effet du bruit cocktail	12.6 dB F(1) = 859.3 ***	18.0 dB F(1) = 3564.0 ***	10.7 dB F(1) = 580.8 ***
Effet du type de bruit	-3.8 dB F(1) = 3.9 *	-3.3 dB F(1) = 63.8 ***	-2.2 dB F(1) = 7.4 **

Energie acoustique entre 1 et 2 kHz

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	16.2 dB (soit 23.6%)	27.3 dB (soit 46.4%) F(1) = 2443.2, ***	36.6 dB (soit 71.7%)	Tous les cas	Tous les cas
BD2	19.0 dB (soit 24.3%)	25.2 dB (soit 35.0%) F(1) = 2357.1, ***	30.9 dB (soit 44.9%)	Tous les cas	Tous les cas

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0.2 dB	5.0 dB 1/5ème de l'E.L F(9) = 10.3, ***	14.3 dB	29%	24%	16%
BD2	0.2 dB	4.6 dB 1/6ème de l'E.L F(2) = 108.6, ***	6.9 dB	2 cas / 3	2 cas / 3	Aucun cas

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-11.1 dB	-3.8 dB 1/21ème de l'E.L F(1) = 23.2, ***	0.9 dB	5 cas / 10	3 cas / 10	1 cas / 10
BD2	-6.3 dB	-4.7 dB 1/6ème de l'E.L F(1) = 20.9, ***	-2.6 dB	2 cas / 3	2 cas / 3	Aucun cas

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	36.6 dB F(1) = 298.9 ***	31.7 dB F(1) = 247.2 ***	29.2 dB F(1) = 265.5 ***	31.6 dB F(1) = 169.8 ***	31.7 dB F(1) = 162.2 ***
Effet du bruit cocktail	29.5 dB F(1) = 184.3 ***	30.8 dB F(1) = 294.3 ***	22.7 dB F(1) = 137.3 ***	20.5 dB F(1) = 61.2 ***	24.3 dB F(1) = 89.9 ***
Effet du type de bruit	-7.1 dB F(1) = 16.6 ***	-0.8 dB F(1) = 0.3 NS (p=0.59)	-6.5 dB F(1) = 9.7 **	-11.1 dB F(1) = 20.9 ***	-7.3 dB F(1) = 16.3 ***

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	25.9 dB F(1) = 126.5 ***	28.5 dB F(1) = 204.1 ***	33.2 dB F(1) = 290.3 ***	19.6 dB F(1) = 130.9 ***	23.7 dB F(1) = 107.5 ***
Effet du bruit cocktail	26.2 dB F(1) = 143.9 ***	27.6 dB F(1) = 194.4 ***	34.1 dB F(1) = 371.2 ***	16.2 dB F(1) = 85.9 ***	21.6 dB F(1) = 97.0 ***
Effet du type de bruit	0.4 dB F(1) = 0.04 NS (p=0.83)	-0.9 dB F(1) = 0.3 NS (p=0.58)	0.9 dB F(1) = 0.2 NS (p=0.69)	-3.5 dB F(1) = 6.2 *	-2.1 dB F(1) = 1.1 NS (p=0.30)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	29.9 dB F(1) = 342.1 ***	30.9 dB F(1) = 627.2 ***	21.6 dB F(1) = 270.3 ***
Effet du bruit cocktail	24.9 dB F(1) = 497.7 ***	24.6 dB F(1) = 823.2415, ***	19.0 dB F(1) = 387.6 ***
Effet du type de bruit	-5.1 dB F(1) = 1.7 NS (p=0.19)	-6.3 dB F(1) = 90.9 ***	-2.6 dB F(1) = 5.3 *

Energie acoustique entre 2 et 4 kHz

	Effet de l'exposition au bruit ↔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	20.4 dB (soit 32.7%)	27.7 dB (soit 50.5%) F(1) = 2973.1, ***	38.8 dB (soit 82.5%)	Tous les cas	Tous les cas
BD2	20.8 dB (soit 28.3%)	25.1 dB (soit 37.7%) F(1) = 3489.8, ***	31.0 dB (soit 46.7%)	Tous les cas	Tous les cas

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0.1 dB	5.5 dB 1/5ème de l'E.L F(9) = 11.2, ***	14.6 dB	38%	33%	13%
BD2	2.0 dB	3.3 dB 1/8ème de l'E.L F(2) = 49.3, ***	5.0 dB	Tous les cas	1 cas / 3	Aucun cas

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-8.0 dB	-1.7 dB 1/168ème de l'E.L F(1) = 4.1, *	4.3 dB	4 cas / 10	Aucun cas	Aucun cas
BD2	-6.4 dB	-4.0 dB 1/7ème de l'E.L F(1) = 25., ***	-2.5 dB	2 cas / 3	1 cas / 3	Aucun cas

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	38.8 dB F(1) = 503.8 ***	30.1 dB F(1) = 241.1 ***	26.6 dB F(1) = 171.067, ***	33.9 dB F(1) = 423.7 ***	31.5 dB F(1) = 258.7 ***
Effet du bruit cocktail	32.9 dB F(1) = 358.2 ***	31.1 dB F(1) = 300.3 ***	22.1 dB F(1) = 112.2 ***	25.9 dB F(1) = 125.8 ***	24.9 dB F(1) = 166.5 ***
Effet du type de bruit	-6.0 dB F(1) = 8.8 **	1.0 dB F(1) = 0.4 NS (p=0.54)	-4.5 dB F(1) = 5.7 *	-8.0 dB F(1) = 13.2 **	-6.6 dB F(1) = 13.0 ***

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	22.2 dB F(1) = 140.7 ***	26.8 dB F(1) = 240.3 ***	32.6 dB F(1) = 261.8 ***	21.7 dB F(1) = 195.0 ***	21.8 dB F(1) = 122.0 ***
Effet du bruit cocktail	24.5 dB F(1) = 177.7 ***	28.3 dB F(1) = 308.4 ***	36.9 dB F(1) = 462.6 ***	20.4 dB F(1) = 183.0 ***	21.8 dB F(1) = 120.2 ***
Effet du type de bruit	2.3 dB F(1) = 1.5 NS (p=0.22)	1.6 dB F(1) = 0.9 NS (p=0.36)	4.3 dB F(1) = 4.0 NS (p= 0.05)	-1.3 dB F(1) = 0.7 NS (p=0.42)	0.01 dB F(1) = 0.0005 NS (p=0.99)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	26.4 dB F(1) = 582.8 ***	31.0 dB F(1) = 1285.1 ***	23.8 dB F(1) = 582.2 *** dB
Effet du bruit cocktail	23.9 dB F(1) = 1117.5 ***	24.6 dB F(1) = 2074.9 ***	20.8 F(1) = 766.5 ***
Effet du type de bruit	-2.5 dB F(1) = 0.02 NS (p=0.88)	6.4 dB F(1) = 90.3 ***	-3.0 dB F(1) = 9.4 **

Energie acoustique entre 4 et 6 kHz

	Effet de l'exposition au bruit ↔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	16.8 dB (soit 35.8%)	25.7 dB (soit 54.9%) F(1) = 2795.8, ***	39.0 dB (soit 100.6%)	Tous les cas	Tous les cas
BD2	11.5 dB (soit 21.3%)	23.0 dB (soit 38.9%) F(1) = 1864.3, ***	31.1dB (soit 55.8%)	Tous les cas	Tous les cas

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0 dB	6.5 dB ¼ de l E.L F(9) = 16.2587, ***	18.3 dB	56%	44%	29%
BD2	3.5 dB	8.0 dB 1/3 de l E.L F(2) = 219.3, ***	12.1 dB	Tous les cas	2 cas / 3	2 cas / 3

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-7.6 dB	-1.0 dB 1/95 ^{ème} de l'E.L F(1) = 0.9, ns (p=0.33)	4.4 dB	3 cas / 10	Aucun cas	Aucun cas
BD2	-4.4 dB	-0.2 dB 1/6 ^{ème} de l E.L F(1) = 1.8, ns (p=0.18)	6.9 dB	Tous les cas	2 cas / 3	1 cas / 3

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	39.0 dB F(1) = 578.3 ***	28.9 dB F(1) = 359.9 ***	23.3 dB F(1) = 147.9 ***	31.4 dB F(1) = 466.4 ***	28.8 dB F(1) = 287.5 ***
Effet du bruit cocktail	34.0 dB F(1) = 434.1 ***	29.5 dB F(1) = 348.2 ***	20.6 dB F(1) = 116.5 ***	23.8 dB F(1) = 109.3 ***	23.1 dB F(1) = 189.8 ***
Effet du type de bruit	-5.0 dB F(1) = 3.9 NS (p=0.05)	0.6 dB F(1) = 0.1 NS (p=0.72)	-2.7 dB F(1) = 1.5 NS (p=0.23)	-7.6 dB F(1) = 14.2 ***	-5.6 dB F(1) = 9.7 **

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	16.8 dB F(1) = 102.2 ***	24.8 dB F(1) = 213.8 ***	28.5 dB F(1) = 220.3 ***	20.3 dB F(1) = 240.0 ***	19.6 dB F(1) = 111.8 ***
Effet du bruit cocktail	21.1 dB F(1) = 124.9 ***	28.1 dB F(1) = 297.8 ***	32.9 dB F(1) = 335.5 ***	18.8 dB F(1) = 212.0 ***	19.7 dB F(1) = 114.8 ***
Effet du type de bruit	4.3 dB F(1) = 4.5 *	3.2 dB F(1) = 3.5 NS (p=0.06)	4.4 dB F(1) = 3.8 NS (p=0.06)	-1.5 dB F(1) = 0.9 NS (p=0.36)	0.02 dB F(1) = 0.0001 NS (p=0.99)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	11.5 dB F(1) = 125.3 ***	31.1 dB F(1) = 1071.6 ***	26.6 dB F(1) = 597.4 ***
Effet du bruit cocktail	18.4 dB F(1) = 974.4 ***	26.7 dB F(1) = 2127.3 ***	23.7 dB F(1) = 1037.4 ***
Effet du type de bruit	6.9 dB F(1) = 84.4 ***	-4.4 dB F(1) = 19.1 ***	-2.9 dB F(1) = 4.4 *

Durée syllabique

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	-6 ms (soit -0.3%)	26 ms (soit 17.5%) F(1) = 42.8, ***	59 ms (soit 38.6%)	8 cas / 20	2 cas / 20
BD2	13 ms (soit 9.2%)	39 ms (soit 25.8%) F(1) = 124.1, ***	63 ms (soit 43.6%)	Tous les cas	2 cas / 6

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0 ms	24 ms équivalent de l E.L F(9) = 3.3, ***	64 ms	9 %	2 %	82 %
BD2	11 ms	22 ms moitié de l E.L F(2) = 44.2, ***	34 ms	Tous les cas	1 cas / 3	2 cas / 3

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-44 ms	-16 ms équivalent à l effet F(1) = 8.3, **	28 ms	2 cas / 10	0 cas / 10	10 cas / 10
BD2	-35 ms	-21ms Moitié de l E.L F(1) = 24.9, ***	-14 ms	2 cas / 3	1 cas / 3	2 cas / 3

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	34ms F(1) = 3.34 NS (p=0.07)	51ms F(1) = 13.4 ***	53ms F(1) = 5.9 *	59ms F(1) = 19.8 ***	49ms F(1) = 4.8 *
Effet du bruit cocktail	0 ms F(1) = 0.005 NS (p=0.98)	33ms F(1) = 5.8 *	8ms F(1) = 0.2 NS (p=0.63)	26ms F(1) = 2.7 NS (p=0.11)	35ms F(1) = 1.5 NS (p=0.22)
Effet du type de bruit	-34ms F(1) = 2.5 NS (p=0.12)	-18ms F(1) = 3.7 NS (p=0.06)	-44ms F(1) = 10.7 **	-33ms F(1) = 1.7 NS (p=0.20)	-15ms F(1) = 0.04 NS (p=0.84)

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	10ms F(1) = 0.4 NS (p=0.52)	14ms F(1) = 0.6 NS (p=0.43)	56ms F(1) = 27.7 ***	11ms F(1) = 0.6 NS (p= 0.45)	7ms F(1) = 0.2 NS (p=0.64)
Effet du bruit cocktail	38ms F(1) = 6.0 *	20ms F(1) = 1.3181, NS (p=0.26)	33ms F(1) = 13.1 ***	-6ms F(1) = 0.2 NS (p=0.66)	-5ms F(1) = 0.1 NS (p=0.72)
Effet du type de bruit	28ms F(1) = 2.6 NS (p=0.11)	6ms F(1) = 0.1 NS (p=0.73)	-23ms F(1) = 5.4 *	-17ms F(1) = 2.1 NS (p=0.15)	-12ms F(1) = 1.1 NS (p=0.31)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	60ms F(1) = 67.7 ***	63ms F(1) = 34.5 ***	26ms F(1) = 14.0 ***
Effet du bruit cocktail	46ms F(1) = 40.9 ***	28ms F(1) = 19.1 ***	13ms F(1) = 9.7 **
Effet du type de bruit	-14ms F(1) = 2.5 NS (p=0.12)	-35ms F(1) = 41.3 ***	-14ms F(1) = 5.2 *

Amplitude maximale des mouvements labiaux

		Effet de l'exposition au bruit ↔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
		Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
ΔMax A	[a]	1.7 %	7.6 %, F(1) = 55.4, ***	10.2 %	5 cas / 20	5 cas / 20
	[i]	3.0%	7.1 %, F(1) = 31.4, ***	12.7 %	4 cas / 20	6 cas / 20
	[u]	7.9 %	21.2 %, F(1) = 47.2, ***	37.5 %	4 cas / 20	6 cas / 20
ΔMax B	[a]	10.0 %	17.7 %, F(1) = 65.4, ***	31.0 %	6 cas / 20	6 cas / 20
	[i]	4.6 %	7.8 %, F(1) = 26.7, ***	11.7 %	4 cas / 20	6 cas / 20
	[u]	3.5%	8.3 %, F(1) = 60.1, ***	15.6 %	3 cas / 20	6 cas / 20
ΔMax S	[a]	11.4%	25.1 %, F(1) = 53.3, ***	38.7%	6 cas / 20	6 cas / 20
	[i]	6.9 %	13.6 %, F(1) = 24.2, ***	26.9 %	4 cas / 20	6 cas / 20
	[u]	3.1 %	4.9 %, F(1) = 50.6, ***	8.6 %	3 cas / 20	6 cas / 20
ΔMax P1	[a]	-42.3%	-23.5 %, F(1) = 27.2, ***	-9.4	4 cas / 20	0 cas / 20
	[i]	-21.3 %	-14.0 %, F(1) = 11.0, **	1.4 %	3 cas / 20	0 cas / 20
	[u]	-1.0 %	12.2 %, F(1) = 4.4, *	21.2 %	0 cas / 20	2 cas / 20

		Différence interindividuelle moyenne			Différence interindividuelle supérieure à	
		Min	Moyenne	Max	Variabilité intra-individuelle	Tiers de l'E.L
ΔMax A	[a]	0%	4% de l'éirement maximal, Moitié de l'E.L, F(2) = 10.9, ***	6%	2 cas / 3	2 cas / 3
	[i]	1%	3% de l'éirement maximal, Moitié de l'E.L, F(2) = 4.2, *	4%	2 cas / 3	2 cas / 3
	[u]	1%	9% de l'éirement maximal, Moitié de l'E.L, F(2) = 3.3, ns (p=0.06)	14%	2 cas / 3	2 cas / 3
ΔMax B	[a]	2%	4% de l'ouverture maximale, 1/5 ^{ème} de l'E.L, F(2) = 1.3, ns (p=0.28)	5%	1 cas / 3	Aucun cas
	[i]	1%	1.7% de l'ouverture maximale, 1/5 ^{ème} de l'E.L, F(2) = 1.0, ns (p= 0.39)	3%	Aucun cas	Aucun cas
	[u]	0%	3% de l'ouverture maximale, Tiers de l'E.L, F(2) = 1.3, ns (p=0.29)	4%	Aucun cas	2 cas / 3
ΔMax S	[a]	1%	8% de l'aire intéro- labiale maximale, Tiers de l'E.L, F(2) = 5.2, *	13%	2 cas / 3	2 cas / 3
	[i]	3%	5% de l'aire intéro- labiale maximale, Tiers de l'E.L, F(2) = 2.5, ns (p=0.10)	7%	2 cas / 3	1 cas / 3
	[u]	0.2%	0.3% de l'aire intéro- labiale maximale, 1/15 ^{ème} de l'E.L, F(2) = 0.06, ns (p=0.94)	0.5%	Aucun cas	Aucun cas
ΔMax P1	[a]	1%	5% de la protrusion maximale, 1/5 ^{ème} de l'E.L, F(2) = 0.6, ns (p=0.56)	8%	Aucun cas	Aucun cas
	[i]	6%	10% de la protrusion maximale, équivalent de l'E.L, F(2) = 4.3, *	15%	1 cas / 3	Tous les cas
	[u]	2%	8% de la protrusion maximale, moitié de l'E.L, F(2) = 1.3, ns (p=0.29)	13%	Aucun cas	2 cas / 3

		Effet du type de bruit			Effet du type de bruit supérieur à		
		Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
ΔMax A	[a]	-6%	-3%, 1/8ème de l'E.L F(1) = 0.1, NS (p=0.76)	0%	1 cas / 10	1 cas / 10	1 cas / 10
	[i]	-6%	-2%, Moitié de l'E.L F(1) = 1.4, NS (p=0.26)	2%	0 cas / 10	1 cas / 10	2 cas / 10
	[u]	-21%	-7% Quart de l'E.L F(1) = 0.1, NS (p=0.72)	4%	0 cas / 10	1 cas / 10	2 cas / 10
ΔMax B	[a]	-21%	-12%, Moitié de l'E.L F(1) = 6.0, *	-7%	1 cas / 10	3 cas / 10	3 cas / 10
	[i]	-6%	-4%, Tiers de l'E.L F(1) = 2.0, NS (p=0.18)	-2%	0 cas / 10	2 cas / 10	2 cas / 10
	[u]	-10%	-2%, 1/61ème de l'E.L F(1) = 0, NS (p=1)	2%	0 cas / 10	1 cas / 10	2 cas / 10
ΔMax S	[a]	-27%	-16%, Moitié de l'E.L F(1) = 4.9, NS (p=0.05)	-8%	1 cas / 10	3 cas / 10	3 cas / 10
	[i]	-15%	-8%, Quart de l'E.L F(1) = 2.1, NS (p=0.17)	-1%	0 cas / 10	2 cas / 10	2 cas / 10
	[u]	-6%	-2%, 1/37ème de l'E.L F(1) = 0.001, NS (p=0.94)	1%	0 cas / 10	1 cas / 10	1 cas / 10
ΔMax P1	[a]	20%	24%, Equivalent à l'E.L F(1) = 2.6, NS (p=0.14)	27%	1 cas / 10	3 cas / 10	3 cas / 10
	[i]	2	9%, Quart de l'E.L F(1) = 1.8, NS (p=0.20)	23%	0 cas / 10	1 cas / 10	1 cas / 10
	[u]	-14%	-9%, Moitié de l'E.L F(1) = 0.2, NS (p=0.64)	-6%	0 cas / 10	0 cas / 10	3 cas / 10

			Locutrices de BD2		
			L2	L6	L11
Max A	[a]	Effet du bruit blanc	7.5% F(1) = 186.5, **	9.2% F(1) = 107.3, **	10.2% F(1) = 126.5, **
		Effet du bruit cocktail	1.6% F(1) = 0.9, NS (p=0.37)	8.7% F(1) = 34.1, ***	8.5% F(1) = 17.7, ***
		Effet du type de bruit	-5.8% F(1) = 41.0, *	-0.5% F(1) = 2.8, NS (p=0.23)	-1.8% F(1) = 8.0, NS (p=0.10)
	[i]	Effet du bruit blanc	6.5% F(1) = 3.3, NS (p=0.21)	12.7% F(1) = 81.1, *	5.8% F(1) = 215.4, **
		Effet du bruit cocktail	3.0% F(1) = 3.2, NS (p=0.09)	6.9% F(1) = 17.5, ***	7.9% F(1) = 8.7, *
		Effet du type de bruit	-3.5% F(1) = 3.0, NS (p=0.22)	-5.8% F(1) = 1.6, NS (p=0.33)	2.1% F(1) = 8.2, NS (p=0.10)
	[u]	Effet du bruit blanc	37.5% F(1) = 374.2, **	7.9% F(1) = 10.9, NS (p=0.08)	28.5% F(1) = 11.3, NS (p=0.08)
		Effet du bruit cocktail	16.7% F(1) = 12.8, **	11.8% F(1) = 12.7, **	24.8% F(1) = 13.9, **
		Effet du type de bruit	-20.8% F(1) = 1.0, NS (p=0.42)	3.8% F(1) = 0.58, NS (p=0.53)	-3.7% F(1) = 1.5, NS (p=0.35)

Max B	[a]	Effet du bruit blanc	31.0% F(1) = 72.07, *	23.0% F(1) = 4765.7, ***	17.0% F(1) = 44.6, *
		Effet du bruit cocktail	10.2% F(1) = 12.6, **	15.2% F(1) = 245.2, ***	10.0% F(1) = 16.8, **
		Effet du type de bruit	-20.9% F(1) = 5.7, NS (p=0.14)	-7.8% F(1) = 72.0, *	-7.1% F(1) = 0.002, NS (p=0.97)
	[i]	Effet du bruit blanc	7.8% F(1) = 5.3, NS 0.14741	11.7% F(1) = 58.0, *	9.2% F(1) = 72.0, *
		Effet du bruit cocktail	4.6% F(1) = 4.2, NS (p=0.06)	5.9% F(1) = 23.4, ***	7.3% F(1) = 12.8, **
		Effet du type de bruit	-3.2% F(1) = 3.2, NS (p=0.22)	-5.8% F(1) = 1.05, NS (p=0.41)	-1.9% F(1) = 0.006, NS (p=0.94)
	[u]	Effet du bruit blanc	15.6% F(1) = 8.4, NS (p=0.10)	3.5% F(1) = 0.7, NS (p=0.48)	9.4% F(1) = 6.4, NS (p=0.13)
		Effet du bruit cocktail	6.0% F(1) = 23.0, ***	5.8% F(1) = 17.1, **	9.5% F(1) = 19.5, **
		Effet du type de bruit	-9.6% F(1) = 1.5, NS (0.35)	2.3% F(1) = 0.3, NS (p=0.65)	0.05% F(1) = 1.6, NS (p=0.33)

Max S	[a]	Effet du bruit blanc	38.2% F(1) = 145.0, **	38.7% F(1) = 6009.8, ***	22.4% F(1) = 79.8, *
		Effet du bruit cocktail	11.4% F(1) = 7.8, *	25.8% F(1) = 208.1392, ***	14.0% F(1) = 17.2, ***
		Effet du type de bruit	-26.8% F(1) = 13.6, NS (p=0.07)	-12.9% F(1) = 123.9575, **	-8.5% F(1) = 7.3, NS (p=0.11)
	[i]	Effet du bruit blanc	13.1% F(1) = 4.6, NS (p=0.16)	26.9% F(1) = 876.6, **	12.1% F(1) = 148.8, **
		Effet du bruit cocktail	6.9% F(1) = 4.0, NS (p=0.06)	12.0% F(1) = 27.4, ***	10.6% F(1) = 11.2, **
		Effet du type de bruit	-6.2% F(1) = 9.4, NS p=0.09	-15.0% F(1) = 1.4, NS (p=0.35)	-1.5% F(1) = 0.08, NS (p=0.81)
	[u]	Effet du bruit blanc	8.6% F(1) = 8.2, NS (p=0.10)	4.6% F(1) = 3.0, NS (p=0.22)	3.8% F(1) = 3.1, NS (p=0.22)
		Effet du bruit cocktail	3.1% F(1) = 17.8, ***	4.6% F(1) = 19.5, ***	4.6% F(1) = 19.3, **
		Effet du type de bruit	-5.5% F(1) = 1.1, NS (p=0.40)	0.04% F(1) = 0.1, NS (p=0.76)	0.8% F(1) = 2.1, NS (p=0.28)

Max P1	[a]	Effet du bruit blanc	-42.3% F(1) = 549.6, **	-34.5% F(1) = 1239.2, ***	-29.3% F(1) = 6.7, NS (p=0.12)
		Effet du bruit cocktail	-15.7% F(1) = 10.1, **	-9.9% F(1) = 5.9644, *	-9.4% F(1) = 2.2, NS (p=0.16)
		Effet du type de bruit	26.7% F(1) = 25.5, *	24.5% F(1) = 15.4, NS (p=0.06)	19.8% F(1) = 1.3, NS (p=0.37)
	[i]	Effet du bruit blanc	-21.2% F(1) = 2.7, NS (p=0.24)	-13.7% F(1) = 26.8, *	-21.3% F(1) = 0.9, NS (p=0.44)
		Effet du bruit cocktail	-17.6% F(1) = 16.3, **	-11.5% F(1) = 6.5, *	1.4% F(1) = 0.1, NS (p=0.75)
		Effet du type de bruit	3.6% F(1) = 0.8, NS (p=0.48)	2.2% F(1) = 0.001, NS (p=0.98)	22.7% F(1) = 1.3, NS (p=0.37)
	[u]	Effet du bruit blanc	12.7% F(1) = 0.6, NS (p=0.52)	16.8% F(1) = 0.8, NS (p=0.46)	21.2% F(1) = 2.4, NS (p=0.26)
		Effet du bruit cocktail	-0.9% F(1) = 0.01, NS (p=0.91)	11.0% F(1) = 1.9 NS (p=0.19)	12.6% F(1) = 2.9 NS (p=0.11)
		Effet du type de bruit	-13.6% F(1) = 0.01, NS (p=0.92)	-5.7% F(1) = 0.00005, NS (p=0.99)	-8.6% F(1) = 0.6, NS (p=0.51)

Pic de vitesse des mouvements labiaux

Max Vitesse S	[a]	Effet du bruit blanc	0.45 %·s ⁻¹ F(1) = 0.6, NS (p=0.53)	5.56%·s ⁻¹ F(1) = 21.9, *	2.29%·s ⁻¹ F(1) = 2.3, NS (p=0.27)
		Effet du bruit cocktail	0.94%·s ⁻¹ F(1) = 1.4, NS (p=0.25)	5.28%·s ⁻¹ F(1) = 8.3, *	1.45%·s ⁻¹ F(1) = 5.2, *
		Effet du type de bruit	0.49%·s ⁻¹ F(1) = 9.0, NS (p=0.09)	-0.28%·s ⁻¹ F(1) = 0.5, NS (p=0.54)	-0.85%·s ⁻¹ F(1) = 0.2, NS (p=0.66)
	[i]	Effet du bruit blanc	2.83%·s ⁻¹ F(1) = 1.2, NS (p=0.39)	1.22%·s ⁻¹ F(1) = 84.7, *	-0.005%·s ⁻¹ F(1) = 0.0001, NS (p=0.99)
		Effet du bruit cocktail	2.01 %·s ⁻¹ F(1) = 6.3, *	0.63%·s ⁻¹ F(1) = 0.05, NS (p=0.82)	0.87%·s ⁻¹ F(1) = 1.6, NS (p=0.24)
		Effet du type de bruit	-0.82%·s ⁻¹ F(1) = 0.1, NS (p=0.78)	-0.59%·s ⁻¹ F(1) = 92.2, *	0.88%·s ⁻¹ F(1) = 1.2, NS (p=0.38)
	[u]	Effet du bruit blanc	2.62%·s ⁻¹ F(1) = 1.6, NS (p=0.34)	2.08%·s ⁻¹ F(1) = 5.7, NS (p=0.14)	2.66%·s ⁻¹ F(1) = 1.9, NS (p=0.30)
		Effet du bruit cocktail	-1.61%·s ⁻¹ F(1) = 2.4, NS (p=0.15)	0.06%·s ⁻¹ F(1) = 0.01, NS (p=0.92)	1.85%·s ⁻¹ F(1) = 6.8, *
		Effet du type de bruit	-4.24%·s ⁻¹ F(1) = 0.8, NS (p=0.46)	-2.02 %·s ⁻¹ F(1) = 7.2, NS (p=0.11)	-0.81%·s ⁻¹ F(1) = 0.1, NS (p=0.81)

Max Vitesse P1	[u]	Effet du bruit blanc	2.67%·s ⁻¹ F(1) = 0.5, NS (p=0.54)	2.09%·s ⁻¹ F(1) = 0.2, NS (p=0.70)	-2.03%·s ⁻¹ F(1) = 0.2, NS (p=0.73)
		Effet du bruit cocktail	2.58%·s ⁻¹ F(1) = 1.3, NS (p=0.28)	-1.43%·s ⁻¹ F(1) = 0.2, NS (p=0.69)	2.32%·s ⁻¹ F(1) = 0.7, NS (p=0.43)
		Effet du type de bruit	-0.09%·s ⁻¹ F(1) = 0.04, NS (p=0.86)	-3.53%·s ⁻¹ F(1) = 0.01, NS (p=0.92)	4.35%·s ⁻¹ F(1) = 0.6, NS (p=0.53)

Fréquence centrale du premier formant vocalique (F1)

	BD1					BD2				
	Effet de l'exposition au bruit ↔ Effet Lombard (E.L.)			Effet de l'exposition au bruit supérieur à		Effet de l'exposition au bruit ↔ Effet Lombard (E.L.)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
[a]	93 Hz	247 Hz F(1) = 139.9 ***	407 Hz	20 cas / 20	20 cas / 20	75 Hz	147 Hz F(1) = 27.1 ***	281 Hz	3 cas / 20	6 cas / 20
[i]	- 12 Hz	88 Hz F(1) = 267.1 ***	177 Hz	15 cas / 20	12 cas / 20	64 Hz	111 Hz F(1) = 64.3 ***	157 Hz	3 cas / 20	6 cas / 20
[u]	- 11 Hz	84 Hz F(1) = 85.2 ***	189 Hz	10 cas / 20	10 cas / 20	62 Hz	95 Hz F(1) = 57.8 ***	136 Hz	3 cas / 20	6 cas / 20

	BD1					BD2				
	Différence interindividuelle moyenne			Différence interindividuelle supérieure à		Différence interindividuelle moyenne			Différence interindividuelle supérieure à	
	Min	Moyenne	Max	Variabilité intra-individuelle	Tiers de l'E.L.	Min	Moyenne	Max	Variabilité intra-individuelle	Tiers de l'E.L.
[a]	0 Hz	143 Hz Moitié de l'E.L. F(9) = 9.8, ***	300 Hz	60% des cas	64% des cas	6.3 Hz	65 Hz Moitié de l'E.L. F(2) = 1.8, ns (p=0.18)	98 Hz	2 cas / 3	2 cas / 3
[i]	2 Hz	86 Hz Equivalent à l'E.L. F(9) = 35.8, ***	177 Hz	56% des cas	62% des cas	4 Hz	28 Hz Quart de l'E.L. F(2) = 2.7, ns (p=0.09)	12 Hz	1 cas / 3	1 cas / 3
[u]	2 Hz	83 Hz Equivalent à l'E.L. F(9) = 14.4, ***	173 Hz	56% des cas	69% des cas	1 Hz	22 Hz Quart de l'E.L. F(2) = 1.6, ns (p=0.22)	33 Hz	Aucun cas	2 cas / 3

	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L.	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L.
[a]	- 66 Hz	7 Hz 1/39ème de l'E.L. F(1) = 0.1, NS (p=0.79)	46 Hz	0 cas / 10	0 cas / 10	0 cas / 10	31 Hz	103 Hz Moitié de l'E.L. F(1) = 1.8, NS (p=0.22)	206 Hz	1 cas / 10	1 cas / 10	2 cas / 10
[i]	-4 Hz	14 Hz 1/7ème de l'E.L. F(1) = 2.0, NS (p=0.16)	30 Hz	0 cas / 10	0 cas / 10	2 cas / 10	-53 Hz	-48 Hz Moitié de l'E.L. F(1) = 5.3, *	-41 Hz	0 cas / 10	2 cas / 10	3 cas / 10
[u]	-1 Hz	16 Hz 1/6ème de l'E.L. F(1) = 1.1, NS (p=0.31)	33 Hz	0 cas / 10	0 cas / 10	4 cas / 10	-47 Hz	-24 Hz 1/6ème de l'E.L. F(1) = 1.8, NS (p=0.21)	-8 Hz	0 cas / 10	1 cas / 10	1 cas / 10

		Locutrices de BD1				
		L1	L2	L6	L9	L10
[a]	Effet du bruit blanc	326 Hz F(1) = 20.9 ***	336 Hz F(1) = 112.2 ***	336 Hz F(1) = 30.2 ***	361 Hz F(1) = 66.1 ***	380 Hz F(1) = 59.6 ***
	Effet du bruit cocktail	334 Hz F(1) = 19.7 ***	341 Hz F(1) = 21.9 ***	351 Hz F(1) = 43.9 ***	407 Hz F(1) = 49.0 ***	314 Hz F(1) = 49.3 ***
	Effet du type de bruit	8 Hz F(1) = 0.006 NS (p=0.94)	5 Hz F(1) = 0.004 NS (p=0.95)	16 Hz F(1) = 0.05 NS (p=0.83)	53 Hz F(1) = 0.6 NS (p=0.47)	66 Hz F(1) = 0.8 NS (p=0.37)
[i]	Effet du bruit blanc	150 Hz F(1) = 240.2 ***	138 Hz F(1) = 112.3 ***	129 Hz F(1) = 114.7 ***	160 Hz F(1) = 47.5 ***	147 Hz F(1) = 138.7 ***
	Effet du bruit cocktail	163 Hz F(1) = 164.8 ***	163 Hz F(1) = 152.9 ***	136 Hz F(1) = 191.1 ***	175 Hz F(1) = 57.6 ***	177 Hz F(1) = 168.8 ***
	Effet du type de bruit	13 Hz F(1) = 0.6 NS (p=0.44)	26 Hz F(1) = 2.6 NS (p=0.12)	7 Hz F(1) = 0.2 NS (p=0.67)	18 Hz F(1) = 0.3 NS (p=0.57)	22 Hz F(1) = 1.5 NS (p=0.23)
[u]	Effet du bruit blanc	119 Hz F(1) = 56.4 ***	163 Hz F(1) = 87.4 ***	141 Hz F(1) = 32.9 ***	153 Hz F(1) = 25.0 **	150 Hz F(1) = 175.1 ***
	Effet du bruit cocktail	144 Hz F(1) = 11.0 *	189 Hz F(1) = 41.9 ***	153 Hz F(1) = 96.3 ***	186 Hz F(1) = 76.5 ***	176 Hz F(1) = 47.9 ***
	Effet du type de bruit	25 Hz F(1) = 0.2 NS (p=0.63)	27 Hz F(1) = 0.99 NS (p=0.34)	13 Hz F(1) = 0.2 NS (p=0.68)	33 Hz F(1) = 0.7 NS (p=0.44)	26 Hz F(1) = 0.7 NS (p=0.41)

		Locuteurs de BD1				
		L3	L4	L5	L7	L8
[a]	Effet du bruit blanc	146 Hz F(1) = 61.0 ***	133 Hz F(1) = 60.7 ***	202 Hz F(1) = 47.6 ***	107 Hz F(1) = 56.6 ***	106 Hz F(1) = 32.4 ***
	Effet du bruit cocktail	185 Hz F(1) = 109.4 ***	141 Hz F(1) = 74.2 ***	232 Hz F(1) = 64.3 ***	93 Hz F(1) = 53.2 ***	105 Hz F(1) = 26.8 ***
	Effet du type de bruit	39 Hz F(1) = 1.7 NS (p=0.21)	8 Hz F(1) = 0.2 NS (p=0.64)	30 Hz F(1) = 0.6 NS (p=0.44)	14 Hz F(1) = 0.9 NS (p=0.34)	0.5 Hz F(1) = 0.0004 NS (p=0.98)
[i]	Effet du bruit blanc	8 Hz F(1) = 0.7 NS (p=0.42)	12 Hz F(1) = 0.2 NS (p=0.63)	34 Hz F(1) = 6.5 *	38 Hz F(1) = 5.6 *	14 Hz F(1) = 0.8 NS (p=0.37)
	Effet du bruit cocktail	35 Hz F(1) = 4.6 *	8 Hz F(1) = 0.1 NS (p=0.73)	44 Hz F(1) = 10.8 **	34 Hz F(1) = 6.6 *	16 Hz F(1) = 1.2 NS (p=0.27)
	Effet du type de bruit	27 Hz F(1) = 2.2 NS (p=0.16)	20 Hz F(1) = 1.0 NS (p=0.32)	13 Hz F(1) = 0.6 NS (p=0.44)	0.5 Hz F(1) = 0.001 NS (p=0.98)	2 Hz F(1) = 0.01 NS (p=0.91)
[u]	Effet du bruit blanc	4 Hz F(1) = 0.1 NS (p=0.79)	6 Hz F(1) = 0.2 NS (p=0.63)	9 Hz F(1) = 0.2 NS (p=0.64)	25 Hz F(1) = 1.4 NS (p=0.27)	10 Hz F(1) = 0.5 NS (p=0.49)
	Effet du bruit cocktail	11 Hz F(1) = 1.2 NS (p=0.29)	22 Hz F(1) = 1.5 NS (p=0.25)	14 Hz F(1) = 0.7 NS (p=0.41)	24 Hz F(1) = 1.9 NS (p=0.19)	4 Hz F(1) = 0.1 NS (p=0.79)
	Effet du type de bruit	7 Hz F(1) = 0.1 NS (p=0.74)	16 Hz F(1) = 0.4 NS (p=0.54)	6 Hz F(1) = 0.04 NS (p=0.84)	7 Hz F(1) = 0.04 NS (p=0.84)	14 Hz F(1) = 0.5 NS (p=0.51)

		Locutrices de BD2		
		L2	L6	L11
[a]	Effet du bruit cocktail	154 Hz F(1) = 13.7 **	160 Hz F(1) = 14.0 **	281 Hz F(1) = 17.0 **
[i]	Effet du bruit cocktail	104 Hz F(1) = 17.8 ***	100 Hz F(1) = 20.0 ***	63 Hz F(1) = 17.6 ***
[u]	Effet du bruit cocktail	97 Hz F(1) = 12.4 **	97 Hz F(1) = 30.2 ***	65 Hz F(1) = 10.8 **

Fréquence centrale du deuxième formant vocalique (F2)

		Locutrices de BD1				
		L1	L2	L6	L9	L10
[a]	Effet du bruit blanc	41 Hz F(1) = 0.7 NS (p=0.42)	87 Hz F(1) = 3.8 NS (p=0.07)	39 Hz F(1) = 0.8 NS (p=0.38)	28 Hz F(1) = 0.6 NS (p=0.46)	89 Hz F(1) = 2.3478, NS (p=0.14)
	Effet du bruit cocktail	53 Hz F(1) = 1.1 NS (p=0.31)	113 Hz F(1) = 2.9 NS (p=0.10)	7 Hz F(1) = 0.02 NS (p=0.87)	55 Hz F(1) = 1.4 NS (p=0.25)	70 Hz F(1) = 1.2969, NS (p=0.27)
	Effet du type de bruit	12 Hz F(1) = 0.05 NS (p=0.83)	26 Hz F(1) = 0.1 NS (p=0.73)	31 Hz F(1) = 0.7 NS (p=0.41)	13 Hz F(1) = 0.1 NS (p=0.81)	19 Hz F(1) = 0.1 NS (p=0.74)
[i]	Effet du bruit blanc	80 Hz F(1) = 6.6 *	15 Hz F(1) = 0.2 NS (p=0.69)	1 Hz F(1) = 0.0004 NS (p=0.98)	3 Hz F(1) = 0.004 NS (p=0.95)	55 Hz F(1) = 2.8 NS (p=0.11)
	Effet du bruit cocktail	27 Hz F(1) = 0.5 NS (p=0.48)	14 Hz F(1) = 0.2 NS (p=0.64)	29 Hz F(1) = 0.7 NS (p=0.42)	33 Hz F(1) = 0.6 NS (p=0.44)	38 Hz F(1) = 1.2 NS (p=0.28)
	Effet du type de bruit	53 Hz F(1) = 1.6 NS (p=0.22)	34 Hz F(1) = 0.9 NS (p=0.35)	28 Hz F(1) = 0.6 NS (p=0.46)	19 Hz F(1) = 0.1 NS (p=0.74)	9 Hz F(1) = 0.04 NS (p=0.84)
[u]	Effet du bruit blanc	171 Hz F(1) = 1.1 NS (p=0.32)	190 Hz F(1) = 4.2 ns (p=0.07)	30 Hz F(1) = 0.2 NS (p=0.66)	100 Hz F(1) = 0.2 NS (p=0.64)	146 Hz F(1) = 2.9 NS (p=0.12)
	Effet du bruit cocktail	154 Hz F(1) = 0.8 NS (p=0.39)	152 Hz F(1) = 2.3 NS (p=0.16)	50 Hz F(1) = 0.6 NS (p=0.45)	146 Hz F(1) = 0.5 NS (p=0.50)	191 Hz F(1) = 4.0 NJS (p=0.07)
	Effet du type de bruit	17 Hz F(1) = 0.005 NS (p=0.94)	38 Hz F(1) = 0.1 NS (p=0.76)	21 Hz F(1) = 0.1 NS (p=0.79)	46 Hz F(1) = 0.02 NS (p=0.88)	45 Hz F(1) = 0.2 NS (p=0.63)

		Locuteurs de BD1				
		L3	L4	L5	L7	L8
[a]	Effet du bruit blanc	7 Hz F(1) = 0.02 NS (p=0.89)	102 Hz F(1) = 3.5 NS (p=0.08)	4 Hz F(1) = 0.01 NS (p=0.93)	14 Hz F(1) = 0.1 NS (p=0.79)	46 Hz F(1) = 0.7 NS (p=0.42)
	Effet du bruit cocktail	26 Hz F(1) = 0.3 NS (p=0.59)	75 Hz F(1) = 2.0 NS (p=0.18)	30 Hz F(1) = 0.5 NS (p=0.50)	11 Hz F(1) = 0.04 NS (p=0.85)	54 Hz F(1) = 0.8 NS (p=0.38)
	Effet du type de bruit	33 Hz F(1) = 2.2 NS (p=0.15)	28 Hz F(1) = 0.6 NS (p=0.46)	34 Hz F(1) = 1.2 NS (p=0.29)	25 Hz F(1) = 0.2 NS (p=0.66)	8 Hz F(1) = 0.02 NS (p=0.89)
[i]	Effet du bruit blanc	176 Hz F(1) = 9.4 **	86 Hz F(1) = 2.6 NS (p=0.12)	49 Hz F(1) = 1.3 NS (p=0.26)	14 Hz F(1) = 0.3 NS (p=0.59)	48 Hz F(1) = 1.1 NS (p=0.29)
	Effet du bruit cocktail	142 Hz F(1) = 5.9 *	112 Hz F(1) = 4.6 *	66 Hz F(1) = 3.0 NS (p=0.09)	21 Hz F(1) = 0.6 NS (p=0.46)	47 Hz F(1) = 1.3 NS (p=0.26)
	Effet du type de bruit	34 Hz F(1) = 0.2 NS (p=0.68)	26 Hz F(1) = 0.2 NS (p=0.69)	24 Hz F(1) = 0.2 NS (p=0.65)	5 Hz F(1) = 0.04 NS (p=0.85)	1 Hz F(1) = 0.0002 NS (p=0.99)
[u]	Effet du bruit blanc	4 Hz F(1) = 0.001 NS (p=0.97)	115 Hz F(1) = 2.1 NS (p=0.18)	24 Hz F(1) = 0.2 NS (p=0.68)	31 Hz F(1) = 0.2 NS (p=0.64)	33 Hz F(1) = 0.2 NS (p=0.66)
	Effet du bruit cocktail	37 Hz F(1) = 0.1 NS (p=0.71)	84 Hz F(1) = 1.1 NS (p=0.31)	21 Hz F(1) = 0.1 NS (p=0.73)	60 Hz F(1) = 1.0 NS (p=0.34)	35 Hz F(1) = 0.1 NS (p=0.72)
	Effet du type de bruit	33 Hz F(1) = 0.04 NS (p=0.84)	30 Hz F(1) = 0.1 NS (p=0.81)	45 Hz F(1) = 0.3 NS (p=0.59)	71 Hz F(1) = 0.6 NS (p=0.46)	68 Hz F(1) = 0.8 NS (p=0.39)

		Locutrices de BD2		
		L2	L6	L11
[a]	Effet du bruit cocktail	-18 Hz F(1) = 0.1 NS (p=0.82)	43 Hz F(1) = 0.4 NS (p=0.51)	47 Hz F(1) = 0.4 NS (p=0.55)
[i]	Effet du bruit cocktail	2 Hz F(1) = 0.004 NS (p=0.95)	0 Hz F(1) = 0.00001 NS (p=0.99)	2 Hz F(1) = 0.01 NS (p=0.93)
[u]	Effet du bruit cocktail	77 Hz F(1) = 0.5 NS (p=0.49)	145 Hz F(1) = 12.3 **	395 Hz F(1) = 32.6 ***

Fréquence centrale du troisième formant vocalique (F3)

		Locutrices de BD1				
		L1	L2	L6	L9	L10
[a]	Effet du bruit blanc	23 Hz F(1) = 0.2 NS (p=0.62)	163 Hz F(1) = 13.1 **	143 Hz F(1) = 14.5 **	213 Hz F(1) = 25.6 ***	115 Hz F(1) = 5.6 *
	Effet du bruit cocktail	69 Hz F(1) = 1.4 NS (p=0.25)	171 Hz F(1) = 16.2 ***	113 Hz F(1) = 7.0 *	70 Hz F(1) = 2.1 NS (p=0.17)	223 Hz F(1) = 17.8 ***
	Effet du type de bruit	46 Hz F(1) = 1.1 NS (p=0.30)	8 Hz F(1) = 0.03 NS (p=0.87)	29 Hz F(1) = 0.4 NS (p=0.52)	130 Hz F(1) = 3.2 NS (p=0.09)	108 Hz F(1) = 3.4 NS (p=0.08)
[i]	Effet du bruit blanc	82 Hz F(1) = 4.1 NS (p=0.05)	41 Hz F(1) = 0.7 NS (p=0.42)	90 Hz F(1) = 3.0 ns (p=0.09)	8 Hz F(1) = 0.02 NS (p=0.90)	122 Hz F(1) = 9.9 **
	Effet du bruit cocktail	77 Hz F(1) = 3.4 NS (p=0.07)	101 Hz F(1) = 5.6 *	37 Hz F(1) = 0.7 NS (p=0.42)	126 Hz F(1) = 3.9 NS (p=0.06)	84 Hz F(1) = 4.1 NS (p=0.05)
	Effet du type de bruit	5 Hz F(1) = 0.01 NS (p=0.92)	55 Hz F(1) = 1.4 NS (p=0.25)	53 Hz F(1) = 0.6 NS (p=0.43)	106 Hz F(1) = 1.2 NS (p=0.28)	41 Hz F(1) = 0.8 NS (p=0.37)
[u]	Effet du bruit blanc	142 Hz F(1) = 1.7 NS (p=0.24)	107 Hz F(1) = 1.8 NS (p=0.21)	49 Hz F(1) = 0.3 NS (p=0.61)	75 Hz F(1) = 0.8 NS (p=0.40)	245 Hz F(1) = 10.3 **
	Effet du bruit cocktail	228 Hz F(1) = 4.4 NS (p=0.08)	20 Hz F(1) = 0.1 NS (p=0.81)	90 Hz F(1) = 1.0 NS (p=0.33)	134 Hz F(1) = 2.6 NS (p=0.14)	316 Hz F(1) = 13.0 **
	Effet du type de bruit	86 Hz F(1) = 0.4 NS (p=0.57)	87 Hz F(1) = 0.8 NS (p=0.38)	40 Hz F(1) = 0.2 NS (p=0.69)	60 Hz F(1) = 0.4 NS (p=0.54)	71 Hz F(1) = 0.4 NS (p=0.52)

		Locuteurs de BD1				
		L3	L4	L5	L7	L8
[a]	Effet du bruit blanc	68 Hz F(1) = 4.1 NS (p=0.06)	74 Hz F(1) = 1.3 NS (p=0.26)	36 Hz F(1) = 1.0 NS (p=0.33)	88 Hz F(1) = 4.4 *	44 Hz F(1) = 2.5 NS (p=0.13)
	Effet du bruit cocktail	67 Hz F(1) = 3.8 NS (p=0.07)	85 Hz F(1) = 2.0 NS (p=0.17)	25 Hz F(1) = 0.5 NS (p=0.49)	62 Hz F(1) = 2.3 NS (p=0.14)	61 Hz F(1) = 3.2 NS (p=0.09)
	Effet du type de bruit	1 Hz F(1) = 0.0004 NS (p=0.98)	11 Hz F(1) = 0.03 NS (p=0.86)	12 Hz F(1) = 0.1 NS (p=0.77)	25 Hz F(1) = 0.5 NS (p=0.48)	17 Hz F(1) = 0.3 NS (p=0.58)
[i]	Effet du bruit blanc	184 Hz F(1) = 19.5 ***	32 Hz F(1) = 0.4 NS (p=0.52)	73 Hz F(1) = 3.5 NS (p=0.073)	91 Hz F(1) = 7.0 *	105 Hz F(1) = 3.1 NS (p=0.09)
	Effet du bruit cocktail	202 Hz F(1) = 17.3 ***	12 Hz F(1) = 0.08 NS (p=0.77)	92 Hz F(1) = 6.5 *	12 Hz F(1) = 0.2 NS (p=0.67)	127 Hz F(1) = 6.0 *
	Effet du type de bruit	18 Hz F(1) = 0.1 NS (p=0.80)	44 Hz F(1) = 0.7 NS (p=0.42)	15 Hz F(1) = 0.1 NS (p=0.79)	69 Hz F(1) = 2.6 NS (p=0.12)	22 Hz F(1) = 0.1 NS (p=0.75)
[u]	Effet du bruit blanc	124 Hz F(1) = 1.7 NS (p=0.23)	110 Hz F(1) = 2.7 NS (p=0.13)	209 Hz F(1) = 10.7 **	88 Hz F(1) = 9.2 *	96 Hz F(1) = 3.5 ns (p=0.09)
	Effet du bruit cocktail	110 Hz F(1) = 2.0 NS (p=0.19)	113 Hz F(1) = 3.0 NS (p=0.11)	251 Hz F(1) = 12.3 **	104. Hz F(1) = 5.7 *	73 Hz F(1) = 1.3 NS (p=0.29)
	Effet du type de bruit	13 Hz F(1) = 0.01 NS (p=0.92)	3 Hz F(1) = 0.001 NS (p=0.97)	43 Hz F(1) = 0.2 NS (p=0.66)	16 Hz F(1) = 0.1 NS (p=0.73)	23 Hz F(1) = 0.1 NS (p=0.80)

		Locutrices de BD2		
		L2	L6	L11
[a]	Effet du bruit cocktail	64 Hz F(1) = 1.9 NS (p=0.19)	8 Hz F(1) = 0.05 NS (p=0.82)	177 Hz F(1) = 12.9 **
[i]	Effet du bruit cocktail	-150 Hz F(1) = 2.4 NS (p=0.15)	44 Hz F(1) = 0.4 NS (p=0.49)	59 Hz F(1) = 0.5 NS (p=0.47)
[u]	Effet du bruit cocktail	16 Hz F(1) = 0.06 NS (p=0.81)	119 Hz F(1) = 3.6 NS (p=0.08)	-67 Hz F(1) = 0.4 NS (p=0.52)

Quotient ouvert laryngé

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	-0.11 (soit -17.9%)	-0.03 (soit -4.2%) F(1) = 23.0, ***	0.15 (soit 26.4%)	15 cas / 20	2 cas / 20
BD2	-0.10 (soit -12.3 %)	-0.04 (soit -4.5%) F(1) = 195.5, ***	0.05 (soit 7.2%)	Tous les cas	2 cas / 6

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	0	0.06 équivalent de l E.L F(9) = 28.0, ***	0.23	33 %	29 %	71 %
BD2	0.01	0.08 équivalent à l'E.L F(2) = 690.4, ***	0.12	Tous les cas	2 cas / 3	2 cas / 3

	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle	Tiers de l'E.L
BD1	-0.04	0 F(1) = 0.004, ns (p=0.95)	0.02	Aucun cas	Aucun cas	3 cas / 10
BD2	-0.01	0.01 1/6 ^{ème} de l E.L F(1) = 2.6, ns (p=0.11)	0.04	1 cas / 3	1 cas / 3	1 cas / 3

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	-0.004 F(1) = 0.1 NS (p=0.71)	0.15 F(1) = 189.2 ***	-0.04 F(1) = 21.7 ***	-0.04 F(1) = 23.3 ***	-0.01 F(1) = 1.2 NS (p=0.29)
Effet du bruit cocktail	0.01 F(1) = 2.3 NS (p=0.13)	0.15 F(1) = 270.1 ***	-0.04 F(1) = 11.5 **	-0.08 F(1) = 19.9 ***	-0.02 F(1) = 0.4 NS (p=0.54)
Effet du type de bruit	0.02 F(1) = 0.9 NS (p=0.35)	-0.001 F(1) = 0.01 NS (p=0.92)	0.003 F(1) = 0.04 NS (p=0.85)	-0.04 F(1) = 3.2 NS 0.09)	-0.003 F(1) = 0.5 NS (p=0.48)

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	-0.11 F(1) = 78.0 ***	-0.02 F(1) = 4.5 *	-0.05 F(1) = 15.0 ***	-0.05 F(1) = 43.8 ***	-0.05 F(1) = 48.6 ***
Effet du bruit cocktail	-0.11 F(1) = 92.9 ***	-0.01 F(1) = 0.4 NS (p=0.50)	-0.06 F(1) = 22.9 ***	-0.05 F(1) = 36.3 ***	-0.06 F(1) = 8.7 **
Effet du type de bruit	-0.005 F(1) = 0.1 NS (p=0.78)	0.02 F(1) = 1.0 NS (p=0.32)	-0.01 F(1) = 0.5 NS (p=0.48)	0.002 F(1) = 0.02 NS (p=0.88)	-0.005 F(1) = 0.05 NS (p=0.82)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	0.05 F(1) = 159.8 ***	-0.10 F(1) = 959.1 ***	-0.09 F(1) = 266.6 ***
Effet du bruit cocktail	0.04 F(1) = 134.2 ***	-0.05 F(1) = 396.9 ***	-0.08 F(1) = 501.3 ***
Effet du type de bruit	-0.01 F(1) = 0.6 NS (p=0.44)	0.04 F(1) = 185.3 ***	0.01 F(1) = 1.1 NS (p=0.30)

Vitesse de fermeture laryngée

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L.)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	0.2 u.a.	2.0 u.a. F(1) = 540.1 ***	5.1 u.a.	19 cas / 20	19 cas / 20
BD2	0.008 u.a.	0.014 u.a. F(1) = 386.1 ***	0.023	6 cas / 20	6 cas / 20

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	1.21 u.a. F(1) = 54.4 ***	0.76 u.a. F(1) = 40.2 ***	1.29 u.a. F(1) = 39.8 ***	0.97 u.a. F(1) = 33.2 ***	1.40 u.a. F(1) = 81.8 ***
Effet du bruit cocktail	0.91 u.a. F(1) = 31.5 ***	0.94 u.a. F(1) = 50.9 ***	1.25 u.a. F(1) = 47.3 ***	0.23 u.a. F(1) = 1.6 NS (p=0.22)	0.78 u.a. F(1) = 8.5 **
Effet du type de bruit	-0.30 u.a. F(1) = 2.3 NS (p=0.13)	0.18 u.a. F(1) = 1.5 NS (p=0.22)	-0.04 u.a. F(1) = 0.03 NS (p=0.86)	-0.74 u.a. F(1) = 20.7 ***	-0.62 u.a. F(1) = 2.1 NS (p=0.16)

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	2.20 u.a. F(1) = 50.1 ***	2.99 u.a. F(1) = 131.6 ***	3.80 u.a. F(1) = 99.5 ***	2.12 u.a. F(1) = 34.0 ***	3.11 u.a. F(1) = 44.6 ***
Effet du bruit cocktail	2.00 u.a. F(1) = 67.0 ***	3.76 u.a. F(1) = 244.2 ***	5.10 u.a. F(1) = 267.9 ***	1.53 u.a. F(1) = 24.8 ***	3.36 u.a. F(1) = 51.4 ***
Effet du type de bruit	-0.20 u.a. F(1) = 0.37 NS (p=0.54)	0.77 u.a. F(1) = 7.4 **	1.30 u.a. F(1) = 7.7 **	-0.59 u.a. F(1) = 2.4 NS (p=0.13)	0.25 u.a. F(1) = 0.27 NS (p=0.60)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	0.0114 u.a. F(1) = 64.4 ***	0.0075 u.a. F(1) = 112.0 ***	0.0234 u.a. F(1) = 247.0 ***
Effet du bruit cocktail	0.0139 u.a. F(1) = 118.5 ***	0.0092 u.a. F(1) = 295.3 ***	0.0214 u.a. F(1) = 218.1 ***
Effet du type de bruit	0.0025 u.a. F(1) = 12.3 ***	0.0017 u.a. F(1) = 2.6 NS (p=0.11)	-0.0019 u.a. F(1) = 11.7 ***

Différence d'amplitude entre les deux premiers harmoniques de l'ODGD (H1-H2)

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD1	-22.3 dB	-9.0 dB F(1) = 329.3 ***	-1.1 dB	18 cas / 20	0 cas / 20
BD2	-5.9 dB	-3.7 dB F(1) = 171.8 ***	0.7 dB	5 cas / 20	0 cas / 20

	Locutrices de BD1				
	L1	L2	L6	L9	L10
Effet du bruit blanc	-8.0 dB F(1) = 8.6 **	-5.4 dB F(1) = 16.0 ***	-21.0 dB F(1) = 56.3 ***	-16.0 dB F(1) = 65.8 ***	-22.3 dB F(1) = 62.9 ***
Effet du bruit cocktail	-8.0 dB F(1) = 31.1 ***	-7.1 dB F(1) = 54.2 ***	-5.0 dB F(1) = 9.9 **	-8.4 dB F(1) = 15.2 ***	-1.8 dB F(1) = 0.1 NS (p=0.76)
Effet du type de bruit	-0.01 dB F(1) = 0.01 NS (p=0.91)	-1.7 dB F(1) = 1.7 NS (p=0.20)	16.0 dB F(1) = 39.7 ***	7.6 dB F(1) = 7.2 *	20.5 dB F(1) = 7.2 *

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
Effet du bruit blanc	-7.4 dB F(1) = 44.3 ***	-15.2 dB F(1) = 51.5 ***	-8.5 dB F(1) = 45.1 ***	-5.8 dB F(1) = 14.9 ***	-7.7 dB F(1) = 30.3 ***
Effet du bruit cocktail	-8.4 dB F(1) = 55.3 ***	-10.5 dB F(1) = 32.3 ***	-6.6 dB F(1) = 38.8 ***	-1.1 dB F(1) = 1.3 NS (p=0.26)	-4.89 F(1) = 6.3 *
Effet du type de bruit	-0.9 dB F(1) = 0.5 NS (p=0.48)	4.7 dB F(1) = 3.3 NS (p=0.07)	2.0 dB F(1) = 1.9 NS (p=0.17)	4.8 dB F(1) = 11.3 **	2.8 dB F(1) = 2.3 NS (p=0.14)

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	0.7 dB F(1) = 0.5 NS (p=0.46)	-5.8 dB F(1) = 55.1 ***	-5.4 dB F(1) = 62.5 ***
Effet du bruit cocktail	-1.7 dB F(1) = 10.7 **	-4.2 dB F(1) = 67.9 ***	-5.6 dB F(1) = 96.2 ***
Effet du type de bruit	-2.4 dB F(1) = 9.2 **	1.6 dB F(1) = 3.0 NS (p=0.08)	-0.2 dB F(1) = 1.0 NS (p=0.33)

Amplitude maximale des mouvements verticaux du larynx

	Effet de l'exposition au bruit ⇔ Effet Lombard (E.L)			Effet de l'exposition au bruit supérieur à	
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra-individuelle dans le silence et dans le bruit
BD2	-3.1%	16.9 % F(1) = 239.8, ***	41.3%	5 cas / 6	3 cas / 6

	Différence entre deux locuteurs			Différence entre deux locuteur supérieure à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra- individuelle	Tiers de l'E.L
BD2	12.7 %	23.2% équivalent de l E.L F(2) = 136.3, ***	34.7%	Tous les cas	2 cas / 3	Tous les cas

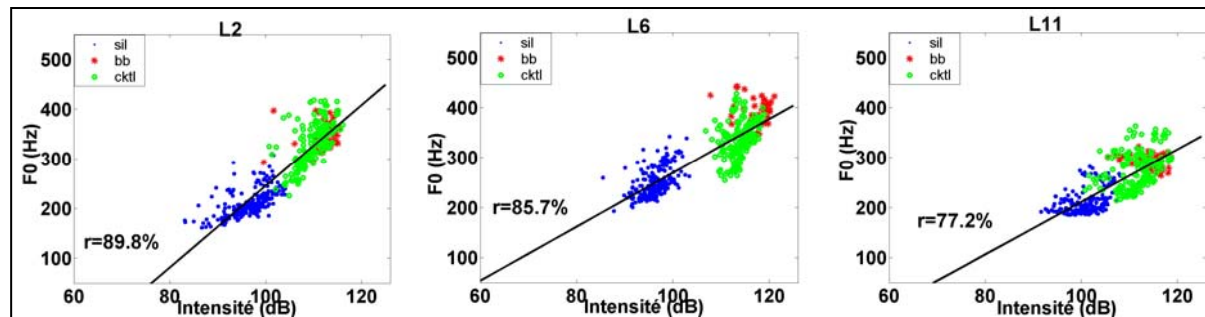
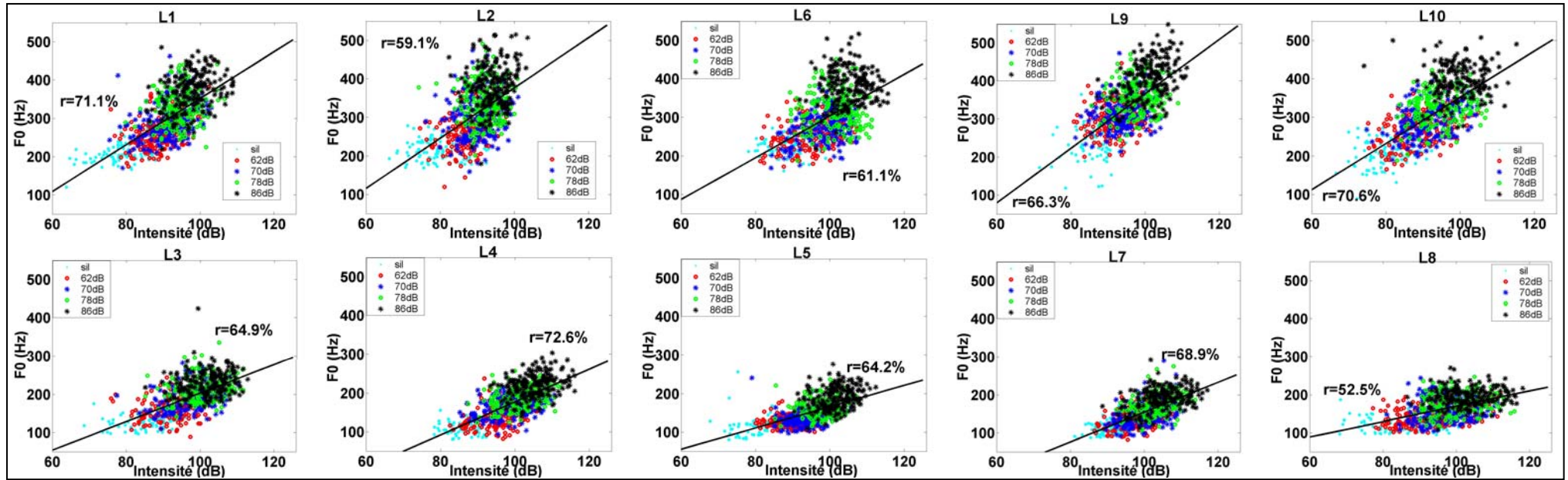
	Effet du type de bruit			Effet du type de bruit supérieur à		
	Min	Moyenne	Max	Seuil de significativité	Variabilité intra- individuelle	Tiers de l'E.L
BD2	6.0%	8.8% Moitié de l E.L F(1) = 17.7, ***	12.0%	Tous les cas	Aucun cas	Tous les cas

	Locutrices de BD2		
	L2	L6	L11
Effet du bruit blanc	-3.1% F(1) = 0.7 NS (p=0.40)	29.3% F(1) = 188.3 ***	11.3% F(1) = 161.8 ***
Effet du bruit cocktail	5.4% F(1) = 4.9 *	41.3% F(1) = 437.3 ***	17.3% F(1) = 59.9 ***
Effet du type de bruit	8.5% F(1) = 6.9 *	12.0% F(1) = 5.8 *	6.0% F(1) = 47.7 ***

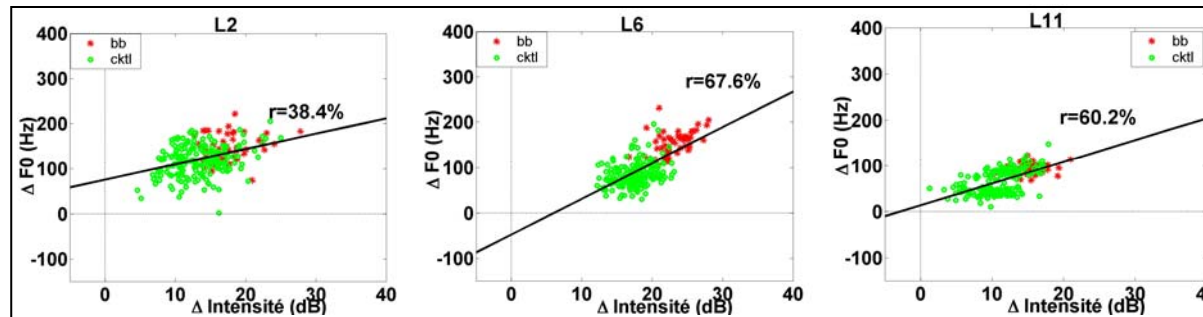
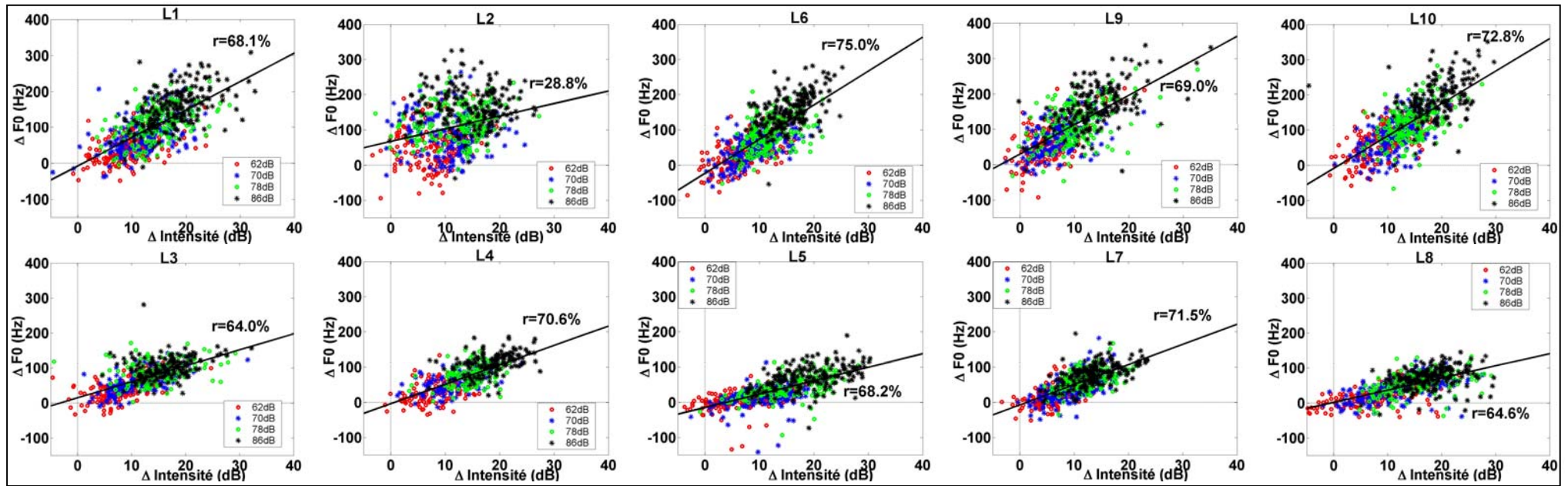
Annexe Res6 :
Résultats statistiques
détaillés du chapitre 6

Corrélation entre l'intensité vocale et la fréquence fondamentale

Entre leurs valeurs dans le silence et dans le bruit



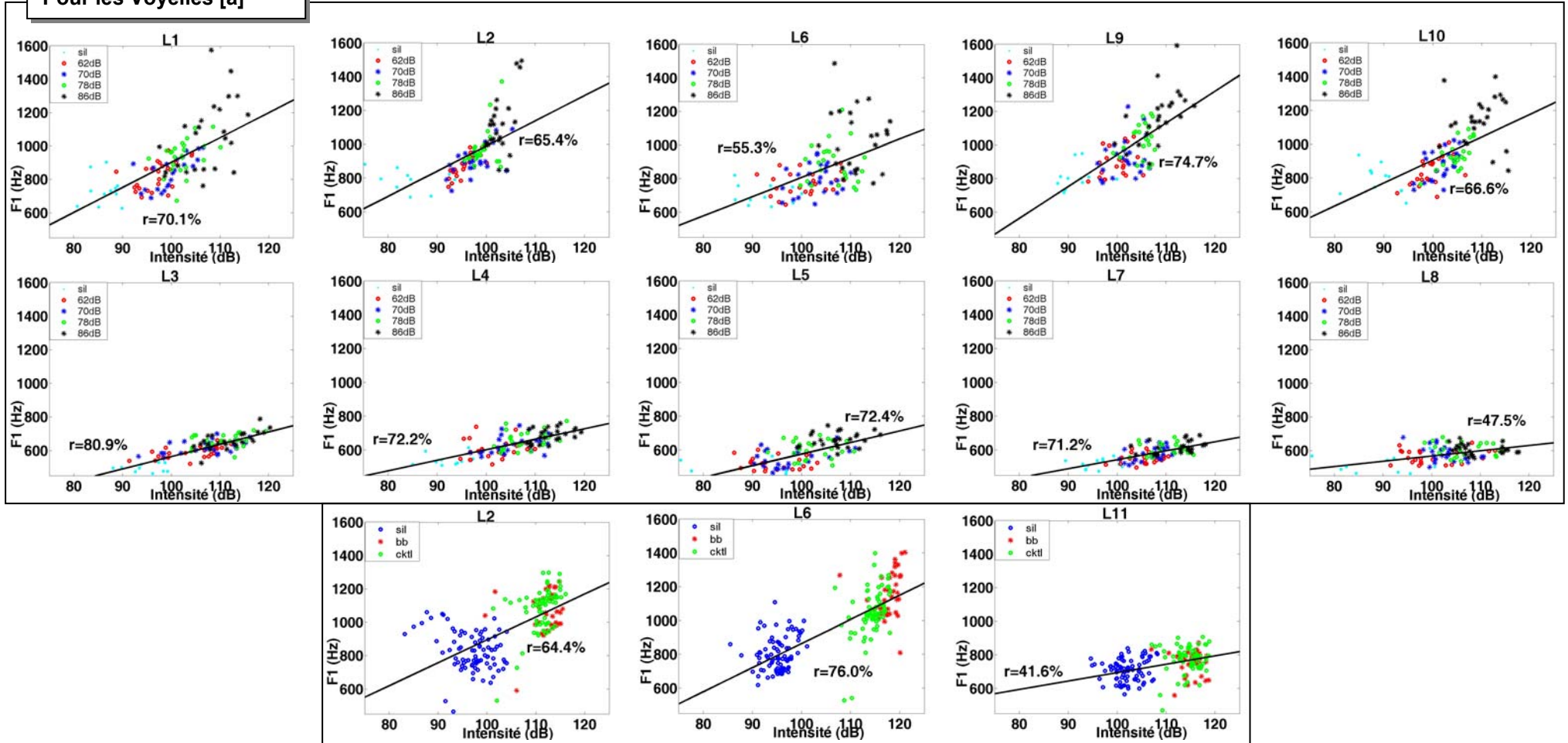
Entre leurs évolutions du silence au bruit



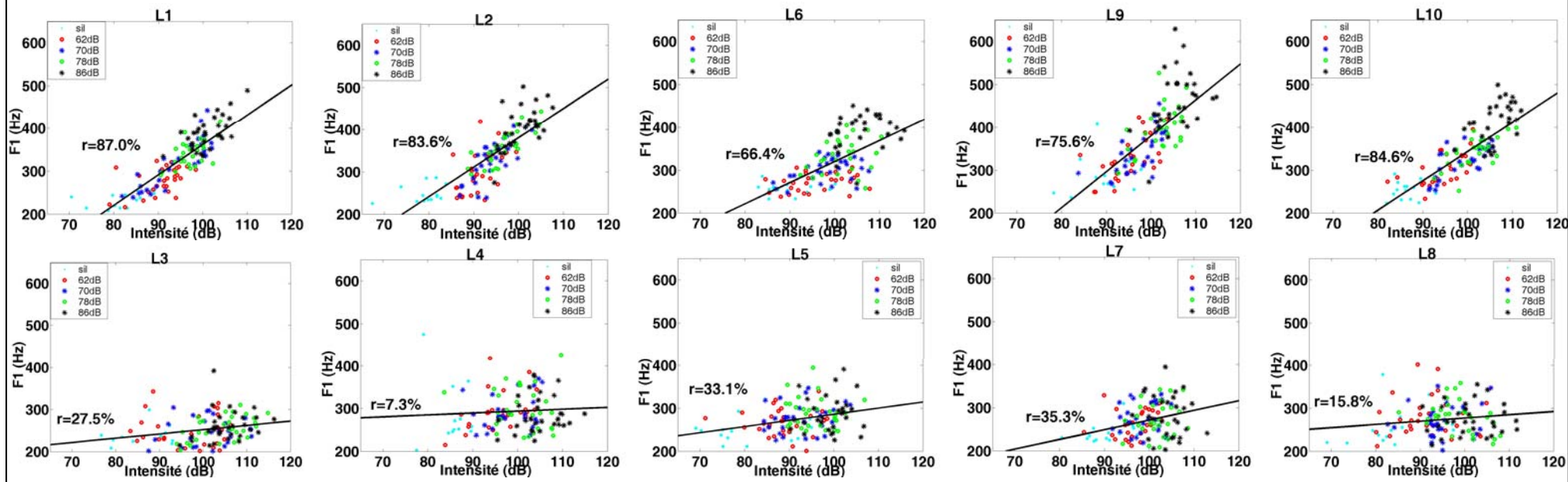
Corrélation entre l'intensité vocale et le premier formant vocalique

Entre leurs valeurs dans le silence et dans le bruit

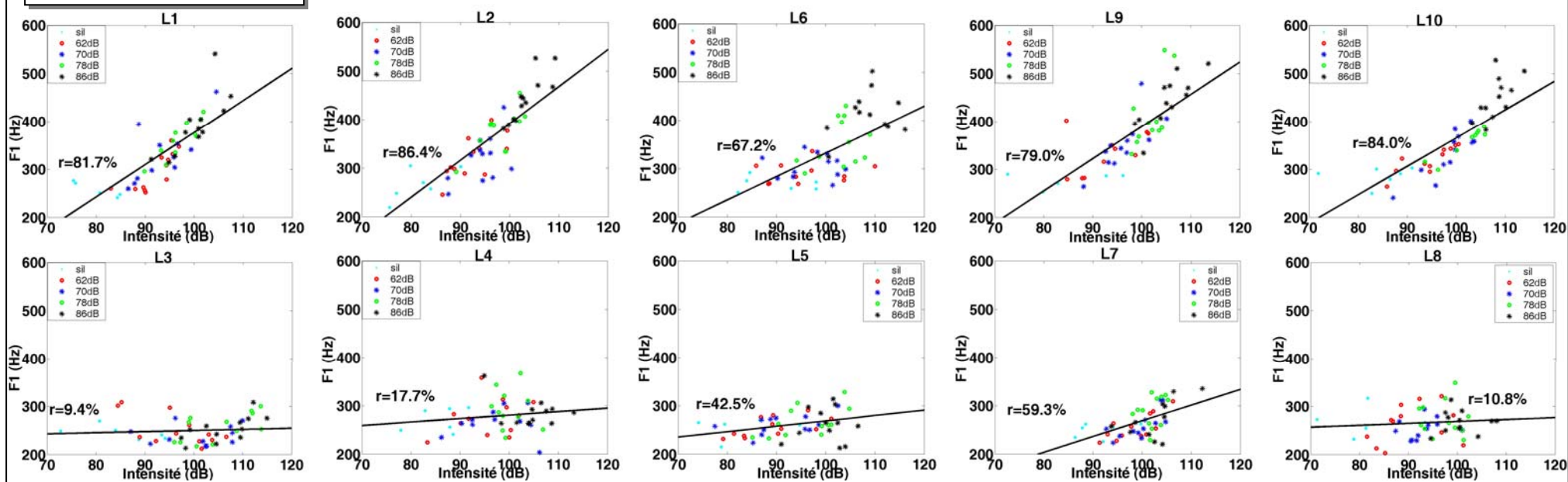
Pour les Voyelles [a]



Pour les Vovelles [i]

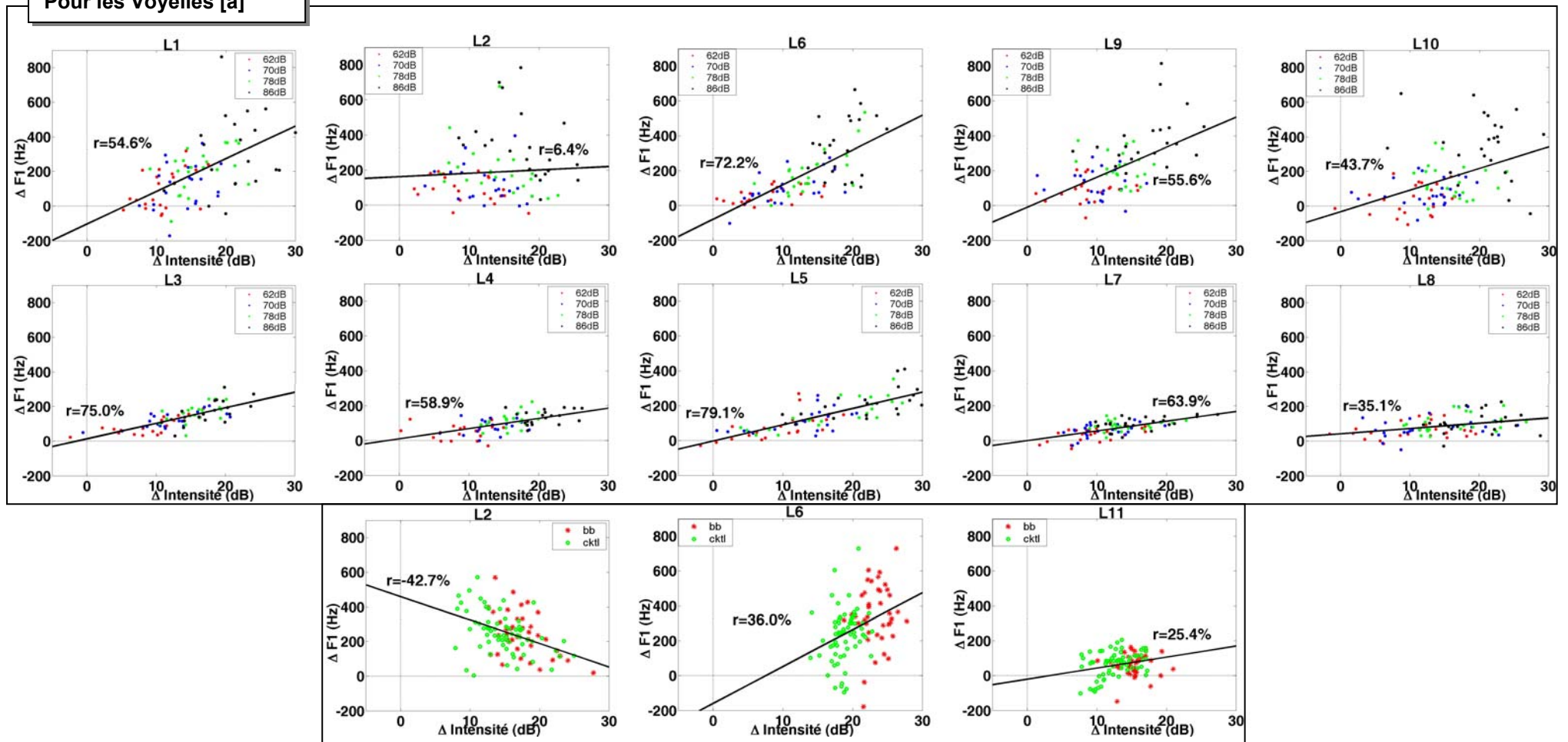


Pour les Vovelles [u]

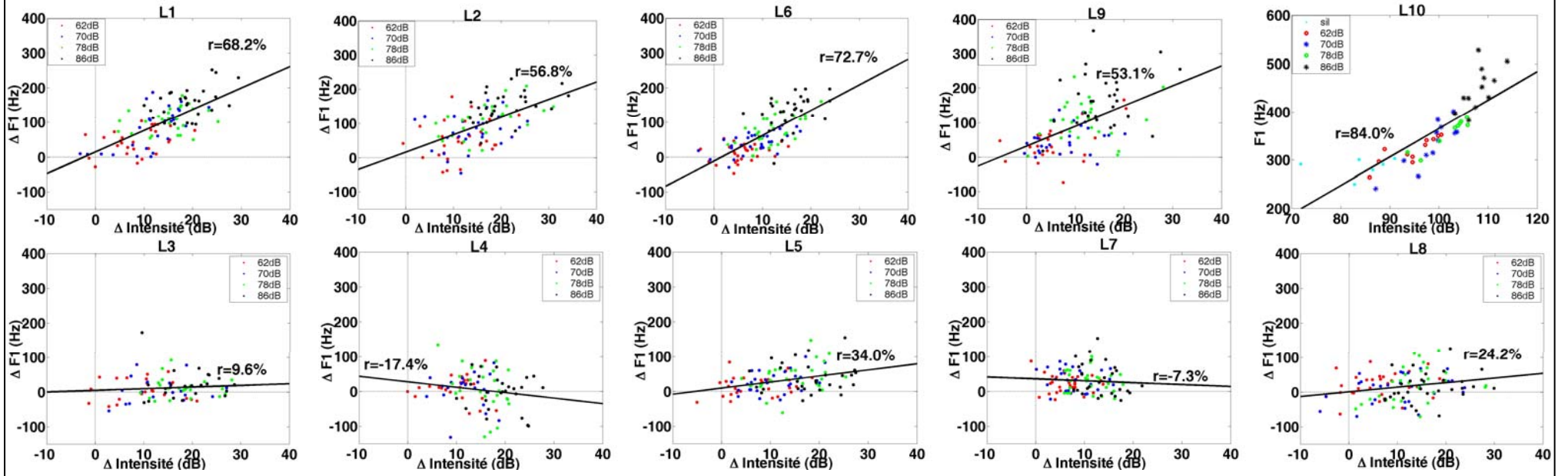


Entre leurs évolutions du silence au bruit

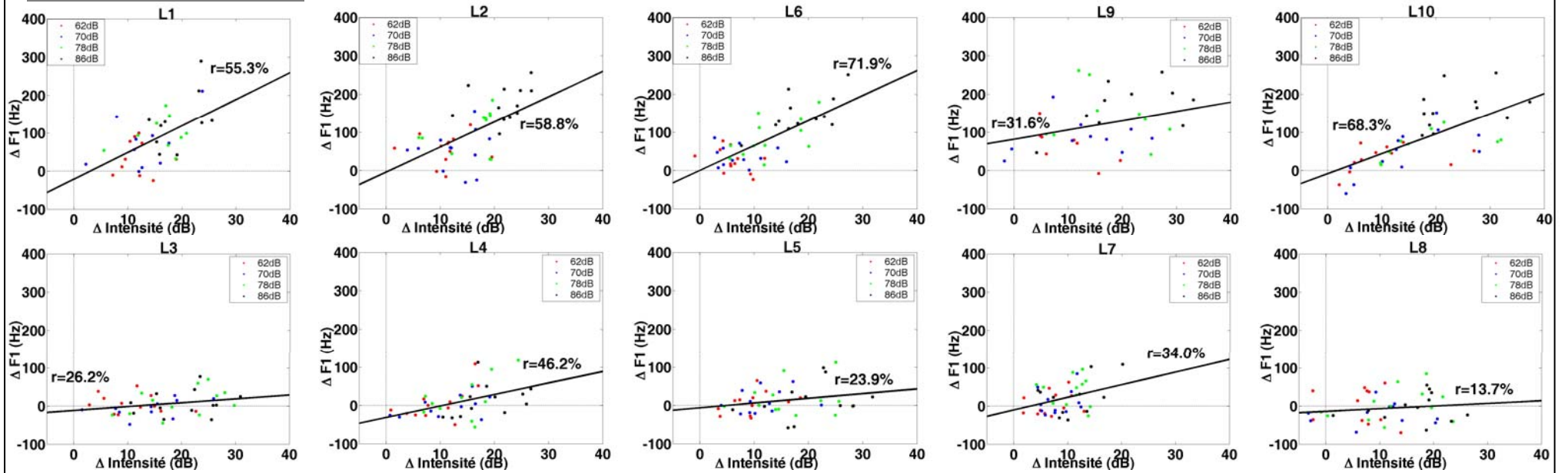
Pour les Voyelles [a]



Pour les Vowelles [i]

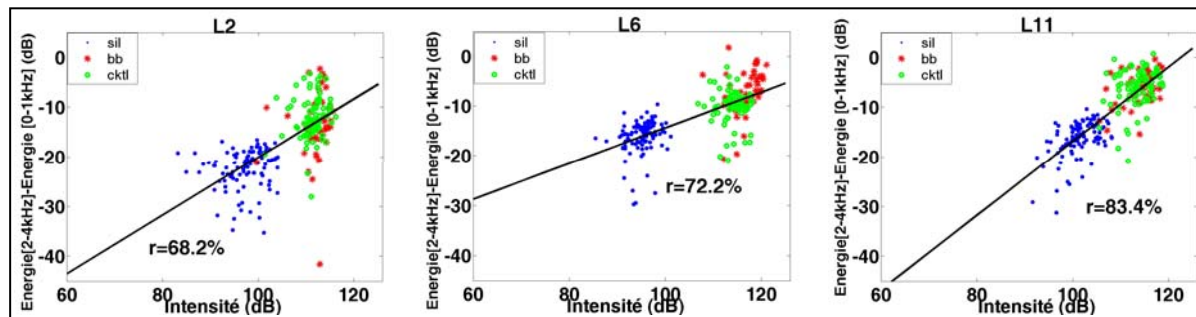
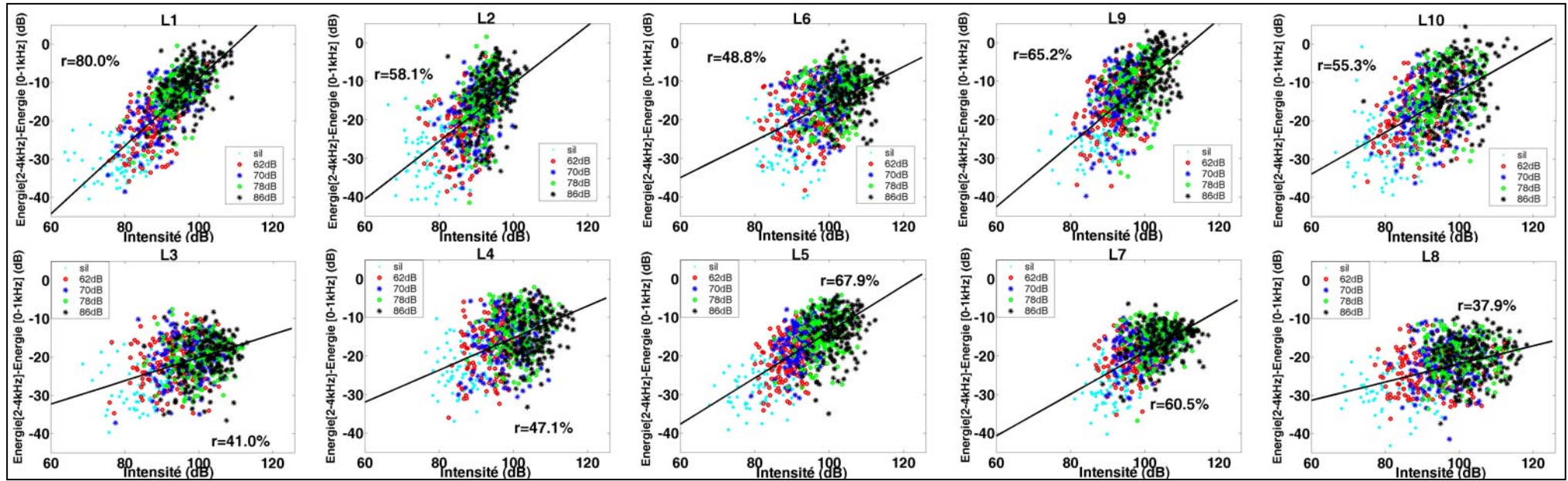


Pour les Vowelles [u]

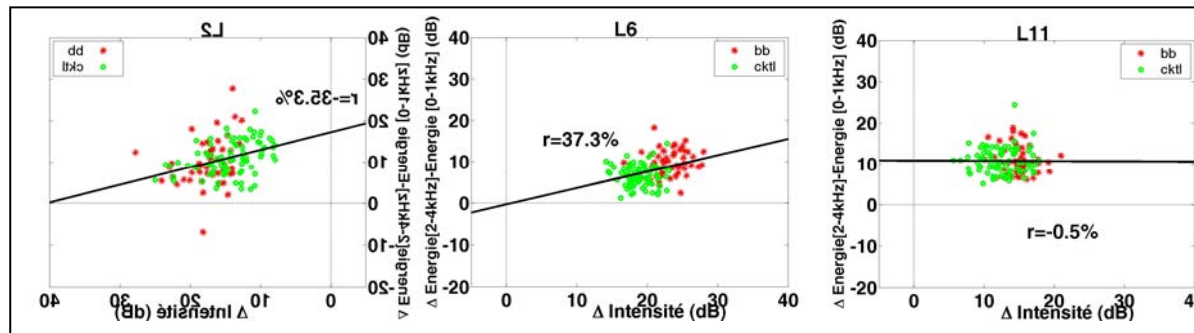
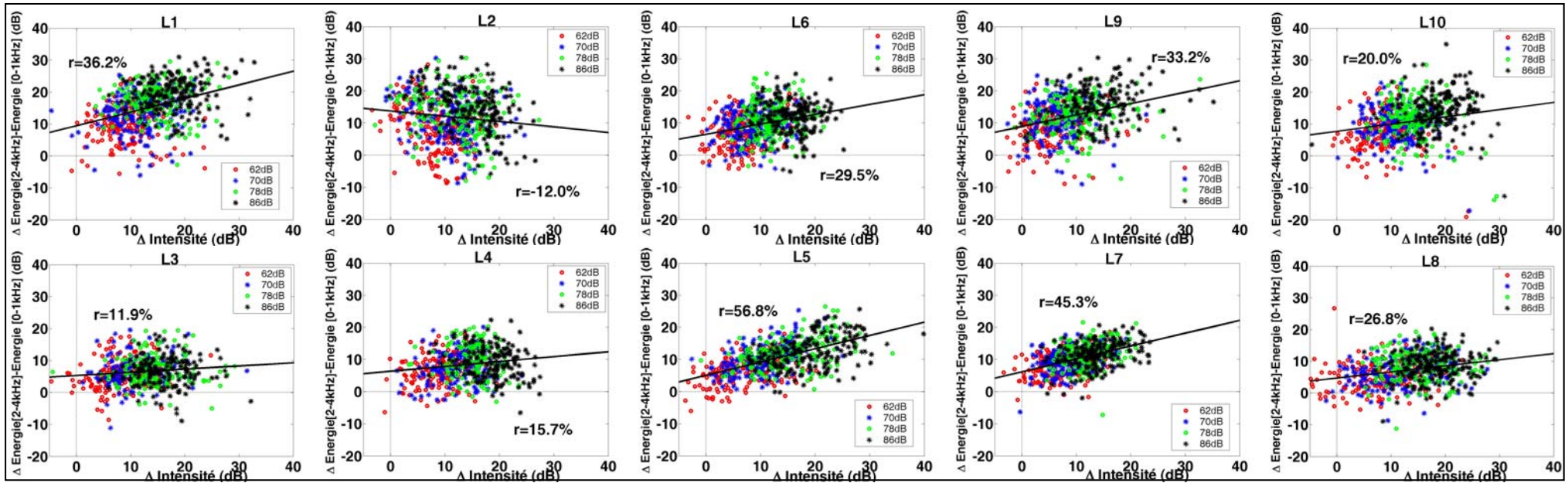


Corrélation entre l'intensité vocale et la différence d'énergie entre la bande [2-4kHz] et la bande [0-1kHz]

Entre leurs valeurs dans le silence et dans le bruit



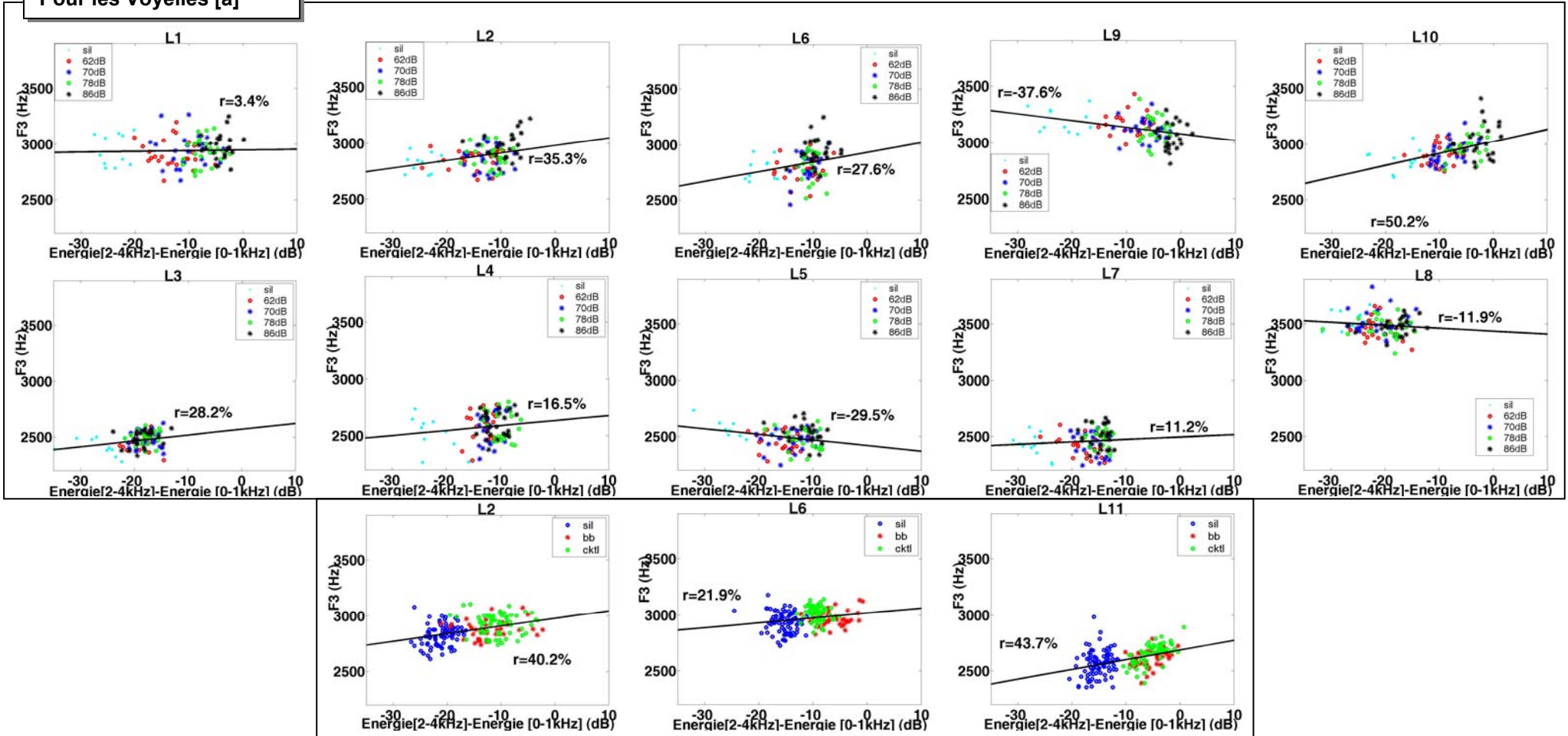
Entre leurs évolutions du silence au bruit



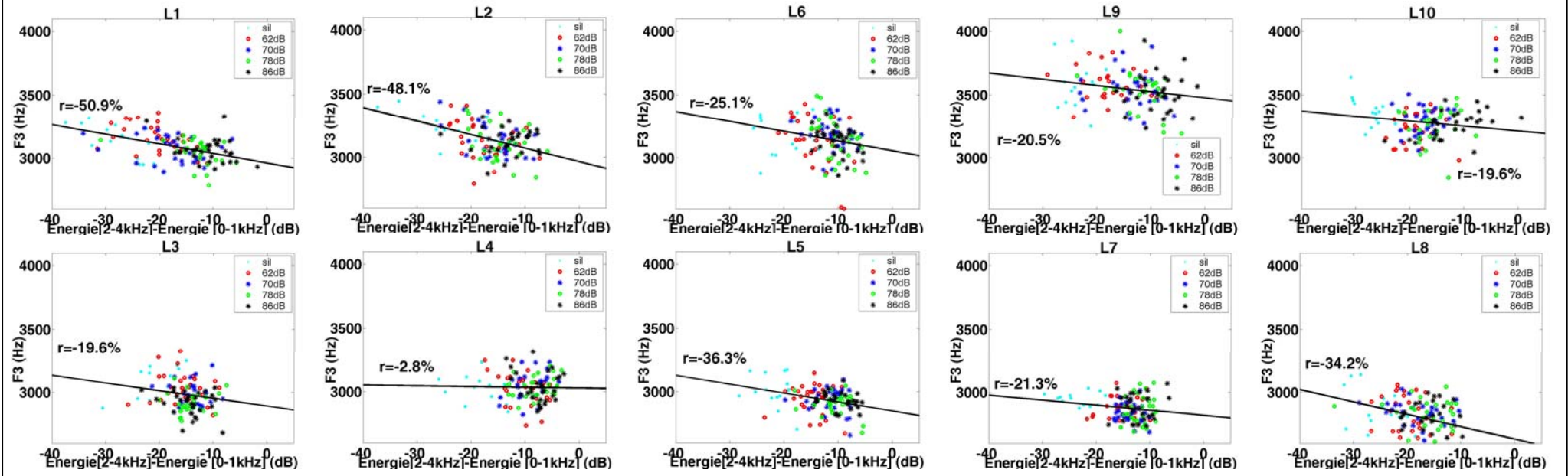
Corrélation entre le 3^{ème} formant vocalique et le timbrage de la voix dans les médiums

Entre leurs valeurs dans le silence et dans le bruit

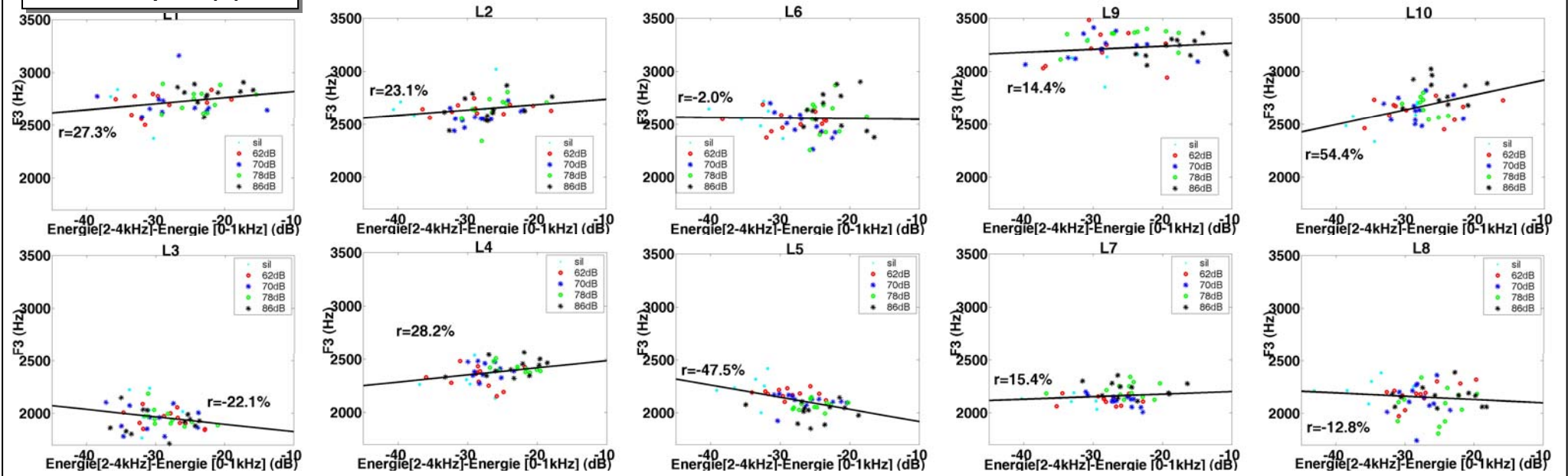
Pour les Voyelles [a]



Pour les Vowelles [i]

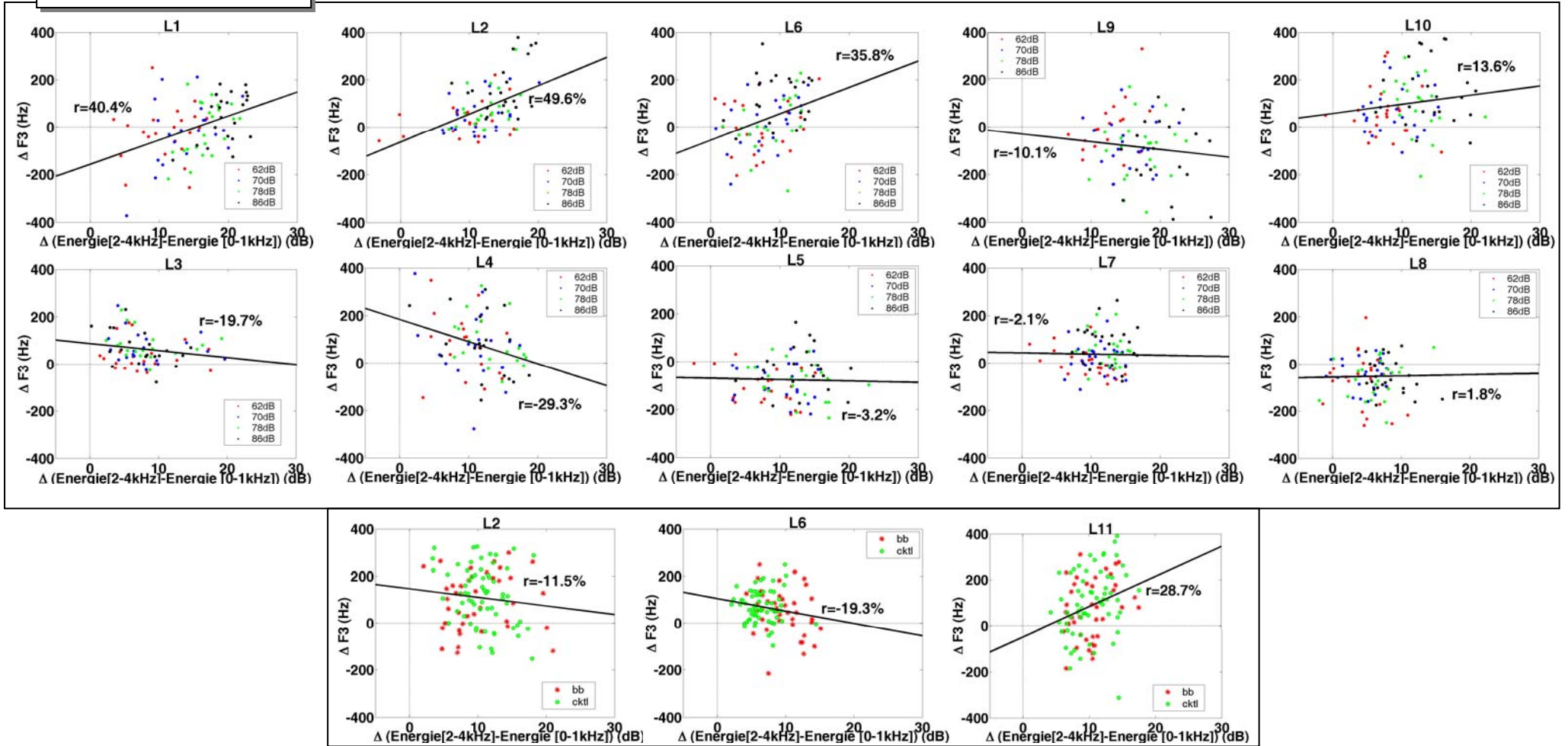


Pour les Vowelles [u]

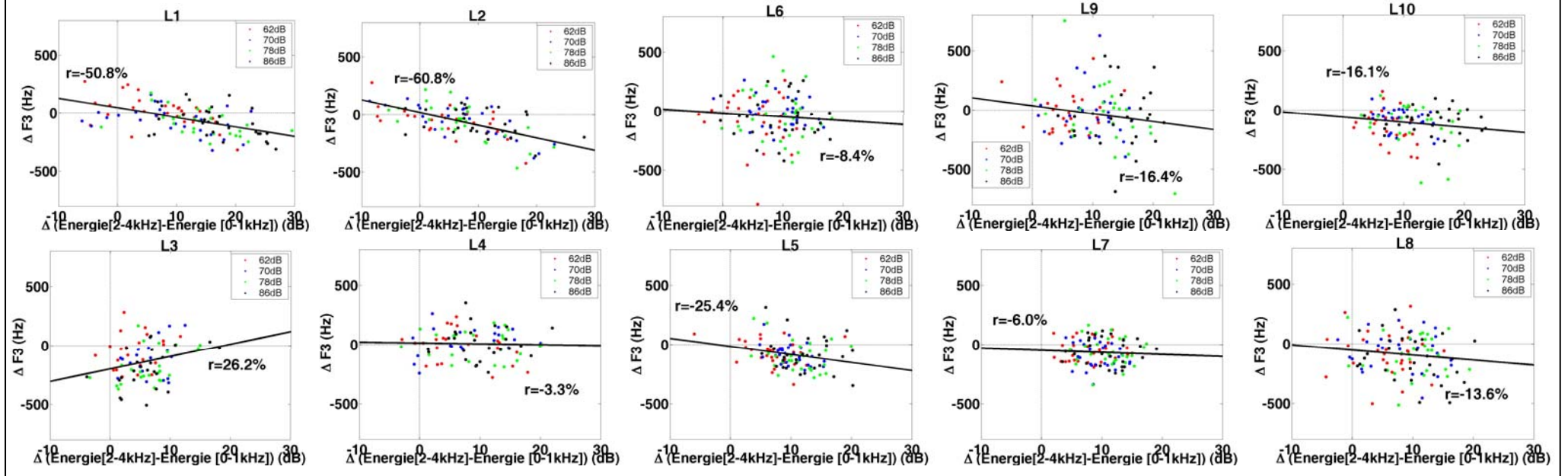


Entre leurs évolutions du silence au bruit

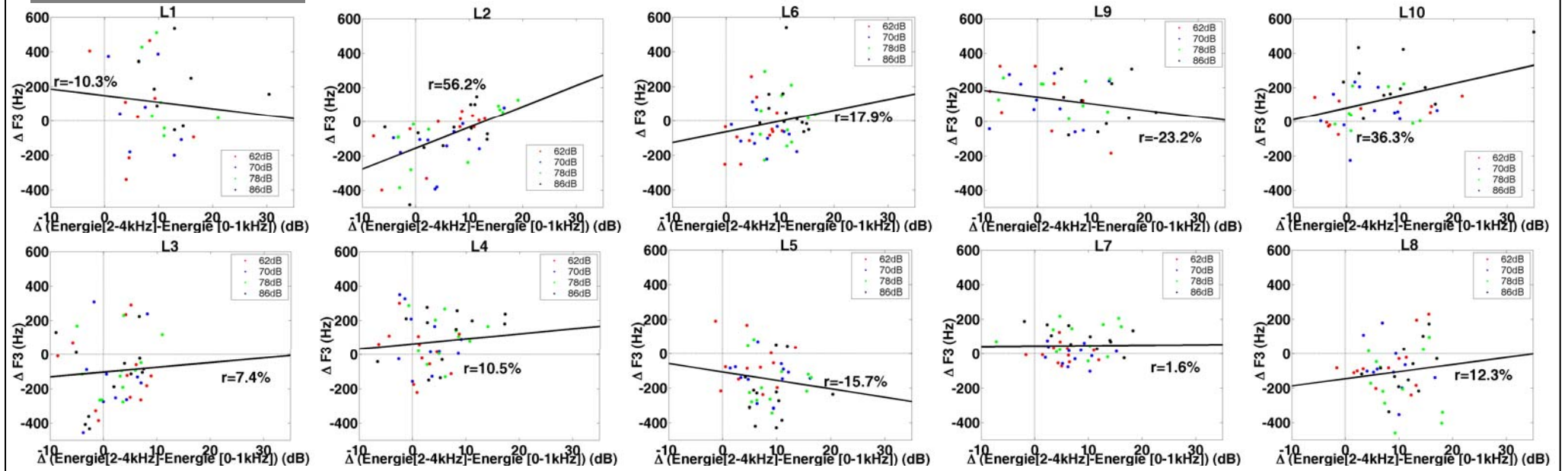
Pour les Voyelles [a]



Pour les Vowelles iï

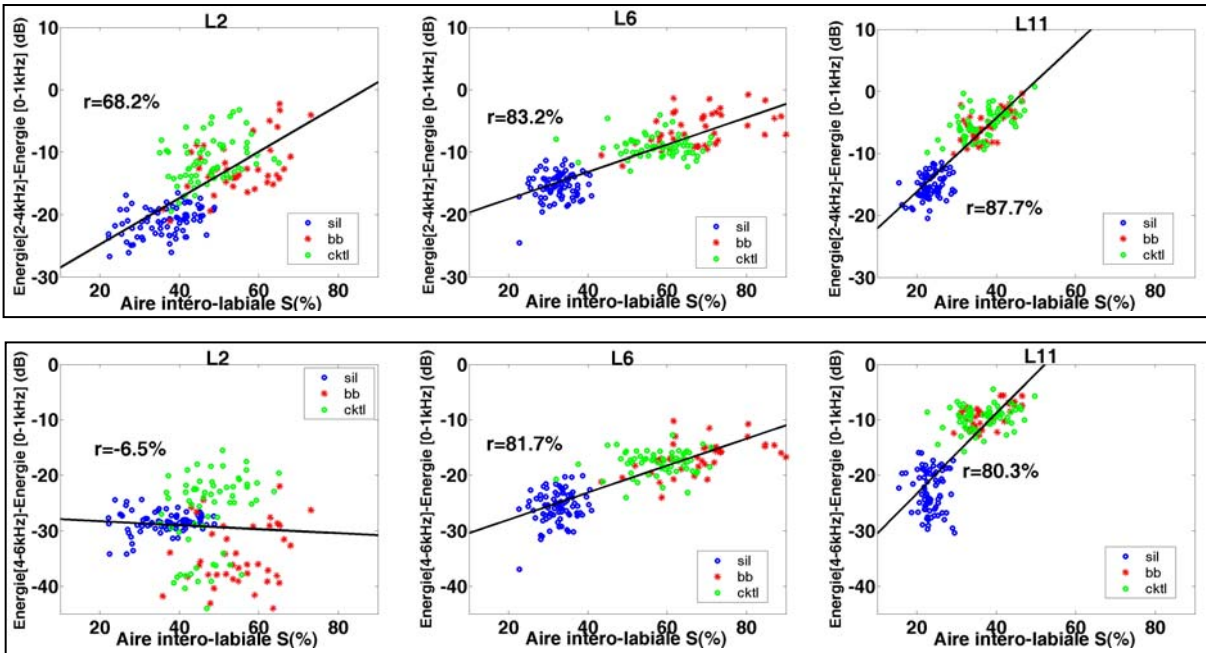


Pour les Vowelles uï

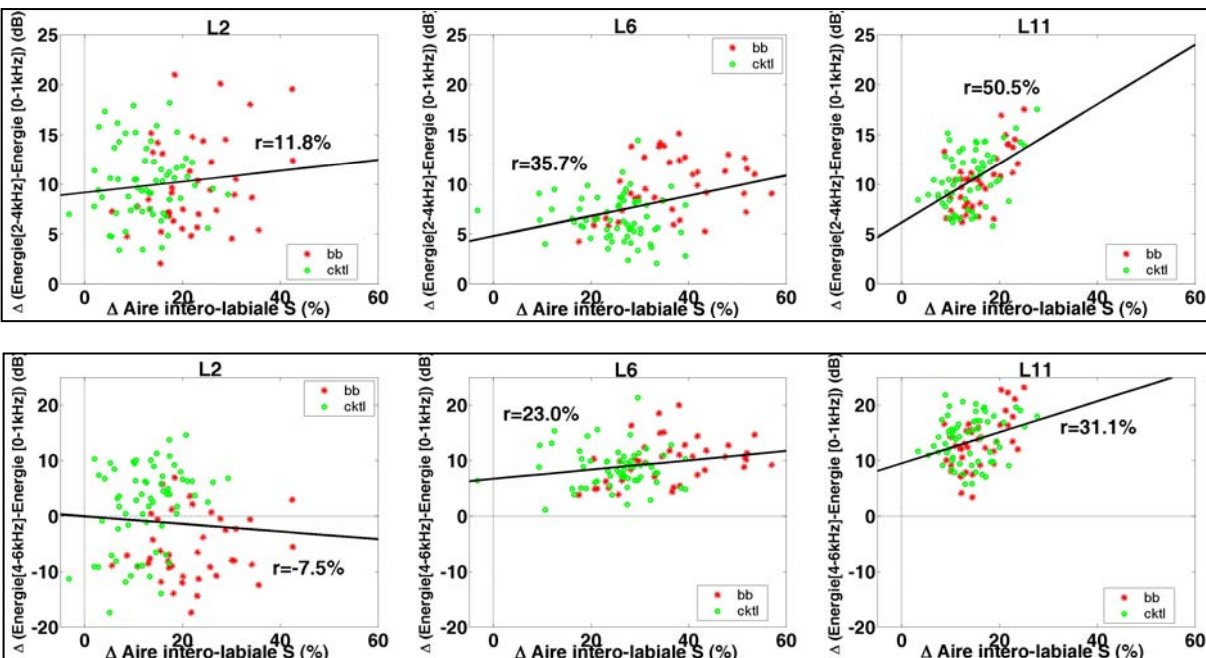


Corrélation entre l'aire intéro-labiale et la répartition spectrale de l'énergie

Entre leurs dans le silence et dans le bruit

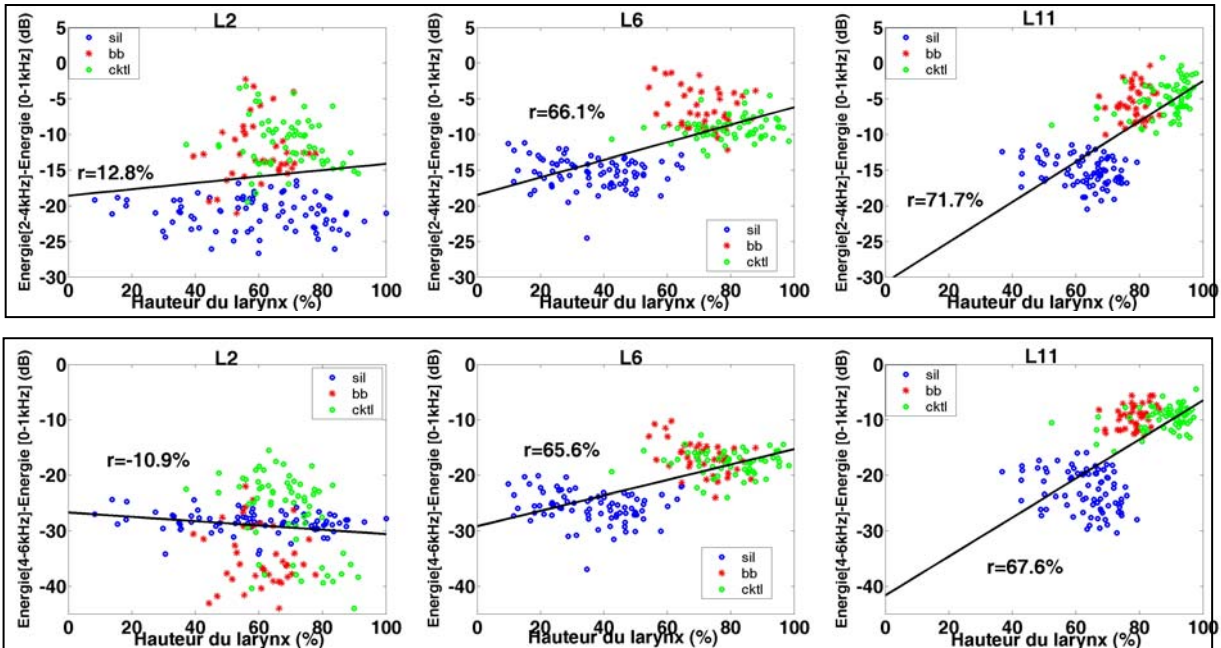


Entre leurs évolutions du silence au bruit

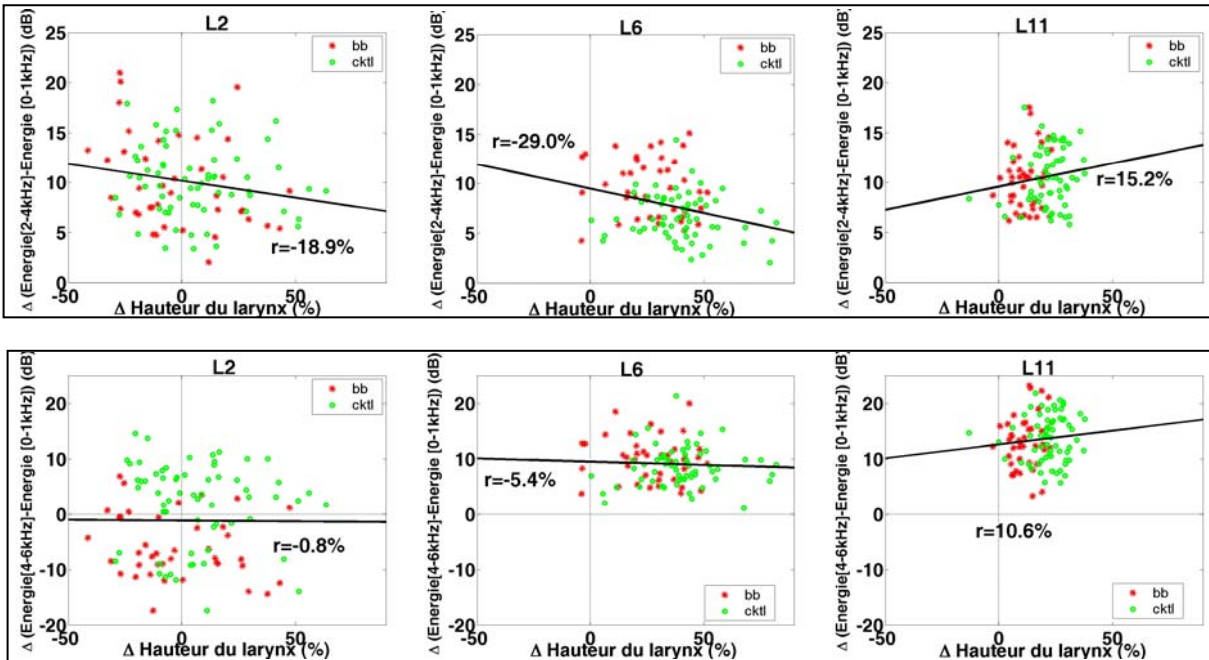


Corrélation entre la position du larynx et la répartition spectrale de l'énergie

Entre leurs valeurs dans le silence et dans le bruit



Entre leurs évolutions du silence au bruit



Annexe Res7 :
Résultats statistiques
détaillés du chapitre 7

Distinctivité des voyelles selon la dimension (F1-F0)_{barks}

	Locutrices de BD1									
	L1		L2		L6		L9		L10	
	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]
Dans le silence	3.8 barks F(1) = 168.9 ***	4.2 barks F(1) = 229.2 ***	4.6 barks F(1) = 1072.7 ***	4.4 barks F(1) = 626.3 ***	3.8 barks F(1) = 731.2 ***	3.8 barks F(1) = 983.0 ***	5.1 barks F(1) = 181.4 ***	4.8 barks F(1) = 125.7 ***	4.0 barks F(1) = 145.3 ***	4.4 barks F(1) = 166.9 ***
Dans le bruit cocktail	4.5 barks F(1) = 101.1 ***	4.6 barks F(1) = 120.1 ***	4.7 barks F(1) = 112.0 ***	4.8 barks F(1) = 110.0 ***	4.9 barks F(1) = 152.9 ***	5.2 barks F(1) = 166.7 ***	5.5 barks F(1) = 715.0 ***	5.3 barks F(1) = 376.4 ***	5.1 barks F(1) = 61.3 ***	5.2 barks F(1) = 63.6 ***
Evolution du silence au bruit	0.5 barks F(1) = 0.3 NS (p=0.60)	0.6 barks F(1) = 0.7 NS (p=0.44)	0.5 barks F(1) = 2.2 NS (p=0.18)	0.5 barks F(1) = 1.4 NS (p=0.29)	1.2 barks F(1) = 5.7 *	1.4 F(1) = 17.9 **	0.3 barks F(1) = 0.3 NS (p=0.58)	0.5 barks F(1) = 0.8 NS (p=0.40)	0.8 barks F(1) = 0.8 NS (p=0.40)	0.4 barks F(1) = 0.3 NS (p=0.62)

	Locuteurs de BD1									
	L3		L4		L5		L7		L8	
	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]
Dans le silence	2.6 barks F(1) = 187.8931, ***	2.5 barks F(1) = 1126.2163, ***	2.7 barks F(1) = 150.8 ***	2.6 barks F(1) = 107.0 ***	1.9 barks F(1) = 151.6 ***	1.8 barks F(1) = 127.3 ***	2.5 barks F(1) = 263.7 ***	2.4 barks F(1) = 416.2 ***	2.0 barks F(1) = 144.8 ***	2.2 barks F(1) = 216.2 ***
Dans le bruit cocktail	3.704157, F(1) = 746.013, ***	3.7 barks F(1) = 783.1227, ***	3.4 barks F(1) = 91.2 ***	3.8 barks F(1) = 403.6 ***	3.7 barks F(1) = 271.5 ***	3.6 barks F(1) = 390.6 ***	3.1 barks F(1) = 212.5 ***	3.0 barks F(1) = 86.2 ***	3.0 barks F(1) = 202.9 ***	3.0 barks F(1) = 161.8 ***
Evolution du silence au bruit	1.2 barks F(1) = 21.5539, **	1.3 barks F(1) = 70.1626, ***	0.6 barks F(1) = 1.4 NS (p=0.26)	1.2 barks F(1) = 9.6 *	2.0 barks F(1) = 93.9 ***	2.0 barks F(1) = 41.8 ***	0.6 barks F(1) = 4.0 NS (p=0.07)	0.7 barks F(1) = 2.5 NS (p=0.15)	1.0 barks F(1) = 15.4 2.0 **	0.8 barks F(1) = 6.9 *

	Locutrices de BD2					
	L2		L6		L11	
	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]
Dans le silence	5.0 barks F(1) = 66.5 ***	5.0 barks F(1) = 63.7 ***	5.1 barks F(1) = 996.1 ***	5.1 barks F(1) = 982.8 ***	4.0 barks F(1) = 246.5 ***	4.3 barks F(1) = 761.8 ***
Dans le bruit cocktail	5.4 barks F(1) = 727.5 ***	5.6 barks F(1) = 800.9 ***	4.6 barks F(1) = 58.2 ***	4.6 barks F(1) = 59.9 ***	3.5 barks F(1) = 143.1 ***	3.6 barks F(1) = 143.7 ***
Evolution du silence au bruit	0.4 barks F(1) = 0.4 NS (p=0.57)	0.6 barks F(1) = 0.5 NS (p=0.50)	-0.5 barks F(1) = 0.4 NS (p=0.53)	-0.5 barks F(1) = 0.5 NS (p=0.50)	-0.4 barks F(1) = 1.2 NS (p= 0.31)	-0.7 barks F(1) = 4.2 NS (p= 0.09)

Distinctivité des voyelles selon la dimension (F2-F1)_{barks}

	Locutrices de BD1									
	L1		L2		L6		L9		L10	
	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]
Dans le silence	5.0 barks 6.2 NS (p=0.07)	5.6 barks F(1) = 248.4 ***	6.4 barks F(1) = 153.9 ***	6.3 barks F(1) = 1226.7 ***	4.9 barks F(1) = 189.9 ***	5.5 barks F(1) = 665.4 ***	6.2 barks F(1) = 34.8 **	6.1 barks F(1) = 804.5 ***	6.2 barks F(1) = 197.7 ***	5.7 barks F(1) = 342.7 ***
Dans le bruit cocktail	4.7 barks F(1) = 117.2 ***	5.7 barks F(1) = 285.0 ***	5.5 barks F(1) = 641.3 ***	6.1 barks F(1) = 258.1 ***	4.9 barks F(1) = 520.7 ***	6.7 barks F(1) = 586.7 ***	4.9 barks F(1) = 172.8 ***	6.7 barks F(1) = 305.4 ***	5.2 barks F(1) = 156.6 ***	6.2 barks F(1) = 435.2 ***
Evolution du silence au bruit	-0.1 barks F(1) = 0.001 NS (p=0.97)	-0.1 barks F(1) = 0.02 NS (p=0.90)	-0.8 barks F(1) = 1.8 NS (p=0.22)	0.1 barks F(1) = 0.04 NS (p=0.86)	-0.1 barks F(1) = 0.04 NS (p=0.85)	1.2 barks F(1) = 14.6 **	-1.2 barks F(1) = 1.0 NS (p=0.35)	0.5 barks F(1) = 2.3 NS (p=0.17)	-1.1 barks F(1) = 3.3 NS (p=0.10)	0.5 barks F(1) = 2.8 NS (p=0.13)

	Locuteurs de BD1									
	L3		L4		L5		L7		L8	
	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]
Dans le silence	4.3 barks F(1) = 56.8 **	5.0 barks F(1) = 193.7 ***	5.4 barks F(1) = 85.4 ***	4.7 barks F(1) = 225.3 ***	4.6 barks F(1) = 157.7 ***	3.4 barks F(1) = 158.4 ***	4.0 barks F(1) = 154.6 ***	4.1 barks F(1) = 844.8 ***	4.4 barks F(1) = 83.5 ***	4.6 barks F(1) = 335.7 ***
Dans le bruit cocktail	4.6 barks F(1) = 548.7 ***	5.6 barks F(1) = 782.9 ***	6.2 barks F(1) = 1509.4 ***	5.8 barks F(1) = 1122.4 ***	4.0 barks F(1) = 129.0 ***	5.1 barks F(1) = 447.7 ***	4.5 barks F(1) = 190.6 ***	5.0 barks F(1) = 350.5 ***	3.8 barks F(1) = 57.6 ***	4.8 barks F(1) = 282.1 ***
Evolution du silence au bruit	0.3 barks F(1) = 0.5 NS (p=0.53)	0.7 barks F(1) = 3.3 NS (p=0.12)	0.7 barks F(1) = 1.3 NS (p=0.28)	1.0 barks F(1) = 8.6 *	-0.6 barks F(1) = 2.1 NS (p=0.18)	1.9 barks F(1) = 28.1 ***	0.5 barks F(1) = 1.4 NS (p=0.27)	0.9 barks F(1) = 8.0 *	-0.7 barks F(1) = 1.1 NS (p=0.32)	0.3 barks F(1) = 1.0 NS (p=0.34)

	Locutrices de BD2					
	L2		L6		L11	
	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]
Dans le silence	6.2 barks F(1) = 57.8 ***	7.2 barks F(1) = 75.5 ***	6.1 barks F(1) = 335.9 ***	7.8 barks F(1) = 348.1 ***	7.2 barks F(1) = 191.0 ***	6.7 barks F(1) = 227.6 ***
Dans le bruit cocktail	5.6 barks F(1) = 109.3 ***	7.9 barks F(1) = 219.1 ***	5.4 barks F(1) = 478.1 ***	6.8 barks F(1) = 127.2 ***	5.2 barks F(1) = 167.7 ***	5.8 barks F(1) = 432.4 ***
Evolution du silence au bruit	-0.6 barks F(1) = 0.2 NS (p=0.64)	0.6 barks F(1) = 0.3 NS (p=0.61)	-0.7 barks F(1) = 1.8 NS (p=0.22)	-0.9 barks F(1) = 1.3 NS (p=0.30)	-2.0 barks F(1) = 8.2 *	-0.9 barks F(1) = 2.1 NS (p=0.20)

Distinctivité des voyelles selon la dimension F2' barks

	Locutrices de BD1				
	L1	L2	L6	L9	L10
	[a]-[u]	[a]-[u]	[a]-[u]	[a]-[u]	[a]-[u]
Effet du bruit blanc	3.4 barks F(1) = 21.2, **	4.2 barks F(1) = 159.3 ***	3.3 barks F(1) = 334.3 ***	4.1 barks F(1) = 71.3 ***	4.3 barks F(1) = 287.3 ***
Effet du bruit cocktail	3.9 barks F(1) = 121.2 ***	3.9 barks F(1) = 80.5 ***	3.6 barks F(1) = 278.4 ***	4.1 barks F(1) = 139.9 ***	4.0 barks F(1) = 181.9 ***
Effet du type de bruit	-0.5 barks F(1) = 0.1 NS (p=0.82)	-0.6 barks F(1) = 0.2 NS (p=0.67)	-0.3 barks F(1) = 3.4 NS (p=0.09)	-0.4 barks F(1) = 0.001 NS (p=0.98)	-0.5 barks F(1) = 0.6 NS (p=0.47)

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
	[a]-[u]	[a]-[u]	[a]-[u]	[a]-[u]	[a]-[u]
Effet du bruit blanc	2.4 barks F(1) = 46.0 ***	2.8 barks F(1) = 90.7 ***	2.2 barks F(1) = 171.1 ***	2.3 barks F(1) = 157.0 ***	1.8 barks F(1) = 116.3 ***
Effet du bruit cocktail	3.2 barks F(1) = 822.5 ***	3.6 barks F(1) = 377.6 ***	3.1 barks F(1) = 610.5 ***	2.7 barks F(1) = 231.7 ***	2.2 barks F(1) = 85.1 ***
Effet du type de bruit	-0.1 barks F(1) = 3.3 NS (p=0.12)	0.2 barks F(1) = 4.0 NS (p=0.08)	0.003 barks F(1) = 12.7 **	0.2 barks F(1) = 3.8 NS (p=0.08)	-0.1 barks F(1) = 3.3 NS (p=0.10)

	Locutrices de BD2		
	L2	L6	L11
	[a]-[u]	[a]-[u]	[a]-[u]
Effet du bruit blanc	4.3 barks F(1) = 87.8 ***	4.0 barks F(1) = 182.8 ***	4.2 barks F(1) = 107.6 ***
Effet du bruit cocktail	4.0 barks F(1) = 103.5 ***	3.6 barks F(1) = 80.4 ***	3.1 barks F(1) = 141.2 ***
Effet du type de bruit	-0.3 barks F(1) = 0.2 NS 0.63	-0.4 barks F(1) = 0.6 NS 0.46	-1.1 barks F(1) = 9.0 *

Evolution dans le bruit de l'aire du triangle vocalique (dans l'espace (F1-F0)*F2*F3)

	Locutrices de BD1				
	L1	L2	L6	L9	L10
	Dans le silence	8.7 barks ³	13.9 barks ³	9.9 barks ³	15.4 barks ³
Dans le bruit cocktail	10.5 barks ³	12.5 barks ³	12.1 barks ³	14.7 barks ³	13.2 barks ³
Evolution du silence au bruit	1.8 barks ³	-1.5 barks ³	2.2 barks ³	-0.7 barks ³	0.0 barks ³

	Locuteurs de BD1				
	L3	L4	L5	L7	L8
	Dans le silence	7.8 barks ³	7.8 barks ³	4.7 barks ³	5.5 barks ³
Dans le bruit cocktail	9.8 barks ³	11.5 barks ³	9.0 barks ³	7.5 barks ³	7.5 barks ³
Evolution du silence au bruit	2.0 barks ³	3.7 barks ³	4.2 barks ³	1.9 barks ³	0.8 barks ³

	Locutrices de BD2		
	L2	L6	L11
	Dans le silence	15.8 barks ³	15.8 barks ³
Dans le bruit cocktail	15.5 barks ³	12.7 barks ³	9.4 barks ³
Evolution du silence au bruit	-0.3 barks ³	-3.1 barks ³	-5.9 barks ³

Evolution de l'inertie inter-catégories vocaliques dans le bruit (dans l'espace (F1-F0)*F2*F3)

Locutrices de BD1					
	L1	L2	L6	L9	L10
[a]-[i]	0.2 barks	0.1 barks	1.2 barks	0.4 barks	0.8 barks
[a]-[u]	1.1 barks	-0.2 barks	0.8 barks	-0.1 barks	0.1 barks
[i]-[u]	0.1 barks	-0.8 barks	-0.4 barks	-0.7 barks	-1.1 barks
Moyenne	0.5 barks	-0.3 barks	0.5 barks	-0.2 barks	-0.1 barks

Locuteurs de BD1					
	L3	L4	L5	L7	L8
[a]-[i]	0.2 barks	0.7 barks	1.3 barks	0.6 barks	0.4 barks
[a]-[u]	1.0 barks	1.1 barks	1.2 barks	0.5 barks	0.6 barks
[i]-[u]	-0.8 barks	0.7 barks	0.0 barks	0.4 barks	-0.4 barks
Moyenne	0.1 barks	0.8 barks	0.9 barks	0.5 barks	0.2 barks

Locutrices de BD2			
	L2	L6	L11
[a]-[i]	0.5 barks	-0.6 barks	-0.9 barks
[a]-[u]	-0.1 barks	-0.6 barks	-1.4 barks
[i]-[u]	-0.6 barks	-0.7 barks	-1.9 barks
Moyenne	-0.1 barks	-0.6 barks	-1.4 barks

Evolution de l'inertie intra-catégories vocaliques dans le bruit (dans l'espace (F1-F0)*F2*F3)

Locutrices de BD1					
	L1	L2	L6	L9	L10
[a]	0.4 barks	0.7 barks	0.5 barks	-0.02 barks	0.5 barks
[i]	-0.2 barks	-0.04 barks	0.04 barks	-0.3 barks	-0.2 barks
[u]	-1.7 barks	-0.3 barks	-0.3 barks	-0.9 barks	-0.2 barks
Moyenne	-0.5 barks	0.1 barks	0.1 barks	-0.4 barks	0.03 barks

Locuteurs de BD1					
	L3	L4	L5	L7	L8
[a]	0.2 barks	0.01 barks	0.1 barks	0.04 barks	0.1 barks
[i]	-0.1 barks	-0.1 barks	0.05 barks	0.4 barks	-0.02 barks
[u]	-0.8 barks	-0.4 barks	0.02 barks	-0.1 barks	0.2 barks
Moyenne	-0.2 barks	-0.2 barks	0.05 barks	0.1 barks	0.1 barks

Locutrices de BD2			
	L2	L6	L11
[a]	-0.6 barks	0.4 barks	-0.1 barks
[i]	0.03 barks	0.1 barks	-0.1 barks
[u]	-0.2 barks	-0.05 barks	-0.4 barks
Moyenne	-0.3 barks	0.1 barks	-0.2 barks

Evolution du rapport d'inertie inter sur l'inertie intra-catégories vocaliques dans le bruit (dans l'espace (F1-F0)*F2*F3)

Locutrices de BD1					Locuteurs de BD1				
L1	L2	L6	L9	L10	L3	L4	L5	L7	L8
+4.2	-2.2	-0.8	+4.4	-0.4	+5.7	+3.7	+0.9	-0.6	-1.0

Locutrices de BD2		
L2	L6	L11
+2.9	-2.6	+3.7

Distinctivité des voyelles selon la dimension de l'ouverture labiale

	Locutrices de BD2					
	L2		L6		L11	
	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]	[a]-[u]	[a]-[i]
Dans le silence	29.8 % F(1) = 225.2 ***	7.1 % F(1) = 8.6854, *	23.4 % F(1) = 470.1 ***	6.6 % F(1) = 57.8 ***	23.4 % F(1) = 215.6 ***	5.6 % F(1) = 16.9 **
Dans le bruit cocktail	34.4 % F(1) = 172.4 ***	12.7 % F(1) = 21.6 ***	32.8 % F(1) = 619.2 ***	15.9 % F(1) = 150.8 ***	24.4 % F(1) = 90.8 ***	8.2 % F(1) = 8.2 *
Evolution du silence au bruit	4.7 % F(1) = 3.2 NS (p=0.10)	5.6 % F(1) = 2.7 NS (p=0.12)	9.4 % F(1) = 50.0 ***	9.3 % F(1) = 56.6 ***	2.4 % F(1) = 1.9 NS (p=0.21)	2.6 % F(1) = 1.7 NS (p=0.21)

Distinctivité des voyelles selon la dimension de l'étirement labial

	Locutrices de BD2					
	L2		L6		L11	
	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]	[i]-[u]	[i]-[a]
Dans le silence	57.8 % F(1) = 166.8 ***	2.3 % F(1) = 1.7 NS (p=0.21)	42.9 % F(1) = 172.3 ***	-0.2 % F(1) = 0.02 NS (p=0.89)	54.5 % F(1) = 65.8 ***	0.2 % F(1) = 0.02 NS (p=0.88)
Dans le bruit cocktail	44.2 % F(1) = 438.1 ***	3.7 % F(1) = 4.7 *	38.0 % F(1) = 497.6 ***	-2.0 % F(1) = 1.3 NS (p=0.28)	37.7 % F(1) = 190.05 ***	-0.3 % F(1) = 0.01 NS (p=0.92)
Evolution du silence au bruit	-13.6 % F(1) = 7.3 *	1.4 % F(1) = 0.4 NS (p=0.55)	-4.9 % F(1) = 2.1 NS (p=0.17)	-1.9 % F(1) = 1.6 NS (p=0.23)	-16.0 % F(1) = 6.06 *	-0.5 % F(1) = 0.1 NS (p=0.75)

Distinctivité des voyelles selon la dimension de protrusion labiale

	Locutrices de BD2					
	L2		L6		L11	
	[u]-[i]	[u]-[a]	[u]-[i]	[u]-[a]	[u]-[i]	[u]-[a]
Dans le silence	68.0 % F(1) = 138.2 ***	67.2 % F(1) = 186.1 ***	34.0 % F(1) = 37.6 ***	38.9 % F(1) = 47.4 ***	43.5 % F(1) = 155.3 ***	50.6 % F(1) = 123.3 ***
Dans le bruit cocktail	84.7 % F(1) = 128.2 ***	81.9 % F(1) = 95.3 ***	56.6 % F(1) = 60.4 ***	59.8 % F(1) = 74.4 ***	54.669403, F(1) = 49.5 ***	72.6 % F(1) = 72.1 ***
Evolution du silence au bruit	16.7 % F(1) = 2.4 NS (p=0.14)	14.7 % F(1) = 1.6 NS (p=0.22)	22.6 % F(1) = 4.1 ns (p=0.06)	21.0 % F(1) = 5.4 *	11.2 % F(1) = 2.7 NS (p=0.12)	22.0 % F(1) = 9.1 **

Distinctivité des voyelles selon la dimension de l'arrondi labial

	Locutrices de BD2					
	L2		L6		L11	
	[u]-[i]	[u]-[a]	[u]-[i]	[u]-[a]	[u]-[i]	[u]-[a]
Dans le silence	15.1 % F(1) = 46.1 ***	12.3 % F(1) = 32.7 ***	9.3 % F(1) = 24.9 ***	6.3 % F(1) = 11.7 **	11.9 % F(1) = 122.2 ***	8.1 % F(1) = 87.0 ***
Dans le bruit cocktail	20.0 % F(1) = 116.9 ***	15.8 % F(1) = 32.8 ***	18.4 % F(1) = 30.7 ***	17.6 % F(1) = 26.3 ***	19.3 % F(1) = 28.3 ***	16.2 % F(1) = 32.6 ***
Evolution du silence au bruit	4.9 % F(1) = 8.5 *	3.5 % F(1) = 1.6 NS (p=0.23)	9.1 % F(1) = 3.4 NS (p=0.09)	11.452531, F(1) = 7.6 *	5.4 % F(1) = 1.8 NS (p=0.20)	8.1 % F(1) = 5.5 *

Evolution du pincement labial dans le bruit

	Locutrices de BD2		
	L2	L6	L11
Δ Amplitude	7.7 % F(1) = 14.6 **	26.0 % F(1) = 10.2 **	21.3 % F(1) = 28.4 ***
Δ Durée	20 ms F(1) = 3.9 NS (p=0.06)	18 ms F(1) = 1.2 NS (p=0.28)	2 ms F(1) = 0.05 NS (p=0.82)

Evolution de la durée des cibles et des transitions vocaliques dans le bruit

Δ Durée des phases stables		
L2	L6	L11
22ms	8ms	19ms
F(1) = 7.2 **	F(1) = 1.0 NS (p=0.31)	F(1) = 8.6 **

Δ Durée des transitions formantiques		
L2	L6	L11
28 ms	37 ms	3 ms
F(1) = 32.3 ***	F(1) = 87.9 ***	F(1) = 0.3 NS (p=0.55)

Annexe Res8 :
Résultats statistiques
détaillés du chapitre 8

Evolution dans le bruit de la dynamique d'intensité par syllabe

Locutrices de BD1					Locuteurs dde BD1				
L1	L2	L6	L9	L10	L3	L4	L5	L7	L8
5.4 dB F(1) = 17.1 ***	6.8 dB F(1) = 21.9 ***	4.0 dB F(1) = 14.4 ***	3.5 dB F(1) = 13.0 ***	4.1 dB F(1) = 10.4 **	6.4 dB F(1) = 22.9 ***	4.6 dB F(1) = 13.4 ***	2.4 dB F(1) = 5.8 *	1.8 dB F(1) = 3.4 NS (p=0.07)	2.1 dB F(1) = 5.2 *

Locutrices de BD2		
L2	L6	L11
4.6 dB F(1) = 24.4 ***	2.7 dB F(1) = 22.3 ***	2.7 dB F(1) = 27.2 ***

Durée de la consonne initiale d'un mot de contenu (C1) par rapport à une consonne non initiale (C2)

	Locutrices de BD1									
	L1		L2		L6		L9		L10	
	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée
Dans le silence	-24ms F(1) = 15.0 ***	25ms F(1) = 6.3 *	-33ms F(1) = 11.2 **	29ms F(1) = 5.6 *	-22ms F(1) = 12.0 **	8ms F(1) = 0.8 NS (p=0.38)	-21ms F(1) = 8.5 **	21ms F(1) = 4.6 *	-28ms F(1) = 10.4 **	10ms F(1) = 0.6 NS (p=0.46)
Dans le bruit cocktail	-20ms F(1) = 8.7 **	9ms F(1) = 0.7 NS (p=0.40)	-20ms F(1) = 6.3 *	14ms F(1) = 1.3 NS (p=0.26)	-15ms F(1) = 5.5 *	11ms F(1) = 1.4 NS (p=0.24)	-19ms F(1) = 11.6 **	6ms F(1) = 0.3 NS (p=0.58)	-21ms F(1) = 6.2 *	20ms F(1) = 2.0 NS (p=0.17)
Evolution du silence au bruit	4ms F(1) = 0.08 NS (p=0.77)	-16ms F(1) = 0.5 NS (p=0.50)	12ms F(1) = 0.5 NS (p=0.49)	-14ms F(1) = 0.6 NS (p=0.43)	6ms F(1) = 0.4 NS (p=0.53)	3ms F(1) = 0.04 NS (p=0.85)	2ms F(1) = 0.04 NS (p=0.84)	-15ms F(1) = 0.9 NS (p=0.36)	7ms F(1) = 0.6 NS (p=0.43)	10ms F(1) = 0.007 NS (p=0.93)

	Locuteurs de BD1									
	L3		L4		L5		L7		L8	
	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée
Dans le silence	-8ms F(1) = 0.6 NS (p=0.44)	0ms F(1) = 0.0003 NS (p=0.99)	-11ms F(1) = 1.4 NS (p=0.25)	21ms F(1) = 1.5 NS (p=0.23)	-4ms F(1) = 0.3 NS (p=0.56)	33ms F(1) = 9.4 **	-8ms F(1) = 1.2 NS (p=0.28)	8ms F(1) = 0.35 NS (p=0.56)	-19ms F(1) = 8.5 **	24ms F(1) = 6.2 *
Dans le bruit cocktail	13ms F(1) = 2.0 NS (p=0.16)	42ms F(1) = 18.4 ***	-6ms F(1) = 1.0 NS (p=0.31)	34ms F(1) = 7.8 **	-3ms F(1) = 0.2 NS (p=0.62)	27ms F(1) = 8.0 **	-9ms F(1) = 2.4 NS (p=0.13)	9ms F(1) = 0.8 NS (p=0.38)	-10ms F(1) = 2.2 NS (p=0.15)	24ms F(1) = 5.4 *
Evolution du silence au bruit	20ms F(1) = 3.0 NS (p=0.09)	42ms F(1) = 3.7 NS (p= 0.07)	4ms F(1) = 0.4 NS (p=0.52)	13ms F(1) = 0.2 NS (p=0.70)	2ms F(1) = 0.04 NS (p=0.84)	-6ms F(1) = 0.4 NS (p=0.52)	2ms F(1) = 0.001 NS (p=0.98)	7ms F(1) = 0.2 NS (p=0.63)	9ms F(1) = 1.4 NS (p=0.25)	0ms F(1) = 0.00002 NS (p=0.99)

Intensité de la consonne initiale d'un mot de contenu (C1) par rapport à une consonne non initiale (C2)

	Locutrices de BD1									
	L1		L2		L6		L9		L10	
	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée
Dans le silence	2.1 dB F(1) = 0.03 NS (p=0.85)	-5.4 dB F(1) = 0.1 NS (p=0.71)	3.9 dB F(1) = 0.1 NS (p=0.72)	-4.3 dB F(1) = 0.1 NS (p=0.75)	0.9 dB F(1) = 0.01 NS (p=0.94)	-1.1dB F(1) = 0.01 NS (p=0.94)	0.3 dB F(1) = 0.001 NS (p=0.98)	-1.9 dB F(1) = 0.02 NS (p=0.90)	2.2 dB F(1) = 0.03 NS (p=0.86)	-1.9 dB F(1) = 0.02 NS (p=0.90)
Dans le bruit cocktail	5.2 dB F(1) = 9.6 **	-6.7 dB F(1) = 12.1 **	5.0 dB F(1) = 11.6 **	-7.1 dB F(1) = 9.8 **	3.5 dB F(1) = 3.8 NS (p=0.05)	-6.4 dB F(1) = 7.5 **	0.6 dB F(1) = 0.2 NS (p=0.69)	-3.4 dB F(1) = 4.0 NS (p=0.05)	1.3 dB F(1) = 0.6 NS (p=0.45)	-3.1 dB F(1) = 2.9 NS (p=0.10)
Evolution du silence au bruit	3.1 dB F(1) = 2.9 NS (p=0.09)	-1.3dB F(1) = 0.2 NS (p=0.65)	1.1 dB F(1) = 0.4 NS (p=0.55)	-2.8 dB F(1) = 1.6 NS (p=0.21)	2.5 dB F(1) = 2.2 NS (p=0.14)	-5.2 dB F(1) = 9.3 **	0.3 dB F(1) = 0.01 NS (p=0.91)	-1.5 dB F(1) = 0.7 NS (p=0.40)	-0.9dB F(1) = 0.2 NS (p=0.67)	-1.2 dB F(1) = 0.7 NS (p=0.42)

	Locuteurs de BD1									
	L3		L4		L5		L7		L8	
	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée	voisée	non voisée
Dans le silence	0.1 dB F(1) = 0.0001 NS (p=0.99)	-4.2 dB F(1) = 0.07 NS (p=0.79)	1.7 dB F(1) = 0.02 NS (p=0.89)	-5.7 dB F(1) = 0.1 NS (p=0.72)	1.6 dB F(1) = 0.02 NS (p=0.89)	-4.5 dB F(1) = 0.1 NS (p=0.76)	0.6 dB F(1) = 0.002 NS (p=0.96)	-2.7 dB F(1) = 0.03 NS (p=0.87)	1.3 dB F(1) = 0.01 NS (p=0.92)	-3.1 dB F(1) = 0.04 NS (p=0.84)
Dans le bruit cocktail	0.2 dB F(1) = 0.02 NS (p=0.89)	-5.0 dB F(1) = 6.5 *	1.2 dB F(1) = 0.4 NS 0.5376	-8.3 dB F(1) = 11.5 **	2.0 dB F(1) = 3.1 NS (p=0.08)	-6.8 dB F(1) = 20.0 ***	1.9 dB F(1) = 2.3 NS (p=0.13)	-2.3 dB F(1) = 2.1 NS (p=0.15)	2.6 dB F(1) = 2.8 NS (p=0.10)	-4.3 dB F(1) = 6.3 *
Evolution du silence au bruit	0.1 dB F(1) = 0.1 NS (p=0.75)	-0.9 dB F(1) = 1.6 NS (p=0.21)	-0.5 dB F(1) = 0 NS (p=1)	-2.6 dB F(1) = 2.4 NS (p=0.14)	0.4 dB F(1) = 0.1 NS (p=0.79)	-2.3 dB F(1) = 1.2 NS (p=0.29)	1.3 dB F(1) = 0.8 NS (p=0.38)	0.3 dB F(1) = 0.01 NS (p=0.90)	1.2 dB F(1) = 1.0 NS (p=0.32)	-1.3 dB F(1) = 0.5 NS (p=0.47)

Pincement labial de la consonne initiale d'un mot de contenu (C1) par rapport à une consonne non initiale (C2)

	MaxPince(C1)-MaxPince(C2)			PicVitessePince(C1)-PicVitessePince(C2)		
	L2	L6	L11	L2	L6	L11
Dans le silence	-0.8% F(1) = 0.1 NS (p=0.77)	3.2% F(1) = 3.1 NS (p=0.10)	-3.7% F(1) = 0.9 NS (p=0.36)	5.4 %·s ⁻¹ F(1) = 26.6 ***	2.8 %·s ⁻¹ F(1) = 13.3 **	3.6 %·s ⁻¹ F(1) = 4.7 *
Dans le bruit cocktail	6.5% F(1) = 6.2 *	12.8% F(1) = 4.0 NS (p=0.07)	10.1% F(1) = 2.3 NS (p=0.15)	5.6 %·s ⁻¹ F(1) = 10.6 **	21.9 %·s ⁻¹ F(1) = 20.1 ***	10.8 %·s ⁻¹ F(1) = 17.7 ***
En évolution du silence au bruit	7.3% F(1) = 1.8 NS (p=0.26)	9.5% F(1) = 1.7 NS (p=0.22)	13.8% F(1) = 3.0 NS (p=0.13)	0.2 %·s ⁻¹ F(1) = 0.1 NS (p=0.73)	19.0 %·s ⁻¹ F(1) = 28.7 ***	7.2 %·s ⁻¹ F(1) = 11.5 **

Durée de la syllabe initiale d'un énoncé (S1) par rapport à une syllabe non initiale (S2)

		L2	L6	L11
Durée(S1)- Durée(S2)	Dans le silence	10ms F(1) = 1.9 NS (p=0.18)	19ms F(1) = 18.0 ***	27ms F(1) = 32.4 ***
	Dans le bruit cocktail	21ms F(1) = 10.8 **	28ms F(1) = 24.6 ***	33ms F(1) = 27.1 ***
	En évolution du silence au bruit	11ms F(1) = 1.3 NS (p=0.27)	9ms F(1) = 2.3 NS (p=0.14)	7ms F(1) = 1.4 NS (p=0.25)

Durée de la syllabe finale d'un énoncé (S1) par rapport à une consonne non finale (S2)

		L2	L6	L11
Durée(S1)- Durée(S2)	Dans le silence	63ms F(1) = 75.4 ***	59ms F(1) = 68.7 ***	22ms F(1) = 9.9 **
	Dans le bruit cocktail	76ms F(1) = 55.3 ***	74ms F(1) = 93.7 ***	53ms F(1) = 116.5 ***
	En évolution du silence au bruit	13ms F(1) = 1.8 NS (p=0.18)	14ms F(1) = 2.9 NS (p=0.10)	30ms F(1) = 14.5 ***

Intensité de la syllabe finale d'un énoncé (S1) par rapport à une consonne non finale (S2)

		L2	L6	L11
Intensité(S1)- Intensité(S2)	Dans le silence	5.0 dB F(1) = 9.9 **	2.1 dB F(1) = 6.0 *	2.4 dB F(1) = 7.9 **
	Dans le bruit cocktail	1.9 dB F(1) = 3.3 NS (p=0.08)	1.2 dB F(1) = 1.9 NS (p=0.18)	0.5 dB F(1) = 0.3 NS (p=0.57)
	En évolution du silence au bruit	3.0 dB F(1) = 7.5 *	0.9 dB F(1) = 2.1 NS (p=0.16)	1.8 dB F(1) = 5.8 *

Articulation de la syllabe finale (S1) d'un énoncé par rapport à une consonne non finale (S2)

		L2	L6	L11
MaxS(S1)- IMaxS(S2)	Dans le silence	2.8% F(1) = 0.6 NS (p=0.45)	1.8% F(1) = 0.9 NS (p=0.35)	1.7% F(1) = 4.6 *
	Dans le bruit cocktail	2.2% F(1) = 0.4 NS (p=0.55)	0.7% F(1) = 0.04 NS (p=0.85)	0.2% F(1) = 0.003 NS (p=0.96)
	En évolution du silence au bruit	0.6% F(1) = 0.03 NS (p=0.87)	2.5% F(1) = 0.9 NS (p=0.37)	1.5% F(1) = 0.5 NS (p=0.49)

Evolution dans le bruit de l'ambitus tonal de la chute mélodique en fin d'énoncé

L2	L6	L11
0.7 tons F(1) = 9.6 **	1.0 ton F(1) = 15.1 ***	1.0 ton F(1) = 3.1 NS (p=0.09)

Annexe Res9 :
Résultats statistiques
détaillés du chapitre 9

Différence de durée entre la syllabe d'un mot de contenu et celle d'un mot de fonction

	L2	L6	L11
Dans le silence	23ms F(1) = 4.1 *	63ms F(1) = 83.2 ***	30ms F(1) = 48.7 ***
Dans le bruit cocktail	37ms F(1) = 10.3 **	62ms F(1) = 64.3 ***	30ms F(1) = 23.4 ***
En évolution du silence au bruit	14ms F(1) = 1.1 NS (p=0.28)	1 ms F(1) = 0.01 NS (p=0.91)	1ms F(1) = 0.01 NS (p=0.94)

Différence d'intensité entre la syllabe d'un mot de contenu et celle d'un mot de fonction

	L2	L6	L11
Dans le silence	-0.1dB F(1) = 0.005 NS (p=0.92)	1.5dB F(1) = 7.0 **	0.7dB F(1) = 1.3 NS (p=0.23)
Dans le bruit cocktail	2.1dB F(1) = 24.8 ***	1.1dB F(1) = 7.9 **	-0.1dB F(1) = 0.1 NS (p=0.75)
En évolution du silence au bruit	2.0dB F(1) = 8.8 **	-0.4dB F(1) = 0.5 NS (p=0.40)	-0.8dB F(1) = 2.5 NS (p=0.12)

Différence de timbrage entre la syllabe d'un mot de contenu et celle d'un mot de fonction

	L2	L6	L11
Dans le silence	-0.3dB F(1) = 0.4 NS (p=0.55)	-0.5dB F(1) = 1.5 NS (p=0.22)	-1.0dB F(1) = 5.1 *
Dans le bruit cocktail	0.8dB F(1) = 1.1 NS (p=0.30)	2.0dB F(1) = 52.1 ***	2.6dB F(1) = 29.2 ***
En évolution du silence au bruit	1.1dB F(1) = 2.3 NS (p= 0.14)	2.5dB F(1) = 11.3 **	3.6dB F(1) = 39.7 ***

Différence de F0 entre la syllabe d'un mot de contenu et celle d'un mot de fonction

	L2	L6	L11
Dans le silence	0.6 ton F(1) = 6.3 *	0.9 ton F(1) = 28.1 ***	0.9 ton F(1) = 24.0 ***
Dans le bruit cocktail	1.4 tons F(1) = 114.5 ***	1.1 tons F(1) = 46.6 ***	0.8 ton F(1) = 22.6 ***
En évolution du silence au bruit	0.8 ton F(1) = 10.2 **	0.2 ton F(1) = 1.1 NS (p=0.31)	0.1 ton F(1) = 0.1 NS(p= 0.79)

Différence d'ouverture labiale sur la syllabe d'un mot de contenu et celle d'un mot de fonction

	L2	L6	L11
Dans le silence	-1.8% F(1) = 2.7 NS (p=0.10)	0.8% F(1) = 1.3 NS (p=0.26)	-0.8% F(1) = 1.7 NS (p=0.19)
Dans le bruit cocktail	2.1% F(1) = 2.8 NS (p=0.09)	4.9% F(1) = 19.4 ***	3.2% F(1) = 27.9 ***
En évolution du silence au bruit	3.9% F(1) = 13.9 ***	4.1% F(1) = 14.3 ***	4.0% F(1) = 36.3 ***

***Références
bibliographiques***

- [1] "Trésor de la langue française informatisé." from <http://atilf.atilf.fr/tlf.htm>.
- [2] (1969). "Harvard sentences. Appendix of: IEEE subcommittee on subjective measurements." IEEE transactions on Audio and Electroacoustics 17: 227-246.
- [3] Abbs, J. H. and Gracco, V. L. (1984). "Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech." Journal of Neurophysiology 51(4): 705-723.
- [4] Acker, B. F. (1987). "Vocal tract adjustments for the projected voice." Journal of Voice 1: 83-87.
- [5] Alku, P. and Vilkmán, E. (1996). "A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers." Folia Phoniatrica Logopaedica 48(5): 240-254.
- [6] Alku, P., Vintturi, J. and Vilkmán, E. (1999). "On the linearity of the relationship between the sound pressure level and the negative peak amplitude of the differentiated glottal flow in vowel production." Speech Communication 28(4): 269-281.
- [7] Alku, P., Vintturi, J. and Vilkmán, E. (2002). "Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation." Speech Communication 38(3-4): 321-334.
- [8] Amazi, D. K. and Garber, S. R. (1982). "The Lombard sign as a function of age and task." Journal of Speech and Hearing Research 25(4): 581-585.
- [9] Amphoux, P. (1991). Aux ecoutes de la ville. Rapport n° 94, Cresson, Grenoble.
- [10] Amy de la Bretèque, B. (2003). *L'équilibre et le rayonnement de la voix*. Solal, Marseille.
- [11] Arons, B. (1992). "A Review of the Cocktail Party Effect." Journal of the American Voice I/O Society . 12 35-50.
- [12] Aronsson, C., Södersten, M., Bohman, M. and Ternström, S. (2005). "Loud voice produced during realistic environmental noise exposure a comparison of acoustic data, subglottal pressure and subjective reports between vocal nodule patients and healthy female subjects". Actes de PVOC, London.
- [13] Atal, B. S. and Hanauer, S. L. (1971). "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave." The Journal of the Acoustical Society of America 50(2): 637-655.
- [14] Aubergé, V., Rillard, A. and Audibert, N. (2005). "De E-Wiz à E-Clone : méthodologie expérimentale pour la modélisation des émotions et affects authentiques". Actes de Workshop sur les Agents Conversationnels Animés, Grenoble.
- [15] Audouy, M. (2000). *Logiciel de traitement d'images video pour la détermination de mouvements des lèvres*. Projet de fin d'études, ENSIMAG, Grenoble.
- [16] Autesserre, D., Charpy, N., Crevier-Buchman, L., Dejonckere, P., Eluard, F., Fresnel, E., Giovanni, A., Schoentgen, J. and Timmermans, B. (2006). *Troubles de la voix chez les enseignants*, Institut National de la Santé et de la Recherche Médicale pour la Mutuelle générale de l'Education Nationale.
- [17] Ayers, G. M. (1994). *Discourse functions of pitch range in spontaneous and read speech*. Working papers in linguistics. Columbus, Ohio.
- [18] Badaoui, L. (1997). *L'effet Lombard dans le chant: une approche de la théorie de la régulation des paramètres de la phonation par boucle de rétroaction externe*. Mémoire de maîtrise de phonétique. Université Sorbonne Nouvelle, Paris.

- [19] Bagou, O., Fougeron, C. and Frauenfelder, U. H. (2002). "Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language". Actes de Speech Prosody, Aix-en-Provence, 159-162.
- [20] Bagou, O. and Frauenfelder, U. H. (2006). "Stratégie de segmentation prosodique: rôle des prééminences initiales et finales dans l'acquisition d'une langue artificielle". Actes de XXVIèmes Journées d'Etude sur la Parole, Dinard, 571-574.
- [21] Bailly, L. (2005). *Etude articulatoire de la parole en environnement bruyant*. Mémoire de Master ATIAM. Université Paris VI.
- [22] Baum, S. R. and McFarland, D. H. (1997). "The development of speech adaptation to an artificial palate." *Journal of the Acoustic Society of America* 102(4): 2353-2359.
- [23] Bele, I. V. (2006). "The Speaker's Formant." *Journal of Voice* 20(4): 555-578.
- [24] Benoit, C., Mohamadi, T. and Kandel, S. (1994). "Effects of phonetic context on audio-visual intelligibility of French." *Journal of Speech and Hearing Research* 37(5): 1195-1203.
- [25] Berthoz, A. (1997). *Le sens du mouvement*. Odile Jacob,
- [26] Bloothoof, G. and Plomp, R. (1986). "The sound level of the singer's formant in professional singing." *Journal of the Acoustic Society of America* 79(6): 2028-2033.
- [27] Blum, A., Afonso, A., Katz, B. F. G. and Jacquemin, C. (2004). "Expérimentation sur la perception de l'espace en réalité virtuelle immersive audio." Actes de 16ème Conférence Francophone sur l'Interaction Homme-Machine, Namur, Belgique.
- [28] Blumstein, S. and Stevens, K. (1979). "Acoustic invariance in speech production." *Journal of the Acoustic Society of America* 66: 1001-1017.
- [29] Boë, L.-J., Perrier, P. and Bailly, G. (1992). "The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion." *Journal of Phonetics* 20: 27-38.
- [30] Boersma, P. and Weenink, D. Praat: doing phonetics by computer (Version 4.2.28) [Computer program]: récupéré depuis <http://www.praat.org/> en 2004.
- [31] Bond, Z. S., Moore, T. J. and Gable, B. (1989). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask." *Journal of the Acoustic Society of America* 85(2): 907-912.
- [32] Boril, H. and Pollak, P. (2005). "Design and collection of czech lombard database". Actes de ICSLP, Lisbon, Portugal, 1577-1580.
- [33] Bougnoux, D. (2002). *Introduction aux sciences de la communication*. La découverte, Paris.
- [34] Briolle, F. and Canevet, G. (1993). *Evaluation de la qualité sonore de casque d'écoute par simulation numérique des fonctions de transfert* These de doctorat. Université d'Aix-Marseille 2.
- [35] Browman, C. P. and Goldstein, L. (1989). "Articulatory gestures as phonological units." *Phonology* 6: 201-252.
- [36] Brown, G., Anderson, A., Yule, G. and Shillcock, R. (1983). *Teaching Talk*. Cambridge University Press, UK,
- [37] Brown, W. S., Jr., Rothman, H. B. and Sapienza, C. M. (2000). "Perceptual and acoustic study of professionally trained versus untrained voices." *Journal of Voice* 14(3): 301-309.

- [38] Buekers, R. (1998). "Are voice endurance tests able to assess vocal fatigue?" *Clinical Otolaryngology and Allied Sciences* 23(6): 533-538.
- [39] Burnett, T. A., Freedland, M. B., Larson, C. R. and Hain, T. C. (1998). "Voice F0 responses to manipulations in pitch feedback." *Journal of the Acoustic Society of America* 103(6): 3153-3161.
- [40] Burzynski, C. M. and Starr, C. D. (1985). "Effects of feedback filtering on nasalization and self-perception of nasality." *Journal of Speech and Hearing Research* 28(4): 585-588.
- [41] Calliope (1989). *La parole et son traitement automatique*. Masson, Paris.
- [42] Carlson, R., Granström, B. and Fant, G. (1970). "Some studies concerning perception of isolated vowels." *STL-QPSR* 2-3: 19-35.
- [43] Carlsson, G. and Sundberg, J. (1992). "Formant frequency tuning in singing." *Journal of Voice* 6: 256-260.
- [44] Carre, R. and Divenyi, P. L. (2000). "Modeling and perception of 'gesture reduction'." *Phonetica* 57(2-4): 152-69.
- [45] Carson, C. P., Ingrisano, D. R. and Eggleston, K. D. (2003). "The effect of noise on computer-aided measures of voice: a comparison of CSpeechSP and the Multi-Dimensional Voice Program software using the CSL 4300B Module and Multi-Speech for Windows." *Journal of Voice* 17(1): 12-20.
- [46] Castellanos, A., JM.Benedi and Casacuberta, F. (1996). "An analysis of general acoustic-phonetic features for spanish speech produced with the lombard effect." *Speech Communication* 20: 23-35.
- [47] Castellengo, M. (1973). "Particularités acoustiques de la voix des chanteurs professionnels." *Bulletin du GAM* 67.
- [48] Chan, R. M. M. and Yiu, E. M. L. (2003). "Effect of Hydration and Vocal Rest on the Vocal Fatigue in Amateur Karaoke Singers." *Journal of Voice* 17(2): 216-227.
- [49] Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears." *Journal of Acoustic Society of America* 25: 975-979.
- [50] Childers, D. G. and Lee, C. K. (1991). "Vocal quality factors: analysis, synthesis, and perception." *Journal of the Acoustic Society of America* 90(5): 2394-2410.
- [51] Chistovich, L. A. and Lublinskaya, V. V. (1979). "The 'center of gravity' effect in vowel spectra and critical distance between the formants: ." *Hearing Research* 1 185-195.
- [52] Christophe, A. (1993). *Rôle de la prosodie dans la segmentation en mots*. These de doctorat. EHESS, Paris.
- [53] Chuberre, B. (2000). *Les registres et passages dans la voix chantée*. Mémoire de phoniatrie, Université de Nantes.
- [54] Clark, G., Mitra, S. and Parker, S. (1981). "Block implementation of adaptative digital filters." *IEEE Transactions on circuits and systems* 28: 584-592.
- [55] Clark, H. H. and Wilkes-Gibbs, D. (1986). "Referring as a collaborative process." *Cognition* 22(1): 1-39.
- [56] Cleveland, T. F., Sundberg, J. and Stone, R. E. (2001). "Long-term-average spectrum characteristics of country singers during speaking and singing." *Journal of Voice* 15(1): 54-60.

- [57] Coleman, R. F. (1971). "Effect of waveform changes upon roughness perception." *Folia Phoniatica (Basel)* 23(5): 314-322.
- [58] Cooper, D. S. and Rice, D. H. (1990). "Fatigue resistance of canine vocal fold muscle." *Annals of Otology, Rhinology and Laryngology* 99(3 Pt 1): 228-233.
- [59] Cox, R. M., Alexander, G. C. and Gilmore, C. (1987). "Intelligibility of average talkers in typical listening environments." *Journal of Acoustical Society of America* 81(5): 1598-1608.
- [60] Cutler, A. (1993). "Phonological cues to open- and closed-class words in the processing of spoken sentences." *Journal of Psycholinguistics Research* 22: 109-131.
- [61] Dang, J. and Honda, K. (1997). "Acoustic characteristics of the piriform fossa in models and humans." *The Journal of the Acoustical Society of America* 101(1): 456-465.
- [62] Davis, C., Kim, J., Grauwinkel, K. and Mixdorff, H. (2006). "Lombard speech: Auditory(A), Visual(V) and AV effects". *Actes de Speech prosody, Dresden, Germany*.
- [63] Dejonckere, P. (1979). "L'effet Lombard-Tarneau objectif." *Revue Electrodiagnostic-Therapie* 16(2): 87-95.
- [64] Dejonckere, P. H., Obbens, C., de Moor, G. M. and Wieneke, G. H. (1993). "Perceptual evaluation of dysphonia: reliability and relevance." *Folia Phoniatica (Basel)* 45(2): 76-83.
- [65] Dejonckere, P. H. and Pepin, F. (1983). "Etude de l'effet Lombard par la mesure du niveau sonore équivalent." *Folia Phoniatica (Basel)* 35(6): 310-315.
- [66] Dejonckere, P. H., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchman, L. and Millet, B. (1996). "Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements." *Revue de Laryngologie, d'Otologie et de Rhinologie (Bordeaux)* 117(3): 219-224.
- [67] Delattre, P. (1966). "Les Dix Intonations de base du français." *The French Review* 40(1): 1-14.
- [68] Deliyski, D. D., Shaw, H. S. and Evans, M. K. (2005). "Adverse effects of environmental noise on acoustic voice quality measurements." *Journal of Voice* 19(1): 15-28.
- [69] Diedrich, F. J. and Warren, W. H. J. (1995). "Why change gaits? Dynamics of the walk-run transition." *Journal of experimental psychology. Human perception and performance* 21(1): 183-202.
- [70] Dieroff, H. G. and Siegert, C. (1966). "[Tone height shifts under noise conditions]." *Folia Phoniatr (Basel)* 18(4): 247-55.
- [71] Dieroff, H. G. and Siegert, C. (1966). "Tonhöhenverschiebung unter Lärm- belästigung [Tone height shifts under noise conditions]." *Folia Phoniatica (Basel)* 18(4): 247-55.
- [72] Dodane, C., Blanc, J. M. and Dominey, P. F. (2006). "Différenciation des mots de fonction et des mots de contenu par la prosodie: analyse d'un corpus trilingue de langue adressée à l'enfant et à l'adulte". *Actes de XXVIèmes Journées d'Etude sur la Parole, Dinard, 255-258*.
- [73] Dohalska, M. and Mejvaldova, J. (2000). "Rôle de la prosodie dans la communication du milieu bruite". *Actes de XXIIIèmes Journées d'Etudes sur la Parole, Aussois, France, 265-268*.
- [74] Dohen, M. (2005). *Deixis prosodique multisensorielle : production et perception audiovisuelle de la focalisation contrastive en français*. These de doctorat. INPG, Grenoble.
- [75] Donohue, A. (2005). *A comparison of the effects of vocal warm-up exercises and hydration on vocal fatigue*. Department of Communication Disorders, University of massachusetts, Amherst.

- [76] Draegert, G. L. (1951). "Relationships between voice variables and speech intelligibility in high level noise." *Speech Monographs* (18): 272-278.
- [77] Dreher, J. J. and O'Neill, J. (1957). "Effects of Ambient Noise on Speaker Intelligibility for Words and Phrases " *Journal of the Acoustical Society of America* 29: 1320-1323.
- [78] Dubois, D., Guastavino, C., Maffiolo, V. and Raimbault, M. (2001). "A cognitive approach to soundscape research." *Journal of the Acoustic Society of America* 115(5): 2495-2495.
- [79] Egan, J. J. (1971). "The Lombard Reflex. Historical perspective." *Archives of Otolaryngology* 94: 310-312.
- [80] Egan, J. J. (1972). "Psychoacoustics of the Lombard voice response." *Journal of Auditory Research* 12: 318-324.
- [81] Eliades, S. J. and Wang, X. (2003). "Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations." *Journal of Neurophysiology* 89(4): 2194-2207.
- [82] Elliot, N., Sundberg, J. and Gramming, P. (1995). "What happens during vocal warm-up?" *Journal of Voice* 9(1): 37-44.
- [83] Elman, J. L. (1981). "Effects of frequency-shifted feedback on the pitch of vocal productions." *Journal of the Acoustic Society of America* 70(1): 45-50.
- [84] Eluard, P. F. (2004). *Troubles de la voix chez les enseignants. Document pour le médecin du travail*, Institut National de Recherche et de Sécurité.
- [85] Eustace, C. S., Stemple, J. C. and Lee, L. (1996). "Objective measures of voice production in patients complaining of laryngeal fatigue." *Journal of Voice* 10(2): 146-154.
- [86] Fairbanks, G. (1954). "Systematic research in experimental phonetics: A theory of speech mechanism as a sevosystem." *Journal of speech and Hearing Disorders* 19: 133-139.
- [87] Fairbanks, G. and Miron, M. S. (1957). "Effects of vocal effort upon consonant-vowel ratio within the syllable." *Journal of the Acoustic Society of America* 29(5): 621-626.
- [88] Fant, G. (1960). *Acoustic theory of speech production*. Mouton, La Haye.
- [89] Fant, G. (1995). "The LF-model revisited. Transformations and frequency domain analysis." *STL-QPSR* 2-3: 119-156.
- [90] Faure, M. A. (1988). "Dynamique respiratoire et qualités acoustiques de la voix." *Bulletin d'audiophonologie. Annales scientifiques de l'Université de Franche-Comté*. 4: 95-106.
- [91] Feng, G. and Kottenkof, C. (2006). "New Considerations for Vowel Nasalization Based on Separate Mouth-Nose Recording". *Actes de ICSLP, Pittsburgh*, 2242-2245.
- [92] Ferrara, E. (1980). "Fast implementation of LMS Adaptive filters." *IEEE Transactions on Acoustic, Speech and Signal Processing* 28: 474-475.
- [93] Flanagan, J. (1972). *Speech analysis, synthesis and perception* Springer-Verlag, New York.
- [94] Fonagy, I. and Fonagy, J. (1966). "Sound pressure level and duration." *Phonetica* 15 14-21.
- [95] Fougeron, C. and Keating, P. A. (1997). "Articulatory strengthening at edges of prosodic domains." *Journal of the Acoustic Society of America* 101(6): 3728-3740.
- [96] Fowler, C. A. (1986). "An event approach to the study of speech perception from a direct-realist perspective " *Journal of Phonetics* 14: 3-28.

- [97] Frank, T. and Tufts, J. (2003). "Speech production in noise with and without wearing hearing protection." *Journal of the Acoustic Society of America* 114(2): 1069-1080.
- [98] Fucci, D. (1972). "Oral vibrotactile sensation: an evaluation of normal and defective speakers." *Journal of Speech and Hearing Research* 15(1): 179-184.
- [99] Garber, S. R., Siegel, G. M. and Pick, H. L., Jr. (1981). "Regulation of vocal intensity in the presence of feedback filtering and amplification." *Journal of Speech and Hearing Research* 24(1): 104-108.
- [100] Garber, S. R., Speidel, T. M. and Siegel, G. M. (1980). "The effects of noise and palatal appliances on the speech of five-year-old children." *Journal of Speech and Hearing Research* 23(4): 853-862.
- [101] Garber, S. R., Speidel, T. M., Siegel, G. M., Miller, E. and Glass, L. (1980). "The effects of presentation of noise and dental appliances on speech." *Journal of Speech and Hearing Research* 23(4): 838-852.
- [102] Garnier, M., Bailly, L., Dohen, M., Welby, P. and Løevenbruck, H. (2006). "An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance". Actes de ICSLP, Pittsburgh, PA, USA, 17-22.
- [103] Garnier, M., Bailly, L., Dohen, M., Welby, P. and Løevenbruck, H. (2006). "Étude acoustique et articulatoire de la parole Lombard : Effets globaux sur l'énoncé entier". Actes de XXVIèmes Journées d'Etude sur la Parole, Dinard, France, 69-72.
- [104] Garnier, M., Dohen, M., Løevenbruck, H., Welby, P. and Bailly, L. (2006). "The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement?" Actes de 7th International Seminar on Speech Production Ubatuba, Brazil, 255-262.
- [105] Garnier, M., Dubois, D. and Henrich, N. (2006). "Constitution de corpus de parole semi-spontanée en environnement bruyant : intérêts et applications d'une telle méthodologie". Actes de Rencontres Jeunes Chercheurs de l'ED 268, Paris.
- [106] Garnier, M., Henrich, N., Dubois, D., Poitevineau, J., Castellengo, M. and Sotiropoulos, D. (2005). "Etude de la qualité vocale dans le chant lyrique." *Scolia* 20: 151-169.
- [107] Garnier, M., Henrich, N., Dubois, D. and Polack, J. D. (2006). "Est-il valide de considérer l'effet Lombard comme un phénomène linéaire en fonction du niveau de bruit ?" Actes de VIIIème Congrès Français d'Acoustique, Tours, France.
- [108] Gaudrain, E., Grimault, N., Healy, E. W. and Béra, J. C. (2006). "Ségrégation de séquences de voyelles avec ou sans simulation de perte auditive". Actes de Congrès Français d'Acoustique, Tours, France.
- [109] Gauffin, J. and Sundberg, J. (1989). "Spectral correlates of glottal voice source waveform characteristics." *Journal of Speech and Hearing Research* 32(3): 556-565.
- [110] Gauvain, J. L., Lamel, L. and Adda, G. (2002). "The Limsi Broadcast News Transcription System." *Speech Communication*, 37(1-2): 89-108.
- [111] Gelfer, M. P., Andrews, M. L. and Schmidt, C. P. (1991). "Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers." *Journal of Voice* 5(2): 158-167.
- [112] Gendrot, C. (2005). *Aspects perceptifs, physiologiques et acoustiques de différentes catégories prosodiques en français*. These de doctorat. Université Sorbonne Nouvelle, Paris.
- [113] Giovanni, A. (2006). "Les traumatismes vocaux." (A paraître, communication personnelle).

- [114] Giovanni, A. (2007). "Traits de personnalité et lésions nodulaires." (à paraître, communication personnelle).
- [115] Giovanni, A., Aumelas, E., Chapus, E., Lassalle, A., Remacle, M. and Ouaknine, M. (2004). "Le forçage vocal et ses conséquences." *Annales d'Otolaryngologie et de Chirurgie Cervicofaciale* 121: 187-196.
- [116] Goldin, A. A. (2004). "Close talking autodirective dual microphone". Actes de 116th Convention of the Audio Engineering society, Berlin, Germany.
- [117] Gotaas, C. and Starr, C. D. (1993). "Vocal fatigue among teachers." *Folia Phoniatrica (Basel)* 45(3): 120-129.
- [118] Graciarena, M., Franco, H., Myers, G., Cowan, C., Cesari, F. and Abrash, V. (2004). "Combination of standard and throat microphones for robust speech recognition in highly noisy environments". Actes de Interpseech, Jeju island, Korea.
- [119] Gramming, P. and Sundberg, J. (1988). "Spectrum factors relevant to phonetogram measurement." *Journal of the Acoustic Society of America* 83(6): 2352-2360.
- [120] Granqvist, S. (2003). "The self-to-other ratio applied as a phonation detector for voice accumulation." *STL-QPSR* 45: 23-33.
- [121] Grant, K. W. and Seitz, P. F. (2000). "The use of visible speech cues for improving auditory detection of spoken sentences." *Journal of the Acoustical Society of America* 108 (3): 1197-1208.
- [122] Gravano, A. and Hirschberg, J. (2006). "Effect of Genre, Speaker, and Word Class on the Realization of Given and New Information". Actes de ICSLP, Pittsburgh, USA.
- [123] Grieser, D. and Kuhl, P. K. (1989). "Categorization of Speech by Infants: Support for Speech-Sound Prototypes." *Developmental Psychology* 25(4): 577-588.
- [124] Grini, M. N., Ouaknine, M. and Giovanni, A. (1998). "Modifications posturales et segmentaires contemporaines du forçage vocal [Contemporary postural and segmental modification of forced voice]." *Revue de Laryngologie Otol Rhinol (Bordeaux)* 119(4): 253-257.
- [125] Grosjean, F. and Deschamps, A. (1975). "Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation." *Phonetica* 31: 144-184.
- [126] Guastavino, C. (2003). *Étude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain*. These de doctorat. Université Paris 6, Paris.
- [127] Guastavino, C., Katz, B., Dubois, D. and Polack, J. D. (2005). "Ecological Validity of Soundscape Reproduction." *Acta Acustica* 91.
- [128] Guerin, C. (2004). "De l'intérêt de la technique vocale pour une voix parlée tonique ou comment ne pas entrer dans le processus de forçage vocal." from <http://www.sosvoix.asso.fr/sosvoix/orth0014.htm>.
- [129] Guerin, C. (2006). "Gérer sa voix." from <http://www.sosvoix.asso.fr/sosvoix/GERERSAVOIX.htm>.
- [130] Guerrero Lopez, H. A. (2004). "Le forçage vocal chez l'enfant implanté cochléaire." from <http://icar.univ-lyon2.fr/membres/haguerrero/>.
- [131] Haddad, K. and Benoit, D. (2002). "Capabilities of a beamforming technique for acoustic measurements inside a moving car". Actes de Internoise, Dearborn, USA.

- [132] Hagerman, B. and Olofsson, A. (2004). "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise." *Acta Acustica United with Acustica* 90(2): 356-361.
- [133] Hanley, T. D. and Steer, M. D. (1949). "Effect of level of distracting noise upon speaking rate, duration and intensity." *Journal of Speech Disorders* 14(4): 363-368.
- [134] Hanson, B. A. and Applebaum, T. H. (1990). "Robust speaker-independent word recognition using static, dynamic and acceleration feature : experiments with monard and noisy speech." *IEEE Transactions on Speech and Audio Processing*: 857-860.
- [135] Hanson, H. M. (1997). "Glottal characteristics of female speakers: acoustic correlates." *Journal of the Acoustic Society of America* 101(1): 466-481.
- [136] Hashimoto, Y. and Sakai, K. L. (2003). "Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study." *Human Brain Mapping* 20(1): 22-28.
- [137] Haxer, M. J., Guinn, L. W. and Hogikyan, N. D. (2001). "Use of speech recognition software: a vocal endurance test for the new millennium?" *Journal of Voice* 15(2): 231-236.
- [138] Hemler, R. J., Wieneke, G. H. and Dejonckere, P. H. (1997). "The effect of relative humidity of inhaled air on acoustic parameters of voice in normal subjects." *Journal of Voice* 11(3): 295-300.
- [139] Henrich, N. (2001). *Etude de la source glottique en voix parlée et chantée*. These de doctorat. Université Paris 6, Paris.
- [140] Henrich, N. (2006). "Mirroring the voice from Garcia to the present day: some insights into singing voice registers." *Logopedics Phoniatrics Vocology* 31(1): 3-14.
- [141] Henrich, N., d'Alessandro, C., Doval, B. and Castellengo, M. (2005). "Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency." *The Journal of the Acoustical Society of America* 117(3): 1417-1430.
- [142] Henrich, N., Kiek, M., Smith, J. and Wolfe, J. (2006). "Resonance strategies used in Bulgarian women's singing style: a pilot study." *Logopedics Phoniatrics Vocology* in press.
- [143] Henton, C. G. and Bladon, R. A. W. (1985). "Breathiness in normal female speech: Inefficiency versus desirability." *Language and Communication* 5: 221-227.
- [144] Hess, M. M. and Ludwigs, M. (2000). "Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns." *Journal of Voice* 14(2): 255-271.
- [145] Hillenbrand, J., Cleveland, R. A. and Erickson, R. L. (1994). "Acoustic Correlates of Breathy Vocal Quality." *Journal of Speech and Hearing Research* 37: 769-778.
- [146] Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M. and Vaughan, C. (1989). "Objective assessment of vocal hyperfunction: an experimental framework and initial results." *Journal of Speech and Hearing Research* 32(2): 373-392.
- [147] Hirano, M. (1981). *Psycho-acoustic evaluation of voice: GRBAS scale for evaluating the hoarse voice*. Clinical Evaluation of Voice. Springer Verlag. Wien.
- [148] Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E. and Benson, R. W. (1952). "Development of materials for speech audiometry." *Journal of Speech and Hearing Disorders* 17(3): 321-337.
- [149] Hofler, H. (1984). "[Comparative long-term spectral studies of the voice using the Lombard effect]." *Laryngology Rhinology Otolaryngology (Stuttgart)* 63(11): 589-595.

- [150] Hogikyan, N. D. and Sethuraman, G. (1999). "Validation of an instrument to measure voice-related quality of life (V-RQOL)." *Journal of Voice* 13(4): 557-569.
- [151] Hoit, J. D. (1995). "Influence of body position on breathing and its implications for the evaluation and treatment of speech and voice disorders." *Journal of Voice* 9(4): 341.
- [152] Hollien, H. and Michel, J. F. (1968). "Vocal fry as a phonational register." *Journal of Speech and Hearing Research* 11(3): 600-604.
- [153] Holmberg, E. B., Hillman, R. E. and Perkell, J. S. (1988). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice." *Journal of the Acoustic Society of America* 84(2): 511-529.
- [154] Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C. and Goldman, S. L. (1995). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice." *Journal of Speech and Hearing Research* 38(6): 1212-1223.
- [155] Hood, J. D. (1962). "Bone conduction: a review of the present position with especial reference to the contributions of Dr. Georg von Békésy." *Journal of the Acoustic Society of America* 34: 1325-1332.
- [156] Houde, J. F. and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production." *Science* 279(5354): 1213-1216.
- [157] Houde, J. F. and Jordan, M. I. (2002). "Sensorimotor adaptation of speech I: Compensation and adaptation." *Journal of Speech, Language and Hearing Research* 45(2): 295-310.
- [158] Howard, D. M. (1995). "Variation of electroglottographically derived closed quotient for trained and untrained adult female singers." *Journal of Voice* 9(2): 163-172.
- [159] Howes, D. W. (1957). "On the relationship between intelligibility and the frequency of occurrence of English words." *Journal of the Acoustical Society of America* 29: 296-305.
- [160] Huang, D. Z., Minifie, F. D., Kasuya, H. and Lin, S. X. (1995). "Measures of vocal function during changes in vocal effort level." *Journal of Voice* 9(4): 429-438.
- [161] Hugonnet, C. and Walder, P. (1995). *Stereophonic Sound Recording*. John Wiley and Sons, West Sussex, England.
- [162] Husson, R. (1962). *Physiologie de la phonation*. Masson, Paris.
- [163] Isshiki, N. (1964). "Regulatory Mechanism Of Voice Intensity Variation." *Journal of Speech and Hearing Research* 128: 17-29.
- [164] Isshiki, N. (1981). *Vocal efficiency index*. *Vocal Fold Physiology*. T. U. Press. Stevens and Hirano. Tokyo: 193-207.
- [165] Jacobson, B. H., Johnson, A., Grywalski, C., Jacobson, G., Benninger, M. S. and Newman, C. W. (1997). "The voice handicap index (vhi): Development and validation." *American Journal of Speech-Language Pathology* 6(3): 66-70.
- [166] Javkin, H., Hanson, B. and Kaun, A. "The effects of breathy voice on intelligibility." *Speech Communication* 10(5-6): 539 -543
- [167] Johnson, K., Flemming, E. and Wright, R. (1993). "The Hyperspace Effect: Phonetic Targets Are Hyperarticulated." *Language* 69(3): 505-528.
- [168] Joliveau, E., Smith, J. and Wolfe, J., . (2004). "Vocal tract resonances in singing: the soprano voice." *Journal of the Acoustic Society of America* 116(2434-2439).

- [169] Jones, J. A. and Munhall, K. G. (2003). "Learning to produce speech with an altered vocal tract: the role of auditory feedback." *Journal of the Acoustic Society of America* 113(1): 532-543.
- [170] Jun, S. A. and Fougeron, C. (1995). "The accentual phrase and the prosodic structure of French". *Actes de ICPhS Stockholm, Sweden*, 722-725.
- [171] Junqua, J. (1993). "The lombard reflex and its role on human listener and automatic speech recognizers." *Journal of the Acoustic Society of America* 93(1): 510-524.
- [172] Junqua, J. (1996). "The influence of acoustics on speech production : a noise-induced stress phenomenon known as the Lombard reflex." *Speech Communication* 20: 13-22.
- [173] Junqua, J. C. (1992). "Acoustic and production pilot studies of speech vowels produced in noise". *Actes de ICSLP, Banff, Canada*.
- [174] Junqua, J. C. and Anglade, Y. (1990). "Acoustic and perceptual studies of lombard speech: application to isolated-words automatic speech recognition." *IEEE Transactions on Acoustic, Speech and Signal Processing*: 841-844.
- [175] Jurafsky, D., Bell, A., Gregory, M. and Raymond, W. D. (2001). *Probabilistic relations between words: Evidence from reduction in lexical production*. Frequency and the emergence of linguistic structure. J. Bybee and P. Hopper. Benjamins. Amsterdam.
- [176] Kadiri, N. (1998). *Conséquences d'un environnement bruyé sur la production de la parole*. These de doctorat. Université Paul Sabatier, Toulouse.
- [177] Kasuya, H., Ogawa, S., Mashima, K. and Ebihara, S. (1986). "Normalised noise energy as an acoustic measure to evaluate pathologic voice." *Journal of the Acoustical Society of America* 80: 1329-1334.
- [178] Keating, P. A., MacEachern, P., Shryock, A. and Dominguez, S. (1994). "A manual for Phonetic Transcription : Segmentation and Labeling of Words in Spontaneous Speech." *UCLA Working Papers in Phonetics* 88: 91-120.
- [179] Kim, J., Davis, C., Vignali, G. and Hill, H. (2005). "A visual concomitant of the Lombard reflex". *Actes de AVSP, Vancouver, Canada*, 17-21.
- [180] Kim, S. (2005). "Durational characteristics of korean Lombard speech". *Actes de ICSLP, Lisbon, Portugal*, 2901-2904.
- [181] Kitch, J. A. and Oates, J. (1994). "The perceptual features of vocal fatigue as self-reported by a group of actors and singers." *Journal of Voice* 8(3): 207-214.
- [182] Kitch, J. A., Oates, J. and Greenwood, K. (1996). "Performance effects on the voices of 10 choral tenors: acoustic and perceptual findings." *Journal of Voice* 10(3): 217-227.
- [183] Klatt, D. H. and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers." *Journal of the Acoustic Society of America* 87(2): 820-857.
- [184] Koopmans-van Beinum, F. (1980). *Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions*. These de doctorat. Universiteit van Amsterdam.
- [185] Korn (1954). "Effect of psychological feedback on conversational noise reduction in rooms." *Journal of the Acoustic Society of America* 26(5): 793-794.
- [186] Kostyk, B. E. and Rochet, A. P. (1998). "Laryngeal airway resistance in teachers with vocal fatigue: a preliminary study." *Journal of Voice* 12(3): 287-299.

- [187] Koufman, J. A. and Blalock, P. D. (1988). "Vocal fatigue and dysphonia in the professional voice user: Bogart-Bacall syndrome." *Laryngoscop* 98(5): 493-498.
- [188] Kryter, K. D. (1946). "Effect of ear protective devices on the intelligibility of speech in noise." *Journal of the Acoustic Society of America* 18(2): 413-417.
- [189] La Machhia, E. (2005). *Modifications posturales en relation avec le forçage vocal: études expérimentales sur des sujets sains*. Mémoire d'orthophonie. Université de la méditerranée, Marseille.
- [190] Lallouache, T. (1991). *Un poste "Visage-Parole" couleur. Acquisition et traitement automatique des contours des lèvres*. These de doctorat. INPG.
- [191] Lamprecht, A. (1988). "[Changes of the voice in the Lombard reflex]." *Laryngology Rhinology Otology (Stuttgart)* 67(7): 350-354.
- [192] Lane, H. and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech." *Journal of Speech and Hearing Research* 14: 677-709.
- [193] Lane, H. L., Catania, A. C. and Stevens, S. S. (1961). "Voice level: autophonic scale, perceived loudness and effects of sidetone." *Journal of the Acoustic Society of America* 33(2): 160-167.
- [194] Lane, H. L., Tranel, B. and Sisson, C. (1970). "Regulation of voice communication by sensory dynamics." *Journal of the Acoustic Society of America* 47(2): 618-624.
- [195] Larcher, V. and Polack, J. D. (2001). *Techniques de spatialisation des sons pour la réalité virtuelle* These de doctorat. Université Paris 6.
- [196] Larson, C. R., Burnett, T. A., Kiran, S. and Hain, T. C. (2000). "Effects of pitch-shift velocity on voice Fo responses." *Journal of the Acoustic Society of America* 107(1): 559-564.
- [197] Laukkanen, A. M., Jarvinen, K., Artkoski, M., Waaramaa-Maki-Kulmala, T., Kankare, E., Sippola, S., Syrja, T. and Salo, A. (2004). "Changes in voice and subjective sensations during a 45-min vocal loading test in female subjects with vocal training." *Folia Phoniatica et Logopaedica* 56(6): 335-346.
- [198] Laukkanen, A. M., Lindholm, P. and Vilkmán, E. (1995). "On the effects of various vocal training methods on glottal resistance and efficiency." *Folia Phoniatica Logopedica* 47: 324-330.
- [199] Laukkanen, A. M., Lindholm, P., Vilkmán, E., Haataja, K. and Alku, P. (1996). "A physiological and acoustic study on voiced bilabial fricative/beta:/as a vocal exercise." *Journal of Voice* 10(1): 67-77.
- [200] Laukkanen, A. M., Mickelson, N. P., Laitala, M., Syrja, T., Salo, A. and Sihvo, M. (2004). "Effects of HearFones on speaking and singing voice quality." *Journal of Voice* 18(4): 475-487.
- [201] Lauri, E. R., Alku, P., Vilkmán, E., Sala, E. and Sihvo, M. (1997). "Effects of prolonged oral reading on time-based glottal flow waveform parameters with special reference to gender differences." *Folia Phoniatica et Logopaedica* 49(5): 234-246.
- [202] Le Huche, F. (1984). *La voix*. Masson, Paris.
- [203] Lee, B. (1950). "Effects of delayed speech feedback." *Journal of the Acoustic Society of America* 22(6): 824-827.
- [204] Lee, B., Hasegawa-Johnson, M., Goudeseune, C., Kamdar, S., Borys, S., Liu, M. and Huang, T. (2004). "AVICAR: Audio-Visual Speech Corpus in a Car Environment". Actes de ICSLP, Jeju Island, Korea.

- [205] Legent , F., Perlemutier, L. and Vanderbrouck, C. (1975). *Cahiers d'anatomie O.R.L.* . Masson, Paris.
- [206] Lehto, L., Rantala, L., Vilkmán, E., Alku, P. and Backström, T. (2003). "Experiences of a short vocal training course for call-centre customer service advisors." *Folia Phoniatrica et Logopaedica* 55(4): 163-176.
- [207] Leino, T. (1993). "Long-term average spectrum study on speaking voice quality in male actors". Actes de SMAC, Stockholm, Sweden, 206-210.
- [208] Leipp, E. (1977). *La machine à écouter: essai de psycho-acoustique*. Masson, Paris.
- [209] Leothaud, G. (2005). "Théorie de la phonation." from <http://www.paris4.sorbonne.fr/e-cursus/texte/CEC/Gleothaud/Th%E9orie%20de%20la%20Phonation.pdf>.
- [210] Levelt, W. and Cutler, A. (1983). "Prosodic marking in speech repair." *Journal of Semantics* 2(1): 205-218.
- [211] Leydon, C., Bauer, J. J. and Larson, C. R. (2003). "The role of auditory feedback in sustaining vocal vibrato." *Journal of the Acoustic Society of America* 114(3): 1575-1581.
- [212] Liberman, G. and Mattingly, A. M. (1985). "The motor theory of speech perception revised." *Cognition* 21: 1-36.
- [213] Lieberman, P., Katz, W., Jongman, A., Zimmerman, R. and Miller, M. (1985). " Measures of the sentence intonation of read and spontaneous speech in American English." *Journal of the Acoustic Society of America* 77: 649-657.
- [214] Liebermann, P. (1961). "Perturbation in vocal pitch." *Journal of the Acoustic Society of America* 33: 697-703.
- [215] Lienard, J. S. and Di Benedetto, M. G. (1999). "Effect of vocal effort on spectral properties of vowels." *Journal of the Acoustic Society of America* 106(1): 411-422.
- [216] Lindblom, B. (1963). "Spectrographic Study of Vowel Reduction." 35(5): 783.
- [217] Lindblom, B. (1990). *Explaining phonetic variation: a sketch of the H&H theory*. Speech production and speech modeling. Dordrecht. Kluwer. : 403-439.
- [218] Lindblom, B. (1996). "Role of articulation in speech perception: clues from production." *Journal of the Acoustic Society of America* 99(3): 1683-1694.
- [219] Lindblom, B., Brownlee, S., Davis, B. and Moon, S. J. (1992). "Speech transforms." *Speech Communication* 11: 357-368.
- [220] Lindblom, B. E. and Sundberg, J. E. (1971). "Acoustical consequences of lip, tongue, jaw, and larynx movement." *Journal of the Acoustic Society of America* 50(4): 1166-1179.
- [221] Lindblom, B. E. F. and Studdert-Kennedy, M. (1967). "On the Rôle of Formant Transitions in Vowel Recognition." *Journal of the Acoustic Society of America* 42(4): 830-843.
- [222] Linville, S. E. (1995). "Changes in glottal configuration in women after loud talking." *Journal of Voice* 9(1): 57-65.
- [223] Loevenbruck, H. (2000). "Effets articulatoires de l'emphase contrastive sur la Phrase Accentuelle en français". Actes de XXIIIèmes Journées d'Etude sur la Parole, Aussois, France, 165-168.
- [224] Lombard, E. (1911). "Le signe de l'elevation de la voix." *Annales des maladies de l'oreille et du larynx* 37: 101-119.

- [225] Long, J., Williford, H. N., Olson, M. S. and Wolfe, V. (1998). "Voice problems and risk factors among aerobics instructors." *Journal of Voice* 12(2): 197.
- [226] Lozanski, L. (2004). "Les troubles de la voix." *Bulletin de santé et de sécurité au travail* (4): <http://www.caut.ca/fr/publications/healthandsafety/4-voix.pdf>
- [227] Maclagan, M. A. and Welham, N. V. (2004). "Vocal fatigue in young trained singers across a solo performance: a preliminary study." *Logopedics Phoniatrics Vocology* 29(1): 3-12.
- [228] Maeda, S. (1990). *Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tractshapes using an articulatory model*. Speech Production and Speech Modelling Hardcastle and Marchal. Kluwer Academic. 131-149.
- [229] Maeda, S. and Honda, K. (1994). "From EMG to formant patterns of vowels: the implication of vowel spaces." *Phonetica* 51(1-3): 17-29.
- [230] Maffiolo, V. (1999). *Caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain*. These de doctorat. Université du Maine, Le Mans.
- [231] Manabe, K., Sadr, E. I. and Dooling, R. J. (1998). "Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard effect." *Journal of the Acoustical Society of America* 103(2): 1190-1198.
- [232] Mann, E. A., McClean, M. D., Gurevich-Uvena, J., Barkmeier, J., McKenzie-Garner, P., Paffrath, J. and Patow, C. (1999). "The effects of excessive vocalization on acoustic and videostroboscopic measures of vocal fold condition." *Journal of Voice* 13(2): 294-302.
- [233] Margolis, R. H. and Popelka, G. R. (1975). "Loudness and the acoustic reflex." *Journal of the Acoustic Society of America* 58(6): 1330-1332.
- [234] Maurer, D. and Landis, T. (1990). "Role of bone conduction in the self-perception of speech." *Folia Phoniatrica (Basel)* 42(5): 226-229.
- [235] McCabe, S. L. and Denham, M. J. (1997). "A model of auditory streaming." *Journal of the Acoustic Society of America* 101: 1611-1621.
- [236] McHugh-Munier, C., Scherer, K. R., Lehmann, W. and Scherer, U. (1997). "Coping strategies, personality, and voice quality in patients with vocal fold nodules and polyps." *Journal of Voice* 11(4): 452-461.
- [237] Meltzner, G. and Hillman, R. (2003). "Impact of abnormal acoustic properties on the perceived quality of electrolaryngeal speech". *Actes de ITRW VOQUAL*, 73-78.
- [238] Meunier, C., Espesser, R. and Frenck-Mestre, C. (2006). "Aspects phonologique et dynamique de la distinctivité au sein des système vocaliques: une étude inter-langue". *Actes de XXVIème Journées d'Etude sur la Parole, Dinard*, 333-336.
- [239] Milbrath, R. L. and Solomon, N. P. (2003). "Do vocal warm-up exercises alleviate vocal fatigue?" *Journal of Speech, Language and Hearing Research* 46(2): 422-436.
- [240] Miller, D. G. and Schutte, H. K. (1990). "Formant tuning in a professional baritone." *Journal of Voice* 4(3): 231-237.
- [241] Miller, G. A., Heise, G. A. and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials." *Journal of Experimental Psychology*(41): 328-335.
- [242] Miller, G. A. and Nicely, P. E. (1955). "An Analysis of Perceptual Confusions Among Some English Consonants." *Journal of the Acoustic Society of America* 27: 338-352

- [243] Miller, R. (1990). *La structure du chant: pédagogie systématique de l'art du chant*. Ed. Ipcm, Paris.
- [244] Mixdorff, H., Grauwinkel, K. and Vainio, M. (2006). "Time-domain noise subtraction applied in the analysis of Lombard speech". Actes de Speech prosody, Dresden.
- [245] Mokbel, C. (1992). *Reconnaissance de la parole dans le bruit: bruitage / débruitage*. These de doctorat. Ecole nationale supérieure des télécommunications, Paris.
- [246] Morrison, M. (1997). "Pattern recognition in muscle misuse voice disorders: How I do it." *Journal of Voice* 11(1): 108.
- [247] Motel, T., Fisher, K. V. and Leydon, C. (2003). "Vocal warm-up increases phonation threshold pressure in soprano singers at high pitch." *Journal of Voice* 17(2): 160-167.
- [248] Murakami, T. and Yoshihisa, I. (2006). "Adaptive Filtering for Attenuating Musical Noise Caused by Spectral Subtraction". Actes de ICSLP, Pittsburgh, 1443-1446.
- [249] Nabelek, A. K. (1990). "Parameters governing speech intelligibility." *Environment international* 16: 399-404.
- [250] Nawka, T., Anders, L. C., Cebulla, M. and Zurakowski, D. (1997). "The speaker's formant in male voices." *Journal of Voice* 11(4): 422-428.
- [251] Neely, K. K. (1956). "Effects of visual factors on the intelligibility of speech." *Journal of the Acoustic Society of America* 28: 1275-1277.
- [252] Neils, L. R. and Yairi, E. (1987). "Effects of speaking in noise on vocal fatigue and vocal recovery." *Folia Phoniatica (Basel)* 39(2): 104-112.
- [253] Nelson, W. L. (1983). "Physical principles for economies of skilled movements." *Biological Cybernetics* 46(2): 135-147.
- [254] New, B. (2006). "Lexique 3 : Une nouvelle base de données lexicales." Actes de TALN, Louvain, Belgique, <http://www.lexique.org/>.
- [255] Nicely, P. E. and Miller, G. A. (1955). "An analysis of perceptual confusions among some english consonants." *Journal of the Acoustic Society of America* 27: 338-352.
- [256] Nonaka, S., Takahashi, R., Enomoto, K., Katada, A. and Unno, T. (1997). "Lombard reflex during PAG-induced vocalization in decerebrate cats." *Neuroscience Research* 29(4): 283-289.
- [257] Novak, A., Dlouha, O., Capkova, B. and Vohradnik, M. (1991). "Voice fatigue after theater performance in actors." *Folia Phoniatica (Basel)* 43(2): 74-78.
- [258] Novak, A. and Vokral, J. (1995). "Acoustic parameters for the evaluation of voice of future voice professionals." *Folia Phoniatica Logopedica* 47: 279-285.
- [259] Ohde, R. N. (1988). "Revisiting stop-consonant perception for two-formant stimuli." *Journal of the Acoustic Society of America* 84(4): 1551-1555.
- [260] Oliveira Barrichelo, V. M., Heuer, R. J., Dean, C. M. and Sataloff, R. T. (2001). "Comparison of singer's formant, speaker's ring, and LTA spectrum among classical singers and untrained normal speakers." *Journal of Voice* 15(3): 344-350.
- [261] Olsen, C. C. and Brandt, J. F. (1976). "Middle ear muscle activity during speech in stapedectomized and laryngectomized subjects." *Journal of the American Audiology Society* 1(5): 215-220.

- [262] Ormezzano, Y. (2000). *Le guide de la voix*. O. Jacob, Paris.
- [263] Pabon, J. P. and Plomp, R. (1988). "Automatic phonetogram recording supplemented with acoustical voice-quality parameters." *Journal of Speech and Hearing Research* 31(4): 710-722.
- [264] Pabon, P. (2006). "Voice Profiler. Manual." from <http://www.let.uu.nl/~Peter.Pabon/personal/OtherActivities/VoiceProfiler/>.
- [265] Papon, L. (2006). *Modifications posturales segmentaires contemporaines du forçage vocal*. Mémoire d'orthophonie. Université de la méditerranée, Marseille.
- [266] Peng, S. H. and Beckman, M. E. "Annotation conventions and corpus design in the investigation of spontaneous speech prosody in Taiwanese." from <http://www.ling.ohio-state.edu/>.
- [267] Perkell, J. S., Klatt, D. and K.N., S. (1986). *Invariance and Variability in Speech Processes*. Erlbaum, Hillsdale.
- [268] Perrier, P. (2005). "Control and representation in speech production." *ZAS Papers in linguistics* 40: 109-132.
- [269] Picheny, M. A., Durlach, N. I. and Braidà, L. D. (1985). "Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech." *Journal of Speech and Hearing Research* 28: 96-103.
- [270] Picheny, M. A., Durlach, N. I. and Braidà, L. D. (1986). "Speaking Clearly for the Hard of Hearing II : Acoustic Characteristics of Clear and Conversational Speech" *Journal of Speech and Hearing Research* 29: 434-446.
- [271] Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R. and Kearney, J. K. (1989). "Inhibiting the Lombard effect." *Journal of the Acoustic Society of America* 85(2): 894-900.
- [272] Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C. and Yuchtman, M. (1985). "Some acoustic-phonetic correlates of speech produced in noise". *Actes de ICASSP, Tampa, Florida, 1581-1584*.
- [273] Polack, J. D. and Pelorson, X. (1988). "Spatial Impression evaluation with omnidirectional microphones". *Actes de Acoustics'88, Cambridge, 383-389*.
- [274] Pörschmann, C. (2000). "Influences of bone conduction and air conduction on the sound of one's own voice." *Acta Acustica* 86: 1038-1045.
- [275] Portmann, M. and Portmann, C. (1959). *Précis d'audiométrie clinique*. Masson, Paris.
- [276] Prado, J. and Moulines, E. (1994). "Frequency-domain adaptive filtering with applications to acoustic echo cancellation." *Ann. Telecommun* 49(7-8): 414-428.
- [277] Pruszevicz, A., Jassem, W. and Waclawik, W. (1974). "Effect of noise on some acoustical parameters of speech." *Folia Phoniatrica* 26: 307-318.
- [278] Raimbault, M. (2002). *Caractérisation fonctionnelle d'un simulateur des ambiances sonores urbaines*. These de doctorat. Université de Nantes, Ecole d'Architecture de Nantes.
- [279] Rantala, L., Paavola, L., Korkko, P. and Viikman, E. (1998). "Working-day effects on the spectral characteristics of teaching voice." *Folia Phoniatrica et Logopaedica* 50(4): 205-211.
- [280] Rantala, L. and Viikman, E. (1999). "Relationship between subjective voice complaints and acoustic parameters in female teachers' voices." *Journal of Voice* 13(4): 484-495.

- [281] Rantala, L., Viikman, E. and Bloigu, R. (2002). "Voice changes during work: subjective complaints and objective measurements for female primary and secondary schoolteachers." *Journal of Voice* 16(3): 344-355.
- [282] Raphael, B. N. and Scherer, R. (1987). "Voice modifications of stage actors: acoustic analyses." *Journal of Voice* 1(1): 83-87.
- [283] Revis, J. (2004). *L'analyse perceptive des dysphonies*. Le bilan d'une dysphonie. Solal.
- [284] Rialland, A. (2003). *La syllabe ou la more en tonologie africaine, ou comment se fait l'interface entre segments et tons*. Phonologie : Champs et Perspectives. J. P. Angoujard and S. Wauquier. Presses de l'ENS. Lyon: 105-130.
- [285] Rialland, A. (2005). "Phonological and phonetic aspects of whistled languages." *Phonology* 22(2): 237-271.
- [286] Ringel, R. L., House, A. S., Burk, K. W., Dolinsky, J. P. and Scott, C. M. (1970). "Some relations between orosensory discrimination and articulatory aspects of speech production." *Journal of Speech and Hearing Research* 35(1): 3-11.
- [287] Rolland, G. and Loevenbruck, H. (2002). "Characteristics of the Accentual Phrase in French: an Acoustic, Articulatory and Perceptual Study". *Actes de Speech Prosody*, Aix-en-Provence, France, 611-614.
- [288] Rostolland, D. (1982). "Acoustic features of shouted voice." *Acta Acustica* 50: 118-125.
- [289] Rostolland, D. (1982). "Phonetic structure of shouted voice." *Acta Acustica* 51: 80-89.
- [290] Rostolland, D. and Parant, C. (1973). "Distorsion and intelligibility of shouted voice". *Actes de Symposium Speech Intelligibility* Liège, 293-304.
- [291] Rothman, H. B., Brown, W. S., Jr. and LaFond, J. R. (2002). "Spectral changes due to performance environment in singers, nonsingers, and actors." *Journal of Voice* 16(3): 323-332.
- [292] Roubeau, B., Castellengo, M., Bodin, P. and Ragot, M. (2004). "Phonétogramme par registre laryngé." *Folia Phoniatrica et Logopaedica* 56(5): 321-333.
- [293] Roubeau, B., Chevré-Muller, C., Arabia, C. and Arragon, C. (1993). "[Change in laryngeal vibratory mechanism: a physiological entity]." *Revue de Laryngologie Otologie Rhinologie (Bordeaux)* 114(4): 267-274.
- [294] Roy, J. P. (2005). "Visual perception of anticipatory rounding gestures in French". *Actes de ICSLP*, Lisbon, Portugal.
- [295] Roy, N., Bless, D. M. and Heisey, D. (2000). "Personality and voice disorders: A multitrait-multidisorder analysis." *Journal of Voice* 14(4): 521.
- [296] Rumsey, H. S. (1950). "Voice strain." *Journal of Laryngology and Otology* 64(11): 708-713.
- [297] Russell, A., Oates, J. and Greenwood, K. M. (1998). "Prevalence of voice problems in teachers." *Journal of Voice* 12(4): 467-479.
- [298] Safran, M. R., Seaber, A. V. and Garrett, W. E., Jr. (1989). "Warm-up and muscular injury prevention. An update." *Sports Med* 8(4): 239-249.
- [299] Sander, E. K. and Ripich, D. E. (1983). "Vocal fatigue." *Annals of Otology Rhinology and Laryngology* 92(2 Pt 1): 141-145.
- [300] Sarfati, J. (1998). *Soigner la voix*. Ed. Solal, Paris.

- [301] Savariaux, C. and Perrier, P. (1995). "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production"
" Journal of the Acoustic Society of America 98(5): 2428-2442.
- [302] Savariaux, C., Perrier, P., Orliaguet, J. P. and Schwartz, J. L. (1999). "Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis." Journal of the Acoustic Society of America 106(1): 381-393.
- [303] Schafer, A. J., Speer, S. R., Warren, P. and White, S. D. (2001). "Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task". Actes de Fourteenth Annual CUNY Conference on Human Sentence Processing, Philadelphia.
- [304] Schulman, R. (1989). "Articulatory dynamics of loud and normal speech." Journal of the Acoustic Society of America 85(1): 295-312.
- [305] Schultz-Coulon, H. J. and Fues, C. P. (1976). "[The lombard reflex as a test of vocal function (author's transl)]." Hno 24(6): 200-204.
- [306] Schutte, H. K. and Seidner, W. (1983). "Recommendation by the Union of European Phoniaticians (UEP): standardizing voice area measurement/phonetography." Folia Phoniatica (Basel) 35(6): 286-288.
- [307] Schwartz, J. L. (2004). "La parole multisensorielle : Plaidoyer, problèmes, perspective". Actes de XXVème Journées d'Etudes sur la Parole, Fès, Maroc, 11-18.
- [308] Schwartz, J. L., Abry, C., Boë, L. J. and Cathiard, M. (2002). *Phonology in a theory of perception-for-action-control* Phonology: from Phonetics to Cognition. J. Durand and B. Laks. Oxford University Press. Oxford: 255-280.
- [309] Schwartz, J. L., Berthommier, F. and Savariaux, C. (2004). "Seeing to hear better: evidence for early audio-visual interactions in speech identification." Cognition 93(2): B69-78.
- [310] Schwartz, J. L. and Escudier, P. (1989). "A strong evidence for existence of a large scale integrated spectral representation in vowel perception." Speech Communication 8: 235-259.
- [311] Scotto di Carlo, N. (194). "Internal voice sensitivities in opera singers." Folia Phoniatica Logopedica 46: 79-85.
- [312] Selkirk, E. (1996). *The prosodic structure of function words* Signal to Syntax. M. Demuth. 187-213.
- [313] Sherman, D. and Jensen, P. J. (1962). "Harshness and oral-reading time." Journal of Speech and Hearing Disorders 27: 172-177.
- [314] Shi, R., Weker, J. and Morgan, J. (1999). "Newborn infants sensitivity to perceptual cues to lexical and grammatical words." Cognition 72: B11-B21.
- [315] Siegel, G. M. and Pick, H. L., Jr. (1974). "Auditory feedback in the regulation of voice." Journal of the Acoustic Society of America 56(5): 1618-1624.
- [316] Siegel, G. M., Pick, H. L., Jr., Olsen, M. G. and Sawin, L. (1976). "Auditory feedback in the regulation of vocal intensity of preschool children." Dev. Psychol.(12): 255-261.
- [317] Sihvo, M. and Sala, E. (1996). "Sound level variation findings for pianissimo and fortissimo phonations in repeated measurements." Journal of Voice 10(3): 262-268.
- [318] Sinnott, J. M., Stebbins, W. C. and Moody, D. B. (1975). "Regulation of voice amplitude by the monkey." Journal of the Acoustic Society of America 58: 412-414.

- [319] Sluijter, A. M. C. v. H., V.J. (1996). "Spectral balance as an acoustic correlate of linguistic stress." *Journal of the Acoustical Society of America* 100(4): 2471-2485.
- [320] Smith, E., Kirchner, H. L., Taylor, M., Hoffman, H. and Lemke, J. H. "Voice problems among teachers: Differences by gender and teaching characteristics." *Journal of Voice* 12: 328-334.
- [321] Sodersten, M., Granqvist, S., Hammarberg, B. and Szabo, A. (2002). "Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural DAT recordings." *Journal of Voice* 16(3): 356-371.
- [322] Sodersten, M., Ternstrom, S. and Bohman, M. (2005). "Loud speech in realistic environmental noise: phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers." *Journal of Voice* 19(1): 29-46.
- [323] Solomon, N. P. and DiMattia, M. S. (2000). "Effects of a vocally fatiguing task and systemic hydration on phonation threshold pressure." *Journal of Voice* 14(3): 341-362.
- [324] Sonninen, A., Damste, P. H., Jol, J. and Fokkens, J. (1972). "On vocal strain." *Folia Phoniatica (Basel)* 24(5): 321-336.
- [325] Sonninen, A. and Hurme, P. (1998). "Vocal fold strain and vocal pitch in singing: radiographic observations of singers and nonsingers." *Journal of Voice* 12(3): 274-86.
- [326] Soo, J. (1990). "Multidelay block frequency domain adaptive filters." *IEEE Trans on ASSP* 38: 373-376.
- [327] Stanton (1988). "Acoustics-Phonetic analysis of loud and Lombard speech in simulated cockpit conditions." 331-334.
- [328] Stemple, J. C., Lee, L., D'Amico, B. and Pickup, B. (1994). "Efficacy of vocal function exercises as a method of improving voice production." *Journal of Voice* 8(3): 271-278.
- [329] Stemple, J. C., Stanley, J. and Lee, L. (1995). "Objective measures of voice production in normal subjects following prolonged voice use." *Journal of Voice* 9(2): 127-33.
- [330] Stenfelt, S. and Hakansson, B. (2002). "Air versus bone conduction: an equal loudness investigation." *Hearing Research* 167(1-2): 1-12.
- [331] Stephenson, L. S. (2004). *An electropalatographic and acoustic analysis of frequency effects in the lexicon*. These de doctorat. Mcquarie university, Sydney, Australia.
- [332] Stone, R. E., Jr. and Sharf, D. J. (1973). "Vocal change associated with the use of atypical pitch and intensity levels." *Folia Phoniatica (Basel)* 25(1): 91-103.
- [333] Stuart, A., Kalinowski, J., Rastatter, M. P. and Lynch, K. (2002). "Effect of delayed auditory feedback on normal speakers at two speech rates." *Journal of the Acoustic Society of America* 111(5 Pt 1): 2237-2241.
- [334] Sulter, A. M., Wit, H. P., Schutte, H. K. and Miller, D. G. (1994). "A structure approach to voice range profile (Phonetogram) analysis." *Journal of Speech and Hearing Research* 37: 1076-1085.
- [335] Sumbly, H. and Pollack, I. W. (1954). "Visual Contribution to Speech Intelligibility in Noise." *Journal of the Acoustic Society of America* 26: 212-215.
- [336] Sundberg, J. (1970). "Formant structure and articulation of spoken and sung vowels." *Folia Phoniatica (Basel)* 22(1): 28-48.
- [337] Sundberg, J. (1972). "An articulatory interpretation of the singing formant." *STL-QPSR* 1: 45-53.

- [338] Sundberg, J., Titze, I. and Scherer, R. (1993). "Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source." *Journal of Voice* 7(1): 15-29.
- [339] Sussman, J. E. (1993). "A preliminary test of prototype theory for a [ba]-to-[da] continuum (A)." *Journal of the Acoustic Society of America* 93(4): 2392.
- [340] Szabo, A., Hammarberg, B., Granqvist, S. and Sodersten, M. (2003). "Methods to study pre-school teachers' voice at work: simultaneous recordings with a voice accumulator and a DAT recorder." *Logopedics Phoniatrics Vocology* 28(1): 29-39.
- [341] Tabain, M. (2003). "Effects of prosodic boundary on /aC/ sequences: articulatory results." *Journal of the Acoustic Society of America* 113(5): 2834-2849.
- [342] Takizawa, Y. and Hamada, M. (1990). "Lombard Speech Recognition by Formant-Frequency-Shifted LPC Cepstrum". *Actes de ICSLP, Kobe, Japan*, 293-296.
- [343] Tarnaud, J. (1935). *Le nodule de la corde vocale*. Maloine, Paris.
- [344] Ternström, S. (1994). "Hearing myself with others: sound levels in choral performance measured with separation of one's own voice from the rest of the choir." *Journal of Voice* 8(4): 293-302.
- [345] Ternström, S., Bohman, M. and Sodersten, M. (2006). "Loud speech over noise: Some spectral attributes, with gender differences." *Journal of the Acoustical Society of America* 119(3): 1648-1665.
- [346] Ternström, S., Sodersten, M. and Bohman, M. (2002). "Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure." *Journal of Voice* 16(2): 195-206.
- [347] Ternström, S., Sundberg, J. and Colden, A. (1988). "Articulatory Fo perturbations and auditory feedback." *Journal of Speech and Hearing Research* 31(2): 187-192.
- [348] Teston, B. (2001). "L'évaluation objective des dysfonctionnements de la voix et de la parole. 2ème partie : les dysphonies." *Travaux Interdisciplinaires du Laboratoire Parole et Langage* 20: 169-232.
- [349] Thunberg, G. (2003). "Balancing spectra between different speaking styles." *Phonum* 9: 137-140.
- [350] Titze, I. R. (1984). *Vocal fatigue: some biomechanical considerations*. *Transcripts of the Twelfth Symposium: Care of the professional voice*. V. Foundation. Lawrence V. ed. New York: 97-104.
- [351] Titze, I. R. (1992). "Acoustic interpretation of the voice range profile (phonetogram)." *Journal of Speech and Hearing Research* 35(1): 21-34.
- [352] Tomatis, A. (1987). *L'oreille et la voix*. Robert Laffont, Paris.
- [353] Tonkinson, S. (1994). "The Lombard Effect in Choral Singing." *Journal of Voice* 8(1): 24-29.
- [354] Tonndorf, J. (1966). "Bone conduction. Studies in experimental animals." *Acta Otolaryngologica: Suppl* 213.
- [355] Traunmüller, H. (1981). "Perceptual dimension of openness in vowels." *Journal of the Acoustic Society of America* 69(5): 1465-1475.
- [356] Traunmüller, H. and Eriksson, A. (2000). "Acoustic effects of variation in vocal effort by men, women, and children." *Journal of the Acoustic Society of America* 107(6): 3438-3451.

- [357] Vaissière, J. (1983). *Language Independent Prosodic Features*. Prosody: Models and Measurements. A. Cutler and R. Ladd. Springer Verlag. 53-65.
- [358] Vaissière, J. (1997). "Phonological use of the larynx: a tutorial". Actes de Larynx 97, Marseille, 115-126.
- [359] Vaissière, J. (2006). *La phonétique*. Presses Universitaires de France, Paris.
- [360] Van Heusden, E., Plomp, R. and Pols, L. C. W. (1979). "Effect of ambient noise on the vocal output and the preferred listening level of conversational speech." Applied Acoustics 12(1): 31.
- [361] Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. and Stokes, M. A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses." Journal of the Acoustic Society of America 84(3): 917-928.
- [362] Vaxelaire, B. and Sock, R. (1998). *Perturbing phonetic gestures using speech rate : X-ray and acoustic data from french*. Oralité et gestualité. Communication multimodale, interaction. S. Santi, I. Guaitella, C. Cavé and G. Konopczynski. L'Harmattan. Paris: 447-453.
- [363] Verdolini, K., Titze, I. R. and Fennell, A. (1994). "Dependence of phonatory effort on hydration level." Journal of Speech and Hearing Research 37(5): 1001-1007.
- [364] Verstraete, J., Forrez, G., Mertens, P. and Debruyne, F. (1993). "The effect of sustained phonation at high and low pitch on vocal jitter and shimmer." Folia Phoniatrica (Basel) 45(5): 223-228.
- [365] Vilkman, E. (2004). "Occupational safety and health aspects of voice and speech professions." Folia Phoniatrica et Logopaedica 56(4): 220-253.
- [366] Vilkman, E., Lauri, E. R., Alku, P., Sala, E. and Sihvo, M. (1997). "Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions." Folia Phoniatrica et Logopaedica 49(5): 247-263.
- [367] Vilkman, E., Lauri, E. R., Alku, P., Sala, E. and Sihvo, M. (1999). "Effects of prolonged oral reading on F0, SPL, subglottal pressure and amplitude characteristics of glottal flow waveforms." Journal of Voice 13(2): 303-312.
- [368] Vinturi (2001). *Studies on voice production with a special emphasis on vocal loading, gender, some exposure factors and intensity regulation*. These de doctorat. University of Helsinki.
- [369] Vogel, C. (1999). *Etude sémiotique et acoustique de l'identification des signaux sonores d'avertissement en contexte urbain* These de doctorat. Université Paris VI.
- [370] Von Bekesy, G. (1960). *Conductive processes*. Experiments in hearing Mc Graw Hill. New York.
- [371] Wauquier-Gravelines, S. (1999). "Segmentation lexicale de la parole continue : la linéarité en question." Recherches linguistiques de Vincennes 28: la linéarité: 133-156.
- [372] Webster, R. K. J. (1962). "Effects of ambient noise and nearby talkers on a face-to-face communication task." Journal of the Acoustic Society of America 34(7): 936-941.
- [373] Welby, P. (2003). "French intonational rises and their role in speech segmentation". Actes de Eurospeech, Geneva, 2125-2128.
- [374] Welby, P. (2005). "Pitch changes in noisy conditions: Data from French". Actes de Conference Making Europe More Attractive for Researchers, Pisa/Livorno, Italy, 28-30.
- [375] Welby, P. (2006). "Changements intonatifs dans la parole Lombard: au delà de l'étendue de F0". Actes de XXVIèmes Journées d'Etude sur la Parole Dinard, France, 271-274.

- [376] Welby, P. (2006). "French intonational structure: Evidence from tonal alignment." *The Journal of Phonetics* 34(3): 343-371.
- [377] Welby, P. (2007). "The role of early fundamental frequency rises and elbows in French word segmentation." *Speech Communication* 49(1): 28-48.
- [378] Wenk, G. and Wiolland, F. (1982). "Is French really syllable-timed?" *Journal of Phonetics* 10: 177-193.
- [379] Yanagihara, N. (1967). "Significance of harmonic changes and noise components in hoarseness." *Journal of Speech and Hearing Research* 10: 531-541.
- [380] Yanagisawa, E., Estill, J., Mambrino, L. and Talkin, D. (1991). "Supraglottic contributions to pitch raising. Videoendoscopic study with spectroanalysis." *Annals of Oto Rhino Laryngology* 100(1): 19-30.
- [381] Yumoto, E., Gould, W. J. and Baer, T. (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness." *Journal of the Acoustic Society of America* 71(6): 1544-1549.
- [382] Zeiliger, J., Serignat, J. F., Autresserre, D. and Meunier, C. (1994). "BD_Bruit, une base de données de parole de locuteurs soumis à du bruit". *Actes de Xèmes Journées d'Etude de la Parole, Grenoble, France, 287-290.*
- [383] Zeine, L. and Brandt, J. F. (1988). "The Lombard effect on alaryngeal speech." *Journal of Communication Disorders* 21(5): 373-383.
- [384] Zimpfer, V. and Buck, K. (2006). "Bouchon d'oreille à atténuation active adaptée aux bruits d'hélicoptère". *Actes de Congrès français d'Acoustique, Tours, 1045-1049.*

