

وزارة التعليم والبحث العلمي

MINISTÈRE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE

ÉCOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT

ELECTRONIQUE

PROJET DE FIN D'ÉTUDES

المدرسة الوطنية
المكنية
ÉCOLE NATIONALE POLYTECHNIQUE
BIBLIOTHÈQUE

SUJET

MODÈLE ARTICULATOIRE POUR
LA PRODUCTION DE LA PAROLE
(SIMULATION D'UN CONDUIT VOCAL)

Proposé par :

B. BOUSSEKSOU

Étudié par :

TELLACHE Med.

& NAMANE Abd.

Dirigé par :

B. BOUSSEKSOU

PROMOTION :
JUN 1985

ELECTRONIQUE

MODÈLE ARTICULATOIRE POUR
LA PRODUCTION DE LA PAROLE
(SIMULATION D'UN CONDUIT VOCAL)

B.BOUSSEKSOU TELLACHE Med. & NAMANE Abd. B.BOUSSEKSOU

JUIN 1985

*** REMERCIEMENT ***

Le travail présenté dans ce mémoire de these a été effectué à l'école POLYTECHNIQUE d'EL-HARRACH.

Nous tenons à exprimer nos plus vifs remerciements à notre promoteur monsieur B. BOUSSEKSOU , pour son aide et ses conseils tout au long de l'élaboration de ce projet de fin d'étude.

Que tous les professeurs qui ont contribué à notre formation trouvent ici l'expression de notre profonde gratitude.

Et que tous les gens qui nous ont aidé de près ou de loin , trouvent ici, l'expression de notre sincère remerciement.

*** TABLE DES MATIERES ***

INTRODUCTION

CHAPITRE I ETUDE DE LA PAROLE

- 1.1- Introduction
- 1.2- L'appareil vocal
 - 1.2.1- Définition
 - 1.2.2- Description
 - 1.2.3- Types de son et fonctionnement
- 1.3- La parole phénomène physique
 - 1.3.1- Caractéristique spectrale
 - 1.3.2- Propriétés des sons
 - 1.3.3- Caractéristique de la parole

CHAPITRE II ANALYSE et TRAITEMENT du SIGNAL PAROLE

- 2.1- Introduction
- 2.2- Analyse spectrale
- 2.3- Traitement numérique
- 2.4- Fenêtrage
- 2.5- Filtrage numérique

CHAPITRE III LE MODELE HARMONIQUE DU CONDUIT VOCAL

- 3.1- Introduction
- 3.2- Etude générale de la propagation de l'onde dans le conduit vocal
 - 3.2.1- Introduction
 - 3.2.2- Equations générales
 - 3.2.3- Modèles- Etudes théoriques
 - 1- Cas du tube homogène sans pertes
 - 2- Les pertes dans le conduit et leurs effets
 - a- Les vibrations des parois et leurs effets

- b- Les effets de viscosité et de chaleur
 - c- Effet du rayonnement aux lèvres
 - d- Conclusions
- 3.2.4 - Excitation du son dans le conduit vocal
 - 3.2.5- Descretisation du conduit vocal
 - 3.2.6- Le modèle harmonique
 - 1- Introduction
 - 2- Calcul des coefficients de reflexion
 - 3- Réponse du système
 - 3.2.7- Etude d'un modèle discret de tubes sans pertes
 - 3.2.8- MOdèle numérique du signal parole
 - 1- COnduit vocal
 - 2- Radiation
 - 3- Excitation
 - 3.2.9- Modèle complet
 - 3.2.10- Conclusion

CHAPITRE IV APPLICATION - SYNTHETISEUR A FORMANTS

- 4.1- Introduction
- 4.2- Types de structure de synthetiseur à Formants
- 4.3- Résonateurs
- 4.4- Blocs de synthetiseurs
- 4.5. Modèle du conduit vocal
- 4.6- Conclusion

CONCLUSION GENERALE

- ANNEXE "A"
- ANNEXE "B"
- ANNEXE "C"
- ANNEXE "D"

INTRODUCTION

De puis fort longtemps le phénomène de production de la parole suscite la curiosité des hommes.

Des la fin du XVIIIe siècle "WOLEGANG VON KEMPELEN "(1771) construisait une "machine parlante" imitation mécanique de l'ensemble vocal humain qui pourrait produire une vingtaine de sens différents. Cette première réalisation a été suivie de nombreux travaux. Mais il faut attendre l'avènement des ordinateurs pour que les recherches dans cette voie s'intensifient.

Aujourd'hui des préoccupations d'un intérêt immédiat stimulent ces efforts.

Comment transmettre la parole avec le minimum d'information ? Comment faire parler les ordinateurs ? Comment reconnaître automatiquement la parole ? Comment identifier un locuteur ?

La connaissance approfondie des mécanismes de production et de perception de la parole chez l'homme devrait nous aider à répondre à ces questions et permettre ainsi à l'homme d'établir un véritable dialogue avec les machines.

CHIBA, KAJIYAMA (1941), DUNN (1950), FANT (1960), HOUSE et STEVENS (1961) et FLANAGAN (1972) ont traité l'appareil phonatoire comme un système acoustique linéaire excité par une source ayant des caractéristiques déterminées.

Ce système acoustique est défini entre autres par la fonction d'aire qui spécifie l'aire transversale en chacun des points du conduit depuis la glotte jusqu'aux lèvres.

L'excitation acoustique se fait soit par une source quasi-périodique au niveau de la glotte, soit par un bruit acoustique qui peut être produit au niveau d'une constriction dans le conduit vocal par

L'effet de turbulence du flux d'air.

L'instrument fondamental dans toutes ces études est le synthétiseur de parole, qui permet d'étudier le rôle et l'importance des diverses composantes du signal détecté par l'analyse acoustique.

Le travail que nous présenterons ici consiste à faire une étude d'un modèle articulatoire pour la production de la parole: Simulateur du conduit vocal.

Dans le chapitre I, nous précisons quelques points essentiels de la théorie de fonctionnement de l'appareil vocal.

Le chapitre II, présente les notions théoriques et les mathématiques essentiellement appliquées pour le traitement numérique du signal.

Le chapitre III, est consacré à l'étude de la propagation des ondes acoustiques dans le conduit vocal, au terme de laquelle nous définissons le modèle harmonique du système phonatoire.

Une étude détaillée du modèle harmonique nous a permis de mettre en évidence l'utilisation des coefficients de réflexion .

Dans le chapitre IV, nous avons décrit le fonctionnement d'un synthétiseur à Formant, utilisant deux configurations "parallèle/serie".

CHAPITRE I : ETUDE DE LA PAROLE

1.1 - Introduction:

Nous allons préciser dans ce chapitre quelques points essentiels de la théorie de fonctionnement de l'appareil vocal, nous faisons un bref exposé sur les organes de phonations et le mécanisme de fonctionnement ainsi que les caractéristiques spectrales de la parole.

1.2 - L'Appareil vocal:

1.2.1 - Définition:

Du point de vue acoustique la parole représente un son très complexe qui serait la conséquence d'une excitation des cavités de l'appareil vocal humain.

1.2.2 - Description:

Le système vocal se compose essentiellement des poumons, de la trachée artère et du larynx, celui-ci supportant deux muscles appelés " cordes vocales ". Enfin le conduit vocal qui est formé par le pharynx les cavités buccale et nasale " fig 1.1 ".

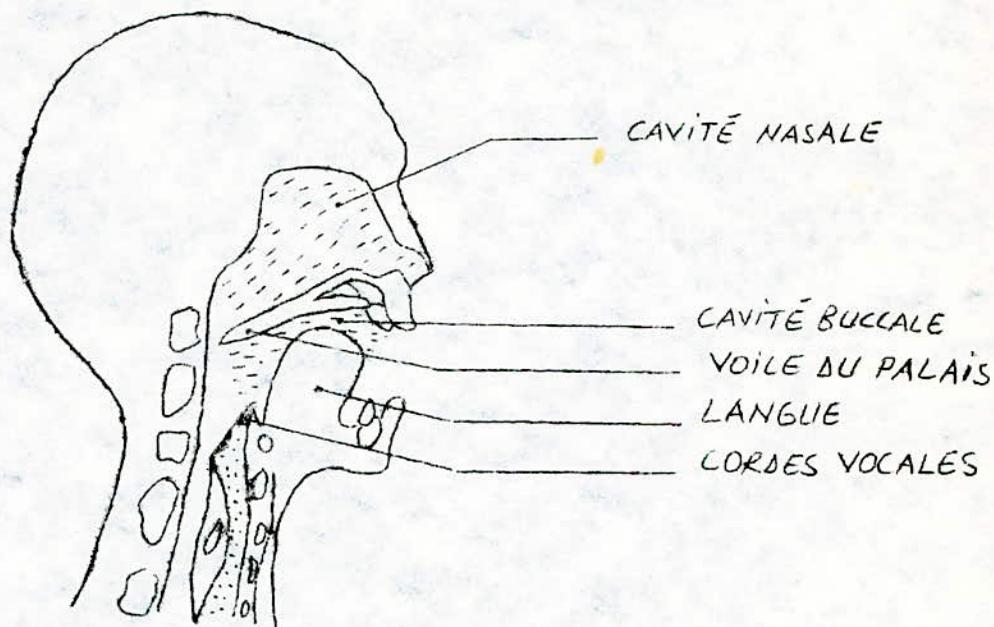
" fig 1.2 " on schématise ici les trois cavités couplées qui constituent l'appareil vocal humain.

La cavité nasale indéformable est couplée au pharynx par le voile du palais. La position de la langue règle le volume et le couplage de deux autres cavités.

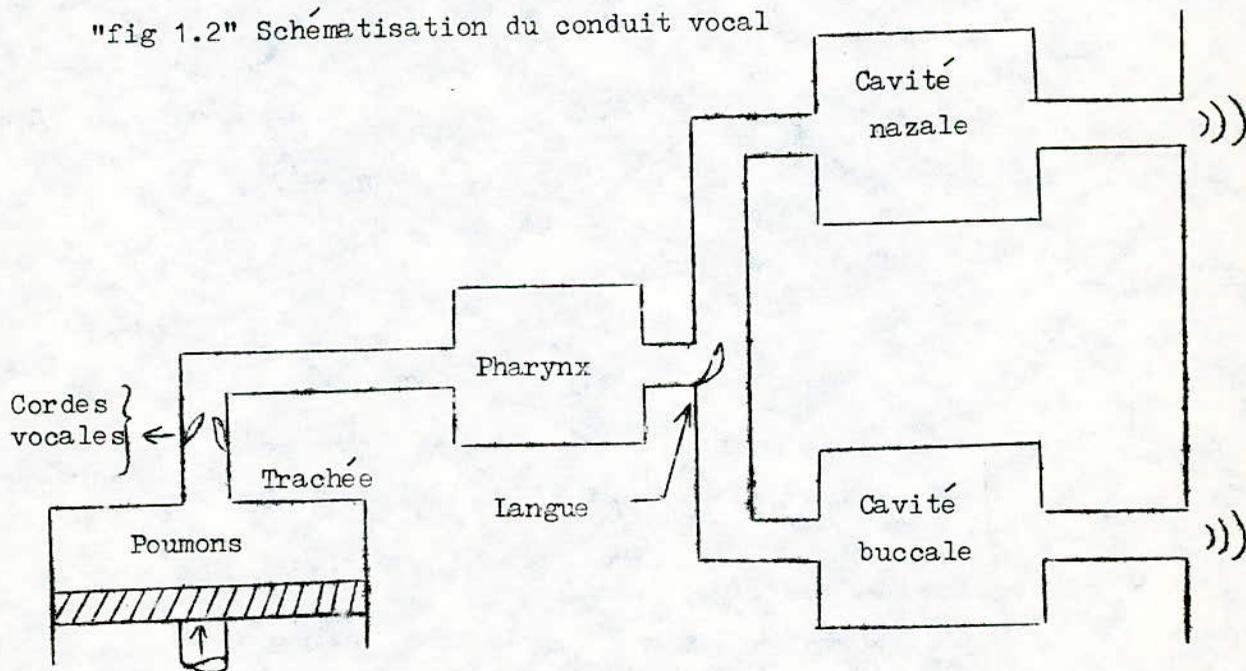
1.2.3 - Types de sons et fonctionnement:

La parole en tant que phénomène physique résulte de l'excitation du conduit vocal par deux types de sons.

"fig 1.1" Coupe du conduit vocal



"fig 1.2" Schématisation du conduit vocal



La production des sons dit "voisés" comme les voyelles fait intervenir une source d'impulsions périodiques constituée par l'ensemble poumons-cordes vocales. Les sons "non voisés" comme certaines consonnes "s" "ch" et "f" sont engendrées par une source de bruit, le bruit est produit par l'action du courant d'air issu des poumons sur les parois de l'appareil vocal, souvent au niveau d'un resserrement du conduit.

La production de certains sons nécessite l'intervention simultanée de deux phénomènes, il en est ainsi par exemple de "Z", "J" et "V", qui correspondent respectivement aux consonnes "S", "CH" et "F", le conduit vocal conservant la même forme lors de l'émission de chacun des deux groupes. La formation des consonnes nazales comme "M" ou "N" utilise de surcroît la cavité nazale limitée d'un côté par les narines et de l'autre par le voile du palais, qui règle son couplage acoustique avec le conduit vocal "Fig 1.2"

Lors de la production de "M" ou "N" le conduit vocal est fermé l'air s'écoule par la cavité nazale.

Les voyelles nazales du type "ON" et "IN" sont obtenues lorsque le conduit nasal est couplé au conduit vocal celui-ci étant ouvert au niveau des lèvres.

1.3. - La parole phénomène physique:

1.3.1. - Caractéristique spectrale de la parole;

La connaissance du "spectre de la parole", c'est-à-dire des différentes fréquences acoustiques qu'elle contient est indispensable pour réaliser des synthétiseurs. Les deux sources sonores que nous avons décrites, possèdent des caractéristiques spectrales différentes. La source périodique ou source sonore délivre des oscillations de "relaxation" qui prennent naissance au niveau des cordes vocales, ces deux muscles, plus ou moins tendus côte à côte dans la phase de phonation, se séparent et se rapprochent alternativement sous l'action de la pression d'air assuré par les poumons dans la trachée.

La fréquence de ce phénomène de relaxation est essentiellement déterminée par la tension et la masse des cordes vocales ainsi par la pression subglottique. Le débit d'air au niveau des cordes vocales est évidemment modulé à la même fréquence, celle-ci dite "fondamentale" appelé aussi "pitch" caractérise la hauteur du son de la voix.

La fréquence fondamentale est en moyenne de 120 Hz pour un homme, et de 220 Hz pour une femme.

Comme tout phénomène périodique le signal émis par la source vocal peut être décomposé en série de Fourier c'est-à-dire en une somme d'oscillations sinusoidales dont les fréquences dite "harmonique" sont toutes multiples de la fréquence fondamentale.

Le spectre se compose donc d'une série de raies équidistantes dont l'amplitude dans le cas présent diminue au fur et à mesure que le rang croît "fig. 1.3".

Ce spectre continu contient toutes les fréquences comprises dans un interval de plusieurs milliers de hertz.

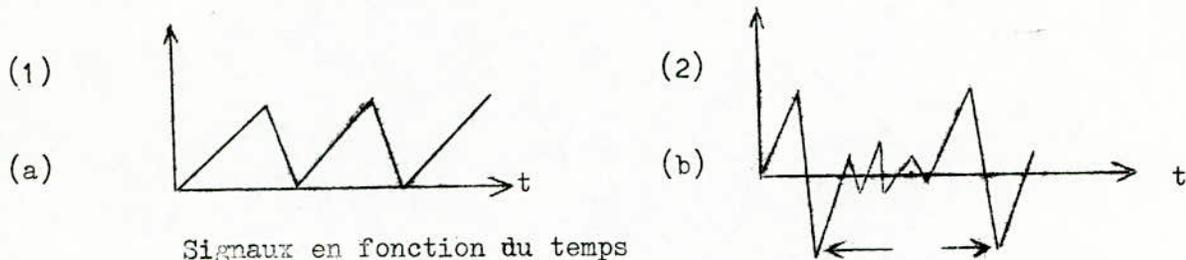
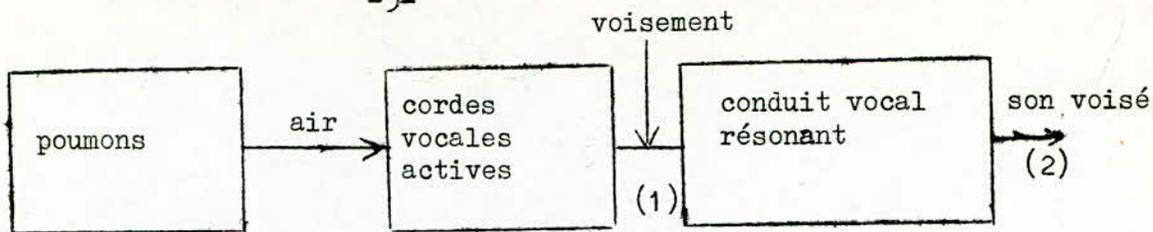
Le signal issu de la source vocale est modifié lors de la traversée du conduit vocal celui-ci est constitué pour l'essentiel de deux cavités couplées, la cavité du pharynx et la cavité buccale; leurs volumes respectifs ainsi que leurs degrés de couplage sont déterminés par la position de la langue. Ces cavités jouent le rôle de caisses de résonances et vont donc renforcer dans le spectre sonore. Les composantes voisines de leurs fréquences propres "fig. 1.3". Les maximums de la "courbe de réponse" en fréquence du conduit vocal sont appelés "formants".

La fréquence du premier formant peut varier de 200 à 900 Hz environ, celle du deuxième de 500 à 2500 Hz. Il existe des formants d'ordre supérieur, mais leur rôle sur le plan perceptif est limité, ils contribuent à caractériser la qualité de la voix.

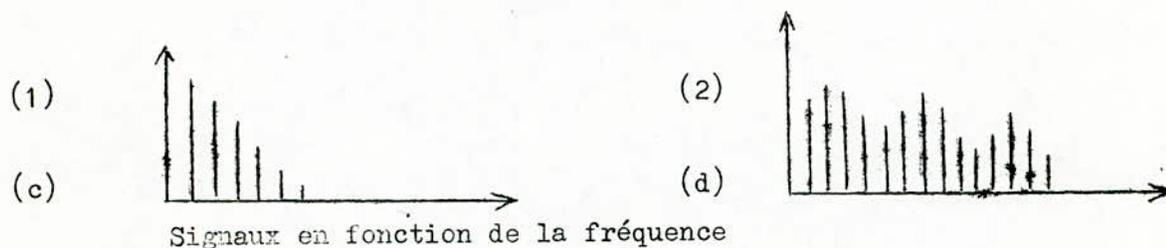
Lors de la production d'une voyelle, le conduit vocal se déforme très peu, et ses fréquences de résonances sont stables et dépendent de la voyelle émise. En revanche la forme du conduit vocal évolue rapidement dans le temps lors de la production de certaines consonnes, ce qui entraîne un changement des fréquences de formants.

Les caractéristiques des transitions sont alors représentées pour une consonne prononcée "fig. 1.4".

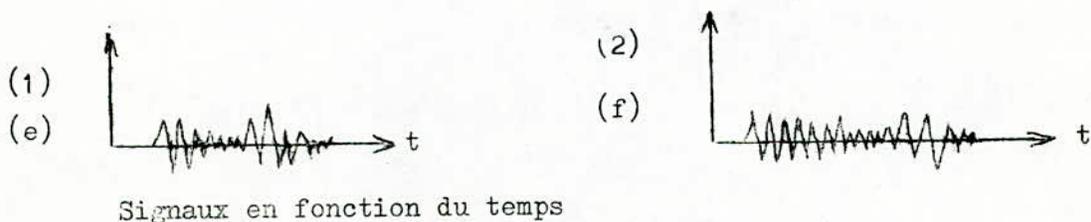
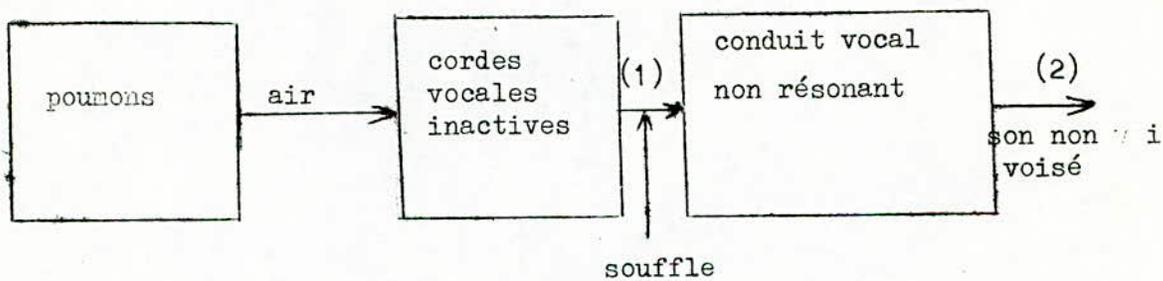
Dans le cas de la production des sons de type "S", "CH", "F" à partir d'une sonore de bruit, les formants se situent vers les hautes fréquences, les consonnes nasales sont au contraire caractérisées par un affaiblissement des formants à fréquences élevées.



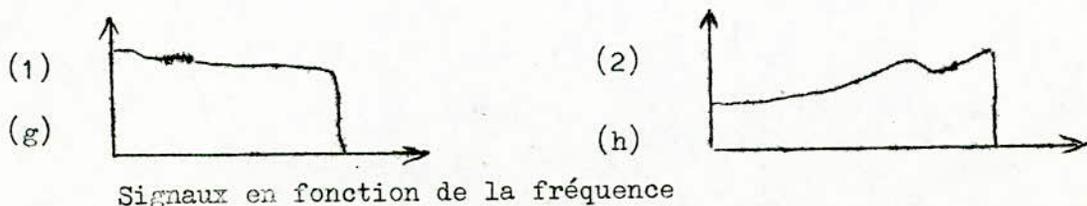
Signaux en fonction du temps



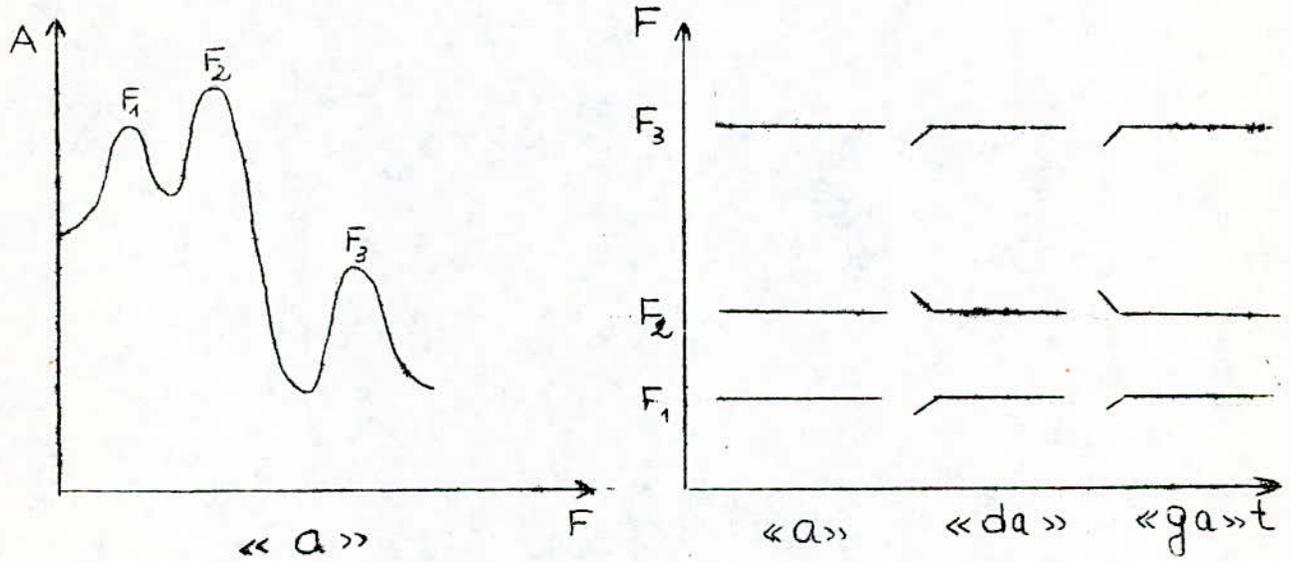
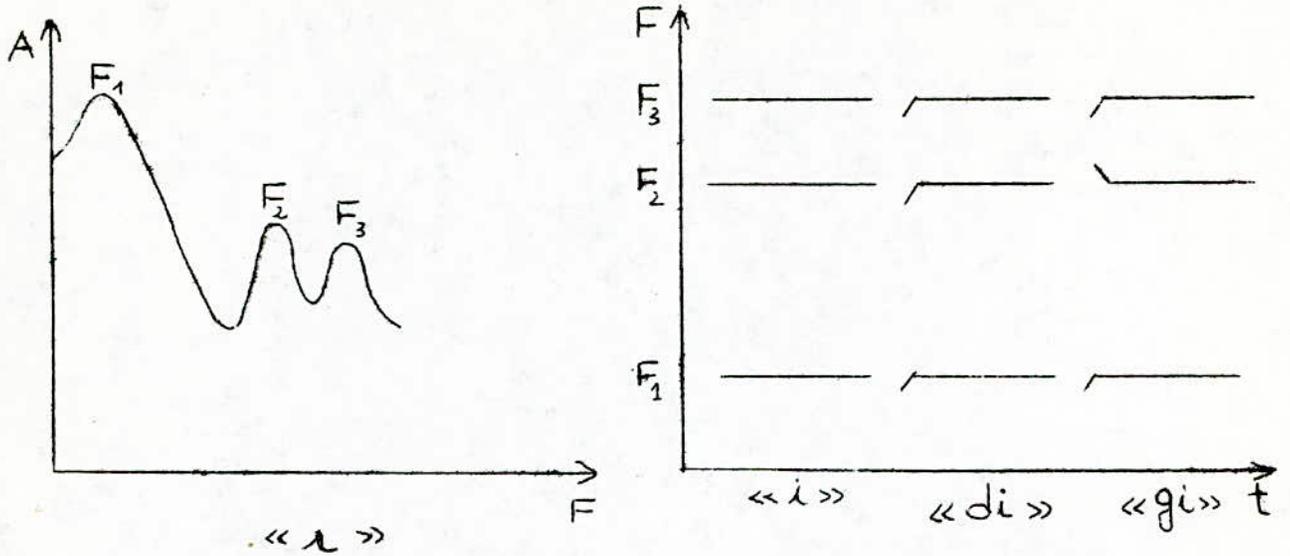
Signaux en fonction de la fréquence



Signaux en fonction du temps



Signaux en fonction de la fréquence



« fig 1.4 »

Les voyelles sont produites pour différentes positions du conduit vocal correspondant à différentes valeurs des formants".

Les consonnes sont caractérisées par les évolutions des fréquences des formants, dans le temps comme le montre la figure pour les sons.

- "di" "gi"
- "da" "ga"

1.3.2 - Propriétés des sons:

1) Son:

Les messages naturels ou provoqués perçus par l'intermédiaire de l'ouïe portant le nom général de son.

L'écriture imprimée ou manuscrite peut être décomposée en lettres, mots et phrases il en est même pour la parole.

Elle peut être caractérisée en.

2) Phonèmes:

La phonétique considère depuis longtemps l'existence de sons élémentaires appelés " phonèmes ". En français on dénombre 37 phonèmes dont 16 voyelles, 18 consonnes et 3 semi-voyelles.

3) Phonatomes:

L'association de plusieurs phonèmes ne donne pas nécessairement un mot compréhensible dans le langage parlé ainsi un " ch"...", " ... a" n'est pas un "chat". La juxtaposition de deux phonèmes conduit seulement à la perception de deux sons séparés par un temps mort; C'est la transition d'un phonème vers un second qui rend le mot intelligible.

Le diphonème ou phonatome, désigne l'ensemble formé par l'association de deux phonèmes, c'est ce que l'on appelle " atome phonatoire ". Pour reconstituer un mot ou une phrase, il suffit d'assembler les diphonèmes nécessaires par leur extrémité commune comme on assemble les dominos. Une analyse statistique réalisée au lins à Orsay a montré qu'à partir d'un répertoire de 627 diphonèmes, il était possible de reconstituer n'importe quelle phrase en français.

1.3.3- Caractéristique de la parole:

1) Hauteur:

La fréquence la plus basse dans le signal parole est la fréquence fondamentale de vibrations des cordes vocales, elle varie entre 70 et 150 Hz pour les hommes, et de 150 à 300 Hz pour les femmes, et peut dépasser 400 Hz pour les enfants.

2) Timbre;

C'est la caractéristique de deux sons de même fréquence. Il peut être considéré comme un paramètre du son, il se caractérise par l'amplitude harmonique, chaque harmonique à son amplitude propre, donc pour retrouver l'intégrité du timbre, il faudra que chaque harmonique soit respectée.

3) Bande passante:

Quand les fréquences entre 100 Hz et 10 KHz sont correctement rendus, on peut admettre qu'une transmission de parole est fidèle.

4) Dynamique:

C'est la variation de l'énergie phonétique d'un individu et qui s'exprime en décibel c'est-à-dire que la dynamique d'un signal peut être définie comme étant l'écart en " dB " entre le niveau maximum et le niveau minimum de ce signal.

$$Ds = (N_{max} - N_{min}) (db)$$

Globalement l'énergie de la parole varie en plus entre 1/10 et 1000, soit une dynamique de 40 dB.

CHAPITRE II

ANALYSE ET TRAITEMENT DU SIGNAL DE LA PAROLE

2.1- Introduction:

L'étude de la parole en tant que phénomène physique nécessite l'utilisation des systèmes numériques pour avoir des résultats adéquats. Les systèmes numériques ont déjà remplacé beaucoup de machines analogiques et il semble que cette tendance aille en s'accroissant, les raisons de cette évolution sont nombreuses: Les systèmes numériques sont très rapides et peuvent remplir des fonctions très diverses sans changer ses composants, il suffit de modifier le programme.

Aujourd'hui des dizaines de méthodes analogiques et numériques d'analyse du signal de parole ont été décrites.

Les analyseurs utilisent tous les notions d'analyse spectrale du signal rayonné par la bouche pour le traitement fréquentiel, et d'analyse par auto-correlation pour le traitement temporel.

2.2- Analyse spectrale:

2.2.1 Introduction:

Parmi les méthodes d'analyse spectrales du type numérique on trouve la transformée de Fourier.

Celle-ci consiste à décomposer un signal $h(t)$ sous la forme d'une somme infinie de sinusoides de fréquence f et d'amplitude $H(f)$, d'où une interprétation intéressante pour l'analyse des systèmes linéaires invariants.

L'objectif est de :

- Recherche une approximation de la T.F de façon à permettre une implantation aisée sur ordinateur.

- Utiliser un algorithme réduisant considérablement le temps de calcul d'où la T.F.R

2.2.2 Serie de fourier :

soit une fonction $x(t)$ périodique de période T alors $x(t) = x(t+T)$

Le point de départ de toute analyse de signal est le théorème de fourier : "Toute fonction périodique peut être décomposée en une somme de fonctions sinusoidales".

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos 2\pi n f t + b_n \sin 2\pi n f t)$$

avec

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos 2\pi n f t dt$$

et
$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin 2\pi n f t dt$$

on défini le spectre de fréquence comme suit :

soit

$$X(nf) = \frac{1}{2} (a_n - j b_n) = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j 2\pi n f t} dt$$

ce dernier peut être décomposée en :

1 - Spectre d'amplitude. :

$$|X(nf)| = \frac{1}{2} \sqrt{a_n^2 + b_n^2}$$

2 - Spectre de phase :

$$\varphi(nf) = \text{Arctg}(-b_n/a_n)$$

Il est important de remarquer que le spectre d'une fonction périodique se décompose de raies, qui correspondent à des fréquences multiples les unes des autres (harmoniques) et dont le plus petit multiple (fondamental) est responsable de la sensation de hauteur, lors de l'audition du signal

Les signaux à traiter sont très rarement périodiques, lorsqu'on a affaire à des signaux quelconques, on est amené à généraliser les notions précédentes et à remplacer tout ce

qui est "somme" par des termes d'intégration. Aussi parle t'on "d'intégrale de Fourier".

2.2.2 : Transformée de fourier (Intégrale de fourier) :

La T.F $X(f)$ d'un signal continu $x(t)$ est donnée par :

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt$$

La transformée inverses est donnée par :

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(f) e^{+j2\pi ft} df$$

2.2.3 Produit de convolution;

Soient deux fonctions de temps $f_1(t)$ et $f_2(t)$ possédant les transformées de Fourier $F_1(\omega)$ et $F_2(\omega)$. La transformée $F(\omega)$ du produit $f_1(t) \cdot f_2(t)$ est égale à :

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_1(u) \cdot F_2(\omega - u) du$$

Cette relation est nommée "produit de convolution". Réciproquement dans le domaine temporel, la fonction $f(t)$ correspondant au produit $F_1(\omega) \cdot F_2(\omega)$ est donnée par :

$$f(t) = \int_{-\infty}^{+\infty} f_1(\lambda) f_2(t - \lambda) d\lambda$$

On parle aussi d'autocorrelation du signal, pour comparer le signal à un instant donné à ce même signal considéré à un autre instant.

L'autocorrelation s'écrit :

$$R(\tau) = \lim_{T \rightarrow \infty} \left[\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) f(t - \tau) dt \right]$$

Si maintenant $|F_T(\omega)|$ désigne le module de la transformée de fourier du signal $f(t)$ dans un interval $[-\frac{T}{2}, \frac{T}{2}]$, on définit enfin" le spectre de densité de puissance moyenne" ou "spectre de puissance" $S(\omega)$:

$$S(\omega) = \lim_{T \rightarrow \infty} \left[\frac{|F_T(\omega)|^2}{T} \right]$$

2.3 Traitement numérique :

2.3.1 Echantillonnage :

Pour appliquer les méthodes de traitement du signal numérique sur un signal analogique comme celui de la parole il est nécessaire de représenter le signal sous forme de séquence de nombres. Ceci est généralement fait par échantillonnage du signal analogique.

Définition :

C'est l'observation d'un signal analogique non pas d'une manière continue mais à certains instants seulement.

L'échantillonnage est réalisé en multipliant le signal analogique $x(t)$ par une suite périodique d'impulsions de dirac de période Δt .

$$x_e(t) = x_a(t) \cdot e(t)$$

avec

$$e(t) = \sum_{k=-\infty}^{+\infty} \delta(t - k\Delta t) = \delta_{\Delta t}(t)$$

$$x_e(t) = x_a(t) \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t) = x_a(t) \cdot \delta_{\Delta t}(t)$$

où

$$\delta(t - k\Delta t) = \begin{cases} 1 & \text{si } t = k\Delta t \\ 0 & \text{si } t \neq k\Delta t \end{cases} \quad k \in \mathbb{N}$$

Dans le domaine fréquentiel, la transformée de fourier $X_e(f)$ d'un signal échantillonné

est donné par le produit de convolution :

$$X_e(f) = X_a(f) * E(f) = \frac{1}{\Delta t} \sum_{n=-\infty}^{+\infty} X_a(f - n/\Delta t)$$

L'échantillonnage est un processus exact (et non une approximation) si deux conditions sont respectées :

- 1 - La fréquence du signal d'entrée est plus petite que la moitié de la fréquence d'échantillonnage.
- 2 - La fréquence du signal de sortie est plus petite que la moitié de la fréquence d'échantillonnage.

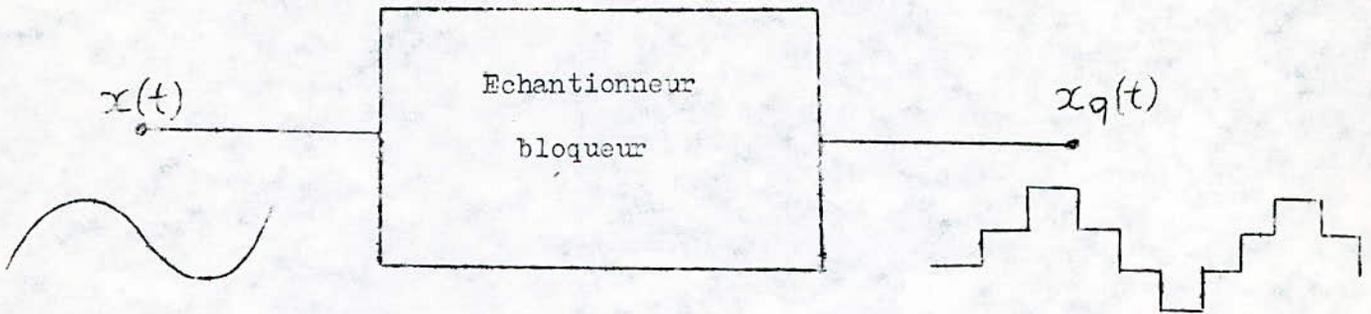
Pour réaliser ces deux conditions il faut encadrer le système par des filtres passe bas.

Remarque :

Il est commode de considérer les procédés d'échantillonnage et de quantification séparément même s'il est impossible de les distinguer au moment de l'exécution de représentations numériques du signal de la parole.

2.3.2 Quantification :

La quantification est l'approximation de chaque valeur du signal $x(t)$ par un multiple entier d'une quantité élémentaire q appelée échelon de quantification. Si q est constant quelque soit l'amplitude du signal, la quantification est dite uniforme. Cette opération revient à faire passer le signal dans un organe qui possède une caractéristique en marche d'escalier, comme le montre " la figure " pour $q = 1$ et fournit le signal $x_q(t)$.



2.3.3 : Transformée de fourier discrète (TFD) :

Définition :

La TFD est une représentation fréquentielle de suites temporelles périodiques, elle possède trois caractères fondamentaux :

- 1 - Le signal est une suite périodique finie
- 2 - Son spectre est également une suite périodique finie
- 3 - Les périodes du signal et les périodes du spectre contiennent le même nombre d'échantillons.

On appelle transformée de fourier discrète de N valeurs discrètes :

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{-j 2\pi \frac{kn}{N}} \quad \text{pour } k=0,1,\dots,N-1$$

La transformée inverse existe, elle a pour expression :

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(n) e^{+j 2\pi \frac{kn}{N}} \quad \text{pour } n=0,1,\dots,N-1$$

Cependant le calcul d'une telle transformée est long et fastidieux et l'obtention du résultat en "temps réel" se révélait impossible jusqu'à la découverte d'un algorithme mathématique qui permet d'effectuer la célèbre FFT ou "Fast Fourier Transform"

2.3.4 Transformée de Fourier rapide (TFR) :

La TFR est un algorithme de calcul permettant d'obtenir la TFD après un nombre d'opérations arithmétiques considérablement inférieur à celui nécessaire dans le cas d'un calcul direct.

Les équations de définition de la TFD fournissent une relation entre deux ensembles de N nombres complexes qui s'écrit d'une manière commode sous une forme matricielle.

on pose

$$W = e^{-j\frac{2\pi}{N}}$$
$$X(n) = \sum_{k=0}^{N-1} x(k) W^{nk}$$

or on a

$$e^{-j2\pi} = 1 \quad ; \quad W^N = 1$$
$$e^{-j\pi} = -1 \quad ; \quad W^{N/2} = -1$$

La TFD⁻¹ s'écrit alors :

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(n) W^{-nk}$$

Du fait que $W^N = 1$ ce qui veut dire que les coefficients W^{nk} sont périodiques de période N.

En exploitant les propriétés de symétrie de W^{nk} on aboutira à une série d'algorithmes dont le premier a été élaboré par "COOLEY - TUKEY".

Ce dernier nécessite $\frac{N}{2} \log_2 \frac{N}{2}$ multiplications et $N \cdot \log_2(N)$ additions, alors que le calcul direct de la TFD pour N valeurs de $\{x(k)\}$ nécessite N^2 multiplications et $N(N-1)$ additions, ce qui correspond à une quantité énorme de calcul si N devient grand.

2.4- Fenêtrage :

2.4.1 Introduction :

Le rôle principal de la fenêtre est de limiter la durée d'un signal infini. La position de la fenêtre est choisie de manière à conserver les échantillons importants du signal à traiter.

Quant au choix de la forme, il dépend principalement de la largeur B_1 du pic central $W(f)$ et d'amplitude des lobes secondaires.

Nous allons étudier la fenêtre de Hamming, celle-ci par ses caractéristiques a été choisie dans l'analyse de notre signal de parole.

2.4.2- Caractérisation des fenêtres spectrales :

Nous remarquons que cette fonction possède un pic central et des lobes secondaires. Deux paramètres principaux permettent de les caractériser. Le premier est la largeur de base du pic central et le second est le rapport de l'amplitude (du premier lobe secondaire et du pic central). Ce dernier est exprimé en decibels de la manière suivante :

$$A_i = 20 \log_{10} \left| \frac{W_i(f_s)}{W_i(0)} \right|$$

ou f_s est la fréquence au milieu du lobe secondaire de la fonction $W_i(f)$.

2.4.3- Fenêtre de Hamming :

L'atténuation des lobes secondaires sera plus importante en choisissant une fenêtre :

$$W_h(k) = -\frac{1}{4} \exp \left[j \frac{2\pi k}{N} + 2 + e^{j \frac{2\pi k}{N}} \right]$$
$$= \frac{1}{2} \left(1 + \cos \frac{2\pi k}{N} \right) \quad \text{pour} \quad |k| \leq \frac{N}{2}$$

Cette fenêtre s'appelle "fenêtre de Hamming". On peut généraliser cette fonction de la manière suivante :

$$W_H(k) = \begin{cases} \alpha + (1-\alpha) \cos \left(\frac{2\pi k}{N} \right) & |k| \leq \frac{N}{2} \\ 0 & \text{partout ailleurs} \end{cases}$$

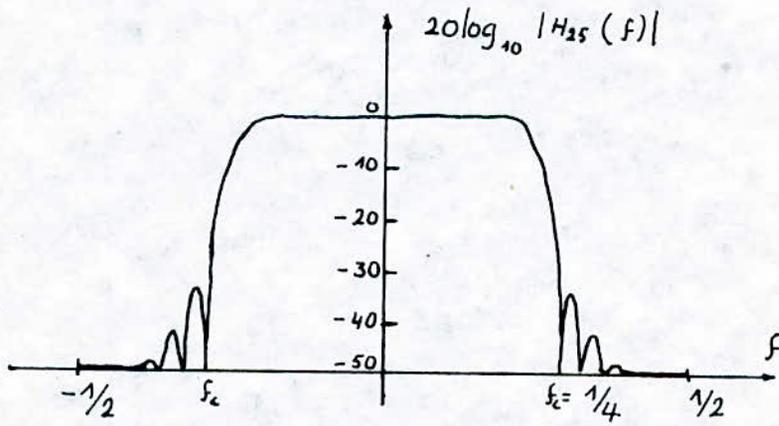
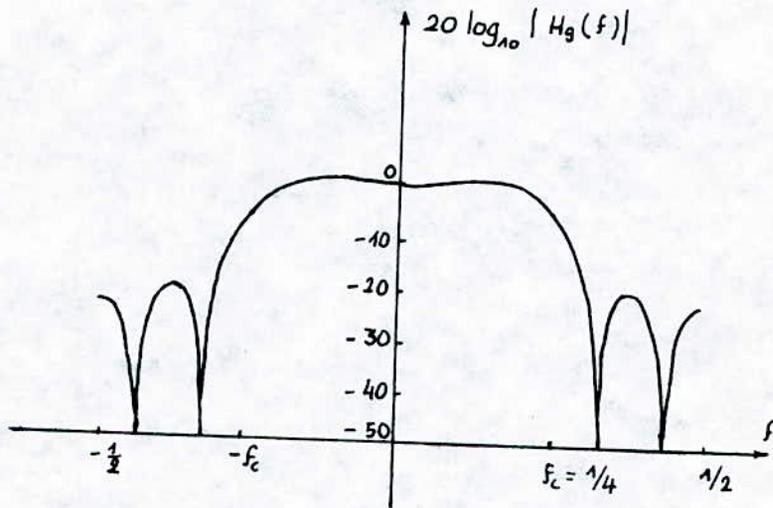
Pour $\alpha = \frac{1}{2}$, on obtient la fenêtre de Hamming. La forme générale ci-dessus dépendant du paramètre α est appelé fonction de fenêtre de Hamming généralisée. Si $\alpha = 0,54$; on obtient la fenêtre de Hamming.

2.4.4 Conclusion :

L'étude comparative des différentes fenêtres d'analyse montre clairement sur quoi se porte notre choix.

Les éléments déterminants sont :

- La position.
- La forme.
- La taille (nombre d'échantillons).



Nous remarquons que la position est choisie de manière à conserver les échantillons important du signal à traiter.

Quant à la forme, son choix dépend de la largeur du pic central et de l'amplitude des lobes secondaires.

En ce qui concerne la taille des échantillons, il suffit de réduire les oscillations en limitant N (nombre d'échantillons de la fenêtre).

Aussi, c'est la fenêtre de Hamming que nous utiliserons à cause de ses propriétés plus favorables.

2.5- Filtrage numérique :

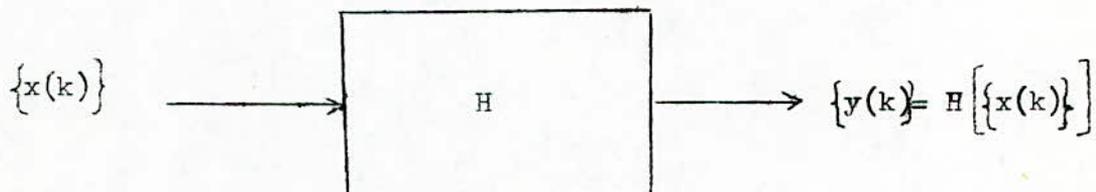
2.5.1 Introduction:

Le rôle d'un filtre analogique est de sélectionner parmi toutes les composantes spectrales du signal celles qui remplissent les conditions désirées par contre le rôle d'un filtre numérique consiste à associer à des échantillons d'entrée $\{x(k)\}$ séquence $\{y(k)\}$ en sortie.

C'est à dire que le filtre numérique n'est d'autre qu'une unité de calcul disposant des opérateurs "Additionneur, multiplicateur et retard "

2.5.2- Définition:

C'est un algorithme de calcul qui associe à une séquence de nombre $\{x(k)\}$ dite d'entrée une séquence de nombre $\{y(k)\}$ dite de sortie.



a- Filtre récursif:

C'est un filtre dont chaque échantillon de sortie dépend des échantillons présents et passés d'entrées ainsi que des échantillons de sortie.

Ce type est caractérisé par l'équation de récurrence suivante.

$$y(k) = \sum_{l=0}^M a_l x(k-l) - \sum_{l=1}^N b_l y(k-l)$$

b- Filtre non récursif:

C'est un filtre dont chaque échantillon dépend des échantillons présents et passés d'entrée .

Il est caractérisé par l'équation suivante:

$$y(k) = \sum_{l=0}^M a_l x(k-l)$$

2.5.3- Fonction de transfert:

Soient $x(k)$ et $y(k)$ les séquences d'entrée et de sortie d'un filtre, soient $X(z)$ et $Y(z)$ leurs transformées en z respectives.

On appelle fonction de transfert $H(z)$ d'un filtre numérique le rapport de ces deux transformées.

$$H(z) = \frac{Y(z)}{X(z)}$$

a- Filtre récursif:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{l=0}^M a_l z^{-l}}{1 + \sum_{l=1}^N b_l z^{-l}}$$

b- Filtre non récursif ;

Pour ce dernier, il suffit de mettre les $b_i = 0$

$$H(z) = \sum_{l=0}^M a_l z^{-l}$$

2.5.4- Structure d'un filtre numérique:

a- Filtre non récursif:

Leur structure découle directement de l'équation de récurrence :

$$y(k) = \sum_{l=0}^M a_l x(k-l)$$

Cette structure nécessite (M + 1) emplacements mémoires, la gestion de la table x (k - i) et (M + 1) multiplication "fig. 5.1"

b - Filtre récursif :

Leur structure découle directement de l'équation de récurrence.

$$y(k) = \sum_{l=0}^M a_l x(k-l) - \sum_{l=1}^N b_l y(k-l)$$

Structure réalisée sous forme directe "1".

On remarque que cette structure nécessite la gestion de deux tables x(k-i) et y(k-i) (M + N) emplacements mémoires et (M + N + 1) multiplications. "fig. 5.2)

Pour l'amélioration de cette dernière structure, on introduit une variable intermédiaire W (z) dans la fonction de transfert soit :

on pose $G(z) = \frac{Y(z)}{X(z)} = \frac{Y(z)}{W(z)} \cdot \frac{W(z)}{X(z)} = \frac{N(z)}{D(z)}$

$N(z) = \frac{Y(z)}{W(z)}$ avec $N(z) = \sum_{l=0}^M a_l z^{-l}$

d'où $y(k) = \sum_{l=0}^M a_l w(k-l)$ (1)

$\frac{W(z)}{X(z)} = \frac{1}{D(z)}$ avec $D(z) = 1 + \sum b_i z^{-i}$

$W(z) = \frac{X(z)}{D(z)}$ d'où $w(k) = x(k) - \sum_{l=1}^N b_l w(k-l)$

(1) et (2) nous donne la structure appelée forme directe "2" donnée par la "fig. 5.3".

Cette forme permet de réduire le nombre d'emplacements mémoires à N et la gestion d'une seule table W (k - i).

La forme parallèle :

Elle est obtenue en décomposant H(z) en fractions du premier ordre ou du deuxième ordre "fig.. 5.4".

$$H(z) = H_1(z) + H_2(z) + \dots + H_L(z) = \sum_{l=1}^L H_l(z)$$

- La forme cascade :

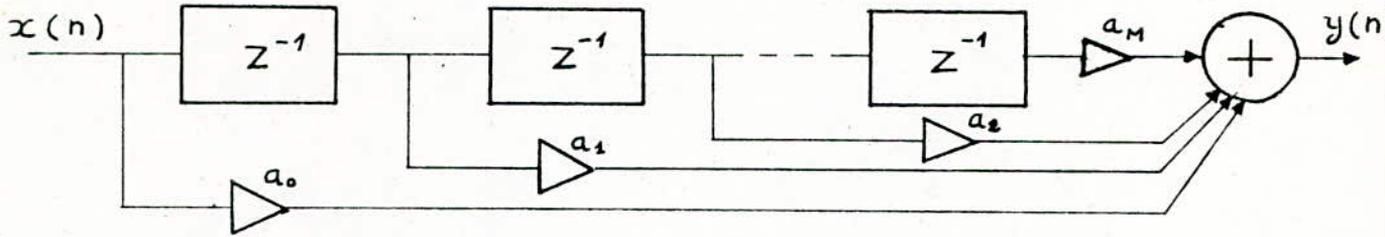
Elle est obtenue en décomposant la fonction de transfert en produit de fonctions simples du premier ou deuxième ordre "fig. 8.5".

$$H(z) = H_1(z) \cdot H_2(z) \cdot \dots \cdot H_L(z) = \prod_{l=1}^L H_l(z)$$

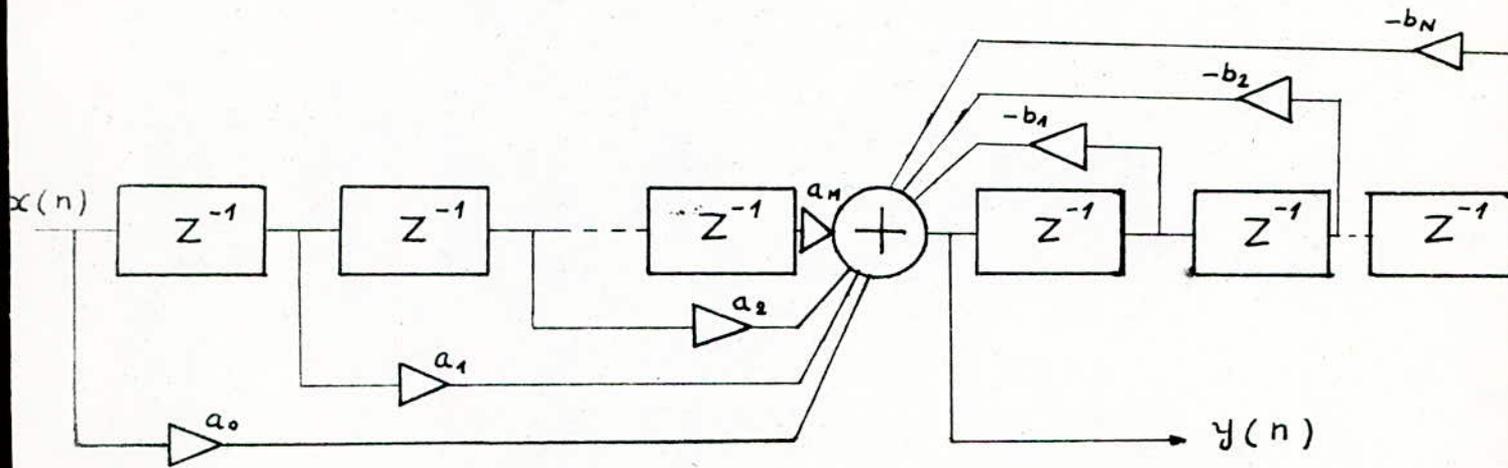
2. .5 Conclusion :

Les structures de réalisations récursives sont connues par leur sensibilité aux erreurs d'arrondi dans les opérations arithmétiques à précision limitée, car l'erreur sur un coefficient agit sur la position de tous les zéros et tous les pôles, cette sensibilité augmente avec l'ordre du système c'est à dire avec l'ordre de l'équation aux différences.

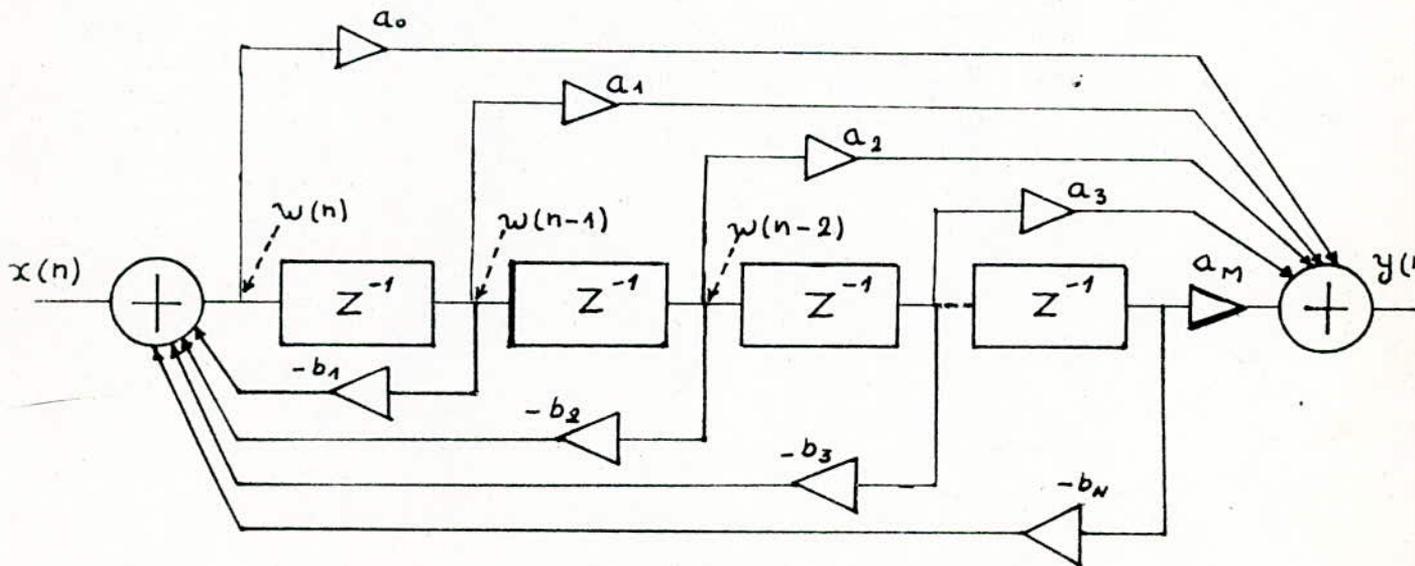
Par contre la réalisation sous forme cascade ou parallèle est moins sensible aux erreurs de quantification parce que la quantification d'un coefficient affectera au plus, la position de deux pôles. C'est pour cette raison que la réalisation des filtres numériques se fait généralement sous forme cascades ou parallèles.



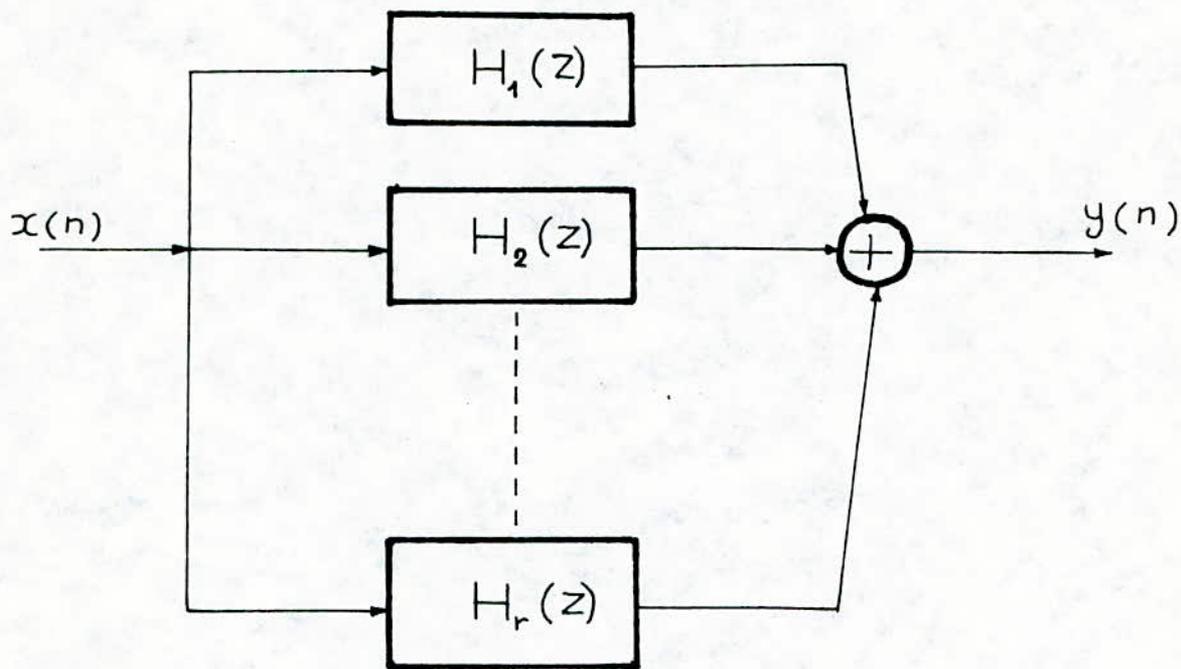
"fig 5.1"



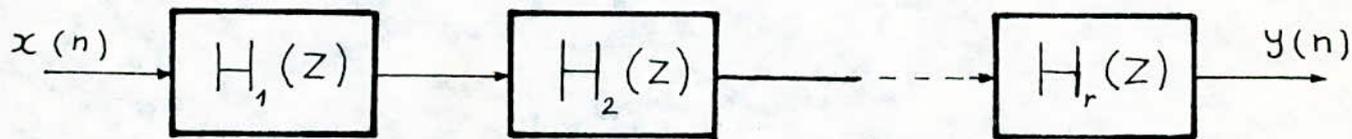
"fig 5.2"



"Fig 5.3"



"fig 5.4"



"fig 5.5"

CHAPITRE III :

LE MODELE HARMONIQUE DU CONDUIT VOCAL

3.1. - INTRODUCTION :

Les analogues du conduit vocal peuvent être classés en deux catégories selon que leur fonctionnement est statique ou dynamique.

Cette technique permet de reproduire par simulation numérique ou analogique, un signal représentatif de la pression sonore issue des lèvres à partir de la forme du conduit vocal.

Dans cette technique la phase d'analyse consiste à accumuler un grand nombre de données physiologiques et acoustiques sur la production des sons de la langue étudiée.

L'étude de la propagation acoustique dans le conduit vocal nous amène à établir un système d'équations différentielles de deux variables (x, t). La résolution de ce système n'étant pas simple, nous utiliserons une méthode de résolution faisant intervenir une discrétisation du conduit vocal. Nous étudierons alors un modèle harmonique grâce auquel nous évaluerons, à partir de la fonction d'aire discrétisée, tous les indices acoustiques.

Une étude détaillée du modèle de tubes sans pertes nous amène à établir un modèle général discret pour la production de la parole.

3.2. - Etude générale de la propagation d'onde acoustique dans le conduit vocal :

3.2.1. Généralités :

Les ondes sonores se propagent dans l'air ou dans un autre milieu par la vibration des particules du milieu.

Les lois physiques constituent la base pour décrire la génération et la propagation du son dans le conduit vocal.

En utilisant ces principales lois de la physique, un ensemble d'équations différentielles partielles peuvent être obtenu pour décrire le mouvement de l'air dans le système vocal, la formule et la solution de ces opérations est extrêmement difficile.

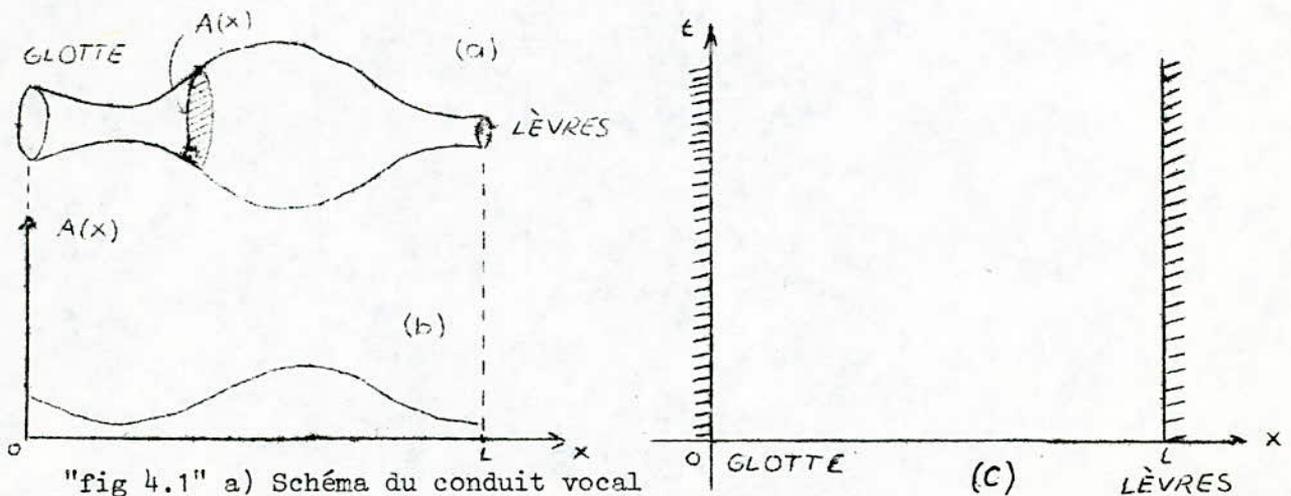
Le détail de la théorie acoustique doit considérer les effets suivants :

1. Variation de la pression et de débit dans le temps du conduit vocal ;
2. Non rigidité des parois du conduit vocal ;
3. Pertes par effet de viscosité et chaleur ;
4. Rayonnement des sons au niveau des lèvres ;
5. Excitation du son dans le conduit vocal.

3.2.2. Equations générales :

Pour des fréquences inférieures à 4000 Hz, l'onde sonore dans le tube "fig 4.1a" satisfait aux équations (4.1).

Le conduit vocal est un tube acoustique de sections variables et ayant des parois non rigides "fig 4.1 a, b, c".



"fig 4.1" a) Schéma du conduit vocal

b) La fonction d'aire correspondante

c) Plan des solutions des équations d'onde.

La fig 4.1 b" indique la fonction d'aire du tube de la "fig 4.1 a" à un instant particulier.

Pour les sons continus il est raisonnable de supposer que $A(x, t)$ ne change pas dans le temps ce qui n'est pas le cas pour les sons non continus. Les mesures détaillées de $A(x, t)$ sont extrêmement difficiles pour les sons continus. Une approche pour telle mesure est l'utilisation des photos en rayon X.

Considérons un élément de masse $dm = \rho dr$

- . ρ étant la densité de l'air
- . dr , étant un élément de volume.

Les lois de la physique générale impliquent que sous la pression P l'élément dm est soumis à la force.

$$F = - \frac{\partial P}{\partial x} dr$$

Par ailleurs la lois de Newton permet d'écrire

$$F = \frac{d}{dt} (v \rho \cdot dr) \quad \text{où } v \text{ est la vitesse corpusculaire}$$

L'élément de masse étant constant au cours du temps.

On peut écrire :

$$\frac{d}{dt} (\rho \cdot dr) = 0$$

Comme :

$$\frac{d}{dt} v = \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \quad \text{et} \quad - \frac{\partial P}{\partial x} dr = \rho \cdot dr \cdot \frac{dv}{dt}$$

On peut tirer l'expression

$$- \frac{\partial P}{\partial x} = \rho \frac{\partial v}{\partial t} + \rho v \frac{\partial v}{\partial x} \quad (1)$$

La masse d'air, située entre les aires $A(x)$ et $A(x + dx)$ s'exprime de la façon suivante :

$$m = \iint_{A(x)} \rho \cdot dA$$

La variation du flux est donnée par :

$$- \frac{\partial}{\partial x} \iint_{A(x)} \rho \cdot v \cdot dA = \Delta \varphi$$

La conservation de la masse permet d'écrire :

$$-\frac{\partial}{\partial x} \iint_{A(x)} \rho \cdot v \, dA = \frac{\partial}{\partial t} \iint_{A(x)} \rho \cdot dA \quad (2)$$

Soient ρ_0, v_0, ρ_0 les paramètres décrivant l'état d'équilibre du système, un état perturbé peut-être représenté par :

$$P = P_0 + P; \quad v = v_0 + v; \quad \rho = \rho_0 + \rho'; \quad P, v, \rho' \quad \text{sont des perturbations.}$$

L'air étant un gaz parfait, par conséquent ;

$$\rho = \rho_0 \left[1 + \left(\frac{1}{\rho_0 c^2} \right) \cdot P \right]$$

Portons cette expression dans les équations [(1) et (2)] et en ne tenant compte que des termes du premier ordre, on obtient :

$$-\frac{\partial P}{\partial x} = \rho_0 \frac{\partial v}{\partial t} + \rho_0 v_0 \frac{\partial v}{\partial t} \quad (3)$$

$$-\frac{\partial}{\partial x} \iint_{A(x)} \left(v + \frac{1}{\rho_0 c^2} \rho_0 v_0 \right) dA = \frac{\partial}{\partial t} \iint_{A(x)} \left(1 + \frac{1}{\rho_0 c^2} P \right) dA \quad (4)$$

Soit U la vitesse volumique ou le débit d'air ; on a alors :

$$U = \iint_{A(x)} v \cdot dA; \quad PA = \iint_{A(x)} P \cdot dA; \quad v = U/A$$

Intégrons les relations (3) et (4), on obtient :

$$-\frac{\partial P}{\partial x} = \rho_0 \frac{\partial}{\partial t} \left[\frac{U}{A} \right] + \rho_0 v_0 \frac{\partial}{\partial t} \left[\frac{U}{A} \right] \quad (5)$$

$$-\frac{\partial U}{\partial x} - \frac{1}{\rho_0 c^2} v_0 \cdot \frac{\partial}{\partial t} (P \cdot A) = \frac{\partial A}{\partial t} + \frac{1}{\rho_0 c^2} \cdot \frac{\partial (P \cdot A)}{\partial t} \quad (6)$$

La dérivée de P ou U par rapport au temps, contrairement à la dérivée par rapport à x , fait apparaître un facteur multiplicatif égal à la vitesse du son C et comme en générale, $v_0 \ll C$, on néglige v_0 devant C dans les relations (5) et (6) ; on aura alors :

$$-\frac{\partial P}{\partial x} = \rho_0 \frac{\partial}{\partial t} (U/A) \quad (7)$$

$$-\frac{\partial U}{\partial x} = \frac{\partial A}{\partial t} + \frac{1}{\rho_0 c^2} \cdot \frac{\partial (PA)}{\partial t} \quad (8)$$

où $P = P(x, t)$: variation de la pression sonore dans le tube à une position x et à un instant t .

$U = U(x, t)$: variation de la vitesse volumique dans le tube à une position x et à un instant t .

ρ_0 : densité de l'air dans le tube.

C : vitesse du son.

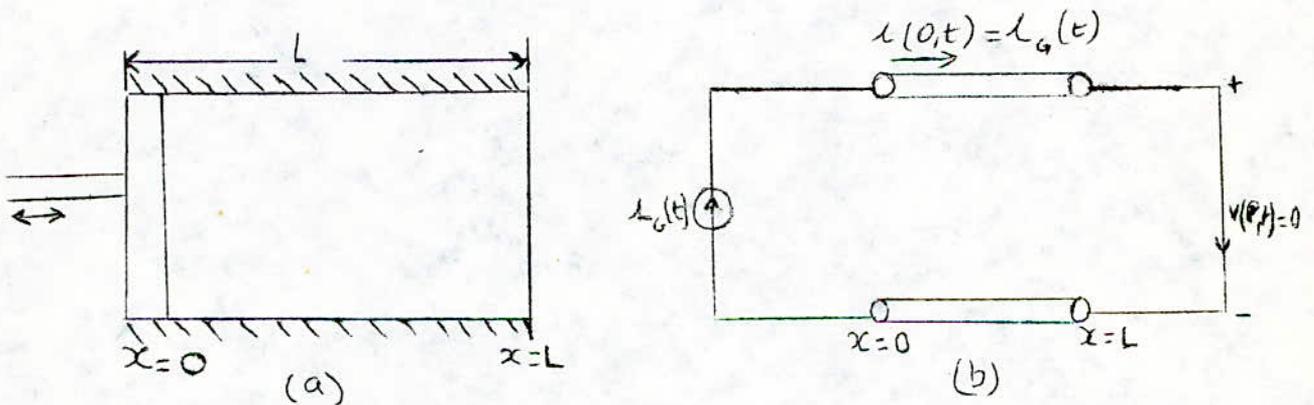
$A = A(x, t)$: fonction d'aire du tube.

3.2.3. Modèles - Etudes théorique :

3.2.3.1. Cas du tube homogène sans pertes :

Le tube homogène sans pertes à parois rigides, représente le modèle le plus simple dont la fonction d'aire est considérée comme constante.

Ce modèle qui est le noyau de base de la discrétisation du conduit vocal, peut être schématisé comme un tube acoustique de section droite homogène excité par une source idéale de débit "fig 4.2".



"fig 4.2" a) Tube uniforme sans pertes
b) Analogue électrique du tube.

Cette source est représentée par un piston dont le mouvement est indépendant de la variation de pression dans le tube. On considère par ailleurs que l'autre bout est ouvert, de cette façon on obtient une variation de débit mais pas de variation de pression.

Cette considération est faite pour les besoins de la cause mais en réalité cela n'est pas possible, on justifie cette démarche par l'intérêt que nous portons à un bon nombre de résultats communs à un modèle plus réaliste.

Posons $A(x, t) = A = \text{Constante}$; les équations différentielles peuvent être écrites sous la forme :

$$\begin{aligned} -\frac{\partial P}{\partial x} &= \rho/A \cdot \frac{\partial U}{\partial t} \\ -\frac{\partial U}{\partial x} &= A/\rho c^2 \cdot \frac{\partial P}{\partial t} \end{aligned} \quad (4.1)$$

On montre que les solutions de ces équations sont de la forme

$$u(x, t) = U^+(t - x/c) - U^-(t + x/c) \quad (9)$$

$$P(x, t) = [U^+(t - x/c) + U^-(t + x/c)] \rho c/A \quad (10)$$

Les fonctions $U^+(t - x/c)$ et $U^-(t + x/c)$ sont respectivement les ondes progressives et régressives se propageant dans le tube acoustique.

Pour une ligne uniforme la tension $V(x, t)$ et le courant $i(x, t)$ satisfont aux équations :

$$\begin{cases} -\frac{\partial U}{\partial x} = L \frac{\partial i}{\partial t} \\ -\frac{\partial i}{\partial x} = C \frac{\partial V}{\partial t} \end{cases}$$

Où L et C sont l'inductance et la capacité par unité de longueur respectivement.

Pour un tube acoustique uniforme, l'équivalence est donnée par le tableau suivant :

GRANDEURS ACOUSTIQUES	GRANDEURS ELECTRIQUES
P : Pression	V : Tension
U : Débit d'air	I : Courant
ρ/A : Inductance Acoustique	L : Inductance
$A/\rho c^2$: Capacité Acoustique	C : Capacité.

Il existe des relations entre ces deux ondes, déterminées par les conditions aux limites. Par analogie ou système linéaire représenté par une ligne électrique on peut envisager le passage à la représentation dans le domaine fréquentiel.

En utilisant cette analogie (équivalence acoustique-électrique) le tube acoustique uniforme est équivalente à une ligne de transmission terminée par un court-circuit $[v(l,t) = 0]$

La ligne excitée par une source de courant $i(0,t) = i_G(t)$ "fig 4.1b"

La représentation dans le domaine fréquentiel d'un système linéaire (tube uniforme) est obtenue par les conditions:

$$(11) \quad U(0,t) = U_G(t) = U_G(\omega) e^{j\omega t} \quad (\text{au niveau de la glotte})$$

Le tube excité par une variation complexe du débit d'air de pulsation et d'amplitude complexe $U_G(\omega)$ de l'équation linéaire (4.1).

La solution de $U^+(t-x/c)$ et $U^-(t+x/c)$ doit avoir la forme suivante :

$$U^+(t-x/c) = k^+ e^{j\omega(t-x/c)} \quad (12)$$

$$U^-(t+x/c) = k^- e^{j\omega(t+x/c)} \quad (13)$$

Portons ces expressions dans les équations (12) et (13), on aura avec la condition limite aux lèvres $P(l,t) = 0$ et celle de la glotte (11), il vient:

$$P(x,t) = \frac{jZ_0 \sin[\omega(l-x)/c]}{\cos \omega l/c} U_G(\omega) e^{j\omega t} \quad (14)$$

$$U(x,t) = \frac{\cos[\omega(l-x)/c]}{\cos \omega l/c} U_G(\omega) e^{j\omega t} \quad (15)$$

avec $Z_0 = \rho c/A$ "impédance caractéristique acoustique "

En d'autres termes cela revient à poser :

$$P(x,t) = P(x,\omega) e^{j\omega t} \quad (16)$$

$$U(x,t) = U(x,\omega) e^{j\omega t} \quad (17)$$

Et d'après les équations (7) et (8) on tire les expressions

$$-\frac{dP}{dx} = Z \cdot U \quad (18)$$

$$-\frac{dU}{dx} = Y \cdot P \quad (19)$$

avec $Z = j\omega \rho/A$ "impédance acoustique par unité de longueur "

$Y = j\omega A/\rho c^2$ "Admittance acoustique par unité de longueur "

La forme générale des solutions des équations (18) et (19) est :

$$P(x, \Omega) = a e^{+\delta x} + b e^{-\delta x} \quad (20)$$

$$u(x, \Omega) = c e^{+\delta x} + d e^{-\delta x} \quad (21)$$

avec $\delta = (ZY)^{1/2} = j\Omega/c$, et a, b, c et d sont des constantes des constantes déterminées par les conditions limites :

$$P(l, \Omega) = 0 \quad (22)$$

$$U(0, \Omega) = U_g(\Omega) \quad (23)$$

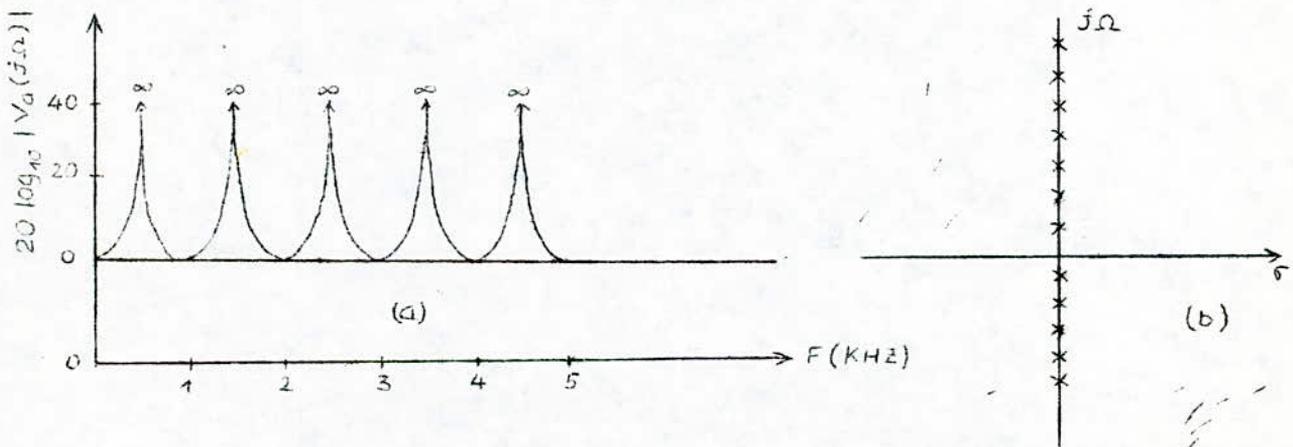
Il en résulte que ces équations sont tout à fait similaires à celles de (14) et (15), en particulier si nous considérons le débit à la glotte et aux lèvres on peut écrire les relations suivantes :

$$U(l, t) = U(l, \Omega) e^{j\Omega t} = \frac{1}{\cos(\Omega l/c)} U_g(\Omega) e^{j\Omega t} \quad (24)$$

$$\frac{U(l, \Omega)}{U_g(\Omega)} = \frac{1}{\cos(\Omega l/c)} = V_a(j\Omega) \quad (25)$$

$V_a(j\Omega)$ est la réponse en fréquence du système (c'est-à-dire que ce rapport représente la réponse en fréquence reliant le débit d'air entrant et sortant).

Cette fonction est donnée sur la figure 4.3 pour un tube de 17,5 cm et $C = 35.000 \text{ cm/}$



"fig 4.3" a) réponse en fréquence du tube homogène sans pertes
b) Lieux des poles de cette fonction.

Dans le plan de la place :

$$\Omega = s/j, \quad V_a(s) = \frac{2e^{-s\ell/c}}{1 + e^{-2s\ell/c}} \quad (26)$$

On remarque que $V(s)$ à une infinité de pôles également distants et placés, sur l'axe imaginaire, ces pôles sont donnés par :

$$S_n = \pm j \left[\frac{(2n+1)\pi c}{2l} \right] , \quad n = 0, \pm 1, \pm 2, \pm 3, \dots \quad (27)$$

Ces pôles correspondent aux fréquences de résonance du système qu'on appelle fréquences de formant dans le cas de la parole.

Comme nous allons le voir, un effet similaire se produit dans le conduit vocal.

Nous venons de voir une méthode d'analyse dans le cas d'un tube idéal, nous allons donc voir maintenant comment passer à un modèle plus réaliste en étudiant les effets des pertes.

3.2.3.2. Tube avec pertes :

Les pertes dans le conduit vocal et leurs effets :

Jusqu'à présent toutes les équations de propagation du son que nous avons décrites considéraient qu'il n'y a pas de pertes d'énergie dans le tube.

En réalité dans le conduit vocal, les parois n'étant pas rigides, elles vibrent sous l'action des forces produites par la vibration de pression le long du système phonatoire.

Les pertes d'énergie dans le conduit vocal se résument ; en pertes par viscosité, pertes sous forme de chaleur et par vibration des parois.

a) Les vibrations de parois et leurs effets :

Nous venons de voir que grâce à l'élasticité des parois, la section droite du tube varie avec la pression, une petite variation de pression entraîne une perturbation de l'aire transversale A

$$A(x, t) = A_0(x, t) + \Delta A(x, t) \quad (28)$$

A_0 est l'aire transversale nominale

A est la perturbation causée par la variation de la pression ; son action est régie par l'équation différentielle :

$$m_w \frac{d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w (\delta A) = p(x,t) \quad (29)$$

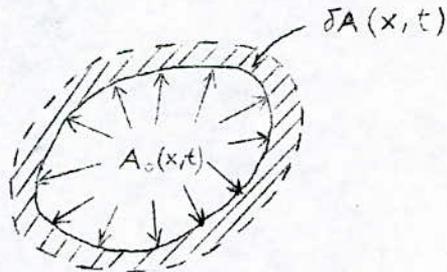
avec :

m_w : la masse des parois par unité de longueur.

b_w et k_w sont les coefficients de pertes par unités de longueur.

"fig 4.4"

Illustration des effets de vibration de parois



En portant cette expression dans les équations (7) et (8) et en négligeant les termes du second ordre en δA et δP , on peut écrire :

$$-\frac{\partial P}{\partial x} = \rho \frac{\partial (U/A_0)}{\partial t} \quad (30)$$

$$-\frac{\partial U}{\partial x} = \frac{1}{\rho c^2} \frac{\partial (PA_0)}{\partial t} + \frac{\partial A_0}{\partial t} + \frac{\partial (\delta A)}{\partial t} \quad (31)$$

Les équations (29), (30) et (31) décrivent la propagation du son dans un tube à parois élastiques.

La condition limite au niveau de la glotte s'écrit alors :

$$U(0,t) = U_g(\omega) e^{j\omega t} \quad (32)$$

Les équations (30) et (31) sont linéaires et ne dépendent pas du temps, dans ce cas le débit et la pression sont :

$$P(x,t) = P(x,\omega) e^{j\omega t} \quad (a) \quad (33)$$

$$U(x,t) = U(x,\omega) e^{j\omega t} \quad (b)$$

En portant (33) dans (29), (30) et (31) il vient que :

$$-\frac{dP}{dx} = z \cdot U \quad (a) \quad (34)$$

$$-\frac{dU}{dx} = y \cdot P + y_w P \quad (b)$$

avec $Z(x, \omega) = j\omega \rho / A_0(x)$ et $Y(x, \omega) = j\omega \frac{A_0(x)}{\rho c^2}$

$$Y_w(x, \omega) = \frac{1}{j\omega m_w(x) + b_w(x) + k_w(x)/j\omega}$$

Nous remarquons que les équations (34) sont identiques aux équations (18) et (19) sauf qu'on a ajouté le terme Y_w et que dans ce cas l'impédance et l'admittance acoustique dépendent de x .

Des mesures faites sur des tissus humains ont permis le calcul de Y_w et les équations différentielles (34) ont été résolues avec la condition limite $P(\ell, t) = 0$ aux lèvres (FLANAGAN).

On a porté le rapport :

$$V_0(j\omega) = \frac{U(\ell, \omega)}{U_0(\omega)}$$

En "fig 4.5" pour un tube à parois non rigides et homogènes.

Les résultats sont voisins de ceux de la "figure 4.4" mais on remarque ici que les pôles ne sont pas situés sur l'axe imaginaire ; par conséquent, il est évident que la réponse en fréquence n'est pas infinie en ces joints (500 Hz, 1500 Hz, 2500 Hz, ...etc), on a porté sur la table de la "figure 4.5" les valeurs des fréquences centrales ainsi que celles de leurs bandes passantes.

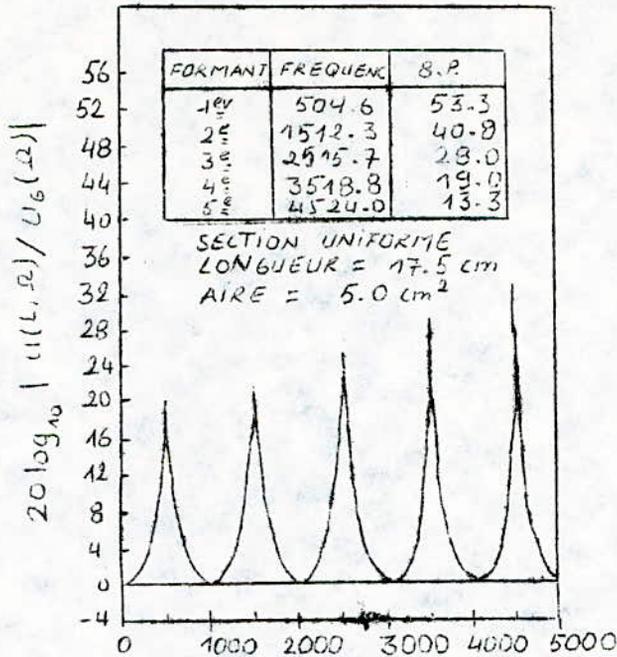
CONCLUSION :

Les fréquences centrales sont un peu plus élevées que celles du tube rigide, les bandes passantes sont plus larges et différentes de zéro. Les effets des pertes dans les parois sont plus prononcés aux fréquences basses et s'atténuent aux fréquences élevées.

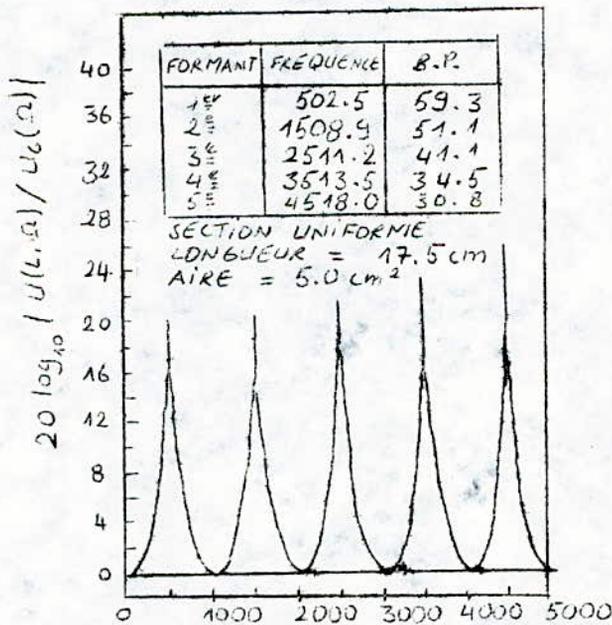
Ces effets se retrouvent bien dans les études des vibrations des parois du conduit vocal.

b) Les effets de viscosité et de chaleur :

Les pertes par viscosité et par chaleur sont moins importantes que celles des vibrations de parois.



"fig 4.5" reponse en frequence du tube a parois non rigides



"fig 4.6" reponse en frequence d'un tube uniforme en incluant les pertes par chaleur, par viscosite et par vibration des parois, termine par un court-circuit $P(1,t) = 0$ (PORTNOFF)

FLANAGAN a considéré ces pertes en détails et a montré que les pertes par viscosité correspondant à la partie réelle de l'impédance acoustique alors que les pertes par chaleur peuvent être incluses dans la partie réelle de l'admittance acoustique.

Dans le domaine des fréquences on peut écrire :

$$Z(x, \Omega) = \frac{S(x)}{[A_0(x)]^2} \sqrt{\Omega \rho \mu / 2} + j \Omega \frac{\rho}{A_0(x)} \quad (35)$$

$S(x)$: Section droite du tube

μ : Coefficient de frottement

ρ : Densité de l'air dans le tube.

$$Y(x, \Omega) = \frac{S(x) (\eta - 1)}{\rho c^2} \sqrt{\frac{\lambda \Omega}{2 c_p \rho}} + j \Omega \frac{A_0(x)}{\rho c^2} \quad (36)$$

c_p : Chaleur spécifique à pression constante

η : Rapport entre chaleur spécifique et le volume

λ : Coefficient de chaleur de conduction.

FLANAGAN a résolu les équations 35 et 36, en donnant les valeurs aux constantes ; il a remarqué que les pertes par viscosité sont proportionnelles à la partie réelle de $Z(x, \Omega)$, par conséquent varient comme $\Omega^{1/2}$ et les pertes par chaleur qui sont proportionnelles à la partie réelle de l'admittance varient aussi comme $\Omega^{1/2}$.

En portant ces nouvelles valeurs dans l'équation différentielle (34) et après résolution en tenant compte de la condition $P(l, t) = 0$, on obtient les résultats de la figure 4.6.

En comparant les figures 4.5. et 4.6, on constate que les fortes perturbations sont du côté des fréquences élevées ce qui est attendu puisque les pertes sont proportionnelles à

CONCLUSION :

Les pertes par viscosité et chaleur augmentent avec la fréquence ; donc leur effets sont plus importants aux fréquences de résonances élevées ; tandis que les pertes dues aux vibrations des parois apparaissent plutôt aux fréquences basses.

c) Effets de rayonnement aux lèvres :

Nous avons toujours considéré les conditions limite aux lèvres $P(\ell, t) = 0$, en réalité le conduit vocal se termine par une ouverture des lèvres et le modèle le plus réaliste est celui représenté par l'orifice d'une sphère "fig 4.7".

En basse fréquence cette ouverture peut être prise comme une surface rayonnante et l'onde sonore rayonnée est diffractée par le baffle sphérique ; les effets qui résultent de la diffraction sont difficiles à représenter mais ce qui nous intéresse le plus c'est de déterminer la condition aux limites au niveau la pression et le débit volumique "fig 4.7a".

L'ouverture des lèvres étant petite devant la taille de la sphère, on peut approcher cette surface sphérique par celle d'un baffle plan infini "fig 4.7b".

Dans ces conditions on définit la relation suivante :

$$P(\ell, \Omega) = Z_L(\Omega) U(\ell, \Omega) \quad (37)$$

Z_L : Impédance de rayonnement aux lèvres.

Elle peut se mettre sous la forme approximative :

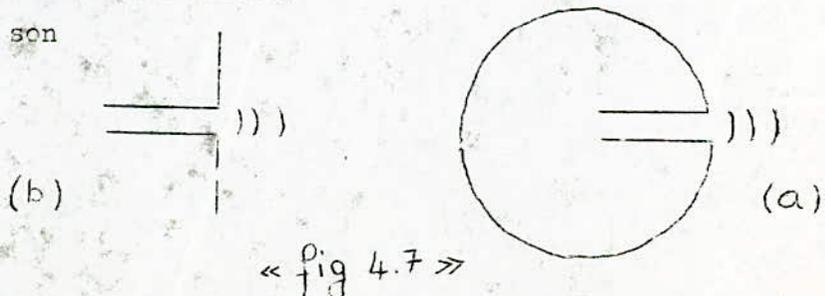
$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r}$$

Ce qui signifie que Z_L est l'impédance équivalente de la résistance et de l'inductance de rayonnement montées en parallèle R_r et L_r sont calculées à partir d'un baffle plan infini :

$$R_r = \frac{128}{9\pi^2} ; \quad L_r = \frac{8a}{3\pi c}$$

a : Rayon d'ouverture du baffle

c : Vitesse du son



"fig 4.7" a) modèle sphérique
b) modèle plan infini.

- Aux basses fréquences $Z_L(\Omega) = 0$ ce qui peut se remplacer par un court circuit.
- Aux fréquences moyennes, on a $\Omega L_r \ll R_r$, alors $Z_L(\Omega) = j\Omega L_r$ et l'impédance se comporte comme une inductance de rayonnement.
- Aux hautes fréquences $\Omega L_r \gg R_r$ alors $Z_L(\Omega) = R_r$
L'impédance se comporte alors comme une résistance de rayonnement ; on a porté sur la figure 4.8 la partie réelle et la partie imaginaire de l'impédance Z en fonction de la fréquence Ω .
- * L'énergie dissipée est proportionnelle à la partie réelle de cette impédance ; on note que pour le conduit vocal, les pertes par rayonnement sont plus importantes aux fréquences élevées.

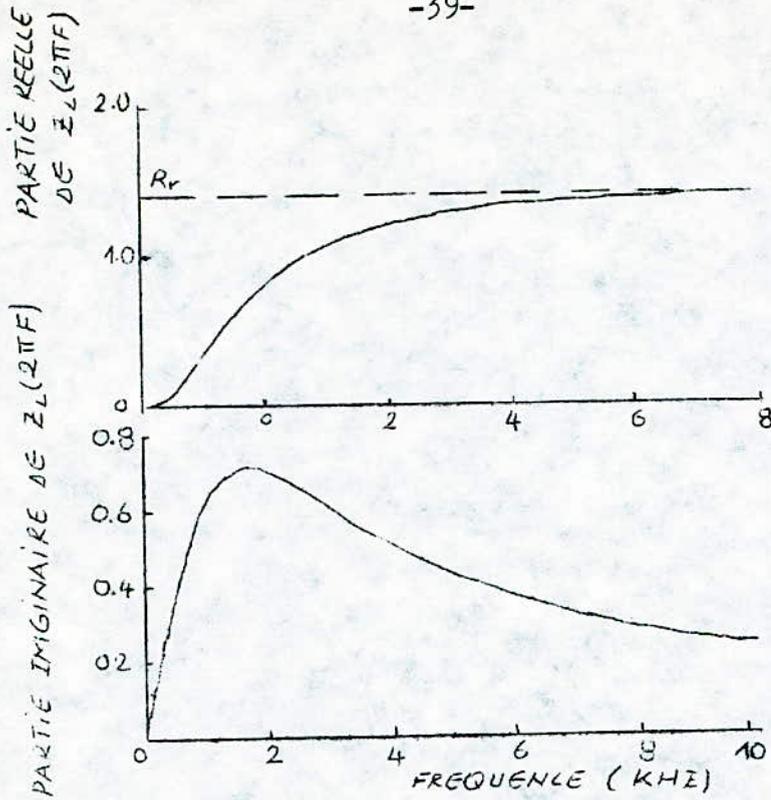
La résolution des équations 34, en tenant compte des pertes par viscosité, par chaleur, par vibration des parois et par rayonnement (baffle plan infini) donne la réponse en fréquence "fig 4.9"

- * L'examen des figures 4.5, 4.6 et 4.9 montre que les bandes augmentent et les fréquences de formant diminuent quand les pertes augmentent.

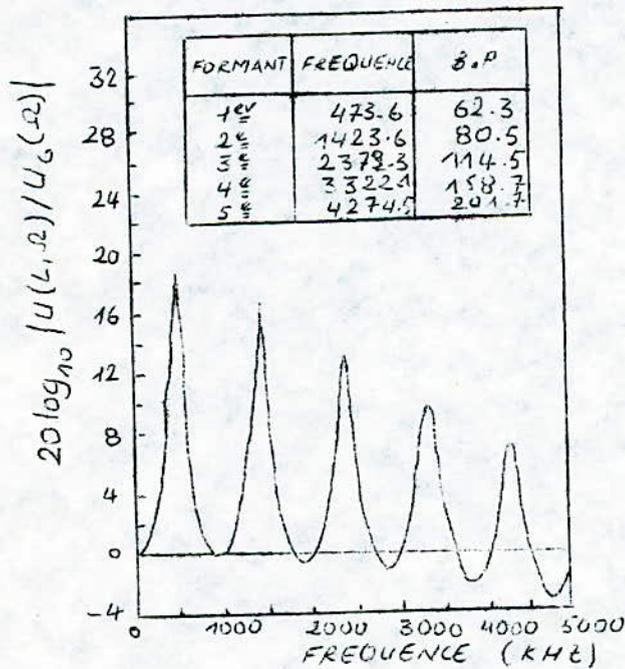
CONCLUSION :

Dans l'ensemble on peut dire que :

- 1) Le système vocal est caractérisé par les fréquences des formants qui dépendent tout d'abord de la fonction d'aire du conduit mais aussi des pertes.



"fig 4.8" partie réelle et imaginaire de Z_1



"fig 4.9" reponse en frequence, tenant compte des pertes par chaleur, vibration, viscosite et par rayonnement (d'apres PORTNOFF)

- 2) Les bandes passantes aux fréquences basses (premier et deuxième formant) dépendent au premier lieu des pertes par vibrations des parois.
- 3) Les bandes passantes aux fréquences élevées dépendent surtout des pertes par viscosité et chaleur et par rayonnement.

3.2.3.3. Excitation du son dans le conduit vocal :

La section précédente décrivait comment les lois physiques peuvent être appliquées à la description de la propagation et la radiation dans la production de la parole, pour compléter la discussion concernant les principes acoustiques, on doit considérer le mécanisme pour lequel les ondes sonores sont générées dans le système vocal.

Un diagramme schématique du conduit vocal (adapté par J.L FLANAGAN) est indiqué par la figure 4.10a, les cordes vocales composent un ensemble d'équations différentielles non linéaires compliquées. Le couplage de ces équations différentielles aux équations différentielles partielles décrivant la transmission de l'onde, peut être représentée par la variation d'une inductance et d'une capacité acoustique dans le temps.

De cette manière le flux glottique est automatiquement coupé en impulsions. Un ensemble de signaux générés par tel modèle est indiqué par la figure 4.11, l'onde supérieure est le débit et l'onde inférieure est la pression au niveau des lèvres pour une configuration du conduit vocal pour la voyelle (a).

Le couplage entre le conduit vocal et la glotte est faible, et il est commun de négliger cette interaction, ceci nous amène à une séparation et une linéarisation du système d'excitation et de transmission donné par la "figure 4.10b".

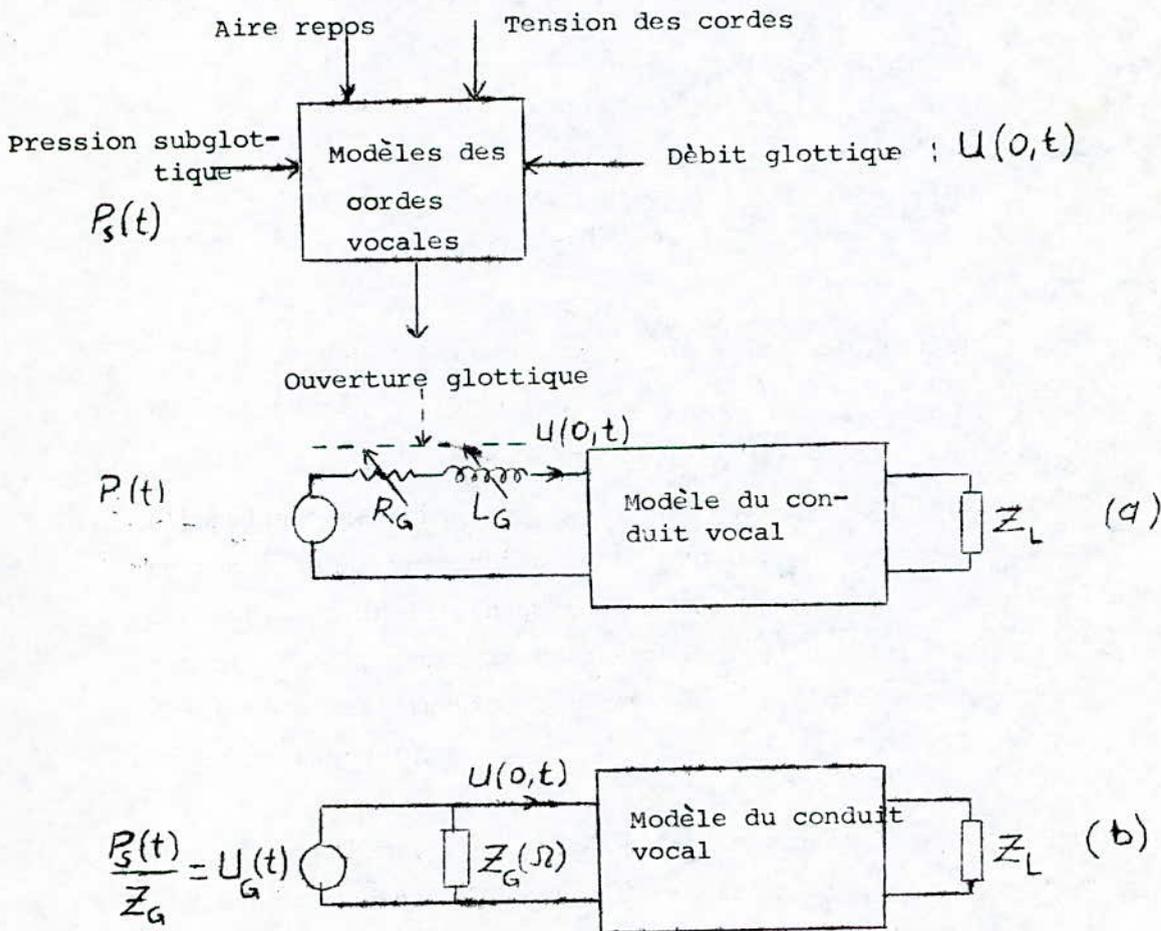
Dans ce cas $U_g(t)$ est la source de débit dont la forme est la même que celle de la figure (4.11) ; d'impédance acoustique glottique Z_g est obtenue par la linéarisation des relations entre la pression et le débit dans la glotte ; cette impédance est de la forme :

$Z_G(\Omega) = R_G + j\Omega L_G$
 R_G et L_G sont des constantes, avec cette configuration la condition aux limites idéal $U(0, \Omega) = U_G(\Omega)$ est remplacé par

$$U(0, \Omega) = U_G(\Omega) - P(0, \Omega) / Z_G(\Omega)$$

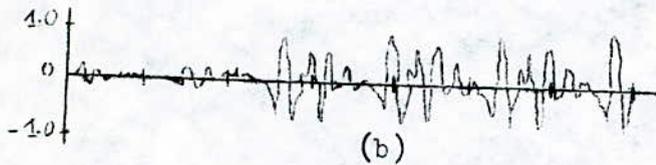
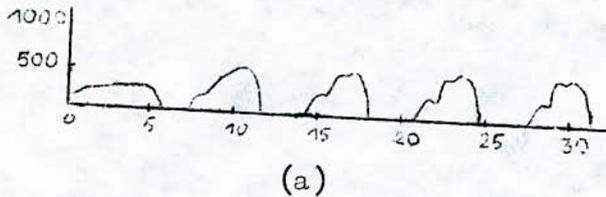
L'impédance glottique Z_G a des effets sur les bandes passantes pour les systèmes de la production de la parole. Le majeur effet est l'élargissement des bandes pour les fréquences de résonances basses, ceci à cause de l'accroissement de Z_G avec la fréquence. Alors aux fréquences élevées Z_G apparait comme un court circuit et toute la source glottique circule dans le système du conduit vocal.

De cette manière, l'élasticité des parois et les pertes glottiques contrôlent les bandes passantes des bas formants, alors que les pertes par radiation contrôlent la bande passante des formants élevés.



"fig 4.10" a) Diagramme du modèle du conduit vocal.

b) Approximation du modèle avec les cordes vocales.



"fig 4.11" Debit glottique (a) et Pression sonore^(b) pour la voyelle |a|
(d'après ISHIZAKA et FLANAGAN)

3.2.4. Discrétisation du conduit vocal :

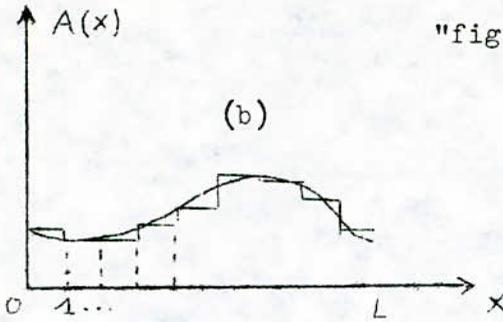
Plusieurs résolutions des équations 7 et 8 ont été proposées mais aucune des méthodes n'a abouti à un résultat satisfaisant.

Ces équations étant à deux variables x et t , il faudra connaître deux conditions limites.

L'état des variables au niveau des cordes vocales et l'impédance de rayonnement au niveau des lèvres.

La résolution numérique impose que l'on transforme 7 et 8 en équations aux différences partielles finies, fonction de variables discrètes, la discrétisation de x et de t est délicate (FORSYTHE - WASON 1960).

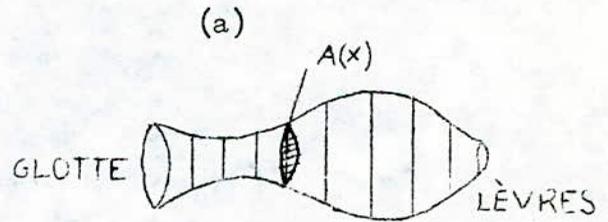
Dans le cas du conduit vocal statique, on préfère utiliser les méthodes d'approximation, c'est-à-dire qu'on discrétise le conduit vocal en une suite de tube ayant des longueurs généralement fixes et de sections variables "fig 4.12".



"fig 4.12"

(b)

Echantillonnage de la fonction d'aire



Discretisation du conduit vocal

Le pas d'échantillonnage est fixé par la fréquence maximale que l'on veut passer dans le conduit vocal.

F_e : la fréquence d'échantillonnage du signal doit être supérieure ou égale à la fréquence de coupure du noyau élémentaire. On admet que l'onde soit unidirectionnelle pour $F < 5$ KHz.

Dans la pratique on considère que le conduit vocale mesure 17 à 20 cm et la fréquence de coupure est de 8 KHz, il est donc suffisant de prendre la longueur des tubes égales à 1 cm pour satisfaire à ces hypothèses.

La discrétisation du conduit implique deux modélisations possibles :

- Le modèle à réflexion, dans lequel on utilise les coefficients aux jonctions des sections élémentaires pour déterminer les variations de pression et de débit en tout point du conduit qui est l'objet de notre étude.
- Le modèle analogue électrique, qui est représenté sous forme d'une ligne électrique artificielle.

La quantification de la fonction d'aire introduit une erreur qui est d'autant plus petite que le nombre de tubes est grand ; le seul critère à respecter est que la longueur d'un tube soit petite devant la longueur d'onde de la plus haute.

3.2.5. Le modèle harmonique :

1. Introduction :

Le conduit vocal peut être représenté, comme nous verrons de le voir, par une concaténation de tubes élémentaires. Cette approche introduit au niveau des jonctions des différents éléments, une discontinuité dans la propagation de l'onde qu'on a supposé plane.

Ce phénomène de discontinuité est évalués à l'aide des coefficients de réflexion.

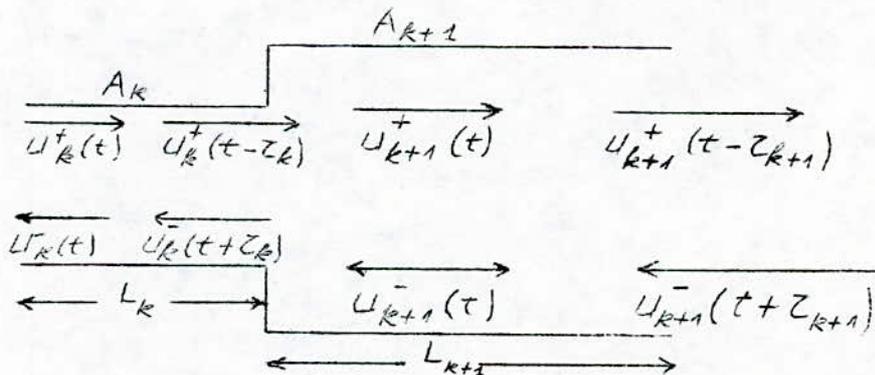
2. Calcul des coefficients de réflexion :

Ecrivons les équations de propagation

$$\begin{aligned}
 -\frac{dP}{dx} &= \rho/A \frac{du}{dt} & (a) \\
 -\frac{du}{dx} &= A/\rho c^2 \frac{dP}{dt} & (b)
 \end{aligned}
 \tag{38}$$

La solution générale de ces équations (au niveau de la jonction K) est de la forme

$$\begin{aligned}
 P_K(x,t) &= \rho c/A_K \left[U_K^+ \left(t - \frac{x}{c} \right) + U_K^- \left(t + \frac{x}{c} \right) \right] & (a) \\
 U_K(x,t) &= U_K^+ \left(t - \frac{x}{c} \right) - U_K^- \left(t + \frac{x}{c} \right) & (b)
 \end{aligned}
 \tag{39}$$



"fig 4.13" Jonction entre les tubes k et k+1.

A_k est la section du tube K

U^+ et U^- sont respectivement l'onde progressive et l'onde régressive.

P_k et U_k représente respectivement la pression et le débit acoustique au niveau du k^{ieme} tube $0 < x < l_k$

Conditions de continuité au niveau de la jonction

$$P_k(l_k, t) = P_{k+1}(0, t) \quad (a) \quad (40)$$

$$U_k(l_k, t) = U_{k+1}(0, t) \quad (b)$$

Des équations 39 et 40 on peut tirer

$$(41) \quad A_{k+1}/A_k [U_k^+(t-z_k) + U_k^-(t+z_k)] = U_{k+1}^+(t) + U_{k+1}^-(t) \quad (a) \quad (41)$$

$$U_k^+(t-z_k) - U_k^-(t+z_k) = U_{k+1}^+(t) - U_{k+1}^-(t) \quad (b)$$

$z_k = l_k/c$: Temps mis pour l'onde pour traverser le tube k.

A partir des équation 41 on peut tirer les expressions de U_{k+1}^+ et U_{k+1}^- fonction de U_k^+ et de U_k^- :

$$U_{k+1}^+(t) = \frac{2A_{k+1}}{A_{k+1} + A_k} U_k^+(t-z_k) + \frac{A_{k+1} - A_k}{A_{k+1} + A_k} U_k^-(t) \quad (42)$$

$$U_{k+1}^-(t) = -\frac{A_{k+1} - A_k}{A_{k+1} + A_k} U_k^+(t-z_k) + \frac{2A_k}{A_{k+1} + A_k} U_k^-(t)$$

La quantité $r_k = \frac{A_{k+1} - A_k}{A_{k+1} + A_k}$ représente le coefficient de réflexion (43)

$$-1 \leq r_k \leq +1$$

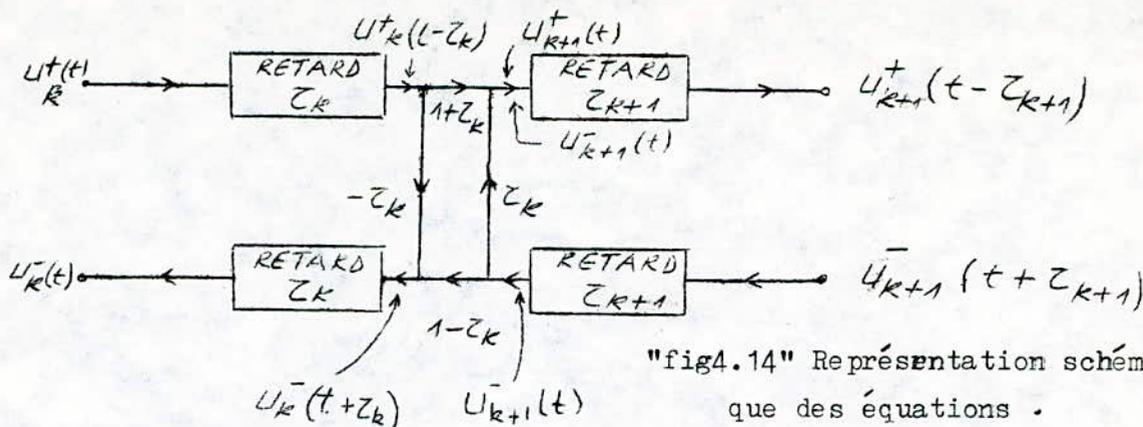
En utilisant la définition de r_k les équations 42 peuvent être écrites sous la forme.

$$U_{k+1}^+(t) = (1+r_k) U_k^+(t-z_k) + r_k U_k^-(t) \quad (44)$$

$$U_{k+1}^-(t) = -r_k U_k^+(t-z_k) + (1-r_k) U_k^-(t)$$

3. Réponse du système :

Schématiquement on peut représenter les équations 44 de la façon suivante :



"fig4.14" Représentation schématique des équations .

Supposons que le conduit vocal soit discrétisé en N tubes élémentaires ; les conditions aux limites sont :

$$\begin{array}{lll}
 P_N(l_N, t) & U_N(l_N, t) & \text{au niveau des lèvres} \\
 P_1(0, t) & U_1(0, t) & \text{au niveau de la glotte.}
 \end{array}$$

Or on sait que :

$$P_N(l_N, t) = Z_L U_N(l_N, t)$$

Pour simplifier les calculs prenons Z réelle et d'après les équations (39) on arrive à :

$$(45) \quad \rho c / A \left[U_N^+(t - z_N) + U_N^-(t + z_N) \right] = Z_L \left[U_N^+(t - z_N) - U_N^-(t + z_N) \right]$$

d'où $U_N^-(t + z_N) = -r_L U_N^+(t - z_N)$

avec

$$r_L = \frac{\rho c / A_N - Z_L}{\rho c / A_N + Z_L} \quad (46)$$

r_L : Coefficient de réflexion au niveau des lèvres.

Les équations du débit d'air sortant par la bouche s'écrivent :

$$\begin{aligned}
 U_N(l_N, t) &= U_N^+(t - z_N) - U_N^-(t + z_N) \\
 U_N(l_N, t) &= (1 + r_L) U_N^+(t - z_N)
 \end{aligned} \quad (47)$$

Au niveau du premier élément (glotte) on a la relation

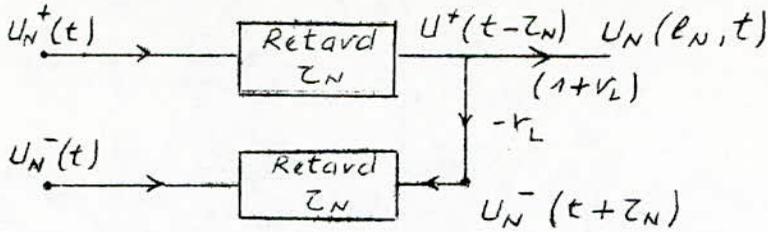
$$U_1(0, t) = U_g(t) - P_1(0, t) / Z_g$$

dans le cas où Z_g est réelle et d'après les équations (36)

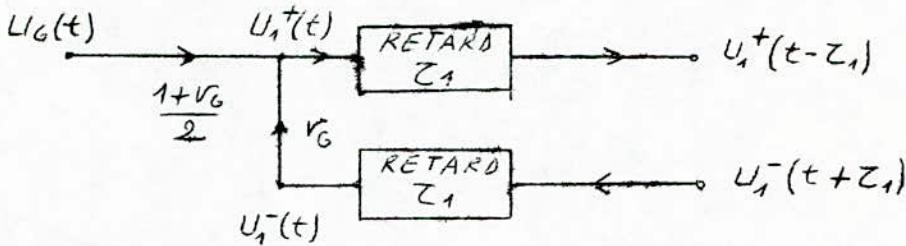
$$U_1^+(t) - U_1^-(t) = U_g(t) - \frac{c}{A_1} \left[\frac{U_1^+(t) + U_1^-(t)}{Z_g} \right] \quad (48)$$

d'où on tire le coefficient de réflexion au niveau de la glotte :

$$r_g = \frac{Z_g - \rho c / A_1}{Z_g + \rho c / A_1}$$



"fig 4.15" Représentation schématique des équations (47).



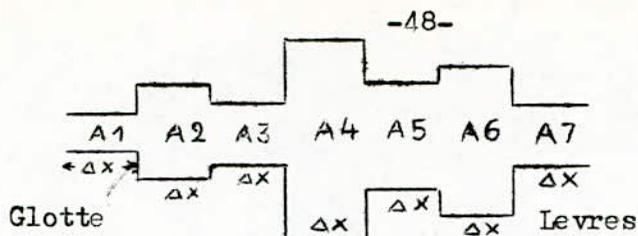
"fig 4.16" Représentation schématique des équations (48).

3.2.7. Etude d'un modèle discret de tubes sans pertes :

a) Généralités :

L'utilisation des modèles pour la production de la parole est basée sur l'hypothèse que le conduit vocal peut être représenté par une succession de tubes acoustiques sans pertes donné par la "figure 4.17". Les aires de la section (A_k) des tubes sont choisies de telle façon à approximer la fonction d'aire $A(x)$ du conduit vocal. Cette approximation réglie les pertes dues à la viscosité, chaleur et les vibrations des parois ; dans ce cas les bandes passantes des fréquences de résonances sont différentes de celle du modèle qui inclut les pertes. Cependant les pertes au niveau de la glotte et lèvres existent.

Soit un système composé de N tubes sans pertes, chacun de longueur $\Delta x = \frac{l}{N}$ avec l : longueur totale du conduit vocal.



"fig 4.17" concaténation de ($N = 7$) tubes sans pertes de même longueur.

La propagation de l'onde peut être représentée comme dans la "figure 4.14" avec tous les retards égaux à $z = \frac{\Delta x}{c}$ qui est le temps de propagation le long d'un tube.

Il est instructif de commencer par considérer la réponse du système à une source d'impulsions unitées, $U_G(t) = \delta(t)$

L'impulsion se propageant dans les séries de tubes est partiellement réfléchiée et propagée au niveau de la jonction. La réponse impulsionnelle est donnée par :

$$U_a(t) = \alpha_0 \delta(t - Nz) + \sum_{k=1}^{\infty} \alpha_k \delta(t - Nz - 2kz)$$

Le temps de propagation d'une impulsion le long de N tubes est de Nz secondes. En outre des impulsions successives dues à la réflexion au niveau de la jonction atteignent la sortie du tube, à un multiple de $2z$ secondes. La quantité $2z$ est le temps nécessaire de propagation sur deux chemins d'une section.

La fonction d'un tel système est sous la forme :

$$V_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-s(N+2k)z} = e^{-sNz} \sum_{k=0}^{\infty} \alpha_k e^{-s2zk}$$

Le facteur e^{-sNz} correspond au retard de propagation à travers N sections ; la quantité :

$$U_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-s2zk}$$

est une fonction d'un système linéaire dont la réponse impulsionnelle est simplement :

$$\hat{U}_a(t) = U_a(t + Nz)$$

La "figure 4.18a" représenté le diagramme d'un modèle sans perte indiquant la séparation du système $\hat{U}_a(t)$ du retard,

La réponse en fréquence $\hat{V}_a(\omega)$ est :

$$\hat{V}_a(\omega) = \sum_{k=0}^{\infty} \alpha_k e^{-j\omega k 2\tau}$$

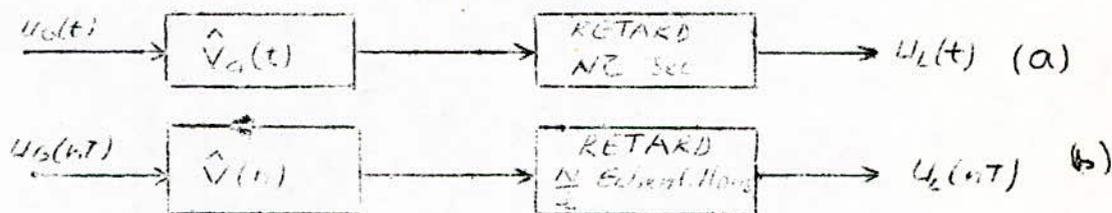
avec

$$\hat{V}_a(\omega + 2\pi/2\tau) = \hat{V}_a(\omega)$$

qui est la réponse fréquentielle d'un système numérique. En fait l'entrée du système c'est-à-dire l'excitation) est limitée à des fréquences inférieures à $\pi/2\tau$, alors on peut échantillonner l'entrée avec une période $T=2\tau$ et filtrer le signal échantillonné avec un filtre numérique dont la réponse impulsionnelle.

$$\begin{cases} \hat{U}(n) = \alpha_n & n \geq 0 \\ \hat{U}(n) = 0 & n < 0 \end{cases}$$

Pour une période d'échantillonnage $T=2\tau$, le retard de $N\tau$ secondes correspond à un changement de $\frac{N}{2}$ échantillons. De cette manière l'équivalent du système discret pour des entrées dont la bande est limitée comme l'indique la "figure 4.18b".



"fig 4.18" a) Diagramme de représentation du modèle acoustique sans pertes.

b) Système discret équivalent.

La transformée en Z de $\hat{U}(n)$ est :

$$\hat{V}(z) = \sum_{k=0}^{\infty} \alpha_k z^{-k}$$

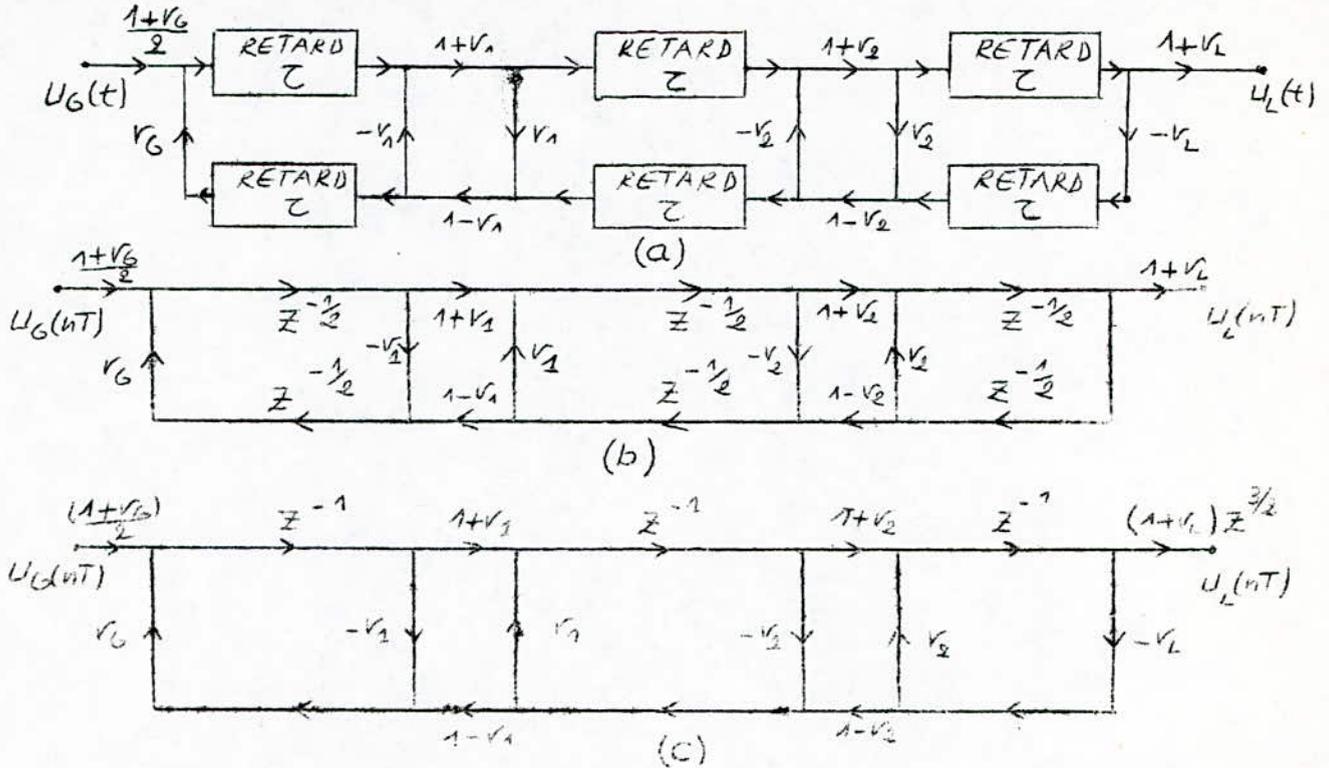
Chaque retard τ secondes est remplacé par un retard de $\frac{1}{2}$ échantillon

Un exemple est donné par la "figure 4.19" on note en particulier que le retard de propagation est représenté par une transmittance de $z^{-1/2}$ comme le montre la "figure 4.19b"

"fig 4.19" a) Schéma du signal pour le modèle du tube sans perte du conduit vocal.

b) Système discret équivalent.

c) Système discret équivalent utilisant seulement des retards entiers.



On peut voir que le retard autour de quelques fermetures du chemin "fig 4.19b" sera conservé si les retards dans les branches inférieures sont transférés en haut aux branches correspondantes directement. Le retard total (entrée/sortie) n'est pas correct, mais cette configuration minutieuse en pratique, peut être compensée par l'insertion d'un élément d'avance (en générale $z^{\frac{N}{2}}$). La figure 4.19c indique comment cette dernière est faite pour un exemple de trois tubes. Ce réseau numérique peut être utilisé pour évaluer les échantillons d'un signal parole produits à partir des échantillons du signal d'excitation.

b) Fonction de transfert :

Pour compléter notre discussion sur les modèles discrétisés de tubes sans pertes, pour la production de la parole, il est instructif de trouver une expression générale de la fonction de transfert avec les termes de coefficient de réflexion.

Les équations qu'on va trouver sont obtenues avant par ATAL, HANAVER, MARKEL, GRAY et WAKITA.

Notons que nous allons chercher la fonction de transfert :

$$V(z) = \frac{U_L(z)}{U_G(z)}$$

Pour trouver $V(z)$, il est plus convenable d'exprimer $U_G(z)$ en termes de $U_L(z)$ et d'en déduire le rapport. Considérons la "figure 4.20" qui schématise une jonction du modèle d'un tube sans pertes. La transformée en Z des équations de cette jonction est donnée par :

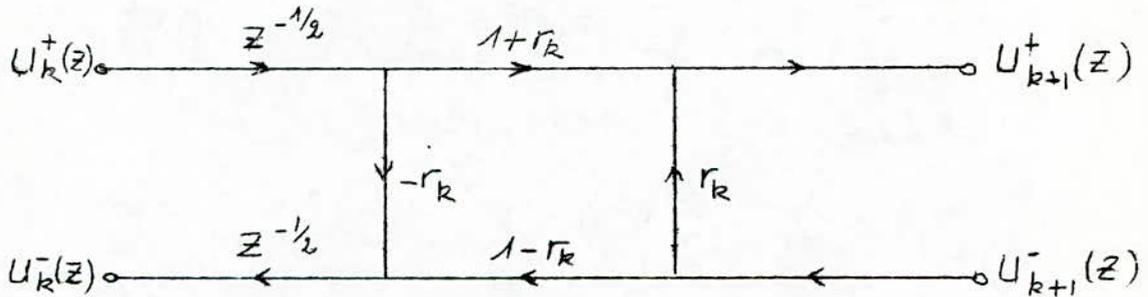
$$U_{k+1}^+(z) = (1+r_k)z^{-\frac{1}{2}} U_k^+(z) + r_k U_{k+1}^-(z)$$

$$U_k^-(z) = -r_k z^{-\frac{1}{2}} U_k^+(z) + (1-r_k)z^{-\frac{1}{2}} U_{k+1}^-(z)$$

d'où on tire :

$$U_k^+(z) = \frac{z^{1/2}}{1+r_k} U_{k+1}^+(z) - \frac{r_k z^{1/2}}{1+r_k} U_{k+1}^-(z) \quad (49)$$

$$U_k^-(z) = \frac{-r_k z^{-1/2}}{1+r_k} U_{k+1}^+(z) + \frac{z^{-1/2}}{1+r_k} U_{k+1}^-(z)$$



"fig 4.20"

Pour mettre le résultat plus compact, il est utile de représenter les conditions aux limites au niveau des lèvres, ainsi que pour toutes les jonctions dans le système.

Soit $U_{N+1}(z)$ la transformée en Z à l'entrée du $(N+1)^{\text{ième}}$ tube imaginaire infiniment long, alors il n'y a pas d'onde régressive dans le $(N+1)^{\text{ième}}$ tube. Un équivalent point de vue est que ce dernier est terminé par son impédance caractéristique.

On note que dans certain cas on a :

$$U_{N+1}^+(z) = U_L(z)$$

$$U_{N+1}^-(z) = 0$$

A partir des équations 43 et 46 :

Si $AN + 1 = \rho c / z_L$ Alors $r_N = r_L$

Les équations (44) peuvent être écrites sous formes matricielle :

$$U_k = Q_k U_{k+1}$$

$$U_k = \begin{bmatrix} U_k^+(z) \\ U_k^-(z) \end{bmatrix} \quad Q_k = \begin{bmatrix} \frac{z^{1/2}}{1+r_k} & \frac{-r_k z^{1/2}}{1+r_k} \\ \frac{r_k z^{-1/2}}{1+r_k} & \frac{z^{-1/2}}{1+r_k} \end{bmatrix}$$

A partir de l'équation (45) on peut exprimer les variables d'entrée du premier tube en terme des variables de sortie par la matrice :

$$U_1 = Q_1 \cdot Q_2 \cdot Q_3 \dots Q_N U_{N+1}$$

$$U_1 = \prod_{k=1}^N Q_k U_{N+1}$$

A partir de la "figure 4.16" on peut voir que les conditions aux limites au niveau de la glotte peuvent être exprimées par :

$$U_G(z) = \frac{2}{1+r_G} U_1^+(z) - \frac{2r_G}{1+r_G} U_1^-(z)$$

qui peut être écrite sous la forme :

$$U_G(z) = \begin{bmatrix} \frac{2}{1+r_G} & \frac{-2r_G}{1+r_G} \end{bmatrix} U_1$$

De même

$$U_{N+1} = \begin{bmatrix} U_L(z) \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_L(z)$$

Enfin on peut écrire :

$$(50) \quad \frac{U_G(z)}{U_L(z)} = \begin{bmatrix} 2 & -2r_G \\ 1+r_G & 1+r_G \end{bmatrix} \prod_{k=1}^N Q_k \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{V(z)}$$

Pour examiner les propriétés de $V(z)$, il est utile d'exprimer Q_k sous la forme :

$$Q_k = z^{1/2} \begin{bmatrix} \frac{1}{1+r_k} & \frac{-r_k}{1+r_k} \\ \frac{-r_k z^{-1}}{1+r_k} & \frac{z^{-1}}{1+r_k} \end{bmatrix}$$

$Q_k : z$

$$Q_k = z^{1/2} \hat{Q}_k \quad (51)$$

De cette façon l'équation (50) peut être écrite sous la forme :

$$\frac{1}{V(z)} = z^{N/2} \begin{bmatrix} z & -2r_0 \\ 1+r_0 & 1+r_0 \end{bmatrix} \prod_{k=1}^N \hat{Q}_k \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Par exemple pour $N = 2$:

$$\frac{1}{V(z)} = \frac{z(1+r_1r_2z^{-1} + r_1r_0z^{-1} + r_2r_0z^{-2})z}{(1+r_0)(1+r_1)(1+r_2)}$$

ou bien

$$V(z) = \frac{0,5(1+r_0)(1+r_1)(1+r_2)z^{-1}}{1+(r_1r_2+r_1r_0)z^{-1}+r_2r_0z^{-2}}$$

En général, à partir des équations 51 et 52 pour un modèle de tubes sans pertes, la fonction de transfert peut être toujours exprimée sous la forme :

$$V(z) = \frac{0,5(1+r_0) \prod_{k=1}^N (1+r_k) z^{-N/2}}{D(z)}$$

avec $D(z)$ est un polynôme en z^{-1} donné par la relation suivante :

$$D(z) = \begin{bmatrix} 1 & -r_0 \end{bmatrix} \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

qui peut être mis sous la forme :

$$D(z) = 1 - \sum_{k=1}^N \alpha_k z^{-k} \quad (56)$$

Remarque :

- 1) Pour un modèle de tubes sans pertes la fonction de transfert possède que des pôles (pas de zéros), ces pôles définissent les résonances (formants) du modèle.
- 2) Dans un cas spécial $Q = 1$ ($ZG = \infty$), le polynôme $D(z)$ peut être trouvé en utilisant la formule de récurrence, qui dérive de l'équation (55) :

$$P_1 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} = \left[(1+r_1 z^{-1}), -(r_1 + z^{-1}) \right]$$

On pose

$$D_1(z) = 1 + r_1 z^{-1}$$

Alors :

$$P_1 = \left[D_1(z), -z^{-1} D_1(z^{-1}) \right]$$

De même, le rang de la matrice P_2 est défini par :

$$P_2 = P_1 \begin{bmatrix} 1 & -r_2 \\ -r_2 Z^{-1} & Z^{-1} \end{bmatrix} = \left[D_2(Z), Z^{-2} D_2(Z^{-1}) \right]$$

avec $D_2(Z) = D_1(Z) + r_2 Z^{-2} D_1(Z^{-1})$

Par conséquent on obtient :

$$P_k = P_{k-1} \begin{bmatrix} 1 & -r_k \\ -r_k Z^{-1} & Z^{-1} \end{bmatrix} = \left[D_k(Z), -Z^{-k} D(Z^{-1}) \right]$$

avec $D_k(Z) = D_{k-1}(Z) + r_k Z^{-k} D_{k-1}(Z^{-1})$

Finalement le polynôme désiré $D(Z)$ est :

$$D(Z) = P_N \begin{bmatrix} 1 \\ 0 \end{bmatrix} = D_N(Z)$$

De cette façon, on peut voir qu'il n'est pas nécessaire d'exécuter toutes les multiplications des matrices mais utilisé simplement les relations de récurrence suivantes :

$$D_0(Z) = 1$$

$$D_k(Z) = D_{k-1}(Z) + r_k Z^{-k} D_{k-1}(Z) \quad K = 1, 2, \dots, N$$

$$D(Z) = D_N(Z)$$

Le nombre de section dépend du choix de la fréquence d'échantillonnage, pour représenter le signal de parole. Rappelons que la réponse en fréquence d'un modèle du tubes sans pertes est périodique, ce modèle peut approximer le conduit vocal dans la bande de fréquence $f = \frac{1}{2T}$; avec T période d'échantillonnage. Or on sait que $z = \frac{L}{cN}$

avec L (longueur du modèle) = 17,5cm ; c (vitesse de propagation) : 35.000cm/s

$$\text{dmc} : \frac{1}{2T} = \frac{1}{4z} = \frac{cN}{4L} = \frac{N}{2} \left(\frac{c}{2L} \right) = \frac{N}{2} (1000) \text{ Hz}$$

Ce qui implique qu'on a $\frac{N}{2}$ fréquences de résonances (formants) par 1000 Hz

Pour expliquer ce cas prenons cet exemple :

soit $\frac{1}{T} = 10.000 \text{ Hz}$: fréquence d'échantillonnage

$F = \frac{1}{2T} = 5000 \text{ Hz}$: fréquence maximale du signal

d'où $N = 10$

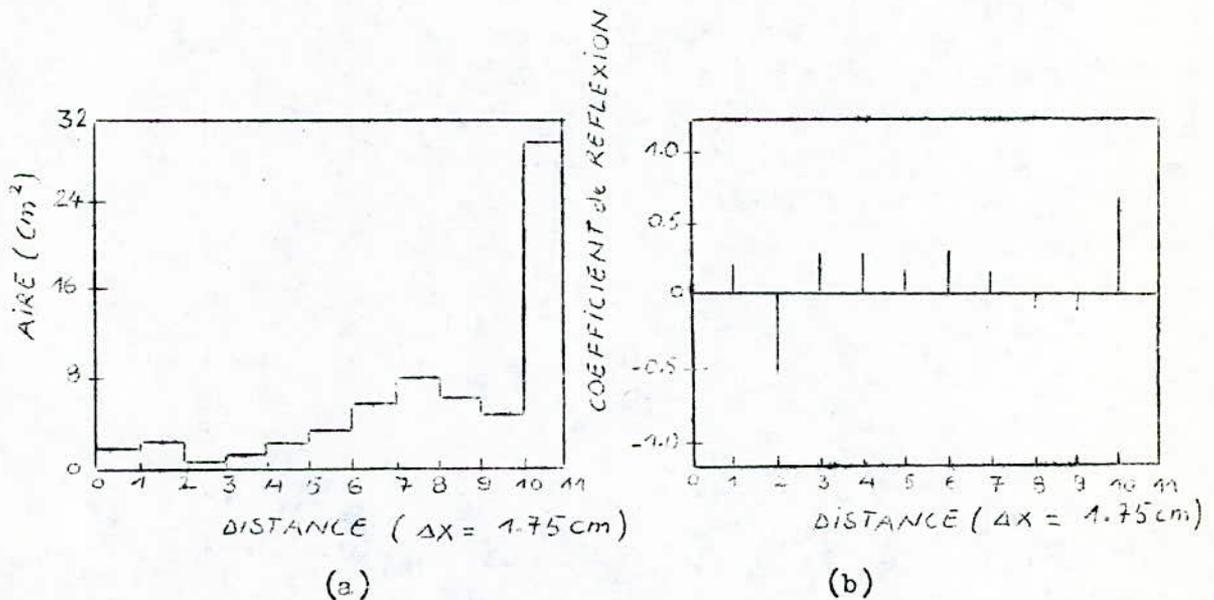
donc on a 5 formants c'est-à-dire une densité d'un formant par 1000 Hz, un court circuit possède peu de formants par 1K Hz et vice versa.

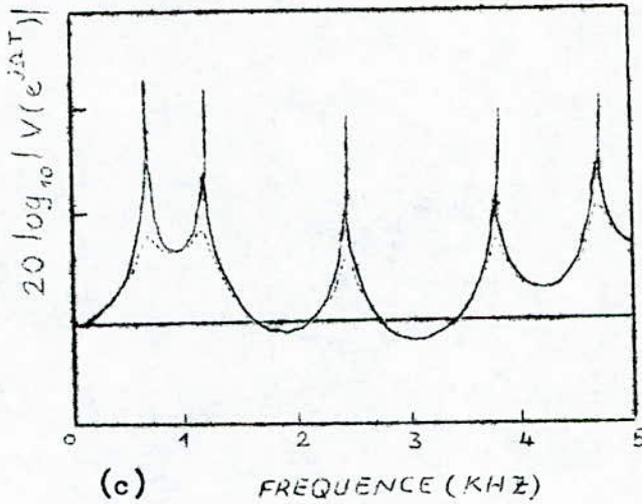
La "figure 4.22" indique un exemple pour $N = 10$ et $\frac{1}{T} = 10$ KHz

La "figure 4.22a", indique les fonctions d'aire de la "figure 4.22a" échantillonnées pour donner dix tubes pour l'approximation de la voyelle (a).

La "figure 4.21b", indique le résultat d'un ensemble de dix coefficients de réflexion pour $A_{11} = 30$ cm², ce qui nous donne un coefficient de réflexion au niveau des lèvres $r_N = 0,714$.

La "figure 4.21c", indique la réponse en fréquence pour $r_N = 1$ et $r_N = 0,714$ (en pointillé). La comparaison entre la courbe en pointillé de la "figure 4.21c et la "figure 4.22b" incluant les pertes au niveau des lèvres, la réponse en fréquence d'un modèle du tube sans pertes ressemble beaucoup à celui du modèle plus détaillé.





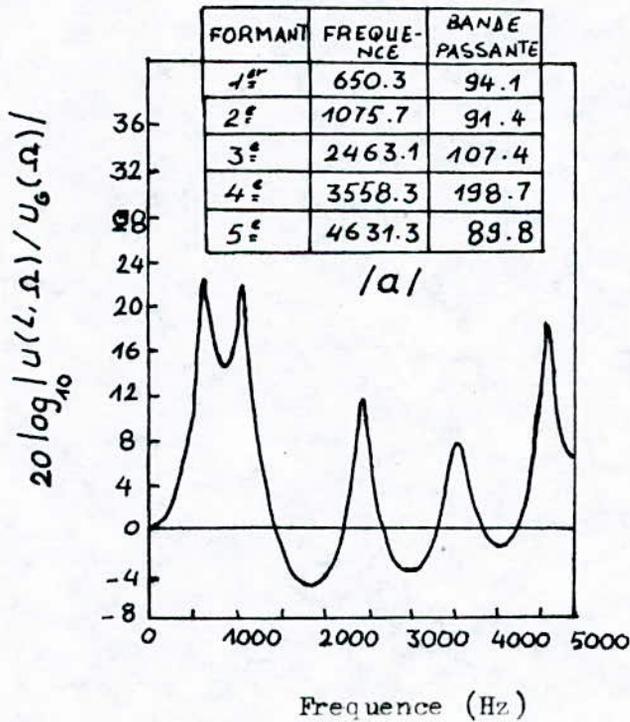
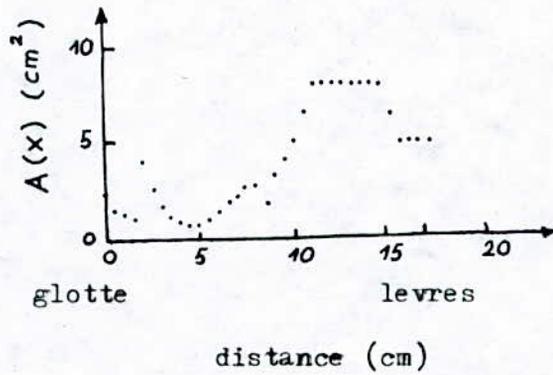
"figure 4.21"

3.2.8. Modèle numérique du signal parole :

Pour produire un signal de parole, le mode d'excitation et les propriétés de de résonance du système linéaire doivent changer dans le temps. Pour quelques sous, il est raisonnable de supposer que les propriétés générales de l'excitation et le conduit vocal restent fixes pour des périodes de 10 à 20 ms. Donc un modèle terminal analogue entraine un système linéaire variant lentement dans le temps, par une excitation dont la nature change du quasi-périodique pour les sous voisés, au bruit pour les sous non voisés.

Le modèle du tube discrétisé sans perte du paragraphe précédent sert comme exemple de ce qu'on vient de signifier.

Le schéma essentiel de ce modèle est celui de la "figure 4.23a" rappelons que le système vocal est caractérisé par un ensemble de fonctions d'aire ou de coefficients de réflexions.



"fig 4.22" pour la voyelle /a/

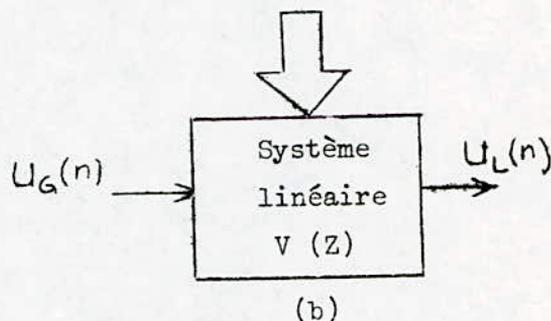
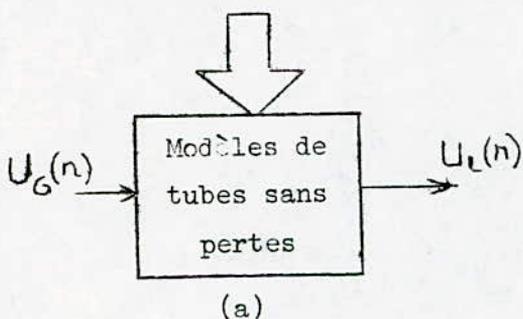
Fonction d'aire (Fant) et la reponse en frequence (PORTNOFF)
pour la voyelle /a/.

Les systèmes de la figure 4.19c peuvent être utilisés pour évaluer la parole en sortie, à partir de la relation entre l'excitation et la réponse qui est représenté par la fonction de transfert $V(Z)$ de la forme :

$$(57) \quad V(Z) = \frac{G}{1 - \sum_{k=1}^N \alpha_k Z^{-k}}, \quad G \text{ et } \alpha_k \text{ dépendent de la fonction d'aire.}$$

(Fonctions d'aire
coefficient de réflexions)

(Paramètres)



"fig 4.23" a) Représentation d'un modèle de tubes sans pertes
b) Modèle analogue terminal

1) Conduit vocal

Un modèle qui contient tous les pôles est une bonne représentation des effets du conduit vocal pour la majorité des sons, cependant la théorie acoustique confirme que les nasals et les fricatifs nécessitent des résonances et anti-résonances (pôles et zéros), dans ces cas on doit inclure les zéros dans la fonction de transfert, ou bien d'après ATAL (35) l'effet de zéros peut être atteint par inclure plusieurs pôles, dans la plupart des cas cette approche est préférable.

Les coefficients du dénominateur de l'équation (57) sont réels, les racines sont aussi réelles ou complexes conjuguées pairs.

$$S_k, S_k^* = -\delta_k \pm j 2\pi F_k$$

La représentation du pôle complexe conjugué en numérique est

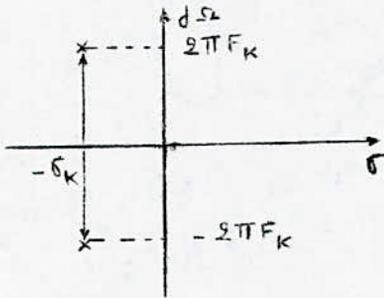
$$\begin{aligned} Z_k, Z_k^* &= e^{-\sigma_k T} e^{\pm j 2\pi F_k T} \\ &= e^{-\sigma_k T} \cos(2\pi F_k T) \pm j e^{-\sigma_k T} \sin(2\pi F_k T) \end{aligned}$$

La bande passante du conduit vocal est approximativement σ_k/π et le centre est de $2\pi F_k$, dans le plan des Z le rayon de l'origine est au pôle détermine la bande passante c'est-à-dire :

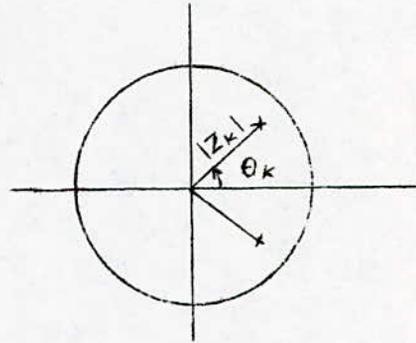
$$|Z_k| = e^{-\sigma_k T}$$

et l'angle est :

$$\theta_k = 2\pi F_k T$$



plan s



plan z

avec $|Z_k| < 1$ qui est la condition de stabilité c'est-à-dire que les pôles sont à l'intérieur du cercle unité.

Remarque :

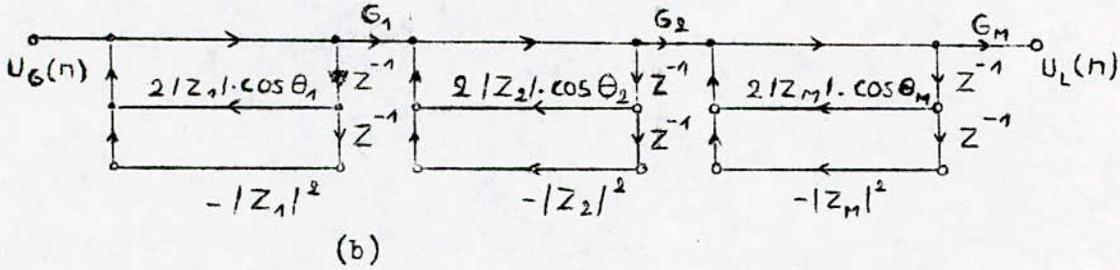
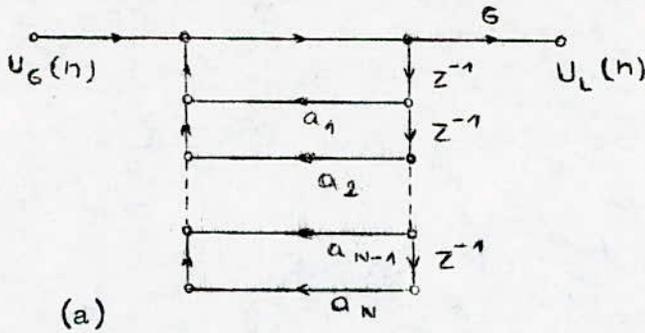
On peut représenter $V(Z)$ par des systèmes en cascade du deuxième ordre (résonateurs) c'est-à-dire :

$$V(Z) = \prod_{k=1}^M V_k(Z), \quad M : \text{nombre entier.}$$

$$V_k(Z) = \frac{1 - 2|Z_k| \cos(2\pi F_k T) + |Z_k|^2}{1 - 2|Z_k| \cos(2\pi F_k T) Z^{-1} + |Z_k|^2 Z^{-2}}$$

Le numérateur de $V_k(Z)$ est choisit de telle façon que le produit aura le même résultat que dans le cas du modèle de tubes sans pertes.

Un modèle en cascade est donné par la "fig 4.24". Une autre approche Une autre approche de mettre en oeuvre le système $V(Z)$ c'est la représentation sous la forme parallèle.



"figure 4.24"

2. RADIATION :

On a considéré que la fonction de transfert $V(Z)$ qui relie le débit au niveau de la glotte à celui des lèvres.

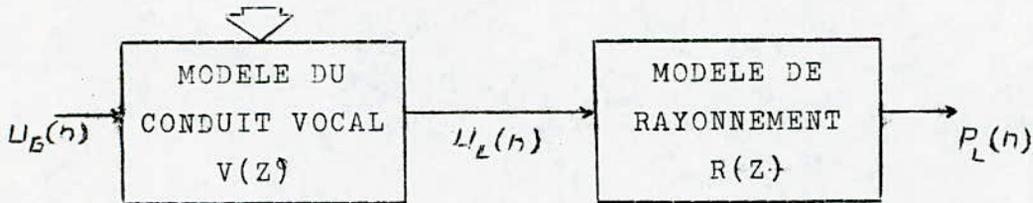
La pression et le débit sont reliés par la relation donné par l'équation (37), la transformé en Z est sous la forme :

A partir de la section 4.2.3.2c et la figure 4.8 la pression est reliée au débit par une opération de filtrage (passé haut), en fait en basse fréquence la pression est approximativement une dérivation du débit. Donc pour obtenir une représentation discrète de cette relation, on doit utiliser la technique de numération qui évite cette approximation, par exemple en utilisant la transformation des filtres numériques qui indique qu'une approximation raisonnable de l'effet de la radiation est obtenue par :

$$R(Z) = R_0(1 - Z^{-1})$$

Cette charge de rayonnement peut être mise en cascade avec le modèle du conduit vocal comme l'indique la "figure 4.25".

PARAMETRES



"Fig 4.25" Terminal analog model including the effects of radiation. $V(Z)$ can be realized by certain convenient cases and the parameters required. For the configuration chosen; functions of areas for a lossless tube model or formants and passbands for the cascade model.

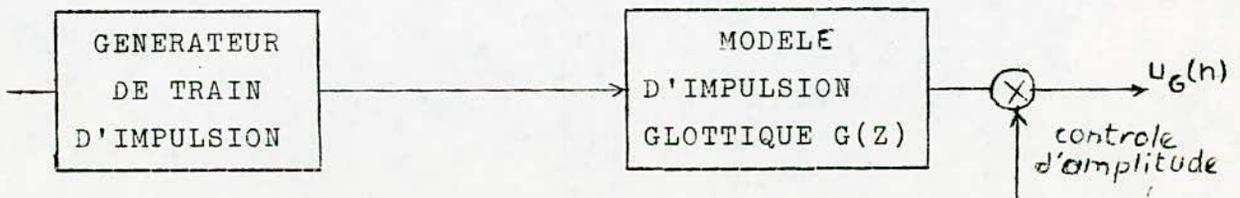
3. EXCITATION :

To complete our study, we must talk about excitation, this latter includes a train of impulse generator which produces a sequence of impulses unit spaced by the period of the fundamental desired.

And of a generator for non voiced sounds which requires a noise source and a gain parameter to control the intensity of the excitation.

A convenient representation to generate glottal waves is indicated by the "figure 4.26":

PERIODE DU FONDAMENTAL



"fig 4.26" Generator of the excitation signal for voiced sounds.

Le signal issu du générateur de train d'impulsions excité un système linéaire, dont la réponse impulsionnelle est $g(n)$, dont la réponse

Au contrôle l'intensité de la voix.

D'après ROSENBERG (39) pour augmenter la qualité de la parole $g(n)$ doit avoir la forme suivante : "fig 4.27".

$$\begin{aligned} g(n) &= \frac{1}{2} \left[1 - \cos\left(\frac{\pi n}{N_1}\right) \right] & 0 \leq n \leq N_1 \\ &= \cos \left[\frac{\pi(n - N_1)}{2N_2} \right] & N_1 \leq n \leq N_1 + N_2 \\ &= 0 & \text{partout ailleurs.} \end{aligned}$$

La forme de cette onde est similaire à celle de la "figure 4.11".

La "figure 4.27" indique la forme de $g(n)$ et sa transformée de fourier. Puisque $g(n)$ est de longueur finie, sa transformée en Z $G(Z)$ à seulement des pôles.

En pratique il est préférable d'étudier un modèle à deux pôles pour $G(Z)$ (J.D MARKEL).

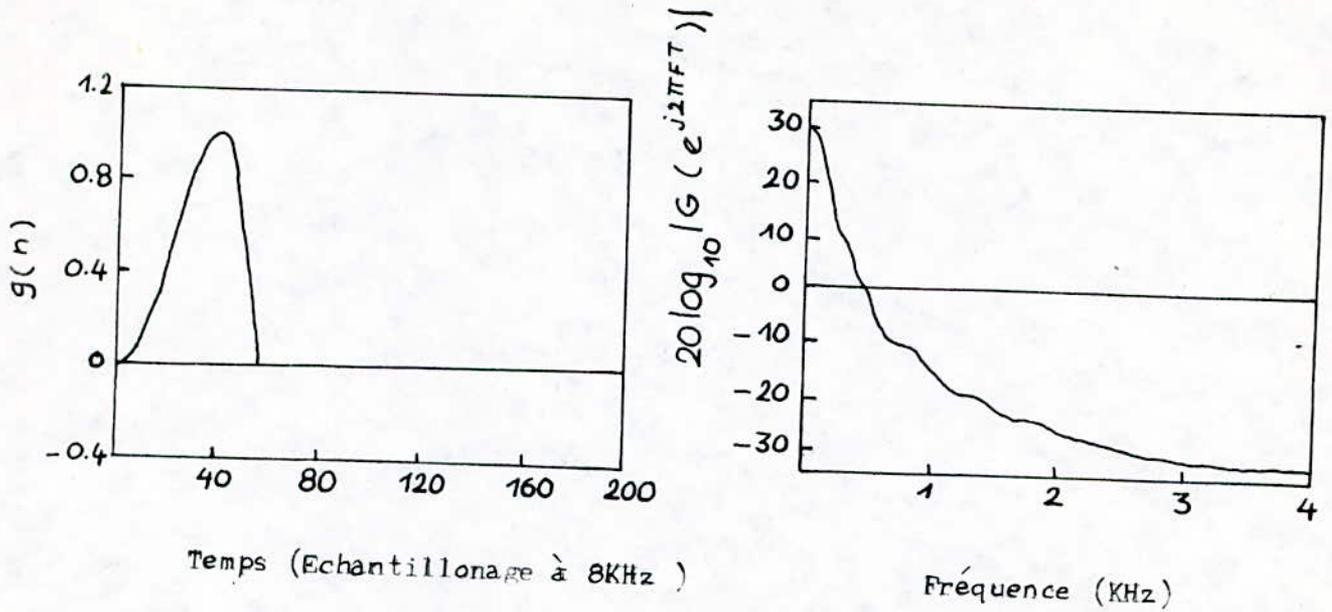
Pour les modèles discret un générateur de nombre aléatoire fournit une source de spectre de bruit plat.

* Dans certains cas il est convenable de combiner le modèle glottique avec les modèles du conduit vocal par un seul système (cas de la synthèse avec prédiction linéaire)

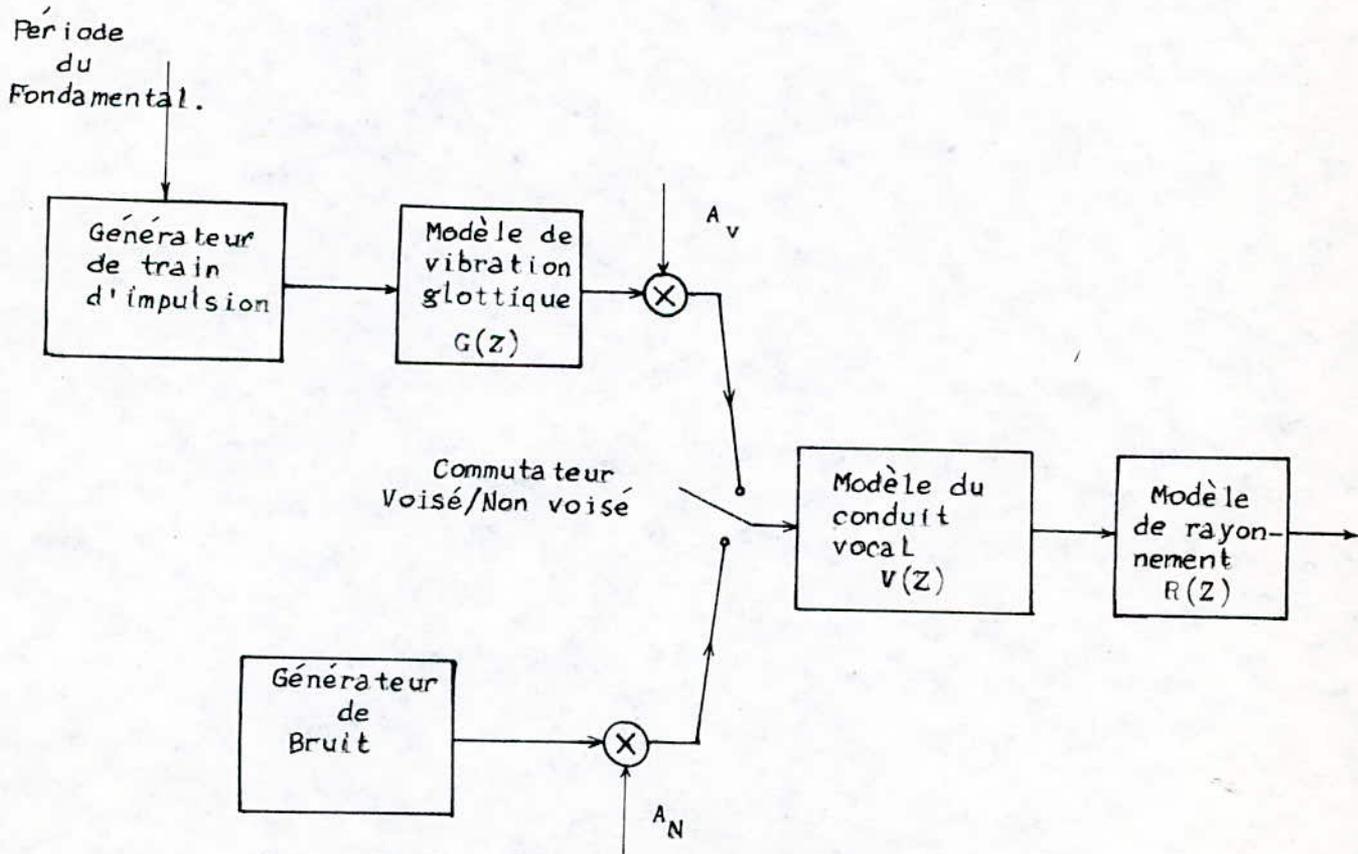
$$H(Z) = G(Z) * V(Z) * R(Z).$$

* On remarque une variation des paramètres dans le temps dans le cas des sons continus par exemple les voyelles, les paramètres changent très lentement et le modèle fonctionne en bonne état, ce qui n'est pas le cas pour les sons transitoires.

(61)



"fig 4.27"



"fig 4.28" Modèle général discret pour la production de la parole. Cette figure indique le modèle complet avec un commutateur entre les sons voisés et non voisés pour permettre le changement du mode d'excitation.

Par ailleurs l'utilisation de la fonction de transfert et la réponse en fréquence suppose qu'on peut représenter le signal de parole en "courte durée", les paramètres de ce modèle sont supposés constants sur des intervalles de temps typique de 10 à 20 ans.

2.10. CONCLUSION :

On peut utiliser les coefficients de réflexions aux jonctions des différentes sections élémentaires (KELLY et LOCKBAUM 1960).

Pour calculer les variations de pression et de débit ainsi que la fonction de transfert du conduit vocal.

Les coefficients de réflexions sont donnés par la formule :

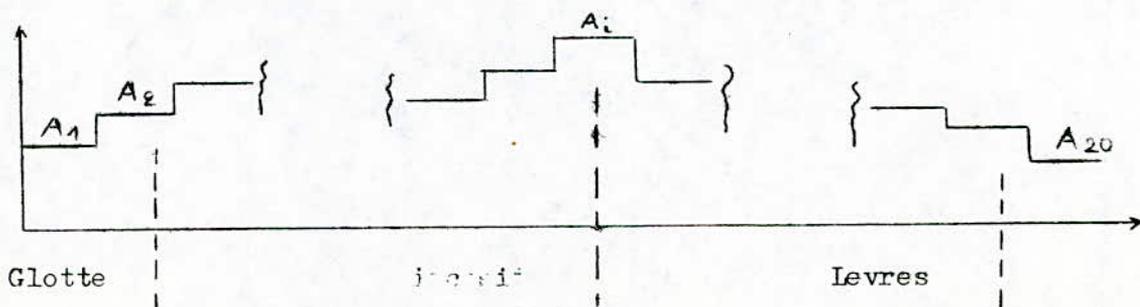
$$r_k = \frac{A_k - A_{k+1}}{A_k + A_{k+1}}$$

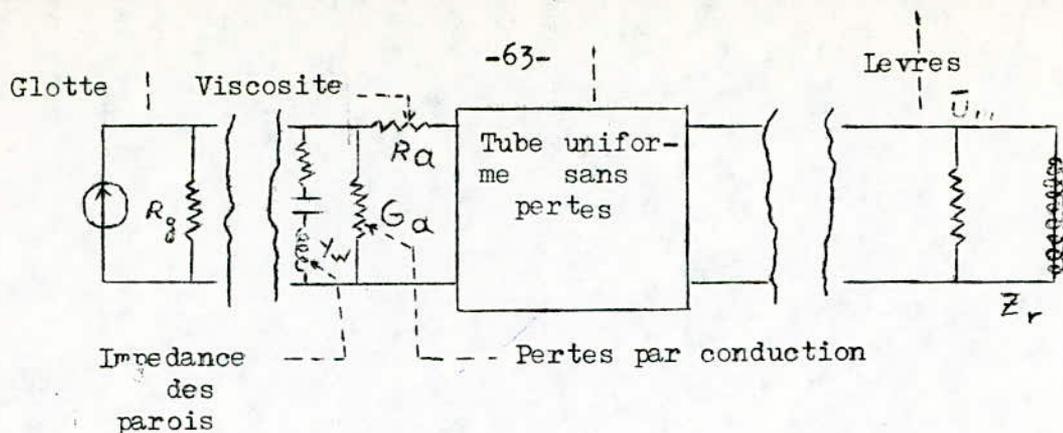
où A_k est l'aire du tube K .

Dans ce modèle, on ne tient pas compte des pertes les seules qui peuvent être introduites sont celles dues à l'impédance de rayonnement.

:

"ATAL et AL" a étudié ce modèle en introduisant des éléments de pertes appropriés ainsi que le montre la "figure 4.29":





"fig. 4.29" modèle avec pertes (d'après ATAL et AL).

Le modèle analogue électrique :

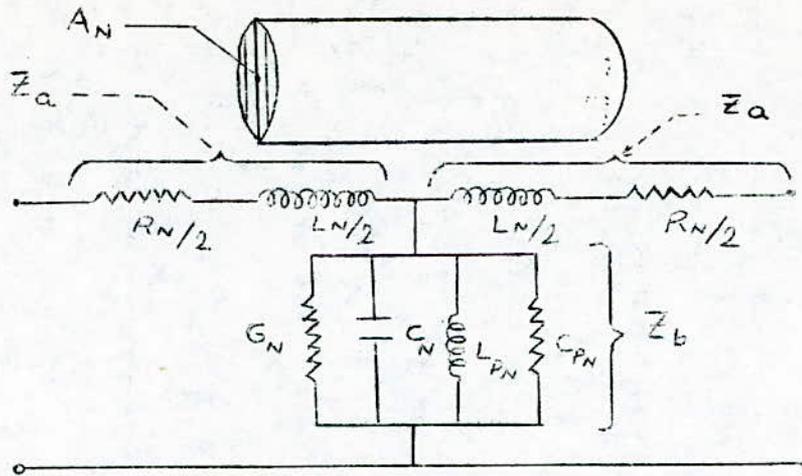
Ici aussi nous utilisons également la discrétisation du conduit vocal en une suite de tubes élémentaires de formes cylindriques et le seul critère qu'il faut respecter, est que la longueur d'un tube soit petite devant la longueur d'onde de la plus haute fréquence traitée.

Par ailleurs la quantification de la fonction d'aire introduit une erreur qui est d'autant plus petite que le nombre de tubes est grand ; on considère que 20 tubes suffisent pour minimiser cette erreur.

Ainsi la simulation du conduit vocal dans sa forme et ses dimensions peut être effectuée au moyen d'une ligne électrique.

A chaque section du conduit est associée une cellule dont les éléments sont accordables, on simule ainsi la variation de la fonction d'aire du système de phonation lorsqu'on se déplace de la glotte aux lèvres. Le conduit nasal dont l'entrée est ici condamné par le voile du palais relevé, pourrait être simulé par un circuit en parallèle sur la ligne.

Dans le cas d'un tube élémentaire, l'analogie acoustique-électrique permet la correspondance suivante "fig.4.30".



"Fig 4.30" - Analogie électrique du tube élémentaire du conduit vocal.

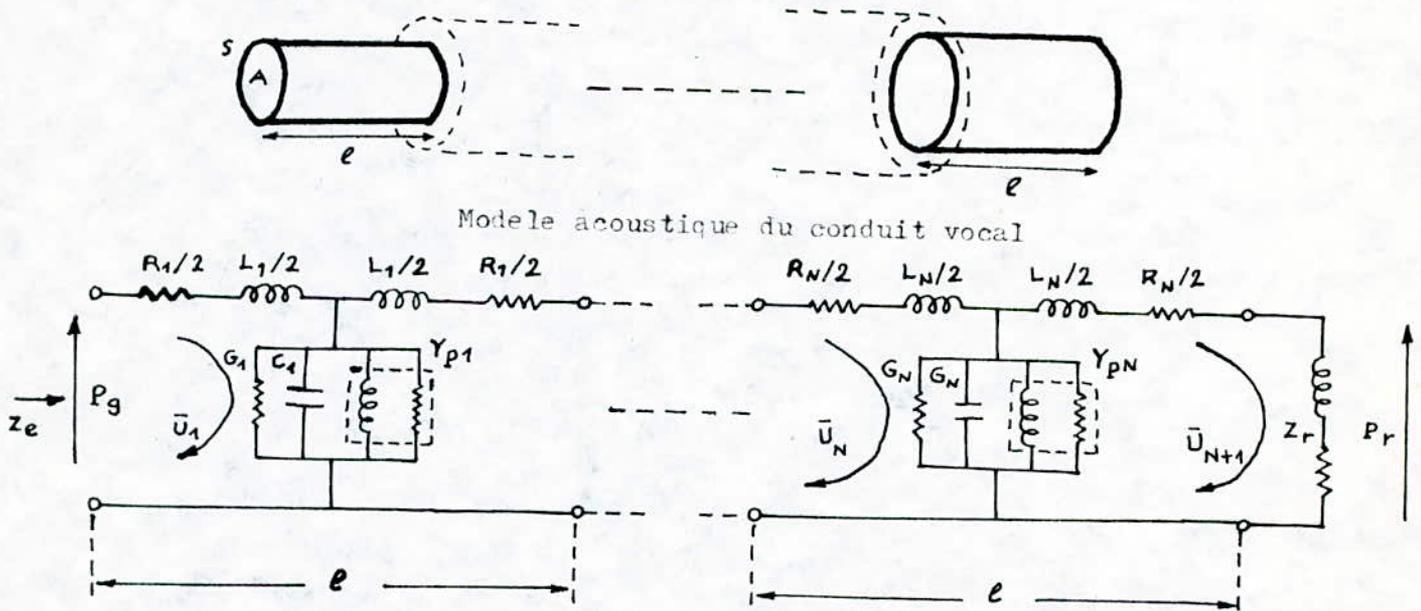
Les équations d'onde dans le tube s'écrivent :

$$\begin{aligned}
 -\frac{\partial p}{\partial x} &= \frac{\rho}{A} \frac{\partial u}{\partial t} & -1- \\
 -\frac{\partial u}{\partial x} &= \frac{A}{\rho c^2} \frac{\partial p}{\partial t}
 \end{aligned}$$

En substituant les quantités acoustiques par leurs analogues électriques dans les équations -1- (tableau des analogues utiles est donné précédemment) on obtient :

$$\begin{aligned}
 -\frac{\partial v}{\partial x} &= L \frac{\partial i}{\partial t} & , \quad Z = j\omega L = j\omega \rho / A \\
 -\frac{\partial i}{\partial x} &= C \frac{\partial v}{\partial t} & , \quad Y = j\omega A / \rho c^2 = j\omega C
 \end{aligned}$$

"Fig 4.31" Modèle acoustique et son analogue électrique complet (MRAYATI 76).



Modèle électrique de simulation du conduit vocal

avec : $R_n = \frac{S_n}{A_n^2} \sqrt{\frac{\rho_0}{2}} e$; $L_n = \frac{\rho_0}{A_n} e$; $C_n = \frac{A_n}{\rho_0 c^2} e$; $G = S_n \frac{n-1}{\rho_0 c^2} \sqrt{\frac{\lambda \omega}{2 c_p \rho_0}} e$;

$Y_p = \frac{b_p S_n}{b_p^2 + (\omega M_p)^2} e - j \frac{M_p S_n}{b_p^2 + (\omega M_p)^2} e$; ($Z_p = b_p + j \omega M_p = 1600 + j \omega 1,4$ par unité de surface)

$Z_r = \frac{\omega^2 \rho_0}{2 \pi c} + j \omega \frac{8 \rho_0}{3 \pi \sqrt{\pi} A_N}$; $n = 1, 2, \dots, N$; $N = \frac{\text{longueur du conduit vocal}}{e}$

APPLICATION - SYNTHETISEUR à FORMANTS

IV-1- Introduction:

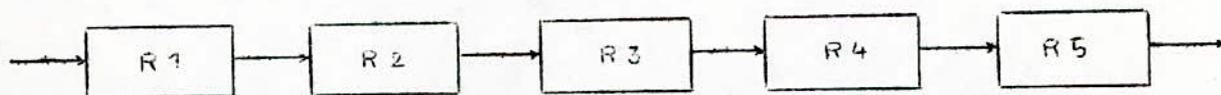
La conception du synthetiseur à Formants, repose sur la réalisation d'un ensemble de circuits dont la courbe de réponse en fréquence se rapproche le plus possible à celle du conduit vocal. Mais cette structure, séduisante à plus d'un titre, ne rend pas compte, en particulier de la nature articulatoire de la parole.

Le synthetiseur à Formants, cascade / parallèle peut être simulé sur un ordinateur numérique à usage général. En utilisant le programme "HANDSY-FOR" vingt paramètres peuvent varier dans le temps, qui servent comme entrée pour les sous programmes "PARCOE -FOR et COEWAV-FOR". A la sortie les échantillons sont calculés en temps différé et stockés sur disquette.

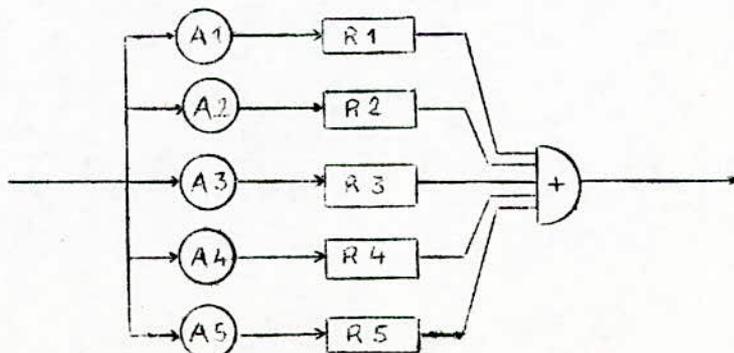
4.2- Structures de synthetiseurs à Formants :

On distingue deux structures de synthetiseurs à Formants.

- * Structure parallèle " fig 4.1a "
- * Structure serie " fig 4.1b "



"fig 4.1b"



"fig 4.1a"

Le principal avantage de la structure serie est de réduire le nombre des paramètres de commande. Cependant la forme parallèle présente l'inconvénient de nécessiter davantage d'informations, on doit spécifier l'amplitude et la fréquence des Formants, ce qui, tous compte faits est un bien, car on contrôle mieux ainsi le synthetiseur.

Pas d'échantillonnage :

Beaucoup d'expériences (test par filtre passe bande) ont montré que l'intelligibilité est inchangeable, si l'énergie pour des fréquences supérieures à 5000 Hz est supprimée. Les sons sont parfaitement naturels, donc on a sélectionné 5000 Hz, ce qui correspond à une fréquence d'échantillonnage de 10.000 échantillons / seconde.

4.3- Les résonateurs:

La base du synthetiseur est un résonateur numérique ayant les propriétés illustrées dans la " figure4.2 ".

Deux paramètres sont utilisés pour spécifier les caractéristiques " entrée-sortie " du résonateur ; la fréquence de résonance F et la bande passante BW.

Les échantillons de sortie $y(nT)$ sont évalués à partir de la séquence d'entrée $x(nT)$ par l'équation aux différences du 2^{ieme} ordre.

$$y(nT) = A x(nT) + B y(nT-T) + C y(nT- 2T) \quad (1)$$

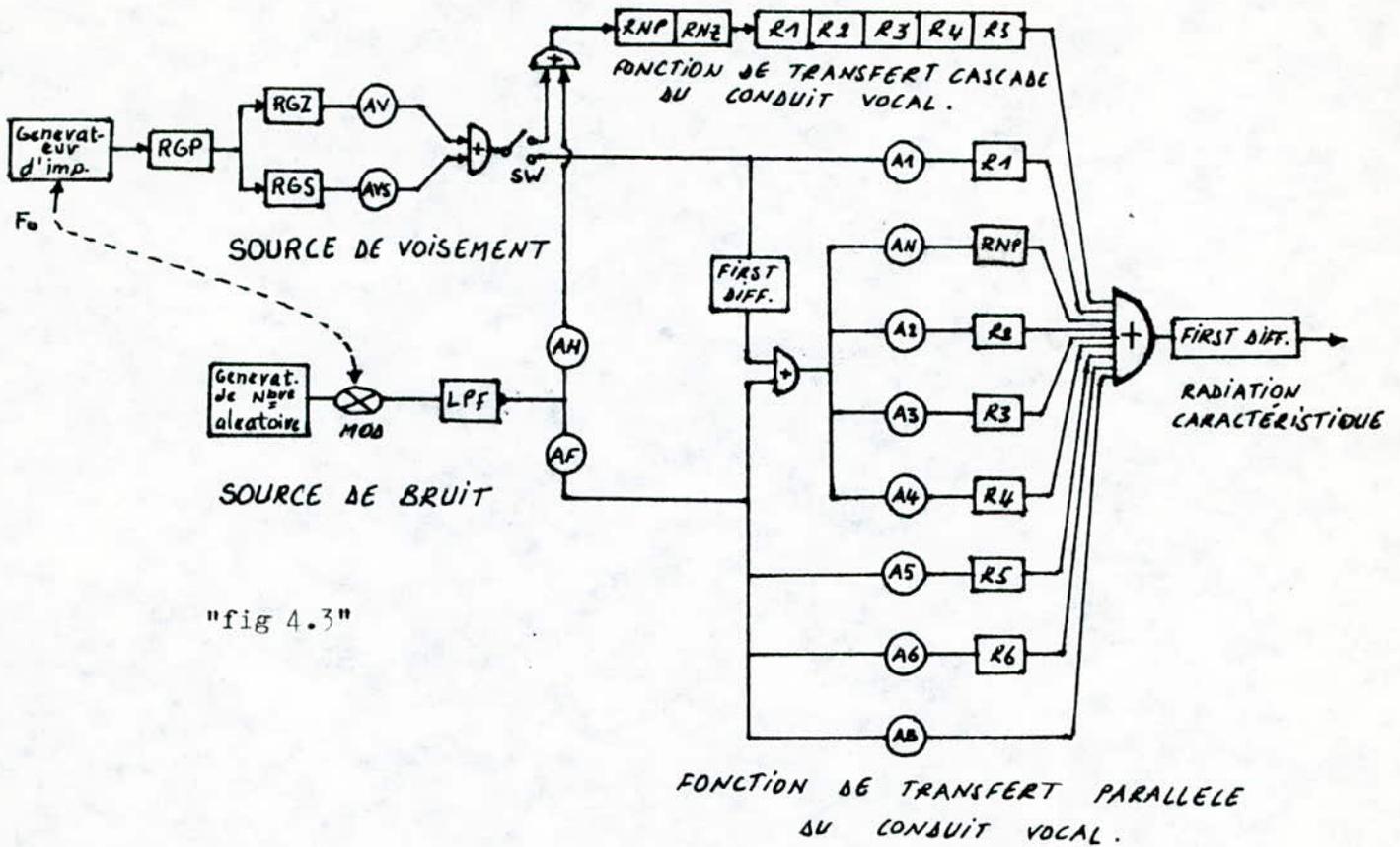
avec $y(nT-T)$ et $y(nT-2T)$ sont les échantillons passés de la séquence $y(nT)$. Les constantes A, B et C sont reliées à la fréquence de résonance et la bande passante d'un résonateur par une transformation impulsionnelle invariante (GOLD et RABINER 1968)

$$C = - \exp(2\pi BWT)$$

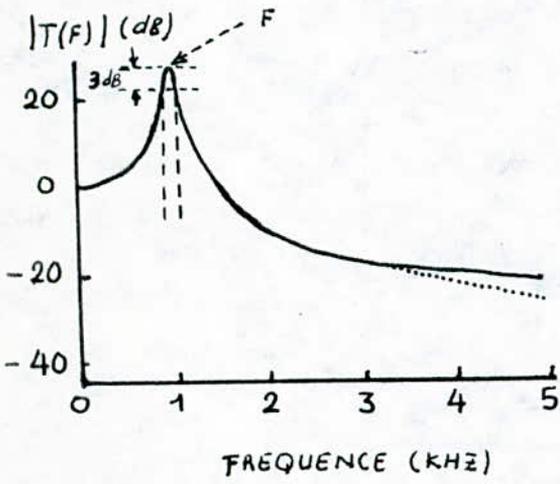
$$B = 2 \exp(-\pi BWT). \cos (2\pi FT)$$

$$A = 1-B^2A$$

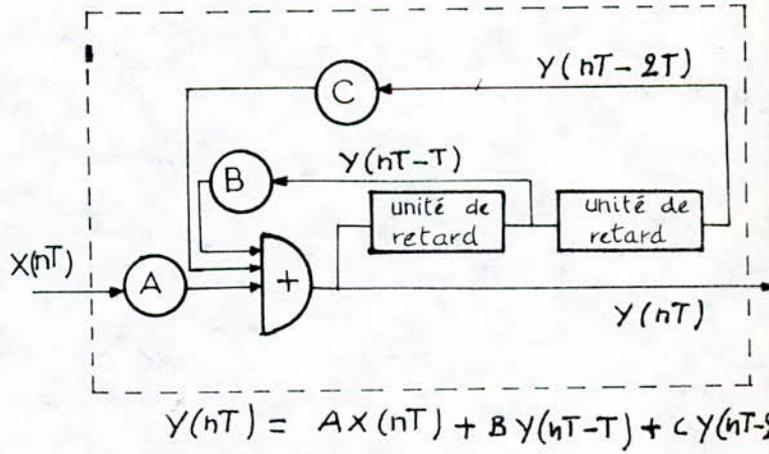
La constante T = 0,0001s étant le pas d'échantillonnage, les valeurs de paramètres de commande F et BW sont injectées de nouveau chaque 5ms, ce qui



"fig 4.3"



(a)



(b)

"fig 4.2"

correspond aux constantes A, B et C.

La fonction de transfert d'un tel résonateur est donnée par :

$$T(f) = \frac{A}{1 - Bz^{-1} - Cz^{-2}}$$

$$T(f) = \frac{1 - 2 \exp(-\pi B \omega T) \cos(2\pi f T) + \exp(-2\pi B \omega T)}{1 - 2 \exp(-\pi B \omega T) \cos(2\pi f T) z^{-1} + \exp(-2\pi B \omega T) z^{-2}}$$

$$|z_k| = e^{-\pi B \omega T} = e^{-\sigma_k T}, \quad B \omega = \sigma_k / \pi$$

$$T(z) = \frac{1 - 2 |z_k| \cos(2\pi f T) + |z_k|^2}{1 - 2 |z_k| \cos(2\pi f T) z^{-1} + |z_k|^2 z^{-2}}$$

* Anti-résonateur numérique :

Un anti-résonateur (anti-Formant) peut être réalisé par une légère modification de l'équation (1). Sa réponse en fréquence est une image de la courbe de réponse donnée en pointillé "fig 4.2a" (on remplace dB par -dB).

Il est utilisé dans le synthétiseur pour donner une forme de source de voisement, pour simuler les effets de nasalité dans le modèle en cascade.

Les séquences de sortie sont reliées à celles d'entrée par l'équation:

$$y(nT) = A' x(nT) + B' x(nT-T) + C' x(nT-2T)$$

$x(nT-T)$ et $x(nT-2T)$ sont les échantillons passés de la séquence $x(nT)$ et les constantes A' , B' et C' sont définies par:

$$A' = 1/A ; B' = -B/A ; C' = C.A$$

* Résonateur passe-bas :

On a pris un résonateur numérique dont la fréquence de résonance est nulle, l'atténuation nominale est de -12dB/octave, et la bande passante est de $BW/2$ (à -3dB).

La source de son voisé contient le résonateur numérique " RGP " , utilisé comme un filtre passe bas qui transforme les impulsions glottiques en impulsions ayant une forme d'onde et un spectre similaire à celle de la voix normale.

4.4- Blocs de synthetiseur :

Ces blocs sont commandés par 39 paramètres, qui déterminent la caractéristique de sortie, parmi ces 39 paramètres, 20 variables et le reste sont des constantes (voir le tableau et ANNEXE).

a- Source de son voisé :

Sa structure est donnée par la "figure 4.3". Elle utilise des paramètres variables, pour spécifier la fréquence du fondamentale " F_0 " et la commande d'amplitude "AV".

Quand " F_0 " est supérieure à zéro , un train d'impulsions est généré. La source est éliminée lorsque "AV" est nulle; pour une voyelle forte "AV" varie aux environs de 60 dB.

Ce train d'impulsions est envoyé à travers un filtre passe bas " RGP " et de bande passante " BGP ". A la sortie de " RGP " on obtient une forme d'onde lisse qui ressemble à la forme d'onde du débit glottique typique (Flanagan 1959). Le spectre d'impulsions filtre devient approximativement lisse (-12dB/octave) à partir de 50 Hz.

La forme d'onde issue de " RGP " n'a pas le spectre de phase identique à celui de l'impulsion glottique typique, et ne contient pas de zéros qui apparaît souvent dans le voisement.

Pour remédier à une solution correcte , on a inséré un anti-résonateur "RGZ" de zéro glottique "FGZ" et de bande passante "BGZ".

La forme d'onde et le spectre d'un voisement sont donnés par la "figure 4.4a" qui sont produits par l'envoi d'un train d'impulsions à travers "RGP et RGZ".

b- Source quasi-sinusoidale :

Le paramètre de commande "AVS" , détermine le degré de voisement généré durant les frications, aspirations et plosions.

Une forme d'onde appropriée du voisement quasi-sinusoidal est obtenue par les résonateurs passe bas " RGP et RGS " .

f- Commande du fondamentale:

Souvent une impulsion glottique passe dans le synthetiseur au temps spécifié par la valeur de paramètre de commande de la fréquence du fondamentale. Si " AV et F₀ " sont à zéro, il n'y aura pas d'impulsion glottique.

g- Commande d'échantillons de bruit:

Un générateur de nombre pseudo-aléatoire est utilisé pour générer les fricatives.

$$x(k+1) = [A \cdot x(k)] \text{ mod } P \quad (*)$$

A et P sont des entiers particuliers.

(*) est periodique de periode P-1.

4.5- Modèle du conduit vocal:

4.5.1- Modèle cascade:

Ignorant pour le moment le pôle nasal du résonateur "RNP" et le zéro nasal de l'anti-résonateur "RNZ". Le modèle en cascade de la "figure 4.3" est constitué de 5 Formants résonateurs.

La fonction de transfert est:

$$T(f) = \frac{A(n)}{1 - B(n)z^{-1} - C(n)z^{-2}}$$

Les constantes A(n), B(n) et C(n) sont déterminées par les valeurs du N^{ieme} fréquence du Formant F(n) et bande passante BW(n) données par les relations precedentes.

a- Fréquence de Formants:

La fréquence des trois premiers Formants varie considerablement avec le changement d'articulation.

Pour un locuteur masculin F₁ varie de 180 a 750Hz, F₂ varie de 600 a 2300 Hz, F₃ varie de 1300 a 3100 Hz.

Les fréquences et bandes passantes du quatrieme et cinquieme Formants ne varient pas beaucoup et peuvent être prises comme constantes, ce qui entraine

La fréquence de commande "FGS" est fixée à zéro et "BGS" à 200Hz qui détermine la fréquence de coupure, au dessus de laquelle les harmoniques sont très atténués. La forme d'onde et l'enveloppe du spectre sont données par la "figure 4.4b".

c- Source de friction:

Dans notre synthétiseur, la source de bruit est simulée par un générateur de nombre pseudo-aléatoire, un modulateur, une amplitude de commande "AF" et un filtre numérique passe bas de -6dB/octave "LPF", indiquée par la "figure 4.3".

On doit estimer la source de débit de bruit, pour déterminer la sortie au niveau des lèvres, le calcul est fastidieux, mais on suppose que la source du débit est proportionnelle à l'intégral de la source de pression. L'intégral est approximé par un filtre numérique passe bas du 1^{er} ordre; "LPF". Les échantillons d'entrée et de sortie respectivement $Y(nT)$ et $X(nT)$ sont reliés par:

$$Y(nT) = X(nT) + Y(nT-T)$$

D- Source d'aspiration:

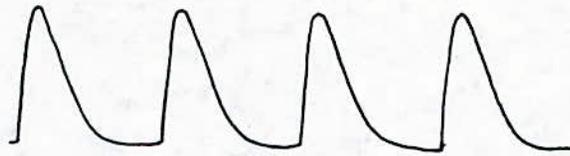
L'aspiration est identique à la friction, la seule différence est que la première est générée au niveau du larynx.

L'aspiration est envoyée à travers la configuration cascade pour les sources de son laryngien. L'amplitude d'aspiration envoyée à travers la configuration cascade est déterminée par "AH"; égale à 60dB pour des aspirations fortes et nulle pour éliminer la source.

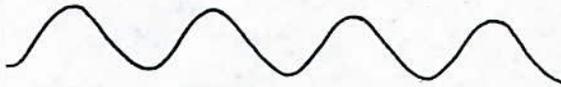
e- Commande d'amplitude de la source:

Les paramètres de commande "AV, AVS, AF et AH" spécifiant la source sont ajustés par l'utilisateur à des nouvelles valeurs chaque 5ms.

Les commandes d'amplitude de bruit "AF et AH" utilisées pour interpoler l'intensité de la source de bruit linéairement dans l'intervalle de temps de 5ms.

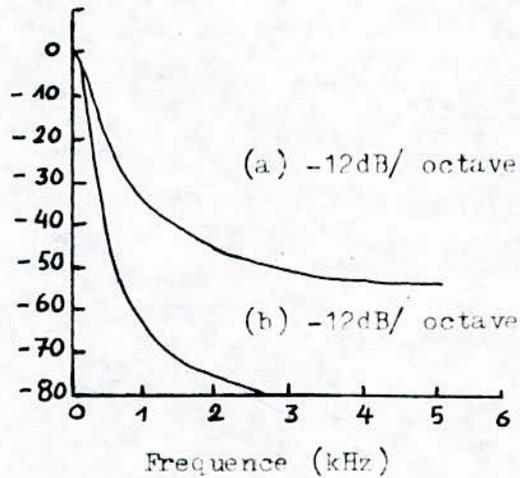


(a) Forme d'onde du voisement normal.



(b) forme d'onde du voisement quasi-sinusoidal.

"fig 4.4"



(c) Spectre de la source de voisement.

"fig 4.4"

une decrementation dans la qualité du son a la sortie du synthetiseur. Ces fréquences élevées nous aident à définir la forme totale du spectre et contribuent peu à l'intelligibilité des voyelles.

b- Bande passante:

Les bandes passantes dépendent des pertes d'énergie dues à la chaleur, viscosité, vibration des parois, rayonnement aux lèvres et la partie réelle d'impédance de la source glottique.

Les bandes passantes sont difficiles à déduire à partir de l'analyse d'une parole à cause de l'irrégularité du spectre de la source glottique.

c- Sons nasales:

On remarque la présence des zéros dans la fonction de transfert pour ce type de sons. Les sons nasales sont approximés par l'insertion d'un résonateur "RNP" et d'un anti-résonateur "RNZ" dans le modèle cascade du conduit vocal.

La fréquence du pôle nasal "FNP" est fixée à une valeur environ 270Hz.

La fréquence du zéro nasal "FNZ" à 270Hz durant les sons non nazals, mais cette dernière doit augmenter durant la production des sons nazals.

4.5.2- MODÈLE Parallèle du conduit vocal:

Durant l'excitation fricative, la fonction de transfert du conduit vocal contient des pôles et des zéros. Les zéros introduisent des crans dans le spectre et modifient les amplitudes des Formants. L'approximation satisfaisante de la fonction de transfert du conduit vocal pour l'excitation fricative est accomplie avec un ensemble de résonateurs numériques placés en parallèle, ayant des commandes d'amplitudes individuelles. Les amplitudes de commande sont faites pour l'excitation des Formants sélectionnés, souvent ceux-ci associés avec la cavité en aval de la constriction (STEVENS 1972).

Cette configuration est constituée de six résonateurs, le sixième est utilisé pour synthétiser la fréquence élevée du bruit dans (S, Z).

On note la présence d'une commande d'amplitude AB, car la fonction de transfert ne contient pas d'importants sommets de résonances (F, V, P, B), et le synthétiseur doit inclure un nombre de résonateurs passe tout pour produire une fonction de transfert plate.

* Caractéristique de la radiation

La radiation aux lèvres transforme l'onde de vitesse volumique $U_1(t)$ en une onde de pression $S(t)$; elle est modélisée par un intégrateur (pente asymptotique +6dB/octave).

$$L(Z) = R_0(1-Z^{-1}) \quad , \quad R_0 = \text{Cte}$$

On a utilisé un anti-résonateur du 1^{er} ordre dont l'équation aux différences est:

$$y(nT) = A x(nT) + B x(nT-T)$$

IV.6- Conclusion:

On a décrit un synthétiseur à Formants à software flexible, qui peut être exécuté sur quelques ordinateurs de laboratoire ; parmi lesquelles on a utilisé le VAX 11/750.

Le synthétiseur et les commandes associées peuvent être utilisés comme un outil de recherche.

Le synthétiseur peut aussi trouver une application avec programme pour la production de la parole (KLATT 1976).

Les expériences montrent que le synthétiseur est suffisamment flexible pour générer une bonne imitation de la voix humaine.

* CONCLUSION GENERALE *

GRACE AU PROGRES DE L'INFORMATIQUE ET DE L'ELECTRONIQUE, NOUS ASSISTONS ACTUELLEMENT AU DEVELOPPEMENT DE TRES NOMBREUX SYSTEMES UTILISANT LA PAROLE.

CE TRAVAIL NOUS A PERMIS D'ELARGIR NOS CONNAISSANCE DANS LE DOMAINE DU TRAITEMENT DU SIGNAL DE LA PAROLE QUI EST ACTUELLEMENT UN NOUVEAU FOCYER DE RECHERCHES.

CETTE ETUDE MET EN EVIDENCE LE ROLE DE LA THEORIE ACOUSTIQUE APPLIQUEE SUR LE MODELE ARTICULATOIRE POUR LA PRODUCTION DE LA PAROLE.

CE PROCESSUS TIENT COMPTE DES VARIATIONS DE PRESSION ET DE DEBIT A TRAVERS UNE CONCATENATION DE TUBES REPRESENTANT LE CONDUIT VOCAL.

LA SIMULATION DONNE DE MEILLEURS RESULTATS CONCERNANT L'INTELLIGIBILITE ET LA QUALITE DE LA REPRODUCTION.

CELLE-CI PAR CE MODELE PERMET D'OBSERVER LES RECHERCHES AVANCEES DANS LA THEORIE ACOUSTIQUE ET SA VALIDITE EN DETERMINANT SES PARAMETRES PERTINENTS .

LE MODELE PEUT ETRE UTILISE DANS DIFFERENTS DOMAINES:

- * LA RECONSTITUTION DE LA PAROLE .
- * L'AIDE AUX HANDICAPES ET LA REEDUCATION QUI SONT PARMIS LES APPLICATIONS LES PLUS UTILES.
- * LA COMMUNICATION AVEC LES AGENCES AERIENNE ET METEOROLOGIQUE...etc

*** ANNEXE "A" ***

* TRANSFORMEE EN "Z" *

SEQUENCE X(n)	TRANSFORMEE EN "Z"
-1- $\delta(n) = \begin{cases} 1 & , n=0 \\ 0 & , n \neq 0 \end{cases}$	1
-2- $\delta(n-k) = \begin{cases} 1 & , n=k \\ 0 & , n \neq k \end{cases}$	z^{-k}
-3- $U(n) = \begin{cases} 1 & , n \geq 0 \\ 0 & , n < 0 \end{cases}$	$\frac{z}{z-1}$
-4- $a^n \cdot U(n)$	$\frac{z}{z-a}$
-5- $x(n-k)$	$X(z) \cdot z^{-k}$
-6- $x(n) \cdot a^n$	$X(a \cdot z)$
-7- $x(n+k)$	$X(z) \cdot z^k - \sum_{j=0}^{k-1} x(j) \cdot z^{k-j}$

*** ANNEXE "B" ***

* TRANSFORMEE DE FOURIER NORMAL "T.F.N" *

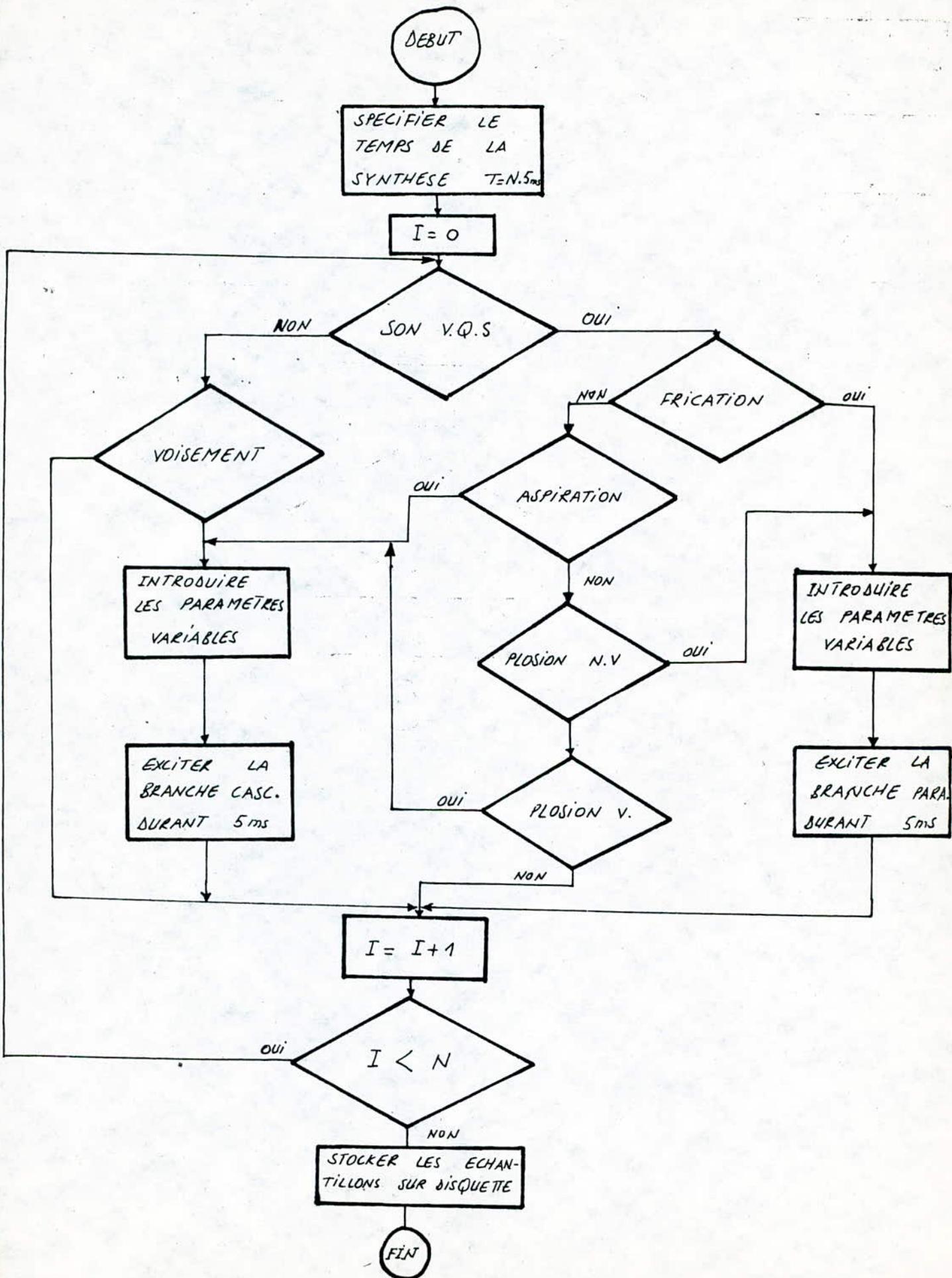
2- ORIGINAL	T.F.N
-1- $f(t), g(t)$	$F(w), G(w)$
-2- $f(t) + g(t)$	$F(w) + G(w)$
-3- $f'(t)$	$jwF(w)$
-4- $f(t - t_0)$	$F(w) \cdot \exp(-jt_0 w)$
-5- $\int_{-\infty}^t f(\tau) d\tau$	$\frac{1}{jw} \cdot F(w)$
-6- $f(t) \cdot \exp(jw_0 t)$	$F(w - w_0)$
-7- $f(a \cdot t)$	$\frac{1}{ a } \cdot F\left(\frac{w}{a}\right)$
-8- $f(t) * g(t) = \int_{-\infty}^{+\infty} f(\tau) \cdot g(t - \tau) d\tau$	$F(w) \cdot G(w)$
-9- $f(t) \cdot g(t)$	$\frac{1}{2\pi} \int F(y) \cdot G(w - y) dy$
-10- $F(t)$	$2\pi \cdot f(-w)$

$$-11- \int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |F(w)|^2 dw$$

*** ANNEXE "C" ***

* TABLEAU DE PARAMETRES DE COMMANDE *

N	V/C	SYM.	NOM	Min	Max	Typ
1	V	AV	Amplitude de voisement N. (dB)	0	80	0
2	V	AF	Amplitude de friction (dB)	0	80	0
3	V	AH	Amplitude d'aspiration (dB)	0	80	0
4	V	AVS	Amplitude de voisement Q.S. (dB)	0	80	0
5	V	F0	Frequence du Fondamental (Hz)	0	500	0
6	V	F1	Frequence du 1 Formant (Hz)	150	900	450
7	V	F2	Frequence du 2 Formant (Hz)	500	2500	1450
8	V	F3	Frequence du 3 Formant (Hz)	1300	3500	2450
9	V	F4	Frequence du 4 Formant (Hz)	2500	4500	3300
10	V	FNZ	Zero nasal (Hz)	200	700	250
11	C	AN	Amplitude du formant nasal (dB)	0	80	0
12	C	A1	Amplitude du 1 Formant (dB)	0	80	0
13	V	A2	Amplitude du 2 Formant (dB)	0	80	0
14	V	A3	Amplitude du 3 Formant (dB)	0	80	0
15	V	A4	Amplitude du 4 Formant (dB)	0	80	0
16	V	A5	Amplitude du 5 Formant (dB)	0	80	0
17	V	A6	Amplitude du 6 Formant (dB)	0	80	0
18	V	AB	Amplitude (dB)	0	80	0
19	V	B1	B.P. du 1 Formant (Hz)	40	500	50
20	V	B2	B.P. du 2 Formant (Hz)	40	500	70
21	V	B3	B.P. du 3 Formant (Hz)	40	500	110
22	V	B4	B.P. du 4 Formant (Hz)	100	500	250
23	C	FGP	Frequence du resonateur glott. (Hz)	0	600	0
24	C	BGP	B.P. du resonateur glott. (Hz)	100	2000	100
25	C	FGZ	Zero glottique (Hz)	0	5000	1500
26	C	BGZ	B.P. du zero glott. (Hz)	100	9000	6000
27	c	SW	Commutateur cascade/parallele	0(CASC)	1(PARA)	0
28	V	F5	Frequence du 5 Formant (Hz)	3500	4900	3750
29	C	B5	B.P. du 5 Formant (Hz)	150	700	200
30	C	F6	Frequence du 6 Formant (Hz)	4000	4999	4900
31	C	B6	B.P. du 6 Formant (Hz)	200	2000	1000
32	C	FNP	Frequence du pole nasal (Hz)	200	500	250
33	C	BNP	B.P. du pole nasal (Hz)	50	500	100
34	C	BNZ	B.P. du zero nasal (Hz)	50	500	100
35	C	BGS	B.P. du resonateur glott. (Hz)	100	1000	200
36	C	SR	Pas d'echantillonnage	5000	20000	10000
37	C	NWS	Nbre. d'echantillons par interval	1	200	50
38	C	GO	Commande de gain total (dB)	0	80	47
39	c	NFC	Nombre de Formants cascade	4	6	5



DEBUT

SPECIFIER LE TEMPS DE LA SYNTHESE T=N.5ms

I = 0

SON V.Q.S

VOISEMENT

INTRODUIRE LES PARAMETRES VARIABLES

EXCITER LA BRANCHE CASC. DURANT 5ms

I = I + 1

I < N

STOCKER LES ECHANTILLONS SUR DISQUETTE

FIN

FRICATION

ASPIRATION

PLOSION N.V

PLOSION V.

INTRODUIRE LES PARAMETRES VARIABLES

EXCITER LA BRANCHE PARA. DURANT 5ms

*** BIBLIOGRAPHIE ***

- 1- L.R. RABINER - R.W SCHAFFER
Digital processing of speech signals. PRENTICE HALL (1980).
- 2- J.D. MARKEL - A.H. GRAY JR.
Linear prediction of speech of communication and cybernetics (1976).
- 3- J. LIFERMANN
Les methodes rapides de transformation du signal -MASSON ,1980
- 4- M. KUNT
Traitement numerique des signaux -DUNOD , ' 1981
- 5- J. MAX
Methodes et techniques de traitement du signal-MASSON,1981-VOL 1 th.
- 6- J. MAX
Methodes et techniques de traitement du signal-MASSON,1981-VOL 2 app
- 7- M. BELANGER
Traitement numerique du signal -MASSON,1981.
- 8- J. AUVRAY
Electronique des signaux echantillonnes et numeriques-DUNOD,1976.
- 9- M. FERRETI - F. CINARE
Synthese et reconnaissance de la parole- EDIT-tests ,1983.
- 10- Revue Algerienne de linguistique AL- LISSANIYAT-VOL 1 N⁰ 2,1971 -VOL 2
N⁰1,1972.
- 11- I. EL MALAWANY
Synthese et evaluation comparative de melographes numeriques- Note
technique CNET ,NT / DAS / SST / 10- AOUT 1978.
- 12- R. DESCOUT
Les techniques de synthese de la parole -AVRIL 1982.
- 13- D. FROSSARD
Influence de l'excitation sur la qualite de la parole dans une chaine
d'analyse, synthese par prediction lineaire -CNET - JUILLET 1982.
- 14- The journal of the acoustical society of AMERICA.
VOL .67 ,N⁰ 3 , MARS 1980.