

1971.

SEVENTH INTERNATIONAL CONGRESS ON ACOUSTICS, BUDAPEST 1971 **21 C 1**

LES ANAMORPHOSES DE LA PAROLE ET LEURS SIMULATION À L'ORDINATEUR

E. LEIPP—M. CASTELLENGO—J. SAPALY—J. S. LIENARD *Faculté des Sciences, Dép. Acoustique, Paris, France*

12) INTRODUCTION. L'information sémantique contenue dans un mot ou une phrase est supportée par des "formes", des "gestalts" acoustiques que les sonagrammes de parole chuchotée mettent bien en évidence (bib 1). Ces formes sont décomposables en éléments insécables: les diphonèmes ou "phonatomes", traduisant le passage d'un phonème à un autre. Si on admet 35 phonèmes on a donc $35 \times 35 = 1225$ phonatomes, dont la moitié environ est utilisée en pratique. Si on tire les sonagrammes des diphonèmes chuchotés, on obtient des graphismes élémentaires, des "digrammes phonétiques", que l'on peut accoler les uns aux autres pour reconstituer la super-forme d'un mot ou d'une phrase. Ces super-formes, relues à l'aide d'un synthétiseur optique (bib 2), l'ICOPHONE, fournissent une parole chuchotée intelligible.

Les digrammes phonétiques, normalisés dans un cadre fréquence-temps fixe (4400 Hz - 100 ms), peuvent être aisément quantifiés et numérisés, puis entrés en mémoire machine. (IBM II30). On peut alors les appeler, et grâce à des mémoires-tampon, les associer côte à côte en super-formes. Un ICOPHONE DIGITAL, spécialement construit à cet usage, constitue la "sortie son" de l'ordinateur et fournit, en temps réel, une parole synthétique chuchotée normalisée d'intelligibilité supérieure à celle du téléphone.

Si les 44 générateurs de sinusoides de l'ICOPHONE digital sont accordés en série harmonique, la parole n'est plus chuchotée, mais voisée en recto-tono (hauteur fixe). Cette parole est peu agréable et son intelligibilité peut être largement améliorée si on peut régler la position des formants et la hauteur du spectre harmonique, comme en parole normale.

Une recherche systématique a donc été faite pour étudier les fluctuations fréquentielles de la parole normale (bib 3 et 4).

‡ Recherches faites en collaboration par le Laboratoire de Mécanique (Prof. SIESTRUNCK), le Centre de Calcul Analogique (Prof. MALAVARD et M. RENARD) avec l'aide de la DRME.

SEVENTH INTERNATIONAL CONGRESS ON ACOUSTICS, BUDAPEST 1971

LES ANAMORPHOSES DE LA PAROLE

2

2*) LES ANAMORPHOSES DE LA PAROLE NORMALE ET LEUR SIMULATION.

Deux problèmes distincts sont à considérer :

- Le problème des formes sémantiques mises en évidence par les sonagrammes de parole chuchotée. Ces formes changent d'un locuteur à l'autre, selon l'anatomo-physiologie de l'appareil phonatoire du sujet considéré et selon ses habitudes articulatoires, prises dans son milieu sociologique. Si on prend comme référence, le stéréotype de notre parole normalisée pour le mot "toilette", par exemple (fig 1a), on note trois types d'anamorphoses des formes sémantiques: étirement (ou compression) du temps (Ib) ou de la fréquence (Ic); ou encore anamorphose rythmique (inégalité de durée des digrammes). En parole normale, ces anamorphoses conditionnent l'identification du locuteur.

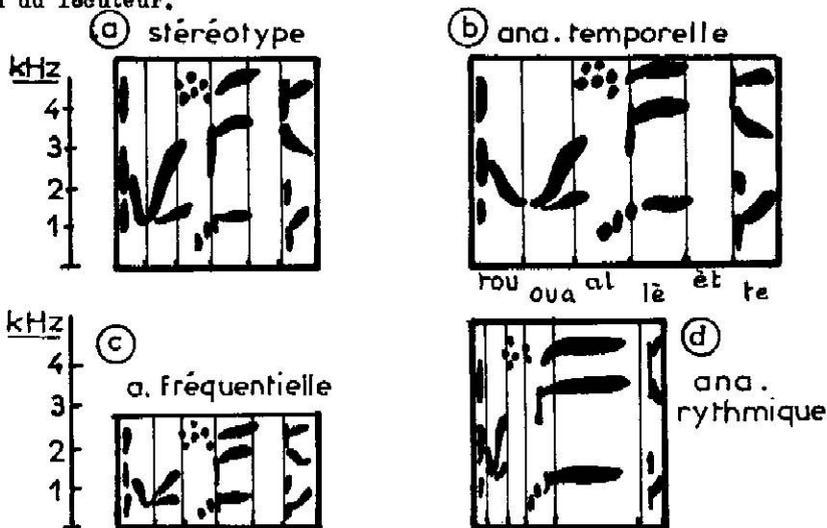


fig.1

Pour simuler ces anamorphoses à l'ordinateur, un ICOPHONE particulier a été conçu et construit (voir Communication SAPALY: "L'icophone considéré comme périphérique de calculateur" dans le 7^e ICA). Une équipe de chercheurs du CCA (MM. LIENARD-TEIL- MILOUKA - CALINET ; voir communication au 7^e ICA sous ces noms) a réalisé les programmes permettant d'anamorphoser les formes sémantiques normalisées dans les mêmes proportions que la parole normale. On peut ainsi simuler le "timbre" d'une voix d'homme, de femme ou d'enfant.

- Le problème de l'intonation. Les données de base pour simuler l'intonation ont été clairement définies par l'un de nous (bib 4). Lorsque les cordes vocales entrent en jeu, le spectre de raies harmoniques qu'elles délivrent, quantifie automatiquement les formes sémantiques de la parole. Prenons le cas de la séquence "mais oui" (fig 2); pour une voix aiguë, les

SEVENTH INTERNATIONAL CONGRESS
ON ACOUSTICS, BUDAPEST 1971

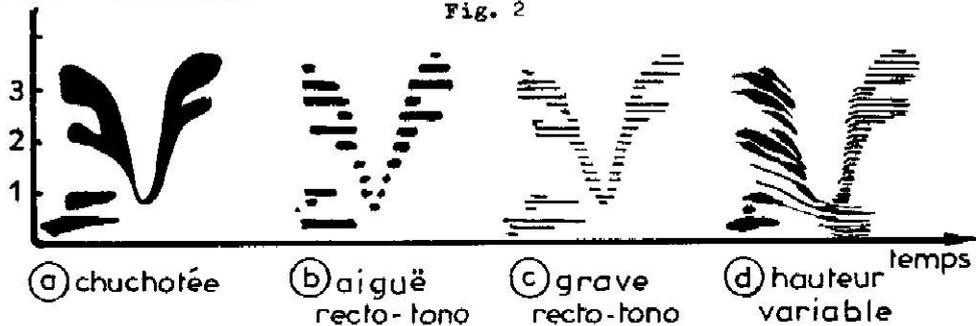
LES ANAMORPHOSES DE LA PAROLE

3

raies sont écartées (2b); elles sont serrées pour une voix grave (2c) et l'écartement varie continuellement et continuellement pour une voix normale (2d); mais on reconnaît toujours la forme sémantique (2a): on "comprend" toujours le mot.

KHZ "mais oui" en parole:

Fig. 2



On notera en passant que la sensation de hauteur de la voix ne résulte pas d'un éventuel comptage de fréquence du fondamental; on peut couper celui-ci par filtrage sans que la sensation de hauteur change. La sensation de hauteur résulte en fait de l'appréhension par notre cerveau du degré de serrage des raies harmoniques. C'est précisément cette opération de réglage du degré de serrage que simule l'ICOPHONE IV, piloté par l'ordinateur, grâce à un écran de visualisation sur lequel on inscrit au préalable la "ligne mélodique" à suivre.

3°) RESULTATS. Une bande sonore sera présentée au 7^o ICA permettant d'apprécier les résultats obtenus. On a réalisé ici en parole synthétique une scène à trois personnages: une voix grave d'homme en recto-ono (sans intonation), une voix moyenne d'homme en recto-ono et une voix de femme également sans intonation. A un moment donné, on introduit des modulations d'intonation dans la voix moyenne d'homme et dans la voix de femme. Les échantillons de parole réalisés ici ne sont pas de qualité esthétique suffisante, mais ils montrent que la simulation des diverses anamorphoses est correctement réalisée, car la parole est intelligible. Il s'agit en fait d'un premier essai et des travaux sont en cours pour améliorer les résultats.

4°) CONCLUSIONS. L'ensemble des dispositions utilisées (ICOPHONE IV et ordinateur IBM II30) permet donc d'anamorphoser par programme les formes sémantiques et, simultanément, de régler à loisir l'écartement des raies au spectre. On peut donc simuler le timbre des diverses voix normales et aussi l'intonation dans un discours synthétisé par ce procédé. La synthèse ne

SEVENTH INTERNATIONAL CONGRESS ON ACOUSTICS, BUDAPEST 1971

LES ANAMORPHOSES DE LA PAROLE

4

représente d'ailleurs qu'une étape dans nos recherches, mais une étape indispensable. En effet notre but lointain est la reconnaissance automatique de la parole d'un locuteur quelconque par ordinateur, problème difficile, en raison même de l'existence des anamorphoses en parole normale. Il fallait donc au préalable étudier celles-ci et montrer qu'on en avait compris le mécanisme en les reconstituant par synthèse. Le problème de la reconnaissance des formes sémantiques d'un discours de locuteur quelconque déterminant l'intelligibilité de celui-ci peut dès lors être attaqué avec quelques chances de succès. Il s'agira de capter le discours par microphone, de mettre en évidence par un appareillage approprié (en cours d'élaboration au Laboratoire de Mécanique) les formes sémantiques particulières au locuteur, puis d'anamorphoser celles-ci pour les contraindre dans le cadre normalisé de la parole synthétique qui est actuellement en mémoire machine. Cette anamorphose réalisée, des opérations de corrélation permettront d'établir le taux de similitude entre le signal délivré par le microphone et les digrammes normalisés stockés en mémoire machine. Le problème de l'information esthétique de la parole, liée aux modulations du spectre de raies sera abordé de la même façon. Les recherches préparatoires pour y atteindre sont en cours.

BIBLIOGRAPHIE

limitée à quelques-unes de nos publications relatives à la parole.

- 1^a) LEIPP(E) - CASTELLENGO(M) - LIENARD (J.S) . La synthèse de la parole à partir de digrammes phonétiques . Comptes Rendus du 6^e ICA (TKIO) 1968 n° C 5-6
- 2^a) LEIPP(E) - CASTELLENGO(M) - LIENARD(J.S) - SAPALY(J) . Structure physique et contenu sémantique de la parole. Colloque GALT de Grenoble. Revue d'Acoustique n° 3-4 (Paris) 1968.
- 3^a) LEIPP(E) . Les recherches sur la parole au Laboratoire d'Acoustique de la Faculté des Sciences de Paris. Colloque GAM (Groupe d'Acoustique Musicale de la Faculté des Sciences) organisé en janvier 1971. Compte rendu en cours d'impression; Bulletin GAM n°53 (Paris) 1971
Ce bulletin comprendra également les communications de Melle CASTELLENGO(M) et de MM. LIENARD (J.S) -SAPALY (J) - TEIL (D) MLOUKA (M) et GALINET (A)
- 4^a) LEIPP (E) . Le contenu informatif de la parole. Comptes Rendus du IV^e Congrès d'Acoustique de BUDAPEST (1967)