THE 6TH INTERNATIONAL CONGRESS ON ACOUSTICS
TOKYO, JAPAN, AUGUST 21-28, 1968

C-5-6

LA SYNTHESE DE LA PAROLE A PARTIR DE DIGRAMMES PHONETIQUES

LEIPP (E) Maître de recherche au CNRS
CASTELLENGO (M) ; LIENARD (J.S.)

Laboratoire d'Acoustique de la Faculté des Sciences de Paris*

1°) INTRODUCTION . Une recherche systématique sur la structure physique des signaux acoustiques en général nous a conduit à nous intéresser aux problèmes de la parole.Une approche originale du problème nous a permis d'aboutir à une synthèse de parole intelligible avec des moyens très limités.Notre procédé est à présent suffisamment au point pour que nous puissions en donner les grandes lignes.

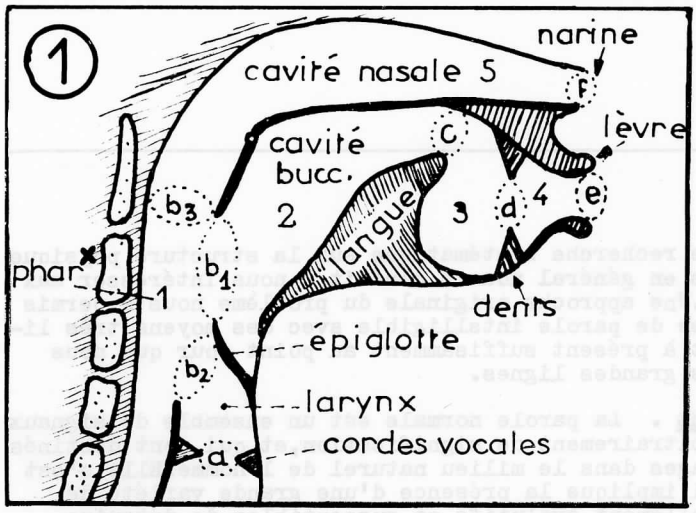
2°) POSITION DU PROBLEME . La parole normale est un ensemble de signaux acoustiques chargés arbitrairement de signification,et qui sont destinés à communiquer des messages dans le milieu naturel de l'homme.Elle s'est adaptée à ce milieu qui implique la présence d'une grande variété de bruits de fond essentiellement évolutifs et susceptibles de détruire plus ou moins les signaux.On comprend dès lors la nécessité d'une redondance du signal physique de la parole.Mais si on se place dans des conditions particulières où le bruit de fond n'intervient plus,cette redondance représente une perte inutile d'énergie.Il était donc intéressant de chercher à définir cette redondance de façon précise,puis de tenter de l'éliminer pour ne conserver que le squelette informatif,c'est à dire le minimum nécessaire et suffisant du signal.Nous y avons largement réussi, et,du même coup,nous avons simplifié le problème de la synthèse de la parole au point de pouvoir la réaliser avec des moyens très économiques.

Pour réduire au maximum la redondance,nous avons commencé par n'utiliser que la bande téléphonique (entre 100 et 4000 Hz),dont l'expérience montre qu'elle est suffisante.D'autre part,nous avions observé que le spectre de raies délivré par les cordes vocales et contenant la plus grosse partie de l'énergie acoustique de la parole,n'avait d'autre utilité que de faire porter la voix au loin.La meilleure preuve en est qu'on peut le supprimer en déconnectant les cordes vocales par un artifice très simple:il suffit d'utiliser la voix chuchotée dont on vérifie que l'intelligibilité est totale.

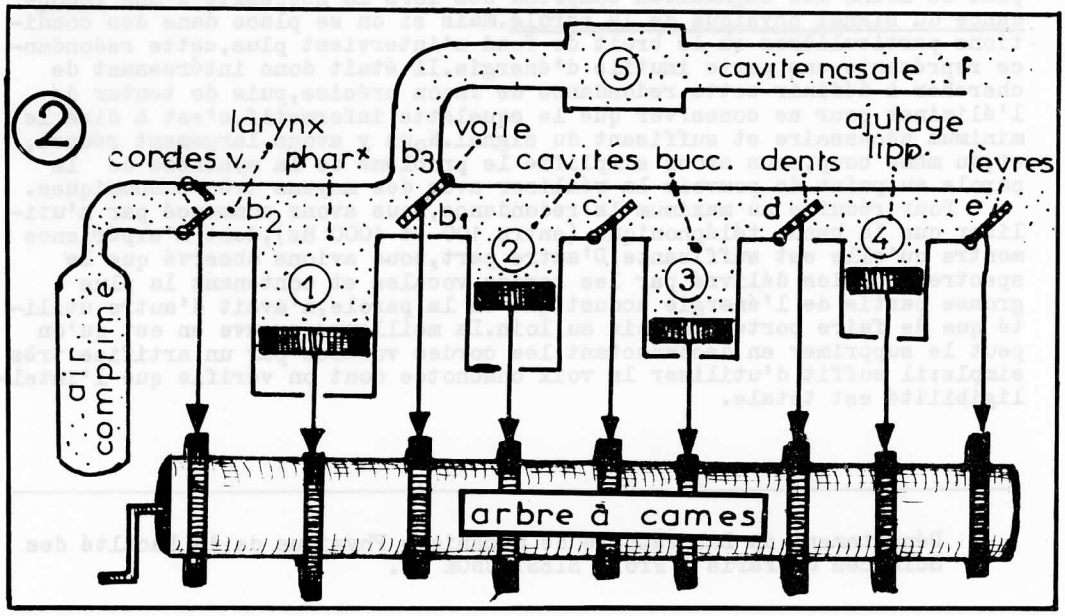
---

* Département du Laboratoire de Mécanique Physique de la Faculté des Sciences de Paris (Prof. SIESTRUNCK) .

3°) MECANISME ET STRUCTURE PHYSIQUE DE LA PAROLE CHUCHOTEE. L'appareil phonatoire normal est constitué d'une série de cavités couplées,de volume et d'ouvertures réglables à volonté dans une large mesure .(Fig 1). Il s'agit d'une véritable machine à faire des sons dont on peut aisément réaliser un modèle mécanique de fonctionnement (Fig 2) .Le mouvement compliqué de l'appareil phonatoire normal lors de l'articulation d'un mot est simulé ici grâce à un arbre à cames qui synchronise les mouvements élémentaires des cavités et ouvertures.Chaque came représente un sous-programme.Si l'on excite ce dispositif en insufflant de l'air comprimé on produit des bruits de bande évolutifs entrecoupés d'explosions plus ou moins brèves,colorées par les cavités voisines de l'explosion.

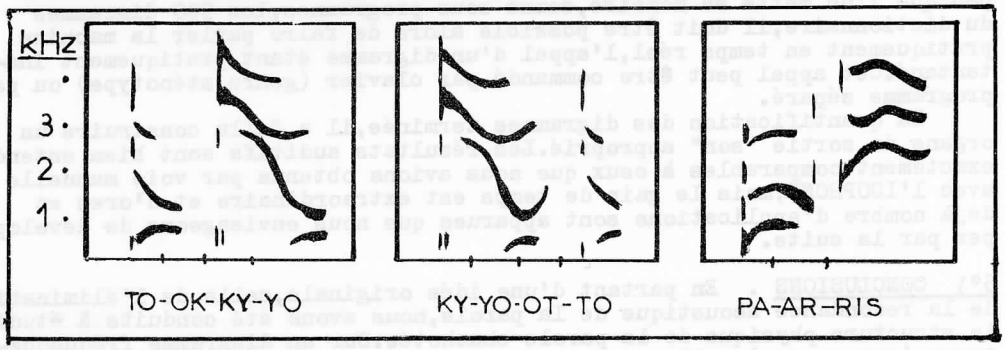


Du point de vue acoustique,le programme de l'arbre à cames se traduit donc par un signal acoustique complexe à trois dimensions:temps, fréquence,niveau. L'expérience montre que l'une de ces grandeurs, le niveau,est d'importance secondaire du point de vue du contenu sémantique du signal. C'est pourquoi la représentation sonagraphique du phénomène est hautement significative.En fait,le sonagramme véhicule l'essentiel de l'information.



LA SYNTHESE DE LA PAROLE A PARTIR DE DIGRAMMES PHONETIQUES.

Le sonagramme d'un mot chuchoté se présente sous l'aspect d'un graphisme mettant en évidence une forme globale,originale pour chaque mot,et,par conséquent parfaitement reconnaissable dans chaque cas sans ambiguïté. Par exemple,on ne confondra jamais "TOKIO" avec "KIOTO" ou "PARIS",pas plus qu'on ne confond le dessin d'un chat et d'une maison.(fig 3)



Comprendre un mot,c'est donc capter,reconnaitre et décoder une forme acoustique;celle-ci apparait le plus clairement possible dans la parole chuchotée,mais se retrouve bien entendu dans la parole normale.Nous avons longuement étudié cette question et observé que chaque forme globale pouvait être décomposée en éléments raccordables.Chacun de ceux-ci correspond non à un phonème,mais au mouvement de l'appareil phonatoire entre deux phonèmes voisins,en bref à un sous-programme de la machine à parler (mouvement d'une,ou de plusieurs cames simultanément).Un mot n'est donc pas décomposable phonétiquement en phonèmes,mais en éléments phonétiques qui sont des digrammes (association de deux phonèmes) et que nous appelons "phonatomes".Ainsi le mot "TOKIO" n'est pas la somme de cinq phonèmes mais l'enchainement de quatre phonatomes: TO OK KI IO ,qui se raccordent par définition,donnant de la parole fluide.

En français on admet l'existence d'environ 30 phonèmes principaux; leur association deux à deux donnera donc 900 digrammes,dont nous avons réalisé le dictionnaire complet.On notera que la forme de beaucoup de digrammes est réversible dans le temps ("iou" inversé donne "oui").Il suffit en fait de quelque 400 digrammes pour dessiner la forme globale de la totalité des mots utilisés.On dispose ainsi d'une véritable "sténographie acoustique" dont l'intérêt est évident puisqu'on peut,sans analyse préalable d'un mot,en dessiner la forme graphique.

Pour vérifier le bien fondé de nos hypothèses de départ,nous avons construit un synthétiseur du type play-back,l'ICOPHONE,qui est exactement l'inverse du sonagraphe.Tout graphisme est lu par un groupe de 44 photodiodes et transformé en signal acoustique global instantanément audible.

Nous avons vérifié avec une centaine de sujets que les digrammes de notre dictionnaire étaient auditivement reconnaissables;les mots entiers et les textes réalisés sont intelligibles à fortiori,et,point important, pour des personnes non suggestionnées,quoiqu'il s'agisse d'une voix chuchotée assez particulière,similaire à une bonne voix de laryngectomisé. Insistons sur le fait que nous n'avons pas cherché à reproduire fidèlement la voix normale,mais à réaliser une voix intelligible aux moindres frais. Nous avons à présent une assez longue expérience en ce domaine:dictionnaire et méthode sont suffisamment au point pour que nous ayons pu envisager d'entrer nos résultats en machine.

4°) SYNTHESE DE LA PAROLE A L'ORDINATEUR (*) Le digramme phonétique, graphisme simple,est facilement quantifiable et numérisable;d'où l'idée de les utiliser pour faire parler un ordinateur.L'intérêt de l'opération est évident:manuellement,avec l'ICOPHONE,il faut quelque 10 secondes pour dessiner un digramme,soit environ 40 secondes pour écrire "TOKIO".Pour peu que l'on mette en mémoire,comme sous-programmes,les 900 digrammes du dictionnaire,il doit être possible alors de faire parler la machine pratiquement en temps réel,l'appel d'un digramme étant pratiquement instantané.Cet appel peut être commandé par clavier (genre sténotype) ou par programme séparé.

La quantification des digrammes terminée,il a fallu construire un organe de sortie "son" approprié.Les résultats auditifs sont bien entendu exactement comparables à ceux que nous avions obtenus par voie manuelle avec l'ICOPHONE;mais le gain de temps est extraordinaire et d'ores et déjà nombre d'applications sont apparues que nous envisageons de développer par la suite.

5°) CONCLUSIONS . En partant d'une idée originale,celle de l'élimination de la redondance acoustique de la parole,nous avons été conduits à étudier la structure physique de la parole chuchotée.Sur un diagramme fréquence-temps (sonagramme),celle-ci se présente sous l'aspect de graphismes simples,de formes globales que l'on peut découper en éléments dont chacun correspond à un digramme phonétique (association de deux phonèmes).La combinatoire entre les 900 éléments phonétiques du français,permet de reconstituer sans analyse préalable n'importe quel mot ou discours en parole intelligible.Pour d'autres langues (allemand,anglais etc),l'adjonction d'un très petit nombre de digrammes suffit pour arriver au même résultat.La lecture des formes représentant les mots est faite par un playback que nous avons construit ,l'ICOPHONE.Le dictionnaire complet des digrammes phonétiques a été quantifié et numérisé pour entrée en ordinateur;ce dernier est ainsi capable de parler pratiquement en temps réel, grâce à une sortie son spécialement construite à cet effet.De nombreuses applications sont à l'étude.

BIBLIOGRAPHIE

LEIPP (E); CASTELLENGO (M) ; LIENARD (J.S) . PAROLE ET GESTALTTHEORIE
Colloque sur la parole (GALF) . Lannion Juin I965
Tirage interne de la Faculté des Sciences

LEIPP (E) INFORMATION SEMANTIQUE ET PAROLE
Bulletin G.A.M N°22 Paris Juin I966
Edition interne Faculté des Sciences de Paris.

LEIPP (E) ; CASTELLENGO (M) ; LIENARD (J.S) ; SAPALY (J)
Structure physique et contenu sémantique de la parole
Colloque du GALF.Grenoble (Avril I967)
A paraitre dans la Revue Française d'Acoustique

LEIPP (E) LE CONTENU INFORMATIF DE LA PAROLE
Comptes rendus du IV° Congrès d'Acoustique de Budapest(1967)

LEIPP (E) MECANIQUE ET ACOUSTIQUE DE L'APPAREIL PHONATOIRE
Bulletin G.A.M. N°32 Paris d cembre I967
Edition interne Faculté des Sciences de Paris.

(*) Etude conduite en collaboration avec le Centre de Calcul Analogique du Centre National de la Recherche Scientifique,dans le cadre des recherches sur le calcul hybride (Prof. L. MALAVARD) .