Synthèse de la parole : description et utilisation d'un synthétiseur du type “ à formants ”

J. Paillé, J.P. Beauviala, R. Carré

► To cite this version:

J. Paillé, J.P. Beauviala, R. Carré. Synthèse de la parole : description et utilisation d'un synthétiseur du type “ à formants ”. *Revue de Physique Appliquée*, 1970, 5 (5), pp.785-793. 10.1051/rphysap:0197000505078500 . jpa-00243456

HAL Id: jpa-00243456

<https://hal.science/jpa-00243456>

Submitted on 1 Jan 1970

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SYNTHÈSE DE LA PAROLE : DESCRIPTION ET UTILISATION D'UN SYNTHÉTISEUR DU TYPE « A FORMANTS »

par J. PAILLÉ, J. P. BEAUVIALA, R. CARRÉ
Ecole Nationale Supérieure d'Electronique et de Radioélectricité
23, rue des Martyrs, Grenoble

(Reçu le 9 juin 1970)

Résumé. — Après avoir rappelé la théorie du fonctionnement de l'appareil vocal, on décrit un synthétiseur de parole du type « à formants » dont la fonction de transfert est équivalente à celle de l'ensemble des cavités du conduit vocal. Puis on développe quelques applications de cet appareil.

Abstract. — The theory of speech production is reminded. Then, a formant speech synthesizer the transfert function of which is equivalent to the vocal tract transfert function is described with some applications.

Introduction. — Les premiers synthétiseurs de parole furent construits pour vérifier l'hypothèse suivant laquelle le riche spectre de la parole pouvait être, sans qu'il en découle une dégradation de l'intelligibilité du message parlé, réduit à un petit nombre de composantes.

La parole est très redondante aussi bien sur le plan du signal lui-même que sur le plan linguistique. Nous savons que 40 000 unités d'information par seconde environ sont transmises par une ligne téléphonique classique au cours d'une conversation alors que notre cerveau, selon certaines hypothèses, ne pourrait décoder, en fin de traitement qu'un maximum de 50 unités d'information par seconde, ces 50 unités d'information étant porteuses de données sur le message ainsi que sur la qualité et sur les caractéristiques individuelles de la voix.

Le rôle et l'importance de telle ou telle composante du signal peuvent être étudiés au moyen de la synthèse.

Aussi, comme compléments indispensables des analyseurs, les synthétiseurs de parole sont des instruments fondamentaux en particulier dans le domaine des recherches sur les mécanismes de la production et de la perception de la parole.

Il existe divers types de synthétiseurs. Certains ne tiennent pas compte du mode de fonctionnement de notre appareil vocal. D'autres, au contraire, en tiennent étroitement compte.

Le synthétiseur de parole que nous allons décrire est du type « à formants ». La fonction de transfert des principaux circuits de cet appareil est équivalente à celle de l'ensemble des cavités en couplage de la bouche.

Nous pensons qu'un synthétiseur dont le fonctionnement simule au mieux celui de notre appareil vocal

est bien adapté pour la recherche des composantes fondamentales ou paramètres de la parole.

Après avoir rappelé quelques aspects de la théorie du fonctionnement de l'appareil vocal, nous décrirons le synthétiseur que nous avons construit, puis nous évoquerons diverses utilisations de cet appareil en recherche fondamentale et dans le domaine technique.

Rappels sur la théorie du fonctionnement de l'appareil vocal. — La parole en tant que phénomène physique 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 deuxième cas, le bruit est produit par l'action du courant d'air sur certaines parois de l'appareil vocal (à l'endroit d'un resserrement du conduit vocal par exemple). Ces deux phénomènes peuvent intervenir simultanément avec des intensités relatives plus ou moins grandes.

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

Nous allons brièvement indiquer les caractéristiques principales des sources ainsi que celles du conduit vocal.

SOURCE VOCALE ET SOURCE DE BRUIT. — La source vocale est essentiellement constituée par les poumons qui assurent une pression d'air à l'intérieur des bronches et de la trachée et par les cordes vocales, composées

de deux muscles plus ou moins tendus côte à côte dans la phase de phonation.

Ces deux muscles se séparent et se rapprochent alternativement sous l'action de la pression de l'air à l'intérieur de la trachée. La fréquence de ce phénomène de relaxation est essentiellement déterminée par la tension et la masse des cordes vocales ainsi que par la pression subglottique.

Le débit d'air D au niveau des cordes vocales est donc modulé à la fréquence de vibration de ces dernières (voir Fig. 1).

FIG. 1.

La forme du signal image du débit ne subit pas de modifications importantes en cours de phonation et l'enveloppe du spectre de ce signal de source présente un affaiblissement de l'ordre de 12 dB/octave (voir Fig. 2).

FIG. 2.

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

Le bruit peut être généré par un rétrécissement du conduit vocal, dans la mesure où certaines conditions de pression interne et de section du conduit sont réalisées.

La source de bruit est localisée sur le parcours du conduit vocal et sa situation peut varier avec le son prononcé.

L'impédance interne de cette source est considérée comme faible et purement réelle. Quant au spectre, on s'accorde à le trouver plat aux moyennes fréquences,

l'affaiblissement se manifestant côté basses fréquences au-dessous de 1 kHz et côté hautes fréquences au-dessus de 8 kHz [1].

CONDUIT VOCAL. — Le calcul de la fonction de transfert du conduit vocal $T(p) = D_s(p)/D_e(p)$ ($D_s(p)$ et $D_e(p)$ sont les transformées de Laplace du débit en sortie $D_s(t)$ et du débit à l'entrée $D_e(t)$ respectivement) a été effectué par G. Fant [2].

Dans le cas de pertes faibles et dans le cas d'une propagation suivant une seule dimension, la fonction de transfert peut être mise sous la forme :

$$T(p) = \prod_{n=1}^{\infty} \frac{1}{(\tau_n^2 p^2 + 2 \zeta_n \tau_n p + 1)}$$

avec $\tau_n = 1/\omega_n$ (ω_n est une pulsation propre).

Cette expression peut être simplifiée en ne tenant compte que des quatre premiers termes du produit et en faisant intervenir un terme correctif :

$$T(p) = \prod_{n=1}^4 \frac{1}{\tau_n^2 p^2 + 2 \zeta_n \tau_n p + 1} \times C_{r4}.$$

Le terme correctif C_{r4} a été déterminé par Fant ; il correspond à la fonction de transfert d'un circuit résonnant et d'un filtre passe haut montés en cascade.

Les cavités du conduit vocal modifient le spectre du signal transmis en favorisant certaines composantes qui correspondent aux fréquences des pôles de la fonction de transfert $T(p)$.

Sur la figure 3a, nous avons représenté le module de la fonction de transfert du conduit vocal et sur la figure 3b, le spectre du signal résultant de l'action de la source sur le conduit vocal.

FIG. 3a.

FIG. 3b.

Les régions fréquentielles correspondant à des maximums d'amplitude sont appelées 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 donc :

$$F_n = \frac{1}{2\pi\tau_n}.$$

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

$$\Delta F_n = \frac{\zeta}{\pi\tau_n}.$$

Cette bande passante est, en première approximation, constante dans la plage de variation d'un formant déterminé.

Dans le cas de production de bruit, la fonction de transfert à considérer est de la forme :

$$T(p) = \prod_{n=1}^{\infty} \frac{1}{\tau_n^2 p^2 + 2\zeta\tau_n p + 1} \times \left\{ \prod_{m=1}^{\infty} \tau_m^2 p^2 + 2\zeta\tau_m p + 1 \right\} \times (p - \tau_g).$$

L'apparition de zéros dans la fonction de transfert est due aux cavités en amont de la source de bruit. Le terme τ_g est un zéro réel qui correspond aux pertes glottiques.

Dans le cas d'une excitation des cavités nasales, la fonction de transfert peut se mettre sous la forme suivante :

$$T(p) = K \prod_{n=1}^f \frac{(\tau_n^2 p^2 + 2\zeta\tau_n p + 1)}{(\tau_n'^2 p^2 + 2\zeta\tau_n' p + 1)} \times \frac{C_{rg}}{\prod_{m=1}^g (\tau_m^2 p^2 + 2\zeta\tau_m p + 1)}.$$

Les fonctions de transfert évoquées précédemment permettent la détermination du débit D_s au niveau des lèvres en fonction du débit de la source.

Normalement, dans le cadre du processus phonation-audition, l'oreille réceptrice n'est pas excitée par le débit D_s au niveau des lèvres du locuteur mais par la pression P_s , qui en résulte, au niveau de l'oreille, à une distance l des lèvres. Il apparaît donc nécessaire d'associer aux fonctions de transfert précédentes celle de rayonnement $(P_s(p)/D_s(p))$ de la bouche.

Si on assimile la tête à une sphère, on a :

$$\frac{P_s}{D_s} = \frac{\rho p}{4\pi l} C_r(p).$$

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

Description d'un synthétiseur de parole. — Dans le synthétiseur de parole que nous avons réalisé, les caractéristiques des sources d'impulsions et de bruit

approchent au mieux celles des sources originales et la fonction de transfert de l'ensemble des circuits de transmission correspond à celle du conduit vocal.

SCHEMA SYNOPTIQUE. — Le schéma synoptique du synthétiseur que nous avons réalisé est représenté figure 4. La structure de l'appareil correspond, pour l'essentiel, à celle du synthétiseur OVE II construit par Fant [3].

FIG. 4.

Le synthétiseur comporte trois canaux : l'un pour la synthèse des sons vocaux, un deuxième pour la synthèse des sons non vocaux et un troisième pour la nasalité.

Ces canaux sont attaqués par deux sources : l'une est une source d'impulsions dont la fréquence F_s est commandée par une tension, l'autre est une source de bruit blanc. Des commandes de gain A_v , A_{BV} , A_N , A_B permettent le réglage des amplitudes des signaux de sources excitant les différents canaux et des commandes spéciales permettent de modifier les caractéristiques des fonctions de transfert des canaux selon la nature de la parole à synthétiser.

CIRCUITS DE SOURCE. — Le circuit de source vocale utilisée est classique (voir Fig. 5).

FIG. 5.

Un générateur délivre des impulsions de largeur 100 μ s. L'enveloppe du spectre de ces impulsions est

donc pratiquement plate dans la région fréquentielle de fonctionnement (50-5 000 Hz).

Un filtre adéquat modifie ensuite la forme de cette enveloppe.

Le générateur comporte un intégrateur, un double comparateur et une bascule R. S., ces circuits étant suivis d'un monovibrateur délivrant des impulsions de largeur 100 μ s.

En sortie de l'intégrateur, en supposant le signal d'entrée constant et positif, nous avons une montée linéaire de la tension. Lorsque cette tension atteint la valeur V , la bascule RS change d'état, par action du comparateur C_1 ; un signal négatif apparaît alors à l'entrée de l'intégrateur d'où résulte une décroissance linéaire en sortie jusqu'à la valeur 0 volt. Pour cette valeur, le comparateur C_2 agit sur la bascule RS ; la tension en sortie de cette bascule redevient positive et le cycle peut recommencer. La fréquence du générateur est donnée par la formule suivante :

$$F_s = \frac{k}{R_1 + R_2}.$$

On agit sur R_2 par découpage à haute fréquence, comme nous le verrons ultérieurement, pour faire varier linéairement la fréquence du générateur.

Le filtre disposé en sortie est un filtre passe bas « coupant en - 6 dB/oct. », de fréquence de transition égale à 200 Hz. Les caractéristiques de ce filtre correspondent à la décroissance moyenne du spectre de la source (chute en 12 dB/oct.) et à la croissance moyenne en + 6 dB/oct. due au rayonnement.

On n'a pas tenu ici des zéros de la source ni du facteur $C_r(p)$ de l'impédance de rayonnement.

Les caractéristiques de cette source approchent, en moyenne, celles de la source vocale mais une véritable simulation du fonctionnement des cordes vocales n'est pas envisagée ici. Les problèmes posés par cette simulation ont été récemment étudiés [4] [5].

La source de bruit est un générateur de bruit blanc constitué, pour l'essentiel, d'une diode Zener traversée par un faible courant. Son spectre est pratiquement plat dans la plage de fréquence 100-10 000 Hz.

CIRCUITS DE FORMANTS. — La fonction de transfert d'un circuit de formant doit être de la forme :

$$T(p) = \frac{1}{\tau^2 p^2 + 2 \zeta \tau p + 1}$$

avec

$$F = \frac{1}{2 \pi \tau} \quad \text{et} \quad \Delta F = \frac{\zeta}{\pi \tau}.$$

C'est, en particulier, la fonction de transfert d'un circuit RLC tel que représenté figure 6.

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

FIG. 6.

avec

$$F = \frac{1}{2 \pi \sqrt{LC}} \quad \text{et} \quad \Delta F = \frac{R}{2 \pi L}.$$

Pour faire varier F il faut modifier C , la bande passante ΔF étant conservée constante. Mais il n'est pas aisé de faire varier C dans un rapport 100 (F varie dans un rapport 10 pour le premier formant) et la relation liant C à F n'est pas simple.

Aussi avons-nous expérimenté un nouveau type de circuit [8] qui ne présente pas ces caractères défavorables (Fig. 7).

FIG. 7.

Sa fonction de transfert est :

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

Si $R' > 10 R$, le facteur de surtension est plus grand que 5 (ce qui est toujours vrai pour cette application) et on a, avec une approximation meilleure que 0,5 % :

$$F = \frac{1}{2 \pi (R_1 + R_2) C} \quad \text{et} \quad \Delta F = \frac{1}{\pi (R'_1 + R'_2) C}.$$

On agit soit sur R_2 soit sur R'_2 par découpage à haute fréquence pour faire varier linéairement soit F soit ΔF .

Le rapport signal/bruit de ce circuit est de l'ordre de 60 dB.

Un circuit de même type mais dont la fonction de transfert comporte un zéro complexe a été mis au point (Fig. 8).

FIG. 8.

La fonction de transfert de ce circuit est la suivante :

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

Comme dans le cas du circuit de formant, si $R' > 10 R$ on a, avec une approximation meilleure que 0,5 % :

$$F = \frac{1}{2 \pi R C} \quad \text{et} \quad \Delta F = \frac{1}{\pi R' C}.$$

Pour simuler le canal vocal, il faut associer aux circuits de formants un circuit de correction de fonction de transfert C_{r4} (si 4 circuits de formant sont utilisés). Cette fonction de transfert est égale à [1] :

$$C_{r4} = \frac{(p + 2 \pi \cdot 2000)^2}{((p + \pi \cdot 4000)^2 + (2 \pi \cdot 4000)^2)} \times \frac{(2 \pi \cdot 4000)^2}{((p + \pi \cdot 400)^2 + (2 \pi \cdot 4000)^2)}.$$

FIG. 9.

La fonction de transfert du circuit représenté figure 9, où $C = 10^{-8} F$, $R_1 = 10^5/4 \pi \Omega$ et $R_2 = 10^5/8 \pi \Omega$, est la suivante :

$$T_1 = \frac{(p + 2 \pi \cdot 2000)^2}{((p + \pi \cdot 4000)^2 + (2 \pi \cdot 4000)^2)} \quad (\text{premier facteur de } C_{r4})$$

FIG. 10.

La fonction de transfert du circuit représenté figure 10, où $A = 2,9$, $C = 10^{-8} F$ et $R = 10^5/8 \pi \Omega$, est la suivante :

$$T_2 = \frac{(2 \pi \cdot 4000)^2}{((p + \pi \cdot 400)^2 + (2 \pi \cdot 4000)^2)}$$

(deuxième facteur de C_{r4}).

COMMANDES DE GAIN. — Les commandes de gain réglant les amplitudes des signaux de source dans les différents canaux suivent une loi exponentielle ; ainsi les tensions de commande appliquées à l'entrée sont des images des gains en décibels.

FIG. 11.

On a alors exploité le schéma représenté figure 11. Ce circuit, constitué pour l'essentiel d'une diode Zener Z à caractéristique logarithmique, effectue la transformation : $V_s = e^{V_e}$ sur 60 dB.

Les résistances R_1 , R_2 et la tension V sont ajustées pour faire travailler la diode Z dans sa plage de fonctionnement. La résistance R_3 contrôle l'amplitude maximale de la tension de sortie.

FIG. 12.

Le schéma représenté figure 12 est le circuit de commande de gain proprement dit. On a :

$$G = (R'/R_1 + R_2).$$

On agit sur R_2 par découpage à haute fréquence pour faire varier linéairement le gain.

PRINCIPE DES CIRCUITS DE COMMANDE. — Pour assurer la commande automatique des fréquences de résonance des filtres, des bandes passantes ou bien des gains, il faut agir sur les parties R_2 des résistances R notées sur les différents schémas (Fig. 13).

FIG. 13.

Pour cela, R_2 est découpée à haute fréquence (par rapport à la plus haute fréquence de travail du synthétiseur) par un signal rectangulaire de rapport cyclique τ/T . On peut montrer que la conductance moyenne de R est proportionnelle à τ/T .

$$\frac{1}{R_1 + R_2} = \frac{1}{R_1 + R_2} + \left[\frac{1}{R_1} - \frac{1}{R_1 + R_2} \right] \frac{\tau}{T}.$$

On a donc :

$$F = \frac{1}{2\pi C} \left[\frac{1}{R_1 + R_2} + \left(\frac{1}{R_1} - \frac{1}{R_1 + R_2} \right) \frac{\tau}{T} \right]$$

si $\tau/T = 0$

$$F = \frac{1}{2\pi C(R_1 + R_2)}$$

si $\tau/T = 1$

$$F = \frac{1}{2\pi C R_1}.$$

La fréquence varie donc linéairement avec τ et nous pouvons fixer d'une manière très précise les limites de variation en jouant sur R_1 et $R_1 + R_2$.

De même, nous avons :

$$\Delta F = \frac{1}{\pi C} \left[\frac{1}{R'_1 + R'_2} + \left(\frac{1}{R'_1} - \frac{1}{R'_1 + R'_2} \right) \frac{\tau}{T} \right]$$

F_s (fréquence du générateur d'impulsions)

$$= k \left[\frac{1}{R_1 + R_2} + \left(\frac{1}{R_1} - \frac{1}{R_1 + R_2} \right) \frac{\tau}{T} \right]$$

$$\text{et } G = R' \left[\frac{1}{R_1 + R_2} + \left(\frac{1}{R_1} - \frac{1}{R_1 + R_2} \right) \frac{\tau}{T} \right].$$

COMMANDE ANALOGIQUE. — Pour élaborer un signal rectangulaire de rapport τ/T variable nous avons utilisé le circuit représenté figure 14.

Nous disposons d'un générateur de dents de scie dont le principe a été décrit précédemment.

Cette dent de scie est comparée à la tension analogique de commande V_{ca} comprise entre 0 et V volts.

A la sortie du comparateur, le signal est de forme rectangulaire et de rapport cyclique $\tau/T = V_{ca}/V$.

FIG. 14.

COMMANDE DIGITALE. — Le circuit représenté figure 15 délivre un signal rectangulaire de rapport cyclique τ/T proportionnel à l'état numérique imposé aux entrées E .

FIG. 15.

L'état d'un compteur dont l'avance est commandée par une horloge de fréquence $F_H = 2^n/T$ est comparé au signal de commande présenté sous forme digitale $E_1 \dots E_{2n}$.

Lors de l'égalité une impulsion est délivrée par le comparateur et commande l'une des entrées d'une bascule RS, l'impulsion de passage à zéro commandant l'autre entrée.

A titre d'exemple, signalons que l'on attribue 64 niveaux de quantification pour le paramètre fréquence d'excitation de la source.

ORGANE D'ENTRÉE. — Le synthétiseur que nous venons de décrire peut être commandé soit par un ordinateur soit par tout appareil pouvant délivrer une dizaine de signaux simultanés images des paramètres de la parole.

Pour effectuer cette commande, nous avons construit un lecteur de courbes à caméra de télévision. Les courbes sont dessinées sur du papier et représentent

les évolutions dans le temps des paramètres de la parole (mélodie, formants, amplitudes, etc...). Ces évolutions ont été déterminées, au préalable, par analyse d'une parole originale.

FIG. 16.

Le lecteur est un ensemble de télévision en circuit fermé constitué d'une caméra à tube vidicon et d'un téléviseur de contrôle [7]. La durée d'un balayage ligne est de $46 \mu\text{s}$, la durée d'un balayage trame est de 20 ms. Le signal vidéo est exploité par deux circuits échantillonneurs.

Nous avons représenté figure 16 une trame de balayage ainsi que les courbes images de paramètres. A chacune de ces courbes est associé un repère tracé à gauche du dessin et qui représente le début de la plage de variation du paramètre considéré.

Les deux échantillonneurs transmettent le signal vidéo un temps déterminé Δt après les impulsions de

synchronisation ligne. Pour le 1^{er} échantillonneur, le temps Δt est constant et égal à T_0 pour permettre la détection des repères $T_1, T_2, \dots$; pour le 2^e échantillonneur, Δt est soit constant (pour l'étude de sons soutenus) soit lentement variable (pour l'étude de phrases), avec $\Delta t = T = kt$, l'exploitation totale de l'image pouvant être effectuée en 2,4 secondes, durée d'une analyse effectuée avec l'appareil classique « Sona-graph ».

On a donc, après échantillonnage, un signal image des informations se trouvant sur une ligne verticale déterminée (voir Fig. 17). Les maximums ($T_1, T_2, \dots$; $T'_1, T'_2, \dots$) des signaux échantillonnés sont détectés par des circuits spéciaux [8].

FIG. 17.

La mesure des temps $T_n - T'_n$, avec

$$T_n - T'_n = k \times 46 \mu\text{s},$$

k étant un entier, donne une image des valeurs des paramètres. Cette mesure est effectuée par un compteur digital dénombrant les k lignes entre le repère T'_n et le paramètre T_n .

FIG. 18.

Enfin des convertisseurs digitaux analogiques délivrent des tensions images des paramètres (Fig. 18).

Le lecteur de courbes que nous avons construit peut détecter 9 paramètres. Sa précision est limitée par la qualité de la prise de vue et par le nombre de lignes par trame.

PERFORMANCES ET UTILISATIONS. — La construction d'un synthétiseur à formants présente de sérieuses difficultés. En effet, la commande automatique des fréquences de circuits résonnants pose de difficiles

problèmes et les solutions qui furent adoptées jusqu'à présent sont ou bien très complexes et conduisent à des matériels encombrants ou bien peu précises.

Nous avons mis au point de nouveaux circuits de formants lesquels nous ont permis la réalisation d'un synthétiseur très précis, compact, pouvant être commandé, soit par des signaux analogiques, soit par des signaux digitaux.

Les paramètres de commande peuvent être réduits à 8 ou 10 (dans un vocoder classique on en compte

au moins 16), la parole de synthèse demeurant intelligible.

Dans ce cas, il faut une ligne de capacité 1 000 bits/s pour transmettre les signaux relatifs à ces paramètres (dans un vocoder cette capacité est de 2 400 bits/s). Pour des travaux particuliers, on pourra augmenter le nombre des commandes (commandes des bandes passantes par exemple) et effectuer des synthèses plus fines.

La parole de synthèse obtenue à l'aide de notre appareil est de bonne qualité et paraît naturelle. Mais il est très difficile d'effectuer des comparaisons globales. Il faudrait faire des tests de perception dans des conditions réputées identiques et ceci pose de nombreux problèmes de psychoacoustique.

En tout état de cause, on ne peut compter sur le jugement de l'opérateur. Ce dernier entend ce qu'il compte entendre ; il est habitué à la parole de synthèse et aux défauts de son appareil.

Les synthétiseurs du type « à formants » tels celui que nous avons construit sont utilisés pour effectuer de nombreux travaux de recherches.

La théorie de Fant sur la production de la parole [2] a été vérifiée à l'aide de synthétiseurs à formants. Des problèmes plus limités peuvent aussi être étudiés tels ceux posés par la formation de certains sons. Ces travaux permettent de préciser certaines lois, certaines contraintes particulières sur les transitions et, en fin de compte, de proposer des modèles standards permettant d'effectuer des synthèses par règles.

Les synthétiseurs peuvent aussi être utilisés pour préciser l'analyse. Il est parfois difficile de mesurer un paramètre, un formant par exemple. C'est souvent le cas pour l'étude des voix de femmes dont le spectre est caractérisé par un nombre réduit de composantes d'où ne se dégage pas une image suffisamment précise de la structure formantique. On peut faire une première mesure entachée d'erreurs et effectuer ensuite une synthèse. Les spectres de la parole originale et de la parole de synthèse sont alors comparés ; en tenant compte des différences observées, on peut arriver, par retouches successives, à une bonne identité. Mais si, en définitive, une différence apparemment irréductible subsiste, cette différence peut être à l'origine de la détection de traits intéressants qu'il faudra interpréter. Si un trait pertinent nouveau est révélé, il pourra être, à son tour, simulé et intégré à la synthèse. Ainsi, l'analyse puis, en conséquence, la synthèse, peuvent être progressivement améliorées.

Ce processus d'« analyse par la synthèse » est à noter puisque, selon certains, un processus analogue interviendrait au niveau de notre ensemble de production et de perception de la parole.

Les synthétiseurs peuvent être utilisés pour étudier le rôle et l'importance des paramètres : quel est le rôle du 3^e formant, par exemple, quelle doit être la fréquence minimale d'échantillonnage de tel paramètre, quelle quantification faut-il adopter pour coder

chacun des paramètres ? Des travaux sur synthétiseurs peuvent conduire à des résultats très intéressants.

Enfin, des tests de perception peuvent être effectués à l'aide d'un synthétiseur. Les sons de synthèse étant rigoureusement connus et « calibrés », des tests peuvent être entrepris pour étudier le comportement de notre organe de perception. Des travaux de ce type ont déjà été effectués mais ces recherches doivent être considérablement développées.

Outre l'utilisation des synthétiseurs en recherche fondamentale, leur utilisation dans le domaine technique tend à devenir une réalité. Les synthétiseurs à formants peuvent être intégrés à des ensembles d'analyse-synthèse pour télécommunications à bande étroite. Les travaux sur de tels ensembles en sont encore au stade expérimental, les problèmes d'analyse en temps réel étant très difficiles à résoudre. Mais une autre application directement exploitable actuellement consiste en l'utilisation du synthétiseur à formants comme organe de sortie parlée d'ordinateur. La capacité des mémoires utilisées pour stocker l'information parlée est alors limitée puisqu'un débit d'informations de 1 000 bits/s suffit pour commander le synthétiseur. Ce débit pourrait d'ailleurs être encore réduit par la mise en œuvre de techniques de synthèses par règles en cours de développement.

Conclusions. — Dans les années à venir, les synthétiseurs de parole seront des appareils couramment utilisés dans des ensembles de sortie parlée d'ordinateurs et dans les laboratoires effectuant des recherches sur la parole.

Une progression des études fondamentales sur la production et la perception de la parole est liée au développement de ces appareils de synthèse. Le synthétiseur que nous avons construit paraît adapté, dès maintenant, à de nombreux travaux. Il est précis, souple d'emploi et peut être commandé par les paramètres jugés intéressants présentés soit sous forme analogique soit sous forme digitale.

Cependant, ce type d'appareil doit constamment évoluer en fonction des résultats obtenus par son exploitation.

D'ores et déjà, nous disposons d'un nombre suffisant d'éléments pour pouvoir apprécier le rôle et l'importance d'une simulation correcte de la source vocale.

Enfin, signalons l'existence des synthétiseurs de type différent du type « à formants » et dont les caractéristiques sont, elles aussi, intéressantes. C'est le cas, en particulier, du simulateur du conduit vocal qui reproduit, centimètre par centimètre, la forme de notre conduit vocal. Les paramètres de commande sont, alors, des paramètres articulatoires. Des travaux de synthèse par règles, effectués à l'aide de cet appareil doivent tout naturellement être associés à des études sur la coordination de l'articulation au niveau du cerveau.

Bibliographie

- [1] FLANAGAN (J. L.), Speech analysis, synthesis and perception. Springer-Verlag, 1965.
- [2] FANT (G.), Acoustic theory of speech production, Mouton, 1960.
- [3] FANT (G.), MARTONY (J.), RENGMAN (U.), RISBERG (A.), Proc. of the Speech Communication Seminar, Stockholm, 1962.
- [4] CARRÉ (R.), BEAUVIALA (J. P.), PAILLÉ (J.), Article à paraître dans la revue *I. E. E. on Audio and Electroacoustics*, sept. 1970.
- [5] BEAUVIALA (J. P.), *Revue d'Acoustique*, 1968, **3-4**, 235.
- [6] CARRÉ (R.), *Electronique*, 1968, **79**, 173.
- [7] PAILLÉ (J.), LANCIA (R.), Colloque International de Téléinformatique, Paper 250, Paris, 1969.
- [8] PAILLÉ (J.), *Revue d'Acoustique*, 1969, **6**, 111.