2 - SYNTHESE DE LA PAROLE A L'ICOPHONE (1964-1974)

1. LES FORMES DE LA PAROLE

Groupe de travail : E. Leipp, J.S. Lienard, J. Sapaly, puis avec la numerisation, D. Teil, A. Calinet, M. Mlouka.

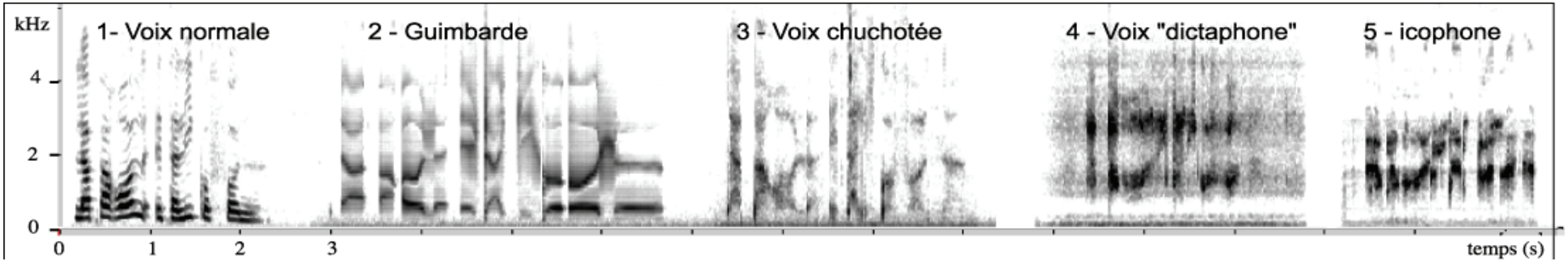
Cette recherche naît dans une période d'exploration systématique de l'univers sonore à l'aide du sonagraphe. L'appareil founit une spectrographie temporelle des signaux acoustiques, donnant ainsi accès à la notion de "forme sonore temporelle". La démarche adoptée est originale et se démarque de celle des autres équipes travaillant sur la synthèse de parole, qui privilégient la mesure précise des paramètres du son, et se fondent sur le concept du "phonème". Nous avons de fréquents contacts avec A.Moles[1] qui a donné en 1964 un séminaire GAM sur la théorie de l'information et la théorie de la forme en 1964.

Conviés à une soutenance de thèse portant sur l'analyse sonagraphique d'un texte complet filmé[2] à destination des sourds-muets, nous sommes frappés par notre incapacité à "décoder" les spectres, malgré notre entraînement pour ce genre de représentation. Dans la suite immédiate de cette expérience nous posons comme objectif de rechercher l'identification des éléments du signal acoustique porteurs du sens de la parole, à l'exclusion des données esthétiques, émotives ou personnalisant le locuteur. La démarche consiste donc à "simplifier" le problème.

Nous constatons rapidement que la parole reste intelligible sans intonation (guimbarde) et même sans l'apport des cordes vocales (voix chuchotée). En effet, quel que soit le support acoustique du signal vocal - bruit ou spectre de raies - **ce sont les formes spectro-temporelles qui portent le code de l'information sémantique pour l'auditeur humain**. Le concept de forme acoustique sémantique, énoncé et développé par E.Leipp va se révéler d'une grande fécondité.

---

1. A. Moles, Théorie informationnelle de la Musique; GAM N°5, 1964
2. Il faut rappeler qu'à cette époque il n'était possible d'analyser qu'une durée de signal de 2,4 secondes, et qu'en conséquence la représentation de 15 minutes de parole était un travail considérable.



Son 1

La parole est à l'icophone.

Phrase produite successivement 1/ enVoix normale; 2/ à la guimbarde; 3/ en voix chuchotée; 4/ enregistrée sur un dictaphone; 5/ synthétisée à l'icophone 01. Chaque exemple est répété 2 fois. Congrès de LANNION Juin 1965

Son 2

Il fait beau aujourd'hui.

Phrase produite en Voix normale et par synthèse globale à l'icophone 02. Congrès de Grenoble 1967.

Son 3

Allo allo, ici l'icophone. Bonjour Monsieur, comment allez-vous? Il fait beau aujourd'hui, je voudrais aller me promener, (chant de merle puis de pinson), les pt'its oiseaux chantent, c'est l'printemps.

Texte Synthétisé par copie de sonagrammes, icophone 02; MC 1968.

Divers essais de filtrage nous conduisent à limiter la bande passante à 100-4000 Hz, celle-là même retenue par les ingénieurs du téléphone.

Mais pour montrer la validité de nos hypothèses, il nous faut envisager une synthèse, c'est à dire transcrire cette forme sémantique réduite, en sons. Nous sommes en 1965...La première tentative imaginée par le groupe est acrobatique. Elle consiste à produire la "lecture" spiralée d'une image fixée sur le cylindre du sonagraphe à l'aide d'une cellule photoélectrique et d'un oscillateur dont la fréquence glissante balaye la bande 100-4400 Hz[3]. On peut entendre ***Ex Son 1*** la 1ère phrase ainsi réalisée, qui tenait de l'acrobatie. Arrivant en 4ème position la phrase est reconnaissable... mais pour nous le résultat était suffisamment convaincant pour que soit décidée la construction d'un synthétiseur analogique multi-oscillateurs. Les lignes directrices de notre recherche et les premières réalisations sont présentées au colloque sur la parole organisé par le CNET en 1966.

Cg: 54. - LEIPP E., LIENARD J.S, CASTELLENGO M., (1966), Parole et gestalttheorie, Colloque GALF - Lannion 1966.

Avec la construction du premier "icophone"[4] disposant d'un lecteur optique et d'un banc de 44 oscillateurs il devenait possible de reproduire plusieurs phrases préalablement analysées au sonagramme ***Ex Son 2***. Notre présentation au colloque GALF de Grenoble en 1967 fait sensation. Nous sommes les seuls à produire une synthèse efficace de phrases variées, en débit naturel ***Ex Son 3*** . Mais l'étape décisive va être la constitution d'un dictionnaire de formes élémentaires

3. Le signal de l'oscillateur produit à chaque tour est enregistré par "re-recording". Comme l'opération nécessite 144 tours.... il faut compenser l'intensité de l'enregistrement de façon empirique. Le premier exemple audible est obtenu après plusieurs semaines d'efforts...
4. Les deux premiers icophones ont été conçus et réalisés au laboratoire de Mécanique Physique de St Cyr l'Ecole, par J. Sapaly et C. Cotin.

Cg: 53. - CASTELLENGO M., (1967), La synthèse de la parole a l'icophone : Les problèmes de la perception d'une voix synthétique, Colloque GALF, Grenoble 1967, Revue d'Acoustique 3-4, Paris.

2. LES DICTIONNAIRES DE DIPHONEMES. 1967-1974.

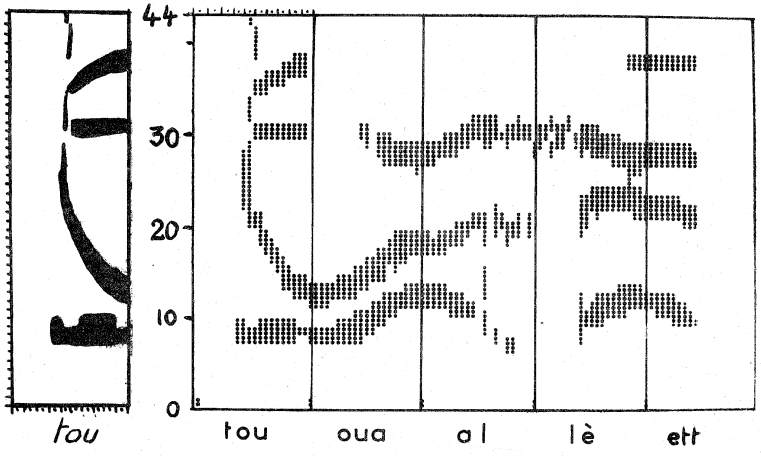


Figure 2.1 Principe de la synthèse par éléments phonétiques normalisés. (à gauche, analogique ; à droite, numérisés et normalisés ; 44 fréqx20 événements temporels)

Jusqu'alors, la synthèse d'une phrase nécessitait, comme c'était le cas pour les systèmes paramétriques des autres centres de recherche, d'analyser cette phrase préalablement énoncée par une voix naturelle. Les premiers essais de reconstitution de mots à partir de l'assemblage de syllabes isolées confirment que **l'unité articulatoire, la forme élémentaire irréductible est bien la syllabe**. Ainsi pour écrire le mot "Paris" il faut associer trois syllabes : PA, AR et RI. En réalisant le catalogue acoustique de toutes les transitions des sons de la langue, il devient possible de reconstituer des mots et des phrases quelconques. Une telle base nous donnera accès à un vocabulaire illimité.

Le travail d'élaboration du dictionnaire des éléments phonétiques a été considérable. Sur la base de 30 phonèmes (13 voyelles et 17 consonnes) ce sont 900 diphonèmes qu'il faut enregistrer, transcrire en dessins, corriger, inclure dans des phrases, et le tout manuellement. C'est à dire que chaque nouvel essai nécessite de redessiner intégralement[5] la forme de l'élément en cours d'étude. Les exemples sonores ***Ex Son 4*** et ***Ex Son 5*** donnent à entendre des mots isolés utilisés dans les tests, et le premier grand texte dessiné à la main.

Son 4

Un ticket, carotte, succès, saucisson.

Synthèse par diphones de mots disyllabiques, pour tests d'intelligibilité. Dictionnaire MC.

Son 5

Bonjour Mesdames, bonjour mesdemoiselles, bonjour Messieurs. Je m'appelle icophone. J'ai appris par coeur 400 elements phonetiques.. (etc).

1er texte synthétisé manuellement à partir du dictionnaire d'éléments phonétiques. Icophone 03. MC 1967.

5. Le dessin, assez acrobatique à pratiquer, était réalisé sur une bande de mylar à l'aide d'une «encre» constituée d'une fine poudre de carbone en suspension dans du benzène...

Son 6

Voix 1- *Bonjour Mademoiselle* - **Voix 2** *-Bonjour Monsieur* - **Voix 1** *Les p'tits oiseaux chantent c'est l'printemps* ...etc.

Texte à 3 voix synthétisé de façon entièrement automatique à l'Icophone 04 numérique, avec variation de l'intonation et voix chantée!!.

Dictionnaire de diphones LMANC. MC 1971.

Heureusement, dès 1968, les dessins peuvent être numérisés (880 nombres binaires par éléments) et après connection de l'icophone III à l'ordinateur du Centre de Calcul Analogique du CNRS[6], la synthèse est réalisée en quasi temps réel, à partir d'un texte écrit sur le clavier de l'ordinateur. Dans l'***Ex Son 6*** on peut entendre un dialogue entre trois personnes, dont l'une chante! La méthode montre ses qualités : extrême simplicité, intelligibilité quasi-totale, vocabulaire illimité, réponse immédiate. Un brevet international est déposé par l'Anvar.

Cg: 50. - LEIPP E., LIENARD J.S ,CASTELLENGO M., (1968), La synthèse de la parole a partir de digrammes phonetiques), Comptes rendus du 6ème ICA, Tokyo.

Pendant toutes ces années j'ai consacré une grande partie de mon activité de recherche à mettre au point les dictionnaires de diphonèmes, d'abord sous forme analogique, puis au CCA. Le travail se faisait "à l'oreille", et était sanctionné par différents tests pratiqués avec des publics divers. J.S. Liénard avait réalisé à cet effet une série de mots «inventés», produits par l'ordinateur, en respectant les règles d'occurence des transitions de la langue française.

Cg: 47. - SAPALY J., TEIL D., CASTELLENGO M., (1972), un périphérique d'ordinateur à réponse vocale, Congrès AFCET – Grenoble.

Cg: 46. - SAPALY J., TEIL D., CASTELLENGO M., (1974), l'Unite à réponse vocale, Icophone 5, 5èmes journées d'Etudes GALF / AFCET – Paris.

Contrairement aux systèmes de synthèse "par règles", le travail nécéssitait d'extraire de l'analyse les éléments pertinents de la forme. Comme l'appareil de synthèse était très réducteur (nombre limité de fréquences; déclenchement en tout ou rien) il s'agissait moins d'une copie que d'une extraction sélective des traits pertinents, du «prototype» dont la validation se faisait à l'écoute. Bénéficiant d'une "oreille" analytique et objective développée lors de la pratique musicale, j'ai pu m'adonner, pendant toutes ces années, à la passion qui m'avait en partie retenue au laboratoire : l'écoute des sons.

3. Conclusion

A l'occasion de ce travail j'ai développé mes premières observations sur la perception sonore et en particulier :

- les relations entre la notion d'intelligibilité et la structure syntaxique;
- l'importance du taux de prévisibilité du signal;
- la prise en compte d'une hiérarchisation des différents niveaux d'écoute (phonème, mot, phrase);
- la capacité des auditeurs à reconnaître des formes avec peu d'indices pertinents.

---

6. Un IBM 1130 : mémoire centrale de 8000 mots de 16 bits!

La notion de forme acoustique spectro-temporelle, ici porteuse de l'information sémantique de la parole, allait se révéler d'une grande fécondité pour tous les aspects de la perception sonore, et plus particulièrement pour la notion de timbre dans sa composante «timbre causal»

Une forme est anamorphosable dans toutes ses dimensions, et reste reconnaissable pour un auditeur humain. L'exemple sonore ***Son 7***fait entendre la même phrase en voix chuchotée, avec quatre réalisations différentes. Entre la première, qu'on peut attribuer à un enfant, et la quatrième qui serait celle d'un homme âgé, l'ensemble des fréquences est dans un rapport 1/2.

Son 7

Le petit chat fait sa toilette.

Présenté avec 4 anamorphoses fréquentielles différentes.

Successivement 1,5 ; 1,25 ; 1 et 0,75. Icophone 02

MC ; 1968.

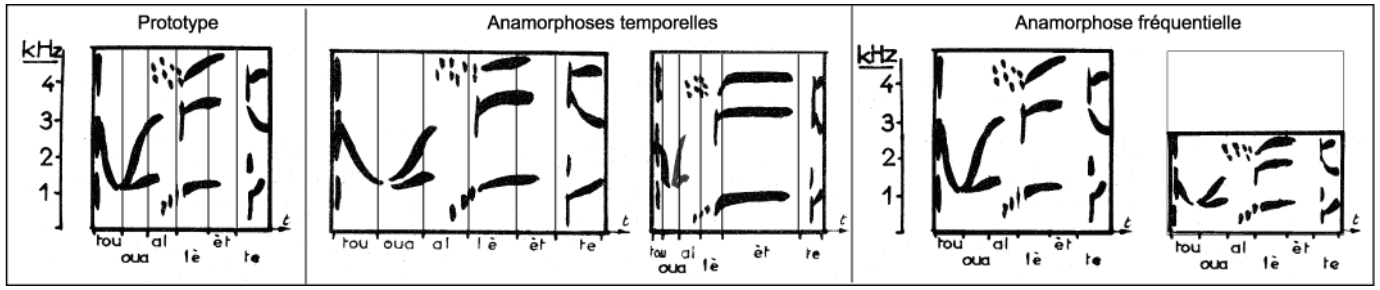


Figure 2.2 La forme spectrale d'un mot « toilette » et ses anamorphoses. Exemple d'anamorphose fréquentielle : Son *7*

En offrant la possibilité de générer des sons parfaitement maîtrisables ne différant que par un trait distinctif dont on peut tester l'effet perceptif, cette première expérience de synthèse sonore s'est révélée être un outil puissant pour tester les hypothèses que l'on pouvait faire sur la perception des sons.

Je poursuivrai ce travail «d'analyse par synthèse» à IRCAM, à propos de l'étude des sons multiphoniques.