

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

37/87

وزارة التعليم و البحث العلمي
Ministère de l'Enseignement et de la Recherche Scientifique

20x

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE - المكتبة
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

SUJET

SIMULATION

de L'ANALYSE de la PAROLE

SUR VAX 11/750

Proposé par :

M^r Beniddir N.

Etudié par :

N. Ouidah
H. Aoudia

Dirigé par :

M^r N. Beniddir

PROMOTION JUIN 87

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم و البحث العلمي
Ministère de l'Enseignement et de la Recherche Scientifique

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

SUJET

SIMULATION
DE L'ANALYSE DE LA PAROLE
SUR VAX 11/750

Proposé par :

M^r N. BENIDDIR

Etudié par :

N. OUIDAH
H. AOUZIA

Dirigé par :

M^r N. BENIDDIR

PROMOTION JUIN 87

oOo Remerciements oOo

Nous tenons à remercier notre promoteur, Mr. BENIDDIR N.
qui nous a aidés tout le long de l'élaboration de ce projet
de fin d'études.

Nous remercions également Melle Y. MOUSSAOUI pour la documentation
qu'elle nous a fournie.

La frappe de ce polycopé a été soigneusement faite par Melle
BAYOU F.Z. (Hina) , qu'elle trouve ici l'expression de notre
grande reconnaissance.

Nous tenons à remercier tous ceux qui ont participé de près ou
de loin et qui ont contribué à la réalisation de cet ouvrage.

oOo Dédicaces oOo

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Je dédie ce Mémoire :

- A ma très chère mère qui s'est tant sacrifiée pour moi,
- A mon frère Abdelkader qui n'a cessé de m'encourager,
- Et à tous ceux qui me sont chers.

Nouria,

Je dédie cet ouvrage :

- A ma mère,
- A mes oncles,
- A tous mes amis

Hamid,

P L A N



INTRODUCTION

PARTIE -1 -

Chapitre 1.

La phonétique

1. - Introduction
2. - L'appareil phonatoire humain
- 3 - Fonctionnement et production des différents sons
- 4 - Caractéristiques de la parole
- 5 - Les différents types de sons
- 6 - Conclusion.

Chapitre -2 -

Traitement numérique des signaux

- 1 - Définition du signal
- 2 - Série de Fourier
- 3 - Transformée de Fourier
- 4 - Transformée en Z
- 5 - Transformée inverse de Fourier
- 6- Corrélation des signaux
- 7 - Systèmes causals
- 8 - Systèmes linéaires
- 9 - Systèmes linéaires invariants
- 10 - Produit de convolution
- 11 - Echantillonnage et codage
- 12 - La transformation de Fourier discrète (T.F.D.)
- 13 - Transformée de Fourier inverse
- 14 - Transformée de Fourier rapide (F.F.T)
- 15 - Les fenêtres de troncature
- 16 - Les filtres numériques

Chapitre -3 -

Analyse

- 1 - Introduction
- 2 - Pré-traitement
- 3 - Méthodes d'analyse
- 4 - Analyse par prédiction linéaire
 - RAPPEL : Transformation homomorphique
- 5 - Analyse cepstrale.

PARTIE -2 - :

Programmation

- A - Organigrammes
- B - Programmes

Simulation

- Généralités
- Programme

CONCLUSION

RECOMMANDATIONS

ANNEXE

BIBLIOGRAPHIE

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

INTRODUCTION



INTRODUCTION /

Depuis l'antiquité, l'homme, ayant observé les différents phénomènes qui se produisaient autour de lui dans la nature, a essayé de les comprendre, les expliquer, allant même jusqu'à les reproduire artificiellement.

Dans le domaine de la parole, beaucoup de choses restent à faire, surtout dans le domaine de la reconnaissance qui est un problème très complexe.

Les chercheurs se heurtent au manque de connaissances scientifiques fondamentales sur les processus physiologiques ou psychologiques mis en œuvre pour la compréhension du message vocal et il y a beaucoup de contraintes financières et technologiques.

La parole artificielle a fait l'objet de réalisations les plus anciennes. Au XVIII^{ème} siècle, Von Kempelen réalisait une machine parlante produisant une vingtaine de sons différents. Il fallut attendre l'arrivée des ordinateurs pour traiter ce problème d'une manière plus fine.

La reconnaissance est la faculté pour une machine de reconnaître puis de réagir à la voix d'un système monolocuteur ou de plusieurs locuteurs.

De manière à représenter le phénomène physique sur un ordinateur et par souci d'épargner le temps réel, on a recours à des modèles algorithmiques très techniques.

Dans notre étude, nous allons appliquer certains de ces algorithmes (analyse par prédiction linéaire, FFT, analyse cepstrale) à la simulation de la reconnaissance de la parole, avec un programme sur le VAX.

Une étude détaillée sera faite pour chaque technique utilisée et une comparaison entre ces méthodes sera donnée d'après les résultats obtenus pour pouvoir affirmer quelle est la plus performante : Pour cela, le plan que l'on propose sera le suivant :

Après l'introduction, la première partie comprend :

- Le chapitre 1 qui sera consacré à la phonétique et à la production de la parole.
- Le chapitre 2 comprendra les outils utilisés dans le traitement du signal.
- Le chapitre 3 : les différentes méthodes d'analyse de la parole. On parlera brièvement sur l'analyse par sonographie, ensuite, on traitera en détail

.../...

l'analyse par prédiction linéaire et l'analyse cepstrale.

La 2ème partie sera réservée à la simulation, la programmation, la comparaison des résultats et une conclusion.

.....

PARTIE 1

CHAPITRE . 1

LA PHONÉTIQUE

1 INTRODUCTION /

Pour faire la reconnaissance de la parole, il faut bien connaître comment elle est produite et les différents organes qui participent dans son élaboration.

2 L'APPAREIL PHONATOIRE HUMAIN

2.1 Description :

L'appareil phonatoire est constitué des organes suivants :

- a) - **Les poumons** : jouent le rôle de générateur d'air.
- b) - **La trachée artère** : qui comporte à son extrémité supérieure le larynx, à la base duquel sont attachées "les cordes vocales" séparées par un espace appelé "glotte" qui varie au rythme de la vibration des cordes.
- c) - **Le conduit vocal**. Il se compose de deux parties :
 - Le conduit buccal qui est formé du pharynx et de la cavité buccale.
 - Le conduit nasal qui se compose de deux cavités fixes "les fosses nasales".

voir fig 1, 2

3 FONCTIONNEMENT ET PRODUCTION DES DIFFERENTS SONS

3.1 Fonctionnement :

Le système vocal humain peut se résumer comme suit :

Un générateur (poumons) alimentant la source d'excitation (cordes vocales dans le cas de sons voisés) qui produit des signaux acoustiques devant être filtrés par le résonateur (conduit vocal).

3.2 Mode de production de la parole :

Sons voisés. fig 3

Sous l'effet de la pression du flux d'air, les cordes vocales vibrent, l'air sortant de la bouche, dans le cas des voyelles, est sous forme d'impulsions quasi-périodiques. Les sons ainsi fabriqués sont les sons voisés. La vibration obtenue est la mélodie qui donne la sensation de hauteur du son "pitch".

Le son résultant de cette vibration est appelé "fondamental" et sa fréquence notée F_0 .

Le spectre de l'onde sonore obtenue par la vibration des cordes vocales est très riche en harmoniques de F_0 .

.../...

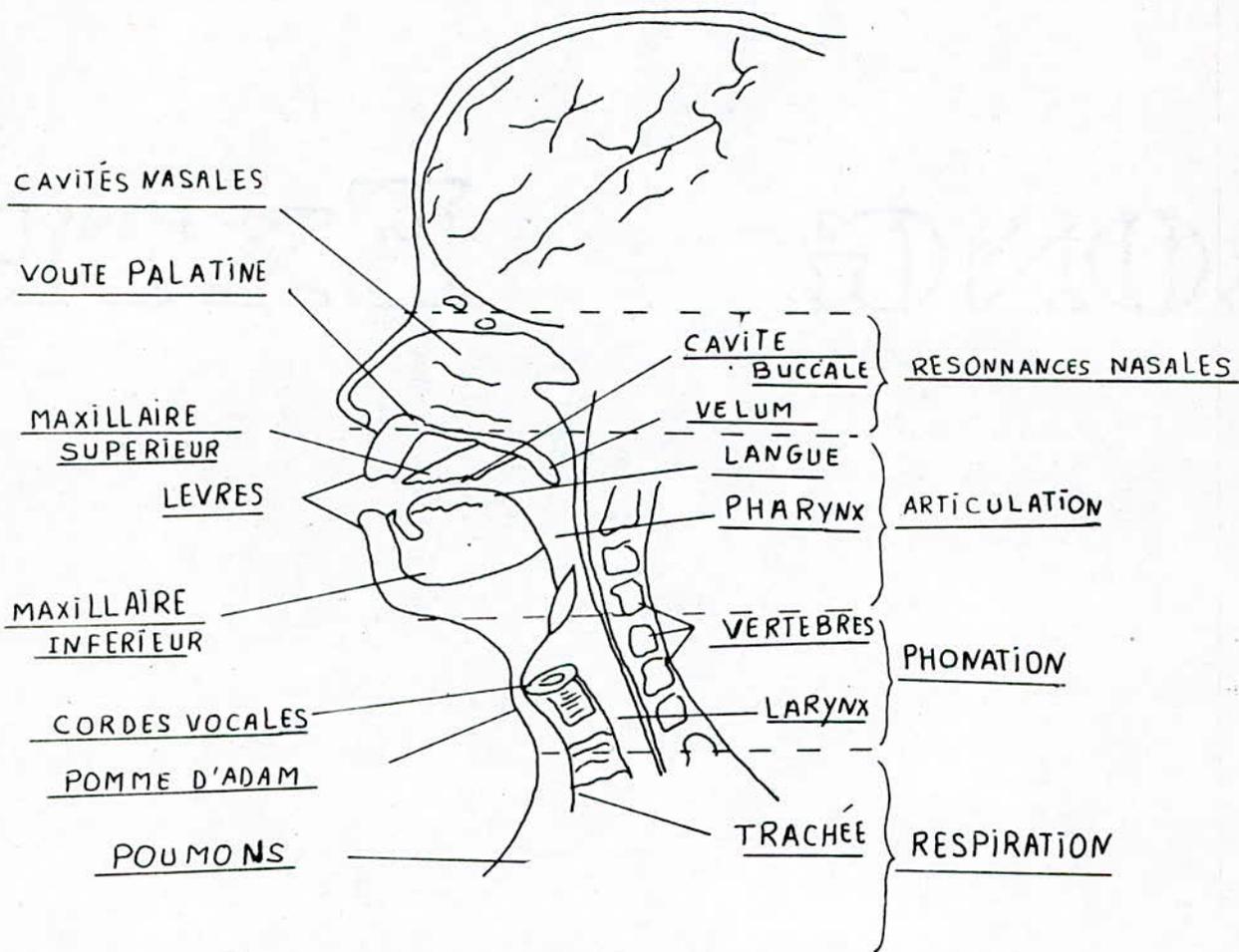


FIG1 ORGANE DE LA PAROLE

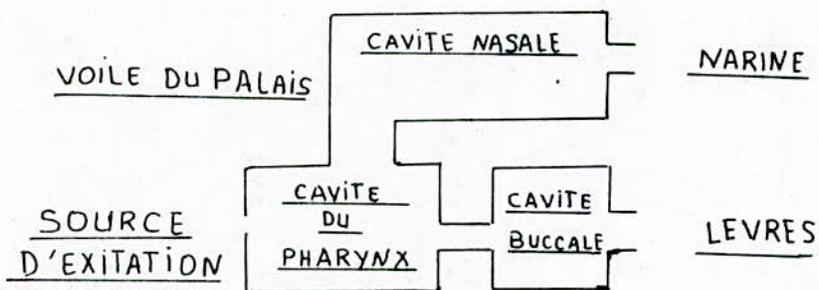


FIG2 L'APPAREIL PHONATOIRE EN TANT QUE
SYSTEME ACOUSTIQUE

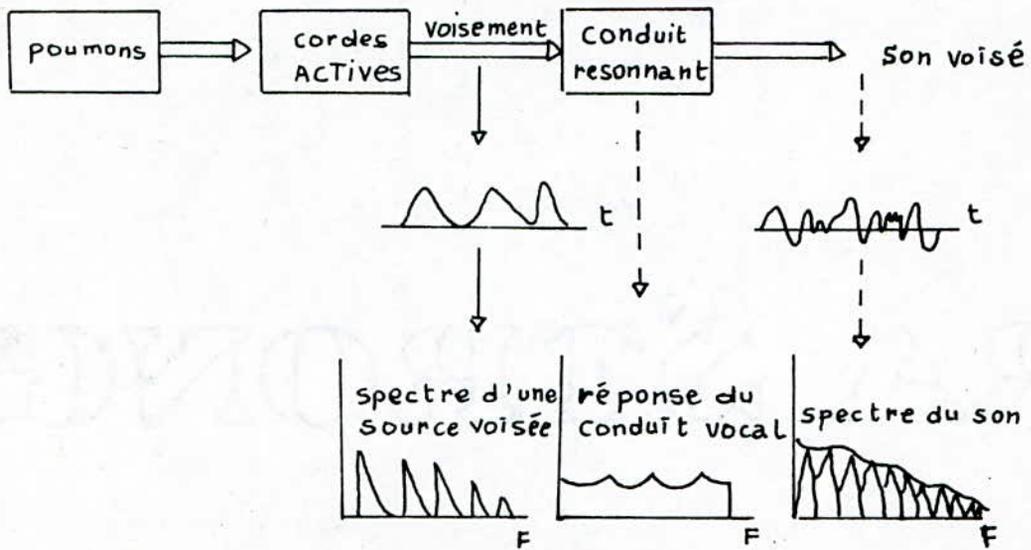


FIG 3 : PRODUCTION D'UN SON VOISÉ

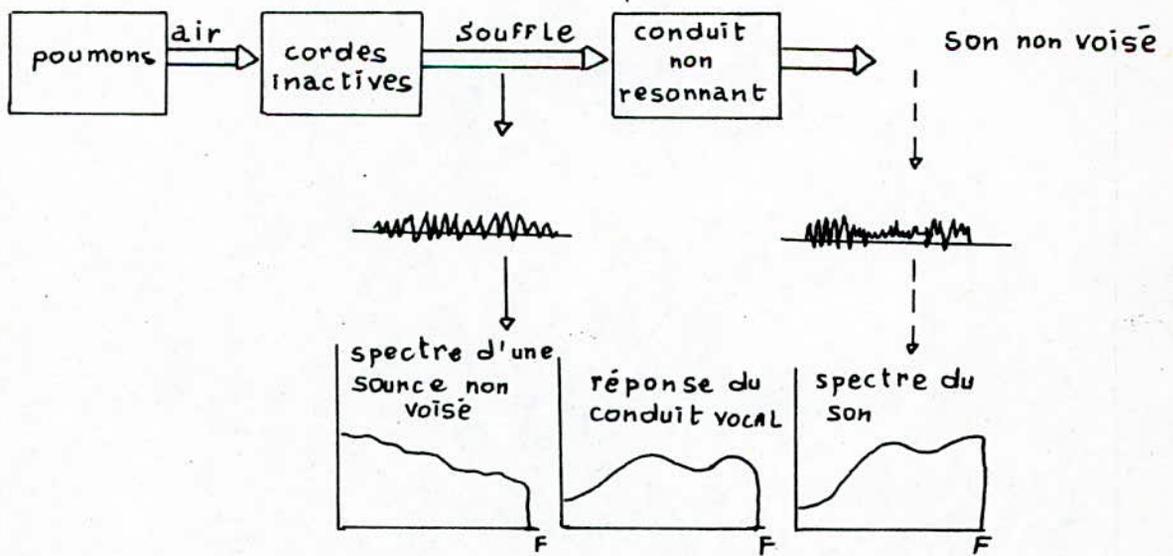


FIG 4 : PRODUCTION d'un son non voisé

Ce son va être filtré avant de parvenir à l'extérieur du conduit vocal.

Sons non voisés : FIG 4

Si les cordes vocales sont relâchées, une turbulence quasi-aléatoire est produite dans la conduite vocale par diminution de sa section. Ce sont les organes articulatoires langue et lèvres qui créent une constriction du canal à un endroit. Le signal sourd qui en résulte a l'aspect d'un bruit à l'origine blanc mais qui est coloré par les cavités résonnantes. Les signaux ainsi formés sont appelés signaux non voisés. Les cordes vocales n'entrent pas en jeu.

Ces sons voisés et non voisés sont respectivement les voyelles et les consonnes dits logatomes, leurs assemblages combinés de manière à transmettre un message constituant la parole.

Remarque

Les sons voisés sont formés par des vibrations sinusoïdales périodiques qu'on appelle harmoniques dont les fréquences d'oscillations sont des multiples entiers de la fréquence fondamentale.

Exemples de sons voisés et non voisés

Sons voisés : [a], [e], [i].

Sons non voisés : [ʒ], [s].

Les sons non voisés ne sont pas périodiques. Ils ne contiennent pas d'harmoniques mais seulement des vibrations partielles.

4 - CARACTERISTIQUES DE LA PAROLE.

On a deux paramètres fondamentaux qui caractérisent la parole :

- La fréquence des cordes vocales qui donne la mélodie de la voix.
- L'intensité des vibrations.

L'analyse acoustique de la parole permet de déterminer les fréquences et les amplitudes des harmoniques.

4.1 Timbre du son : c'est la qualité globale d'un son, qui le distingue d'un autre son de même fréquence fondamentale et de même intensité.

Le timbre dépend de plusieurs facteurs : vocalité, stabilité et la richesse.

4.2 Niveau du son : c'est l'énergie moyenne développée au silence. Elle est de 10 mW environ.

.../...

4-3 Bande passante : La transmission de la parole est fidèle entre 10 Hz et 10 KHz.

4-4 Dynamique: C'est l'écart entre le niveau maximum et le niveau minimum du signal.

$$D_{\text{son}} = [N_{\text{max}} - N_{\text{min}}] \text{ (db)}$$

L'énergie de la parole varie entre 0,1 et 1000 μW , ce qui correspond à une dynamique de 40 db.

4-5 Hauteur du son :

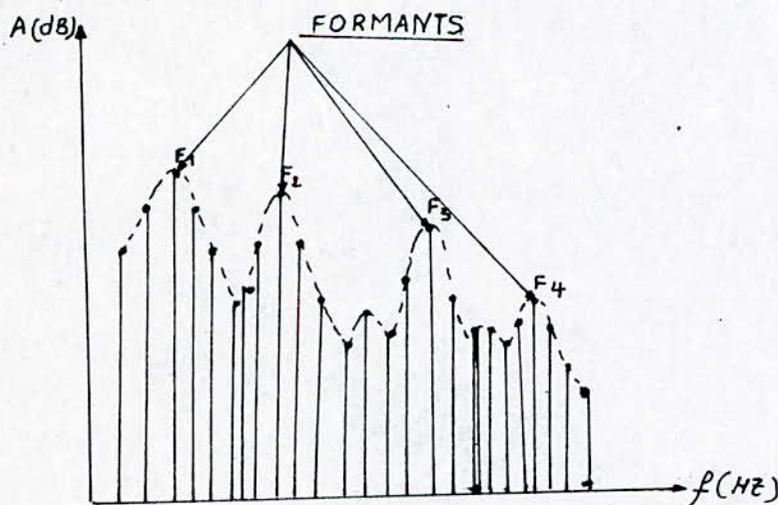
La plus basse fréquence du signal parole est la fréquence fondamentale qui sert à l'identification du locuteur. Elle est de 100 à 150 Hz en moyenne, pour les hommes de 200 à 300 hz pour les femmes et jusqu'à 400 Hz pour l'enfant et au-delà.

4-6 Les formants du son :

Pour chaque son possédant un timbre particulier, on remarque des harmoniques renforcés dans certaines bandes de fréquences particulières à chaque timbre sonore, donc à chaque voyelle : ces renforcements d'harmoniques sont les formants des sons.

Remarque :

Ce sont les fréquences des formants qui déterminent le timbre particulier de chaque voyelle.



Spectre vocalique présentant quatre résonances formantiques

Les résonances du spectre obtenu sont appelées "Formants".

F_1 , F_2 : sont caractéristiques de la voyelle prononcée.

F_3 , F_4 : sont liés au locuteur.

Remarque :

Les voyelles se répartissent dans le plan $F_1 - F_2$ conformément au triangle vocalique des phonéticiens (de P. DALATTRE) **FIG 5**

.../...

UNE CLASSIFICATION ACOUSTIQUE SOMMAIRE
DES VOYELLES PAR LA FREQUENCE MOYENNE
DE LEURS DEUX PREMIERS FORMANTS

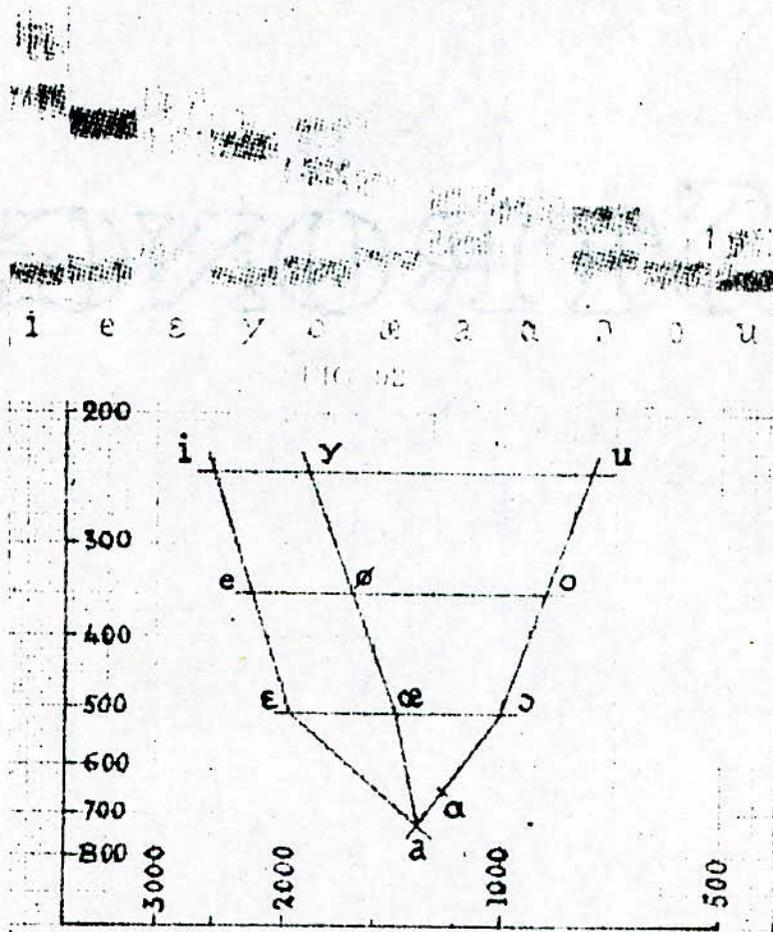
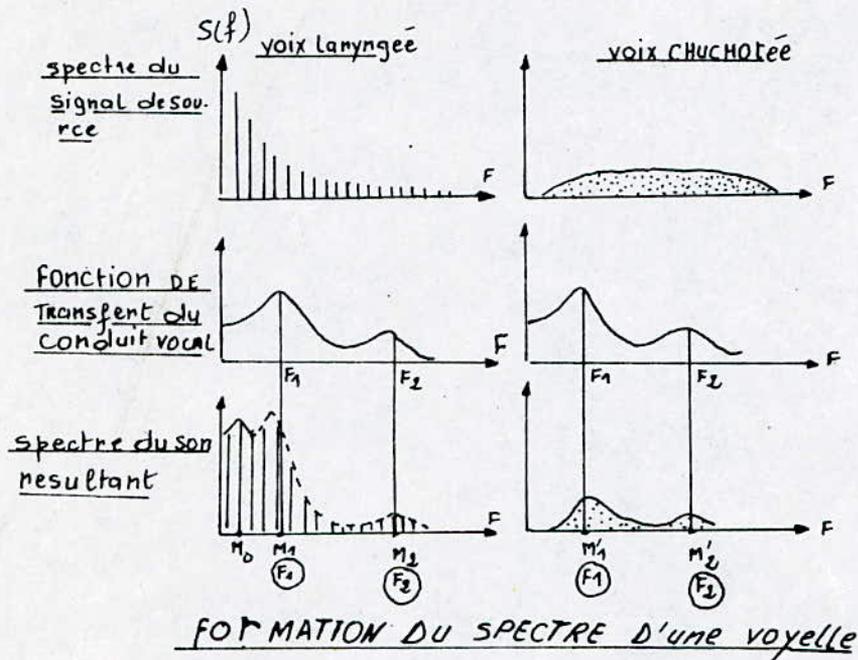


Fig. 92 : Le « triangle vocalique » de P. DELATTRE.
Ces illustrations sont reproduites de « The French Review », XXI, 6,
mai 1948, et extraites du mémorable article de P. DELATTRE : « Un triangle
acoustique des voyelles orales du français » (article réédité dans « Studies in
French and comparative phonetics », du même auteur, Mouton, The Hague
1966, p. 238).

Le spectre de raies de la source vocale $s(f)$ (cordes vocales) est multiplié par le spectre résonnant du conduit et donne le spectre du signal perçu.



5. Les différents types de sons. Fig. 6

a)- Les voyelles

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

On distingue :

- les voyelles orales ex : [a], [u]
- les voyelles nasales : sont obtenues en ouvrant le voile du palais, la communication est faite avec la cavité nasale ex. [ã], [ê].

b)- Les consonnes

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

On distingue :

- les consonnes occlusives (dites "plosives").

Elles sont produites par l'occlusion complète (momentanée) du conduit vocal suivie d'une ouverture brusque.

On a les non voisées telles : [p], [t], [k] et leurs correspondantes voisées [b], [d], [g].

.../...

CLASSIFICATION DES SONS DE PAROLE

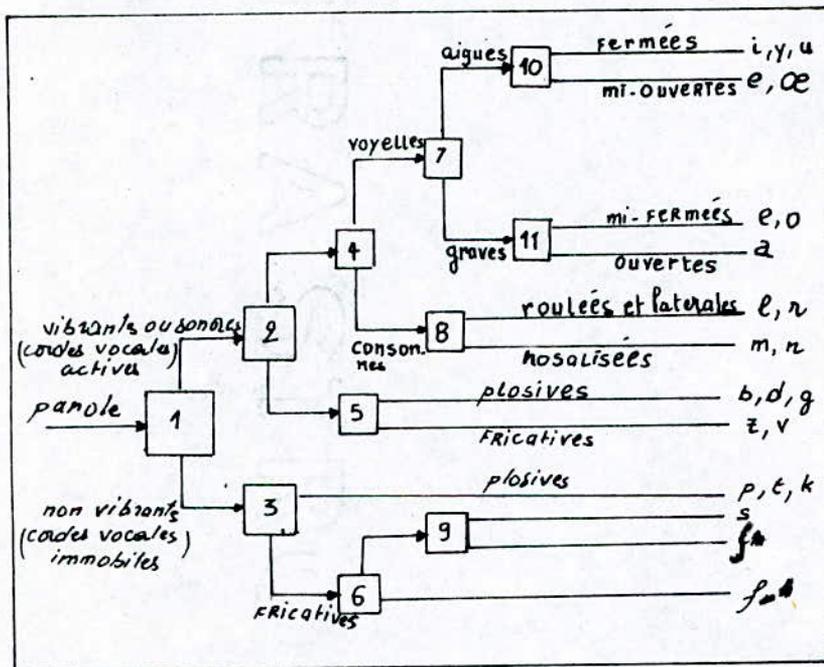


FIG. 6

- Les consonnes fricatives (dites "spirantes").

Une obturation presque totale du conduit vocal est produite soit par les lèvres, soit par la langue.

Ces sons sont non voisés. Ex : [f], [s], [ʃ].

Si on combine cet effet avec le voisement, les sons obtenus sont [v], [z], [ʒ].

- Les consonnes nasales.

Sont obtenues quand le conduit vocal est modifié par un abaissement de l'arrière du voile de palais permettant à l'air expiré de s'échapper par les cavités nasales.

ex. : [m], [n].

- Les semi-voyelles et consonne liquide.

Sont très semblables aux voyelles orales et du fait des transitions rapides du conduit vocal, elles ressemblent aussi aux consonnes. ex. : [w], [y], [l].

- Les affriquées (dites "mi-occlusives").

Ce sont des occlusives dentales suivies instantanément par un son sifflant ou chuintant. Ex. : [ts].

6. CONCLUSION /

Nous avons fait une étude brève de l'ensemble du système vocal en nous limitant aux points susceptibles de jouer un rôle dans la phonation.

Par ailleurs, nous avons examiné la production des sons élémentaires de la parole (voyelles, consonnes, semi-voyelles) considérées isolément. Certaines caractéristiques de la parole ont été également abordées afin d'avoir une vue complète sur ses différents aspects.

.../...

CHAPITRE 2.

TRAITEMENT

N AMERIQUE

DES SIGNAUX

CHAPITRE II /

Dans ce chapitre, on va donner quelques notions générales sur le traitement du signal. Pour plus de détails, il faut se référer aux ouvrages traitant le traitement numérique des signaux (voir dans la bibliographie).

1/. - Définition du signal

Un signal peut être défini, comme le support physique d'une information. Il peut être continu ou discret. Dans le premier cas, il est appelé signal échantillonné ou signal numérique, dans l'autre cas, il est appelé signal analogique.

Le traitement numérique de signaux consiste à traiter des suites de nombres en vue d'extraire l'information utile.

Le signal de la parole contient une quantité importante d'informations imbriquées entre elles, ce qui rend difficile leur extraction.

L'un des caractères particuliers du traitement du signal est la référence importante aux aspects perceptifs : soit qu'il s'agisse d'adopter le signal aux canaux de communications de l'opérateur humain (communication homme/machine) soit que l'on vise une approche mimétique destinée à copier la perception humaine par un dispositif automatique ou automate générant des énoncés et permettant d'interpréter les énoncés d'autrui.

2/. - Série de Fourier

Soit $x(t)$ une fonction périodique de période T

$$x(t + T) = x(t) \quad \forall t$$

Cette fonction peut s'écrire comme la somme de plusieurs fonctions sinusoïdales

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

Le spectre de fréquences de cette fonction est défini par :

$$X(nf) = \frac{1}{2} (a_n - j b_n) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-j2\pi nft} dt \quad n \in \mathbb{N}$$

3/. Transformée de Fourier

La T.F. est un outil très précieux en traitement de signaux, en particulier dans l'analyse spectrale.

La parole se compose de sons et de bruits.

Les sons sont périodiques, tandis que les bruits sont non périodiques.

.../...

Pour cela, on remplace la notion de série de Fourier par l'intégrale de Fourier.

La transformée de Fourier d'un signal non périodique continu $x(t)$ est

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

La transformée de Fourier des signaux numériques est définie par :

$$X(f) = \sum_{k=-\infty}^{+\infty} x(k) \exp(-j2\pi f \cdot k)$$

Le spectre d'un bruit est continu, alors que celui d'un son est discret.

La fonction $X(f)$ est généralement une fonction complexe de la variable continue f .

Condition d'existence de la transformée de Fourier

La condition suffisante pour que $X(f)$ existe est : $\sum_{k=-\infty}^{+\infty} |x(k)| < \infty$

Si cette condition est satisfaite, alors la série converge absolument vers une fonction continue f .

Les signaux qui satisfont cette condition sont tous les signaux à énergie finie et si de plus

$$\sum_{k=-\infty}^{+\infty} |x(k)|^2 \ll \left[\sum_{k=-\infty}^{+\infty} |x(k)| \right]^2$$

↑
énergie du signal $x(k)$

Les signaux sont absolument sommables.

4/. - Transformée en Z.

L'étude des systèmes nécessite une dimension de plus, que la T.F. ne possède pas. Pour cela, on utilise une transformation plus générale qui est la transformée en Z qui est orientée vers l'analyse et la synthèse des systèmes

La transformée en Z, $X(Z)$ d'un signal numérique $x(k)$ est définie par :

$$X(Z) = \sum_{k=-\infty}^{+\infty} x(k) z^{-k}$$

5/. - Transformation inverse de Fourier

Les échantillons $x(k)$ ne peuvent être obtenus à partir de la transformée

$$X(f) \text{ par : } x(k) = \frac{1}{F} \int_{-F/2}^{F/2} X(f) \exp(j2\pi f k) df$$

où la période $F = 1$.

6/. - Corrélation des signaux

En traitement des signaux, il est souvent nécessaire de comparer 2 signaux (par exemple, on décale l'un des signaux par rapport à l'autre et on mesure leur similitude en fonction du décalage)

$$\dot{E}_{xy}(k) = \sum_{l=-\infty}^{+\infty} x(l) y(l+k) = \sum_{l=-\infty}^{+\infty} x(l-k) y(l) = \dot{E}_{xy}(-k)$$

où au moins, un des 2 signaux supposés réels est à énergie finie.

Le signal $\dot{E}_{xy}(k)$ est une fonction d'intercorrélation de $x(k)$ et de $y(k)$.

Si ces 2 signaux sont identiques, le signal $\dot{E}_x(k)$ est une fonction d'autocorrélation de $x(k)$.

Dans le calcul de la fonction de corrélation, on procède par étapes :

1°) Le signal $g(l)$ est décalé d'une certaine quantité k

2°) Le produit $x(l) y(l+k)$ est effectué échantillon par échantillon pour toutes les valeurs de l .

3°) Les valeurs obtenues sont additionnées pour obtenir une valeur $\dot{E}_{xy}(k)$

Pour plus de détails sur :

7 - Systèmes causals

8 - Systèmes linéaires

9 - Systèmes linéaires invariants

Voir KUNT chapitre : 3

10 - Produit de convolution

Le produit de convolution joue un rôle important en traitement numérique des signaux.

notation

$$y(k) = x(k) * g(k)$$

Pour le calcul du produit de convolution, on procède par étapes.

1°). La réponse impulsionnelle est rabattue autour de l'axe des coordonnées. On passe donc de $g(l)$ à $g(-l)$.

2°). Elle est ensuite décalée d'une certaine quantité k .

3°). Le produit $x(l).g(k-l)$ est effectué échantillon par échantillon pour toutes les valeurs de l .

4°). Les valeurs obtenues sont additionnées pour obtenir les valeurs de $y(k)$

.../...

Remarque :

On a une certaine ressemblance, entre un produit de convolution et une fonction de corrélation.

Pour $l = -l'$
$$\mathcal{C}_{xy}(k) = \sum_{l'=-\infty}^{+\infty} x(-l') y(k-l') = x(-k) * y(k)$$

Inversement, le produit de convolution de $x(k)$ et de $g(k)$ peut être considéré comme la fonction d'intercorrélation de $x(-k)$ et de $g(k)$.

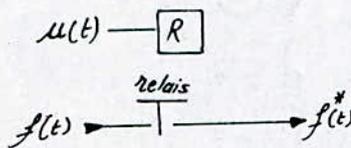
. La T.F. transforme un produit de convolution en un produit simple des transformées.

$$x(y) * h(t) \hat{=} X(f) \cdot H(f).$$

11. Echantillonnage et codage.

a - Echantillonnage :

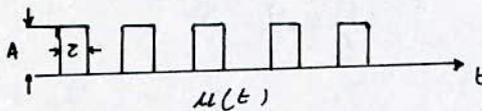
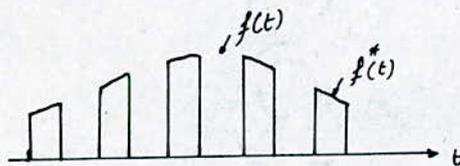
Echantillonner une fonction, c'est prélever cette fonction pendant un certain intervalle de temps. C'est lui substituer une autre fonction obtenue à partir de la fonction continue par découpage de cette dernière.



Le relais R est l'échantillonneur

$u(t)$ est la fonction d'échantillonnage

$f^*(t)$ est la fonction échantillonnée de $f(t)$.



La durée τ doit être infiniment petite devant la période T .

Les impulsions de période T et de durée relatives au signal $f(t)$ constituent les échantillons de cette fonction.

La fonction $u(t)$ peut se décomposer sous la forme d'une série de Fourier.

$$u(t) = \delta A \left[1 + 2 \sum_{n=1}^{\infty} \frac{\sin n \delta \pi}{n \delta \pi} \cos n \Omega t \right]$$

A : amplitude des impulsions.

δ : coefficient de remplissage

$\Omega = 2\pi f$ ($f = 1/T$) Ω : pulsation de récurrence des impulsions ...

Remarque 1 :

Le de substituer à $f(t)$ sa fonction échantillonnée $f^*(t)$ signifie que l'échantillonnage d'une fonction n'entraîne aucune perte d'information.

Pour reconnaître le signal $f(t)$ à partir de son échantillonnage $f^*(t)$, un simple filtrage suffit. On échantillonne toujours à une fréquence supérieure à deux fois la fréquence maximum.

Remarque 2 :

Echantillonner un signal $x(t)$: c'est multiplier ce signal par peigne de DIRAC

$$\tilde{x}(t) = x(t) \sum_{k=-\infty}^{+\infty} \delta(t - kT_0) = X(f) * F_e \sum_{h=-\infty}^{+\infty} \delta(f - hF_e)$$

Energie d'une fonction échantillonnée.

L'énergie d'une fonction échantillonnée correspond à l'énergie continue dans la fonction elle-même. On peut la définir par la moyenne quadratique :

$$E = \int_{-\infty}^{+\infty} [f(t)]^2 dt \quad : \text{Cas Continu}$$

$$E = \sum X_i^2 \quad : \text{cas discret.}$$

Le théorème de Parseval nous permet d'affirmer que la valeur moyenne du carré de la fonction $f(t)$, c'est-à-dire son énergie, est liée à celle de son spectre sous la forme :

$$\int_{-\infty}^{+\infty} [f(t)]^2 dt = \int_{-\infty}^{+\infty} [F(\omega)]^2 d\omega$$

Théorème de Shanon : Si un signal $f(t)$ de spectre borné est échantillonné à une fréquence supérieure à deux fois la fréquence maximale de ce spectre, alors on n'a pas de pertes d'informations.

Remarque :

Si des fréquences élevées existent dans une fonction dont on désire analyser les composantes B.F, il faut filtrer la HF porteuse et n'effectuer l'échantillonnage que sur le signal filtré.

.../...

Théorème de restitution des informations échantillonnées:

Si $x(t)$ et $h(t)$ sont respectivement la fonction à échantillonner et la réponse d'un filtre passe bas idéal de fréquence de coupure $f_r/2$ $\overline{h(t)}^2 = \frac{1}{x(t)}^2$

La fréquence $f_0 = \frac{f_r}{2}$ est appelée la fréquence de Nyquist. C'est la limite extrême de la fréquence maximum admissible à l'entrée.

L'erreur résultant du chevauchement des spectres est appelée erreur d'échantillonnage. On peut l'éliminer en augmentant la fréquence d'échantillonnage.

S'il n'y a pas d'erreur due au chevauchement, un filtre passe bas idéal éliminera les fréquences supérieures à F , restituera intégralement la fonction initiale.

Le filtre de reconstitution (passe bas idéal) peut être défini comme ayant une atténuation nulle, jusqu'à la fréquence de Nyquist et infinie au-delà.

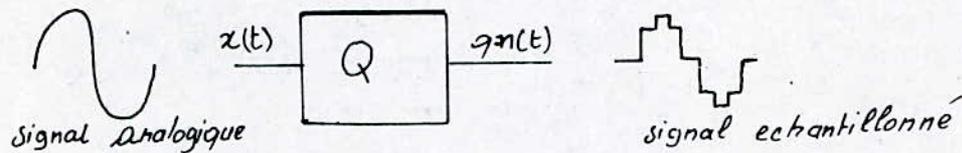
On reconstitue avec ce filtre l'enveloppe du signal original, en éliminant les composantes propres à l'échantillonnage.

b. Co d a g e.

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

. Quantification : C'est l'approximation de chaque valeur du signal $x(t)$ par multiple entier d'une quantité élémentaire q , appelée échelon de quantification.

. Codage : Le signal échantillonné et quantifié en amplitude est représenté par une suite de nombres binaires.



.../...

12. LA TRANSFORMATION DE FOURIER DISCRETE (T.F.D.).

La 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.

La T.F.D. de n valeurs discrètes $g(k)$ pour $k = 0, 1, \dots, N-1$ est l'ensemble des N valeurs données par :

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

En posant $W = e^{-j 2\pi \frac{1}{N}}$

Les affixes des nombres W^n appelés coefficients de la T.F.D. se trouvent sur le cercle unité. Ce sont les racines de l'équation $z^N - 1 = 0$

L'expression matricielle de la T.F.D. est :

Pour les différentes valeurs de $n = 0, 1, \dots, N-1$.

$$\begin{bmatrix} G(0) \\ G(1) \\ \vdots \\ G(N-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W^1 & W^2 & \dots & W^{N-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & W^{(N-1)} & W^{2(N-1)} & \dots & W^{(N-1)^2} \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(N-1) \end{bmatrix}$$

$$\vec{G}(n) = [W] \vec{g}(k).$$

La T.F.D. est une transformation qui, à un vecteur $\vec{g}(k)$ défini dans le domaine temporel, fait correspondre un vecteur $\vec{G}(n)$ dans le domaine fréquentiel, l'opérateur de la transformation est la matrice $[W]$ qui est une matrice carrée symétrique du type $N \times N$.

13. Transformée de Fourier inverse

Elle existe, et a pour expression

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

$$N g(k) = [W] G(n)$$

Le calcul des TFD directes ou inverses peut être effectué au moyen d'un algorithme : la T.F.R.

14. Transformée de Fourier rapide (T.F.R.) ou F.F.T. :

L'ensemble des méthodes permettant d'accélérer le calcul des TFD constitue la TF. Elle ramène le calcul à être plus rapide en diminuant le nombre d'opérations à effectuer de N^2 à **une valeur** de l'ordre de $N \log_2 N$.

Parmi les algorithmes de la TFR, on peut en citer plusieurs :

- Algorithme de Tukey-Cooley
- Algorithme de Sande
- Algorithme de Hadaward.

C'est suivant la décimation de n_k :

. Quand la décimation s'effectue dans le domaine temporel.

Cette méthode constitue l'algorithme de Cooley.

. Quand la décimation s'effectue dans le domaine fréquentiel, elle constitue l'algorithme de sande.

La T.F.R. permet d'effectuer avec une très grande rapidité :

- La détermination d'un spectre .
- La détermination d'une convolution
- La détermination d'une corrélation.

Soit un signal $x(t)$, fonction continue du temps.

Son spectre est défini par $x(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt$

il est de nature complexe :

$$x(f) = A(f) + jB(f)$$

Son module $|X(f)|^2 = A^2(f) + B^2(f)$

Son argument $\angle x(f) = \text{Arc tg } B(f)/A(f)$

La TFR fournit des valeurs approchées de ce module et de cette phase en effectuant sur $x(t)$ trois opérations :

- 1/ - Un échantillonnage satisfaisant le critère de Nyquist.
- 2/ - Une limitation de durée faisant appel à une fenêtre auxiliaire.
- 3/ - Une périodisation fournissant N points pour chacune des périodes du signal et de son spectre qui est discontinue.

$$X(n) = \sum_k x(k) e^{-2\pi j m k / N} = A(n) + j B(n)$$

.../...

Si la fréquence d'échantillonnage du signal ne satisfait pas le critère de Nyquist, on a empiètement des motifs du spectre les uns sur les autres et on ne peut pas isoler correctement l'un d'eux pour récupérer le signal.

Pour cela, on doit filtrer le signal avec un passe bas tel qu'aucune fréquence supérieure à une valeur f_m ne subsiste dans le spectre définitif ou bien utiliser une grande cadence d'échantillonnage.

On doit choisir une fenêtre de troncature dont le spectre est aussi peu étalé que possible.

On doit tester plusieurs fréquences d'échantillonnage pour avoir une fréquence intéressante.

15. Les fenêtres de troncature.

Du fait de la non stationnarité, l'analyse doit se faire sur un nombre d'échantillons suffisamment réduits pour que le signal puisse, sur cette fenêtre, être assimilé à un signal stationnaire.

Pour atténuer la distorsion des spectres obtenus par application de la TFR, on doit choisir judicieusement les fenêtres limitant la durée du signal initial de façon à avoir le moins de pertes possible.

La portion du signal traitée est définie sur une durée θ , ce qui revient à multiplier le signal par une fenêtre temporelle naturelle qui est égale à :

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

Il lui correspond dans le domaine fréquentiel une fenêtre spectrale $Q_0(\nu)$ qui dépend de la durée θ et de la technique adaptée.

L'estimation de la densité spectrale est perturbée par $Q_0(\nu)$. Pour diminuer ce effets, on superpose à la fenêtre temporelle naturelle une autre fenêtre temporelle de pondération $f(t)$.

On a plusieurs sortes de fenêtres de pondération :

$$\psi(t) = 0,54 + 0,46 \cos 2\pi \frac{t}{\theta} : \text{ Fenêtre de Haming}$$

$$0 \leq t \leq \theta ; \theta = NT$$

$$\psi(t) = 0,5 \left(1 + \cos 2\pi \frac{t}{\theta} \right) : \text{ Fenêtre de Hamming.}$$

La fenêtre de Haming a 99,96 % de son énergie dans le lobe principal et le lobe secondaire le plus important est à 40 dB au-dessous du lobe principal.

.../...

Le modèle de Haming est l'un des plus utilisés.

$$\psi(t) = \alpha + (1-\alpha) \cos 2\pi t/\theta \quad t \in \left[-\frac{\theta}{2}, +\frac{\theta}{2}\right]$$

$\alpha = 0,50$: fenêtre de Hamming.

$\alpha = 0,54$: fenêtre de Hanning.

La transformée de Fourier du Modèle général de Haming s'obtient pour :

$$\Psi(f) = \pi(f) [\alpha + (1-\alpha) \cos 2\pi f/\theta]$$

on a le produit de deux fonctions, sa transformée est la convolution des spectres de la fenêtre $\pi(t)$ et du signal entre crochets.

$$\Phi(f) = \alpha \operatorname{sinc}(\pi f\theta) + (1-\alpha)/2 \operatorname{sinc}[\pi(f-\frac{1}{\theta})\theta] - \frac{(1-\alpha)}{2} \operatorname{sinc}[\pi f\theta]$$

Filtres numériques.

Le filtrage est une opération qui permet de recueillir des signaux utiles parmi toutes les composantes spectrales du signal, comme l'extraction des paramètres pertinents et la sélection des fréquences.

Lorsque le filtrage est effectué par un système linéaire numérique, il s'appelle filtrage numérique. Ce filtrage numérique est équivalent à un filtrage analogique, suivi d'une conversion analogique-numérique.

Définition :

Les filtres numériques sont des filtres à réponse impulsionnelle. Leur principe consiste à sommer des nombres qui se présentent à l'entrée, après les avoir pondérés. Les coefficients de sommation pondérés constituent la réponse impulsionnelle du filtre.

Le problème général de ce filtrage est d'élaborer une réponse impulsionnelle $g(k)$ qui possède la réponse fréquentielle et qui soit réalisable efficacement. Les nombres qui sont à l'entrée peuvent dépendre ou non de la sortie. Pour cela, on a des filtres récurrents et des filtres non récurrents.

Filtres non récurrents : Ce sont des filtres à réponse impulsionnelle finie, 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 nombre d'entrées, représentant les échantillons du signal à filtrer.

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

La fonction de transfert du filtre est :

$$H(f) = \sum_{i=0}^{N-1} a_i e^{-j2\pi f i T}$$

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

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

La fonction de transfert de ce filtre est :

$$H(z) = Y(z)/X(z) = \frac{\sum_{l=0}^L a(l) \cdot z^{-l}}{(1 + \sum_{k=1}^K b(k) \cdot z^{-k})}$$

La fonction de transfert présente des pôles et des zéros. ces filtres peuvent être instables à cause de leurs pôles. S'ils sont instables, on doit faire une correction qui consiste à augmenter le rayon du cercle unité afin d'englober les pôles se situant à l'extérieur.

Les filtres en temps réel, filtres en temps différés

- Un filtre est dit en temps réel, quand il détermine une valeur filtrée $\bar{y}(t)$ en fonction d'informations $y(t')$, avec t' antérieur à t .
- Un filtre est dit par contre en temps différé lorsqu'il détermine une valeur filtrée $\bar{y}(t)$ en fonction d'informations antérieures et postérieures à t .

Remarque :

Le conduit vocal est modélisé par un filtre purement récursif d'ordre P .
Chaque type de filtre possède des avantages et des inconvénients.

Sur le tableau ci-dessous, on représente la comparaison générale entre les filtres RIF et RII.

Remarque :

Les filtres RIF sont les filtres non récursifs.
Les filtres RII sont els filtres récursifs.

.../...

Type	Avantages	Inconvénients
Filtres RIF	<ul style="list-style-type: none"> . La fonction de transfert n'a pas de pôles, donc les filtres sont stables. . La formulation est simple. . Les erreurs de calcul ne se cumulent pas. 	<ul style="list-style-type: none"> . Nécessitent un espace mémoire important. . Le temps de calcul est long.
Filtres RII	<ul style="list-style-type: none"> . Nécessitent une capacité mémoire faible. . Les calculs sont simples 	<ul style="list-style-type: none"> . La position dépend de la position des pôles. . Les erreurs se cumulent lors des étapes de calcul.

COMPARAISON ENTRE FILTRES RII ET FILTRES RIF

...../.....

- CHAPITRE - 3 -

- ANALYSE -

1. INTRODUCTION /

Le but de l'analyse est l'extraction des paramètres pertinents caractérisant le signal vocal. Elle suppose que la parole a été délimitée (détection bruit/parole)

2. - PRETRAITEMENT /

Avant d'être analysé, le signal de parole doit d'abord subir un prétraitement que constituent le fenêtrage et la préaccentuation.

2.1. Fenêtrage

Son rôle est ^{de} limiter la durée d'un signal.

La fenêtre communément choisie pour le signal de la parole est la fenêtre de Hamming (1) qui est un cas particulier de celle de Hanning ($\alpha=0,50$) de forme sinusoidale, est préférée et sera choisie par la suite, car elle ne provoque l'apparition que de raies secondaires peu énergétiques dont les maxima décroissent rapidement, toutefois, la largeur du pic principal s'agrandit.

$$W_H = \begin{cases} \alpha + (1-\alpha) \cos(2\pi k/N) & \text{pour } |k| \leq \frac{N}{2} \text{ ET } \alpha=0,54 \\ 0 & \text{AILLEURS} \end{cases}$$

2.2. Préaccentuation

Une préaccentuation numérique de 6 dB/oct AVE compense les effets dus à la source d'excitation du conduit vocal (-12 dB/Octave) et au rayonnement des lèvres (6dB/octave) sur certains sons voisés.

Cette opération permet donc de rétablir le signal de parole tel qu'il était avant de sortir de la bouche.

3. - METHODES D'ANALYSE:

Les méthodes d'analyse sont diverses. Pour des raisons indiquées ci-après, notre choix s'est porté sur la méthode cepstrale et la méthode par prédiction linéaire. Aussi, nous ne nous attarderons pas sur les autres ; néanmoins on peut citer :

- analyse spectrale
 - . type : analogique : méthode par filtrage (vocodateur à canaux, à formants).
 - . type numérique : transformée de Fourier rapide : FFT.
- analyse par sonographie.

.../...

Le premier appareil qui a permis de représenter graphiquement les variations d'amplitude des vibrations sonores en fonction du temps est le kymographe, ensuite, il y'eut l'oscillographe, on a aussi l'enregistreur de hauteur et le Batymètre qui est un enregistreur logarithmique de niveau.

mais ces appareils-là ne donnent pas une vue complète du signal sonore.

Le "Sonograph" qui est un outil indispensable dans un laboratoire de phonétique peu après, a permis de donner plus d'informations et de mieux connaître le phénomène son en traçant automatiquement les spectres sonores qui sont en quelques sortes la carte d'identité du signal son. **VOIR FIGURE.**

3) La sonographie

On utilise le sonographe pour l'analyse du signal son.

1. Spectre du son :

Le son est une vibration complexe de l'air, cette vibration contient dans son mouvement complexe, un mélange d'un certain nombre de vibrations simples (vibrations sinusoïdales).

D'après Fourier, on peut décomposer une fonction périodique en plusieurs fonctions sinusoïdales dont les fréquences sont toutes des multiples entiers de la fréquence fondamentale. La répartition des amplitudes de chacune des oscillations correspond au spectre du son.

2. Le sonographe :

Est un appareil qui sert à analyser la parole à l'aide des sonagrammes.

Le sonographe était composé d'un ensemble de filtres mais ces derniers ont été remplacés par un seul filtre dont la fréquence centrale est variable.

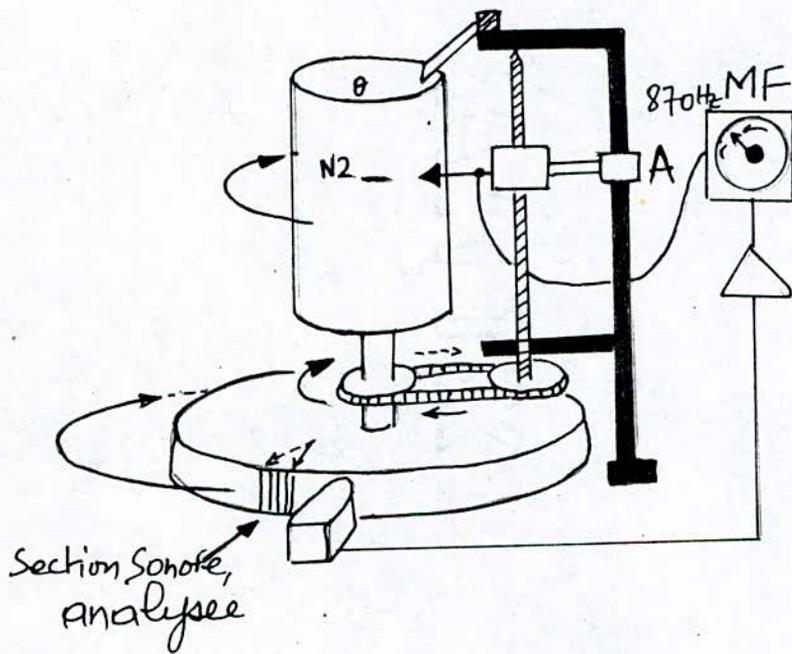
Pour chaque valeur de la fréquence centrale, correspond un tour de disque magnétique. A chaque nouveau tour, le signal de la parole considéré est relu et analysé autour de la fréquence considérée.

- Le sonogramme en bande étroite :

L'inconvénient majeur du spectre sonore est qu'il ne permet d'analyser qu'une "section sonore" très courte car il lui manque la dimension "temps". On a remédié à ce défaut en adoptant le système dit du "SONAGRAMME" FREQUENCE/TEMPS en bande passante étroite. La 3ème dimension, l'amplitude relative des harmoniques ou des "partiels" est rendue par un noircissement plus ou moins grand du papier.

Pour obtenir "paysage sonore" presque aussi complet que ce que l'oreille elle-même peut saisir (jusqu'à 13.000 à 18.000 HZ) on étend l'analyse sonographique à une gamme s'étendant de 85 Hz à 8.000 Hz.

.../...



xe SONOGRAPH

Le sonagramme en bande passante étroite sur 8.000 Hz donne un aperçu global intéressant de la structure acoustique d'une séquence parlée.

Défauts :

- on est mal renseigné sur l'aspect réel des formants,
- "Paresse" du filtre à se mettre en action

donc la "constante de temps" du filtre est grande. *à l'inverse de la bande passante*
. Le sonagramme en bande passante étroite présente une haute définition des détails sonores dans le sens vertical. Par contre, il présente une mauvaise définition dans le sens horizontal.

Sonographie en large bande passante :

La différence qui existe avec le sonagramme en bande passante étroite réside dans le fait que dans ce cas, la bande passante du filtre varie ; elle passe de 45 Hz (dans le 1er cas) à 300 Hz, ce qui donne au filtre une constante de temps très brève.

3.2 L'analyse temporelle

Dans la production de la parole, il existe des phénomènes évoluant rapidement dans le temps. C'est le cas par exemple de la fermeture brusque du conduit vocal lors de la production d'une plosive.

L'intérêt de l'analyse temporelle réside dans l'étude de ce genre de phénomènes mieux caractérisés par leur évolution dans le temps, que par leur spectre.

Les principales méthodes de cette analyse sont :

- La méthode de la fonction d'autocorrelation.
- La méthode de passage par zéro.
- La méthode de la prédiction linéaire.

a) Méthode des passages par zéro.

Le principe de cette méthode est qu'on ne s'intéresse qu'aux changements de signe du signal $S(t)$ considéré.

De ce fait, la méthode représente une grande simplicité de mise en œuvre et une grande rapidité puisqu'elle n'exige pas d'opération arithmétique compliquée. Cette méthode permet de localiser les fréquences des premiers formants ; elle permet aussi la mesure de la fréquence fondamentale "F0" après filtrage passe-ba

.../...

b) Méthode de la fonction d'autocorrelation

Pour que la méthode d'autocorrelation puisse être utilisée, le signal doit être découpé en fenêtres temporelles de 10 à 20 ms ; ces fenêtres doivent être ni trop courtes ni trop longues afin qu'elles puissent permettre la mesure des paramètres cette méthode, appliquée à un signal $s(t)$, fait appel au calcul de la fonction d'autocorrelation définie par :

$$g(\tau) = \int_{-\infty}^{+\infty} s(t) \cdot s(t-\tau) dt$$

où " τ " représente un retard.

De cette fonction, on peut déduire la densité spectrale de puissance

$$P(f) = \int_{-\infty}^{+\infty} g(\tau) \cdot e^{-2\pi j f \tau} d\tau$$

Cette grandeur est utile pour l'évaluation de la puissance totale du signal :

$$P_T = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P(f) \cdot df$$

La fonction d'autocorrelation peut être générée à partir du signal analogique par des lignes à retard, des sommateurs ou des modulateurs.

Elle peut aussi l'être par des procédés numériques à partir du signal échantillonné et ce, à l'aide de calculateurs ou grâce à un matériel spécialisé (autocorrelateur).

Cette méthode permet de mesurer la fréquence fondamentale " F_0 ". Elle présente l'avantage d'être simple, cependant elle exige des calculs trop longs.

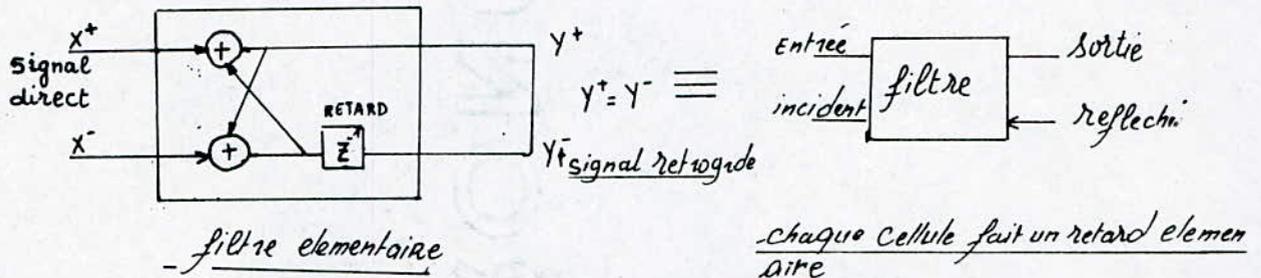
c) Méthode de la prédiction linéaire :

qui sera prise en détail par la suite.

.../...

3.3 Méthode du filtre en treillis

C'est un filtre qui est composé de p. étages (où p est l'ordre du filtre) identiques. Dans chaque étage, on réalise 2 multiplications des signaux par le coefficient de réflexion approprié



Les équations qui régissent le filtre sont :

$$Y^+ = X^+ + k z^{-1} Y^- \Rightarrow X^+ = Y^+ - k z^{-1} Y^- = E \quad (10)$$

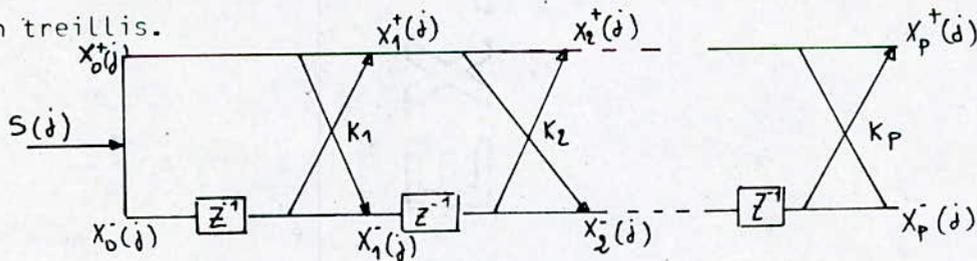
Parcours rétrograde

$$X^- = -k Y^+ + z^{-1} Y^- \quad (11) \quad S = Y^+ = Y^- \quad (12)$$

La fonction de transfert globale est :

$$H(z) = \frac{S(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}}$$

Quand on a plusieurs cellules, le filtre A(z) peut être donné par une structure en treillis.



Les équations de ce filtre sont :

$$X_i^+(j) = X_{i-1}^+(j) + K_i X_{i-1}^-(j) \quad (13)$$

$$X_i^-(j) = X_{i-1}^-(j) + K_i X_{i-1}^+(j) \quad (14)$$

$$X_0^+(j) = X_0^-(j) = S(j) \quad (15)$$

(13) et 14 donnent l'erreur précédente de prédiction à l'ordre "i" et l'erreur suivante de prédiction pour le même ordre.

(15): Prédicteur à l'ordre zéro

Remarque : Les coefficients de réflexion K_i sont aussi appelés coefficients de corrélation.

4. - ANALYSE PAR PREDICTION LINEAIRE /

Les principes méthodes de prédiction linéaire.

En l'absence temporaire d'une source d'excitation, un échantillon de parole peut être prédit avec exactitude à l'aide d'une somme pondérée linéairement d'un nombre fini P d'échantillons le précédant immédiatement. Toutes les méthodes dites de prédiction linéaire sont fondées sur cette hypothèse.

La valeur \hat{s}_n n'est qu'une approximation de la vraie valeur S_n

soit :

$$\hat{S}_n = - \sum_{k=1}^P a_k S_{n-k} \quad (1)$$

et

$$e_n = S_n - \hat{S}_n = S_n + \sum_{k=1}^P a_k S_{n-k} \quad (2)$$

où e_n = erreur de prédiction.

En appliquant la transformée en Z à l'équation (2), nous obtenons :

$$E(Z) = S(Z) \cdot D(Z) \quad (3)$$

où

$$D(Z) = 1 - \sum_{k=1}^P a_k Z^{-k} \quad (4)$$

$$D(Z) = G/F(Z) \quad \text{FILTRE INVERSE}$$

Par conséquent, le problème d'analyse consiste à calculer un ensemble de coefficients a_k du prédicteur, qui soit tel que la relation (1) demeure optimale dans un intervalle de temps déterminé, pendant lequel l'appareil vocal peut être considéré stationnaire.

Nous examinerons trois méthodes d'optimisation qui se sont relevées intéressantes dans ce cadre d'application.

Ces méthodes font appel à un critère de moindres carrés. Nous faisons l'hypothèse que le signal est déterministe ; néanmoins, on parvient à des résultats analogues dans le cas où S_n est un échantillon d'un processus aléatoire.

4.1. - Approches fondées sur la covariance

Ces méthodes sont fondées sur les hypothèses suivantes :

- a) Le signal est défini pour $P + N$ échantillons consécutifs, avec N entier ;
- b) Un échantillon du signal de parole peut être prédit approximativement à partir des p échantillons précédents.
- c) L'hypothèse b) est valable pour les N échantillons consécutifs qui suivent.
- d) L'optimisation dépend d'un critère des moindres carrés, dont la formulation dépend de la méthode qui est soit globale, soit récurrente.

4.1.1. - Méthode de covariance :

Cette méthode est du type global, dont l'optimisation est fondée sur la minimisation de l'écart quadratique total entre le signal original et le signal prédit sur toute l'échelle des temps NT.

Le problème consiste, comme nous l'avons déjà mentionné, à trouver un ensemble de coefficients a_k , $1 \leq k \leq P$, tels que l'erreur en soit minimisée pour l'ensemble des échantillons de la séquence de signal considérée. Le critère utilisé est celui des moindres carrés, qui est le plus approprié. Il suffit donc de minimiser l'erreur quadratique totale E, exprimée par :

$$E = \sum_n C_n^2 = \sum_n \left(s_n + \sum_{k=1}^P a_k s_{n-k} \right)^2 \quad (5)$$

Le problème se ramène donc à trouver la condition qui rend minimale l'erreur quadratique totale E, relativement aux coefficients a_k , $1 \leq k \leq P$.

On obtient cette condition en annulant la dérivée partielle de E par rapport à chaque a_k , ce qui conduit aux équations normales suivantes :

$$\sum_{k=1}^P a_k \psi_{ik} = -\psi_{i0}, \quad 1 \leq i \leq P \quad (6)$$
$$\psi_{ik} = \sum_{n=0}^{N-1} s_{n-i} \cdot s_{n-k}$$

Les équations (6) constituent un ensemble de p équations à p inconnues, que l'on peut résoudre pour obtenir les coefficients a_k . Les coefficients ψ_{ik} dans (6) constituent une matrice de covariance, qui est symétrique. L'erreur quadratique totale minimale de prédiction E_p , qui est la valeur minimale de E, peut être déterminée une fois les coefficients a_k calculés. Cette valeur minimale dépend de l'ordre du prédicteur P, d'où sa désignation E_p . A partir des équations (5) et (6), on peut démontrer que :

$$E_p = \psi_{00} + \sum_{k=1}^P a_k \psi_{0k}$$

4.2. - Approche fondée sur l'autocorrelation

Les hypothèses de base des méthodes d'autocorrelation sont les suivantes :

4.2.1. Le signal est défini pour toutes les valeurs du temps ($-\infty \leq m \leq +\infty$)

Pratiquement, on ne s'intéresse qu'à l'intervalle fini pendant lequel l'appareil vocal peut être considéré stationnaire.

Par conséquent, on applique une fonction d'apodilisation (fenêtre), W_n , au signal de parole tel que le produit $S_n = W_n s_n$ est identiquement nul en dehors d'une séquence de N échantillons ; par souci de simplification, supposons que la fenêtre soit rectangulaire tel que $S_n = s_n$ pour $0 \leq n \leq N$

.../...

4.2.2. - Chaque échantillon peut être prédit approximativement à partir des p échantillons précédents.

4.2.3. - L'hypothèse précédente est valable pour toutes les valeurs du temps (pour tous les échantillons du signal).

4.2.4. - L'écart quadratique total : entre le signal original et le signal prédit est minimalisé sur toute l'échelle des temps.

Par conséquent, le signal S_n prend des valeurs, à priori quelconques, pour $n = 0, 1, \dots, N-1$ et nulles pour tout n en dehors de cet intervalle. La minimalisation de l'erreur quadratique totale E conduit au système d'équation suivant :

$$(7) \quad \sum_{k=1}^p a_k R_{i-k} = R_i \quad 1 \leq i \leq p$$

où

$$R_i = \sum_{n=0}^{N-1-i} S_n \cdot S_{n+i}$$

Ici encore, le système d'équation (7) est de p Equations à p inconnues. Les coefficients R_{i-k} constituent une matrice dite d'autocorrelation et les Equations (7) sont appelées équations normales d'autocorrelation.

4.3. - Méthode de résolution du système linéaire

Les différentes méthodes de prédiction linéaire font appel à une résolution de p equations à p inconnues ($a_k ; 1 \leq k \leq p$) du type $AX = B$. Il existe un nombre considérable d'algorithmes effectuant une telle résolution. Les méthodes itératives du type GAUSS SEIDEL et JACOBI ont l'inconvénient que leur temps de calcul est à priori, indéterminé. Par contre, il existe des méthodes directes, telles que la triangularisation de GAUSS, la méthode GAUSS-JORDAN et la réduction de crout.

La matrice d'autocorrelation constituée des R_{i-k} est de la forme d'une matrice

Toeplitz

$$\begin{bmatrix} R(0) & R(1) & \dots & \dots & R(p-1) \\ R(1) & R(0) & \dots & \dots & R(p-2) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & \dots & R(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ \vdots \\ R(p) \end{bmatrix}$$

.../...

Remarque :

La matrice carrée R (P) est symétrique, de plus les éléments situés sur les diagonales parallèles à la diagonale principale sont égaux.

4.4. - Comparaison des deux méthodes

La comparaison des méthodes décrites n'a pas de signification que par rapport à une application bien déterminée.

Chaque méthode présente des avantages et des inconvénients liés aux conditions d'application et des contraintes de natures différentes, qui sont fixées par le domaine d'application.

Sur un plan théorique, les hypothèses formulées dans le cas de la méthode de covariance sont plus réalistes que celles de la méthode d'autocorrelation. Le signal de parole dans l'intervalle d'analyse est considéré non stationnaire dans le cas de la méthode de covariance, mais stationnaire dans la méthode d'autocorrelation.

Nous savons que la production de la parole est un processus non stationnaire, par conséquent l'hypothèse de non-stationnarité est meilleure. Néanmoins, dans un intervalle de temps relativement court, le signal peut être considéré quasi-stationnaire et, de ce fait, l'hypothèse de stationnarité n'est pas à rejeter.

Propriétés	Méthodes	
	Autocorrelation	Covariance
Fenêtre	Nécessaire	non
Stabilité	Garantie	non

4.5. - Détermination du nombre de coefficients prédicteurs

L'ensemble source vocale, conduit vocal et rayonnement pouvant être modélisé par un filtre linéaire récursif; on s'intéresse alors à la détermination des résonances de ce filtre.

...../.....

En admettant que le spectre de la parole puisse être décrit à l'aide de cinq résonances et qu'à chacune d'elles on puisse associer leurs coefficients, on voit que dix coefficients seront nécessaires pour représenter le conduit vocal.

Par ailleurs, deux coefficients supplémentaires sont nécessaires pour représenter le spectre de la source d'excitation et du rayonnement (au niveau des lèvres), ce qui donne un total d'une douzaine de coefficients de prédiction.

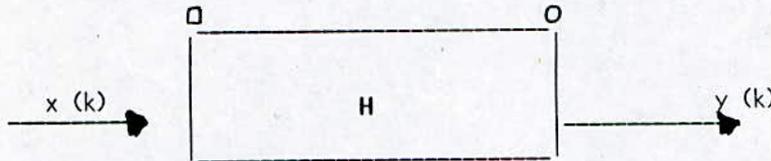
Rappel

Transformation homomorphique.

Pour séparer deux signaux combinés par multiplication ou par convolution dans le cas où le filtrage n'est pas linéaire, comme dans le cas du traitement homomorphique (qui, d'une manière générale trouve son application dans le traitement de signaux acoustiques et le traitement d'image).

Définition :

Les systèmes représentés par une transformation linéaire entre espaces vectoriels sont appelés homomorphiques, on peut les représenter par :



□ : loi d'entrée

○ : loi de sortie

On se limitera aux cas où les lois à l'entrée et à la sortie sont identiques en particulier dans les deux cas de la multiplication et du produit de convolution.

.../...

1°/ Combinaison de signaux par produit.

Pour extraire un signal utile combiné avec d'autres signaux par multiplication, on doit modifier l'un des signaux.

Par exemple, un signal transmis par modulation d'amplitude est multiplié avec une porteuse de H.F. qu'il faut séparer à la réception (voir KUNT chapitre

Système caractéristique :

Les signaux d'entrée possibles sont de la forme :

$$x(k) = [x_1(k)]^{a_1} [x_2(k)]^{a_2} \quad a_1, a_2 = \text{Ctes arbitraires.}$$

Le système caractéristique A. doit satisfaire la relation suivante :

$$A. [x_1(k)]^{a_1} [x_2(k)]^{a_2} = a_1 A. [x_1(k)] + a_2 A. [x_2(k)]$$

La fonction logarithme satisfait cette relation.

pour $\begin{cases} x_1(k) > 0 \\ \text{et} \\ x_2(k) > 0 \end{cases}$ on a $\ln [x_1(k)^{a_1} x_2(k)^{a_2}] = a_1 \ln x_1(k) + a_2 \ln x_2(k)$

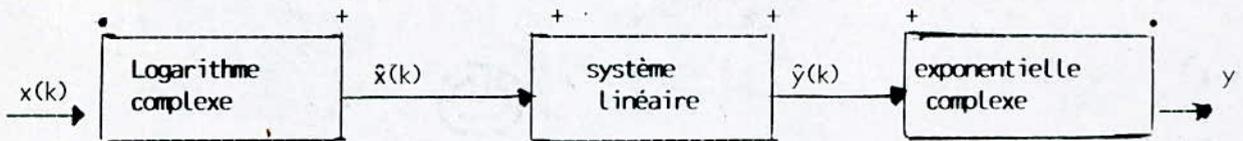
Dans ce cas, les signaux doivent être positifs comme ceux utilisés par exemple en traitement d'images.

On peut lever cette restriction en prenant la fonction logarithme complexe, soit $x(k) = |x(k)| \exp[j \arg [x(k)]]$

Le logarithme complexe est : $\ln [x(k)] = \ln |x(k)| + j \arg [x(k)]$

C'est la fonction inverse de l'exponentielle complexe

Forme canonique du système :



Remarque

Pour utiliser le logarithme complexe comme transformation d'un système caractéristique A, on doit enlever l'ambiguïté sur l'argument (en prenant modulo 2π) et vérifier si le principe de superposition généralisée est respecté, c'est-à-dire :
 si $x(k) = x_1(k) \cdot x_2(k) \Rightarrow \ln [x(k)] = \ln |x_1(k)| + \ln |x_2(k)|$

.../...

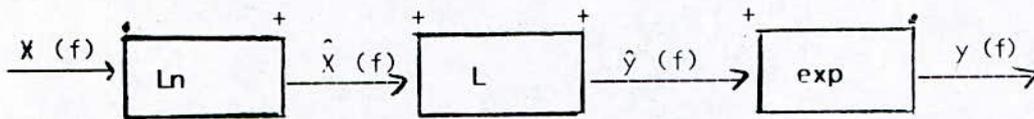
2°/. Combinaison de signaux par produit de convolution.

Comme dans le traitement de la parole, on veut isoler l'excitation de la réponse impulsionnelle de la conduite vocale.

Pour isoler un signal utile convolué avec un autre signal indésiré, on utilise la méthode de déconvolution (le signal composite est filtré avec un filtre dont la réponse fréquentielle est l'inverse de la T.F. du signal perturbateur), mais on connaît mal le signal indésiré. Pour cela, on utilise la méthode des systèmes homomorphiques convolutifs.

Le système caractéristique est de la forme $x(k) = \sum_{l=-\infty}^{+\infty} x_1(l) x_2(k-l)$
 $x(k) = x_1(k) * x_2(k)$ (voir chapitre II). KUNT

Forme canonique des systèmes homomorphiques convolutifs (quand les signaux sont représentés par leur T.F.).



· D'après OPPENHEIM

Les systèmes homomorphiques convolutifs sont utilisés pour estimer les différents paramètres du signal de la parole qui est un signal très complexe et on les utilise surtout pour séparer les divers composants de ce signal.

La parole est formée par l'excitation de la conduite vocale.

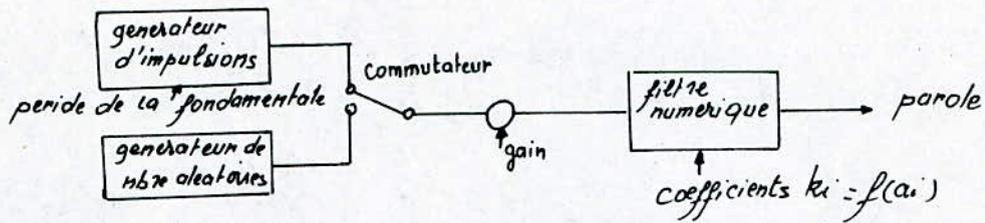
- Si les cordes vocales sont tendues (comme dans les voyelles), les sons sont dits voisés.
- Si les cordes vocales sont écartées, les signaux formés sont appelés signaux non voisés (voir chapitre I phonétique).

Donc, la conduite vocale est comme un système qui, suivant l'excitation qu'il reçoit dans le temps, donne une réponse déterminée.

Remarque

Si l'excitation et la conduite vocale sont indépendantes, la production de la parole peut se résumer dans le modèle donné par SCHAFER.

.../...



k_i : dépend de la section du conduit vocal.

La source d'excitation est :

- soit un générateur d'impulsions périodiques $f = \frac{1}{T}$ fondamentale du son ;
- soit un générateur de nombres aléatoires qui simule les turbulences quasi-aléatoires.

L'une de ces sources est appliquée à l'entrée d'un filtre numérique variant dans le temps qui simule la conduite vocale.

Les coefficients du filtre varient à peu près tous les 10 ms pour donner à chaque fois une nouvelle configuration de la conduite vocale.

- Pour régler le niveau sonore, on a un contrôle de gain entre la source et le système.
 - Pour les sons voisés, le filtre numérique contient un préfiltre qui permet de déterminer la forme des impulsions à durée limitée.
 - Pour atténuer l'effet des discontinuités aux extrémités des intervalles .
- Le signal $S(k)$ est multiplié par une fenêtre $W(k)$, pour un segment de parole on a : $x(k) = S(k) W(k) = [p(k) * g(k) * h(k)] W(k)$.
- $p(k)$: Signal d'excitation sur une courte durée.
- $g(k)$ et $h(k)$: réponses impulsionnelles du préfiltre et du filtre numérique respectivement.

Pour réduire le plus possible les formes de redondance présentes dans le signal de la parole et ne transmettre que l'information utile, on utilise le critère des moindres carrés qui consiste à minimiser l'erreur entre le signal original et la valeur prédite.

.../...

5.- L'ANALYSE CEPSTRALE / .

5.1. - Définition

Pourquoi l'analyse cepstrale ?

Elle apparaît intéressante en reconnaissance grâce à son économie de représentation (peu de paramètres : 8 coefficients au lieu de 12 pour la méthode par prédiction linéaire et 20 pour la méthode spectrale) et à son interprétation physique en terme de distance (comparaison entre paramètres).

Son but est de réaliser une séparation entre l'excitation et l'enveloppe spectrale du conduit vocal.

Un segment de parole peut être décrit par une convolution combinant les effets dûs à la source, au conduit vocal et aux lèvres..

La déconvolution a pour but de séparer ces éléments de convolution, cette séparation est bien effectuée, ce qui représente l'avantage de cette méthode utilisée en reconnaissance de la parole pour la détermination du pitch notamment, lequel but peut être atteint par le biais du traitement homomorphique

Le moyen utilisé est le **ce** pstre.

L'opération de déconvolution vérifie l'équation.

$$\begin{aligned} D(X(n)) &= D(X_1(n) * X_2(n)) \\ &= D(X_1(n)) + D(X_2(n)) \end{aligned}$$

D : étant l'opérateur de déconvolution.

On sait que la transformée en Z d'un produit de convolution est le produit en Z des dérivés qui le constituent, c'est-à-dire :

$$X(Z) = X_1(Z) \cdot X_2(Z)$$

Ce produit peut être transformé en somme à l'aide de la fonction logarithme d'où :

$$\hat{X}(Z) = \text{Log}(X(Z)) = \text{Log} X_1(Z) + \text{Log} X_2(Z).$$

.../.....

Le logarithme utilisé est le logarithme complexe c'est-à-dire :

$$\hat{X}(z) = \hat{X}(e^{j\omega}) = \log |X(e^{j\omega})| + j \text{Arg } X(e^{j\omega})$$

Le cepstre complexe, $\hat{X}(n)$ d'un signal est donné par la transformée de Fourier inverse du logarithme de la transformée de Fourier directe (ou spectre) de ce signal :

$$\hat{X}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{j\omega n} d\omega$$

avec $\hat{X}(e^{j\omega}) = \log X(e^{j\omega})$

Le cepstre qui nous intéresse, est donné par :

$$C(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega \quad -\infty < n < +\infty$$

Il peut être aussi défini par :

$$C(n) = (\hat{X}(n) + \hat{X}(-n)) / 2$$

c'est-à-dire :

$$C(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log (X(e^{j\omega})) \cos(\omega n) d\omega$$

Le cepstre peut donc être obtenu de deux manières :

- soit par deux transformations de Fourier ;
- Soit par une transformation en cosinus du logarithme du spectre.

En pratique, on utilise une approximation pour éviter le calcul intégral :

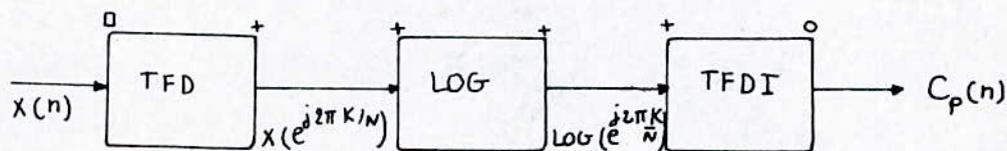
On calcule l'inverse de la transformée de Fourier discrète du logarithme de la transformée de Fourier discrète du signal d'entrée, c'est-à-dire :

$$C_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} \log |X_p(k)| e^{j2\pi k \frac{n}{N}} \quad \text{pour } 0 \leq n \leq N-1$$

où

n : rang du coefficient Cepstral

N : nombre de points



Les coefficients cepstraux décroissent en $\frac{1}{N}$ d'où la nécessité d'un nombre réduit de ceux-ci pour caractériser le signal.

.../...

5.2. - Explication physique du cepstre

1) - Le cepstre d'un signal est une fonction du temps définie soit comme la transformée de Fourier inverse du logarithme de la densité spectrale de puissance, soit comme le module carré de cette transformée. L'intérêt de la méthode est illustré par la Fig.A.

Fig. A.a : on y voit l'allure du signal parole.

En fig. A.b : Le logarithme de $|S(f)|^2$, c'est-à-dire la densité spectrale logarithmique. Il présente une périodicité spectrale, le long de l'axe des fréquences, résultant de la périodicité de la parole due au pitch.

En Fig. A.c : Le cepstre qui est une fonction d'une variable appelée quefrenc (mesurée en seconde).

2) - L'Echelle spectrale dans le calcul du cepstre est linéaire en fréquence. Les études physiologiques et perceptives de l'oreille indiquent que cette dernière est sensible à une échelle quasi-logarithmique (dite échelle MEL) linéaire sur le premier 1 KHZ et logarithmiques au-delà (Fig.B).

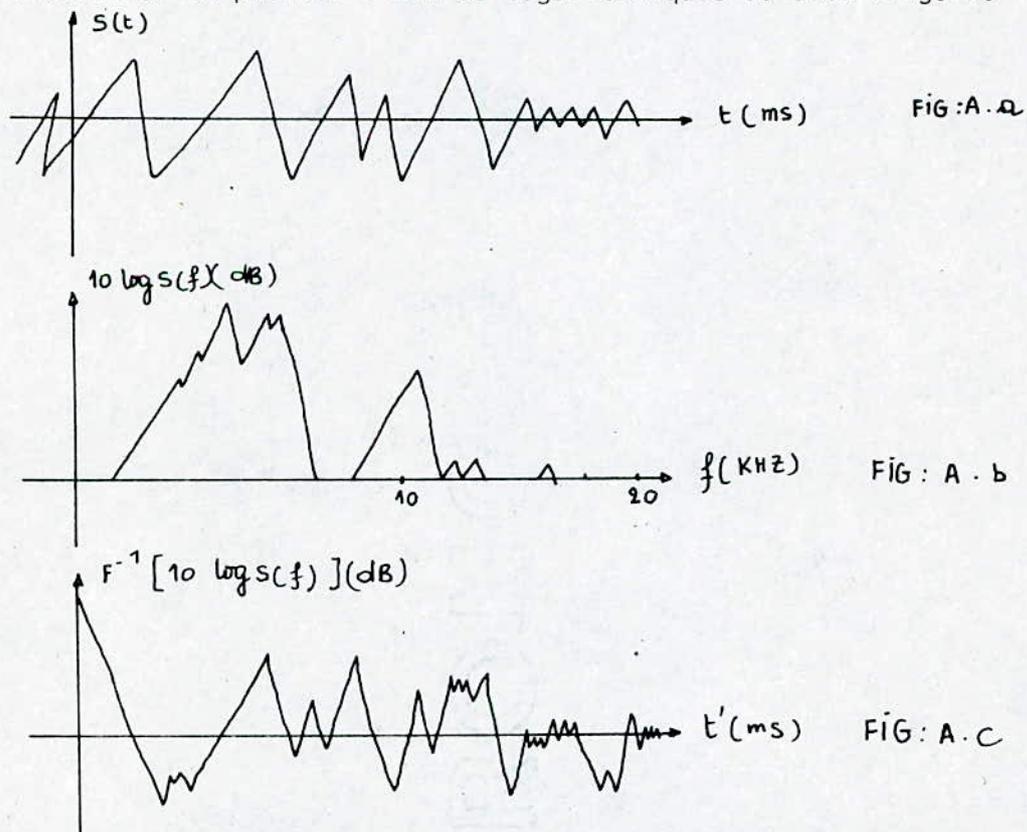


Fig. A. signal, densité spectrale logarithmique et cepstre correspondant.

.../...

5.3. - L'Echelle MEL

L'échelle spectrale dans le calcul du cepstre est linéaire en fréquence. Les études physiologiques et perceptives de l'oreille semblent indiquer qu'elle est sensible à une échelle presque logarithmique de la fréquence.

Donc, les informations pertinentes se trouvent réparties de façon exponentielle dans les zones éloignées.

L'échelle logarithmique tente d'éviter de donner un même poids à des zones de fréquences qui n'ont pas la même densité d'information.

Pour palier à cet inconvénient, on fait appel à une échelle quasi-logarithmique appelée échelle "MEL". Linéaire pour les fréquences inférieures à 1 KHZ et logarithmique au dessus.

Sur le premier KHZ, nous disposons d'un certain nombre de filtres, typiquement une dizaine, de largeurs uniformes.

Les filtres de formes triangulaires sont décalés les uns des autres, de la moitié de la largeur des filtres. Les débuts des filtres suivants sont déterminés en appliquant une raison géométrique au décalage entre les deux filtres précédents.

Cette raison est supérieure à 1 du fait que les filtres ont des largeurs croissantes exponentiellement (voir fig. B).

Les débuts des filtres utilisés sur le système SERAPHINE sont les suivants : 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1150, 1300, 1500, 1700, 2000, 2350, 2700, 3100, 3550, 4000, 4500, 5050, 5600, 6200.

Avec la série de filtres, le cepstre s'écrit comme suit :

$$MFCC(n) = \frac{1}{NF} \sum_{k=1}^{NF} \log [E(k)] \cdot \cos \left[n \cdot (k-1) \frac{\pi}{NF} \right]$$

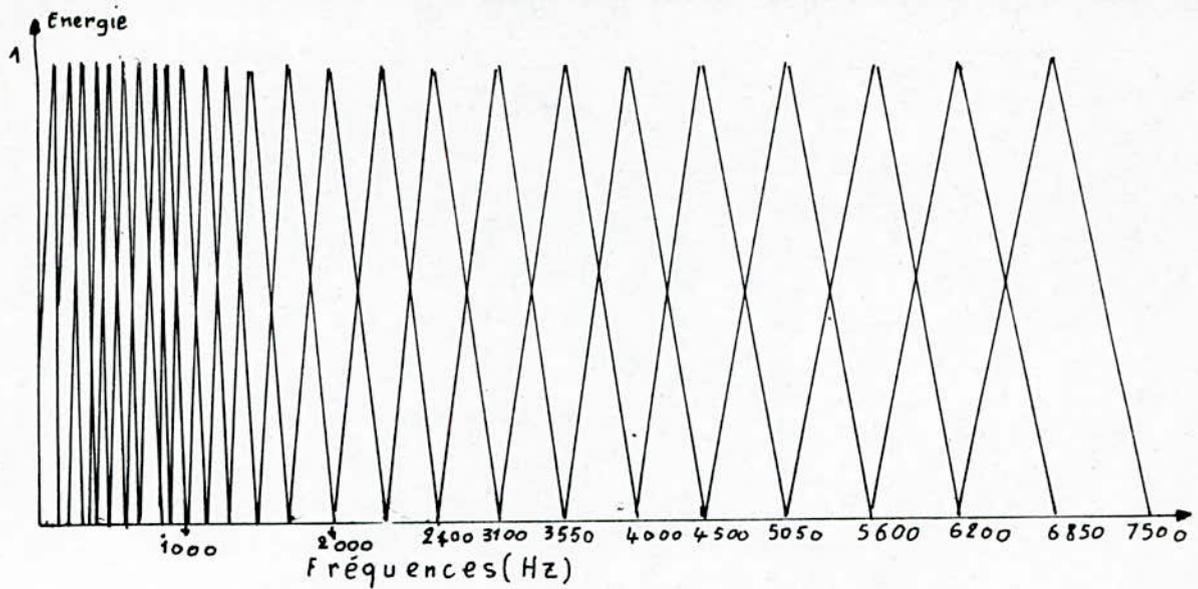
où : n : est le rang du coefficient cepstral

k : le numéro du filtre d'énergie E (k)

NF : le nombre total des filtres triangulaires (en général NF = 25).

Les coefficients cepstraux sont obtenus à l'aide d'un **vocodeur** à canaux (Fig.2) composé de 25 filtres passe-bande de forme triangulaire. Les 10 premiers ont une largeur uniforme et les autres ont une largeur qui suit une progression géométrique.

.../...



FIGB. Banc d'essai de filtres triangulaires

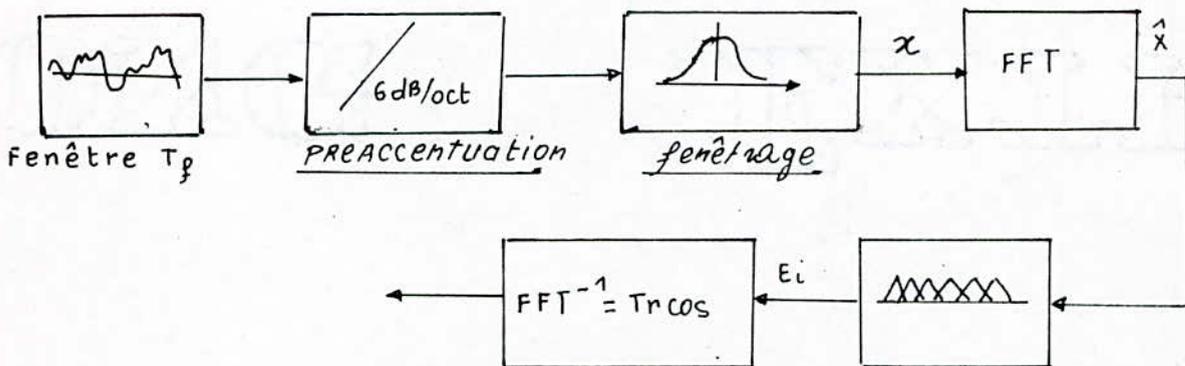
a) Description du traitement du signal :

Le signal de parole, échantillonné et codé, sera analysé en le découpant par zones de petites durées, communément, pour simplifier, des fenêtres d'analyse de longueurs égales sont considérées.

Remarque

Comme on l'a déjà signalé, le signal de la parole n'est pas stationnaire, alors on fait une approximation en prenant le signal sur une durée réduite.

Sur chaque intervalle d'analyse, on effectue plusieurs opérations suivant la figure ci-dessous :



traitement d'une fenetre de signal

...../.....

La largeur de chaque fenêtre est T_f sa durée est comprise entre 20 ms et 40 ms. Le nombre des échantillons compris dans chaque fenêtre est fixé selon la cadence d'échantillonnage.

Pour la cadence 12,8 khz que nous avons choisie, le nombre d'échantillons est 256, ce qui donne la durée de la fenêtre de 20 ms. Pour le décalage qu'on introduit entre 2 fenêtres successives, on prend un recouvrement de 50 % pour pas avoir de pertes.

- Un filtre de préaccentuation de la forme $(z) = 1 - b.z^{-1}$ ($b=96$) (d'après NT/LAA/TSS/79) sert à compenser les effets dus à la source d'excitation du conduit vocal (-12 dB par octave) dans le cas des sons voisés et ce filtre a pour but aussi de compenser le rayonnement des lèvres (6dB/oct).

- Pour diminuer la distorsion du spectre, on utilise la fenêtre de Hamming qui s'adapte bien à notre cas et qui est simple.

- Une transformée de Fourier rapide (FFT).

- Le calcul de l'énergie est en sortie de chaque filtre du banc.

- Transformation en cosinus pour obtenir les coefficients cepstraux

Sur calculateur, pour obtenir les coefficients cepstraux $C_m(k)$, on peut utiliser une autre procédure, puisque les énergies E_i en sortie du banc de filtres est presque la même que celle qu'on obtient en sortie d'un vocodeur à canaux dont la répartition suivrait celle de l'échelle "MEL" des fréquences.

b) Vocodeur à canaux : Fig. 2

- **Structure** du vocodeur à canaux :

Le vocodeur à canaux est constitué d'un organe d'analyse (et d'un organe de synthèse).

L'analyseur du vocodeur :

La fonction d'analyse de l'enveloppe spectrale est effectuée à l'aide d'un banc de filtres passe-bandes couvrant l'étendue de la bande téléphonique (300-3400 Hz).

Ainsi, le message à transmettre est décomposé en bande de largeur d'environ 300 Hz, chaque canal traite une bande de fréquence déterminée.

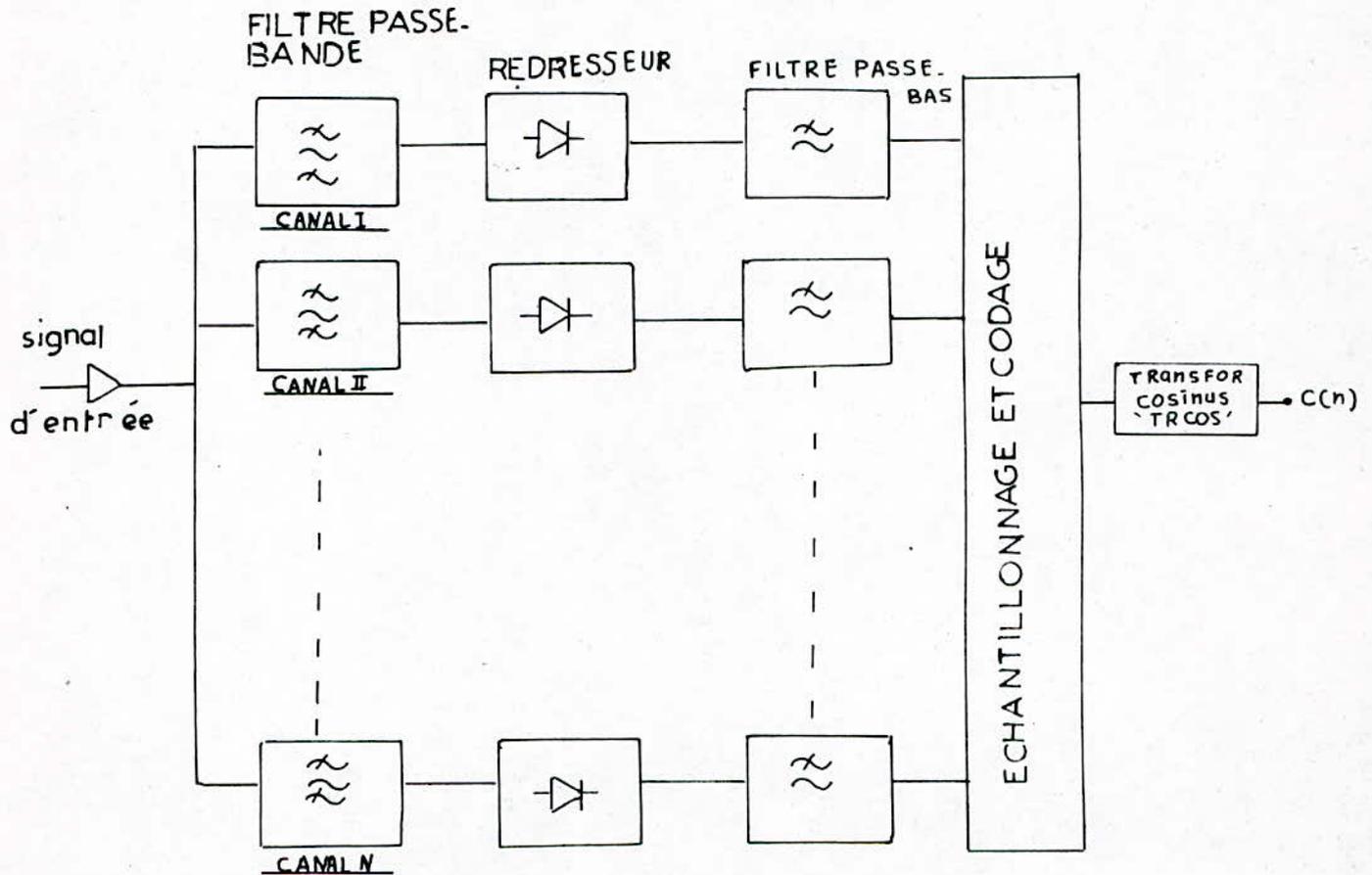
Après chaque filtre, il y a un étage de détection, lui-même suivi d'un étage de filtrage passe-bas limité à 25 Hz environ. Cette limitation des variations

.....

d'énergie à 25HZ a pour effet de modifier ou de supprimer les transitions rapides du niveau d'énergie.

L'analyseur comporte aussi un détecteur de voisement, il permet de différencier les sons voisés des sons non voisés et de donner dans le premier cas la valeur du pitch.

Les signaux issus de la sortie de l'analyseur sont échantillonnés et codés avant d'être transmis.



CALCUL DES COEFFICIENTS CEPSTRAUX - MEL.
PAR

VOCODÉUR A CANAUX

fig 2

TABEAU Comparatif des deux méthodes

Méthode	Nbre de coefficients	Facilité de mise en œuvre	Vitesse de calcul
Méthode Cepstrale	8	Peu de coefficients suffisent à bien représenter le spectre.	Vu les différentes étapes, le calcul des paramètres sera plus long.
Méthode par prédiction linéaire	12	Permet une estimation précise des paramètres de la parole : formant, fréquence fondamentale, etc...	Calcul des paramètres est relativement rapide.

PARTIE 2

2) KONJUGATION



A./ ORGANIGRAMMES /

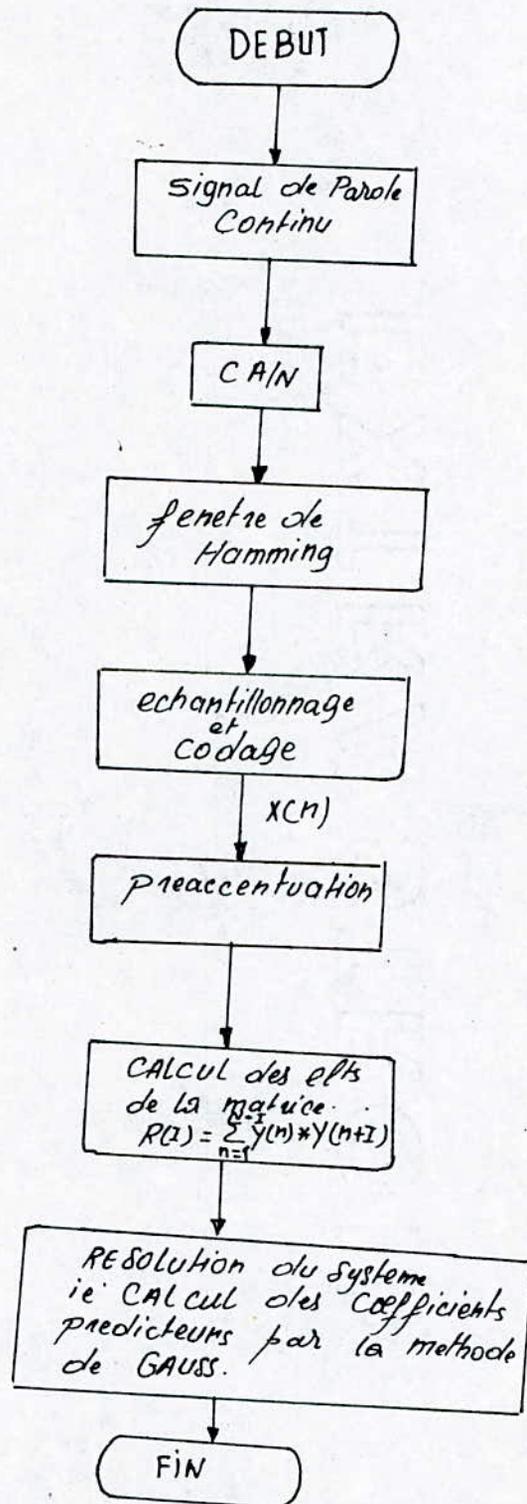
INTRODUCTION :

Les techniques d'analyse exposées précédemment sont données "en langage universel".

.../...

ORGANIGRAMME GENERAL DE L'ANALYSE DE la parole

ANALYSE par PREDICTION Lineaire methode d'autocorrelation



INTERPRETATION DE L'ORGANIGRAMME DE LPC

La résolution des systèmes d'équations est l'un des problèmes les plus courants des applications mathématiques.

Comme dans notre cas où on a à calculer les coefficients prédicateurs par la résolution du système suivant :

$$\begin{array}{l}
 B(1,1) C(1) + B(1,2) C(2) + \dots + B(1,12) C(12) = D(1) \\
 \vdots \\
 B(12,1) C(1) + \dots + B(12,12) C(12) = D(12)
 \end{array}$$

Seulement, notre matrice B a une forme particulière qui est la suivante :

$$B = \begin{bmatrix}
 R(0) & R(1) & \dots & R(11) \\
 R(1) & R(0) & \dots & R(10) \\
 \vdots & \vdots & \ddots & \vdots \\
 R(11) & \dots & \dots & R(0)
 \end{bmatrix}$$

Avec
$$R(I) = \sum_{N=1}^{N-I} \gamma(N) * \gamma(N+I)$$

Et

$$B(1,1) = \dots = B(12,12) = R(0) = \sum_{N=1}^N \gamma(N) * \gamma(N)$$

ce qui simplifie les calculs.

Notre système donc est un système linéaire. Les coefficients B (I,J) étant donnés, ainsi que les nombres D(I), on cherche les valeurs C(I). Il y a un nombre incroyable de méthodes pour effectuer ce genre de calcul, ce qui témoigne de l'importance du problème, ainsi que des nombreuses difficultés qui ont été rencontrées.

6) - Réduire à zéro tous les coefficients de la colonne I, en-deçà du point I,I.
Ceci est un pas vers la triangularisation.

7) - Augmenter I de 1. Si le nouvel I ne dépasse pas NQ-1, retourner à l'étape 4 ; sinon, aller à l'étape 8.

8) - Etape de substitution. Résoudre la NQ^{ième} équation pour C (NQ).
En utilisant cette valeur, résoudre la (NQ-1)^{ième} équation pour C (NQ-1).
En utilisant ces deux valeurs, résoudre la (NQ-2)^{ième} équation pour C (NQ-2).
Continuer ainsi jusqu'à ce que la première équation ait été résolue pour C(1).
Le calcul est alors terminé.

9)- Indiquer que le système donné est singulier.

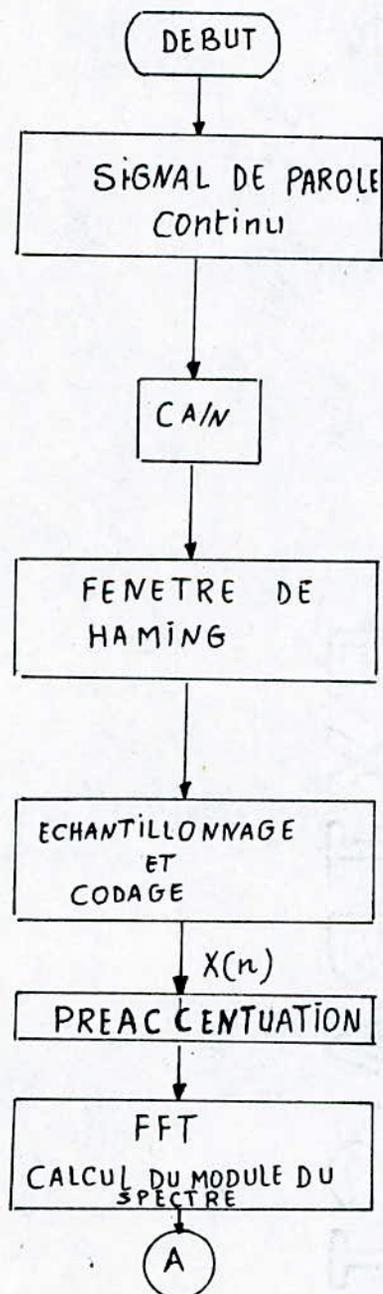
.../...

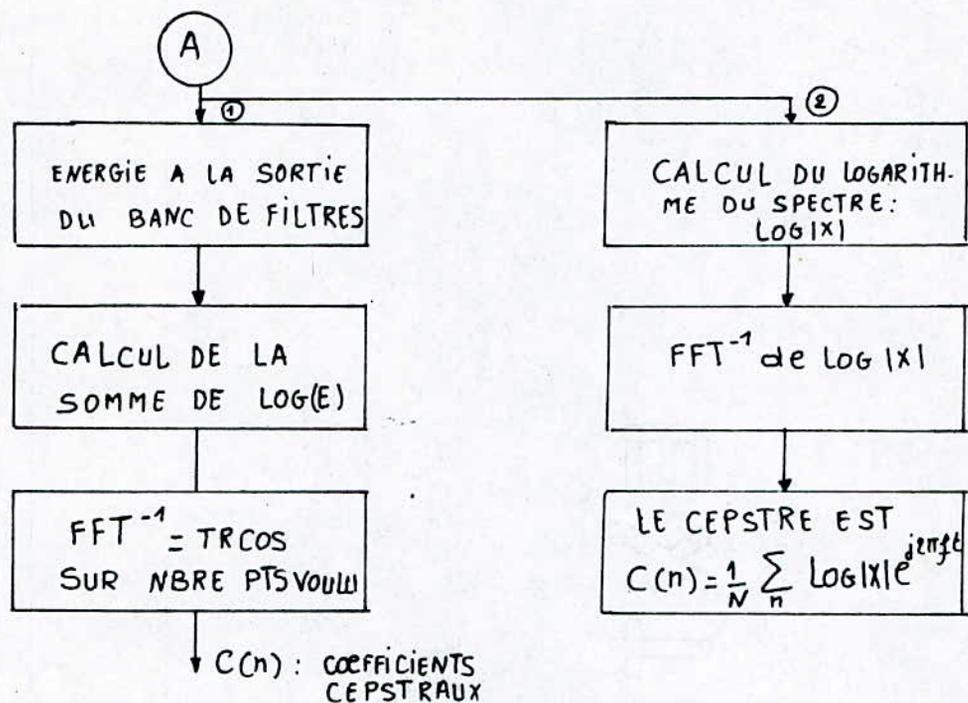
ORGANIGRAMME GENERAL DE L'ANALYSE

DU SIGNAL DE LA PAROLE :

ANALYSE CEPSTRALE

Les différentes étapes sont:





① - Calcul des coefficients cepstraux à partir de l'échelle de fréquences MEL.

② - Obtention du cepstre à partir de la transformée inverse du logarithme *du* spectre.

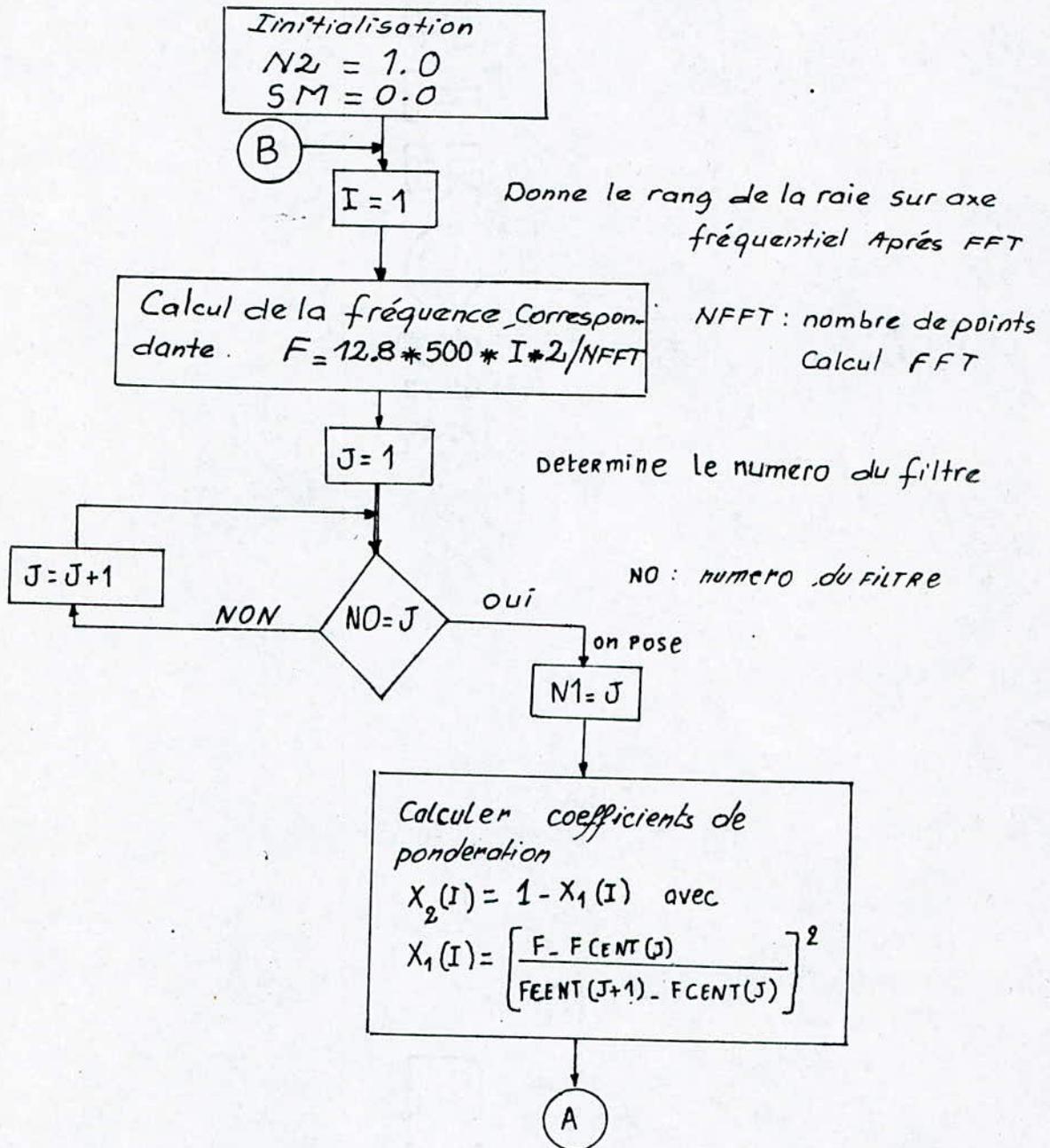
Remarque :

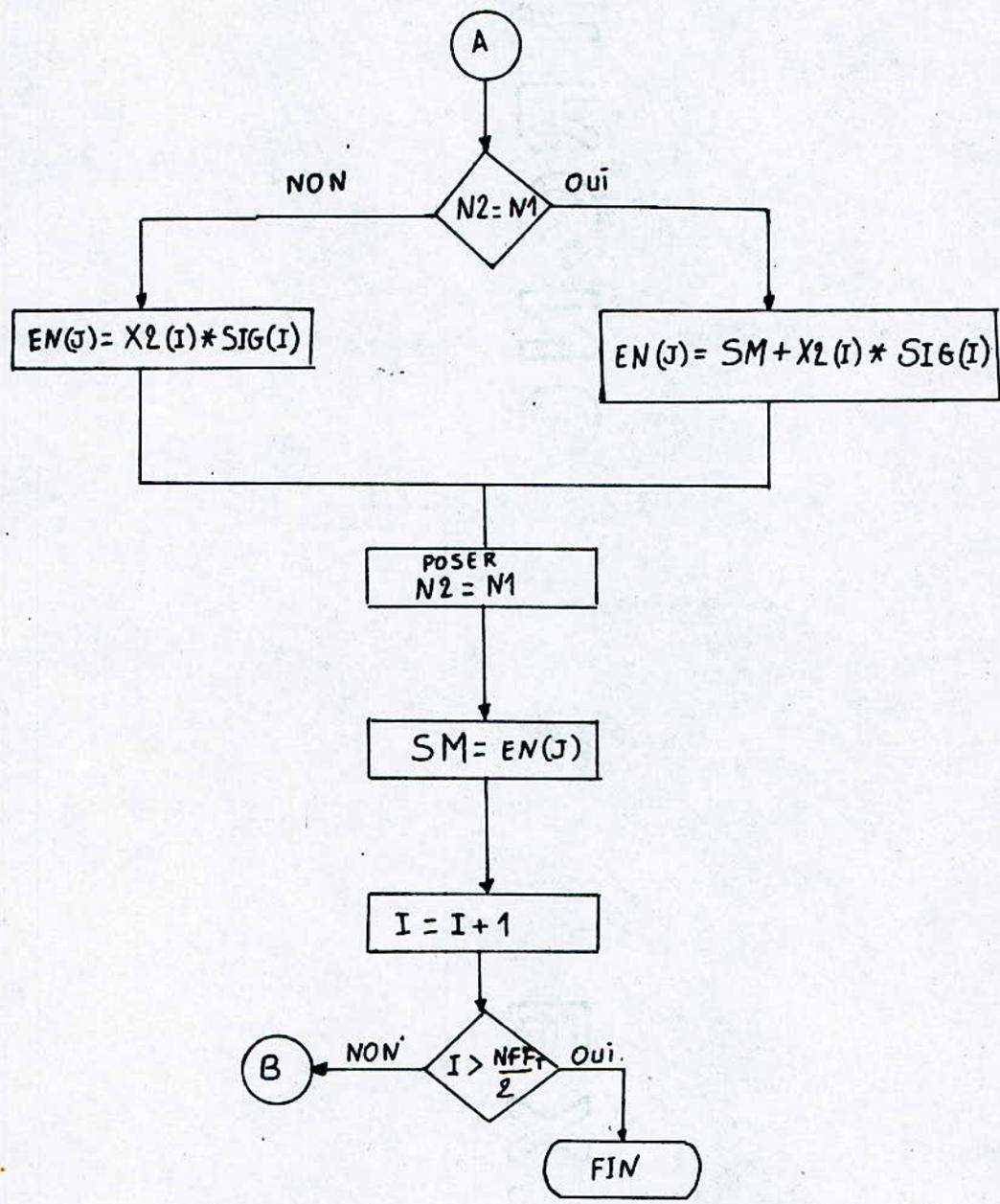
Avec ②, on obtient seulement la forme du cepstre, mais non pas les coefficients cepstraux.

..../....

ORGANIGRAMME DU SOUS PROGRAMME

CALCUL DU NUMERO DU FILTRE SIMULÉ





Explication de l'organigramme de la simulation des filtres triangulaires.

Le but est de calculer l'énergie de chaque raie présente dans le filtre. On utilise un banc de filtres triangulaires avec un recouvrement de 50 %. Ainsi, toutes les raies seront elles contenues dans les filtres et les pertes en énergie seront minimisées.

L'objectif de cette opération est :

- de simuler un vocodeur à canaux
- de suivre une échelle MEL.

Les raies qui se trouvent dans l'intersection des deux filtres peuvent avoir une énergie pondérée par l'un des deux (celui qui peut la pondérer mieux que l'autre).

Après avoir calculé la fréquence correspondante à chaque raie et suivant le numéro du filtre, on calcule le coefficient de pondération X_2 (I).

Quand on pose $N_2 = N_1$, ce n'est qu'une astuce de calcul permettant d'échanger un filtre par un autre.

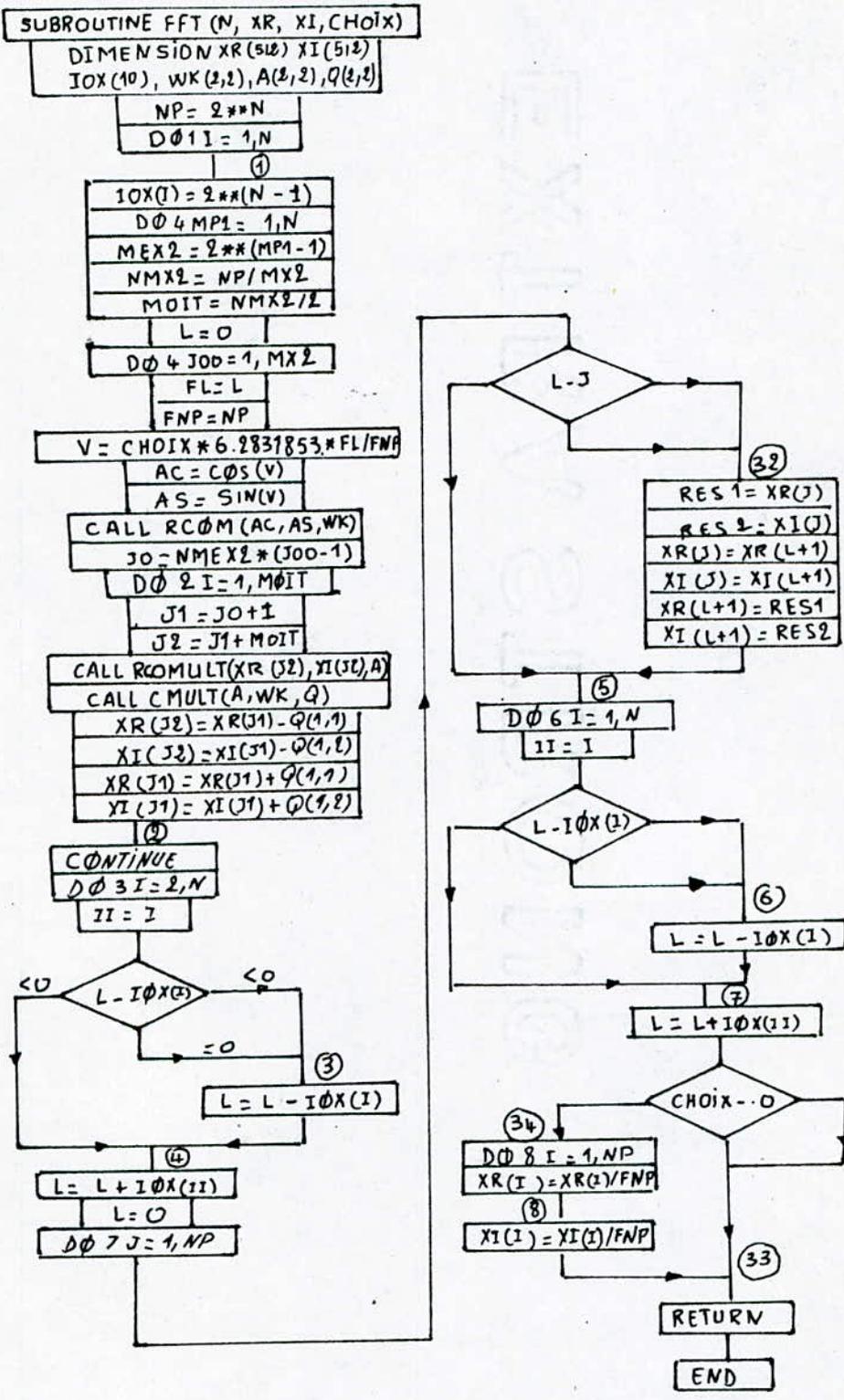
L'énergie de chaque raie est déterminée jusqu'à un nombre égal à la moitié de points de calcul FFT car le spectre est symétrique.

Finalement, les coefficients cepstraux sont obtenus par la transformation en cosinus de la somme des énergies dans chaque filtre en vérifiant la relation :

$$MFc_c(n) = \frac{1}{NF} \sum_{k=1}^{NF} \text{Log}(E(k)) \cos\left(n\left(k - \frac{1}{2}\right) \frac{\pi}{NF}\right)$$

$$1 \leq n \leq M$$

.../...



Interprétation du programme de La F.F.T.

Le graphe de fluence de COOLEY schématise l'organisation méthodique des calculs de la T.F.D. pour donner la F.F.T.

Cette organisation méthodique des calculs a son actif la réduction des calculs de N^2 opérations à $N \log_2 N$ opérations.

Ce programme, à part ses nombreuses qualités, est un graphe de fluence en langage FORTRAN.

Parmi ces avantages, on cite :

- Choix très significatif des variables dans les boucles \emptyset
- En plus de sa simplicité, ce programme comporte beaucoup d'astuces qui lui permettent d'utiliser un minimum d'instructions.

.../...

B. - PROGRAMMES /

INTRODUCTION :

Dans cette partie, la mise en œuvre de l'analyse de la parole est présentée en langage évolué : FORTRAN 77 sur VAX 11/750 disponible à l'ENP.

.../...

```

*****
***** PROGRAMME DE CALCUL DES COEFFICIENTS PREDICTEURS*****
*****
DIMENSION SR(1024),SI(1024),SRE(1024),SIE(1024),Q(12,12),D(12),C(12)
,A(12,12),Y(1024),R(0:11),FH(512),IIS(1024)
TYPE *,'DONNER NBRE PTS N='
READ *,N
DEPI=6.29318531
*****
** ECHANTILLONNAGE DU SIGNAL SUR N PTS**
*****
DO 20 I=1,N
SR(I)=4*(SIN(DEPI*I*10000/N))
SI(I)=0.0
CONTINUE
DO 30 I=1,N
SRE(I)=SR(I)+(20*(SIN(DEPI*I*240/N)))
SIE(I)=0.0
CONTINUE
*****
** CALCUL ECH FENETRE DE HANNING**
*****
DO 501 I=1,N/2
FH(I)=0.54+0.46*COS(DEPI*(I-.4)/FLOAT(N/2))
FH(N+1-I)=FH(I)
TYPE*,FH(I)
CONTINUE
*****
** PRIACCENTUATION**
*****
DO 400 I=1,N-1
SR(I)=SRE(I+1)-SRE(I)
SI(I)=SIE(I+1)-SIE(I)
CONTINUE
*****
** FENETRAGE**
*****
DO 502 I=1,N
Y(I)=SR(I)*FH(I)
IIS(I)=SI(I)*FH(I)
CONTINUE
*****
** CALCUL DES ELTS DE LA MATRICE B**
*****
DO 1 I=1,11
DO 2 J=1,N+1
R(I)=Y(J)*Y(J+I)+R(I-1)
CONTINUE
CONTINUE
DO 70 I=1,11
WRITE(1,*),'R(',I,')=',R(I)
CONTINUE
DO 2 I2=1,12

```

```

DO 4 J=1,N
B(I2,I2)=Y(J)*Y(J)
CONTINUE
CONTINUE
DO 5 I=1,11
  K=I+1
  B(I,K)=R(K-1)
  B(K,I)=B(I,K)
  write(1,*) 'B(',I,',',K,',')=',B(I,K)
DO 71 J=K+1,12
IF(K.NE.11)GO TO 14
  B(I,K)=R(K-1)
  B(K,I)=B(I,K)
  B(I,J)=R(J-1)
  B(J,I)=B(I,J)
CONTINUE
CONTINUE
DO 6 J=1,11
  D(J)=R(J)
CONTINUE
DO 60 J=1,N-12
  D(12)=Y(J)*Y(J+12)
CONTINUE
*****
** RESOLUTION DU SYSTEME PAR ELIMINATION DE GAUSS**
*****
CALL GAUSS(NQ,B,D,C,NON)
END

```

```

SUBROUTINE GAUSS(NQ,B,D,C,NON)
DIMENSION B(12,12),D(12),C(12)
TYPE*, 'DONNER LA VALEUR DE NQ'
READ *,NQ
NON=0
N VAUT 1
IF(NQ-1)32,22,32
IF(B(1,1))24,33,24
C(1)=D(1)/B(1,1)
RETURN
NON=1
RETURN
*****
** CAS GENERAL, TROUVER LE PIVOT**
*****
32 NM1=N-1
DO 19 I=1,NM1
  GRAND=ABS(B(I,I))
  L=I
  IPLUS1=I+1
DO 11 J=IPLUS1,NQ
  IF(ABS(B(J,I))-GRAND)11,11,7

```

```

7      GRAND=ADD(B(J,I))
      L=J
11     CONTINUE
      *****
      ** INTERVERTIR SI NECESSAIRE **
      *****
      IF(GRAND)6,9,8
      NON=1
      RETURN
      IF(L-I)10,17,10
      DO 16 J=1,N
      TEMP=D(L,J)
      B(L,J)=B(I,J)
      B(I,J)=TEMP
      TEMP=D(L)
      D(L)=D(I)
      D(I)=TEMP
      *****
      ** REDUIRE LES COEFFICIENTS A UN **
      *****
7      DO 19 J=IPLUS1,NQ
      QUOT=B(J,I)/B(I,I)
      DO 25 K=IPLUS1,NQ
      B(J,K)=B(J,K)-QUOT*B(I,K)
      D(J)=D(J)-QUOT*D(I)
      *****
      ** ETAPE DE SUBSTITUTION **
      *****
3      IF(B(NQ,NQ))21,23,21
      NON=1
      RETURN
1      C(NQ)=D(NQ)/B(NQ,NQ)
      I=NQ-1
7      SOM=0.
      IPLUS1=I+1
      DO 26 J=IPLUS1,NQ
      SOM=SOM+D(I,J)*C(J)
      C(I)=(D(I)-SOM)/B(I,I)
      write(1,*)C(I,I,C(I,I))=1,C(I,I)
      I=I-1
      IF(I)30,30,27
0      RETURN
      END

```

```

C      *****
C      PROGRAMME D'ANALYSE
C      DE SPECTRALE METHODE
C      DE FILTRES SIMULES
C      *****
C      CALCUL DES COEFFICIENTS SPECTRAUX(FREQUENCES MEL)
C      *****
C      REAL XR(512),XI(512),SIR(512),SII(512),RIS(512)
C      +,IIS(512),SIG(512),FH(512),EN(27)

C      TYPE *, 'DONNER NBRE PTS FFT N= ', PUIS NBRE MOITIE DE N M=N/2
C      READ *,N
C      READ *,M
C      DEFI=C.20313531
C      *****
C      ** ECHANTILLONNAGE DU SIGNAL SUR N POINTS **
C      ** SIMULATION DE DEUX FONCTIONS SINUSOIDALES**
C      *****
C      DO 21 I=1,N
C      SIR(I)=(4*(SIN(DEFI*I*20/N)))+(20*(SIN(DEFI*I*150/N)))
C      SII(I)=0.0
21    CONTINUE
C      CALL ECH (SIR,SII,N)
C      *****
C      ** CALCUL ECH FENETRE HAMMING**
C      *****
C      DO 501 I=1,N/2
C      FH(I)=0.544046 * COS(DEFI*(I-.5)/FLOAT(M))
C      FH(N/2-I)=FH(I)
501  CONTINUE
C      *****
C      ** PENETRATION DU SIGNAL ET PREACCENTUATION**
C      *****
C      CALL PREACC(SIR,SII,XP,XI)
C      CALL FENSIG(XR,XI,FH,SIR,SII,N)
C      *****
C      ** CALCUL DE LA FFT DIRECTE **
C      *****
C      CHOIX=-1
C      CALL FFT(NC,SIR,SII,CHOIX)
C      *****
C      ** CALCUL MODULE DU SPECTRE**
C      *****
C      DO 500 I=1,N
C      SIG(I)=SIR(I)*SIR(I)+SII(I)*SII(I)
500  CONTINUE
C      CALL FILTSIM(SIG,EN)
C      CALL SEPST(EN,CI)
C      END

C      *****
C      * SOUS PROGRAMME DE CALCUL DE PREACCENTUATION**
C      *****

```

```

SUBROUTINE PREAC(SIR,SII,XR,XI)
C SIR:TABLEAU DES ECHANTILLONS DU SIGNAL INITIAL REEL
C SII:TABLEAU DES ECHANTILLONS DU SIGNAL INITIAL IMAGINAIRE
C XR:TABLEAU DES ECHANTILLONS REELS DU SIGNAL ET DE LA FFT
C XI:TABLEAU DES ECHANTILLONS IMAGINAIRES DU SIGNAL ET DE LA FFT
DIMENSION XI(512),XR(512),SIR(512),SII(512)
DO 438 I=1,511
XR(I)=SIR(I+1)-SIR(I)
XI(I)=SII(I+1)-SII(I)
438 CONTINUE
RETURN
END

SUBROUTINE FEN510(XR,XI,FH,SIR,SII,N)
C H: FENETRE DE HANNING
DIMENSION XR(512),XI(512),FH(512),SIR(512),SII(512)
N=512
DO 502 I=1,N
SIR(I)=XR(I)*FH(I)
SII(I)=XI(I)*FH(I)
502 CONTINUE
RETURN
END

C OUVERTURE DU FICHIER POUR VISUALISER
C ECHANTILLONS DU SIGNAL RECONSTITUE
C FERMETURE DU FICHIER DES RESULTATS

SUBROUTINE FFT(NC,XR,XI,CHOIX)
REAL XR(512),XI(512),IOX(10),AK(2,2),A(2,2),Q(2,2)
TYPE *, DONNER VALEUR PUISSANCE DE 2
READ *,NC
NP=2**NC
DO 1 I=1,NC
IOX(I)=2**((NC-I))
CONTINUE
*****
* FIXATION DE LA COLONNE DE CALCUL *
*****
DO 2 COLONE=1,NC
MEX2=2**((COLONE-1))
NMEY2=NP/MEX2
M0IT=MEX2/2
L=0
*****
* FIXATION DU TREILLIS DE CALCUL *
*****
DO 3 JOO=1,MEX2
FL=FLOAT(L)
FNP=FLOAT(NP)

```

V=CHOIX*G,000123*(FLY/FNP)

AC=COS(V)

AS=SIN(V)

CALL ROOM(AC,AS,WK)

J0=N*EX2*(J00-1)

* EFFECTUE CALCUL SUR TREILLIS FIXE *

DO 4 I=1,MOIT

J1=J0+I

J2=J1+MOIT

CALL ROOM(XR(J2),XI(J2),A)

CALL CMULT(A,WK,Q)

XR(J2)=XR(J1)-Q(1,1)

XI(J2)=XI(J1)-Q(1,2)

XR(J1)=XR(J1)+Q(1,1)

XI(J1)=XI(J1)+Q(1,2)

CONTINUE

* CALCUL DE L'EXPOSANT DE W *

DO 5 I=2,NC

II=I

IF(L-IOX(I)) 6,7,7

L=L-IOX(I)

CONTINUE

L=L+IOX(II)

CONTINUE

CONTINUE

L=0

* RENVERSMENT DINAIRE *

DO 8 J=1,NP

IF(L-J)10,9,9

RES1=XR(J)

RES2=XI(J)

XR(J)=XR(L+1)

XI(J)=XI(L+1)

XR(L+1)=RES1

XI(L+1)=RES2

CONTINUE

DO 11 I=1,NC

II=I

IF(L-IOX(I)) 13,12,12

L=L-IOX(I)

CONTINUE

L=L+IOX(II)

CONTINUE

IF(CHOIX-.0)14,15,15

DO 16 I=1,NP

XR(I)=XR(I)/FNP

XI(I)=XI(I)/FNP

```

16 CONTINUE
15 RETURN
END

C *****
C * REPRESENTATION COMPLEXE SOUS *
C * FORME MATRICIELLE *
C *****
SUBROUTINE RCON(R,X,A)
DIMENSION A(2,2)
A(1,1)=R
A(1,2)=X
A(2,1)=-X
A(2,2)=R
RETURN
END

C *****
C * MULTIPLICATION DE MATRICES *
C *****
SUBROUTINE CMULT(A,B,C)
DIMENSION A(2,2),B(2,2),C(2,2)
C(1,1)=A(1,1)*B(1,1)+A(1,2)*B(2,1)
C(1,2)=A(1,1)*B(1,2)+A(1,2)*B(2,2)
C(2,1)=-C(1,2)
C(2,2)=C(1,1)
RETURN
END

C *****
C * CALCUL NUMERO DU FILTRE SIMULE
C * COEFFICIENTS DE PONDERATION X1 ET
C *****
SUBROUTINE FILTSIM(SIG,EN)
SIG=MODULE DU SIGNAL PREACCENTUE ET FENETRE
EN=ENERGIE A LA SORTIE DES FILTRES
REAL X1(250),X2(250),EN(25),ENP(26),SIG(512)
INTEGER NUM(512),FCOUP(27),FCENT(26)
DATA FCENT /50,150,250,350,450,550,650,750,850
+ ,950,1075,1225,1400,1600,1850,2175,2525,2900,3325
+ ,3775,4250,4775,5325,5900,6525,7175 /
DATA FCOUP /0,100,200,300,400,500,600,700,800
+ ,900,1000,1150,1300,1500,1700,2000,2350,2700
+ ,3100,3550,4000,4500,5050,5600,6200,6850,7500 /
N2=1
SM=0.0
DO 10 I = 1,250
F=12.0 * 500 * FLOAT(I) / FLOAT(250)
WRITE(*,*) (FREQUENCE EST ,F
DO 15 J=1,26
X1(I)=0
X2(I)=0
IF (F.LT.FLOAT(FCOUP(J)) .OR. F.GE.FLOAT(FCOUP(J+1))) GOTO 15
N1=J

```

```

17      IF (FCENT(J) - P175) 16,17,20
18      X1(I)=(F-FCENT(J))/(FCENT(J)-FCENT(J-1))
19      X2(I)=1-X1(I)*X1(I)
20      GOTO 18
21      X1(I)=(F-FCENT(J))/(FCENT(J+1)-FCENT(J))
22      X1(I)=X1(I)*X1(I)
23      X2(I)=1-X1(I)
24      TYPE *, 'COEF DE PONDERATION ', X2(I)
25      TYPE *, 'SIG=', SIG(I)
26      IF (N2.EQ.N3) THEN
27      ENF(J)=BN*X2(I)*SIG(I)
28      GOTO 12
29      ELSE
30      ENF(J)=X2(I)*SIG(I)
31      N2=N3
32      END IF
33      GOTO 12
34      CONTINUE
35      EN=ENF(J)
36      CONTINUE
37      DO 20 N=1,25
38      EN(N)=ENF(N)
39      TYPE *, 'ENERGIE', N, EN(N)
40      CONTINUE
41      RETURN
42      END
43      *****
44      * SOUS PROGRAMME DE CALCUL DES COEFFICIENTS CEPSTRAUX*
45      * A PARTIR DES ENERGIES DANS LES FILTRES *
46      *****
47      SUBROUTINE CEPST(EN,CI)
48      EN, ENERGIE DANS LES FILTRES
49      CI, COEFFICIENTS CEPSTRAUX
50      DIMENSION EN(25), CI(17)
51      DATA PI/3.1415926/
52      DO 10 I=1,16
53      CI(I)=0.0
54      DO 10 J=1,25
55      TYPE *, 'ENERGIE', EN(J)
56      CI(I)=CI(I)+ALOG(EN(J))*COS(I*(J-.5)*PI/25)
57      CONTINUE
58      DO 100 I=1,16
59      TYPE *, 'CEPSTRE', I, CI(I)
60      CONTINUE
61      CI(I)=CI(I)/25
62      TYPE *, 'CEPSTRE', I, CI(I)
63      RETURN
64      END

```

DONNER NBR PTB FFT N= , PUIS NBR MOITIE DE N MEM/2
DONNER VALEUR PUISSANCE DE 2

ENERGIE	1	4.4953367E-03
ENERGIE	2	5.6680003E-03
ENERGIE	3	5.6620525E-03
ENERGIE	4	5.6680786E-03
ENERGIE	5	9.2846546E-03
ENERGIE	6	5.5344701E-03
ENERGIE	7	5.6681167E-03
ENERGIE	8	5.6682350E-03
ENERGIE	9	5.6673586E-03
ENERGIE	10	5.8793379E-03
ENERGIE	11	8.5554915E-03
ENERGIE	12	8.7760901E-03
ENERGIE	13	1.1433733E-02
ENERGIE	14	1.1620129E-02
ENERGIE	15	1.7411688E-02
ENERGIE	16	2.0047538E-02
ENERGIE	17	2.0279838E-02
ENERGIE	18	2.3152160E-02
ENERGIE	19	2.5774414E-02
ENERGIE	20	99.64924
ENERGIE	21	2.8910397E-02
ENERGIE	22	3.1528246E-02
ENERGIE	23	3.1778372E-02
ENERGIE	24	3.4627840E-02
ENERGIE	25	1.2246924E-02

DEPSTRE	1	-15.64332
DEPSTRE	2	2.526126
DEPSTRE	3	5.284545
DEPSTRE	4	-11.05928
DEPSTRE	5	5.324340
DEPSTRE	6	-7.416858
DEPSTRE	7	-0.3609142
DEPSTRE	8	6.460425
DEPSTRE	9	-4.866278
DEPSTRE	10	8.050504
DEPSTRE	11	-0.6303654
DEPSTRE	12	-5.236849
DEPSTRE	13	6.407054
DEPSTRE	14	-10.77888
DEPSTRE	15	3.490323
DEPSTRE	16	-0.5065429

SIMULATION



LA SIMULATION /

Généralités.

La simulation est une méthode de vérification expérimentale par un traitement sur ordinateur (qui remplacera les différentes phases d'expérimentation dans un laboratoire sans encourir le risque de détérioration du matériel). On ne testera que le principe, et une fois le montage figé, l'ingénieur le réalisera matériellement.

de manière à représenter le phénomène physique sur ordinateur, on a recours à des modèles algorithmiques très rapides, qui nous font gagner du temps.

On peut simuler un système, un signal ayant les caractéristiques du signal réel, c'est-à-dire on reconstitue mathématiquement le signal. Dans la reconnaissance des formes de la parole, on simule certains mots et on les mémorise et puis on compare l'enregistrement simulé et l'enregistrement réel.

Dans cette étude, on fait un programme de simulation de deux signaux sinusoidaux et suivant les fréquences utilisées (on utilise le triangle vocalique de Delattre Emerit) on va utiliser certaines voyelles.

Le programme de simulation de ces 02 signaux sinusoidaux va permettre de contrôler et de vérifier le programme de l'analyse cepstrale suivant les fréquences utilisées.

Remarque

Pour simuler un signal comme celui de la parole, il faut bien connaître tous ses composants pour le reproduire correctement, ce qui est très difficile car il faut faire une étude approfondie du signal et il n'est pas permis de négliger même les phénomènes parasites.

Programmation de la simulation:

Pour la simulation des voyelles, on considère la bande de fréquences comprise entre 250 et 4.000 HZ : cette large bande est très instructive.

[u] : comporte la presque totalité de ses composantes informatives dans des fréquences inférieures à 1.000 HZ avec une forte concentration vers 350 HZ \Rightarrow voyelle grave.

[a] : Comporte une forte et large concentration des composantes entre 700 et 1.400 HZ (fréquences moyennes) \Rightarrow voyelle médiane.

[i] : Présente un fondamental très ample ainsi qu'un fort taux d'harmoniques dans la bande des 200 HZ, 300 HZ, 1.800 et 2.800 HZ.

Toutes ces voyelles qui comportent un taux important de composantes HF sont les voyelles aiguës.

(C)NC'LLIS'ION

CONCLUSION /

Il n'y a pas de normes précises pour comparer l'efficacité des méthodes d'analyse de la parole.

Seulement, on peut dire que chacune des méthodes a des avantages et des inconvénients par rapport à l'autre.

Si on prend le côté simplicité d'application, la méthode d'autocorrélation est très pratique; elle s'applique facilement dans le cas de la modélisation du signal de la parole et permet de chercher les paramètres pertinents du signal, qui peuvent être utilisés directement pour la synthèse. On modélise entièrement le processus de la production de la parole en tenant compte :

- des particularités du signal vocal ;
- des propriétés de l'appareil auditif (insensibilité à la phase).

La méthode d'analyse cepstrale d'après la décroissance en 1 du cepstre qui est directement liée au pôle de plus fort module (bande laⁿ plus étroite) et il suffit d'un petit nombre de coefficients cepstraux pour caractériser bien le signal de la parole.

La difficulté la plus importante réside dans l'absence d'un critère objectif, représentatif de la qualité du signal reconstruit au plan de la perception auditive.

D'autre part, puisque le nombre de coefficients cepstraux est réduit, donc le nombre de bits occupés est réduit ce qui diminue le temps de calcul par rapport à celui de la prédiction linéaire.

Et pour ne pas encombrer les canaux de transmission par des débits binaires élevés (\Rightarrow élargissement de la bande passante utile) des méthodes de compressions numériques des signaux analogiques ont été développées.

Applications de la reconnaissance de la parole

- Aide aux handicapés
- Aide à la sécurité
- Commande des machines-outils
- Le dialogue avec les machines est moins contraignant que de frapper sur un clavier

La simulation permet de développer les systèmes numériques ou cartes destinées au traitement de la parole.

RECOMMENDATIONS

Recommandations pour la réalisation d'un système complet de reconnaissance de la parole.

Au cours de notre étude, on a élaboré deux programmes d'analyse de la parole en Fortran 77 sur le VAX 11/750

- l'un sur l'analyse cepstrale ;
- l'autre sur la prédiction linéaire (méthode d'autocorrélation).

Ces programmes sont loin de représenter d'une manière précise le signal continu de la parole, puisque ce dernier est non stationnaire et les zones de non stationnarité sont le plus souvent les plus riches en informations.

On ne peut pas se contenter des hypothèses prises.

Donc, beaucoup reste à faire dans le domaine de la reconnaissance de la parole.

Pour faire l'étude complète du signal, il faut un travail de longue haleine.

Dans notre cas, on s'est limité à l'analyse de la parole.

Néanmoins, notre travail peut servir à tous ceux qui veulent continuer dans ce domaine, en leur fournissant les méthodes d'analyses que nous avons choisies et ils peuvent l'enrichir encore plus.

LINNEXE

ANNEXE A

Complexité relative	Taux d'informations K Bits
HIFI	1ère classe
	200 64
<ul style="list-style-type: none"> - <u>Qualité radiophonique</u> . bande passante 0- 7 KHZ 	
Signal reconstruit indispensable du signal original.	2ème classe
	8
<ul style="list-style-type: none"> - <u>Qualité téléphonique</u> . Bande passante 0,3 - 3,4 KHZ . rapport S/B 26 dB (subjectif) . Distorsion harmonique < 3 % 	
"Communication quality" signal reconstruit très intelligible. Distorsions audibles	3ème classe
	8
<ul style="list-style-type: none"> - <u>Qualité sub téléphonique</u> . BP 0,3 - 3,4 KHZ . Très bonne intelligibilité . Distorsions perceptibles 	
<ul style="list-style-type: none"> - 100 vocodeur LPC vocodeur à canaux - 200 vocodeur à transformation orthogonale - 500 vocodeur à formants - 1000 Analyse/Synthèse par règles 	4ème classe
	4 2 1
<ul style="list-style-type: none"> - <u>Qualité synthétique</u> . Résonances peu naturelles et métalliques . Reconnaissance du locuteur souvent difficile . Intelligibilité réduite . Performances très variables (locuteur, micro, bruit, 	

C l a s s e	Bande passante	Débits binaires	S/B	Distorsions
1ère classe (haute fidélité)	8 à 20 KHZ	100 à 300 K bits/s	< 30%	< 3%
2 ème classe	300 à 3.400 HZ	64 à 16 K bits/S		
3ème classe		16 et 5 K bits/S		

Renvoi bibliographique : [11] INRIA .

Pour des débits inférieurs, on utilise des techniques dans lesquelles on effectue une paramétrisation complète du processus de production de la parole (le vocodeur à LPC, le vocodeur à canaux, le vocodeur à formants...).

Codage optimum des paramètres transmis..

Qualité obtenue : Le signal décodé perd son naturel.

La reconnaissance du locuteur est parfois difficile et les performances du vocodeur sont très sensibles aux conditions de numérisation du signal.

Les méthodes de la 4ème classe sont utilisées dans des applications de stockage de la parole, de réponse vocale et de transmission à débit très faible.

.../...

ANNEXE.B :

Comparaison des différentes méthodes

La reconnaissance de la parole est un problème très complexe.

On a deux approches pour la reconnaissance :

1°/ Approche globale :

Elle consiste à reconnaître par mot des phrases au sens d'une distance dynamique. Ces mots sont contenus dans un dictionnaire qui consiste à mémoriser les unités de références.

L'analyseur acoustique permet l'extraction des paramètres pertinents et de réduire le débit pour la mémorisation (l'unité linguistique est le mot) (d'après NT/LAA/TSS/79).

2°/ Approche analytique :

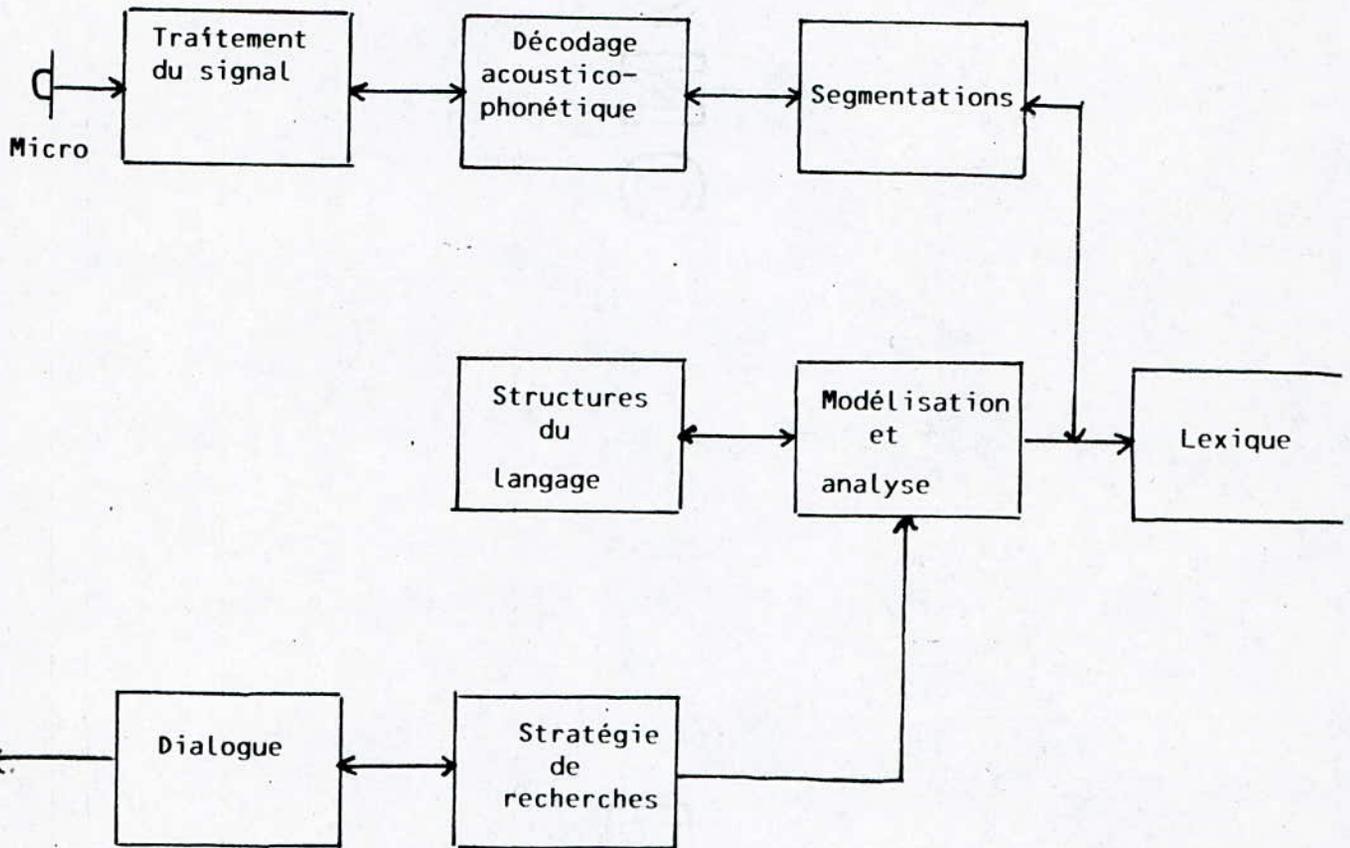
L'unité linguistique ici est le phonème.

La difficulté dans cette approche est la détermination des frontières entre phonèmes (effets de co-articulation).

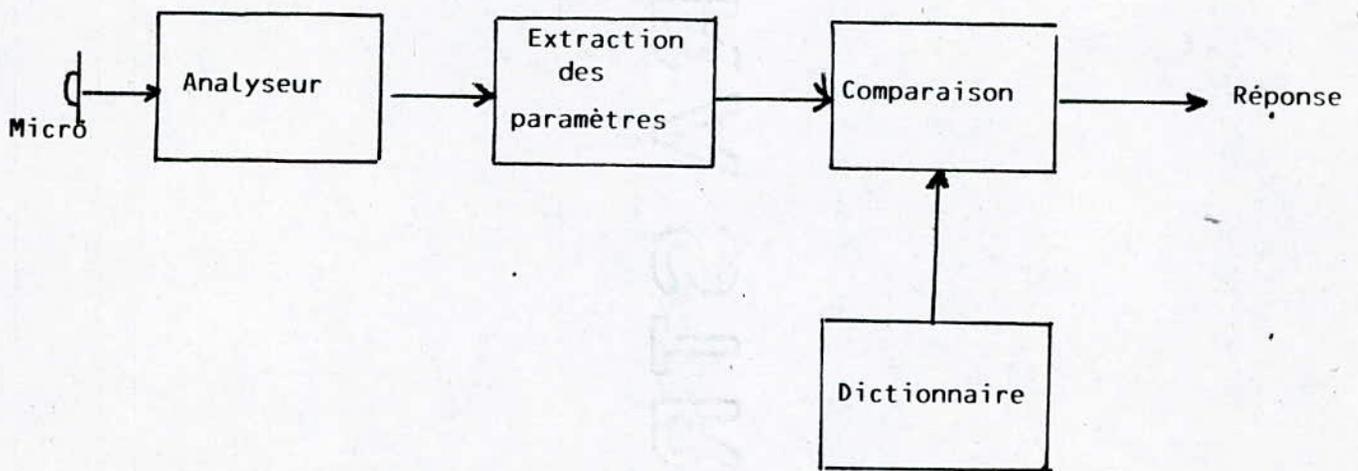
L'étape de reconnaissance

Le reconnaisseur établit des distances entre références et segments de parole dans la phrase et identifie la phrase inconnue.

.../...



Eléments d'un système de reconnaissance de la parole
 (d'après J.P HATON)



Reconnaissance des mots isolés
 (d'après J.J. MARIANI et J.L. GAUVIN)

R E F E R E N C E S B I B L I O G R A P H I Q U E S

- 1 - "Traitements Numériques du signal " M. KUNT MASSON 1981.
- 2 - "Méthodes et techniques de traitement du signal et applications aux mesures physiques". Tome 1 - J. MAX 1981.
- 3 - "Traitement numérique du signal" M. BELLANGER 1981.
- 4 - "Calcul des transformations de Fourier rapides " J. LIFERMAUN
- 5 - "Echantillonnage" J. MARCUS
- 6 - "Les filtres numériques " AUMIAUX
- 7 - "Techniques numériques appliquées au calcul scientifique" PELLETIER.
- 8 - "Méthodes numériques appliquées (avec nombreux problèmes) M. BOUMHRAT
A. GOURDIN
- 9 - "Electronique appliquée à la transmission de l'information". Tome 1
et tome 2. J. HERVE-MASSON 1981, 1982.
- 10 - "Cours de phonétique acoustique" E. EMERIT (1977).Ed. SNED.
- 11 - " Bulletin de liaison de la recherche en informatique et en automatique".
Institut National de Recherche en Informatique et en Automatique
INRIA n° 110/1986

.../...

Reuves

- 12 - Electroniques applications n° 46
- 13 - L'onde Electrique Août-Septembre 1981 "Techniques de codage de la parole".
- 14 - Le nouvel automatisme Septembre-octobre 1979.
- 15 - "Modelisation et reconnaissance de la parole" Yves GRENIER
Ecole Nationale Supérieure des Télécommunications. Département
Système et Communication.
- 16 - THESE DE MAGISTER A. MENACER.

Projets de fin d'étude :

- "Les techniques de la synthèse de la parole". A. MEZAOUI / A. KECHID Janvier 1987
- "Les principales méthodes d'analyse de la parole". SELLAM M. / NOUR F.Z. Juin 1986
- "Simulation software de l'E.E.G". A. AIT AKLI / H. GRAINE Janvier 1987.
- "Etude comparative des Méthodes de traitement numérique de l'E.E.G".
A. AGGOUN
D. AIT BOUDAUD Juin 1986