الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم و البحث العلمي
MINISTERE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT : ELECTRONIQUE

PROJET DE FIN D'ETUDES

SUJET

LES TECHNIQUES DE SYNTHESE DE LA PAROLE

Proposé Par : Melle M.Guerti

Etudié par : A.Mezaoui, A.Kechid

Dirigé par : Melle M.Guerti

PROMOTION : Janvier-1987

E.N.P. 10, Avenue Hacen Badi – EL-HARRACH – ALGER

Dédicaces

A mon père,

A ma mère,

A toute ma famille,

A tous mes amis et particulièrement DJEMAA et MOBAREK.

A. MECHID.

A mon père,

A ma mère,

A mes frères et mes soeurs,

A tous mes amis.

A. MEZAOUI.

REMERCIEMENTS

Nous tenons à exprimer nos remerciements à notre promoteur, Mademoiselle M. GUERTI, pour son aide efficace, et ses précieux conseils tout au long de l'élaboration de ce mémoire.

Nous adressons également nos remerciements à Mademoiselle A. MOUSSAOUI pour son aide bibliographique.

Que Monsieur et Madame SAIDI trouvent ici le témoignage de notre reconnaissance pour leur contribution à la réalisation de ce mémoire.

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

———oOo———

TABLE DES MATIERES

Pages

INTRODUCTION GENERALE I

CHAPITRE I : PHONETIQUE

I.I. L'appareil phonatoire humain 4

I.I.I. Description 4

I.I.2. Fonctionnement et production des différents sons 4

I.2. Transitions phonétiques et formants 7

I.2.I. Définition 7

I.2.2. Transitions phonétiques 7

Conclusion

CHAPITRE 2 : NOTIONS DE TRAITEMENT DU SIGNAL

2.I. Introduction I3

2.2. Analyse de Fourier I3

2.2.I. Introduction I3

2.2.2. Série de Fourier..... I3

2.2.3. Transformée de Fourier I3

2.3. Autocorrélation I4

2.4. Echantillonnage et codage I4

2.4.I. Echantillonnage I5

2.4.2. Codage I5

2.5. Transformée de Fourier discrète I6

2.5.I. Définition I6

2.5.2. Transformée de Fourier rapide I6

2.6. Fenêtrage I7

2.7. Filtres numériques I7

Conclusion

CHAPITRE 3 : L'ANALYSE DU SIGNAL DE LA PAROLE

3.I. Introduction 22
3.2. Analyse spectrale 22
3.2.I. Transformée de Fourier 22
3.2.2. Analyse par filtres 22
3.3. Analyse temporelle 23
3.3.I. Méthode d'autocorrélation 23
3.3.2. Passage par zéro du signal 25
3.3.3. Codage linéaire prédictif 25
3.4. Détection de la fréquence fondamentale 28
3.4.I. Méthode d'intercorrélation avec une fonction peigne... 29
3.4.2. Méthode du cepstre 29
3.4.3. Méthode des harmoniques 29
Conclusion

CHAPITRE 4 : LA SYNTHESE DE LA PAROLE & SES METHODES

4.I. Introduction 32
4.2. Techniques de synthèse 32
4.2.I. Vocodeur à canaux 32
4.2.2. Vocodeur à formants 33
4.2.3. Synthétiseur à formants "parallèle" 34
4.2.4. Synthèse prédictive 35
4.2.5. Synthèse par simulation du conduit vocal 35
4.3. Les méthodes de synthèse 37
4.3.I. Synthèse par phrases 37
4.3.2. Synthèse par mots 37

4.3.3. Synthèse par règles 38

4.3.4. Synthèse par diphones 38

Conclusion

CONCLUSION GENERALE 49

ANNEXE 51

BIBLIOGRAPHIE 56

INTRODUCTION GENERALE

Depuis des millénaires, les hommes d'une même tribu, les peuples entre-eux, communiquent, s'entendent et échangent les idées et les pensées grâce à certains codes de communication; parmi les plus efficaces, reste la parole.

A un certain moment de la vie de l'humanité, la machine a fait son apparition dans le but de nous dispenser de certaines tâches. Elle accomplira, docilement, ses fonctions si on appuyait sur la touche correspondante; mais est-ce que la meilleure façon de la "mettre en marche" ne serait pas encore la parole, et si c'était le cas : comment le faire ?

A vrai dire la question ne s'était pas posée d'une façon aussi directe et simple, mais ce sont les exigences des hommes qui ont formulé ce désir. On a dû attendre l' apparition des ordinateurs. Possédant un moyen de calcul aussi puissant, et maîtrisant des mathématiques de haut niveau, les les chercheurs ont développé un domaine nouveau, celui du traitement du signal. Une grande variété d'applications peut d'ores et déjà en bénéficier comme le traitement automatique de la parole. Celui-ci recouvre les domaines suivants:

— La reconnaissance automatique et la compréhension de la parole qui aboutissent à des machines qui "entendent" et qui "comprenent" la parole. Dans ce même thème, on peut parler aussi de l'identification du locuteur (appliquée comme signature vocale dans les banques...)

— La synthèse automatique de la parole qui conduit à des machines qui "parlent".

Avant l'apprition des ordinateurs, certains chercheurs (Kempelen, l'abbé Nical, Kratzenstein vers la fin du XVIII^{e} siècle) ont déjà produit la parole, en imitant l'appareil phonatoire humain, à l'aide de dispositifs encombrants

où l'homme intervenait manuellement. ces premiers synthétiseurs, de la parole, vont voir leur fonction limitée à rejouer un certain morceau de chant. De telles machines sont appelées, un jour ou un autre, à céder leur places à d'autres.

Effectivement des machines modernes, produisant des sons de meilleure qualité et plus efficaces, ont vu le jour à une époque trés récente. Comprendre leurs principes, faire une étude détaillée pour chaque technique utilisée, va être l'objectif que nous essaierons d'atteindre à travers ce projet, où nous aborderons aussi les méthodes de synthèse. Parmi ces dernières, on a vu :

— La synthèse par phrases

— La synthèse par mots

— La synthèse par règles

— La synthèse par diphones

Pour englober toutes ces notions, la présente étude comporte quatre chapitres:

Le premier présente les différents organes qui contribuent à la formation des différents sons et bruits.

Le deuxième chapitre expose certaines notions, sur le traitement du signal, nécessaires à analyser le signal de la parole.

Le troisième décrit les différentes téchniques d'analyse, pour ne prélever que les informations importantes. La plus grande importance sera donnée à la prédiction linéaire.

Le dernier chapitre sera consacré aux différentes techniques de synthèse et les méthodes de concaténation des éléments phonétiques pour produire de la parole artificielle.

----oOo---O---oOo----

CHAPITRE 1

P H O N E T I Q U E

—oOo—

Nous étudierons dans ce chapitre l'anatomie des organes contribuant à la production de la parole, et leur fonctionnement. Nous étudierons brièvement la production des différents sons ainsi que leurs propriétés spectrales.

1 - 1 L'appareil phonatoire humain :

1 - 1 - 1 Description (fig. 1.1 a)

a - Les poumons :

Mis à part leur fonction de respiration, ils génèrent l'air nécessaire à la production des sons par expiration, produisant le flux qui sera conduit par la trachée-artère.

b- Le larynx :

Est une cavité formée de cartilages. A sa base sont attachées les cordes vocales qui vibrent lors de la production des sons voisés.

c - Le conduit vocal :

Il se compose de deux parties :

- Le conduit buccal qui est formé du pharynx et de la cavité buccale. Ce conduit possède une géométrie variable due à la mobilité de la langue et du maxiliaire.
- Le conduit nasal qui comprend deux fosses communiquant avec la cavité buccale à l'aide du voile du palais "velum".

1 - 1 - 2. Fonctionnement et production des différents sons :

1. Fonctionnement :

On peut décrire le système vocal humain comme suit :
Un générateur (poumons) alimente la source d'excitation (cordes vocales) qui produit des signaux acoutisques devant être filtrés par le résonnateur (conduit vocal) (fig. 1.1.b)

2. Production des sons :

Le système vocal humain produit des sons :

- voisés :

Qui sont dus à une excitation pseudo-périodique où les cordes vocales vibrent sous l'action de la pression du flux d'air (les voyelles par exemple). La fréquence de vibration des cordes vocales est appelée fréquence fondamentale, ou fréquence de "mélodies", ou encore "pitch"(de l'anglais).

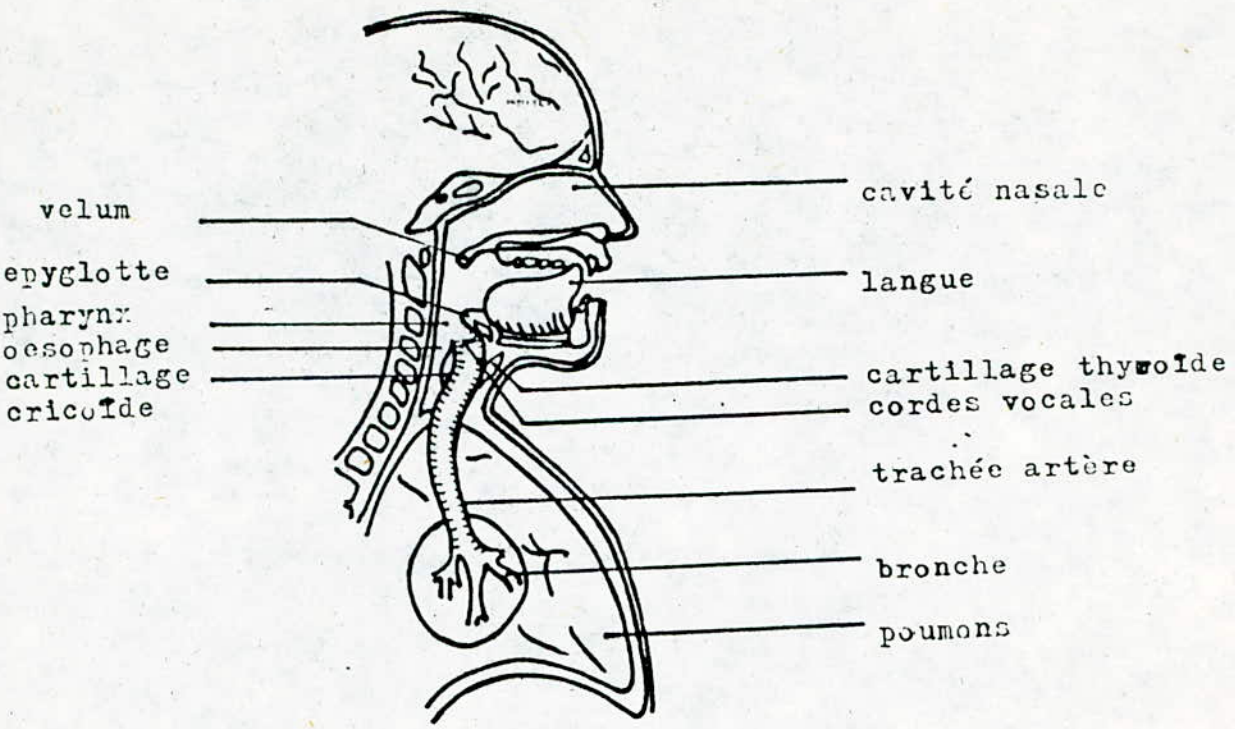


Fig. I. I.a : Appareil phonatoire humain

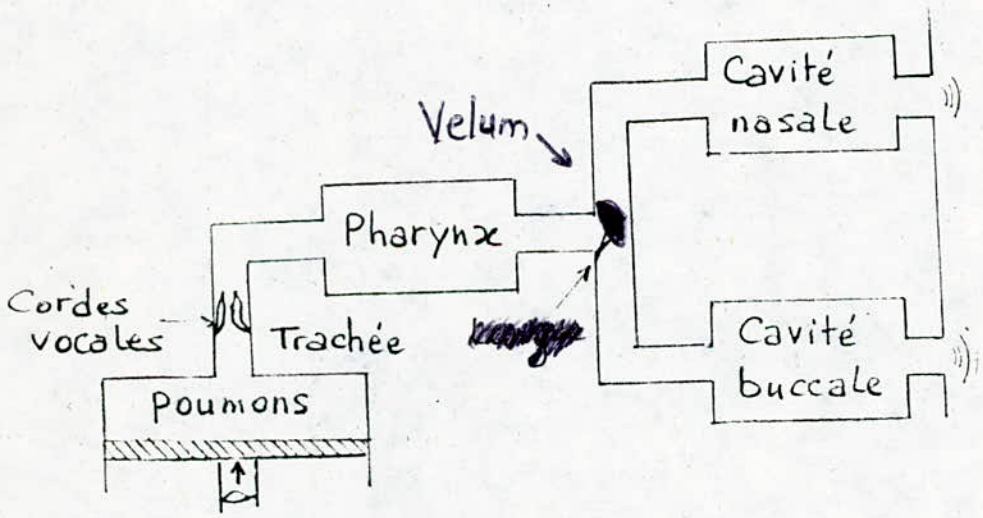


Fig. I. I.b : Schématisation de l'appareil phonatoire humain

- non-voisés :

Qui sont dus à un passage d'un flux d'air turbulent agissant sur la cavité buccale qui pertube son passage (fermeture partielle de la bouche, contact de la langue avec les dents,..) sans cependant que les cordes vocales entrent en jeu.

Ces sons voisés et non voisés sont appelés phonétiquement voyelles et consonnes.

a - les voyelles :

Elles sont caractérisées par le passage libre de l'air. La source d'excitation étant la vibration des cordes vocales.

On distingue :

- Les voyelles orales :

Qui sont dues à un passage de l'air par la cavité buccale seulement (telles que lal, lel, lul,..).

- Les voyelles nasales :

Qui sont obtenues en ouvrant le voile du palais laissant ainsi la communication de la cavité nasale libre (telles que lâl, lêl,..).

b - Les consonnes :

Elles sont produites lors d'une constriction ou fermeture du passage de l'air.

On distingue :

- Les consonnes fricatives :

Elles sont générées pendant le passage de l'air à travers le conduit vocal presque totalement occulté soit par la langue (pour /s/ par exemple) soit par les lèvres (pour /f/ par exemple). Ces consonnes possèdent leurs correspondantes voisées (telles /z/, /v/).

- Les consonnes plosives :

Lors de la fermeture en un point particulier du conduit vocal, la pression augmente pour chuter au moment de l'ouverture produisant ainsi la consonne. Parmis ces plosives, on distingue les non-voisées telles /p/ /t/, /k/, et leurs correspondantes voisées /b/, /d/, /g/.

- Les consonnes nasales (/m/ et /n/) :

Elles sont obtenues par une fermeture partielle à l'avant du conduit buccal avec l'abaissement du voile du palais faisant du conduit nasal le canal de transmission.

- Les semi-voyelles et consonne liquide :

Ce sont des transitions rapides du conduit vocal, ce qui les apparente aux consonnes, mais celui-ci continue à fonctionner en mode résonnant, ce qui les apparente aux voyelles. Ce sont par exemple /w/, /y/, /j/ et /l/.

Un tableau regroupe les voyelles et les consonnes (tableau 1.1) leur mode d'articulation est représentée sur la figure (1.4).

1 - 2 Transitions phonétiques et formants :

1 - 2 - 1. Définition :

Chaque voyelle a ses formants caractéristiques. C'est la concentration des harmoniques renforcés dans certaines bandes de fréquences particulières à chaque voyelle (fig.1.2).

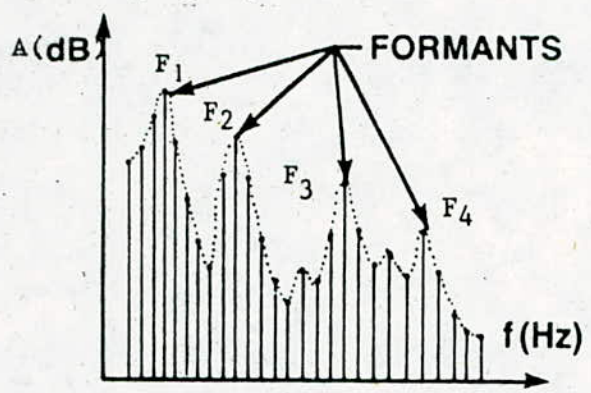


Fig. 1. 2 : spectre vocalique présentant quatre résonnances formantiques.

1 - 2 - 2. Transitions phonétiques :

1. Enchainement des voyelles successives :

La transition d'une voyelle à une autre correspond à un cheminement entre les deux points représentatifs (point d'articulation, ouverture et avancée des lèvres). L'observation montre que l'enchainement respecte un principe de continuité et de simplicité. Les formants "Fn" de rang n subissent l'évolution la plus continue d'après les spectrogrammes (voir chapitre 3) tels que la figure (1. 3).

2. Enchainement des voyelles avec les semi-voyelles et consonnes liquides :

La transition entre les deux positions vocaliques ne prend pas le chemin le plus direct entre les deux positions vocaliques (fig. 1. 3).

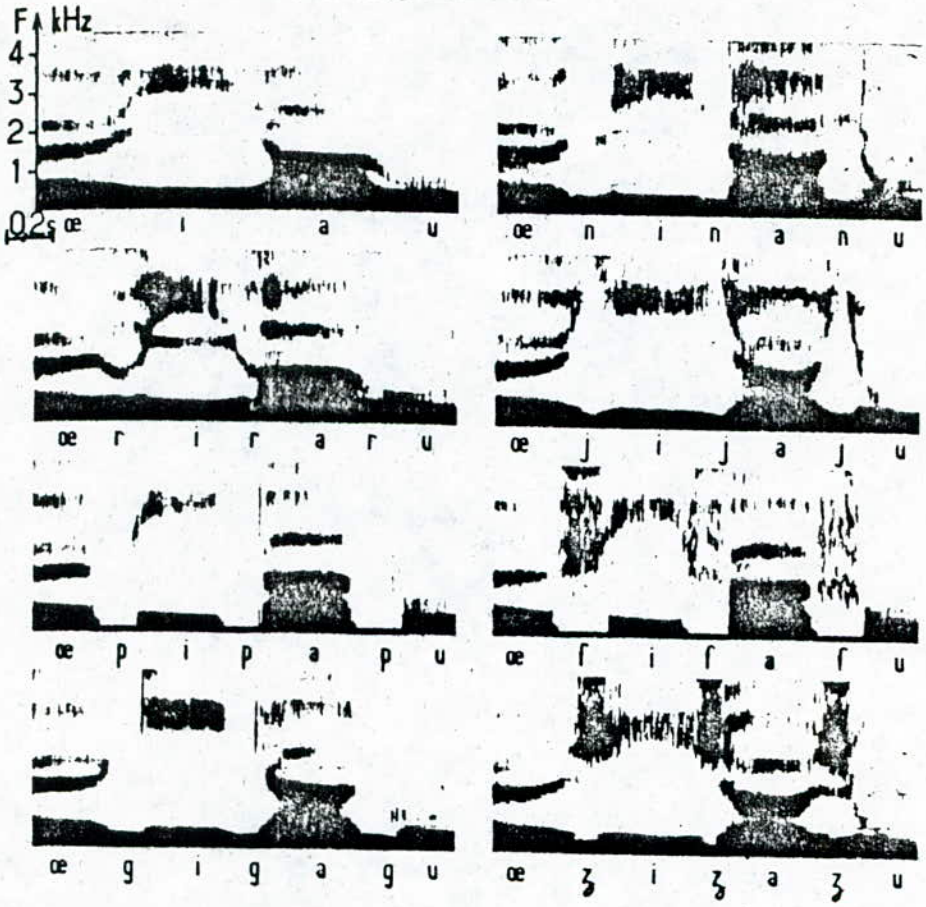


Fig.I.3 : Enchaînement de diverses consonnes (/n/ /r/ /j/ /p/ /ʃ/ / / / /
avec les quatre voyelles principales /œ/ /i/ / / /
à l'enchaînement de ces voyelles entre elles. (
1977)

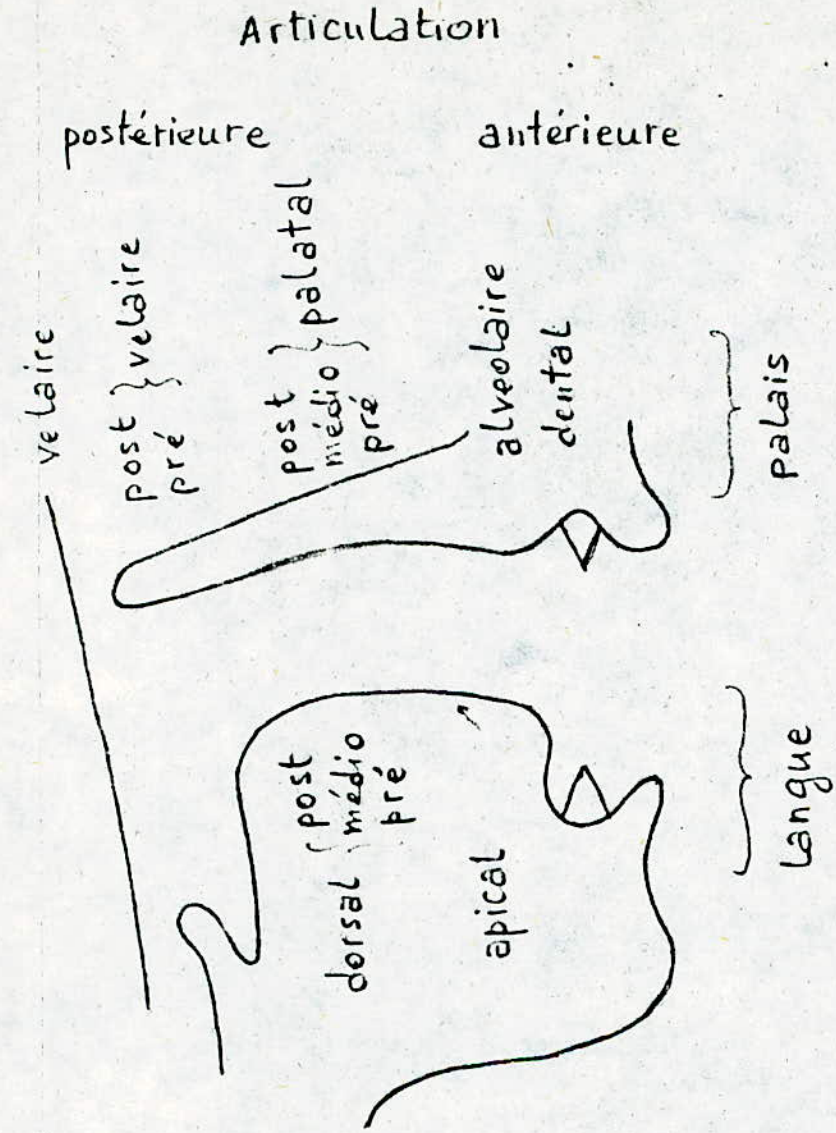


Fig (1.4) : Mode d'articulation de l'appareil vocal.

3.Enchainement des consonnes fricatives avec les voyelles :

Les formants vocaliques subissent une évolution liée au mouvement de l'appareil phonatoire entre position vocalique et position constrictive.

4.Enchainement des consonnes plosives et nasales avec les voyelles :

Le point d'articulation peut être différent pour une même consonne, en particulier lorsque l'articulation est palatale (voir fig. 1-4), il est décalé vers l'avant pour les voyelles antérieures, vers l'arrière pour les voyelles postérieures.

En général, les formants des voyelles sont pertubés par la proximité des consonnes.

CONCLUSION :

Ces notions vont être à la base de toute étude portant sur l'analyse et la synthèse de la parole, sur le plan phonétique et articulatoire.

------•O•------

			bi-labiales	labio-dentales	dentales	palato-alvéolaires	palatales	vélaires	uvulaires
CONSONNES	OCCLUSIVES	Sourdes	p		t			k	
		Sonores	b		d			g	
	CONSTRICTIVES	Sourdes		f	s	ʃ			
		Sonores		v	z	ʒ			
	LIQUIDES	Sonores			l				R
	NASALES	Sonores	m		n				
	SEMI-VOYELLES	Sonores	w/y				j		

VOYELLES	antérieures	centrales	postérieures
FERMEES	i y		u
MI-FERMEES	e ø œ ə		o
MI-OUVERTES	ɛ̃		ɔ̃
OUVERTES		a ɑ	

TABLEAU 1.1 : Les consonnes et les voyelles françaises
(d'après MALMBERG 1974)

CHAPITRE : 2

NOTIONS DE TRAITEMENT DU SIGNAL

---oOo---

2 - 1 - INTRODUCTION :

Le signal, support de l'information émise par une source ou véhicule de l'intelligence dans les systèmes, est particulièrement fragile et doit être manipulé avec beaucoup de soins.

Et, c'est en voulant traiter ce signal, c'est-à-dire le séparer du bruit qui le dégrade et d'extraire les informations les plus importantes pour la compréhension du message, qu'on a développé tout un domaine de recherches : Le traitement du signal.

Le domaine du traitement du signal est très vaste ; dans ce chapitre, nous nous sommes limités aux notions intervenant dans l'analyse et la synthèse de la parole.

2 - 2 - ANALYSE DE FOURIER :

Dans l'analyse fréquentielle, l'outil fondamental reste la transformation de Fourier.

2 - 2 - 1 - Introduction : Le signal de la parole se compose de sons et de bruits. Les sons sont des ondes périodiques, généralement produites en faisant vibrer les cordes vocales, ou par des instruments conçus par l'homme. Par contre, les bruits sont des oscillations complexes non périodiques (expiration de l'air par les poumons sans faire vibrer les cordes vocales).

2 - 2 - 2 - Série de Fourier :

Soit une fonction x (t) périodique, x (t+ T) = x (t). D'après le théorème de Fourier, cette fonction s'écrira comme étant la somme de plusieurs fonctions sinusoïdales :

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(2\pi n f t) + b_n \sin(2\pi n f t) \right] \quad , n \in \mathbb{N} \qquad (2.1)$$

Le spectre de fréquences de cette fonction est défini par :

$$X(nf) = \frac{1}{2}(a_n - j b_n) = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cdot e^{-j 2\pi n f t} \quad , n \in \mathbb{N} \qquad (2.2)$$

2 - 2 - 3 - Transformée de Fourier : (T. F.)

Dans le cas d'un bruit, le signal n'est plus périodique d'où la nécessité de remplacer la notion de série de Fourier par intégrale de Fourier. La transformée de Fourier d'un signal, non périodique continu, x (t) est :

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j2\pi ft} \cdot dt \qquad (2.3)$$

Remarque : Le spectre d'un bruit est continu, par contre celui d'un son est discret.

Propriété de la T. F. : Elle transforme un produit de convolution en un produit simple ; et réciproquement.

$$x(t) * h(t) \rightleftharpoons X(f) \cdot Y(f) \qquad (2.4)$$

(*) produit de convolution ; est défini par :

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(t-\tau) \cdot h(\tau) \cdot d\tau \qquad (2.5)$$

2 - 3 - AUTOCORRELATION :

Le problème qui se pose pertinemment c'est la recherche d'une relation entre deux processus (phénomènes physiques), connus par une grandeur physique traduisant un de leurs paramètres.

Définition : C'est faire une comparaison du signal avec lui-même, décalé de τ dans le temps. La fonction d'autocorrélation est définie par :

$$C_{xx}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} x(t) \cdot x(t-\tau) \cdot dt \qquad (2.6)$$

Cela permet de voir, au moins qualitativement, en quoi la fonction à un instant donné est influencée par "ce qui s'est passé" à un instant τ avant, on voit apparaître une certaine relation avec la mémoire du processus.

Remarque : Cette notion est trop utilisée pour l'extraction de la fréquence fondamentale en analyse prédictive (1).

2 - 4 - ECHANTILLONNAGE ET CODAGE :

Dans de très nombreux cas, on ne traite pas directement les signaux analogiques fournis par les capteurs de mesure, mais on les échantillonne à une fréquence "Fe", c'est-à-dire que l'on observe ces signaux non pas d'une manière continue mais à certains instants seulement.

Ces échantillons seront quantifiés, codés et stockés dans la mémoire des calculateurs, en vue de calculs ultérieurs ou de la reconstitution du signal pour une visualisation ultérieure.

(1) : Voir chapitre 3.

2 - 4 - 1 - Echantillonnage :

C'est la modulation d'un peigne de Dirac par un signal, les échantillons de ce signal sont séparés par la période du peigne de Dirac.

a) Impulsion de Dirac : $\delta(t - t_0)$

C'est une distribution qui assigne à une fonction-test $\varphi(t)$ la valeur numérique $\varphi(t_0)$, selon la relation :

$$\int_{-\infty}^{+\infty} \delta(t-t_0).\varphi(t).dt = \varphi(t_0) \quad , \forall t_0, \forall \varphi \qquad (2.7)$$

La convolution d'un signal avec une impulsion de Dirac fournit une réplique de ce signal, munie d'un retard égal à celui de l'impulsion.

b) Peigne de Dirac :

C'est une suite périodique illimitée d'impulsions de Dirac. Désignons par T_0 la période de répétition des impulsions, nous représenterons alors les peignes unitaires et centrés par la relation :

$$\Pi(t) = \cdots + \delta(t+T_0) + \delta(t) + \delta(t-T_0) + \cdots = \sum_{k=-\infty}^{+\infty} \delta(t-kT_0) \qquad (2.8)$$

La convolution d'un signal avec un peigne de Dirac unitaire et centré fournit une suite périodique :

- dont les motifs sont des répliques du signal ;
- et dont la période est celle du peigne de Dirac.

c) Définition :

L'opération d'échantillonnage revient à multiplier le signal x (t) par une suite d'impulsions de Dirac. Le signal échantillonné est noté:

$$\tilde{x}(t) = x(t) \sum_{k=-\infty}^{+\infty} \delta(t-kT_0) = X(f) * F_e \sum_{n=-\infty}^{+\infty} \delta(f - nF_e) \qquad (2.9)$$

Alors cette opération fait répéter de x (t) le long de l'axe des fréquences avec une période $\frac{"1"}{Fe}$. Pour que la périodicité du spectre ne déforme pas le motif répété, il faut qu'au moins la fréquence d'échantillonnage soit égale à 2 Fm (théorème de Shanon) ; Fm étant la fréquence maximale du spectre de x (t).

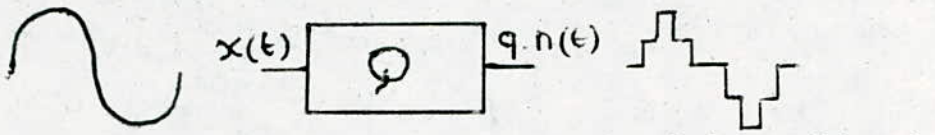
2 - 4 - 2 - Codage :

Le signal échantillonné sera quantifié et codé pour être stocké en mémoire.

a) Quantification : C'est l'approximation de chaque valeur du signal x (t) par un multiple entier d'une quantité élémentaire q,

appelée échelon de quantification (fig. 2. 1).

b) Codage : Le signal, échantillonné et quantifié en amplitude, est représenté par une suite de nombres binaires. Si chaque nombre compte N bits, le nombre maximum d'amplitudes quantifiés qu'il est possible de distinguer s'élève à 2^N.



signal analogique — signal échantillonné

Fig.2.1: Quantificateur

2 - 5 - TRANSFORMEE DE FOURIER DISCRETE (T.F.D.)

Le calcul d'une transformée de Fourier par les moyens électroniques se fait seulement par la variation discrète de la fréquence.

2 - 5 - 1 - Définition : La transformée de Fourier discrète (T. F. D.) est la représentation fréquentielle des 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 toutes le même nombre d'échantillons.

On appelle T.F.D. de "N" valeurs discrètes :

$$X(n) = \sum_{k=0}^{N-1} x(k).e^{-j 2\pi \frac{k\,n}{N}} \quad \text{pour } n = 0, 1, \ldots, N-1 \qquad (2.10)$$

La transformée inverse existe, $x(k) = \frac{1}{N} \sum_{n=0}^{N-1} X(n).e^{j 2\pi \frac{k\,n}{N}}$ (2.11)

2 - 5 - 2 - Transformée de Fourier Rapide (T.F.R.) :

L'emploi de la T.F.D. comporte une sévère limitation dûe à la capacité du calculateur. Une T.F.R. est alors une organisation méthodique du calcul des T.F.D., ramenant le nombre des opérations à effectuer de N^2 à une valeur de l'ordre de $N \log_2 N$.

Cette organisation consiste à utiliser certains algorithmes de calcul (Cooley - Sande...) et à relever les symétries où les ressemblances qui existent à l'intérieur même de la matrice W, $W = \exp(-j\,2\pi/N)$.

2 - 6 - FENETRAGE :

L'enregistrement, ou le traitement en ligne, des signaux représentant le phénomène étudié a une durée limitée. L'appareillage ou l'ordinateur dans le cas d'un traitement différé, impose un temps fini au signal ou au prétraitement. Dans tous les cas, la portion du signal traitée est définie sur une durée Θ. Cela revient à multiplier le signal par une fenêtre temporelle naturelle, qui est égale à :

$$\Pi_{\frac{\theta}{2}}(t) = \begin{cases} 1 & \text{pour } t_0 - \frac{\theta}{2} < t < t_0 + \frac{\theta}{2} \\ 0 & \text{ailleurs} \end{cases} \qquad (2.12)$$

Il lui correspond dans le domaine fréquentielle une fenêtre spectrale $\Phi_0(\nu)$ qui dépend des durées Θ et de la technique adaptée.

L'estimation de la densité spectrale est perturbée par $\Phi_0(\nu)$. Pour diminuer ces effets, on superpose à la fenêtre temporelle naturelle une autre fenêtre temporelle de pondération f (t). Le support de cette nouvelle fenêtre étant au plus égal à Θ, on peut dire que l'on a "remplacé" $\Pi_{\frac{\Theta}{2}}$ (t) par la nouvelle fenêtre.

Les fenêtres utilisées (fig. 2.2) pour la pondération, ont leurs transformées de Fourier qui présentent des ondulateurs plus faibles que celles présentées par la fenêtre naturelle (rectangulaire). Parmi elles, on peut citer :

$$\varphi(t) = \frac{1}{2}\left(1 + \cos 2\pi \frac{t}{NT}\right) \quad : \text{Fenêtre de Hanning} \qquad (2.13)$$

et $$\varphi(t) = 0{,}54 + 0{,}46 \cos 2\pi \frac{t}{NT} \quad : \text{Fenêtre de Hamming} \qquad (2.14)$$

Cette dernière fonction a 99,96 % de son énergie dans le lobe principal et le lobe secondaire le plus important se trouve à environ 40 dB au dessous du lobe principal.

2 - 7 - FILTRES NUMERIQUES :

Dans le traitement numérique du signal, on utilise plutôt des filtres numériques qu'analogiques. Ce filtrage numérique est équivalent à un filtrage analogique, suivi d'une conversion analogique-numérique.

2 - 7 - 1 ; Définition :

Les filtres numériques sont des filtres à réponse impulsionnelle, leur principe consiste à sommer des nombres qui se présentent à

-Pour $\alpha = 0.5$ on a la fenêtre de Hanning

-Pour $\alpha = 0.54$ on a la fenêtre de Hamming

La transformée de Fourier du modèle général de Hamming est :

$$\phi(f) = \alpha \operatorname{sinc}(\pi f\theta) + (I-\alpha)/2 \cdot \operatorname{sinc}(\pi(f-I/\theta)\theta) - (I-\alpha)/2 \cdot \operatorname{sinc}(\pi(f-I/\theta)$$

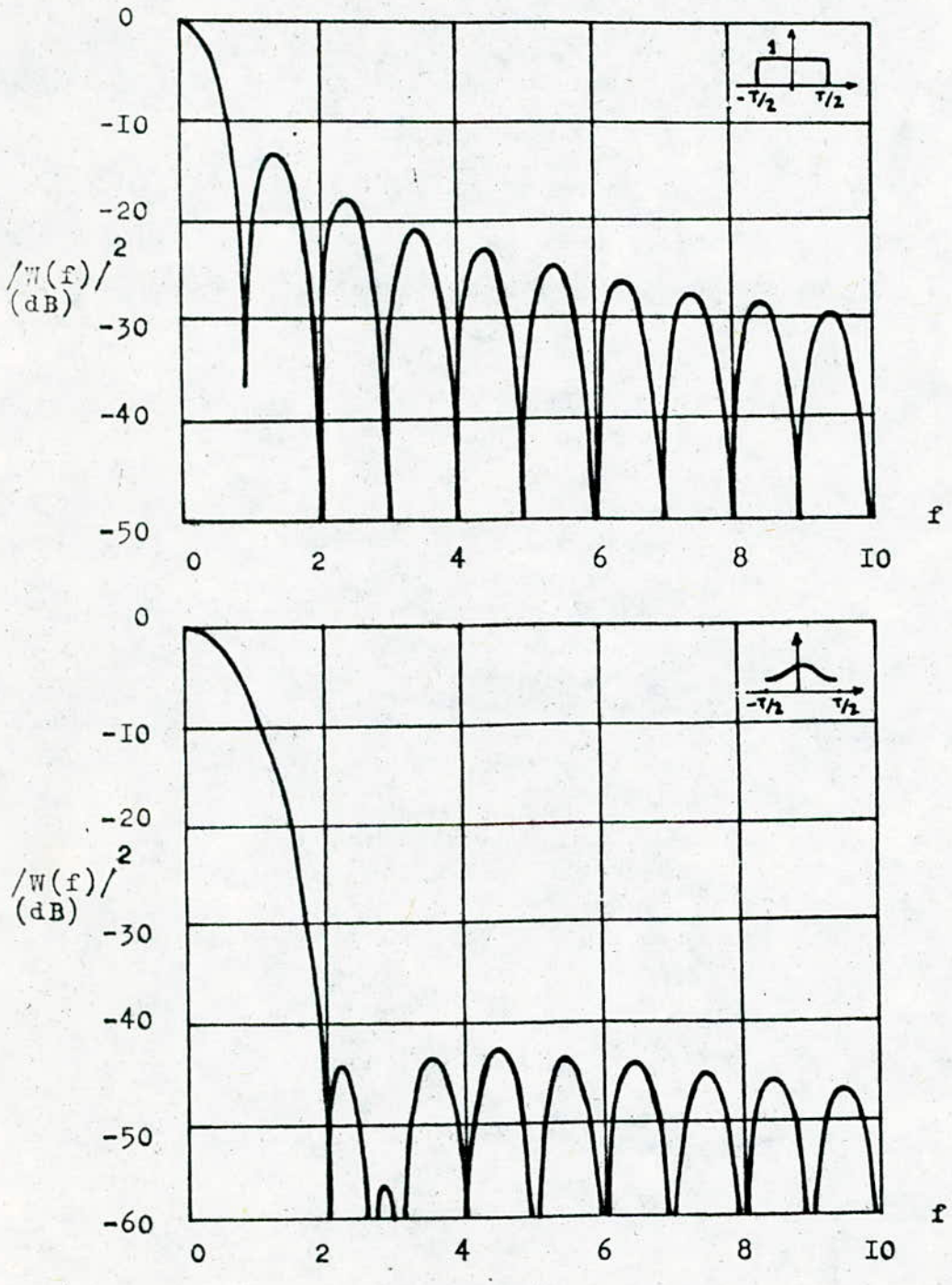


Fig.2.2 : Spectre de la fenêtre rectangulaire et spectre de la fenêtre de Hamming .

l'entrée aprés les avoir pondérés ; Les coefficients de sommation pondérée constitue la réponse impulsionnelle du filtre .

Ces nombres , qui se présentent à l'entrée , peuvent dépendre ou non de la sortie . On distingue selon cette dépendance , des filtres récursifs et non-récursifs .

2 -7 - 2 - Filtres non-récursifs :

Ce sont des filtres à réponse impulsionnelle finie , c'ést-à-dire des systèmes linéaires discrets invariants dans le temps , définis par une équation selon laquelle un nombre de sortie , représentant un échantillon du signal filtré , est obtenu par sommation pondérée d'un ensemble fini de nombres d'entrée , représentant les échantillons du signal à filtrer :

$$y(n) = \sum_{i=0}^{N-1} a_i \, x(n-i) \qquad (2.15)$$

La fonction du filtre (de transfert) s'écrit :

$$H(f) = \sum_{i=0}^{N-1} a_i \, e^{-j 2\pi f i T} \qquad (2.16)$$

2 - 7 - 3 - Filtres récursifs :

Ce sont des filtres à réponse impulsionnelle infinie , c'est-à-dire des systèmes linéaires discrets invariants dans le temps, dont le fonctionnement est régi par une équation de convolution portant sur une infinité de termes :

$$y(n) = \sum_{l=0}^{L} a_l \, x(n-l) - \sum_{k=1}^{K} b_k \, y(n-k) \qquad (2.17)$$

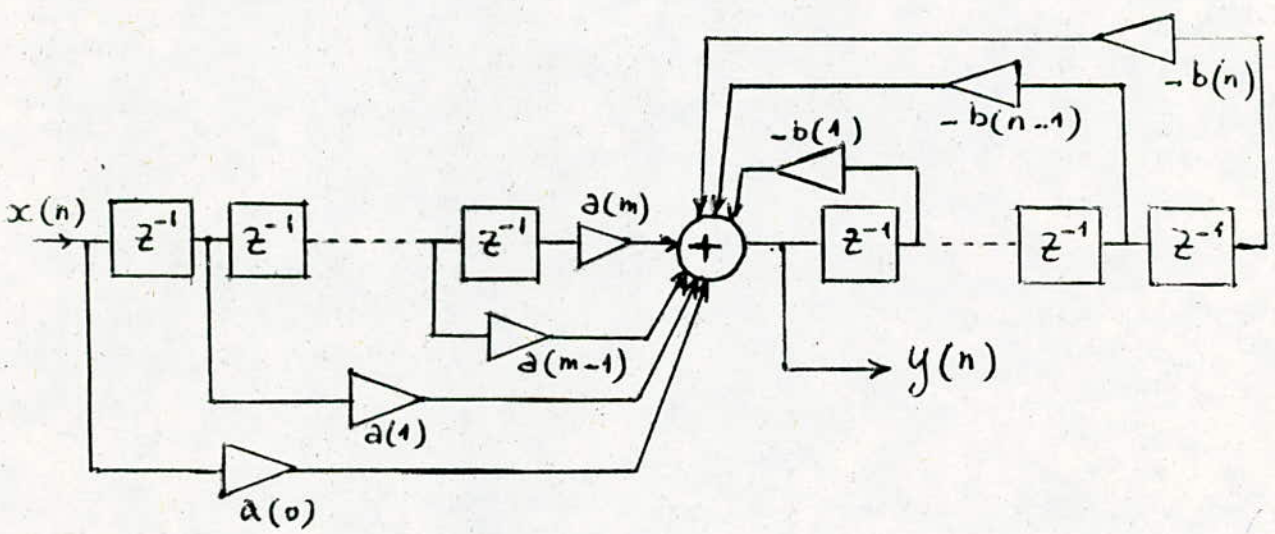


Fig.2.3, Réalisation des filtres récursifs, forme directe.

Pour le calcul des coefficients d'un filtre récursif on fait appel à une fonction modèle (Butterworth, Bessel, Tchebycheff et les fonctions elliptiques:)

La fonction de transfert d'un tel filtre s'écrit :

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{l=o}^{L} a(l).z^{-l} / \left(1 + \sum_{k=1}^{K} b(k).z^{-k}\right) \qquad (2.18)$$

La fonction de transfert présente des zéros (racines de l'équation au numérateur) et des pôles (racines de l'équation au dénominateur) . On dit que ces filtres peuvent être instables à cause de leurs pôles; ceci nous ramène à l'étude de leur stabilité .

Remarque : Si aprés avoir calculé les coefficients d'un filtre récursif, on constate qu'il est instable, on doit alors procéder à une corréction; celle-ci consiste à augmenter le rayon du cercle unité afin d'englober les pôles se situant à l'extérieur . (EL MALAWANY, 1975)

CONCLUSION :

Ces notions sont necessaires pour l'analyse des signaux . Pratiquement, on les utilise pour séparer le signal du bruit qui le dégrade, et d'extraire les paramètres pertinents.

CHAPITRE : 3

L'ANALYSE DU SIGNAL DE LA PAROLE

---oOo---

3 - 1 - INTRODUCTION :

L'étape d'analyse consiste à traiter le signal de parole dans le domaine temporel ou fréquentiel et en tirer des paramètres représentatifs de ce signal.

Dans tout traitement de la parole, on procède d'abord par une analyse du signal suivant certaines méthodes que nous décrirons dans ce chapitre.

Le choix de la méthode se base sur le type d'application, et sur le coût. La qualité de la parole synthétique dépendra en grande partie de la méthode utilisée. Ces méthodes d'analyse utilisent les "outils mathématiques" du traitement de signal (voir ch. 2).

3 - 2 - ANALYSE SPECTRALE :

Dans le cas où le signal ne varie pas très vite, on s'intéresse au spectre du signal plutôt qu'à son évolution temporelle ; on peut noter l'exemple des voyelles qui présentent dans leur partie centrale une certaine stabilité du spectre.

3 - 2 - 1 - Transformée de Fourier (T. F.)

Théoriquement, la transformée de Fourier se calcule sur un intervalle de temps infini, en plus les caractéristiques de la source et du conduit vocal changent. Tout cela nous conduit à appliquer une fenêtre (cf. Ch. 2) dans laquelle le signal peut être considéré comme stationnaire (la configuration du conduit vocal n'évolue pas tellement pendant une durée de 25 ms).

On calcule donc la T. F. D. (cf. ch. 2) sur chaque intervalle définissant ainsi une série de spectres que l'on appelle spectre à court terme. Etant donné que le nombre d'opérations nécessaires dans ce cas s'élève à "N^2" pour N échantillons, on utilise l'algorithme de la T.F.R. (cf. ch. 2) qui ramène le nombre "N^2" à "$N \log_2 N$" (contraintes de capacités des calculateurs).

3 - 2 - 2 - Analyse par filtres :

a) Analyseur du vocodeur (1) à canaux :

Le signal de parole est analysé à l'aide d'un banc de filtres passe-bande (de 12 à 15 filtres) couvrant l'étendue spectrale (bande téléphonique 300-3 000 HZ). Le signal délivré par chacun des filtres d'analyse

(1) de l'anglais "voice coder".

subit une détection puis il traverse un filtre passe-bas (0-50 hz). Le signal issu de ce dernier filtre représente l'évolution de la densité spectrale d'énergie. Les signaux issus de ces derniers étages sont quantifiés et codés pour être transmis.

Le vocodeur comporte aussi un dispositif de décisions voisé/non voisé. Dans le cas d'un son voisé, le mécanisme extrait la fréquence fondamentale. Il génère un signal binaire de décision pour piloter son analogue dans le synthétiseur (cf. fig. 4 - 1).

b) Sonagraphe :

On utilise un seul filtre mais variable (fig. 3-1). Une tranche du signal à analyser est enregistrée sur disque magnétique. Le signal issu d'une tête de lecture est appliqué ensuite à l'entrée d'un filtre passe-bande dont la fréquence centrale est contrôlée par la position verticale d'un stylet inscripteur. Le signal de sortie du filtre règle l'intensité du tracé sur le papier enregistreur. Le déplacement vertical du stylet s'effectue lentement et progressivement du bas vers le haut par l'intermédiaire d'une vis sans fin. Il provoque une variation linéaire, dans la gamme "0 - 8 000 Hz", de la fréquence centrale du filtre. Durant un tour du disque, on peut considérer que la position du stylet, et par conséquent, la fréquence centrale du filtre d'analyse sont fixes. Cinq minutes d'analyse sont nécessaires pour analyser deux secondes de parole. Le diagramme délivré par le sonagraphe est appelé spectrogramme (fig. 3 - 2) ou sonagramme.

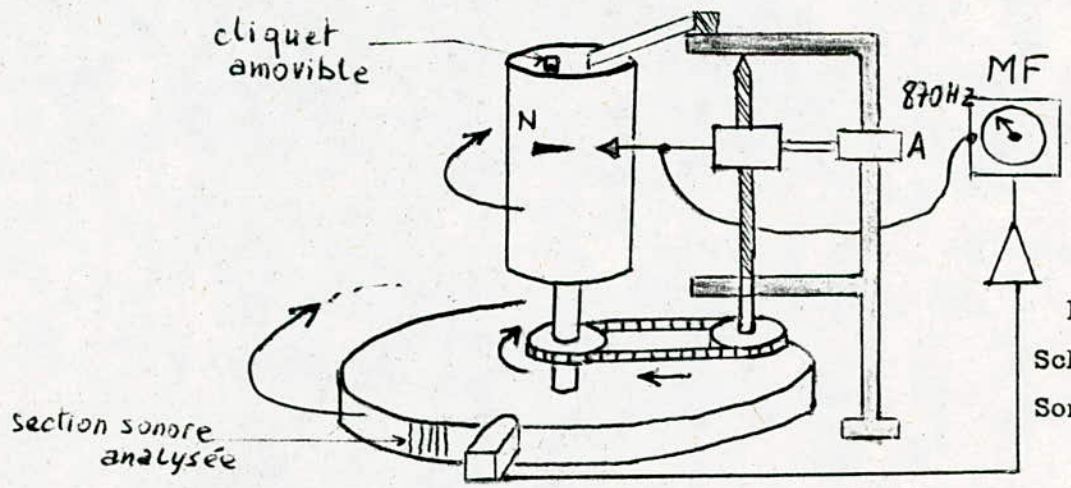


Fig. 3.1 : Schéma d'un Sonagraphe

3 - 3 - ANALYSE TEMPORELLE :

3 - 3 - 1 - Méthode d'autocorrélation :

L'autocorrélation révèle la ressemblance d'un signal avec lui-même (cf. ch. 2).

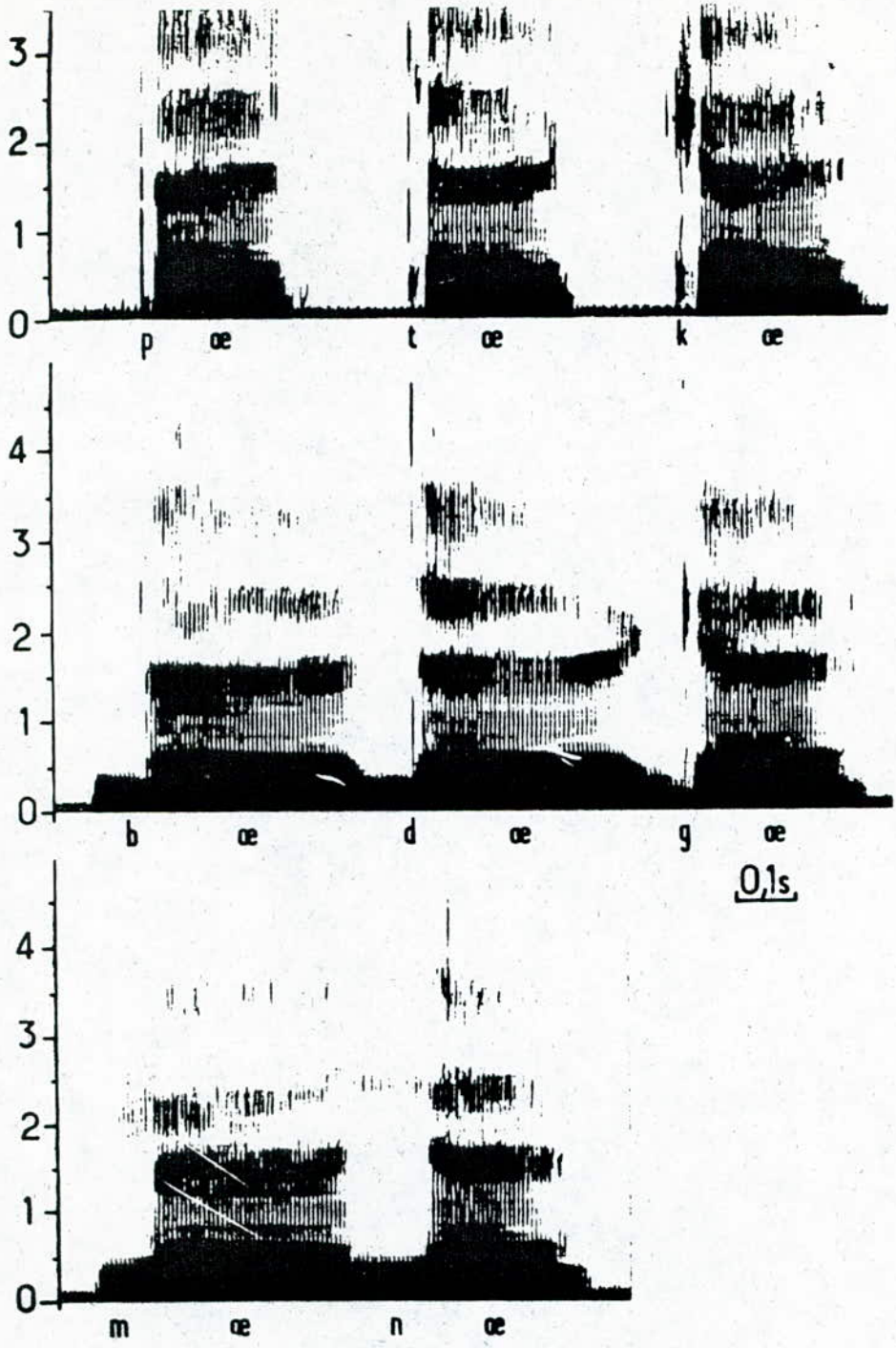


Fig 3.2. : Spectrogrammes réels des consonnes plosives et nasales en association avec la voyelle / œ / .

En plus de l'information fréquentille et temporelle, le spectrogramme montre l'intensité qui est traduite par le noircissement plus ou moins intense.

Cette méthode est utilisée particulièrement pour déterminer la fréquence fondamentale. En effet, la fonction d'autocorrélation présente un maximum pour " $\tau = T_0$" (T_0 = période fondamentale).

3 - 3 - 2 - Passage par zéro du signal :

Le signal S (t) prend la valeur zéro (ou change de signe) à des instants dont la répartition dans le temps est liée à certaines caractéristiques spectrales de S (t). Cependant, l'information relative à l'amplitude est perdue car on ne s'intéresse qu'à son signe.

Cette méthode est utilisée pour la mesure de la fréquence de mélodie. Elle présente une grande simplicité de mise en oeuvre.

3 - 3 - 3 - Codage linéaire prédictif : (LPC)

La prédiction linéaire est une méthode qui a été appliquée au signal de parole pour la première fois en 1967 par F. Itakura et S. Saito, puis développée par de nombreux chercheurs dont J.D. Markel et A. H. Gray en 1976. Cette méthode est considérée à la fois comme temporelle et spectrale.

Dans le domaine temporel, elle considère qu'un échantillon de parole peut être prédit comme fonction linéaire d'un certain nombre d'échantillons précédents :

$$\hat{S}(n) = \sum_{k=1}^{p} a(k) \cdot S(n-k) \qquad (3\_1)$$

p : ordre du prédicteur.

Et ceci, en se basant sur le fait que les variations du conduit vocal peuvent être approchées par une succession de configurations stationnaires qui durent 10 à 25 ms, pendant ce temps la source est constante, et la fonction de transfert du conduit vocal peut être représentée par un filtre numérique récursif.

L'approche d'un phénomène physique par un modèle mathématique se fait toujours avec une erreur, qui est dans ce cas :

$$e(n) = S(n) - \hat{S}(n)$$

et si on remplace $\hat{S}(n)$ par sa valeur dans (3 - 1) on aura :

$$e(n) = S(n) - \sum_{k=1}^{p} a(k).\ S(n-k) \qquad (3-2)$$

le calcul de la transformée en "Z" de (3 - 2) donne :

$$E(Z) = S(Z)\left(1 - \sum_{k=1}^{p} a(k).\ Z^{-k}\right)$$

Si on pose $F(Z) = \sum_{k=1}^{p} a(k).\ Z^{-k}$ (3 - 3)

la relation (3 - 3) devient :

$$E(Z) = S(Z)\ (1 - F(Z)) \qquad (3-4)$$

On schématise le modèle, à partir de la fonction de transfert, comme suit :

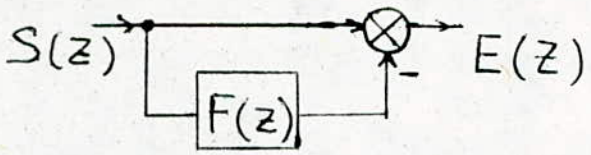


Fig. (3 - 3) : Modèle de prédiction linéaire.

D'une manière générale, on peut exprimer tout signal temporel en terme d'un modèle prédit et d'un signal d'erreur :

$$S(n) = \hat{S}(n) + G\,e(n) \qquad (3-5)$$

G : est une constante d'adaptation d'énergie, lorsque e (n) est exactement égale à l'erreur de prédiction correspondant au signal, on a G = 1.

Dans le cas particulier de la prédiction linéaire on a :

$$S(n) = \sum_{k=1}^{p} a(k)\ S(n-k) + G\,e(n) \qquad (3-6)$$

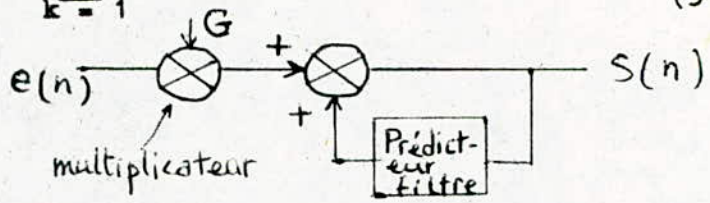


Fig. (3 - 4) : Modèle de production de la parole dans le domaine temporel.

En prenant la transformée en "Z" de l'équation (3 - 6), on obtient :

$$S(Z) = S(Z)\ \left(\sum_{k=1}^{p} a(k)\,Z^{-k}\right) + G.\,E.\,(Z) \qquad (3-7)$$

$$S(Z)\ \left(1 - \sum_{k=1}^{p} a(k)\,Z^{-k}\right) = G.\,E\,(Z) \qquad (3-8)$$

d'où : $$H(z) = \frac{S(z)}{E(z)} = \frac{G}{1 - \sum_{k=1}^{p} a(k)\,Z^{-k}} \quad \text{avec } a_0 = 1. \qquad (3-9)$$

d'où le schéma suivant :

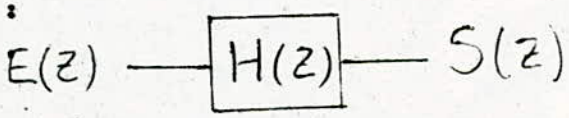


Fig. (3 - 5) : Modèle de production de la parole dans le domaine fréquentiel.

a) Détermination des coefficients a(k) :

En admettant que le spectre de la parole puisse être décrit à l'aide de cinq résonnances (D'aprés Makhoul, 1972) et du fait que chaque formant soit associé à deux coefficients, on aura besoin de dix coefficients pour représenter le conduit vocal. En ajoutant deux supplémentaires caractérisant l'influence de la source vocale et du rayonnement au niveau des lèvres, cela porte à (12) la valeur moyenne du nombre " p " du prédicteur.

Le problème consiste à calculer ces " p " coefficients pour cela, on fait appel à la méthode de covariance ou d'autocorrélation.

- La covariance: Elle est fondée sur la minimisation de l'écart quadratique total (critère des moindres carrés) :

$$E = \sum_{n=1}^{N} e^2(n) = \sum_{n=1}^{N} \left(S(n) - \sum_{k=1}^{p} a(k)\, S(n-k) \right)^2 \qquad (3-9)$$

On annule la dérivée partielle de "E" par rapport aux "a (k)", ce qui conduit aux équations suivantes :

$$\sum_{k=1}^{p} a(k) \cdot \varphi(ik) = -\varphi(i\,0) \qquad (3-10)$$

telles que : $$\varphi(ik) = \sum_{k=0}^{N-1} S(n-i)\; S(n-k) \qquad (3-11)$$

Les coefficients $\varphi(ik)$ constituent une matrice de covariance (symétrique).

- Autocorrélation : Cette méthode consiste aussi à minimiser l'erreur quadratique "E". Cette minimisation conduit au système d'équations suivant :

$$\sum_{k=1}^{p} a(k).\, R\,|\,i-k\,| = R(i) \qquad (3-12)$$

où $$R(i) = \sum_{n=0}^{N-|\,1-i\,|} S(n)\; S(n-|i|) \qquad (3-13)$$

Ici encore, le système d'équation est de "p" équations à "p" inconnues. Les coefficients R (i - k) constituent une matrice dite d'autocorrélation.

Les méthodes, covariance et autocorrélation, avancent des hypothèses différentes. La première suppose que le signal est défini pour "p+n" échantillons, l'optimisation se fait sur toute l'échelle des temps NT (N = nombre d'échantillons, T période d'échantillonnage). La deuxième suppose

que le signal est défini pour toutes les valeurs du temps ($-\infty > n > +\infty$). Pratiquement, on ne s'intéresse qu'à l'intervalle fini pendant lequel l'appareil vocal peut être considéré encore stationnaire. Par conséquent, on applique une fenêtre d'analyse "W (n)" au signal tel que le produit S (n) . W (n) est nul en dehors d'une séquence de N échantillons.

Le mode d'analyse :

Il peut être du type asynchrone c'est-à-dire la séquence du signal analysé et sa position sont indépendants des impulsions d'excitation ou du type synchrone où l'intervalle d'analyse se situe entre deux instants correspondants au début de la fermeture de la glotte, autrement dit sur une période fondamentale (pour les sons sonores), éliminant ainsi les inconvénients de non-stationnarité et de pseudo-périodicité de la parole (EL MALAWANY, 1975).

b) Calcul du gain :

On calcule le gain "G" en se basant sur le critère suivant : L'énergie totale contenue dans le signal de synthèse doit être égale à celle du signal d'analyse. Dans la méthode d'autocorrélation G est donné par :

$$G^2 = R(0) + \sum_{k=1}^{p} a(k)\ R(k) \qquad (3-14)$$

R (k) : sont les coefficients d'autocorrélation.

c) Mesure des formants :

Les formants correspondants aux résonnances du conduit vocal, et le modèle de la prédiction linéaire se base sur un filtre "tous-pôles". Ces pôles correspondent aux formants. Il suffit ainsi de calculer les racines de l'équation :

$$1 - \sum_{k=1}^{p} a(k)\ Z^{-k} = 0 \qquad (3-15)$$

Cependant, seules les racines complexes conjuguées donnent les formants :

$$F(k) = \frac{1}{2\pi T} \operatorname{Arctg} \frac{\Im(ki)}{\Re(kr)} \quad : \text{Fréquence du formant} \qquad (3-16)$$

$$b(k) = \frac{1}{2\pi T} \operatorname{Log}\left(Z^2(ki) + Z^2(kr)\right) \quad : \text{Bande passante du formant} \qquad (3-17)$$

3 - 4 - DETECTION DE LA FREQUENCE FONDAMENTALE :

L'un des paramètres les plus importants issus d'une analyse de la parole est la fréquence de " pitch " . Il existe dans tous les synthétiseur

un générateur de bruit pour produire les consonnes, et un générateur d'impulsion périodiques (pour les sons voisés) de période "$T_o = \frac{1}{F_o}$" détectée dans l'analyse. Plusieurs méthodes de détection existent, nous citerons quelques-unes (spectrales et temporelles) .

3 - 4 - 1 - Méthode d'intercorrélation avec une fonction peigne :

On calcule l'intercorrélation entre le spectre d'amplitude I F(ω) d'un son voisé et une fonction $p(\omega,\omega_p)$. (Fig.3.6)

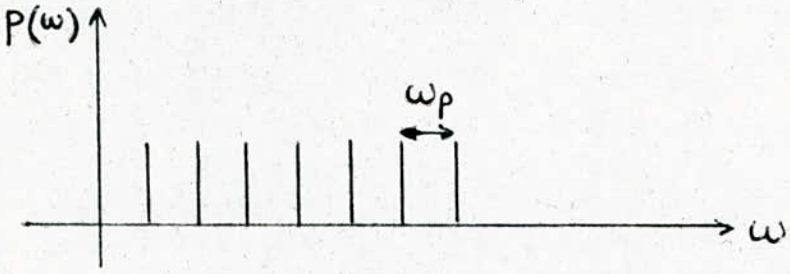


Fig.(3.6): Graphe de p(ω)

Le maximum de la fonction est obtenu lorsque la distance entre deux dents est égale à $\omega_p = 2\pi F_o$.

3 - 4 - 2 - Méthode du cepstre :

On peut considérer que le spectre de puissance de la parole

$$I\ S(f)I^2 = I\ C(f)\ I^2 . I\ G(f)\ I^2$$

où : C(f) est la fonction de transfert du conduit vocal;

I G(f) I est le spectre du signal de la source.

En prenant le logarithme du produit, on obtient une somme de logarithmes. Le retour à la dimension du temps se fait en effectuant une transformée de Fourier inverse. (Fig.3.7)

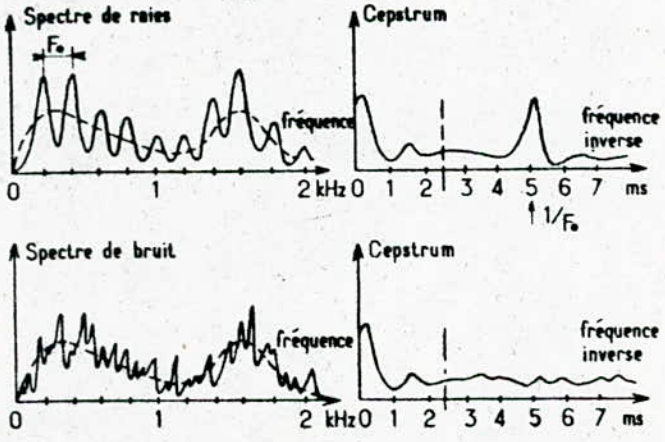


Fig. (3.7) : Calcul du cepstrum.

Une détection du pic sur le cepstre permet l'extraction de "F_o". (Cf.Liénard, 1977)

3 - 4 - 3 - Méthode des harmoniques :

On détecte deux harmoniques F_1 , F_2 par exemple, la fréquence fondamentale" F_o " est le plus grand diviseur commun de F_1, F_2 .

Deux autres méthodes (temporelles) ont été cité dans les paragraphes (3.2.1) et (3.2.2) .

CONCLUSION :

Nous avons cerné, d'une manière globale, dans ce chapitre, les méthodes d'analyse ayant pour objet de tirer des paramètres pertinents du signal de parole. La prédiction linéaire a l'avantage d'être basée sur un modèle simple.

Elle permet de fournir la majorité des paramètres tels que, la mélodie le gain, les formants, ou encore la fonction d'aire du conduit vocal.(Cf.§4.1.5)

------oOo------

CHAPITRE 4

LA SYNTHESE DE LA PAROLE
& SES METHODES

---oOo---

4 - 1 - INTRODUCTION :

Nous examinerons dans ce chapitre, les moyens utilisés par l'homme pour reproduire artificiellement le signal de la parole.

Les premiers essais remontent aux XVIIIe et XIXe siècles (voir J. S. Lienard, 1977 ; J. Guibert 1979). Plus tard, Dudley (1939) mis au point le vocodeur à canaux en essayant de mieux utiliser les lignes téléphoniques, marquant ainsi, le début des recherches modernes.

La qualité de la voix émise par les vocodeurs n'était pas très agréable, d'autres types de synthétiseurs ont été mis au point, tels que le synthétiseur à formants qui représente un progrès dans la qualité de la voix, et les simulateurs du conduit vocal, qui sont en cours d'essai donneront une représentation beaucoup plus fidèle du système vocal.

4 - 2 - TECHNIQUES DE SYNTHESE :

4 - 2 - 1 - Vocodeur à canaux :

a) Description :

Le synthétiseur du vocodeur a une structure de l'analyseur (fig. 4_1). Une source d'impulsions reconstitue un spectre de raies (pour les sons voisés), et une source de bruit (pour les sons non-voisés) fournit un spectre continu.

b) Fonctionnement :

Les données quantifiées des canaux d'analyse sont multiplexées avec celle du détecteur de la fréquence fondamentale pour former une trame.

La synthèse est effectuée par un banc de filtres semblables à ceux de l'analyse. Ces derniers sont attaqués par un signal d'excitation élaboré à partir des données concernant la détection de "F_o". Ce signal d'excitation doit avoir un spectre plat dans la bande de fréquences occupée par le banc de filtres.

L'entrée de chaque filtre passe-bande est modulée en amplitude en fonction de l'énergie mesurée à la sortie du filtre d'analyse correspondant. Le signal final est la somme des sorties des filtres.

Le vocodeur présente des avantages tels que la réduction de la bande passante à 25 Hz ; multipliée par le nombre de canaux (11 X 25 Hz = 275 Hz, par exemple) au lieu de 3 100 Hz requise par le téléphone, ce qui permet un codage inférieur à 5 000 bits/s.

Cependant, sa réalisation est encore lourde et onéreuse, la fréquence des formants n'est définie que par échelon de 300 Hz, par exemple, la limitation à 25 Hz des variations d'énergie a pour effet de supprimer les transitions rapides (durée inférieure à 40 ms), et la décision voisée/non voisée est sujette à des erreurs sachant qu'il y a des fricatives voisées par exemple.

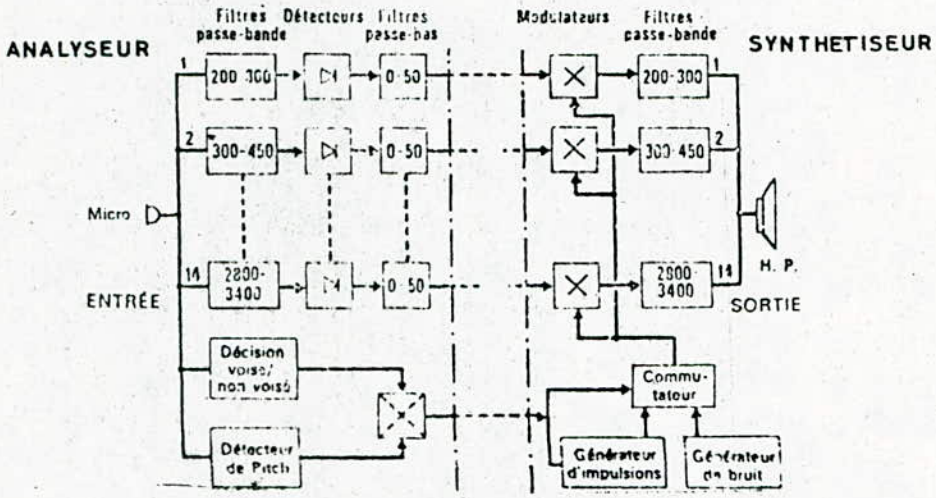


Fig. 4 - 1 : Schéma du vocodeur à 14 canaux.

4 - 2 - 2 : Vocodeur à formants :

Ce type de machines est constitué par des circuits résonnants, reproduisant les résonnances formantiques dues au conduit vocal, disposés en cascade (fig. 4 - 2).

Dans ce synthétiseur, on trouve deux sources d'excitation, un générateur d'impulsions périodiques, et un générateur de bruit, en plus il comporte trois canaux :

- Un canal des formants vocaux comportant trois résonnances variables qui donnent naissance aux trois premiers formants. Ce canal est attaqué par la source d'impulsions de fréquence "F_0.
- Un canal de nasalité contenant des circuits de formants de nasalité se traduisant par des absorptions aux voisinages de certaines fréquences.
- Un canal de bruit, comportant trois circuits de formants de bruit B1, B2 et B3.

Le fonctionnement de ce synthétiseur est commandé à l'aide d'un certain nombre de paramètres :

- Trois fréquences de formants vocaux F1, F2 et F3.
- Trois fréquences de formants de bruit B1, B2 et B3.
- La fréquence fondamentale F_0.
- Les amplitudes A (o), A (b), A (n), A (z) de l'excitation vocale, du bruit, de la nasalité, et du bruit injecté dans le canal vocal.

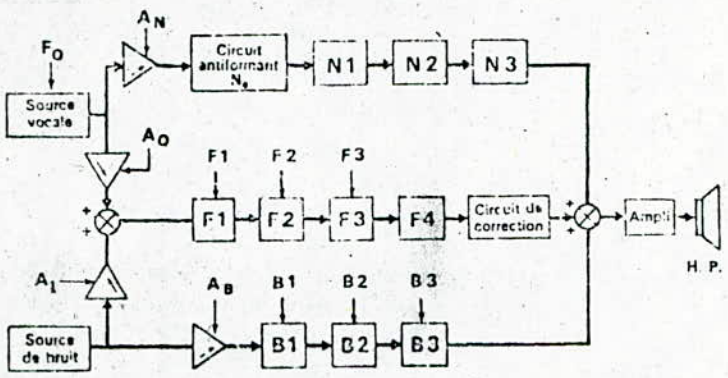


Fig. (4 - 2 a) : Schéma du synthétiseur à formants (série).

4 - 2 - 3 : Synthétiseur à formants "parallèle" :

La différence entre ce synthétiseur et son analogue "série" réside dans la disposition en parallèle des circuits de formants (vocaux et de bruits). (voir fig. 4 - 2 - b)

Comparé au synthétiseur à formants "série", ce dernier permet une meilleure approximation du conduit vocal pour les sons non-voisés.

Il présente en outre l'avantage de ne pas nécessiter des contrôles individuels pour les amplitudes des divers formants.

A l'inverse, le synthétiseur "parallèle" permet, grâce aux contrôles indépendants des différents formants, de simuler les effets de l'effort vocal sur le spectre de la source glottale. Il permet également de reproduire des sons excités en un point quelconque du conduit (important pour la synthèse des consonnes) (Guibert 1979).

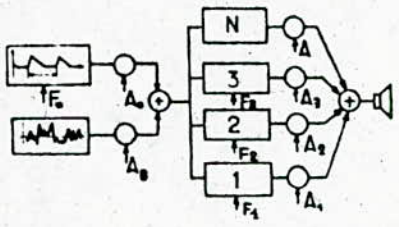


Fig. 4 - 2 b : Synthétiseur à formants parallèle (d'après LIENARD, 1977).

4 - 2 - 4 : Synthèse prédictive :

La figure (4 - 3) montre le principe de la réalisation d'un synthétiseur utilisant la prédiction linéaire. Les paramètres de contrôle fournis au synthétiseur sont :

- la période "T_o" du fondamental ;
- un signal binaire relatif au choix de l'excitation (voisé ou non voisé) ;

les "p" coefficients de prédiction.

Le générateur d'impulsions fournit, au début de chaque période de pitch, une impulsion. Le générateur de bruit blanc produit un signal aléatoire. C'est la commande binaire qui contrôle la commutation de l'un ou de l'autre de ces deux dispositifs d'excitation. L'amplitude du signal d'excitation est ajustée à l'aide de l'amplificateur "G".

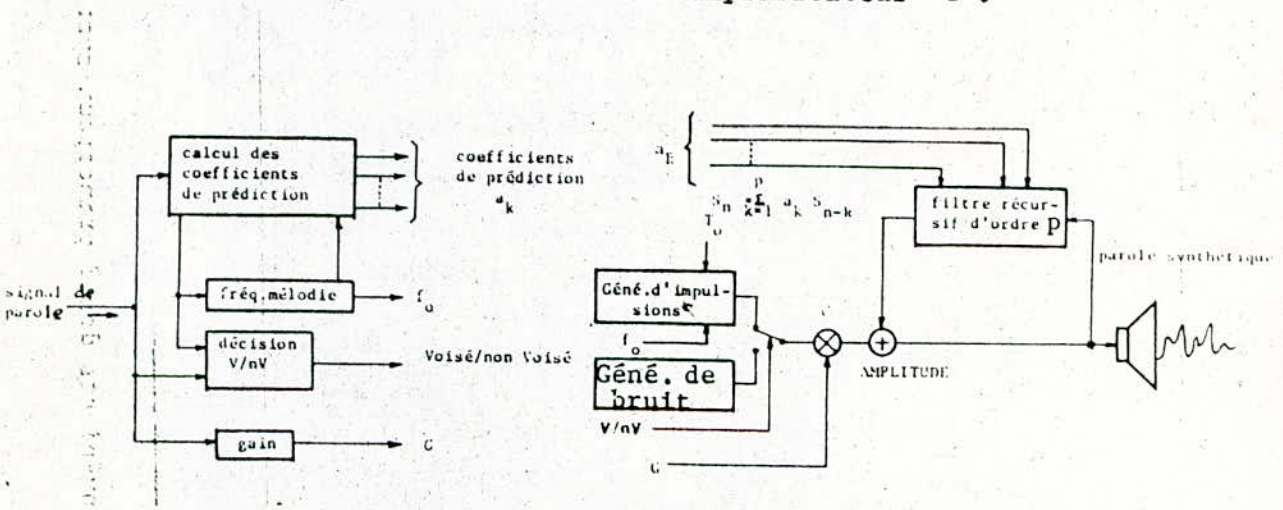


Fig. (4 - 3) : Analyse-synthèse par prédiction linéaire.

La valeur " $\hat{S}$ " du signal à l'instant " nT ", prédite à partir des échantillons précédents, est combinée avec le signal d'excitation "δ_n", lorsque ce dernier est présent; on obtient ainsi le n ième échantillon du signal synthétique. Les échantillons sont enfin soumis à un filtrage passe-bas .

Tous les paramètres de contrôle sont renouvelés au début de chaque période de pitch pour la parole voisée, et, par exemple, tous les (10)ms pour la parole non-voisée.

4 - 2 - 5 : Synthèse par simulation du conduit vocal :

La superiorité du synthétiseur à formants sur le vocodeur à canaux était certainement dûe à ce que les formants constituent une bonne approximation des résonances du conduit vocal. Toutefois, pour une reproduction

plus fidèle du fonctionnement de l'appareil phonatoire, on doit simuler ce dernier par une succession de tuyaux acoustiques ou de circuits électriques élémentaires, chacun des éléments représentant une tranche du conduit (fig. 4 - 4).

D'un point de vue acoustique, la propagation des ondes dans un tuyau de section non-uniforme obéit à l'équation de Webster :

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial P}{\partial x} \cdot \frac{\partial A}{\partial x} \cdot \frac{1}{A} = \frac{1}{o^2} \cdot \frac{\partial^2 P}{\partial t^2} \qquad (4 - 1)$$

P = P (x,t) : Pression en un point d'abscisse x, à l'instant t ;
A(x) : L'aire de section du conduit ;
c : Vitesse du son.

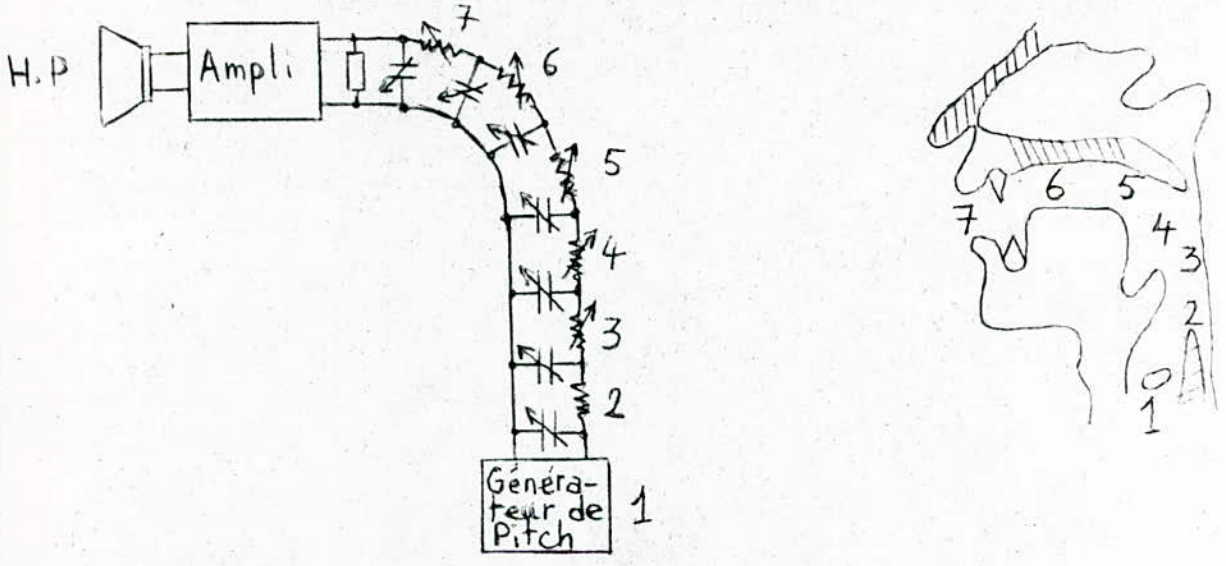


Fig. 4 - 4 : Simulation du conduit vocal à l'aide d'une ligne électrique.

Une résolution numérique de la dernière équation permet de calculer, en supposant que la pression au niveau de la glotte prend une valeur non nulle "P_o", de proche en proche la pression en tous les points du conduit.

La fonction d'aire joue un rôle important dans tout système simulant l'appareil phonatoire. Pour son calcul, on utilise certaines méthodes, telle que la cinéradiographie qui prend des images à rayons X.

Une autre méthode, beaucoup plus simple, est celle de la prédiction linéaire. En supposant le conduit vocal réduit à une succession de tuyaux cylindriques d'égales longueurs, et d'aires A (1), A (2), A (n) et en considérant une onde sonore se propageant à l'intérieur, le coefficient de réflexion r (k) entre les sections contigues A (k) et A (k + 1) est donné par :

$$r(k) = \frac{A(k) - A(k+1)}{A(k) - A(k+1)} \quad . \qquad (4-2)$$

On montre que la connaissance des coefficients de prédiction a (k) permet le calcul des coefficients r (k) et par la suite, de proche en proche, celui des aires successives dès lors que l'on a fixé la première A_0.

En simulant section par section le système phonatoire, les analogues du conduit vocal devraient permettre une reproduction fidèle des caractéristiques de la voix. Les contraintes articulatoires sont assez facilement traduites en lois régissant les changements au cours du temps de la fonction d'aire, ce qui est essentiel pour les transitions interphonémiques. En plus, il est possible de décrire la configuration du conduit vocal à l'aide d'un petit nombre de paramètres.

4 - 3 - LES METHODES DE SYNTHESES :

Si l'on désire, à l'aide d'un synthétiseur donné, produire de la parole continue, on est amené à assembler des éléments phonétiques préalablement analysés, codés, et mis en mémoire sous forme de paramètres qui serviront au contrôle de la synthèse.

4 - 3 - 1 - Synthèse par phrases :

Il ne s'agit pas, proprement dit, de synthèse de la parole, mais d'enregistrements analogiques des phrases sur des bandes magnétiques. Cette méthode est utilisée particulièrement en téléphonie où on peut éviter la présence permanente d'une opératrice rien que pour dire "il n'y a pas d'abonné au numéro demandé" par exemple.

La parole ainsi produite est très fidèle, cependant, le nombre de phrases est très limité.

4 - 3 - 2 - Synthèse par mots :

La mise en service des ordinateurs a permis de concevoir de nouvelles procédures, on utilise des mots ou des membres de phrases que l'on stocke, par exemple, de façon analogique sur disque magnétique. Ce dernier étant relié à un calculateur. A chaque mot correspond une adresse, la composition du message est réalisée par l'intermédiaire du calculateur qui établit l'adressage successif des mots et oriente la tête de lecture vers les mots choisis. La parole restituée est d'excellente qualité, cependant on ne peut pas parler de véritable synthèse puisque les mots sont stockés sans analyse préalable, et pour des raisons d'encombrement et de coût, le vocabulaire de base est limité (le stockage des mots peut être fait numériquement, mais toujours au détriment de la capacité de la

mémoire.)

4 - 3 - 3 - Synthèse par règles :

Partant d'une phrase, si l'on fait une transcription phonétique suivie d'une traduction phonème-son élémentaire, après stockage, la parole reproduite sera inintelligible, et la raison est que les formes sonores changent rarement d'individualité aux endroits que nous serions tentés de considérer comme des frontières interphonémiques (Guibert, 1979) (fig.4-5).

Le problème réside dans le découpage interphonémique, et c'est les transitions formantiques qui constituent l'information la plus importante que nous utilisons dans la perception de très nombreuses consonnes. Un exemple des transitions de formants des consonnes IbI,IdI et IgI avec différentes voyelles est donné au tableau (4 - 2).

La connaissance de l'évolution des formants des différents phonèmes en contact permet d'établir des règles, en plus une connaissance accrue des processus de la phonation permet d'améliorer en retour la commande des synthétiseurs. Les résultats peuvent aider à la mise en oeuvre d'une commande de synthétiseur à partir de "particules élémentaires" du langage étudié (consonnes, voyelles, semi-voyelles). Cette méthode nécessite l'emploi d'un organe de calcul associé à une mémoire pour déterminer, automatiquement à partir des phonèmes, l'évolution des différents paramètres indispensables à la commande du synthétiseur (évolution des formants).

La mise au point d'une synthèse par règles pose des problèmes complexes. En effet, les caractéristiques d'un même phonème dépendant de son environnement. Le synthétiseur à formants semble bien adapté à la synthèse par règles, ainsi que le simulateur du conduit vocal, à condition que les contraintes articulatoires soient déterminées.

Les travaux de synthèse par règles sont suivis avec grand intérêt pour deux raisons principales : d'une part, cette méthode permet de commander un synthétiseur à partir d'un nombre réduit d'information, d'autre part, elle débouche sur la transformation du langage écrit en parole.

4 - 3 - 4 - SYNTHESE PAR DIPHONES :

La synthèse automatique de la parole consiste en le passage de l'écriture orthographique au son. La première opération consiste à savoir faire automatiquement la transcription orthographique-phonétique en adoptant un code, disposant théoriquement, d'un système qui devrait permettre de composer n'importe quelle phrase en combinant les lettres entre elles.

		Time of onset relative to burst ms	Onset freq. Hz	Trans. duration ms	Vowel steady-state freq. Hz	Freq. 95 ms from burst Hz
/bi/	F1	5	310	14	288	279
	F2	5	1816	32	2123	2300
	F3	5	2462	30	2625	2765
/di/	F1	12	298	9	278	275
	F2	12	1920	41	2137	2277
	F3	12	2587	25	2618	2768
/gi/	F1	17	275	7	263	291
	F2	22	2322	12	2226	2270
	F3	17	2835	21	2795	2782
/bɪ/	F1	3	359	9	356	363
	F2	3	1745	24	1860	1951
	F3	5	2424	22	2531	2551
/dɪ/	F1	14	313	17	340	372
	F2	14	1881	17	1916	1966
	F3	14	2587	8	2582	2592
/gɪ/	F1	23	284	25	322	374
	F2	25	2182	33	2084	2053
	F3	23	2679	32	2538	2538
/be/	F1	4	371	27	459	475
	F2	4	1663	25	1804	1913
	F3	7	2402	41	2531	2582
/de/	F1	10	340	20	411	476
	F2	10	1842	9	1847	1904
	F3	10	2556	17	2582	2558
/ge/	F1	18	340	19	385	450
	F2	21	2174	34	2058	2039
	F3	20	2580	43	2538	2541
/bɛ/	F1	3	344	24	460	529
	F2	3	1592	39	1736	1758
	F3	5	2302	32	2478	2514
/dɛ/	F1	11	341	35	450	501
	F2	11	1831	5	1815	1807
	F3	11	2562	27	2466	2495
/gɛ/	F1	15	301	43	427	485
	F2	16	2180	63	1878	1866
	F3	21	2566	38	2470	2464
/bæ/	F1	2	433	30	615	727
	F2	2	1526	14	1571	1631
	F3	2	2156	34	2387	2394
/dæ/	F1	10	370	47	621	652
	F2	10	1786	55	1696	1697
	F3	10	2548	38	2481	2487
/gæ/	F1	18	368	63	606	630
	F2	19	2112	71	1771	1756
	F3	27	2416	49	2456	2441
/bɑ/	F1	2	395	24	646	702
	F2	2	1069	21	1069	1083
	F3	3	2425	38	2590	2646
/dɑ/	F1	9	392	38	669	710
	F2	9	1660	65	1166	1154
	F3	9	2628	30	2523	2522
/gɑ/	F1	20	384	46	645	660
	F2	20	1733	69	1194	1201
	F3	22	2367	35	2512	2518

Tableau 4.2 : Transitions des formants F(I), F(2)et F(3)pour les consonnes /b/,/d/,/g/ avec différentes voyelles(pour l'anglais.)

Suite du tableau 4 - 2.

		Time of onset relative to burst ms	Onset freq, Hz	Trans. duration ms	Vowel steady-state freq. Hz	Freq. 95 ms from burst Hz
/bo/	*F*1	1	378	12	440	465
	*F*2	1	1127	18	1079	993
	*F*3	2	2379	24	2434	2483
/do/	*F*1	11	350	29	441	491
	*F*2	11	1635	79	1190	1179
	*F*3	11	2549	74	2408	2419
/go/	*F*1	26	322	42	447	459
	*F*2	26	1474	67	1134	1142
	*F*3	26	2274	19	2265	2380
/bu/	*F*1	4	327	5	354	341
	*F*2	4	1064	22	1012	983
	*F*3	4	2300	15	2287	2358
/du/	*F*1	11	301	17	327	344
	*F*2	11	1701	34	1536	1340
	*F*3	11	2430	27	2257	2252
/gu/	*F*1	25	334	4	328	351
	*F*2	25	1275	11	1234	1101
	*F*3	27	2203	6	2154	2230

Three-segment *F*3 transition for /do/.

		Onset A	Point B	Point C	SS D
*F*3	Hz	2549	2479	2245	2408
Length	ms		21	10	43

Pour pouvoir effectuer effectuer de la synthèse, il va falloir disposer, de la même façon qu'en paragraphe (4-2-2) mais cette fois-ci non plus à un niveau abstrait (niveau phonétique), mais concrètement d'élèments minimaux acoustiques permettant de constituer n'importe quelle phrase. L'utilisation d'un élément correspondant au phonème suppose que l'on saura reconstituer toutes les transitions lorsqu'il sera en contact avec un autre phonème (synthèse par règles). Cette solution, si séduisante présente cependant de grandes difficultés de réalisation.

Le choix du diphone, ou diphonème ou encore phonatome, consiste à contourner les règles de composition, en intégrant dans les unités choisies, les zônes de transition, et en effectuant le raccordement sur les parties stables. L'assemblage des unités de parole en message est réalisé par simple juxtaposition des unités. Ce sont les résultats des recherches des laboratoires Haskins, pendant les années 50, qui ont conduit à la conception de cette procédure..

- La parole n'est pas constituée d'éléments discrets facilement segmentables en unités fractionnelles, mais se présente sous la forme d'un continuum sonore.

Les rapides changements dans les fréquences formantiques des voyelles au contact des consonnes (transitions) ne constituent pas un phénomène accessoire, mais l'un des indices essentiels pour la perception des syllabes.

L'analyse spectrale permet d'observer dans la réalisation des séquences "consonne - voyelle - consonne" des mouvements de relative stabilité spectrale se situant dans la partie centrale des réalisations vocaliques et consonantiques, et des moments de grande instabilité dans le passage d'une réalisation phonétique à une autre (fig. 4 - 6). La solution consiste donc à isoler et à choisir comme élément de parole, une séquence allant d'une partie stable à une autre, d'où on tire la définition suivante : ON APPELLE DIPHONEME LE SEGMENT QUI S'ETEND DE LA ZONE STABLE D'UNE REALISATION PHONETIQUE A LA ZONE STABLE DE LA REALISATION SUIVANTE ET QUI PROTEGE EN SON CENTRE TOUTE LA ZONE DE TRANSITION. (fig. 4 - 6).

En conséquence, cette méthode nécessite un nombre d'éléments de parole bien plus élevé que pour les méthodes de synthèse par règles.

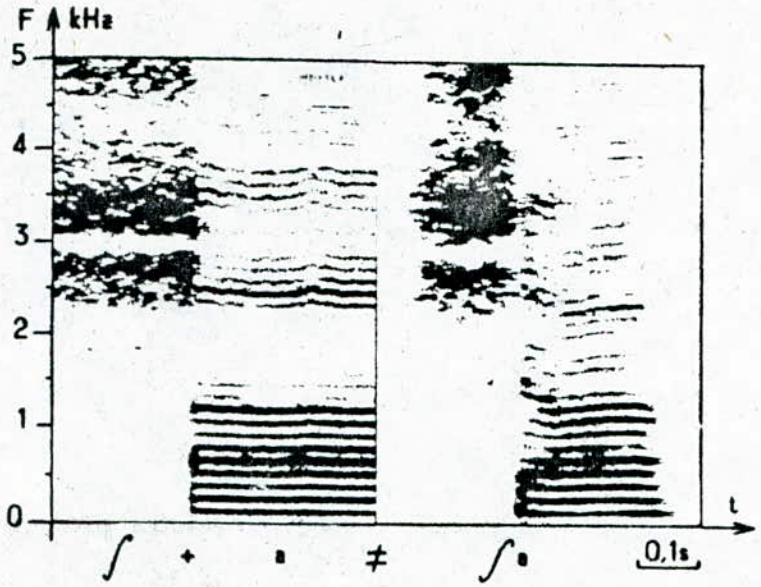


Fig. 4 - 5 : Réalité des transitions phonétiques dans la parole : la juxtaposition de (ʃ) et (a), par collage de bande magnétique, ne fournit pas la syllabe (ʃa) : CH + A = CHA.
(D'après LIENARD, 1977)

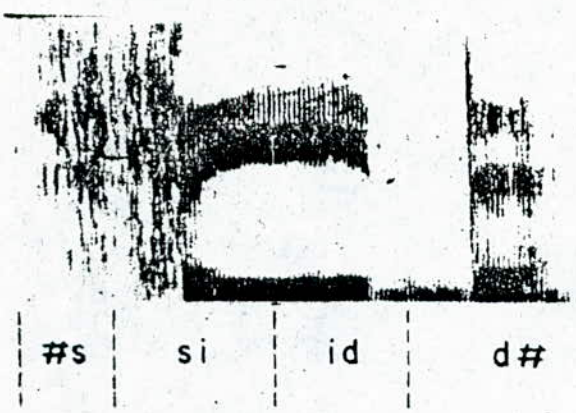


Fig. 4 - 6 : Segmentation du mot seed en diphones //=/s/, /si/, /id/ et /d/=/ / (d'après PETERSON et al., 1958).

En effet, si l'on veut composer n'importe quel message, il faut disposer de toutes les combinaisons deux à deux, c'est-à-dire que pour la synthèse d'une langue comme le français qui compte quelques 33 réalisations phonémiques, il faut envisager de pouvoir disposer théoriquement de $(33)^2$ diphonèmes.

Du point de vue pratique, on doit constituer un dictionnaire de diphones ; pour cela, on suit les étapes suivantes :

- Choix du corpus de mots ;
- Choix du locuteur ;
- Enregistrement de la liste des mots ;
- Numérisation du signal enregistré ;
- Analyse (en prédiction linéaire par exemple) ;
- Segmentation en diphones.

a) Choix du corpus :

La sélection de l'ensemble des mots dépend de la langue qu'on veut synthétiser (nombre de phonèmes,...). Cet ensemble est choisi d'une certaine manière pour contenir toutes les combinaisons possibles des phonèmes dont on éliminera celles qui n'apparaissent pas réellement.

b) Choix du locuteur :

On enregistre les mots du corpus lus par plusieurs locuteurs. Après des tests d'intelligibilité en analyse-synthèse, on retiendra le locuteur donnant les meilleurs résultats.

c) Enregistrement et numérisation :

Les mots lus par le locuteur sont enregistrés sur bande magnétique. L'enregistrement est numérisé pour pouvoir le traiter ultérieurement.

d) Analyse du corpus :

Tous les mots numérisés subissent une analyse pour en tirer les paramètres importants (pitch, gain en LPC, formants,...) qu'on peut visualiser sur des graphes ou sur tableaux numériques d'une analyse par vocodeur à canaux (fig. 4 - 7).

e) Segmentation :

Pour segmenter les diphones, soit on cherche les zônes stationnaires de la fonction de stabilité spectrale, le diphone est alors extrait entre les minimas les plus marqués de cette fonction, soit à partir des

échantillons du vocodeur à canaux, car la méthode est plus précise que celles des courbes (visualisation de la durée) fig. 4 - 7).

Remarque : Une synthèse réduite à une simple juxtaposition d'éléments acoustiques minimaux (phonème ou diphonème) restitue une parole qui manque de "naturel" (cas des phrases énonciatives, interrogatives ou impératives).

Une analyse de la "prosodie" permet d'augmenter la qualité de la parole synthétique. Par prosodie, on entend phénomènes accentués, intonatifs (intensité, durée des pauses, variation de la fréquence fondamentale,..) On peut établir ainsi des "règles" pour les variations de "F_0" par exemple et les durées de pause pour compléter une synthèse par diphones.

CONCLUSION :

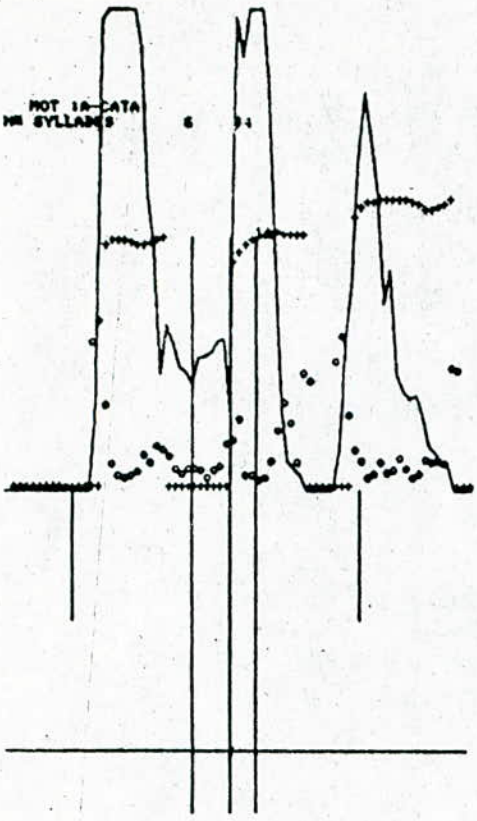
Dans ce chapitre, nous avons essayé de présenter les différentes techniques de synthèse, selon que nous voulons transmettre l'information à faible débit (vocodeur à canaux) ou que nous essayons de copier le mode de production de la parole par l'appareil vocal, c'est-à-dire utiliser les caractéristiques acoustiques de la parole (vocodeur à formants), ou enfin, que nous modélisons la fonction de transfert du conduit vocal (vocodeur à prédiction linéaire). Vu leurs principes différents, ces synthétiseurs présentent des avantages et des inconvénients (cf. tab. 4 - 1).

A l'aide de tous ces synthétiseurs, nous pouvons obtenir une bonne qualité de la parole si nous choisissons la bonne méthode(cf.tab.4-3).

Nous avons insisté sur la méthode de synthèse par diphones, qui consiste à concaténer des diphones tels qu'un diphone doit englober dans son spectre la région de transition. Cette méthode est la plus prometteuse pour la production de la parole synthétique.

Par ailleurs, le tableau 4-4 illustre les différentes applications de la synthèse dans le dialogue homme/machine.

-------oOo-------



21465	14	0	5	3	3	3	3	3	3	3	4	3	4	3	3	3	3	3
21466	12	0	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
21467	15	0	5	3	3	3	3	4	3	3	4	3	3	3	3	3	4	3
21468	309	0	7	7	10	10	10	11	11	10	10	11	11	9	7	7	8	8
21469	1403	96	11	11	14	14	13	13	16	13	13	14	14	11	10	10	10	11
21470	2291	97	14	14	16	15	13	13	16	14	13	14	14	12	11	11	12	12
21471	2061	97	14	14	16	15	13	13	14	14	13	14	14	11	11	11	12	12
21472	2004	97	14	14	16	14	13	13	14	13	12	13	15	13	11	11	12	11
21473	1849	96	14	14	16	14	13	13	14	14	13	13	15	11	10	10	11	11
21474	1679	96	14	14	17	14	13	12	14	14	13	13	14	12	10	10	10	10
21475	1400	96	14	15	16	13	12	11	12	13	12	13	13	11	10	10	10	11
21476	939	96	14	15	14	12	11	11	13	12	12	12	12	10	8	8	9	12
21477	767	97	14	15	12	11	10	10	12	13	12	12	12	10	8	8	8	10
21478	376	98	13	12	12	9	8	9	10	11	10	9	10	10	8	7	8	8
21479	526	0	8	9	12	10	8	10	11	11	11	10	11	11	11	10	9	9
21480	469	0	8	9	12	9	9	10	10	11	11	11	12	10	10	10	9	9
21481	395	0	8	9	11	10	11	10	10	10	10	11	10	10	9	8	9	10
21482	378	0	8	8	9	8	9	10	10	10	12	11	11	10	8	8	9	11
M 21483	342	0	7	7	9	9	9	9	9	9	10	10	11	10	9	9	10	11
21484	420	0	7	7	8	8	9	9	10	11	10	10	11	11	10	10	11	12
21485	424	0	7	7	8	9	9	9	10	11	10	11	11	10	10	11	12	12
21486	438	0	7	7	7	8	9	9	9	11	11	11	12	11	10	11	11	11
21487	475	0	9	8	8	8	9	9	10	11	11	12	11	11	10	10	10	11
21488	481	0	9	9	9	9	9	10	12	12	11	11	12	10	9	8	10	11
M 21489	249	88	11	12	10	8	8	9	10	9	9	9	10	7	6	6	7	9
21490	1495	98	13	15	14	13	12	12	13	13	12	13	13	11	10	10	11	12
21491	1278	96	14	14	15	12	12	11	12	13	12	12	15	11	10	10	11	11
21492	1670	97	13	14	16	13	12	12	13	13	13	13	14	11	10	10	11	11
M 21493	1584	98	13	14	16	13	12	12	14	13	13	13	13	11	10	10	12	11
21494	1674	99	13	14	16	13	12	12	15	13	12	13	14	11	11	11	12	12
21495	1300	99	14	14	15	13	12	12	14	13	12	12	13	11	10	10	11	10
21496	796	100	14	14	14	11	10	11	13	12	11	11	13	9	8	8	9	10
21497	870	98	12	13	11	9	8	8	11	9	8	8	12	7	6	6	7	9
21498	94	99	8	9	11	8	6	6	8	8	7	8	8	6	4	4	6	4
21499	68	99	7	7	11	7	6	5	7	7	6	7	7	5	4	4	5	4
21500	50	99	6	7	11	6	5	4	6	7	5	6	6	5	4	4	4	4
21501	43	99	7	8	10	6	4	4	5	6	6	6	6	4	3	3	4	3
21502	43	99	6	7	9	6	5	5	6	5	5	6	4	4	4	4	4	4
21503	35	99	6	8	10	6	4	5	6	5	4	5	6	4	3	3	4	3
21504	33	99	5	6	10	6	5	4	5	6	4	4	4	4	4	4	4	3
21505	29	100	5	6	8	5	4	4	5	5	5	5	5	4	3	3	3	3
21506	128	0	6	7	9	7	8	9	8	8	7	7	8	9	7	7	7	6
21507	478	0	9	10	11	9	10	11	11	11	10	11	12	10	9	8	9	9
21508	855	106	13	14	14	12	11	11	13	11	10	12	10	8	8	8	9	9
21509	1050	110	13	13	15	12	11	12	13	12	11	12	12	10	10	10	11	10

Fig.4.7: Résultat de l'analyse LPC du mot /=/ A-CATA /=/
(D'aprés M. Guerti , 1983)

\- : Gain LPC

xx : Période fondamentale

oo : Stabilité spectrale

++ : Dérivée du gain .

A l'aide de ce résultat, on peut réaliser la segmentation en diphones du mot /=/ A-CATA /=/, en se basant sur la recherche des zones stationnaires de la fonction de stabilité spectrale.

Le diphone est extrait entre les minima de cette fonction(les plus marqués)pour les deux phonèmes qui le constituent. Son milieu est pris au maximum de la fonction de stabilité.

SYNTHETISEURS ACOUSTIQUES	AVANTAGES	INCONVENIENTS
A CANAUX	- Bonne qualité d'analyse-synthèse - Source et fonction de transfert séparées	- Sonorité assez rocailleuse - Intégration difficile - Qualité variable selon le locuteur
A FORMANTS	- Parole plus naturelle que celle produite par des vocodeurs à canaux - Grande souplesse de variation des paramètres de la source et de la fonction de transfert - Source et fonction de transfert séparées	- Sonorité dans les transitions parfois un peu floue - Qualité variable selon le locuteur - paramètres de commande difficiles à obtenir par une analyse automatique
A PREDICTION LINEAIRE	- Source et fonction de transfert séparées - Théorie mathématique se prêtant mieux aux simulations informatiques - Elimination de la redondance existant dans la forme temporelle du signal de parole. - Intégration facile (Texas Instruments, C.N.E.T., Hitachi, Matsushita, I.T.T,..) - Paramètres obtenus par une analyse automatique. - Simplicité et rapidité d'exécution des algorithmes.	- Manque d'intelligibilité dans certains sons.... (nasals, sonores, liquides) dû au principe même de la méthode. - Qualité variable selon le locuteur

TABLEAU 4 - 1 : Les avantages et les inconvénients des 3 principaux synthétiseurs.

	METHODE DE SYNTHESE	AVANTAGES	INCONVENIENTS	TECHNIQUES UTILISEES
ACOUSTIQUE	Phrases	- parole naturelle - Méthode adéquate pour certaines applications	- Le nombre de phrases synthétiques est limité	- Prédiction linéaire - Vocodeur à canaux
	Mots	- Simple - Qualité de la parole bonne (si les problèmes de coarticulation entre les mots juxtaposés sont pris en compte)	- Peu économique car elle suppose la mise en mémoire préalable de tous les mots du vocabulaire sous forme codée. - Les mots se succèdent avec l'intonation et la durée qu'ils avaient au départ au moment de l'enregistrement. Ce caractère donne un aspect très artificiel à la parole réalisée par concaténation (J.VAISSIERE,1975). - Vocabulaire limité.	- Prédiction linéaire - Vocodeur à canaux
PHONETIQUE	Règles	- parole assez naturelle - En ce qui concerne les transitions, la mélodie et le rythme elle donne une parole de bonne qualité - Facilité avec laquelle la machine pourra répondre aux besoins de n'importe quel utilisateur - Vocabulaire illimité	- Grandes difficultés de réalisation des règles. - Exige davantage de temps de calcul. - Qualité moins bonne que la synthèse acoustique.	- Synthétiseur à formants
	Diphones	- Qualité bonne. - Synthèse assez simple à réaliser. - Possibilité de reconstituer n'importe quel texte d'une langue donnée. - Vocabulaire illimité.	- Nombre d'éléments plus élevé que pour la synthèse par règles (1356 pour l'arabe standard) 1156 pour le français). - Qualité moins bonne que la synthèse acoustique	- Prédiction linéaire - Vocodeur à canaux - Synthétiseur à formants.

TABLEAU 4 - 3 : Les avantages et les inconvénients des méthodes de synthèse.

INDUSTRIELS	- Machines complexes (outils, manutentions,...) - Ascenseurs - Chambre noire - Alarme et sécurité - Mesures - Etats de stocks ou vérification d'état de fichiers - Vente par correspondance
TELECOMMU-NICATION	- Répondeurs et enregistreurs automatiques - Télé-surveillance, télé-signalisation, télé-information - Banque d'informations (journaux, rapports, reportages) - Indications de taxes - Aide aux handicapés - Renseignements téléphoniques - Messages d'alarme ou de déroutage dans les centraux téléphoniques
AUTOMOBILE	- Tableaux de bord - Diagnostics, entretien, dépannage, mode d'emploi - Guidage routier
GRAND PUBLIC	- Machines à laver - Fours et cuisinières - Réveils automatiques - Jeux - Ordinateurs domestiques - Traducteurs de poche parlants.
INFORMATIQUE	- Terminaux de transactions, terminaux bancaires, terminaux d'ordinateurs,...
MILITAIRE/ AERONAUTIQUE	- Calculateur de vol, système d'approche, météo, informations passagers et équipage, systèmes de tirs (chars).
TERTIAIRE	- Publicités, annonces étiquetages, transports en public (Air, Terre, Mer). - Météorologie...
AIDE AUX HANDICAPES	- Machine à lire pour aveugles - Annonces parlées - Utilisations domestiques - renseignements parlés.

Tableau 4.4 : Application de la synthèse dans le dialogue Homme/machine.

CONCLUSION GENERALE

L'objectif de notre travail était de rassembler brièvement les notions et les techniques rencontrées dans la synthèse de la parole.

Parmi les procédés étudiés la prédiction linéaire, d'application assez récente, bien que la notion de prédiction date de Gauss(1795), est traitée par les différents auteurs avec une importance considérable.Cela est dû aux avantages qu'elle présente. En effet, en prédiction linéaire, la théorie mathématique se prête mieux aux simulations informatiques, un grand nombre de paramètres sont obtenus par une analyse automatique. D'un point de vue pratique, son intégration est facile.

Nous avons étudié aussi les différentes méthodes pour synthétiser un message, que ce soit à partir de phrases, ou de mots, ou encore d'éléments plus petits qui sont le phonème et le diphone selon l'application visée.

Dans le cas de synthèse par phrases ou par mots, si la parole réstituée est d'excellente qualité, le nombre de message qu'on peut générer cependant est trés limité. Par contre, ayant les phonèmes comme élément phonétique, la synthèse par règles permet de composer n'importe quelle phrase dés que les règles de transitions de formants sont établies. La production de n'importe quel message de la langue étudiée est possible aussi avec la méthode utilisant le diphone mais d'une manière beaucoup plus facile qui consiste à contourner les transitions formantiques contenues dans la proximité des phonèmes. Ces deux dernières procédures aboutissent à la synthèse à partir du texte écrit.

L'étude de ce sujet nous a permis d'aborder les notions du traitement du signal, domaine si attrayant vu ses champs d'application auxquels la technologie moderne doit une grande part, et demeure cependant assez délicat sachant qu'il est basé essentiellement sur des méthodes mathématiques d'un niveau élevé .

Dans ce présent travail, nous n'avons fait qu'effleurer certaines notions de base, en relation directe avec le traitement du signal de la parole, autour desquelles nous étions obligés de nous réstreindre car le temps ne le permet pas , et sans oublier la difficulté d'accés à la bibliographie . Nous avons parcouru donc les notions de phonétique acoustique, les techniques d'analyse et de synthèse, et enfin les méthodes avec lesquelles les segments de la parole sont assemblés.

Il aurait été préférable d'avoir accés au domaine du traitement de la parole mais à un niveau plus concrèt afin de développer des applications d'une utilité bien définie. Nous souhaitons que ceci soit possible dés que l'ECOLE POLYTECHNIQUE sera dotée d'un laboratoire spécialisé.

——oOo——O——oOo——

A N N E X E

---oOo---

(1)

1. Algorithme de décision voisé/non voisé :

La méthode fait appel au calcul de la fonction d'autocorrélat du signal d'erreur e(t) obtenu par filtrage inverse du signal de parole. l valeur R(i) doit satisfaire à une condition qui permet de décider si la séquence est voisée ou non. Cette condition est que R(i) ne dépasse un se α. Une décision relative à un cadre donné n'est définitive qu'aprés l'an lyse des deux cadres consécutifs.

On trouvera, en fig.1, l'organigramme de l'algorithme utilisé pour parvenir à une décision voisé/non voisé et l'incidence de cette déci sur la période de mélodie τ.

2. Algorithme de " DURBIN " pour la résolution d'un système de "p"équations à "p" inconnues :

La méthode de " DURBIN " assure une rapidité d'exécution et évite l'encombrement mémoire; on va l'utiliser pour résoudre le système d'équations d'autocorrélation. Elle utilise p(p+1) opérations.

Les équations normales d'autocorrélation ont la forme matricielle suivante :

$$\begin{bmatrix} R(o) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(o) & R(1) & \cdots & R(p-2) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(o) \end{bmatrix} \cdot \begin{bmatrix} a(1) \\ a(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ R(p) \end{bmatrix}$$

L'algorithme de "DURBIN" est le suivant :

F(o) = R(o)

Pour k variant de 1 à p faire :

(1) : On l'appelle aussi Algorithme de SIFT

Début

$$C(k) = -\Big(R(k) - \sum_{i=1}^{k-1} a^{(k-1)}(i) \cdot R(k-i) \Big) / E(k-1)$$

$$a^{(k)}(k) = C(k)$$

$$a^{(k)}(i) = a^{(k-1)}(i) + C(k) \cdot a^{(k-1)}(k-i) \qquad 1 \leqslant i \leqslant k-1$$

$$E(k) = (1 - C^2(k)) \cdot E(k-1)$$

Fin

p : nombre de coefficients du filtre ;

a (k) : coefficients du filtre ;

E (k) : erreur de prédiction à l'ordre "k" ;

C (k) : coefficients de corrélation partielle ou coefficients de réflexion

R (k), $0 \leqslant k \leqslant p$: coefficients d'autocorrélation.

3. Algorithme de calcul de la fonction d'aire :

Les étapes de calcul de la fonction d'aire sont les suivantes :

a) Soit Sn la séquence du signal de parole à analyser. La longueur de la séquence dépend du mode d'analyse. Cette séquence est obtenue par l'échantillonnage du signal de parole s (t) à une cadence Fe. Avant l'échantillonnage, il faut filtrer le signal afin d'éviter le problème de repliement du spectre. A cet effet, nous avons utilisé un filtre analogique de Butterwort d'ordre douze dont l'affaiblissement à GdB se situe à Fe/2.

b) L'estimation du nombre, N, de sections nécessaires pour représenter la forme du conduit vocal.

c) L'application de l'égalisation selon la stratégie adoptée (on a le choix suivant les contraintes d'application, entre une stratégie adaptée qui adme une adaptation préalable au locuteur, ou une stratégie générale qui, à défa de pouvoir s'adapter à tous les locuteurs, puisse conduire à des configurations réalistes ou du moins distinctes pour la grande majorité des locuteur possibles).

d) Calcul des coefficients de réflexion, ki, du conduit vocal sur le signal égalisé. Nous avons appliqué tous les algorithmes de prédiction linéaire à des fins de vérification.

e) A partir des N valeurs de ki, on peut calculer les aires de section Ai :

$$k(i) = \frac{A(i) - A(i+1)}{A(i) + A(i+1)}$$

d'où $A(i) = \frac{1 + k(i)}{1 - k(i)} \cdot A(i+1)$; $i = N, N-1, \ldots, 1$.

Du fait que les aires sont calculées de manière relative, la fonction d'aire est obtenue en posant $A(n+1) = 1$.

f) La normalisation de la fonction d'aire que nous avons trouvée être la plus représentative consiste à normaliser la valeur maximale des $A(i)$ à 8 cm^2 . (D'aprés EL MALAWANY, 1975).

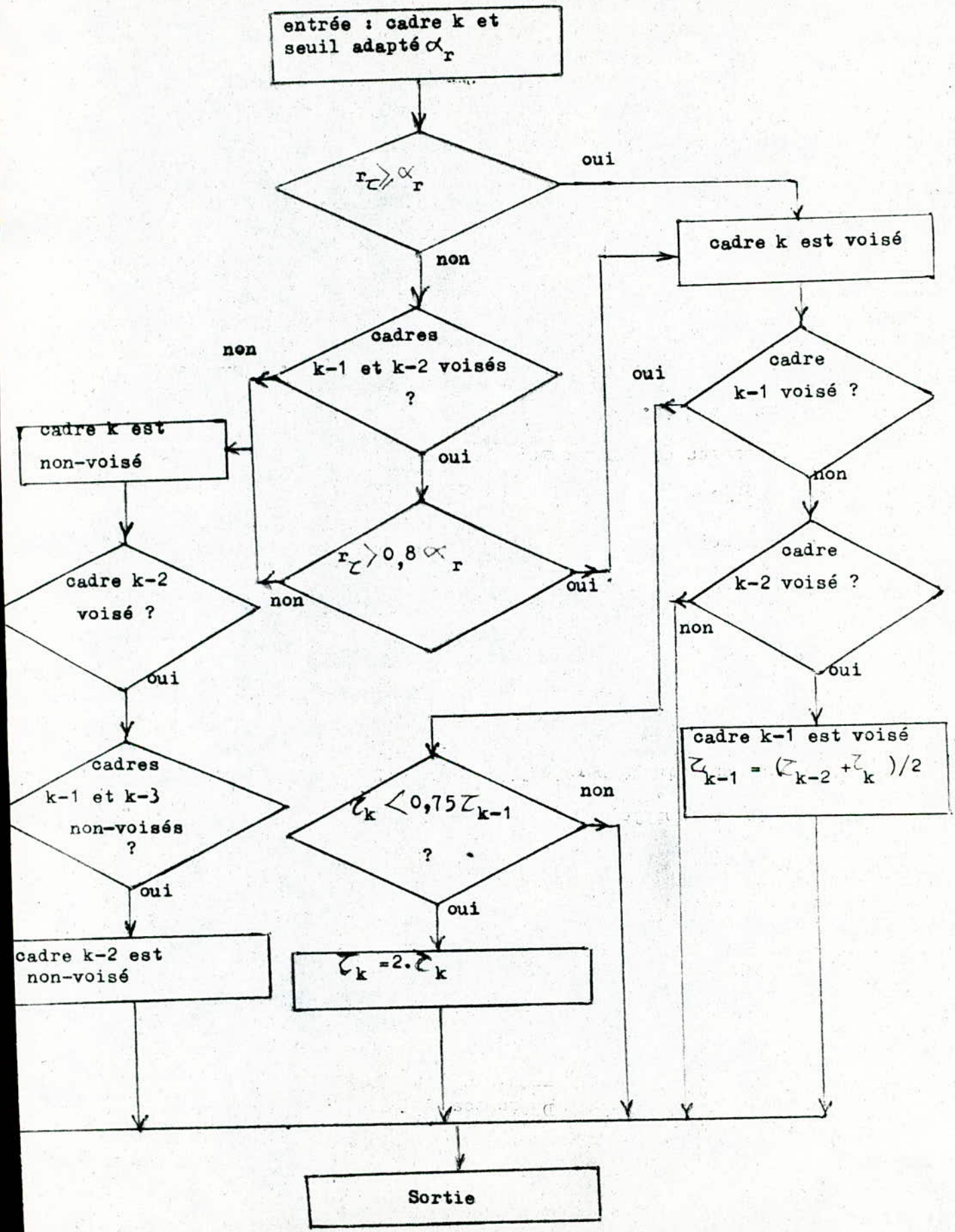


Fig. 4 : Organigramme de l'algorithme de décision voisé / non-voisé .

BIBLIOGRAPHIE

- M. BELLANGER (1981), "Traitement numérique du signal". ENST

- R. DESCOUT (1982), "Les techniques de synthèse de la parole". Documentation bibliographique regroupée; CNET-LANNION.

- I. EL MALAWANY (1975), "Contribuations aux recherches sur la communication parlée : Etude de vocodeurs à prédiction linéaire. Détermination de l'intervalle de fermeture de la glotte. Détection de la mélodie. Extraction de la fonction d'aire du conduit vocal". Thèse Docteur-Ingénieur, Université scientifique et médicale, GRENOBLE.

- F. EMERARD (1977), "Synthèse par diphone et traitement de la prosodie". Thèse de Doctorat 3e cycle, GRENOBLE.

- E. EMERIT (1977), "Cours de phonétique acoustique" Ed. SNED.

- M. GUERTI (1983), "Contribution à la synthèse de la parole en arabe standard". Thèse de magister, Université d'ALGER.

- J. GUIBERT (1979) "La parole, compréhension et synthèse par les ordinateurs" - Ed. PUF.

- D. KEWELEY (1982), "MEASUREMENT OF FORMANT TRANSITIONS" - J. Acoust. Soc. America.

- J. S. LIENARD (1977), Processus de la communication parlée, Introduction à l'analyse et à la synthèse de la parole". Ed. Masson.

-.J. D. MARKEL -A. H. Gray (1976), "Linear prediction of speech". Springer Verlag

- J. Max (1981), "Méthodes et techniques de traitement du signal et applications aux mesures physiques," tome 1, Ed. Masson.