

L'INTELLIGENCE ARTIFICIELLE

comment faire parler un ordinateur?

(Suite : voir n° 1420)

LES premiers synthétiseurs de parole furent construits pour vérifier l'hypothèse suivant laquelle le riche spectre de la parole pouvait être réduit à un petit nombre de composantes, sans qu'il en découle une dégradation de l'intelligibilité du message parlé. La parole est très redondante, puisque 40 000 unités d'information par seconde sont transmises par une ligne téléphonique classique, au cours d'une conversation, alors que notre cerveau ne pourrait décoder, en fin de traitement, qu'un maximum de 50 unités d'information par seconde ; ces 50 unités d'information sont porteuses de données, tant sur le message que sur la qualité et les caractéristiques de la voix. Le rôle et l'importance de telle ou telle composante du signal peuvent être étudiés au moyen de la synthèse.

Il existe plusieurs types de synthétiseurs. Certains ne tiennent pas compte du mode de fonctionnement de notre appareil vocal ; d'autres, au contraire, en tiennent étroitement compte. C'est ainsi que dans le synthétiseur « à formants », on introduit une fonction de transfert équivalente à celle de l'ensemble des cavités en couplage dans la bouche.

L'APPAREIL VOCAL...

La parole résulte de l'excitation des cavités de l'appareil vocal, soit par une source d'impulsions pour les sons vocaux, soit par une source de bruit, pour les sons non vocaux. Dans le premier cas, la source est constituée par l'ensemble « poumons-cordes vocales » ; dans le second cas, le

bruit est produit par l'action de courant d'air sur certaines parois de l'appareil vocal.

En parallèle sur le conduit vocal, est greffée la cavité nasale, limitée par les narines d'un côté, et de l'autre par le voile du palais. Ce dernier règle le couplage acoustique entre le conduit vocal et le conduit nasal.

Les poumons et la larynx constituent l'essentiel du générateur sonore : les poumons assurent une pression d'air à l'intérieur des bronches ; le larynx comprend divers cartilages mobiles reliés entre eux par des ligaments et des muscles striés ; la partie interne est garnie d'une muqueuse formant deux paires de bourrelets membrano-musculaires (cordes vocales et bandes ventriculaires). Une articulation complexe permet de régler en largeur et en hauteur l'ouverture des cordes vocales ; il est possible également d'épaissir les cordes vocales en contractant les muscles qui les sous-tendent (Fig. 6).

L'impédance interne de la source vocale est essentiellement variable et supérieure à 1 000 Ω acoustiques ; l'impédance de charge apportée par le conduit vocal étant inférieure à 100 Ω , on considère souvent la source vocale comme étant un générateur de grande impédance interne, donc non perturbé par l'évolution des cavités de la bouche.

La section sagittale du conduit vocal montre un profil extrêmement variable selon la position des organes qui le délimitent. On distingue quatre cavités principales : le pharynx, les fosses nasales mises en communication avec le pharynx par l'abaissement du voile du palais, les cavités buccales et enfin l'ajutage lippal (Fig. 7). Chacun des

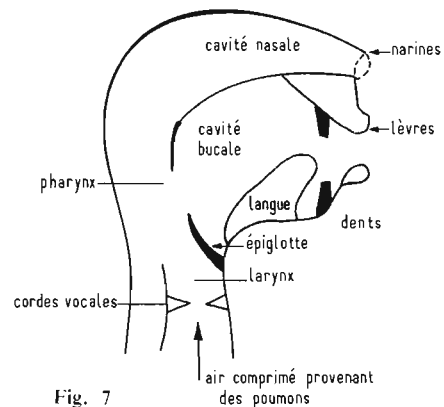
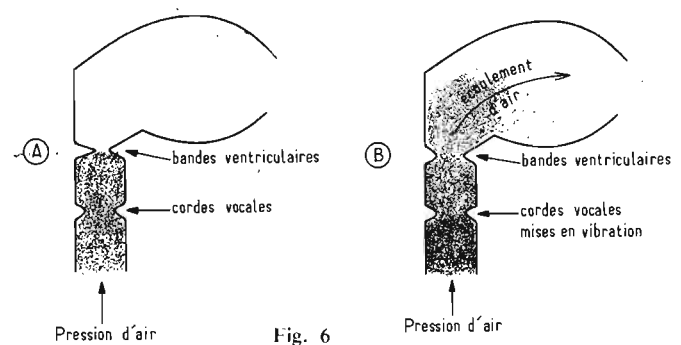
organes en jeu possède un champ de liberté considérable, grâce à la présence de très nombreux muscles : la langue, par exemple, dispose à elle seule de 17 muscles ; la mâchoire inférieure comporte 10 muscles ; les lèvres en comportent également une dizaine... au total, une cinquantaine de muscles participent à la phonation.

Le conduit vocal est un tube acoustique de section variable en fonction du temps et de la dis-

tance au larynx, et de longueur totale légèrement variable. Les parois internes de ce tube sont des tissus acoustiquement très absorbants, donnant lieu à un amortissement considérable des ondes issues de la glotte.

... ET LES FORMANTS

Des modèles mathématiques simplifiés du conduit vocal conduisent à mettre en évidence



plusieurs resonances pour une position donnée du conduit. En 1960, G. Fant a calculé la fonction de transfert du conduit vocal. La fonction de transfert $T(p)$, définie par le rapport du débit en sortie sur débit à l'entrée, peut être considérée comme un produit de plusieurs termes, du type :

$$T_n(p) = \frac{1}{a_n^2 p^2 + 2a_n x_n p + 1} \cdot C_n$$

Cette expression représente la transformée de Laplace de la fonction de transfert ; a_n et x_n représentent des paramètres du conduit vocal, C_n est un terme « correctif » et p est un paramètre (*) complexe.

Cette fonction de transfert correspond à celle d'un circuit résonnant monté en cascade avec un filtre passe-haut.

Les régions fréquentielles correspondant à des maxima d'amplitude sont appelées « des formants ». La fréquence F_n d'un formant est l'une des fréquences de résonance du conduit vocal, et correspond à l'un des pôles de la fonction de transfert $T(p)$. On a alors :

$$F_n = \frac{1}{2\pi a_n}$$

De même, la bande passante « à 3 dB » d'un formant est :

$$B_n = \frac{x_n}{\pi a_n}$$

(*) Les lecteurs intéressés par le calcul des transformées de Laplace, se reporteront à l'article « le Calcul analogique » paru dans le « Haut-Parleur », n° 1360, en pages 133 à 137.

En première approximation cette bande passante est constante dans la plage de variation d'un formant déterminé.

Dans le cas de la production de bruit, la fonction de transfert à considérer est un produit des termes précédents, et de termes de la forme :

$$T_m(p) = a_m^2 p^2 + 2x_m a_m p + 1$$

ainsi que d'un terme correspondant aux pertes glottiques :

$$T_g = p - a_g$$

La présence du terme T_m est due aux cavités en amont de la source de bruit.

Dans le cas d'une excitation des cavités nasales, la fonction de transfert est le produit de deux types de termes :

$$T_n(p) = \frac{a_n^2 p^2 + 2x_n a_n p + 1}{b_n^2 p^2 + 2x_n b_n p + 1}$$

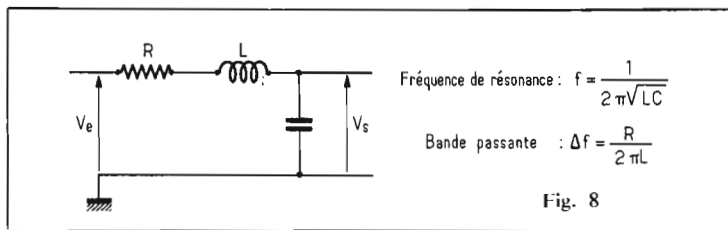
et

$$T_m(p) = \frac{1}{a_m^2 p^2 + 2x_m a_m p + 1}$$

Toutes ces fonctions de transfert permettent de déterminer le débit d'air D_s au niveau des lèvres du locuteur, en fonction du débit de la source. Normalement, l'oreille réceptrice est excitée par la pression P_s qui résulte de ce débit ; on doit associer à ces fonctions de transfert, une fonction de rayonnement de la bouche :

$$\frac{P_s}{D_s} = a \cdot p \cdot C_r(p)$$

a étant une constante et $C_r(p)$ une fonction de transfert comportant un pôle à 1 000 Hz et un zéro à 500 Hz.



LA SYNTHÈSE DE LA PAROLE PAR FORMANTS EN FRANCE

Un synthétiseur de parole, du type « à formants » comportera trois canaux : l'un pour la synthèse des sons vocaux, un second pour la synthèse des sons non-vocaux et un troisième pour la nasalité. On simule ainsi, par des circuits électriques les qualités de résonance du conduit vocal.

Un tel synthétiseur a été réalisé au « Laboratoire de la communication parlée et de l'instrumentation de mesure » de l'Enserg, à Grenoble. La qualité de la parole de synthèse est bonne. Cet appareil ne peut encore être utilisé en télécommunications, mais, par contre, il peut constituer une sortie parlée d'ordinateur.

Aux laboratoires de Marcoussis de la C.G.E., deux synthétiseurs à formants utilisant des circuits analogiques, ont été construits par G. Roger, le second avec l'assistance de R. Carré et J. Paillé de l'Enserg. Il est envisagé de construire un synthétiseur à circuits numériques. Le type de synthèse est une synthèse par mots : chaque mot prononcé est mis en mémoire sous forme d'une suite d'ensembles de paramètres (fréquence fondamentale, fréquences de plusieurs formants, amplitude vocale, nasalité) ; la phrase est alors constituée par la juxtaposition des suites de paramètres, correspondant à la suite des mots à prononcer.

C. Dechaux, chez Thomson-CSF a étudié sous contrat DRME, un synthétiseur à formants, faisant partie d'un codeur-décodeur dénommé Ciphon.

LES CIRCUITS DE FORMANTS

La fonction de transfert d'un circuit de formants doit être de la forme :

$$T(p) = \frac{1}{a^2 p^2 + 2x a p + 1}$$

C'est la fonction de transfert d'un circuit RLC (Fig. 8) :

$$\frac{V_s}{V_e} = \frac{1}{LCp^2 + RCp + 1}$$

Pour faire varier la fréquence de résonance, il faut modifier la capacité ; la bande passante est alors conservée. Mais il n'est

guère aisé de faire varier, dans une plage suffisamment large, cette capacité. J. Paillé, à l'Enserg, a expérimenté un autre type de circuit qui ne présente pas ce caractère défavorable (Fig. 9), et pour lequel la fonction de transfert s'écrit :

$$T(p) = \frac{3}{R^2 C^2 p^2 + \frac{2R^2 C p}{R'} + \frac{R^2}{R'^2} + 1}$$

On agit sur les résistances pour faire varier linéairement, soit la fréquence, soit la bande passante (**).

ON PEUT AUSSI SIMULER LE CONDUIT VOCAL...

Au lieu de reproduire les qualités de résonance du conduit vocal, on peut tenter de reproduire, centimètre par centimètre, la forme de notre conduit vocal. Les paramètres de commande sont alors les paramètres articulatoires. On simule ainsi le comportement acoustique d'un conduit de section variable correspondant à un conduit vocal simplifié, à l'aide de circuits électriques.

La réalisation de tels circuits de simulation devrait permettre de mieux connaître les qualités de résonance du conduit vocal (et donc de construire de meilleurs synthétiseurs à formants) ; en outre, des études parallèles sur la coordination de l'articulation, au niveau du cerveau, sont à mener : il devrait être possible d'en déduire des règles à respecter dans la commande des synthétiseurs à formants.

Deux études sur les synthétiseurs du conduit vocal sont menées en France. A l'Enserg, B. Guérin a mis au point une cellule simulant un élément du conduit vocal, et effectué des travaux sur la détermination d'une fonction d'aire du conduit vocal. Des travaux similaires ont été entrepris par M. Descout, au C.N.E.T. à Lannion : détermination de la fonction d'aire du conduit vocal à partir de la réponse impulsionnelle des lèvres, exploitation de documents cinéradiographiques.

(**) On lira avec intérêt l'article de J. Paillé, J.-P. Beauviola et R. Carré : « Description et utilisation d'un synthétiseur du type « à formants », paru dans la Revue de physique appliquée, en octobre 1970 (pp 785 à 792).



Photo 22. — C.-H. Coker, frappe sur un clavier le texte à dire à haute voix par l'ordinateur. Sur l'écran cathodique apparaît le conduit vocal que l'on simule, et dont on analyse le comportement.

(Cliché Bell Telephone Laboratories)

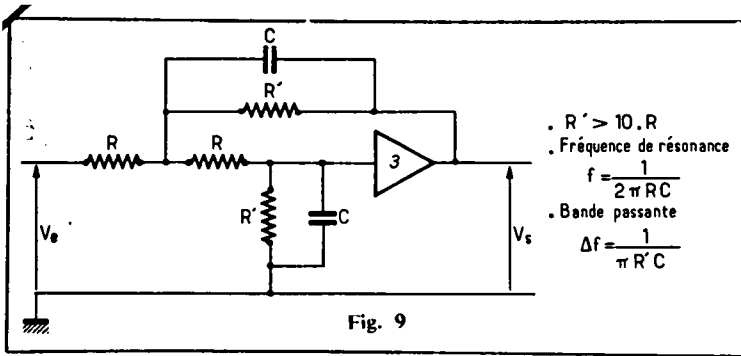


Fig. 9

...OU SA FONCTION DE TRANSFERT

On peut encore simuler, de façon globale, la fonction de transfert du circuit vocal : on a alors affaire aux synthétiseurs utilisant le codage prédictif. Il faut constituer un modèle variable, commandé de façon à minimiser l'écart entre la sortie réelle et celle du modèle. Des études sont en cours, à l'Enserg, au C.N.E.T., chez I.B.M. et à l'E.N.S.T.

(à suivre)
Marc FERRETTI.

Pour les personnes intéressées par l'étude et les applications de la communication parlée, notamment dans les domaines suivants : phonétique, télécommunications, acoustique architecturale, enseignement audiovisuel, dialogue homme-machine, etc., il existe un cours sur les structures acoustiques de la parole, organisé à l'université Paris VI par M. J.-S. Liénard. Une formation scientifique de base (niveau mathématiques élémentaires) est souhaitée.

Le programme de ce cours de formation permanente, pour l'an passé, était le suivant :

1. Généralités

— L'analyse spectrale en acoustique, théorie et pratique.

— Information, forme, structure.

— Les divers niveaux de la communication parlée : acoustique, phonétique, linguistique, sémantique.

— Audition et perception.
— Phonation.

2. Analyse acoustique de la parole

— Recherche du squelette informatif de la parole, invariance sémantique, stationnarité.

— Structure acoustique des éléments de parole; voyelles, consonnes, transitions; notions de phonème et de diagramme phonétique.

— Etude statistique des éléments de parole.

— Champs de liberté du locuteur : type de voix, netteté d'articulation, anamorphoses, mélodies, prosodie, rythme.

— Variabilité du message physique selon le locuteur : différences morphologiques, physiologiques, linguistiques.

— Identification des locuteurs.

3. Transmission et codage de la parole

— Estimation physique de la qualité d'un canal de transmission.

— Estimation subjective : tests d'intelligibilité et de qualité.

— Principales altérations de la parole : en acoustique architecturale, en électro-acoustique et en téléphonie. Les problèmes de la prothèse auditive.

— Les codages de la parole ayant pour but la réduction de sa bande passante, vocoder à canaux, à formants, à bande de base.

— Autres codages, à l'usage des malentendants : appareils transpositeurs, vocoders tactiles et visuels.

4. Synthèse de la parole

— Origines, évolution, intérêt actuel de la synthèse.

— Les synthétiseurs : analogues du conduit vocal, vocoder, relecteurs de sonagrammes, synthétiseurs paramétriques.

— Synthèse par règles, par éléments, réponse vocale automatique, transformation de texte en parole.

5. Reconnaissance automatique de la parole

— Reconnaissance des formes, méthodes générales, intérêt et limites.

— Reconnaissance automatique de la parole, diverses tendances et résultats actuels.

— Dialogue homme-machines, compréhension automatique.

Pour tous renseignements, s'adresser à M. Liénard au Laboratoire d'acoustique de l'université Paris VI, tour 66, 5^e étage, 4, place Jussieu, 75230 Paris Cedex 05.

CHAQUE MOIS LISEZ LES PETITES ANNONCES DU

HAUTE-PARLEUR

★ OFFRES D'EMPLOI

★ DEMANDES D'EMPLOI

★ FONDS DE COMMERCE

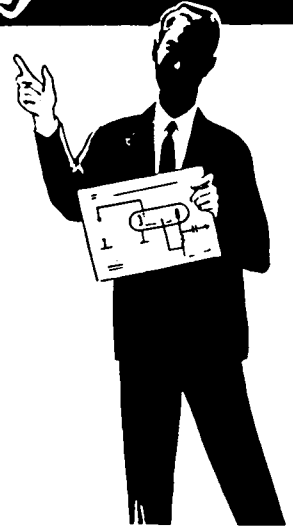
★ ACHAT DE MATÉRIEL

★ VENTE DE MATÉRIEL

LECTEURS DE PROVINCE :

N'oubliez pas de consulter NOTRE CARNET D'ADRESSES VOUS Y TROUVEREZ DU MATÉRIEL DE QUALITÉ !

1^{ère} Leçon gratuite



Sans quitter vos occupations actuelles et en y consacrant 1 ou 2 heures par jour, apprenez

LA RADIO ET LA TELEVISION

qui vous conduiront rapidement à une brillante situation.

● Vous apprendrez Montage, Construction et Dépannage de tous les postes.

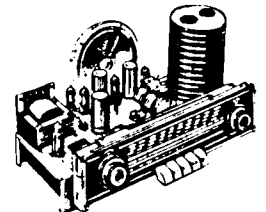
● Vous recevrez un matériel de qualité qui restera votre propriété.

Pour que vous vous rendiez compte, vous aussi, de l'efficacité de notre méthode, demandez aujourd'hui même, sans aucun engagement pour vous, la

Première leçon gratuite!

Si vous êtes satisfait, vous ferez plus tard des versements minimes à la cadence que vous choisirez vous-même. A tout moment, vous pourrez arrêter vos études sans aucune formalité.

SI VOUS HABITEZ EN FRANCE POSSIBILITÉ D'ÉTUDES GRATUITES AU TITRE DE LA FORMATION CONTINUE



Notre enseignement est à la portée de tous et notre méthode VOUS MERVEILLERA.

STAGES PRATIQUES SANS SUPPLEMENT

Documentation seule gratuitement sur demande.

Documentation + 1^{re} leçon gratuite

- contre 2 timbres à 0,50 (France)
- contre 2 coup-réponse (Etranger).

INSTITUT SUPERIEUR DE RADIO-ELECTRICITE

Etablissement privé

Enseignement à distance tous niveaux (Membre du S.N.E.C.)

27 bis, rue du Louvre, 75002 PARIS

(Métro : Sentier)

Téléphone : 231-18-67