Déc 1968 - La Revue d'Acoustique N° 3-4

M. CASTELLENGO
Laboratoire d'Acoustique de la Faculté des Sciences de Paris

LA SYNTHÈSE DE LA PAROLE A L'ICOPHONE LES PROBLÈMES DE LA PERCEPTION D'UNE VOIX SYNTHÉTIQUE

I. — INTRODUCTION

A la suite des travaux d'analyse de la parole entrepris au laboratoire, nous avons été amenés à imaginer un synthétiseur qui nous permettrait de démontrer le bien-fondé de nos hypothèses. Cet appareil, l'Icophone II, sur lequel M. SAPALY donne plus loin des précisions, n'a pu être terminé que très récemment ; ce sont donc les tous premiers résultats de synthèse que nous exposerons, ainsi que les problèmes de perception soulevés par la reconnaissance d'une voix synthétique.

II. — ANALYSE ET SYNTHESE DE LA PAROLE A L'ICOPHONE II

Les idées directrices et les méthodes de travail sont le développement de l'étude réalisée avec l'Icophone I que nous avons présentée à Lannion en Juin 1966.

L'ensemble des signaux accustiques émis par le système phonatoire humain est extrêmement complexe et convoie trois types d'information comme l'a montré M. LEIPP dans la communication précédente. L'information sémantique qui nous intéresse ici est tout entière contenue dans une **forme acoustique** à trois dimensions (fréquence, temps, intensité), originale pour un mot donné. Rappelons que cette forme possède des propriétés remarquables : elle peut être transposée en fréquence, anamorphosée, tronquée, sans cesser d'être reconnue comme telle. C'est un tout, original, différent de la somme des éléments qui le composent.

Le travail que nous nous proposons est donc d'extraire cette forme afin de la fournir à l'appareil de synthèse. Divers essais nous ont amenés à constater qu'elle apparaissait au mieux sur les sonagrammes de voix chuchotée, filtrée (on élimine les composantes supérieures à 4.500 Hz).

La synthèse s'effectue de la façon suivante : on dessine la forme sur un support transparent à l'aide d'une encre opaque. Ce support défile devant une rangée de 44 cellules photoélectriques lesquelles commandent 44 générateurs sinusoïdaux entre 100 et 4.400 Hz.

Pour réaliser de la voix chuchotée avec ce dispositif, nous avons procédé de la façon suivante :

— les générateurs de fréquence sont systématiquement désaccordés à l'oreille, afin d'éviter tout accord consonant qui donnerait à la voix une musicalité indésirable.

— chaque générateur est modulé aléatoirement en fréquence, dans des limites réglables.

— le dessin des formes est peint au moyen de hachures verticales ; de la sorte on réalise autant de « clics » que de traits, donc le maximum de bruit.

On obtient ainsi une voix sans cordes vocales, un peu éraillée, mais parfaitement intelligible. Son timbre particulier déroute l'auditeur nouveau venu, mais celui-ci s'y accoutume très vite au point qu'elle devient pour lui une voix familière.

L'apprentissage du dépouillement des sonagrammes s'effectue parallèlement à celui du

dessin de synthèse, l'un permettant le perfectionnement de l'autre par contrôle mutuel. Deux méthodes de travail sont possibles :

1°) **La synthèse globale**

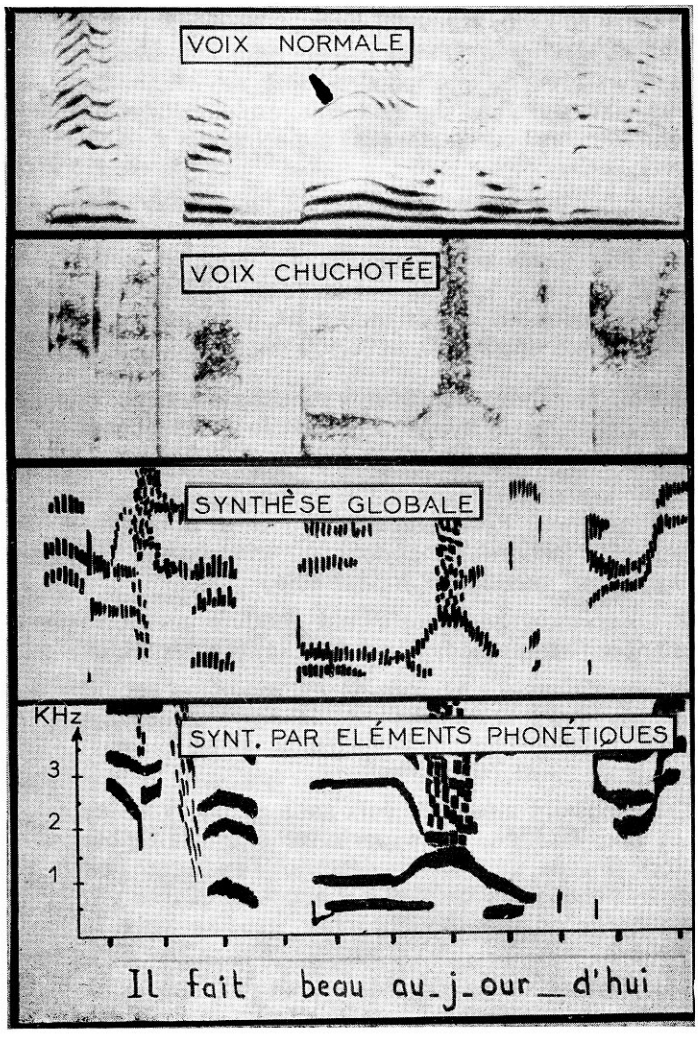
Elle consiste à reproduire le dessin d'un sonagramme fait au préalable. Prenons par exemple la phrase « IL FAIT BEAU AUJOURD'HUI ». La figure 1 montre la comparaison entre le sonagramme de la voix normale, celui de la voix chuchotée et le dessin, en synthèse globale, fourni à l'icophone. On reconnaît particulièrement bien une forme forte comme « aujourd'hui ». Cette méthode a pour avantage de donner de bons résultats même avec un dessin maladroit. En effet, cette « super-forme » qu'est la phrase est reconnue globalement, même si certains éléments sont mal réalisés. D'autre part, l'auditeur dispose d'informations supplémentaires qui aident à la reconnaissance : découpage temporel, rythme de la phrase, etc...

Voici quelques exemples de synthèse :

ALLO ALLO, ICI L'ICOPHONE
IL FAIT BEAU AUJOURD'HUI

(Les autres exemples ne sont pas indiqués dans ce texte afin d'éviter toute suggestion).

FIG. 1



2°) **La synthèse à partir d'éléments phonétiques**

L'élément phonétique est constitué par l'association de deux phonèmes ; c'est une forme originale que l'on ne peut pas décomposer. Il est l'expression du mouvement de l'appareil vocal d'un phonème à un autre. En raccordant les éléments phonétiques, les uns aux autres, on peut reconstituer la forme globale d'un mot, ce que démontrent les exemples que nous allons vous faire entendre :

BUREAU — TICKET — CAROTTE

(D'autres exemples ont été réalisés qui ne sont pas cités ici) cf. fig. 1.

Lorsque nous disposerons des dessins schématisés de tous les éléments phonétiques de la langue française, il nous sera possible d'écrire n'importe quelle phrase, sans analyse préalable. En raison du peu de temps dont nous disposions nous n'avons pu synthétiser que quelques mots, mais l'étude complète est en cours.

III. — LES PROBLEMES POSES PAR LA RECONNAISSANCE DE LA VOIX SYNTHETIQUE

Les échantillons de voix synthétique que nous avons réalisés ont été entendus par de nombreux sujets. Nous avons rapidement constaté les réactions les plus diverses. Les mêmes mots, compris d'emblée par certains individus, restaient totalement inintelligibles pour d'autres. En fait, le mécanisme de reconnaissance de la parole met en jeu un grand nombre de variables dues au récepteur.

Pour réaliser la synthèse d'un mot nous cherchons à extraire la forme la plus simplifiée possible ; on peut se demander jusqu'à quel point il est possible de pousser la simplification d'une forme sans altérer son intelligibilité.

a) **Le degré de schématisation d'une forme dépend du message à transmettre.**

Prenons par exemple l'élément phonétique « MAN ». Le degré de schématisation de cette forme ne sera pas le même selon qu'elle fait partie de la phrase « COMMENT ALLEZ-VOUS ? » ou du mot « ELEMENT » ou du nom propre « M. SEMANTIN ». Dans le premier cas, c'est la forme globale de la phrase qui est perçue, l'élément considéré, n'existe pas pour l'auditeur et pourra être tout juste esquissé, voire déformé, sans pour cela gêner la reconnaissance. Au contraire, dans le cas d'un nom inconnu, chaque élément est capital pour l'intelligibilité. Le dessin doit contenir toutes les indications qui font de cet élément une forme originale. sans confusion possible avec une autre.

b) **Le degré de schématisation d'une forme dépend des caractéristiques du récepteur.**

Comprendre un mot c'est :

— Analyser la forme acoustique projetée au niveau des centres supérieurs,

— Retrouver dans la mémoire une forme comparable.

— Y associer la signification apprise.

Cette opération met en jeu diverses fonctions pour lesquelles les variations sont grandes d'un individu à un autre, tant sur le plan anatomique et physiologique que sur celui du conditionnement.

La forme stockée dans la mémoire est **un stéréotype** contenant les invariants d'une forme (phrase, mot ou élément). Ce stéréotype est extrait par l'individu à partir des formes qu'il mémorise au cours de l'acquisition du langage ; il lui est **personnel** et peut varier dans de grandes proportions d'un individu à un autre et en particulier il dépend de l'accent régional du sujet. Ce stéréotype est d'autant plus net que la forme correspondante est fréquente ; il s'affirme et se simplifie au cours des années. Un enfant, par exemple a besoin de formes plus complètes pour reconnaître la parole ; c'était au cours de nos expériences le sujet le plus dérouté par l'audition de la voix de l'Icophone II.

Enfin la rapidité de reconnaissance d'une forme très schématisée, dépend aussi des capacités intellectuelles du sujet et de son état de fatigue.

Toutes ces données nous permettent de comprendre les réactions individuelles variées que nous avons pu observer lors des tests d'intelligibilité.

Nous voudrions insister, pour terminer, sur l'importance de l'état d'esprit de l'auditeur. Il semble qu'il y ait deux mécanismes totalement différents selon qu'il est ou non prévenu.

— **Un auditeur non prévenu** doit rechercher dans sa mémoire parmi tous les stéréotypes possibles ceux qui s'approchent de la forme proposée, et choisir le bon. Il a d'autant plus de mal que la voix est inhabituelle, qu'il n'y a pas de contexte, que la forme est peu caractéristique, etc...

— **Un auditeur prévenu** extrait d'avance de sa mémoire le stéréotype correspondant ; au moment de l'audition il le compare avec la forme synthétisée et il lui suffit alors d'indices mêmes infimes pour reconnaître le mot. Certains individus même, n'entendent pas la réalité, mais projettent la forme qui est dans leur mémoire. Nous nous heurtons quotidiennement à ce problème en perception de la musique. Malheureusement l'expérimentateur est dans le cas d'un auditeur prévenu, et de ce fait dans les conditions les plus mauvaises. Pour corriger son travail il lui faut sans cesse une « oreille vierge ».

CONCLUSION

L'analyse et la synthèse de la parole tels que nous l'envisageons consiste dans l'extraction d'une forme dont le degré de simplification dépend du message transmis et du récepteur. Les échantillons synthétisés à l'aide de l'icophone II ont 95 % d'intelligibilité ; ceci montre que la forme extraite est nécessaire et suffisante pour véhiculer l'information sémantique. Les recherches en cours ont pour but d'établir le dictionnaire des éléments phonétiques à partir duquel il sera possible de synthétiser une forme acoustique quelconque, donc un discours.