# Synthétiseur vocal à formants

Jean-Louis ZUBER
Lycée technique Jules Siegfried, 76600 Le Havre

## RÉSUMÉ

Après avoir présenté le principe de la parole humaine, on donne le principe d'un circuit intégré à synthèse vocale, le P.C.F. 8200.

L'algorithme des filtres numériques simulant le conduit vocal a été trouvé à partir d'un filtre analogique passe-bas résonnant (méthode par identification des pôles).

A partir de cet algorithme nous avons fait un programme pour visualiser le spectre de quelques phonèmes.

**Mots clefs**

Synthèse vocale - Formants - Filtrage numérique - Résonateurs - Voix - Son - Phonèmes.

## PLAN



## INTRODUCTION

Nos étudiants de T.S. électronique ont été amenés à concevoir un système à synthèse vocale, dans le cadre d'un P.A.E. mené en collaboration avec un laboratoire de l'université du Havre : le Laboratoire d'Analyse et de COmmande des Systèmes, dirigé par le professeur FAURE.

Il a fallu se documenter sur le principe physique de la parole humaine puis sur celui des synthétiseurs vocaux. Cet article a pour but de présenter quelques résultats de nos recherches bibliographiques et des études que nous avons menées sur le circuit intégré choisi pour la réalisation.

## 1. PRODUCTION ET TRANSMISSION DE LA PAROLE

Ce sont les travaux au XIX$^{e}$ siècle, de J. Fourier (décomposition analytique des ondes), et de H. von Helmholtz (Résonateurs) puis de L. Hermann (théorie des formants) qui ont permis de comprendre le mécanisme de la parole.

### 1.1. Structure générale du système vocal humain

La structure est réprésentée schématiquement à la figure 1.

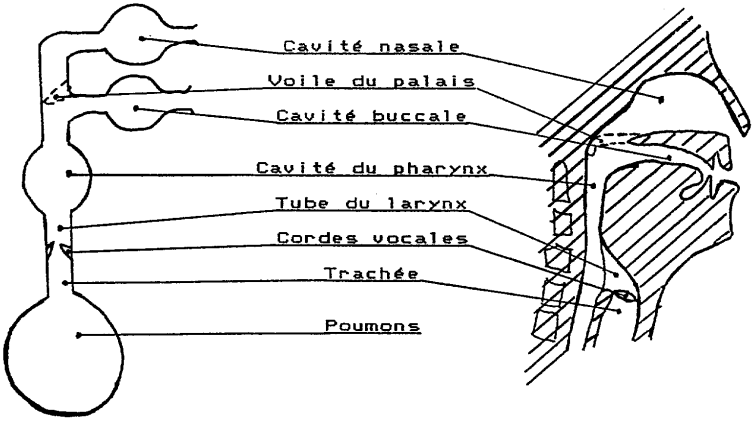



**Figure 1** **Figure 2**

Les poumons servent de soufflerie qui fait vibrer les cordes vocales. L'onde sonore qui en résulte résonne dans les cavités pharyngienne et buccale. C'est en modifiant la géométrie de ces cavités, donc en changeant leurs fréquences de résonance, que nous pouvons produire différents sons ou phonèmes. Les phonème sont les plus petits éléments du langage articulé. Il y en a une quarantaine pour la langue française.

### 1.2. Vibrations des cordes vocales

#### 1.2.1. Les sons voisés

Les sons voisés correspondent aux voyelles et à certaines consonnes dites sonores (b, d, g, l, m, n, r, v, z). Leur production se fait à partir d'un mouvement périodique des cordes vocales (oscillations de relaxation).

Au début les cordes vocales sont fermées et empêchent l'air de passer. Les poumons en augmentant la pression sur les cordes vocales les ouvrent, l'air s'échappe, la pression diminue, et à cause de leur élasticité, les cordes se referment ; puis la pression augmentant, elles s'écartent à nouveau et ainsi de suite. Nous avons la création d'une onde de pression. L'ouverture des cordes vocales est lente et la fermeture est rapide, donc la vibration périodique a à peu près la forme de dents de scie. Or le signal en dents de scie est parmi les signaux périodiques les plus riches en harmoniques. Ainsi les fréquences pouvant exciter les résonateurs (essentiellement la cavité pharyngo-buccale) sont nombreuses.

La fréquence fondamentale de ce signal périodique se situe en moyenne à 100 Hz pour une voix masculine et au double pour la voix féminine.

#### 1.2.2. **Les sons non voisés**

Ces sons correspondent à des consonnes fricatives (ch, f, s,...) ou certaines plosives (k, p, t,...). Elles sont plus sourdes que les précédentes. Pour ces sons les cordes vocales restent entrouvertes, l'air passe de façon continue créant des turbulences dans le conduit vocal qui peut être plus ou moins obstrué par un obstacle (langue, lèvres...). On peut faire une occlusion totale mais momentanée (p, t, ...) ou partielle (ch, f, ...).

On peut sentir en mettant la main au niveau de la gorge que les cordes vocales ne vibrent pas pour les consonnes sourdes.

### 1.3. **Les résonateurs : les cavites pharyngienne et buccale**

#### 1.3.1. **Modèle simple du conduit vocal**

En première approximation, on peut assimiler le conduit vocal à un tube de diamètre constant et de longueur L. Ce tube est fermé du côté des cordes vocales et ouvert du côté des lèvres. Si on appelle c la célérité du son dans l'air, les fréquences de résonance sont :

$$f = \frac{c\,(2\,.\,k + 1)}{4\,.\,L} \qquad k = 0, 1, 2, 3...$$

Si on considère que le conduit vocal est un tube de 17 cm de long et que la vitesse de propagation du son est c = 340 m/s on trouve des fréquences de résonance à 500 Hz, 1 500 Hz, 2 500 Hz, 3 500 Hz etc. ce qui correspond à des fréquences proches du son (œ) de «leur».

Lorsque la célébrité augmente, les fréquences augmentent. Si on parle dans l'hélium, on imite parfaitement la voix de Donald.

### 1.3.2. **Les formants**

Si on prononce un phonème dans un microphone relié à un analyseur de spectre, le spectre d'amplitude présente un certain nombre de pics correspondant aux fréquences de résonance du conduit vocal. Ces pics sont appelés les formants du phonème. Ainsi chaque phonème est caractérisé par ses formants que l'on repère par la fréquence de leur maximum.

Lorsque nous parlons, les cavités résonantes changent sans cesse de forme et de volume, les fréquences de résonance se trouvent sans cesse déplacées et par conséquent ce ne sont pas toujours les mêmes plages de fréquences harmoniques du signal en dents de scies qui sont amplifiées. C'est ce qui fait la richesse des phonèmes émis par l'être humain.

Le premier formant qui correspond à la plage des fréquences les plus basses est lié au volume de la cavité pharyngienne. La cavité buccale étant plus courte amplifie les fréquences harmoniques plus hautes qui forment le deuxième formant.

Jusqu'à 5 kHz (Bande passante du circuit que nous avons étudié), le conduit vocal des hommes donne cinq formants, celui des femmes, plus petit, donne 4 formants (voir figure 3).

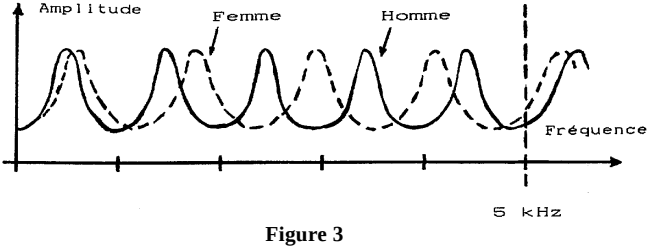



**Figure 3**

### 1.3.3. **La cavité nasale**

La cavité nasale n'intervient pas, lorsqu'on prononce les voyelles dites «orales» (a, e, i, ...). Pour d'autres voyelles, le voile du palais s'abaisse et une partie de l'air s'échappe par la cavité nasale : on obtient ainsi les voyelles «nasales» (an, in, en ,on...). Par exemple à la voyelle orale (ɔ) de «port» correspond la voyelle nasale (ɔ̃) de «pont». La

cavité nasale intervient plutôt comme un filtre antirésonant, atténuant certains formants de la voyelle orale correspondante.

## 2. PRINCIPE DE LA SYNTHÈSE VOCALE À FORMANTS

Le circuit que nous avons étudié est un circuit intégré de 24 broches : le PCF 8200 de RTC ; c'est le jeune frère du MEA 8000. Le PCF 8200 ayant une bande passante de 5 kHz intègre cinq filtres résonants permettant de donner cinq formants pour une voix masculine et quatre pour une voix féminine. (le MEA 8000 n'a que 4 filtres pour une bande passante de 4 kHz).

### 2.1. STRUCTURE GÉNÉRALE DU CIRCUIT INTÉGRÉ

Nous avons indiqué sur le schéma synoptique de la figure 4 les parties analogues au système vocal humain.

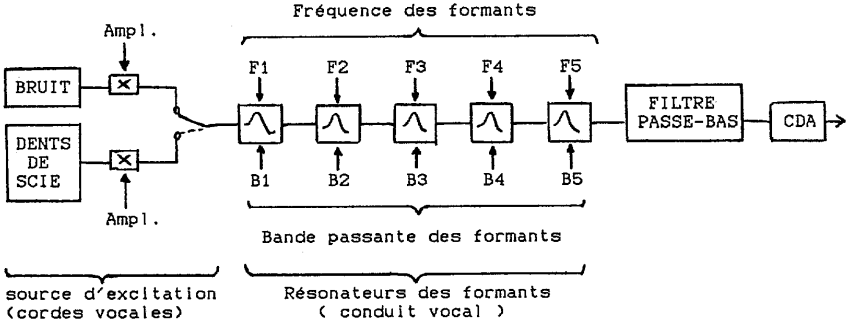



**Figure 4**

Le générateur de bruit blanc est utilisé pour les sons non voisés (consonnes sourdes) et le générateur de dents de scie pour les sons voisés. Une amplification du signal initial permet de moduler les différents phonèmes pour donner l'intonation.

Cinq filtres numériques en cascade donnent les formants du phonème. Ces filtres ont une fréquence d'échantillonnage $f_e$ de 10 kHz. La table des codes correspondant à leur fréquence de résonance (F1, F2...) et leur bande passante (B1, B2...) est dans une mémoire annexe.

Un filtre passe-bas ne garde que la partie utile du spectre jusqu'à 5 kHz ($f_e/2$) en coupant les bandes images qui proviennent de l'échantillonnage du signal (théorème de Shannon).

Après le convertisseur digital-analogique intégré, il faut prévoir un étage d'amplification de puissance pour le haut parleur.

### 2.2. **Générateur pour les sons voisés**

Ce générateur donne un signal en dents de scie (figure 5).

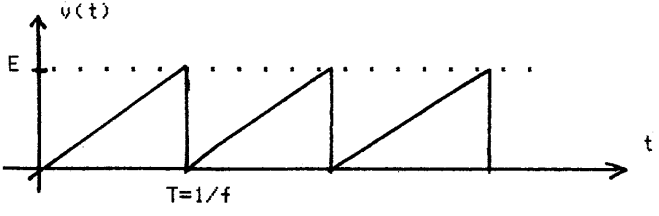



**Figure 5**

La décomposition en série de Fourier d'un tel signal est :

$$v(t) = \frac{E}{2} + \sum_n V_n \,.\, \sin(2\,.\,\pi\, nf\, t) \qquad \text{avec} \quad V_n = -\frac{E}{\pi} \times \frac{1}{n}$$

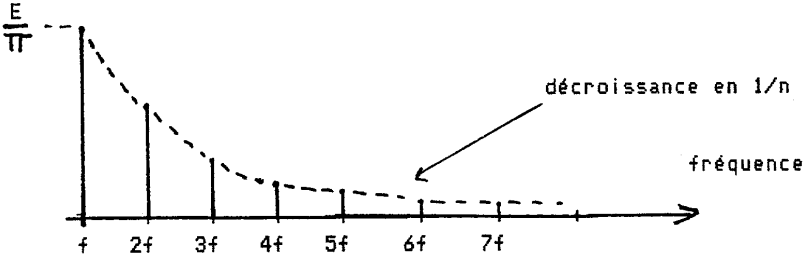



**Figure 6**

Le spectre d'amplitude de la figure 6 montre l'importance des harmoniques par leur nombre et par leur amplitude. La décroissance en fonction de la fréquence est de 6 dB par octave. A titre de comparaison, la série de Fourier d'un signal triangulaire ne comprend que des harmoniques impairs avec une atténuation de 12 dB par octave.

### 2.3. **Étude des filtres numériques des formants**

#### 2.3.1. **Filtre analogique correspondant**

Chaque filtre est utilisé pour un formant. Ce sont des filtres du deuxième ordre du type passe-bas ; à cause de leur facteur de qualité assez grand ils présentent un pic de résonance. Chaque filtre est caractérisé par deux paramètres : sa fréquence de résonance et sa bande passante à – 3 dB.

Le filtre analogique équivalent a pour fonction le transfert (p = jω)

$$H_a(p) = \frac{1}{1 + 2.m.\frac{p}{\omega_0} + \frac{p^2}{{\omega_0}^2}}$$

avec m le coefficient d'amortissement et $\omega_0$ la pulsation propre. On se trouve dans le cas où m est inférieur à 0,7 ; la courbe du gain correspondante a donc la forme de la figure 7.

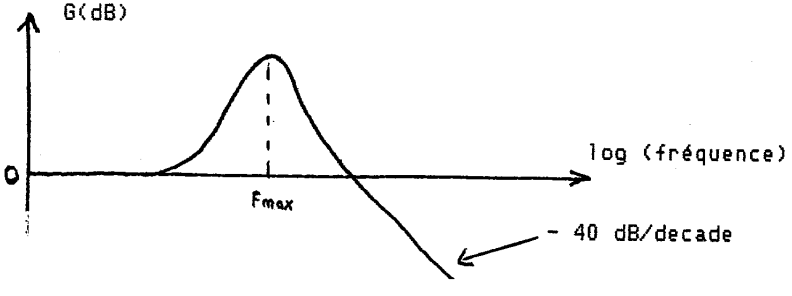



**Figure 7**

Les pôles de cette fonction H a (p) sont :

$$p_1 = -m.\omega_0 + j.\omega_0\sqrt{1-m^2} \text{ et } p_2 = -m.\omega_0 - j.\omega_0\sqrt{1-m^2}$$

Les filtres sont suffisamment résonants pour qu'on puisse négliger $m^2$ devant 1 et identifier la fréquence du maximum avec la fréquence propre $f_0$. De plus, en faisant apparaitre la bande passante à – 3dB notée Δf. ($\Delta f = 2.m.f_0$), on obtient :

$$p_1 = -\pi\Delta f + j\,2.\pi.f_0 \text{ et } p_2 = -\pi.\Delta f - j\,2.\pi.f_0$$

### 2.3.2. **Synthèse du filtre numérique**

Pour synthétiser le filtre numérique équivalent, on cherche celui qui a des pôles correspondants, et donc une fonction de transfert en z qui a la forme suivante :

$$H_n(z) = \frac{1}{[1 - z^{-}1 \,.\, \exp(p_1 \,.\, T_e)]\,[1 - z^{-1} \,.\, \exp(p_2 \,.\, T_e)]}$$

$T_e$ est la période d'échantillonnage du filtre numérique ; on a $1/T_e = f_e = 10$ kHz pour le PCF 8200.

soit

$$H_n(z) = \frac{1}{1 - B \,.\, F \,.\, z^{-1} + B^2 \,.\, z^{-2}}$$

avec $B = \exp(-\pi \,\Delta f \,.\, T_e)$

$F = 2 \,.\, \cos(2 \,.\, \pi \,.\, F_0 \,.\, T_e)$

Ces coefficients correspondent aux codes Bi et Fi (i = 1, 2, 3, 4, 5) du schéma de la figure 4.

On en déduit le schéma synoptique à trois multiplicateurs de la figure 8.

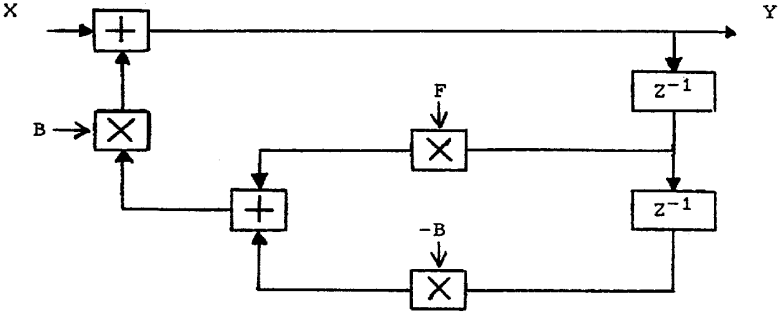



**Figure 8**

**Remarque :** le constructeur aurait pu concevoir un circuit en ne faisant intervenir que 2 multiplicateurs ($A = -B^2$ et $C = B.F$), mais dans ce cas les coefficients C dépendent de la bande passante $\Delta f$ et de la fréquence de résonance $f_0$ il faut donc une table bidimensionnelle pour les mémoriser ce qui augmente la capacité de la mémoire nécessaire.

En utilisant la relation $z^{-1} = \exp(-j \,.\, 2 \,.\, \pi\, f \,.\, T_e)$ on trouve la fonction de transfert isochrone qui a pour amplitude :

$$|\,\underline{H}_n\,(j \,.\, 2\pi f)\,| = \frac{1}{\sqrt{R^2 + I^2}}$$

$$\text{avec} \quad R = 1 - B.F.\cos 2\pi f.T_e + B^2 \cos 4\pi\, f.T_e$$

$$I = B.F.\sin 2\pi f.T_e - B^2 \sin 4\pi f.T_e$$

La courbe de la Figure 9 représente $H_n(j.2\pi f)$ pour des fréquences allant de 0 à $f_e$ = 10 kHz.

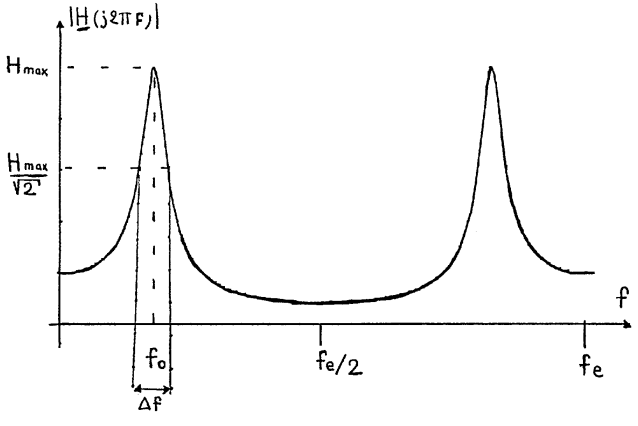



**Figure 9**

On voit l'utilité du filtre passe-bas coupant les fréquence supérieures à la fréquence de Shannon.

### 2.3.3. **Simulation du signal de sortie**

Les filtres étant en cascade il suffit de multiplier les cinq fonctions de transfert des cinq filtres de formants et de tenir compte de la décroissance en 1/n (n étant le numéro de l'harmonique du signal en dents de scie) pour obtenir le spectre théorique du signal de sortie.

Par simulation sur ordinateur, nous avons pu visualiser les spectres théoriques de différents phonèmes. Les valeurs des fréquences de résonance et des bandes passantes des formants sont celles que nous avons programmées pour le PCF 8200 pour obtenir le phonème correspondant.

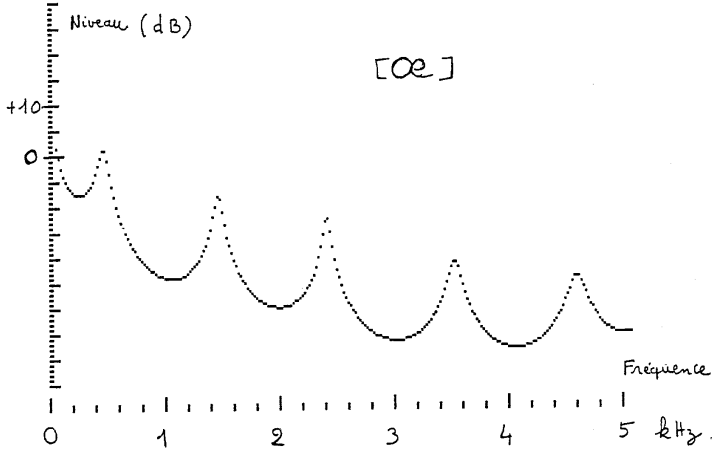
Les spectres réels des phonèmes créés par le synthétiseur sont évidemment des spectres de raies espacées d'un intervalle correspondant au fondamental du générateur de dents de scie. Nous avons préféré la simulation sous forme de spectre continu qui correspond à l'enveloppe du spectre réel, parce que cela permet une meilleure visualisation des formants.

## 3. VISUALISATION DE SPECTRES DE QUELQUES PHONÈMES

Nous ne présentons que des spectres de sons correspondant à des voyelles.

### 3.1. Phonème 'e' ouvert (œ)

On remarque la régularité des fréquences de résonance qui sont assez proches de celles d'un tube de 17 cm de long, fermé à un bout.

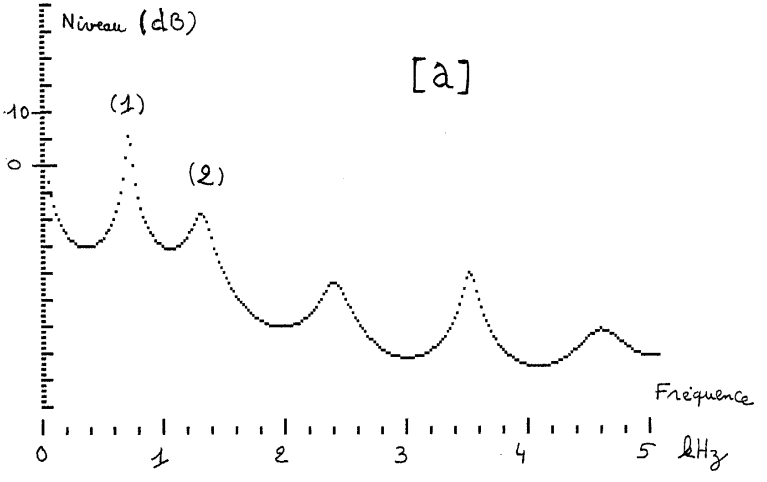


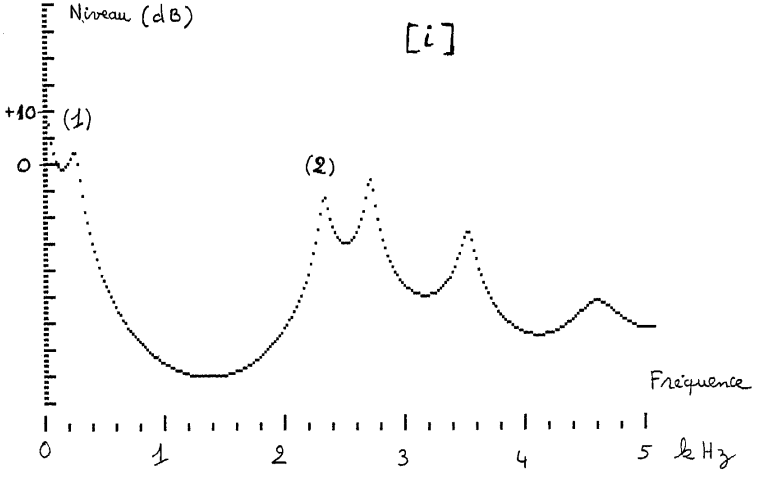

### 3.2. Phonème 'a' et phonème 'i'

Les deux premiers formants sont assez proches pour le 'a' (725 Hz et 1300 Hz) et beaucoup plus écartés pour le 'i' (250 Hz et 2200 Hz).

En effet la cavité pharyngienne est plus grande lorsqu'on prononce le 'i' : donc la fréquence du premier formant est plus petite pour 'i'.

Par contre, la cavité buccale est plus grande lorsqu'on prononce le 'a' : c'est la fréquence du deuxième formant du 'a' qui est plus



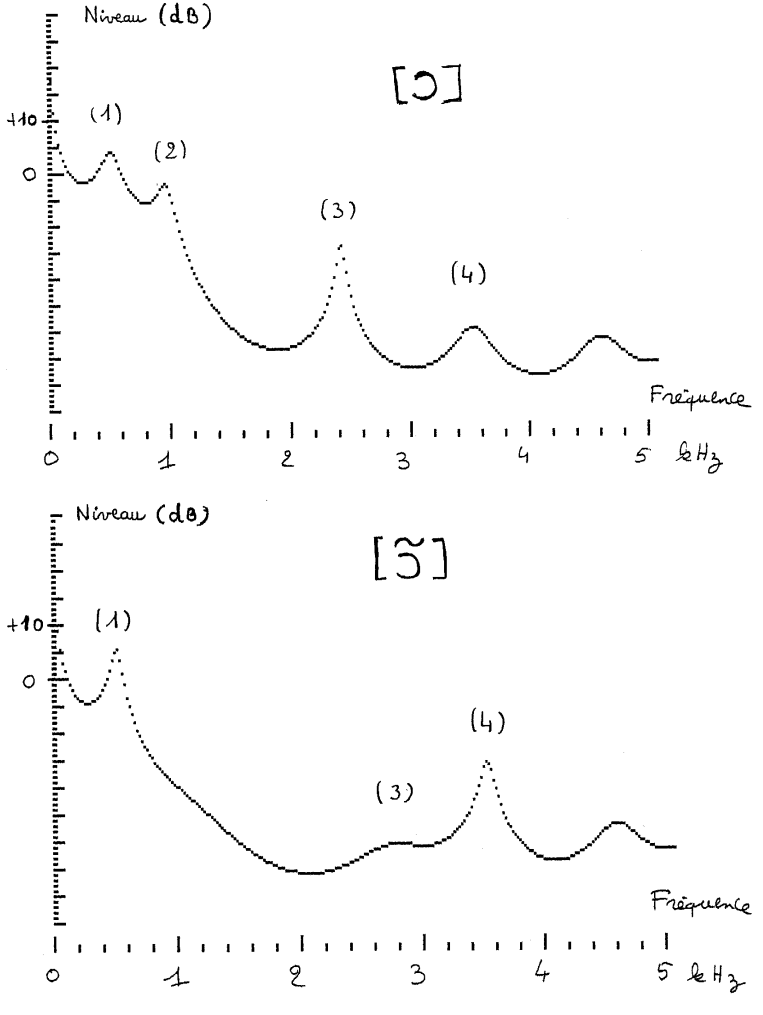




petite.

### 3.3. **Phonème 'o' ouvert** [ɔ] **et 'on'** [ɔ̃]

La voyelle nasale a le deuxième et le troisième formant plus atténués que la voyelle orale. (Le deuxième formant est même complètement estompé).

Ici, on a simplement augmenté la bande passante des filtres correspondants, ce qui donne le même résultat que l'effet d'antirésonance de la cavité nasale.



## CONCLUSION

De nombreux circuits de synthèse vocale existent sur le marché. Le PCF 8200 fait partie des circuits à formants qui génèrent des sons suivant un principe qui se rapproche de celui du conduit vocal humain. Leur réalisation en circuit intégré n'a été possible que par le développement des filtres numériques. On imagine la grandeur et le coût d'un système équivalent constitué de filtres analogiques : Il faudrait autant de filtres résonants que de formants permettant de créer les différents phonèmes. Cette synthèse vocale est un belle application du filtrage numérique.

## BIBLIOGRAPHIE

Encyclopedia Universalis : phonation ; phonétique.

Sons et Musique : Bibliothèque pour la science, Belin.

R. MIQUEL : Le filtrage numérique par microprocesseurs, Editest.

Documents du constructeur de MEA 8000 et du PCF 8200 (RTC).