



**ECOLE NATIONALE POLYTECHNIQUE**

**DEPARTEMENT : ELECTRONIQUE**

**PROJET DE FIN D'ETUDES**

**SUJET**

**synthèse de la parole a l'aide de  
L'EVA III  
(analyse sonographique , synthèse formantique)**

Proposé par :

M<sup>lle</sup> M. GUERTI

Etudié par :

M. B. GOUADFEL  
M. A. RABHI

Dirigé par :

M<sup>lle</sup> M. GUERTI

**PROMOTION : juin 85**

DEDICACES :

-/A mon père

Abdou

-/A)mes parents

à toute ma famille

à tous mes amis(es)

Belaïd

### REMERCIEMENTS :

Nous tenons à remercier vivement notre promoteur M<sup>elle</sup> M. GUERTI, pour toute l'aide qu'elle nous a apportée le long de ce travail.

Nous sommes particulièrement reconnaissants à Fethi et Djamel pour leur contribution à la réalisation du polycopié.

Nous adressons également nos remerciements aux responsables de l'I.L.P (Institut de Linguistique et de Phonétique) qui nous ont permis de visiter leur laboratoire.

Que tous ceux qui ont contribué de près ou de loin à notre formation trouvent ici l'expression de notre profonde gratitude.

## S O M M A I R E

INTRODUCTION	i 1
I. ANATOMIE DE L'APPAREIL PHONATOIRE HUMAIN ET PRODUCTION DE LA PAROLE:	I.1
I.1 Anatomie de l'appareil phonatoire humain	I.1
I.2 Principe de fonct <sup>t</sup> de l'appareil phonat.	I.2
I.3 L'Excitation du conduit vocal	I.2
I.4 Def. et principales caract. acoust.	I.4
I.5 Classification des sons du langage	I.5
I.6 Conclusion	I.7
II. ANALYSE DE LA PAROLE	II.1
II.1 Nature math. du signal de la parole	II.1
II.2 Les "outils math." de l'analyse de la parole et leurs limitations	II.3
II.3 Les méthodes d'analyse	II.4
III. LES FORMANTS ET LES RESONATEURS	III.1
III.1 Def. et caract. des formants	III.1
III.2 Les formants et le timbre d'un son	III.4
III.3 Les voyelles et les foemants	III.5
III.4 La formation des formants et la resonance	III.7
IV. SYNTHESE DE LA PAROLE	IV.1
IV.1 Historique	IV.1
IV.2 Les techniques de synthèse	IV.3
IV.3 Les méthodes de synthèse	IV.15
IV.4 L'Electrical Vocal Analog (E.V.A. III)	IV.18
CONCLUSION.	



## INTRODUCTION

La parole, phénomène, des plus naturels, a toujours été ressentie comme une énigme par les civilisations de tous les âges. Aussi, a-t-on cherché à étudier la voix, sa formation et sa production dès les temps les plus reculés.

Depuis, les démarches analytiques de la plus haute antiquité jusqu'aux "têtes parlantes" de l'ABBE Mical (Moyen Âge), aucun résultat probant n'a été atteint. Il a fallu attendre l'éveil de la pensée scientifique, au XVIII<sup>e</sup> siècle, pour que le rêve tend caressé par l'homme de faire parler la machine, se réalise enfin.

En 1779, le Danois C.G. KRATZENSTEIN présenta les cinq résonateurs qui servaient chacun à reproduire l'une des voyelles [a], [e], [i], [o]; [u]. Après cette réalisation ce fut "l'épopée" des machines parlantes mécaniques jusqu'à 1922. C'est à cette date qu'eurent lieu les premiers essais de simulation électrique de la phonation

et la synthèse  
A partir de là, l'analyse de la parole se sont mises à l'ère de l'électronique. Ainsi les récents progrès de la microélectroniques et de l'informatique font que l'ordinateur n'est pas sans servir le traitement automatique de la parole. Ce dernier recouvre les domaines suivants :

-la synthèse automatique de la parole qui conduit aux machines qui "parlent".

-la reconnaissance automatique et la compréhension de la parole qui aboutissent aux machines qui "entendent" et "comprennent" la voix humaine .

Dans ce dernier domaine, les recherches butent encore contre certains obstacles. Quant à la synthèse des circuits intégrés qui la réalisent sont commercialisés par de nombreuses compagnies internationales.

L'interêt du traitement de la parole n'est plus à démontrer. Pour s'en convaincre, il suffit de jeter un coup d'oeil sur ses diverses domaines d'application (Industrie télécommunications, automobile, grand public, informatique, transport, automatique, ingénierie médicale pour aider les handicapés, etc...).

Dans notre travail nous nous sommes intéressés à l'étude de la production, l'analyse et la synthèse de la parole. Ceci afin de pouvoir étudier et comprendre le synthétiseur E.V.A.III (electrical vocal analog) qui est disponible à l'institut de linguistique et de phonétique de l'université d'Alger.

Pour situer les types d'analyse et de synthèse qui nous intéressent dans ce cas, nous présentons une classification (Lienard, 1977).

On distingue trois systèmes d'analyse et de synthèse qui correspondent à trois démarches différentes:

- D'abord, on peut ne s'intéresser qu'à la transmission du signal de parole sous sa forme temporelle. L'idéal est alors de restituer le signal transmis exactement semblable au signal initial. C'est l'objectif de la "haute-fidélité" en électroacoustique (enregistrement sur bande ou sur disque magnétique), de la diffusion radiophonique etc...:

- Ensuite, on peut envisager la transmission de l'information phonétique; l'analyseur sera alors un système de reconnaissance qui transformera le signal en une suite de symboles phonétiques.

A la synthèse, cette suite phonétique est convertie en parole.



Cependant, la reconnaissance présente encore beaucoup de mystères, d'où la difficulté de réalider l'analyseur dans ce cas.

-Enfin, dans le troisième système, on ne s'intéresse qu'à la transmission du; spectre instantané d'amplitude du signal, qui transporte l'essentiel de l'information perceptible. On trouve donc, côté analyseur un dispositif d'analyse fréquentielle : banc de filtres plus ou moins élaborés, transformateur de Fourier, filtre unique à fréquence centrale variable, ou tout autre système donnant une décomposition fréquentielle du signal. Côté synthèse, le spectre instantané est reconstitué aussi fidèlement que possible puis transformé en son. Il importe peu que le signal synthétisé ressemble au signal initial. L'essentiel est que la différence ne soit pas perceptible pour l'oreille d'un auditeur.

Donc, pour ce troisième système, l'analyse permet une représentation de la parole sous forme d'une suite de spectres instantanés (spectrographie acoustique). Quant à la synthèse, elle transforme en parole la même suite de spectres instantanés. Cette dernière est souvent caractérisée par un nombre réduit de paramètres.

L'analyse sonographique et la synthèse formantique font partie de ce dernier système. De plus, l'E.V.A. III est un synthétiseur à formants et son outil d'analyse est un sonographe.

Alors, ce troisième système fera l'objet de l'essentiel de notre travail sur l'analyse et la synthèse (Chapitres II et IV).

On signalera aussi, que l'étude d'un synthétiseur à formants nécessite une connaissance suffisante de ces derniers. Et ceci car ces paramètres sont des caractéristiques très importantes de la parole.

Ils constituent un "support" de l'intelligibilité du message et des caractéristiques individuelles de la voix. Aussi, nous leur avons consacré tout le chapitre III.

Quant au, chapitre I, on y trouvera l'étude de la production de la parole et la classification des sons.

Enfin, il faut souligner que l'E.V.A.III est d'un grand intérêt. Etant un synthétiseur à formants, il produit de la parole synthétique d'une assez bonne qualité. De plus, il constitue un précieux outil de recherche en phonétique car on peut étudier l'influence perceptive de chaque paramètre de commande (il est possible de modifier un seul paramètre à la fois).



Pour familiariser le lecteur de ce manuscrit avec l'alphabet phonétique dont les signes nous serviront tout au long de notre étude, nous avons jugé utile d'introduire le tableau suivant :

<u>VOYELLES</u>	<u>SEMI-CONSONNES</u>
[ a ] (patte)	[ j ] (abeille) (il y a)
[ a ] (pâte: phonème peu usité, généralement assimilé à [a])	[ ɥ ] (lui)
[ e ] (blé)	[ w ] (oui)
[ ɛ ] (père)	<u>CONSONNES</u>
[ ø ] (feu)	[ b ] (bar)
[ œ ] (peur)	[ k ] (car)
[ a ] (de: phonème inaccentué)	[ d ] (dard)
[ i ] (pire)	[ f ] (fard)
[ o ] (pot)	[ g ] (gare)
[ ɔ ] (port)	[ ʒ ] (jarre)
[ u ] (pour)	[ l ] (laid)
[ y ] (pur)	[ m ] (mare)
[ ɑ̃ ] (rang)	[ n ] (narre)
[ ɔ̃ ] (rond)	[ p ] (par)
[ ɛ̃ ] (rein)	[ R ] (rare)*
[ œ̃ ] (humble: phonème peu usité, généralement assimilé à [ɛ̃]).	[ s ] (Sarre)
	[ z ] (hasard)
	[ t ] (tard)
	[ v ] (Ver)
	[ ʃ ] (char)

\* [R] français, dit "pharyngal", est toujours noté avec une majuscule d'imprimerie pour le différencier d'autres [r] employés dans d'autres langues - Ajouter [ŋ] (ping-pong) d'importation récente.

chapitre I

ANATOMIE DE L'APPAREIL PHONATOIRE HUMAIN ET PRODUCTION  
DE LA PAROLE

- I.1 Anatomie de l'appareil phonatoire humain
- I.2 Principe de fonctionnement de l'appareil phonatoire
- I.3 L'excitation du conduit vocal
- I.4 Définition et principales caractéristiques acoustiques  
de la parole
- I.5 Classification des sons du langage
- I.6 Conclusion

## ANATOMIE DE L'APPAREIL PHONATOIRE HUMAIN ET PRODUCTION NATURELLE DE LA PAROLE

Avant d'aborder le problème de l'analyse et de la synthèse de la parole, il est nécessaire de présenter, ne serait-ce que d'une façon sommaire :

- L'anatomie de l'appareil phonatoire.
- La nature physique et les caractéristiques particulières du signal de la parole.

### I.1 I.1 Anatomie de l'appareil phonatoire humain

L'appareil vocal (fig I.1) est composé principalement de :

I.1.1 Les poumons qui constituent les principaux organes de l'appareil respiratoire. Ils contribuent à la production de la parole en tant que réservoir d'air.

I.1.2 La trachée artère qui comporte à son extrémité supérieure un organe très important : le "larynx" (fig I.2). Ce dernier supporte deux muscles capables de vibrer appelés les cordes vocales. Celles-ci sont séparées par une ouverture : la glotte qui varie au rythme de leurs vibrations.

I.1.3 Le conduit vocal est formé de deux voies :

-La voie orale qui comporte les cavités pharyngale et buccale. Elle commence à la hauteur du larynx et se termine par les lèvres. Grâce aux éléments mobiles qu'elle comporte, tels que le larynx, les mâchoires, les lèvres et surtout la langue, cette voie a de grandes possibilités de changement de forme et de volume. Sa section peut varier de 0 à 20 cm<sup>2</sup> et sa longueur est d'environ 18 cm pour un adulte. Ainsi, une telle cavité permet la production d'un grand nombre de sons différents

-La voie nasale (12 cm de long environ) est composée d'une seule cavité de forme et de volume fixes. Cette voie est située en dérivation sur la voie orale et le



"couplage" entre elles est assuré par le voile du palais (ou velum) ; ce dernier peut se relever à l'horizontale et obstruer complètement la voie nasale, ou s'abaisser à un certain niveau permettant ainsi une nasalisation plus ou moins forte du son produit.

## I.2 Principe de fonctionnement de l'appareil phonatoire :

Le processus de production d'un son par l'appareil vocal (fig I.3 ), peut être divisé en trois étapes :

-La première consiste à produire à l'aide des muscles thoraciques et abdominaux une énergie qui servira à comprimer les poumons pour y chasser de l'air. Cet air cheminera librement jusqu'à l'entrée du larynx.

-Quant à la seconde, elle permet la mise en vibration du débit d'air par un système excitateur se situant soit au niveau du larynx, soit en un point de conduit oral. Ce système 'excitateur est appelé "source d'excitation". Cette dernière peut être sonore ou bruitée.

-Enfin, à la dernière étape, des modifications des caractéristiques acoustiques du signal de source seront apportées par un système de résonateurs (cf chap III). Et l'onde ainsi "construite" sera rayonnée au niveau des lèvres.

## I.3 L'excitation du conduit vocal :

C'est une sorte de "modulation", étroitement liée à la nature du son à produire, du débit d'air issu de la trachée artère. Elle se produit au niveau de la source d'excitation et donne naissance à un signal excitateur. Il existe trois types de source d'excitation :

:



a/ La source sonore :

Cette source située au niveau de la glotte est caractérisée par un mouvement de relaxation des cordes vocales. Ce mouvement impose des variations impulsionnelles, quasi périodiques, de forme presque triangulaire (fig I.4a), du débit d'air à travers la glotte. Le signal excitateur ainsi obtenu est à l'origine de la production des voyelles et de certaines consonnes dites voisées (ou sonores).

b/ La source bruitée :

Elle se situe au niveau d'un resserrement très étroit (constriction) du conduit oral. Elle est caractérisée par un écoulement turbulent de l'air à travers ce rétrécissement. Ceci se traduit par l'émission d'un bruit de friction qui constitue le signal source (ou d'excitation) (fig I.4.b). Ce signal permet la production des consonnes fricatives non voisées (ou sourdes).

c/ La source plosive :

Cette source est une occlusion qui, maintenue en un point du conduit vocal, entraîne un arrêt de l'écoulement de l'air. Ceci se traduit, en amont de l'occlusion, par une augmentation de pression qui après un brusque relâchement provoque une explosion. Le signal ainsi généré a la forme d'une impulsion. Il est à l'origine de la production des consonnes sourdes plosives (ou explosives).

Nous avons présenté les trois types de sources séparément, mais il arrive souvent que deux d'entre elles contribuent à la production d'un même son. Ainsi, on peut avoir une superposition de l'effet d'une des deux dernières sources citées avec celui de la première. Le son produit est alors : soit une fricative voisée dans le cas d'une contribution de la source bruitée, soit une plosive voisée dans le cas de la contribution de la source plosive.

Remarque : Dans le cas de la voix chuchotée, les cordes vocales ne vibrent pas et la glotte ne présente qu'une ouverture partielle. Ainsi, elle est traversée par un débit d'air turbulent et le signal excitateur qui en découle est un bruit prenant naissance au niveau du larynx.

#### I.4 Définition et principales caractéristiques acoustiques de la parole :

##### I.4.1 Définition de la parole :

La parole est une suite de sons assez complexes, produits dans un certain ordre, pour permettre la transmission d'un message parlé.

D'un point de vue acoustique, un son est une vibration provoquée par l'appareil vocal, dans un milieu matériel, dans lequel elle peut se propager pour exciter notre "ouïe".

##### I.4.2 Les principales caractéristiques acoustiques de la parole :

-La fréquence du signal d'excitation issu du larynx et excitant le conduit vocal, est appelée fréquence fondamentale " $F_0$ " (ou pitch en anglais). Elle prend les valeurs 100 à 120 Hz pour un homme, 200 à 300 Hz pour une femme et un peu plus pour un enfant. Le spectre d'un tel signal est riche en harmoniques et présente une pente de -12 dB par octave.

-Les autres signaux d'excitation : les bruits sont aperiodiques, à spectre continu et relativement plat sur une large bande.

-Les limites du spectre de la parole ne sont pas bien déterminées puisqu'elles dépendent du locuteur. Toutefois, dans le cas d'une analyse sonographique\*, 50 à 3000 Hz pour les sons voisés et jusqu'à 8000 Hz les sons non-voisés.

-Les sons peuvent varier par :



.Leur fréquence.

.Leur amplitude qui détermine l'intensité à fréquence constante.

.Leur timbre (cf chap III.2) qui est dû à l'audibilité des harmoniques.

#### I.4 Classification des sons du langage (cf tab I.1)

Cette classification a été réalisée à base de critères liés au type de phonation et au mode d'articulation. Ces critères sont :

-Le mode d'articulation relatif à la manière dont se fait l'écoulement d'air dans le conduit vocal lors de la phonation.

-L'opposition "sonore-sourde" selon que la cavité nasale est obstruée par le velum ou non.

-Le lieu ou point d'articulation : endroit de la constriction maximale du conduit vocal. Partant de ces critères, les sons du langage ont été divisés en deux grandes classes : les voyelles et les consonnes.

##### 1) Les voyelles :

Elles sont caractérisées par :

-un passage libre de l'air à travers le conduit vocal dont la configuration reste quasi stable.

-Une présence de la vibration laryngienne en tant que signal d'excitation.

Il existe des voyelles :

-orales : [a], [u], [i], [e], [o] (par exemple)

-nasales : [ã], [õ], [ẽ] ( " " )

2) Les consonnes :

Leur principale caractéristique est une obstruction du passage de l'air. Elle est due à l'opposition de la langue ou des lèvres qui empêche l'écoulement libre de l'air. Cette obstruction, plus ou moins prononcée, peut être une constriction ou une fermeture (soit momentanée, soit complète) du passage de l'air. Selon sa nature, on distingue diverses sortes de consonnes :

.les fricatives sourdes :

Le signal d'excitation, qui constitue le point de départ de leur production, est un bruit de friction. Il prend naissance au niveau d'une importante constriction du conduit oral.

Exemples : [f] , [s] , [ʃ] .

.les plosives (ou occlusives) sourdes :

Le signal d'excitation qui permet de les produire est aussi un bruit. Mais dans ce cas ce bruit est dû à une source plosive et se présente sous forme d'une impulsion.

Exemples : [p] , [t] et [k] .

.les fricatives et les plosives voisées

Les fricatives et les plosives sourdes ont des correspondantes voisées (ou sonores). L'excitation du conduit vocal est due dans ce cas à une association de l'une des deux sources de bruit avec la source sonore.

Exemples :

les fricatives voisées : [v] , [z] , [ʒ] .

les plosives voisées : [b] , [d] , [g] .

.Les consonnes nasales :

Elles sont caractérisées par la prépondérance du rôle du conduit nasal. Lors de leur formation, la source est sonore. Exemples : [m] et [n] .



.les semi-voyelles (ou semi consonnes) et les liquides :

leurs caractéristiques se rapprochent surtout de celles des voyelles orales. Le conduit vocal présente une certaine stabilité pour les semi-voyelles et un comportement dynamique pour les liquides.

Exemples : [w] , [j] et [ɥ] sont des semi-voyelles.

[l] et [r] sont des liquides.

[r] est aussi appelée une vibrante

Le tableau (I.2) établit les corrélats entre les domaines de l'articulation d'acoustique et de la perception.

#### I.6 Conclusion :

Nous venons de voir les phénomènes qui concourent à la production de la parole, et, la classification des sons qui la composent dans le cas du français. Nous avons alors assez d'éléments pour pouvoir aborder les problèmes d'analyse et de synthèse de la parole.

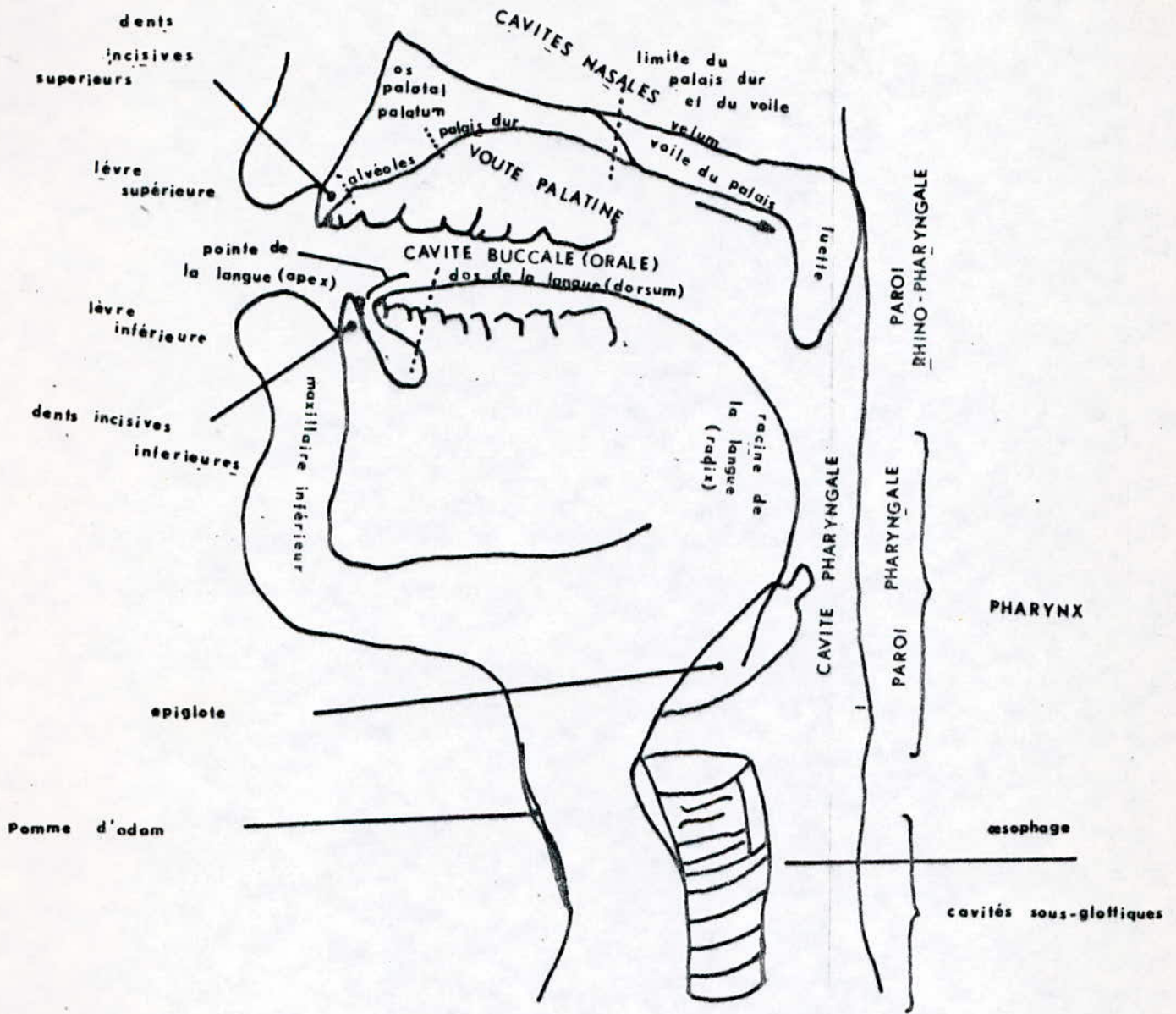


fig-1.1 CONSTITUTION de l'appareil phonatoire humain

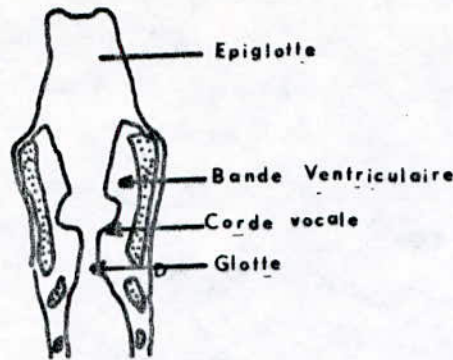


fig:1.2 : Section du larynx.  
(d'après J.GUILBERT. 1979.)

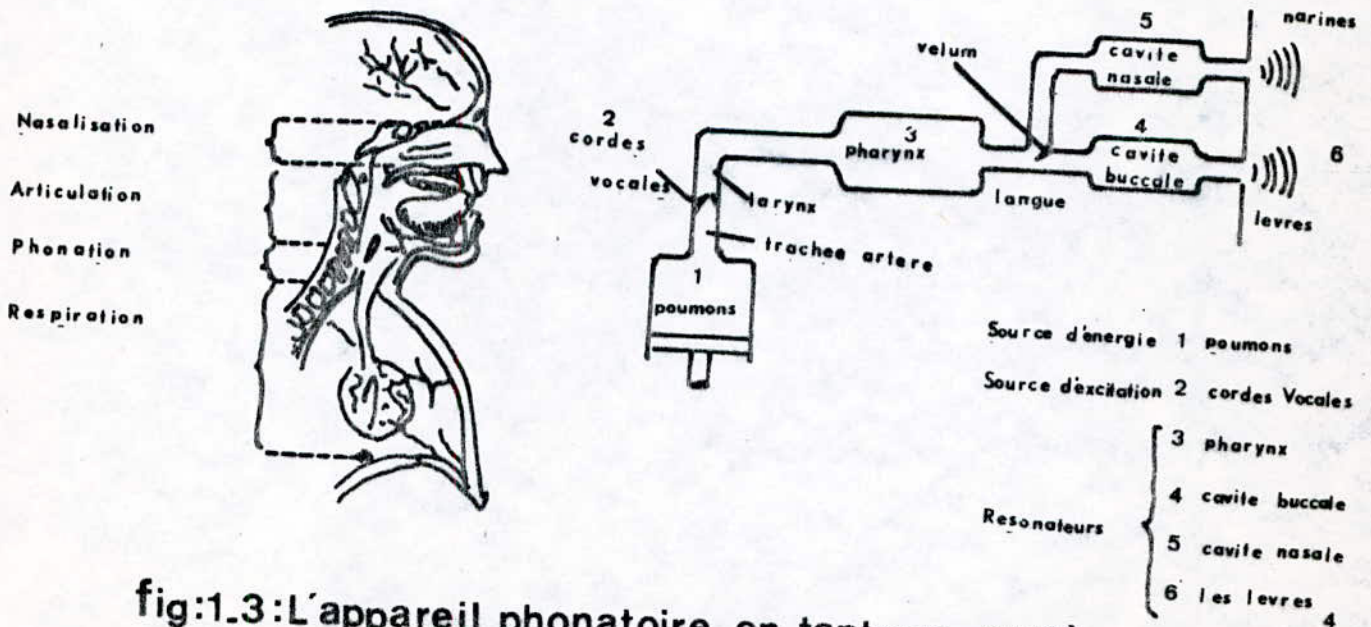


fig:1.3 : L'appareil phonatoire en tant que système acoustique. (GUERTI, M)



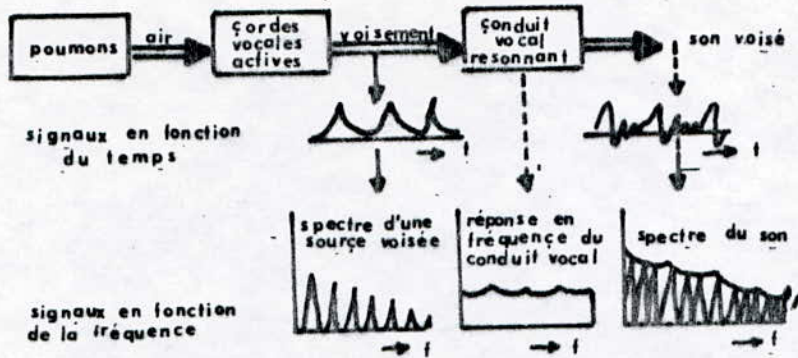


fig.1.4.a Production d'un son voisé

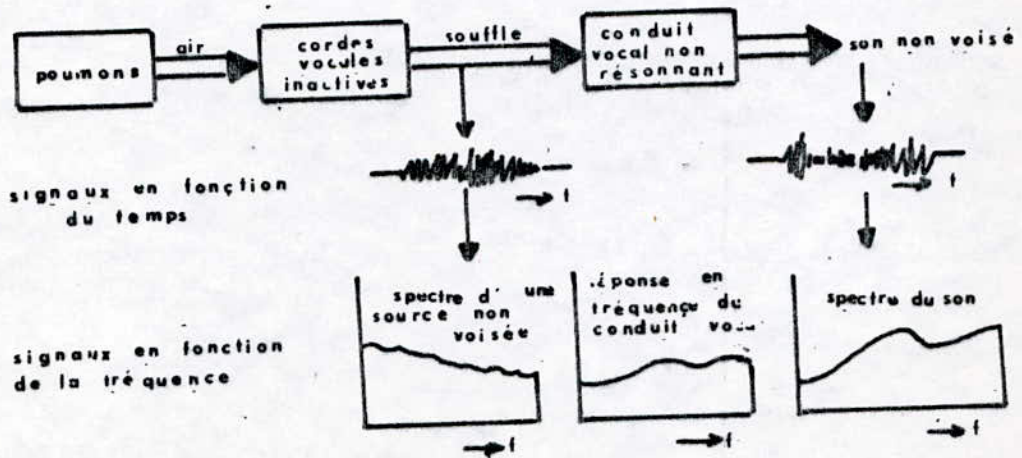


fig-14.b Production d'un son non voisé



Consonnes

		BI-LABIALES	LABIO-DENTALES	DENTALES	PALATO-ALVEOLAIRES	PALATALES	VELAIRES	UVULAIRES
occlusives	SOURDES	p		t			k	
	SONORES	b		d			g	
constrictives	SOURDES		f	s	ʒ : ʃ			
	SONORES		v	z	ʒ̃ : ʒ			
liquides	SONORES			l				R
nasales	SONORES	m		n				
semi-voyelles	SONORES	w/y			j			

Voyelles

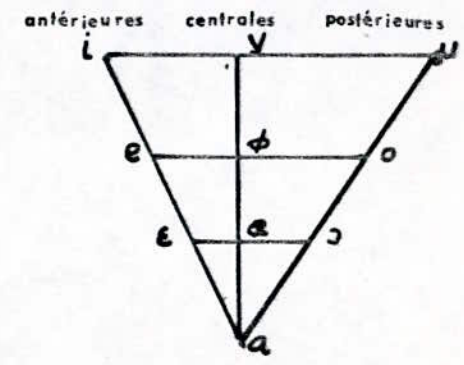
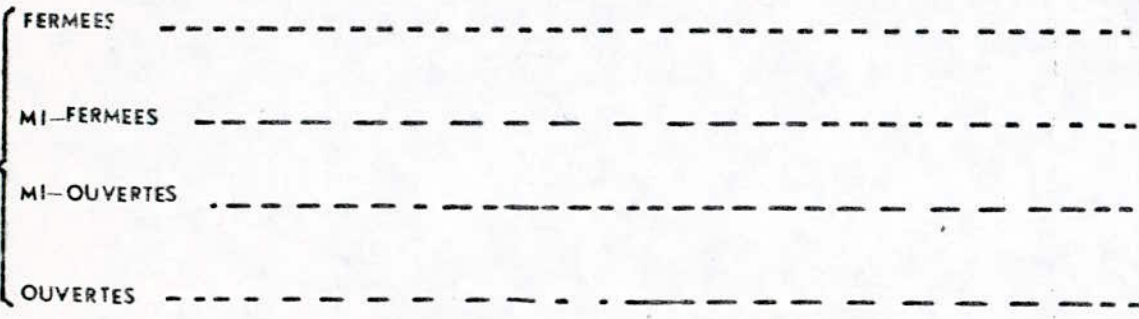


Tableau I.1 Les consonnes et Les voyelles françaises  
( d'après Malmberc 1974 )

TRAIT	ARTICULATION	ACOUSTIQUE	PERCEPTION
Vocalisme	vibration des cordes vocales	source périodique	timbre vocalique quasi musical
Antériorité	localisation de la langue	spectre plus riche en hautes fréquences	timbre plus clair
Postériorité	localisation de la langue	spectre plus riche en basses fréquences	timbre plus sombre
Aperture	écartement des mâchoires	rapprochement de $F_1$ et $F_2$	son plus intense
Consonantisme	entrave au passage de l'air	vibration aperiodique	bruit
Occlusion	fermeture plus tenue plus relâchement brusque	vibration aperiodique impulsionnelle	explosion
Constriction	passage continu de l'air dans un rétrécissement du conduit vocal	vibration aperiodique continue	friction
Labiale	utilisation des lèvres	spectre favorable aux basses fréquences	timbre plus sombre
Dentale	localisation du mouvement aux dents	spectre favorable aux hautes fréquences	timbre plus clair
Palatale	localisation du mouvement au palais	spectre favorable aux moyennes fréquences	timbre médian
Voisement	vibration des cordes vocales	présence de flux laryngé	assombrissement, sonorisation
Intensité	accroissement du courant d'air et de la force articulatoire	intensité plus franche	force, intensité
Nasalisation	abaissement du voile du palais	modification de l'amplitude relative des formants et extraformants	nasalité

TABIEAU 12: Corrélats articulatoires, acoustiques et perceptifs

( D'après LANDERCY et RENARD, 1977 )

CHAPITRE II

ANALYSE DE LA PAROLE

II.1 Nature mathématique du signal de la parole

II.2 Les "outils mathématiques" de l'analyse de la  
parole et leurs limitations

II.3 Les méthodes d'analyse



## L'ANALYSE DE LA PAROLE

Le signal de la parole est très redondant. Il véhicule des informations ne concernant pas uniquement la signification objective du message.

Ainsi, sa transmission intégrale serait difficile et très onéreuse. D'où la nécessité de la réduire à un petit nombre de composantes significatives, sans qu'il en découle une dégradation de l'intelligibilité du message parlé. La connaissance de ces composantes significatives est assujettie à une analyse préalable de ce signal.

C'est ainsi que l'on analyse de la parole a pour principal objectif la description de son signal en fonction de paramètres significatifs. Elle permet d'éliminer les redondances et par conséquent la durée du traitement et son coût.

## II.1 Nature mathématique du signal de la parole

Avant ~~XXXXXXXXXX~~ d'aborder; l'analyse proprement dite du signal de la parole, il est nécessaire de connaître non seulement sa nature physique (cf chap I) mais aussi mathématique. Pour décrire cette dernière, un rappel de la notion de signal, même concis, sera d'un bon apport.

## II.1.1 Notion de signal

## a) Définition :

Un signal est représentatif d'un phénomène physique observé. Il rend compte à l'observateur de l'état et du comportement de ce dernier dans le temps et dans l'espace.

## b) Domaines de description d'un signal

Les deux principaux domaines de description d'un signal sont celui du temps et celui de la fréquence. L'intérêt de cette double description pour un même signal vient de leur complémentarité.

## II.1.2 Le signal de la parole :

C'est un signal composé de juxtapositions ordonnées, selon le message, des signaux de types suivants :

-les signaux quasi-périodiques (fig II.1.b) pour lesquels l'amplitude, la fréquence ou les deux à la fois peuvent varier. On peut prévoir pour un signal de ce type son évolution, en partie et à brève échéance, après l'examen de son passé.

-les signaux aléatoires (fig II.1.b) qui sont à chaque instant totalement imprévisibles. Les valeurs d'un signal aléatoire dépendent du hasard. Il ne saurait être traité que par la théorie probabiliste.

-les signaux stationnaires (fig II.1.c) pour lesquels les caractéristiques statiques se conservent ou évoluent moins vite que le signal lui-même

n -les signaux impulsionnels (fig II.1.d) qui sont caractérisés par leur durée très courte, par rapport au temps d'observation. Un signal de ce type peut être aisément décrit lorsque sa forme géométrique est assez simple. On peut alors conclure que le signal global de la parole est ni stationnaire ni périodique.



## II.2 Les "outils mathématiques" de l'analyse de la parole et leurs limitations

Les "outils mathématiques" utilisés dans l'analyse de la parole sont ceux rencontrés généralement dans la théorie du traitement du signal. On peut en citer par exemple la transformée de Fourier, l'autocorrélation ou encore le filtrage numérique, etc...

Mais l'application de cette théorie du "traitement du signal", à la parole, comporte certaines limitations. Celles-ci sont dues à la non stationnarité et à la pseudo-périodicité du signal de la parole. En effet les changements au cours du temps, des modes phonatoires (signal de source) et de l'articulation (configuration du conduit vocal) font que le signal de la parole continue, n'est pas stationnaire. De plus, la phonation dans le cas des sons voisés, est pseudo-périodique.

Néanmoins, l'évolution de l'articulation et celle de la phonation sont en général relativement lentes. De ce fait on peut représenter l'évolution du conduit vocal par une succession de configurations stationnaires pendant des intervalles de temps de l'ordre de 10 à 20 ms. Il est alors possible d'appliquer les méthodes de "traitement du signal" à l'analyse de la parole. La condition à respecter est de faire le traitement pour des séquences successives de durée inférieure à 25 ms. C'est le type d'analyse dit "asynchrone".

Il existe aussi l'analyse dite "synchrone". Celle-ci ne souffre pas des limitations de l'analyse de la parole car elle s'effectue sur une seule période du signal. Cependant elle est assujettie à la détection de la période du fondamentale. Alors, son traitement est plus long que celui de l'analyse asynchrone.



## II.3 Les méthodes d'analyse :

L'analyse de la parole peut se faire dans le domaine temporel ou spectral.

## II.3.1 L'analyse temporelle :

Il existe des phénomènes évoluant rapidement lors de la production de la parole. Nous citons çà titre d'exemple la fermeture brusque précédant la g n se d'une plosive. Ce genre de ph nom nes est alors mieux caract ris  par son  volution dans le temps. D'o  la n cessit  de l'analyse temporelle.

Les principales techniques permettant d'analyser l'aspect temporel du signal de la parole sont les suivantes :

a) La fonction d'autocorr lation :

Soit  $s(t)$  une fonction quelconque du temps. Sa fonction d'autocorr lation  $g(\tau)$  sera d finie par :

$$g(\tau) = \int_{-\infty}^{+\infty} s(t) \cdot s(t-\tau) \cdot d\tau \quad (\text{II.1})$$

$\tau$  est une variable repr sentant un retard. Si l'on connait  $g(\tau)$  pour toutes les valeurs du retard " $\tau$ ", on peut en d duire la densit  spectrale de puissance du signal  $s(t)$ . En effet, on a :

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

$$\text{inversement } g(\tau) = \int_{-\infty}^{+\infty} P(f) \cdot e^{2\pi j f \tau} \cdot df \quad (\text{II.3})$$

La fonction d'autocorr lation peut  tre g n r e   partir du signal  $s(t)$  gr ce   un ensemble de lignes   retard, de modulateurs et de sommateurs (fig II.2).

## II.5

Elle peut aussi être calculée par procédés numériques, à partir d'un signal échantillonné, soit sur calculateur, soit à l'aide d'un autocorrélateur (matériel spécialisé).

Le signal de la parole est décomposé en tranches "successives" (cf chap II.2); et la fonction d'autocorrélation sera appliquée à chaque tranche. On parlera à ce moment de la fonction d'autocorrélation "instantannée".

### b) Les passages par zéro du signal

Le signal  $s(t)$  prend la valeur zéro à des instants dont la répartition dans le temps est liée à certaines caractéristiques spectrales de  $s(t)$  (fig II.3).

L'information relative à l'amplitude du signal est perdue dans ce cas. Toutefois, cette méthode est simple à mettre en œuvre. Elle est aussi très rapide car elle ne nécessite pas d'opérations arithmétiques compliquées. Cette méthode permet de mesurer la fréquence fondamentale  $F_0$  après un filtrage passe-bas.

### c) Les analyses prédictives

On part du principe que le conduit vocal filtrant le signal d'excitation  $\delta_n$ , peut être assimilé à un filtre récursif; la prédiction linéaire consiste à exprimer la valeur du  $n$ -ième échantillon, en fonction d'un nombre fini  $P$  d'échantillons antérieurs et de l'entrée  $\delta_n$  du système. Si le signal est échantillonné avec une période d'échantillonnage  $T$ , cette valeur est donnée par :

$$s(nT) = \sum_{k=1}^P a_k \cdot s[(n-k)T] + \delta_n \quad (\text{II.4})$$

\*  $\delta_n$  représente le signal d'excitation.

\*  $\sum_{k=1}^P a_k \cdot s[(n-k)T]$  : partie du signal qui peut être prédite d'après les valeurs passées.

\*  $a_k$ ,  $k=1, 2, \dots, P$  : coefficients de prédiction.

## II.6

Les deux principaux avantages de cette méthode sont :

-l'adaptation à l'étude de phénomènes évoluant rapidement.

-l'élimination d'une part très importante de la redondance de la parole.

Le nombre de coefficients nécessaires à la prédiction est de douze au total :

-dix coefficients représentant le conduit vocal.

-deux coefficients pour les effets de la glotte et des lèvres.

Il existe plusieurs méthodes pour calculer les  $p$  coefficients de prédiction. La plupart sont dites optimales au sens d'un certain critère. Ce dernier est très souvent la "minimisation" de l'erreur quadratique moyenne. Sinon, il existe d'autres critères tels que "le maximum de vraisemblance" ou la minimisation de la variance des paramètres du modèle.

L'application de la méthode dépend de la façon avec laquelle on utilise les coefficients prédictifs pour représenter les propriétés du signal. Ainsi, on peut avoir par exemple :

-l'estimation du spectre.

-L'estimation des fréquences des formants (soit à partir du spectre, soit à l'aide des pôles de la fonction de transfert du conduit vocal).

-la détection du fondamental à partir de l'erreur de prédiction qui devient grande au moment de l'excitation  $\delta_n$ .



## II.3.2 L'analyse spectrale

Pour bien comprendre le principe et les techniques de l'analyse spectrale, il est nécessaire de définir d'abord les éléments suivants :

- l'objectif de l'analyse spectrale.
- la transformée de fourier.
- la notion de filtrage.

## II.3.2.1 L'objectif de l'analyse spectrale

L'objectif de l'analyse spectrale d'un signal est de déterminer la composition fréquentielle de celui-ci, autrement-dit son spectre.

Dans le cas de la parole, on peut rechercher soit un spectre sonore, soit un sonagramme.

Le spectre sonore est la représentation du spectre d'un son qui ne dure qu'une fraction de seconde. Il nous renseigne sur les fréquences et les amplitudes des divers harmoniques constituant ce son.

Quand au sonagramme, il constitue la représentation de l'évolution spectrale d'un signal de la parole continu durant un intervalle de temps  $\Delta t$  (fig II.1.3)

Après une analyse spectrale, les paramètres pertinents que l'on peut extraire sont des grandeurs acoustiques. Et parmi ces grandeurs, les plus importantes sont :

- l'énergie du signal qui évolue dans le temps selon l'enchaînement des sons prononcés et la manière dont s'exprime le locuteur.
- les variations de l'harmonique fondamentale (de fréquence fondamentale  $F_0$  ) dans le cas des sons voisés.
- les zones de concentration d'énergie (ou maxima du spectre) de chaque son élémentaire.

## II.8

### II.3.2.2 La transformée de Fourier :

La transformée de Fourier occupe une place de première importance dans la théorie de "traitement du signal". C'est ainsi qu'elle constitue un outil fondamental dans l'analyse fréquentielle d'un signal.

#### a-Définition :

La transformée de Fourier permet de passer du domaine temporel au domaine fréquentiel.

#### b-Transformée de Fourier d'un signal périodique (série de Fourier) :

Soit un signal périodique de période  $T$  :  $s(t+T) = s(t)$ ,  $\forall t \geq 0$

Il peut être représenté (d'après le théorème de Fourier) par une somme de fonctions sinusoïdales, appelées harmoniques, dont les pulsations sont les multiples entiers de  $\omega_0 = 2\pi/T$

$$s(t) = \sum_{n=-\infty}^{+\infty} \alpha_n(\omega_0) \cdot e^{jn\omega_0 t} \quad (\text{II.5})$$

$$\text{avec : } \alpha_n(\omega_0) = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} s(t) \cdot e^{-jn\omega_0 t} \cdot dt \quad (\text{II.6})$$

Ainsi, le signal  $s(t)$  est entièrement caractérisé par sa période  $T$  (ou pulsation  $\omega_0$ ) et les coefficients  $\alpha_n$ .

$\alpha_n(\omega_0)$  est le spectre de fréquence, grandeur en général complexe, qui peut se décomposer en :

-spectre d'amplitude

$$- |\alpha_n(\omega_0)| = \sqrt{\text{Re}[\alpha_n(\omega_0)]^2 + \text{Im}[\alpha_n(\omega_0)]^2} \quad (\text{II.7})$$

-spectre de phase :

$$\varphi_n(\omega_0) = \text{Arctg} \frac{\text{Re}[\alpha_n]}{\text{Im}[\alpha_n]} \quad (\text{II.8})$$

Il est important de remarquer que le spectre d'une fonction périodique de période  $T$  est composé de raies (fig II.4) dont l'écart minimum est sur l'axe des fréquences :  $f_0 = 1/T$ .

Le spectre d'une fonction périodique est donc essentiellement discontinu, il n'existe que pour des fréquences multiples de  $f_0$ .

## II.9

c-Transformée de Fourier d'un signal  
non périodique (intégrale de Fourier)

Une fonction aperiodique peut être considérée comme la limite d'une fonction périodique dont la période tend vers l'infini. Soit alors une fonction périodique  $s_1(t)$  de période  $T$ , égale à la fonction aperiodique  $s(t)$  sur l'intervalle  $[-T/2, +T/2]$

Utilisons le développement en série de Fourier de  $s_1(t)$  :

$$s_1(t) = \sum_{n=-\infty}^{+\infty} \alpha_n(\omega_0) \cdot e^{jn\omega_0 t} \quad (\text{II.9})$$

avec  $\omega_0 = 2\pi/T$

$$\alpha_n = 1/T \cdot \int_{-T/2}^{+T/2} s_1(t) \cdot e^{-jn\omega_0 t} \cdot dt \quad (\text{II.10})$$

Lorsque  $T$  tend vers l'infini, il vient :

$$T \cdot \alpha_n = \int_{-T/2}^{+T/2} s_1(t) \cdot e^{-jn\omega_0 t} \cdot dt \rightarrow \int_{-\infty}^{+\infty} s(t) \cdot e^{-jn\omega_0 t} \cdot dt$$

D'où la transformée de Fourier de  $s(t)$  :

$$S(\omega) = \int_{-\infty}^{+\infty} s(t) \cdot e^{-jn\omega t} \cdot dt \quad (\text{II.11})$$

La transformée de Fourier inverse de  $F(\omega)$  est :

$$s(t) = 1/2\pi \cdot \int_{-\infty}^{+\infty} S(\omega) \cdot e^{j\omega t} \cdot d\omega \quad (\text{II.12})$$

On note :  $s(t) \xleftrightarrow{F} S(\omega)$

$S(\omega)$  est généralement une grandeur complexe qui peut s'écrire :

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{+\infty} s(t) \cdot e^{-j\omega t} \cdot dt \\ &= R(\omega) + jX(\omega) = A(\omega) \cdot e^{j\phi(\omega)} \end{aligned} \quad (\text{II.13})$$



On appelle :

.  $A(\omega)$  le spectre de Fourier de  $s(t)$ .

.  $\Phi(\omega)$  la phase du spectre de Fourier de  $s(t)$ .

Contrairement à une fonction périodique, le spectre d'une fonction apériodique est continu (fig II.5). Il existe pour toutes les valeurs de la fréquence.

Les conditions d'existence de  $S(\omega)$  sont :

- La fonction  $s(t)$  est bornée

-  $\int_{-\infty}^{+\infty} |s(t)| dt$  a une valeur finie.

- Les discontinuités de  $s(t)$ ; ainsi que ses maxima et minima sont en nombre fini.

Les fonctions rencontrées dans la pratique décrivent des phénomènes limités dans le temps et dans l'espace. Ainsi elles vérifient les conditions énoncées, le problème d'existence de la transformée de Fourier ne se pose pas.

d-Notion d'harmoniques et de partiels:

La transformée de Fourier montre que: "toute vibration complexe  $s(t)$  est une composition de vibrations sinusoïdales". Si  $s(t)$  est périodique, ces vibrations sinusoïdales sont appelées "les harmoniques" (leurs fréquences sont des multiples entiers de la fréquence fondamentale  $F_0$ ). Par contre, si  $s(t)$  est apériodique, il n'y aura aucun rapport semblable entre les fréquences de ses composantes. Ces dernières sont alors appelées : "les partiels".

### II.3.2.3 Le filtrage

a-Introduction

Il est souvent difficile et très onéreux de transmettre intégralement une information donnée d'un émetteur à un receveur. Aussi, faut-il toujours essayer d'extraire uniquement les paramètres significatifs du message pour en éliminer la redondance et "rentabiliser" l'opération de transmission ou parfois de traitement.

De plus, tout signal physique est formé généralement d'une partie utile (information pertinente) et d'une partie indésirable (le bruit). Il est souvent indispensable d'éliminer cette deuxième partie ne serait-ce que partiellement.

Pour réaliser l'un de ces objectifs, ou même les deux, l'opération utilisée généralement est le filtrage.

#### b-Définition d'un filtre :

Etant donné un phénomène dont le spectre s'étend sur un intervalle de fréquence quelconque, un filtre est un dispositif susceptible de renforcer certaines composantes fréquentielles du signal tout en affaiblissant les autres.

#### c-Les filtres analogiques :

Il existe essentiellement trois catégories de ces filtres :

- passe-bas.
- passe-haut.
- passe-bande.

Dans chaque catégorie, le filtre est caractérisé par sa courbe de réponse (fig II.6) ou encore sa fonction de transfert.

#### d-Les filtres numériques :

Ils jouent un rôle très important dans le traitement des signaux numériques en général, et de la parole en particulier.

Un filtre numérique n'a pas d'existence physique comme un filtre analogique, il est représenté par un algorithme (méthode de calcul).

Dans le cas fréquent d'un filtre linéaire, celui-ci permet de réaliser les opérations de la relation suivante :

$$y_n = \sum_{k=1}^M a_k \cdot y_{n-k} + \sum_{k=0}^M b_k \cdot x_{n-k} \quad (\text{II.14})$$



avec :  $x_n$  séquence d'entrée du filtre.  
 $y_n$  séquence de sortie du filtre.

Quand aux coefficients  $a_k$  ( $k=1, \dots, M$ ) et  $b_k$  ( $k=1, \dots, M$ ), ils définissent le filtre numérique.

Les conditions de calcul de ces coefficients sont liés aux caractéristiques du filtre analogique dont on veut faire la synthèse numérique.

Par rapport au filtre analogique qu'il remplace dans bien des cas, le filtre numérique possède les avantages suivants:

- la performance, la précision et la très grande variété.
- la souplesse de sa mise en forme et d'adaptation.
- l'absence de problème de réalisation technologique.

Son application est très diverse, on peut le trouver dans de nombreux systèmes physiques : l'audition, la parole, le radar, les télécommunications, l'océanographie, la médecine, etc...

#### II.3.2.4 Principe de l'analyse spectrale

L'objectif de ce type d'analyse est la connaissance de la composition spectrale du signal de la parole. Son principe se ramène à la représentation de ce signal par sa transformée de Fourier.

Or, nous savons que les méthodes du "traitement du signal" ne pourraient s'appliquer au signal de la parole que sur des intervalles de temps n'excédant pas 25 ms (cf chap II.2). La transformée de Fourier faisant partie de cette famille de méthodes, ne sera applicable qu'à des tranches successives du signal. Ceci est appelé "l'analyse spectrale à court terme". On obtient alors une suite de spectres dits "instantanés" (fig II.7). L'ensemble de ces derniers constitue le spectre "évolutif" car il montre l'évolution dans le temps, du spectre du signal.

L'hypothèse sous-jacente de l'analyse spectrale est que sur un ~~XXXXXXX~~ intervalle de temps suffisamment court, le signal



de la parole est stationnaire, car dans ce cas, il est bien représenté par sa transformée de Fourier.

### II.3.2.5 Réalisation de l'analyse spectrale à court terme :

Par définition, la transformée de Fourier d'un signal  $s(t)$  est donnée par la relation (II.11). Alors, le spectre à court terme est donné par :

$$S(\omega) = \int_{-\infty}^t s(\lambda) \cdot h_{\omega}(t-\lambda) \cdot e^{-j\omega\lambda} \cdot d\lambda \quad (\text{II.15})$$

La fenêtre  $h_{\omega}(t)$  détermine la durée de la "tranche" analysée du signal. Son choix est souvent difficile à cause du principe d'incertitude exprimé par :

$$\Delta t \cdot \Delta f = c \quad (\text{II.16})$$

$c$  : étant une constante généralement voisine de l'unité. Cette relation montre l'impossibilité de connaître simultanément, avec précision, le temps et la fréquence. En effet, plus la largeur de la fenêtre temporelle est étroite, plus l'évolution du spectre, dans le temps, est bien définie. Mais dans ce cas, la représentation fréquentielle du spectre manque de précision. Inversement, plus la largeur de la fenêtre temporelle est grande, plus on perd de l'information sur l'évolution temporelle du spectre. Mais ce qui est perdu dans le "temps" est gagné dans la "fréquence".

De plus, la forme de la fenêtre temporelle entraîne une distorsion du spectre instantané, distorsion d'autant plus importante que la largeur de la fenêtre est plus faible.

Les analyseurs physiques permettant la réalisation de l'analyse spectrale du signal de la parole sont soit des systèmes à filtres, soit un équipement spécialisé : "le transformateur de Fourier".

## II.3.2.6 L'analyse spectrale par calculateurs

L'analyse spectrale peut être réalisée sur calculateur. Celui-ci est alors utilisé en filtre numérique ou en "transformateur de Fourier".

Mais pour effectuer une opération arithmétique ou logique, le calculateur numérique a besoin d'un certain temps non négligeable. Par conséquent, pour traiter deux informations dans une machine calculatrice, il faut les introduire séparément; on ne présentera la deuxième information qu'à la fin du traitement de la première.

Ainsi, l'introduction des données numériques dans un calculateur arithmétique est fréquemment précédé de deux opérations : l'échantillonnage et la quantification. D'où la nécessité de rappeler ces deux notions :

a-L'échantillonnage :

C'est le passage d'une fonction  $f(t)$  continue, définie sur un intervalle  $[t_0, t_0+T]$ , à une suite de nombres :  $f(t_1), f(t_2), \dots, f(t_n)$  représentée par  $f^*(t)$  et telle que :  $t_0 < t_1 < t_2 < \dots < t_n < t_0+T$ . (fig II.8.b).

$f^*(t)$  est alors dite fonction échantillonnée. Elle est obtenue à partir de la fonction  $f(t)$  échantillonnée à la fréquence  $f_e = 1/T_e$  par le train d'impulsions de Dirac  $\delta(t-nT)$ . Le  $n^{\text{ème}}$  échantillon est :

$$f(nT) = f^*(t) = f(t) \cdot \sum_{n=-\infty}^{+\infty} \delta(t-nT) \quad (\text{II.17})$$

$$\text{avec : } \delta(t-nT) = \begin{cases} 1, & \text{si } t=nT \\ 0, & \text{si } t \neq nT \end{cases} \quad (\text{II.18})$$

La période d'échantillonnage  $T_e$  n'a rien à voir a priori, avec la période  $T_s$  de la fonction  $f(t)$  (si elle est périodique).



## Efficacité de l'échantillonnage :

Pour que l'information discrète véhiculée par  $f^*(t)$  soit équivalente à l'information continue de  $f(t)$ , il faut que  $f_e$  soit supérieure au double de la fréquence la plus haute contenue dans le spectre de  $f(t)$  (théorème de Shannon).

Quand cette condition est vérifiée, l'efficacité de l'échantillonnage croît avec  $f_e$ . Mais  $f_e$  fixe aussi la rapidité de réponse de l'opération et même du traitement de l'information. Aussi, il faut adopter un compromis selon qu'on désire une bonne précision ou un gain de temps.

## b-La quantification :

Elle consiste à traduire l'information échantillonnée  $[f^*(t)]$  en nombres qui sont introduits dans le calculateur.

La quantification est définie comme étant l'approximation de chacune des valeurs  $f(t_i)$  de  $f^*(t)$  par un multiple d'un certain quantum  $q$  appelé "pas de quantification".

La précision de cette opération croît lorsque l'on choisit  $q$  de plus en plus petit. Mais on est limité par le plus petit nombre pouvant être affiché dans le calculateur.

## c-Exemple de signal échantillonné et quantifié

Soit un signal continu  $s(t)$  (fig II.8).

On l'échantillonne en prenant ses valeurs aux instants  $T, 2T, \dots, nT$ . Pour effectuer la quantification, nous choisissons " $q$ " de longueur fixe sur l'axe des ordonnées. Ensuite, on remplacera chacune des valeurs  $s(t), s(2T), \dots, s(nT)$  par le multiple de  $q$  immédiatement inférieur.

## d-L'analyse de Fourier par ordinateur

La transformée de Fourier d'un signal numérisé est appelée la "transformée de Fourier discrète" (T.F.D). On l'obtient sur calculateur grâce à un algorithme



me appelé "transformée de Fourier rapide" (T.F.R) ou Fast Fourier Transform (F.F.T). Cet algorithme permet de réduire considérablement le nombre d'opérations arithmétiques par rapport à un calcul direct.

### II.3.2.7 Méthode du sepstre :

Le principe de cette méthode est de réaliser une séparation entre le spectre du signal glottal et la réponse en fréquence du conduit vocal.

L'hypothèse de base est que : le spectre de la parole est le produit du spectre de la source par celui des caractéristiques de résonance du conduit vocal. Dans ces conditions, si  $s(t)$  est le signal d'un son voisé, on a :

$$|S(\omega)|^2 = |G(\omega)|^2 \times |C(\omega)|^2 \quad (\text{II.19})$$

avec :  $-S(\ )$  : spectre d'amplitude de  $s(t)$ .

$-G(\ )$  : spectre d'amplitude du signal glottal.

$-C(\ )$  : réponse en fréquence du conduit vocal.

Les produits étant plus difficiles à manipuler que les additions, on transforme l'expression du spectre en somme de deux termes. Pour cela, on prend son logarithme :

$$\text{Log}(|S(\omega)|^2) = \text{Log}(|G(\omega)|^2) + \text{Log}(|C(\omega)|^2) \quad (\text{II.20})$$

Cette méthode permet la détermination du fondamental et l'extraction des formants.

*Analyse instrumentale*

### II.3.2.8 Méthodes d'analyse par synthèse :

L'efficacité de ces méthodes dépend de la connaissance du processus de production de la parole. Elles sont utilisées par exemple en vue de l'extraction du spectre du conduit vocal après élimination des contributions de la source et du rayonnement.

Ainsi, ces méthodes sont utilisées le plus souvent pour l'extraction des formants ou pour déterminer des paramètres articulatoires. Ceci est réalisé en recherchant l'identité entre formants d'un modèle articulatoire et ceux d'un voix

-le vocoder à canaux (fig II.9) a un analyseur constitué d'un banc de 12 à 15 filtres couvrant la bande spectrale de la voix 300 à 3300 Hz.

b-les analyseurs à filtre unique :

Ils sont tous basés sur le même principe. Celui-ci consiste à explorer toute l'étendue du spectre utile du signal de la parole, à l'aide d'un filtre unique. A cet effet, on utilise un filtre passe-bande dont on peut commander la fréquence centrale variable.

L'explication de ce principe sera faite dans l'étude détaillée, qui suivra, d'un exemple de ce genre d'analyseurs : le sonographe. Celui-ci nous intéresse particulièrement car c'est l'appareil utilisé pour l'opération d'analyse, préalable à toute synthèse de la parole par l'E.V.A.III (electrical vocal analog).

### II.3.3 Le sonographe

*exple d'analyse instruments Ce.*

#### II.3.3.1 Description générale :

Le sonographe est composé de trois parties importantes (fig II.10) :

-un système remplissant les fonctions d'enregistrement, de lecture et d'effacement. Il est constitué d'un plateau magnétique cylindrique, d'une tête "d'enregistrement-effacement" et d'une tête de lecture. Le plateau portera sur sa périphérie un enregistrement de durée maximale 2,4 s. Il peut tourner, indéfiniment et à vitesse constante, autour de son axe permettant ainsi à la tête de lecture d'explorer à chaque tour tout l'enregistrement.

-un système de filtrage constitué par le filtre unique, un oscillateur local à fréquence variable et un modulateur (fig II.11).



réelle.

### II.3.2.9 Les analyseurs à filtre

Un des systèmes physiques utilisés pour la réalisation de l'analyse spectrale est le filtre analogique. Dans ce cas, la largeur de la fenêtre temporelle  $h_{\omega}(t)$  (cf chap II.4.4) est fixé par la constante de temps  $\Delta t$  du filtre. Ainsi, la réponse de ce dernier renseigne sur les propriétés moyennes du signal de parole appliqué à l'entrée, dans cet intervalle de temps  $\Delta t$ . La densité spectrale instantanée mesurée à la sortie du filtre sera la quantité  $s(\omega, t)$  donnée par la relation (II.15).

De plus, si " $\Delta f$ " est la bande passante du filtre, le produit  $\Delta t \times \Delta f$  étant voisin de l'unité (cf II.16), un compromis devra être adopté entre selectivité en fréquence et selectivité temporelle. La nature de ce compromis dépendra en fait du problème particulier à résoudre (recherche de détails dans l'évolution temporelle ou fréquentielle du signal (cf chap II.4.4)).

Les analyseurs à filtre sont divisés en deux groupes : les analyseurs à plusieurs filtres et les analyseurs à filtre unique.

#### a- Les analyseurs à plusieurs filtres

Le principal avantage de ce type d'analyseurs est qu'ils permettent une analyse fréquentielle pratiquement instantanée. En effet, on peut appliquer le signal  $s(t)$ , simultanément, à un grand nombre de filtres passe-bande contigus, fixes en fréquence, et montés en parallèle. Chacun de ces filtres analysera la bande de fréquence correspondant à sa propre bande passante. Et, l'ensemble des filtres couvrira toute l'étendue spectrale désirée.

Exemples : - l'analyseur du vocoder (1) à bande de base est constitué par un banc de filtres couvrant la bande de 800 à 6000 Hz environ.



Ces deux derniers permettent de faire varier la fréquence centrale de la bande de filtrage.

-un système de marquage constitué d'un tambour cylindrique sur lequel est enroulé une feuille de papier spécial, et un assemblage de pièces mécaniques portant un stylet.

Cet assemblage comporte essentiellement un bâti supportant un vis sans fin le long de laquelle le stylet se déplacera en cours d'analyse. Cette vis est reliée au tambour par une courroie de transmission ou un système d'engrenage. De plus, le tambour est placé sur le même axe que le plateau magnétique. Ainsi, lorsque ce dernier effectue un tour, le tambour et par la suite la vis font de même.

#### II.3.3.2 Fonctionnement

Une séquence du signal  $s(t)$  de la parole est enregistrée sur le plateau magnétique. Pour analyser cette séquence, la fréquence centrale du filtre part des basses vers les hautes fréquences en augmentant par pas d'environ 20 Hz. L'étendue spectrale à explorer occupe l'intervalle de fréquence allant de 85 à 8000 Hz. Pour chaque valeur de la fréquence centrale de la bande d'analyse le plateau magnétique effectue un tour. Il permet alors la relecture de la séquence de parole et son analyse autour de la fréquence considérée. Lors de cette analyse, le cheminement du signal dans l'appareil, illustré par la figure II.12, peut être décrit comme suit : recueilli par la tête de lecture et renforcé par un amplificateur, le signal attaque l'entrée du système de filtrage. A la sortie, seule la oules composantes correspondant à la bande passante sont présentes. Leur intensité est transformée proportionnellement en un courant envoyé dans le stylet. Ce courant

passé entre la pointe du stylet et le tambour laissant sur le papier un tracé noir. Ce noircissement croît avec l'intensité du courant, donc avec l'amplitude spectrale du signal à l'instant et à la fréquence considérés. En fin d'analyse, on obtient sur le papier une représentation à trois dimensions de la séquence de la parole : le sonagramme.

L'axe des abscisses, horizontal, est celui du temps. L'axe des ordonnées, vertical, sera celui des fréquences. L'intensité (densité spectrale de l'énergie à une fréquence donnée et à un instant donné) est représentée par le noircissement. La figure II.13 nous montre le sonagramme de la séquence : "Un chasseur sachant chasser".

### II.3.3.3 Principe de commande de la fréquence centrale du filtre :

Le feuille de papier comporte une échelle de fréquences. Alors, le tracé du sonagramme doit être réalisé en respectant cette échelle. A cet effet, la valeur de la fréquence centrale du filtre est déterminée par la position du stylet sur la feuille. Ceci est obtenu à l'aide du dispositif décrit par la figure II.11. Quand au principe utilisé, c'est celui de l'hétérodynage.

Le signal à analyser  $s(t)$  est modulé par un autre signal  $r(t)$  délivré par un ~~un~~ oscillateur local dont on peut faire varier la fréquence " $F_L$ ". Soit une composante de rang  $n$  du signal  $s(t)$ , ayant une fréquence " $F_n$ " et une amplitude " $A_n$ ". Le signal obtenu après modulation est :

$$g(t) = A_n \cos(2\pi F_n t + \varphi_n) \times R \cos(2\pi F_L t + \varphi_L)$$

$$g(t) = \frac{A_n R}{2} \cdot \cos[2\pi (F_n - F_L) t + \varphi_n - \varphi_L]$$

$$+ \frac{A_n R}{2} \cdot \cos[2\pi (F_n + F_L) t + \varphi_n + \varphi_L] \quad (\text{II.21})$$



Si  $g(t)$  traverse un filtre passe-bas dont la bande passante  $B$  est telle que :  $(F_n - F_L) < B$ , alors seule la composante en  $(F_n - F_L)$  sera retrouvée en sortie. Car l'autre composante de  $g(t)$ , en  $(F_n + F_L)$ , est de fréquence relativement haute (supérieure à  $B$ ), donc éliminée.

On a ainsi isolé dans le spectre de  $s(t)$  une bande de fréquence de largeur voisine de  $B$ , autour de la fréquence  $F_L$ . Quand le stylet se déplace,  $F_L$  et par la suite  $(F_n - F_L)$  varient. Donc, le système formé par l'oscillateur local, le modulateur et le filtre passe-bas est équivalent à un filtre passe-bande de fréquence variable.

#### II.3.3.4 Dispositifs annexes du sonographe.

Parmi ces dispositifs, les plus importants sont :

- les filtres d'analyse : deux filtres passe-bande différents par leur largeur de bande sont disponibles au sein du sonographe. L'usage a consacré les largeurs de 45 et 300 Hz qui, comme nous le verrons, permettent d'extraire du signal de parole des informations notablement différentes.

- la dilatation de l'échelle de fréquence : on peut ajuster manuellement les limites inférieures et supérieures de l'analyse, de façon à exploiter au mieux la hauteur utile du spectrogramme.

- l'étalonnage en fréquence : un signal périodique de référence, de fréquence précise (500 Hz) et très riche en harmoniques est enregistré pendant un bref instant sur la piste magnétique, et ainsi juxtaposé au signal à étudier. Les harmoniques fournissent des repères de fréquence commodes et précis.



## II.22

-les sections : au vu du sonagramme, on peut juger que le spectre instantané à tel ou tel instant mérite d'être examiné de manière précise. L'appareil permet d'obtenir ces sections de l'objet sonore dont les coordonnées de fréquence en abscisses et d'amplitude en ordonnées

### II.3.5 Le sonagramme en bande passante étroite

C'est un sonagramme obtenu en choisissant le filtre d'analyse de 45 Hz de bande passante. Il montre beaucoup de détails relatifs au domaine fréquentiel (fig II.13). Parmi ces détails, nous avons :

-l'apparition nette de la structure harmonique des voyelles.

-l'absence d'harmoniques dans les zones de bruit des consonnes. Pour cette catégorie de sons, on verra un mélange de vibrations de fréquences irrégulières et continues dans un certain domaine : ce sont les "partiels".

-des discontinuités verticales apparaissant sur toute l'échelle des fréquences. Elles permettent de séparer les phonèmes (plus petites unités distinctives du langage).

-les renforcements de certaines harmoniques dans certaines zones de fréquences particulières à chaque son sonore : les formants (cf chap III).

-l'évolution du fondamental visible grâce à la nette représentation des harmoniques.

-et d'autres détails nécessaires, essentiellement, aux phonéticiens tels que les syllabes et les logatomes.

Cependant, malgré toutes ces performances, le sonagramme en bande passante étroite présente deux défauts importants :

## II.23

-la nette représentation des harmoniques gêne la lecture de l'aspect réel des formants qui sont si importants pour une synthèse formantique (cf chap IV).

-le filtre d'analyse étant très sélectif (bande passante de 45 Hz), présente un temps de montée et un temps de descente assez importants. Ainsi, sa constante de temps  $\Delta t$  est relativement grande ( $1/45^e$  s) par rapport à certaines transitoires, durant moins de  $1/100^e$  s (cas des transitoires de plosives). Il s'ensuit que la représentation horizontale de ces transitions est floue.

On peut donc conclure que le sonagramme en bande passante étroite permet une haute définition des détails sonores sur l'échelle des fréquences (verticale). Par contre, sur l'échelle du temps (horizontale), il présente des insuffisances.

Pour pallier à ces défauts, on utilise la sonographie en large bande passante (300 Hz).

### II.3.3.6 Le sonagramme en large bande passante

L'analyse sonographique en large bande passante est effectuée sur le même sonographe et suivant le même principe que précédemment. La seule différence étant que le filtre utilisé ne présente plus une bande passante de 45 Hz mais de 300 Hz. Ce qui confère à ce filtre une constante de temps très brève :  $1/300^e$  s.

C'est largement suffisant pour lui permettre de tracer les transitoires les plus brutaux avec un maximum de netteté dans le sens horizontal (temps). Mais dans ce cas, on ne distingue plus les harmoniques qui, à chaque nouvelle ligne d'analyse, s'engouffrent à plusieurs dans le filtre. Ceci permet de lire aisément l'aspect des formants.

De plus, on distingue au sein de ces derniers des stries verticales qui nous renseignent ou non sur la nature périodique ou non d'un son élémentaire (fig II.14).

## II.24

L'observation de ces stries permet de calculer la hauteur du fondamental : si on compte sur une longueur de feuille d'environ 13 mm (représentant  $1/10^e$ s) ces stries, on connaît par la même le nombre d'oscillations fondamentales survenues pendant ce laps de temps, il suffit alors de multiplier ce nombre par 10 pour avoir la valeur de la fréquence fondamentale.

La figure II.15 donne une idée des différences caractérisant les deux analyses sonographiques qu'on vient de voir .



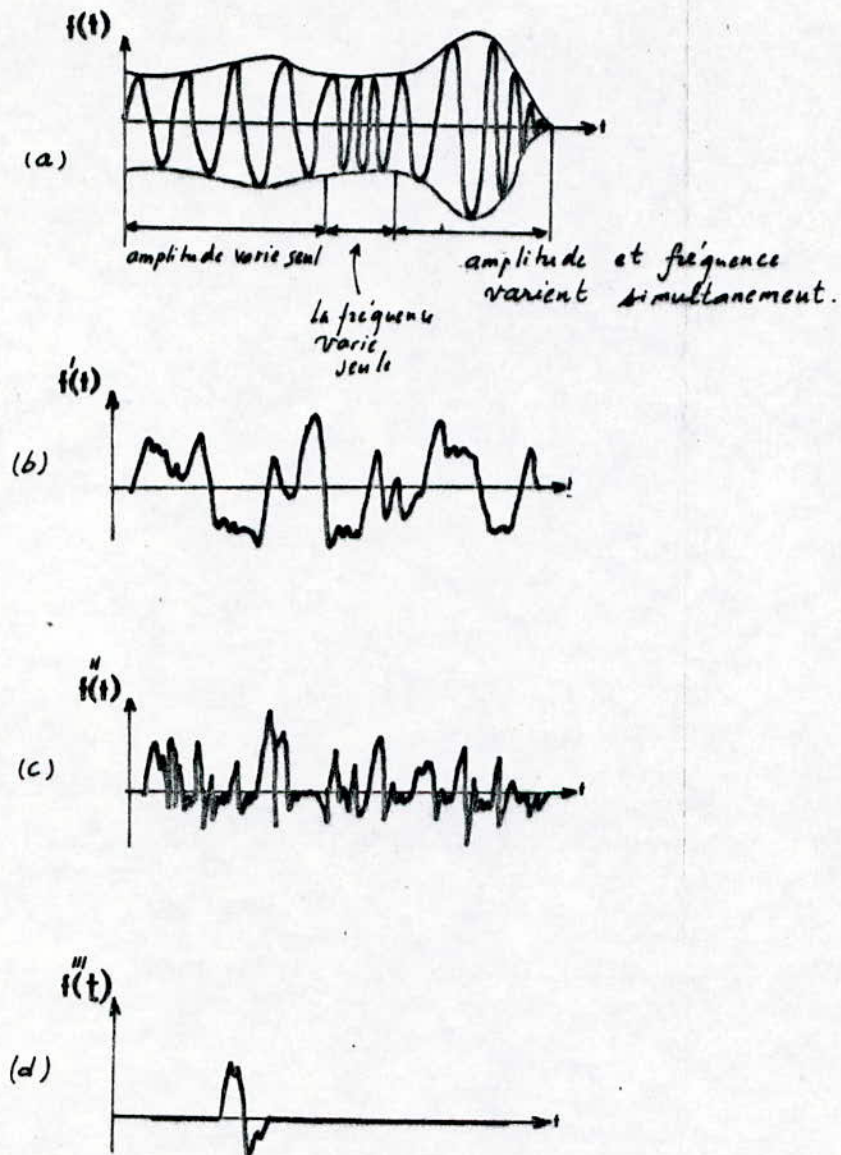


fig II. 1. Les signaux rentrant dans la composition du signal de parole.

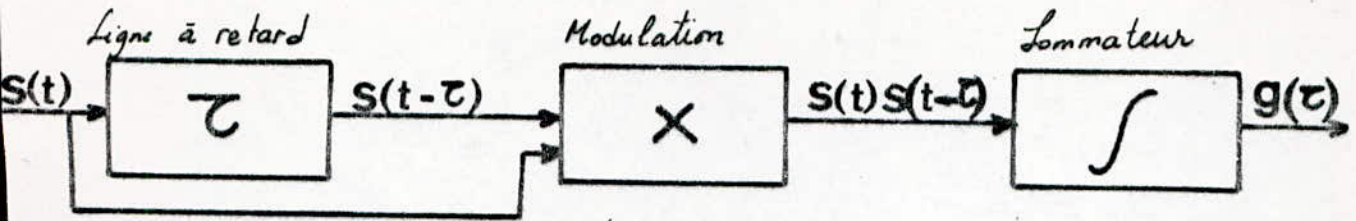


fig II. 2 Generation de la fonction d'autocorrelation  
pour une valeur de  $\tau$  (retard)

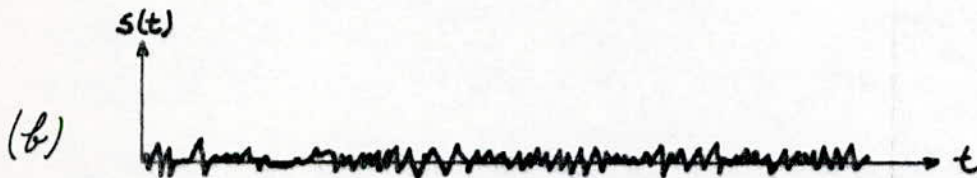


fig II. 3: Les passages par zéro du signal de la parole pour  
(a) : un son voisé  
(b) : un son non voisé.

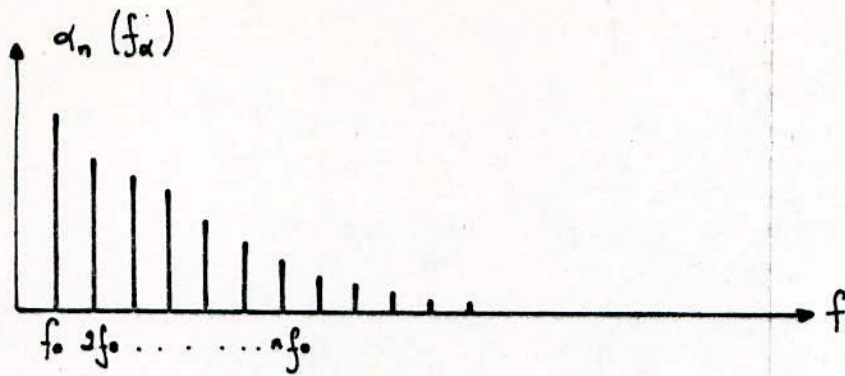


fig II. 4 : Spectre d'amplitude d'un signal périodique  
(triangulaire)

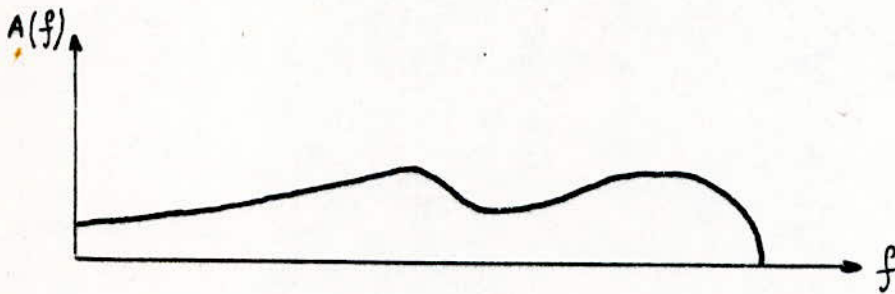
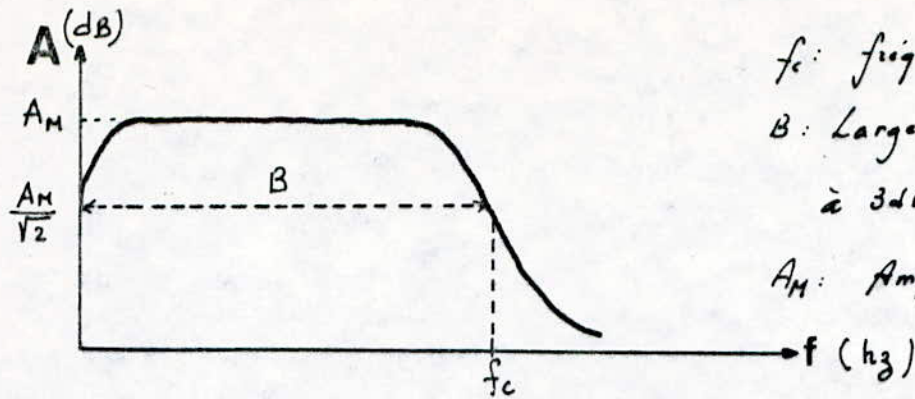


fig II. 5 : Spectre d'amplitude d'un signal aperiodique





$f_c$ : fréquence de coupure  
 $B$ : Largeur de la bande passante à 3dB  
 $A_M$ : Amplitude maximale

fig II. 6.a Courbe de réponse du filtre passe-bas

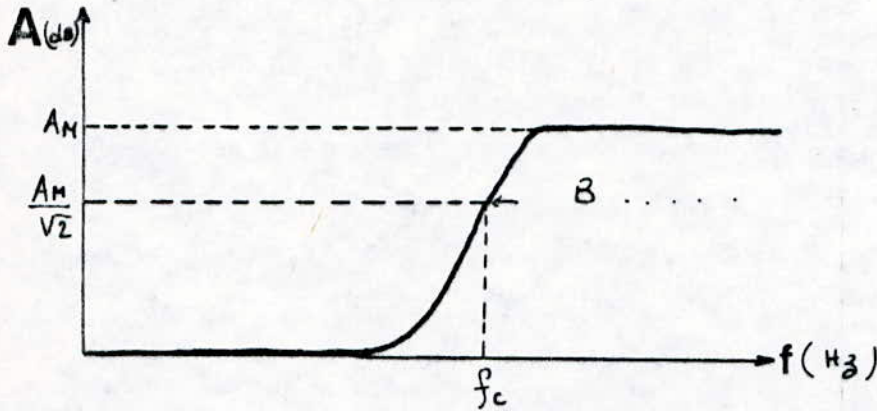
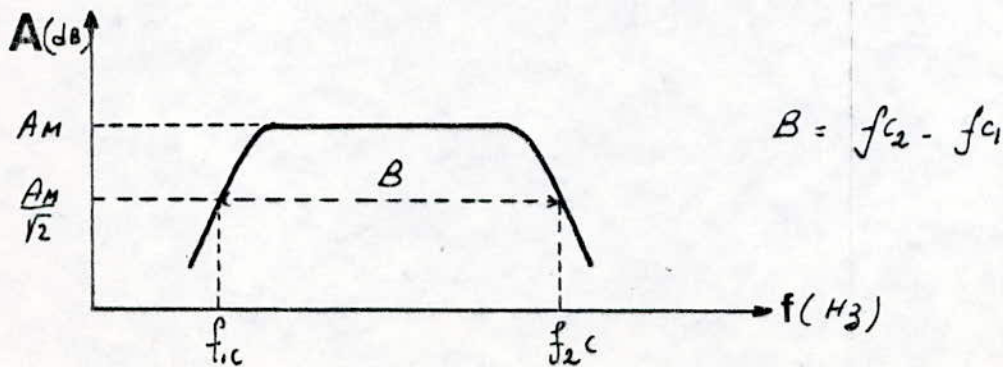


fig II 6.b Courbe de réponse du filtre passe-haut.



$$B = f_{2c} - f_{1c}$$

fig II. 6.c Courbe de réponse du filtre passe-bande.

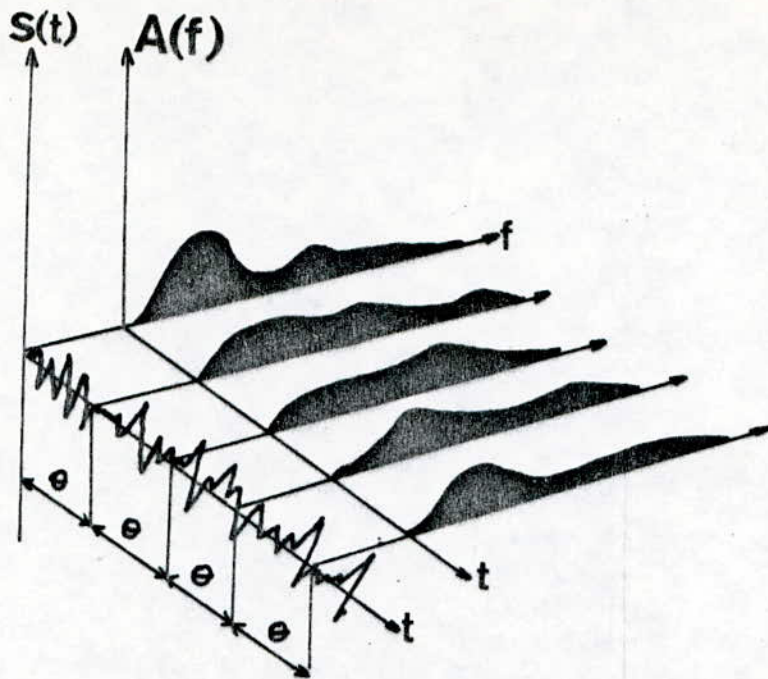
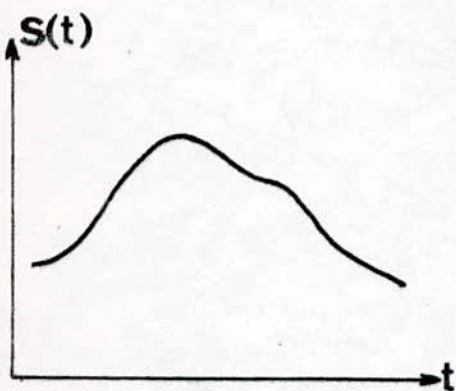
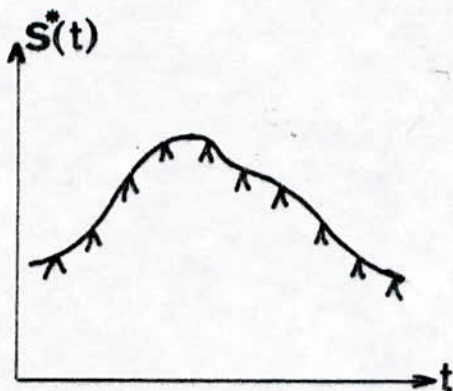


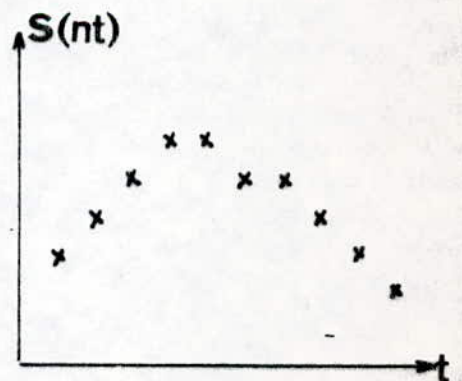
fig II. 7 : Spectre instantané d'un signal  $S(t)$ .



a. Signal continu



b. Signal échantillonné



c. Signal quantifié.

fig II. 8 Échantillonnage et quantification d'un signal  $S(t)$ .

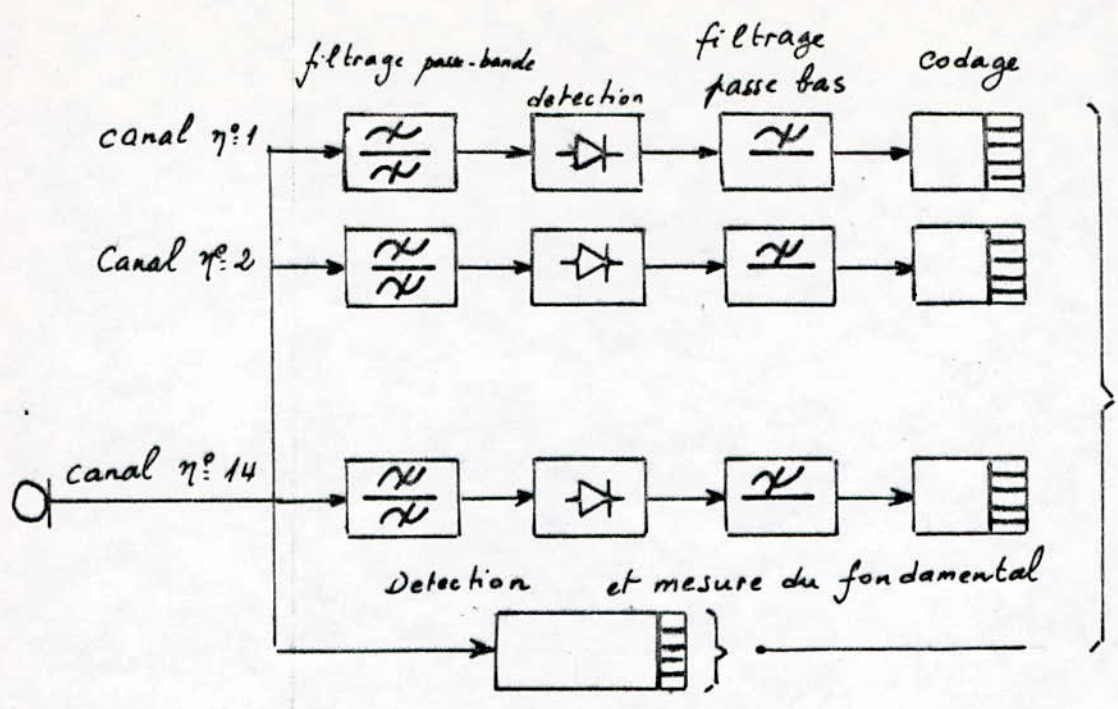


fig II. 9 Analyseur du Vocoder

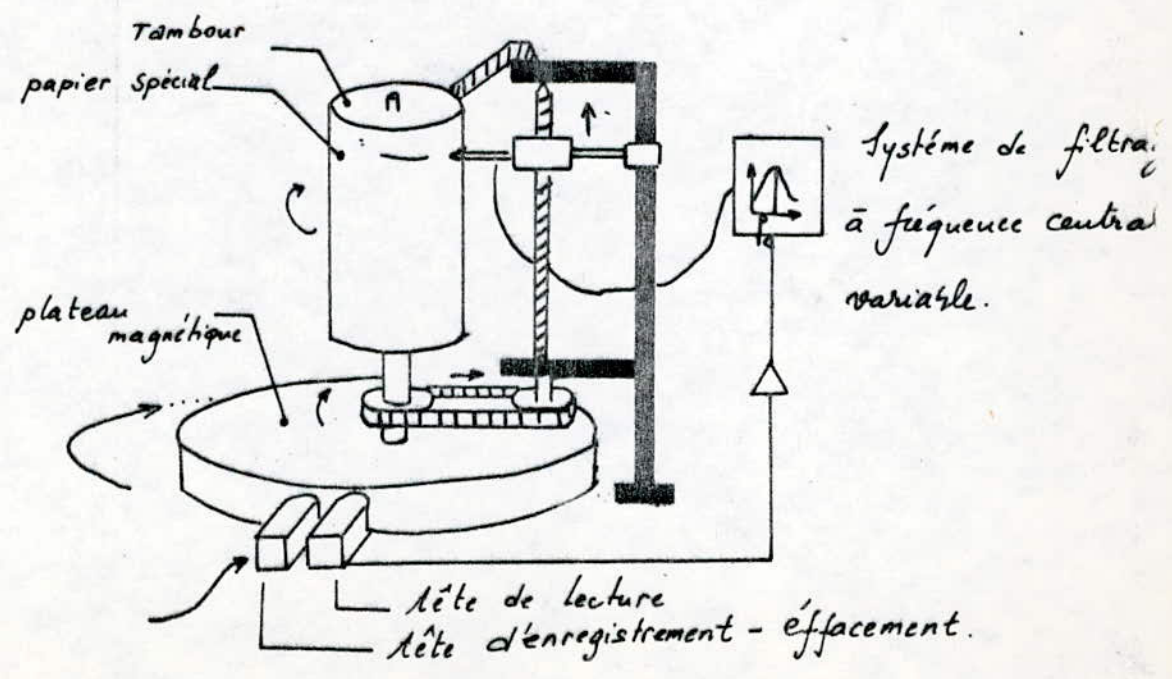


fig II. 10 Schéma de principe du sonaqrphe



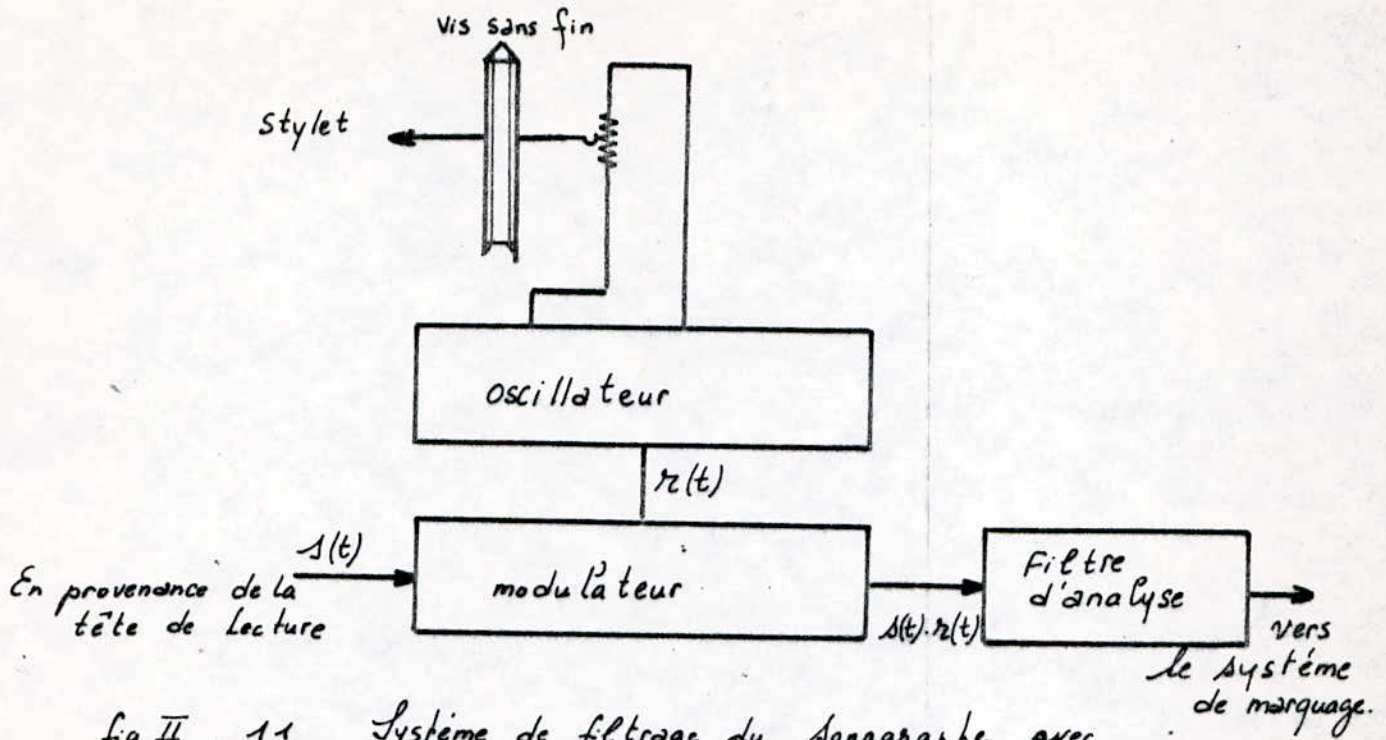


fig II . 11 Système de filtrage du sonographe avec la commande de sa fréquence centrale.

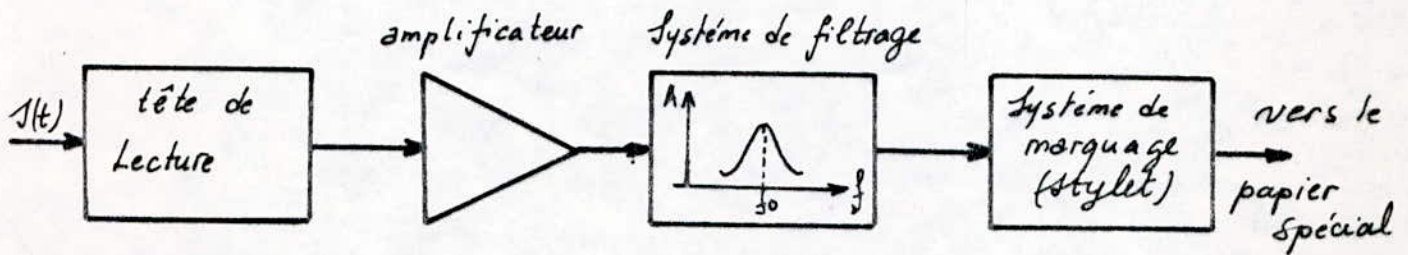


fig II . 12 Cheminement du signal " dans le sonographe "

TYPE B/65 SONAGRAM © KAY ELEMETRICS CO. PINE BROOK, N. J.

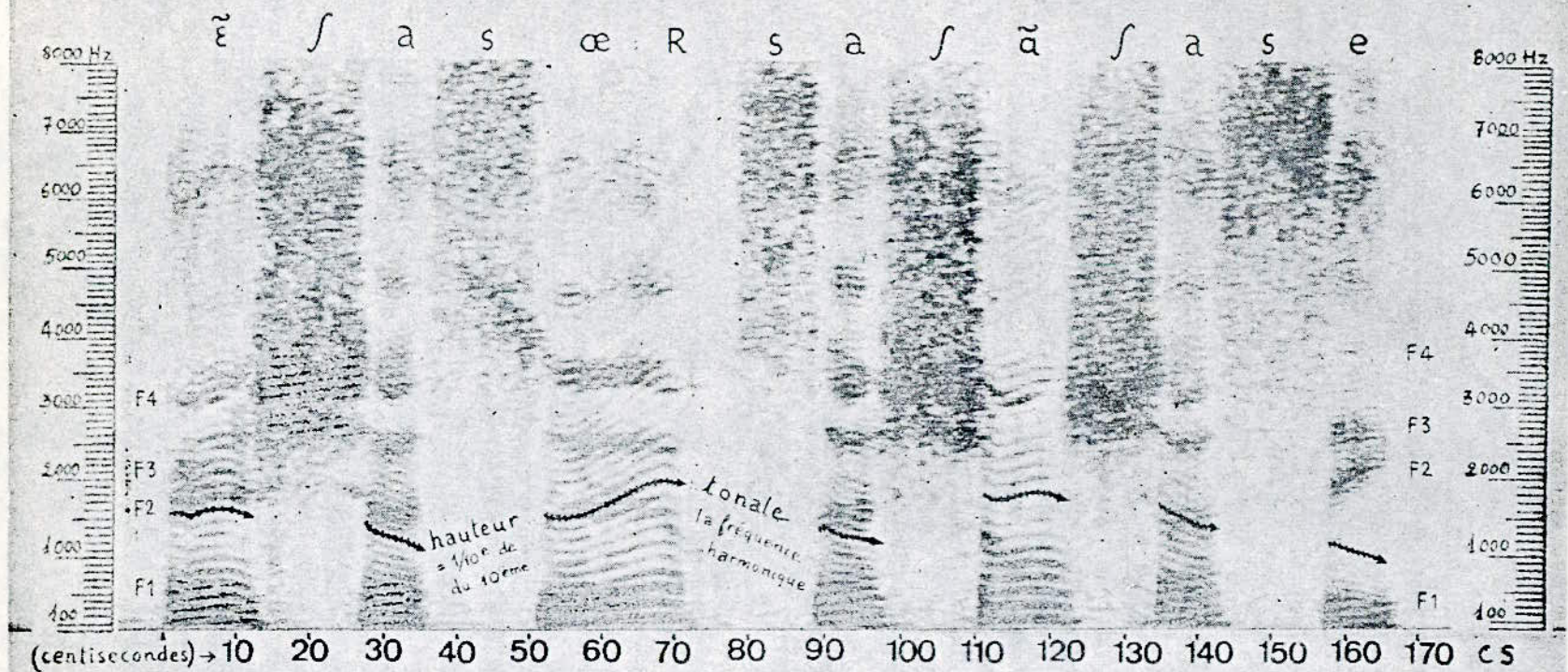


FIG. 87

fig II. 13 : Sonagramme en bande passante étroite de la phrase :  
"un chasseur sachant chasser".

II. 32



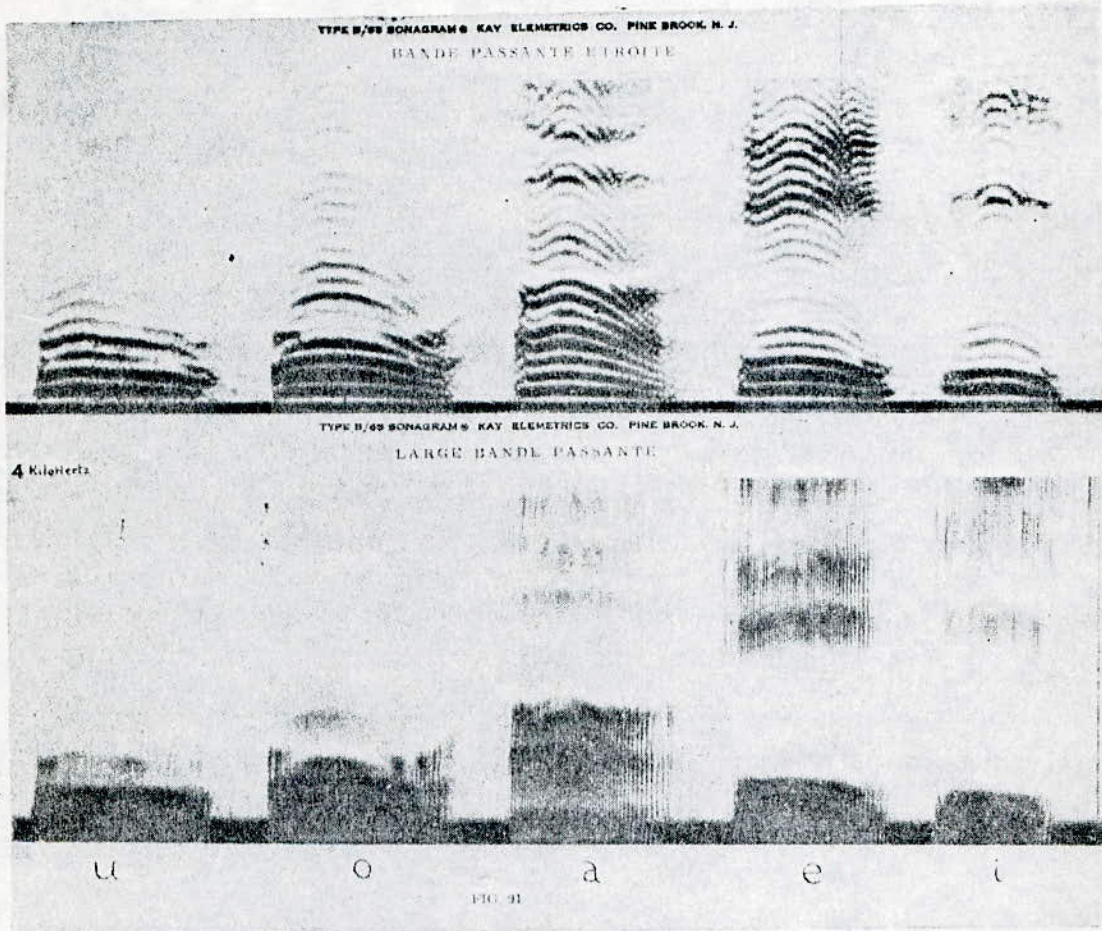


fig. II. 15. Représentation de deux sonagrammes  
 a. bande passante étroite ; b. large bande , en  
 vue d'une comparaison .



CHAPITRE III

LES FORMANTS ET LES RESONATEURS

III.1 Définition et caractéristiques des formants.

III.2 Les formants et le timbre d'un son

III.3 Les voyelles et les formants

III.4 La formation des formants et la résonance

## LES FORMANTS ET LES RESONATEURS

Chacun d'entre nous a une voix présentant un cachet bien particulier, propre à lui. C'est ainsi qu'on peut souvent identifier quelqu'un rien qu'à l'entendre parler.

Ce cachet, si distinctif, est lié à la notion de timbre qui est elle-même liée à la notion de formant. Et, cette dernière s'explique par le phénomène de résonance. L'objet de ce chapitre est d'étudier ces trois notions pour montrer ensuite l'importance des formants en tant que paramètres de la parole.

### III.1 Définition et caractéristiques des formants

#### III.1.1 Définition des formants :

Les formants d'un son élémentaire sont définis comme étant les régions du spectre où l'énergie est fortement concentrée. Ils sont alors représentés, pour une voyelle, par des bouquets d'harmoniques renforcés, proéminents par rapport au reste du spectre (fig III.1). Pour un son non voisé (fricative sourde par exemple), on constate la présence de bandes de fréquences renforcées appelées "formants de bruits".

#### III.1.2 Les caractéristiques des formants

Les principales caractéristiques d'un formant sont :

- l'amplitude, qui est celle de l'harmonique le plus renforcé.

### III.2

-la hauteur, fréquence centrale de la bande, correspondant à l'harmonique le plus intense.

-la bande passante à "3dB" exprimée en Hz.

-le "patrimoine harmonique" qui peut faciliter ou au contraire rendre difficile sa détection; un formant sera d'autant plus visible que le nombre de composantes spectrales le constituant est important. Ainsi, le spectre d'une voix de femme ou d'enfant, étant pauvre, les formants sont peu visibles et difficiles à détecter. Mais, heureusement, ce n'est pas le cas pour une voix d'homme. En effet, une telle voix a une fréquence fondamentale qui est, en moyenne, inférieure d'une octave à celle d'une voix de femme.

Ceci est illustré par les sonagrammes (fig III.2) d'une même phrase prononcée par un homme, puis par une femme (la bande passante du filtre d'analyse est de 300 Hz).

-l'évolution fréquentielle pour un même son: les fréquences des formants restent fixes durant l'élocution d'une voyelle et évoluent pour une consonne. Ceci est justifié par la quasi-stabilité du conduit vocal dans le premier cas, et ses déformations rapides dans le second. Les transitions dues à ces déformations sont alors représentatives de la consonne produite.

-la situation dans le spectre selon la nature du son: les formants de bruits (des sons bruités) se situent dans la région supérieure du spectre. Par contre, les sons nasalisés sont caractérisés par un affaiblissement des formants aux fréquences élevées.

#### III.1.3 Les formants et la fonction de transfert du conduit vocal



### III.3

La fonction de transfert du conduit vocal, calculée par G.Faut (1960) est :

$$T(p) = \frac{Ds(p)}{De(p)} \quad (\text{III.1})$$

avec :  $Ds(p)$  transformée de Laplace du débit  $Ds(t)$  de sortie.

$De(p)$  transformée de Laplace du débit d'entrée  $De(t)$ .

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

$$T(p) = \prod_{n=1}^{\infty} \frac{1}{\tau_n p^2 + 2\xi_n \tau_n p + 1} \quad (\text{III.2})$$

$2\xi_n$  : bande passante d'une résonnance.

avec :  $\tau_n = 1/\omega_n$  ( $\omega_n$  pulsation propre).

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

$$T(p) = \prod_{n=1}^4 \frac{1}{\tau_n p^2 + 2\xi_n \tau_n p + 1} Cr_4 \quad (\text{III.3})$$

Le terme correctif  $Cr_4$  a été déterminé par G.Faut, il correspond à la fonction de transfert d'un circuit résonnant et d'un filtre passe-haut montés en cascade.

Les cavités du conduit vocal modifient le spectre du signal transmis en favorisant certaines composantes correspondant aux fréquences des pôles de la fonction  $T(p)$  (fig III.1.c). On y voit les régions fréquentielles correspondant à des maximums d'amplitude et représentant les formants. La fréquence " $F_n$ " d'un formant correspond à l'un des pôles de la fonction de transfert  $T(p)$ . On a donc :

$$F_n = \frac{1}{2\pi\tau_n} \quad (\text{III.4})$$

De même, la bande passante à "3dB" d'un formant exprimée en Hz est :

$$\Delta F_n = \frac{\xi_n}{\pi \tau_n} \quad (\text{III.5})$$

### III.2 Les formants et le timbre d'un son :

"Il n'est pas de notion plus vague, ni plus controversée que celle du timbre d'un son" (Matras, J.J)

Alors, nous en donnons une définition d'après ce que nous avons lu dans divers ouvrages, sans prétendre définir cette notion au sens mathématique du terme.

Le timbre d'un son représente la qualité globale de celui-ci et permet de le différencier d'un autre son de même fréquence fondamentale et de même intensité.

Dans l'information constituant le timbre d'un son, on trouve :

- la hauteur des formants (ou positions des maxima du spectre).
- l'intensité des formants (ou amplitudes des maxima du spectre).
- la largeur de bande des formants (ou largeur de bande des maxima du spectre).
- / -la largeur totale du spectre.
- la stabilité plus ou moins grande de la fréquence fondamentale.
- la richesse en harmoniques.
- les transitoires et la situation du son dans une séquence parlée.

## III.5

Il apparait alors que le timbre d'un son est caractérisé par plusieurs paramètres de celui-ci, en particulier par les formants

### III.3 Les voyelles et les formants.

#### III.3.1 Importance des formants dans une voyelle

Soit une voyelle prononcée par un même locuteur à des fréquences fondamentales différentes, ou par différents locuteurs. Son enveloppe spectrale sera presque toujours la même, elle montrera des formants dans les mêmes intervalles de fréquences pour toutes les locutions.

Ainsi, chaque voyelle a ses formants caractéristiques. Ils sont jusqu'à aujourd'hui le seul élément presque invariable que l'on ait trouvé pour distinguer toutes les voyelles entre elles.

Nous avons dit "presque invariable" car : l'harmonique le plus renforcé, dans l'intervalle de fréquences constituant un formant, peut varier d'une locution à une autre et à plus forte raison d'un locuteur à un autre. Toutefois, les limites de cet intervalle, appelé "plage du formant", sont remarquablement constantes

L'analyse acoustique d'une voyelle révèle l'existence de plusieurs formants. Seulement, il a été vérifié et admis que les deux premiers, pris ensemble, suffisent pour caractériser chaque voyelle. Les autres étant plutôt responsables des qualités secondaires de cette catégorie de son.

Ainsi, le premier formant étant une caractéristique principale, la limite supérieure de sa zone fréquentielle ne doit pas être dépassée par la hauteur du fondamental;



sinon, ce formant serait supprimé et le son produit sera autre chose que la voyelle désirée.

On peut donc dire que si l'on veut garder son propre timbre à une voyelle, la possibilité d'augmenter la hauteur du fondamental sera limitée. La limite est la fréquence du premier formant de cette voyelle.

### III.3.2 Classification acoustique des voyelles du Français : triangle de Delattre

Une voyelle ne peut être caractérisée que par ses formants. De plus, l'expérience a montré que les deux premiers sont les plus intenses, et donc les plus importants pour la reconnaître. Ces formants permettent de dresser une classification acoustique des voyelles.

En partant de ce principe, P. Delattre (1948) a établi un "triangle vocalique", appelé triangle de Delattre (fig III.3). Il a dégagé les moyennes des fréquences des deux premiers formants pour les voyelles du Français. Puis il les a portées sur deux axes perpendiculaires. En abscisse, il a noté la fréquence moyenne du deuxième formant et en ordonnée celle du premier. Ce travail a été réalisé pour chaque voyelle.

Si l'on désire comparer les structures formantiques des différentes voyelles, une autre illustration des travaux de Delattre existe (fig III.4). En abscisse sont groupées toutes les voyelles selon leur lieu d'articulation, et en ordonnée on peut lire les fréquences moyennes de leurs deux premiers formants. Sur le tableau ainsi obtenu, on peut voir que les premiers formants  $F_1$  et  $F_2$  sont :

-pour la séquence [u, o, ɔ], groupés et à fréquences relativement basses. Ces voyelles sont alors appelées "voyelles compactes" ou "voyelles graves".

-pour la séquence [ɛ, e, i] très écartés et à fréquence de  $F_2$  relativement haute. Ces x voyelles sont alors dites "voyelles diffuses" ou "voyelles aigües".

-ni groupés comme pour la première séquence ni écartés comme pour la seconde. C'est le cas des voyelles [a] et [œ] appelées "voyelles médianes".

On peut remarquer facilement l'évolution régulière de la structure formantique d'une voyelle à la suivante.

Ceci est justifié par le fait que les voyelles ont été classées par ordre d'articulation.

Pour comprendre cette indépendance entre les phénomènes acoustiques et articulatoires, il faudrait étudier la résonance et son rôle dans la production des sons.

### III.4 La formation des formants et la résonance

#### III.4.1 Définition de la résonance :

La résonance est la propriété qu'a un système oscillant, d'osciller plus volontiers à une certaine fréquence qu'à toutes les autres. Cette fréquence privilégiée est sa fréquence naturelle de vibration appelée : "fréquence de résonance".

Tout système ayant cette propriété est dit résonateur. Il présente une impédance réactive, somme d'une inductance et d'une capacitance. Un exemple de ce type de système est le circuit électrique de la figure III.5



### III.8

La valeur de la fréquence de résonance d'un résonateur est obtenue en annulant son impédance réactive.

#### III.4.2 Effet d'une vibration sur un résonateur

Il existe deux types de résonateurs :

-les résonateurs acoustiques (rares et difficiles à observer) jouent un rôle très important dans la production de la parole.

-les résonateurs mécaniques (nombreux et faciles à observer) n'ont pas de lien avec la phonation.

Ces derniers sont basés sur le même principe que les premiers. De ce fait, pour comprendre aisément le comportement des résonateurs acoustiques, on commencera par présenter celui des résonateurs mécaniques.

##### III.4.2.1 Effet d'une vibration sur un résonateur mécanique :

Une corde de guitare, par exemple, est un excellent résonateur mécanique. Car, pincée puis relâchée, elle oscillera à une fréquence bien déterminée. De plus, si on répète l'opération plusieurs fois (sans modifier l'accord) cette fréquence de vibration sera toujours la même et produira la même note. Ainsi, la fréquence d'accord d'un fil de guitare est sa "fréquence de résonance".

Pour avoir l'effet d'une vibration sur un résonateur mécanique, décrivons une expérience menée sur deux cordes A et B d'une même guitare.

Dans un premier temps, les deux cordes sont accordées sur deux notes différentes (leurs fréquences de résonance sont différentes). Si on pince puis on relâche la corde A, la corde B reste apparemment immobile (fig III.6.a).



Dans un deuxième temps, les deux cordes sont accordées sur la même note (leurs fréquences de résonance sont égales). La mise en vibration de la corde A entraîne celle de B à la même fréquence. On dit que B est entrée en résonance avec A (fig III.6.b).

Ce comportement bizarre à priori, est dû à la propriété de résonance que possède la corde B. Cette dernière ne peut osciller aisément jusqu'à sa fréquence caractéristique. Ainsi, dans le premier cas d'accord, elle ne réagit pas à la vibration issue de A (la fréquence de cette dernière étant différente de celle de B).

Au contraire dans le deuxième cas, les deux cordes ont la même fréquence de résonance. De ce fait, la vibration issue de A se présentera vers B sur le rythme le plus favorable pour sa mise en oscillations aisées.

Aussi, suit-elle ce rythme, sans nécessiter la moindre énergie extérieure au système formé par ces deux cordes ?

Conclusion : Un résonateur de fréquence de résonance  $f_0$  oscillera sous l'effet d'une vibration extérieure de fréquence  $f$ , si  $f_0 = f$

#### III.4.2.2 Action d'un résonateur sur une vibration extérieure complexe

Cette action peut être prévue à partir de la précédente conclusion. En effet, une vibration complexe est une composition de plusieurs vibrations simples (cf chap II.3.2.2). Alors, si elle se présente sur un résonateur elle subira les modifications suivantes :

-le résonateur sera mis en oscillation par l'onde simple ayant la même fréquence que sa fréquence de résonance. Par conséquent, il la transmettra à sa sortie.

-au contraire, pour toutes les autres vibrations simples, il restera indifférent ou vibrera d'une façon très faible selon qu'il est parfait ou pas. Ainsi, il les étouffe, sinon tout au moins, il les amortit.

Nous retrouvons dans cette action du résonateur sur les différentes ondes simples, l'effet même du filtrage. Donc, tout résonateur est aussi un filtre.

### III.4.3 Les résonateurs acoustiques

#### III.4.3.1 Description :

Un résonateur acoustique est une cavité presque fermée, de forme ramassée, toutes les dimensions étant du même ordre de grandeur. Il est en relation avec l'extérieur par deux trous ou deux petits tuyaux. L'un reçoit la vibration excitatrice : "l'entrée"; l'autre la transmet : "la sortie".

Les résonateurs les plus connus sont ceux de forme sphérique de Helmholtz ou ceux de forme cylindrique de Kœnig (fig III.7).

#### III.4.3.2 Définition :

Un résonateur acoustique est un système qui permet de discerner dans une onde plus ou moins complexe, la présence d'une composante de fréquence donnée. Pour cela, il devient lui même le siège d'oscillations de même fréquence (cf chap III.4.2.2)

#### III.4.3.3 Fonctionnement :

Les parois d'un résonateur acoustique sont en général rigides. Elles ne peuvent par conséquent entrer en résonance. Mais, la masse d'air qu'elles entourent, est élastique. C'est elle qui se comportera en résonateur.

Une brusque augmentation de pression, à l'entrée du résonateur, comprimera la masse d'air. Puis une fois la perturbation passée, cet air élastique réagira en cherchant à reprendre son volume initial. Il se gonflera jusqu'à heurter les parois du résonateur, par la suite il se recomprimera et le cycle recommence...

Donc, toute perturbation met en oscillations la masse d'air du résonateur. Ces oscillations seront aisées ou difficiles selon leur fréquence. Le résonateur acoustique a donc lui aussi sa fréquence de résonance.

#### III.4.3.4 Propriétés des résonateurs acoustiques

##### 1) Les résonateurs parfaits :

Un résonateur est dit parfait s'il ne vibre qu'à une et une seule fréquence  $f_0$  (( fig III.8.a). La valeur de  $f_0$  est donnée par la relation suivante :

$$f_0 = \frac{v}{2\pi} \sqrt{\frac{\sqrt{S}}{U}} \quad (\text{III.6})$$

- avec : .v : vitesse de propagation de la perturbation dans le résonateur.  
 .S : section de l'ouverture de sortie du résonateur.  
 .U : volume du résonateur.

En général, un résonateur n'est pas parfait. Il vibre pour une bande de fréquences plus ou moins étroite (fig III.8.b). Cette bande est appelée bande passante du "résonateur filtre". La fréquence centrale correspond à une amplitude maximale de vibration. Elle représente alors la fréquence de résonance du résonateur.



## 2) Application en filtre sonore :

Un son complexe est soit un mélange d'harmoniques (s'il est périodique), soit un mélange de partiels (s'il est apériodique) (cf chap II.3.2.2). Alors, si on le présente à l'entrée d'un résonateur, il sera filtré (fig III.8). Seuls les harmoniques (ou les partiels), dont les fréquences sont voisines de celle qui caractérise ce résonateur, seront transmis (cf chap III.4.2.2). En outre, ce dernier produira une amplification de l'harmonique (ou partiel) ayant une fréquence égale à sa fréquence de résonance.

Donc, un résonateur peut jouer non seulement un rôle de "filtre" mais aussi "d'amplificateur". Il en est ainsi des différents résonateurs du conduit vocal.

## 3) l'effet "d'amplification" :

Soit un résonateur de fréquence de résonance " $f_0$ " et une onde arrivant à son entrée avec la même fréquence. Celui-ci vibrera alors naturellement et sans effort; chaque oscillation extérieure qui arrive rencontre et renforce un mouvement déjà amorcé. Il s'ensuit que ce mouvement est de plus en plus amplifié. Donc, l'onde qui se trouve à l'entrée est transmise à la sortie avec une amplification qui augmente à chaque cycle.

D'ailleurs, si cette dernière n'est pas amortie, elle peut même atteindre en quelques cycles, des proportions dangereuses. L'exemple d'un pont qui s'est effondré, lors du passage d'une troupe militaire qui "marchait au pas", est assez connu.

#### 4) l'équivalence résonateurs acoustiques-circuits électriques

Un résonateur acoustique présente depuis son entrée une impédance donnée par :

$$Z_c = \frac{1}{j (U/\rho V^2) \omega} \quad (\text{III.7})$$

avec :

- . U : volume du résonateur.
- . V : vitesse de propagation de la perturbation dans le résonateur.
- .  $\rho$  : masse volumique du gaz dans le résonateur.

Le résonateur se comporte donc comme une capacitance de capacité  $(U/\rho V^2)$ . Quand à son ouverture de sortie, elle peut être assimilée à un petit tuyau de longueur  $l_0 \sqrt{S}$  ("S" : section de cette ouverture).

Ainsi, son impédance sera :

$$Z_L = j \left( \frac{\rho l_0}{S} \right) \omega \quad (\text{III.8})$$

et elle se comportera comme une inductance de coefficient de self  $(\rho l_0/S)$ .

On peut donc admettre qu'un résonateur acoustique parfait est équivalent à un circuit-bouchon (L,C).

L'impédance de l'ensemble "résonateur + ouverture" de sortie est :

$$\begin{aligned} Z_t &= Z_c + Z_L \\ Z_t &= j \left( \frac{\rho l_0}{S} \omega - \frac{\rho V^2}{U \omega} \right) \quad (\text{III.9}) \end{aligned}$$

La pulsation de résonance " $\omega_0$ " du résonateur est obtenue quand l'impédance est nulle :

$$\omega_0 = V \sqrt{\frac{S}{U l_0}} \quad \# \quad V \sqrt{\frac{S}{U}} \quad (\text{III.10}^c)$$

On en déduit la fréquence de résonance donnée par la relation III.6.

4) L'effet de volume sur la fréquence de résonance  $f_0$  :

Si l'on dispose d'un résonateur de volume variable (comme les résonateurs du conduit vocal), sa fréquence de résonance variera inversement à son volume (III.6). L'explication physique de cette relation entre  $f_0$  et  $V$  peut être déduite du principe de fonctionnement du résonateur.

Soit une vibration (suite de compressions-dilatations) qui se présente à l'entrée du résonateur. Lorsque la première compression arrive, elle provoque un déplacement d'air. Cette perturbation ne peut atteindre les différents points des parois du résonateur qu'après une durée " $\Delta t$ ". Cette durée est proportionnelle aux dimensions de l'appareil résonant, donc à son volume. Plus ce dernier est grand, plus " $\Delta t$ " sera grand et inversement.

Pour qu'il y ait résonance, il faut que l'air déplacé heurte les parois du résonateur et revienne de lui-même en arrière. Donc, la compression doit durer exactement  $\Delta t$ , et la période de la vibration excitatrice produisant la résonance sera " $T_e = 2\Delta t$ " (compression dilatation). Comme cette période est égale à la période de résonance " $T_0$ ", on aura :  $T_0 = 2\Delta t$ .

Ainsi, la période de résonance est proportionnelle à  $\Delta t$  qui est proportionnel au volume du résonateur. La fréquence de résonance  $f_0$  est égale à  $1/T_0$ , alors,  $f_0$  varie inversement au volume.

Exemple :

Quand on verse de l'eau dans une bouteille, le volume d'air résonant diminue. Par contre, le son produit est de plus en plus aigu (la fréquence de résonance augmente).



5) Effet de l'ouverture de sortie sur la fréquence de résonance  $f_0$ .

Soit un résonateur dont l'ouverture de sortie, de section  $S$ , est variable. Sa fréquence de résonance sera d'autant plus importante que cette section est grande (cf chap. III.6). Il en est ainsi des cavités buccales et laryngales.

Comme pour l'effet de volume, nous donnons une explication physique de cette relation.

Si une compression est présente à l'entrée du résonateur, elle se propage. À la sortie, elle provoque une poussée sur l'air du milieu voisin. Cette poussée sera d'autant plus importante que l'ouverture de sortie du résonateur est importante.

Examinons l'effet de cette ouverture sur la fréquence de résonance :

a- si l'ouverture de sortie est importante par rapport aux dimensions du résonateur : la poussée exercée sur l'air du milieu extérieur est grande. La vibration extérieure créant cette poussée peut être de basse ou haute fréquence.

b- - si la vibration est à basse fréquence, la compression qu'elle exerce ne cessera pas rapidement. Par conséquent, l'ouverture de sortie étant grande, l'air déplacé trouvera une résistance faible et pourra s'échapper.

- Si la vibration est à haute fréquence (tout en vérifiant la condition sur l'effet de volume), sa compression cessera si rapidement que l'air déplacé reprendra sa place sans s'échapper. Dans ce cas on obtient la résonance.

Donc, si l'ouverture de sortie est importante, la fréquence de résonance est haute.

b- si l'ouverture de sortie est de section faible par rapport aux dimensions du résonateur: dans ce cas, la pression exercée sur l'air du milieu extérieur, en présence d'une compression à l'entrée sera faible. Ainsi, même si cette compression est due à une vibration excitatrice de basse fréquence, l'air du résonateur ne pourra s'échapper facilement. Plus l'ouverture est petite, plus l'air aura des difficultés à s'échapper et la fréquence de résonance sera basse. Elle dépendra surtout de l'effet de volume.

Exemple : Si on produit la séquence [u-o-a] dans cet ordre, on constatera que les sons sont de plus en plus aigus (on monte en fréquence). De même, l'ouverture des lèvres (ouverture de sortie du résonateur buccal) est de plus en plus grande.

Tableau récapitulatif des propriétés des résonateurs acoustiques

Volume du résonateur	Fréquence de formant	Ouverture de sortie
diminue →	Monte	← augmente
augmente →	Baisse	← diminue

Si le volume et l'ouverture de sortie varient dans le même sens, ils produisent des effets inverses. Mais il est à noter que l'effet de l'ouverture l'emporte sur celui du volume en règle générale.

### III.4.3.5 Générateur de formant par un résonateur acoustique.

Un train d'impulsions très brèves peut donner naissance à un son dont l'analyse spectrale montrera un spectre uniforme et très riche en harmoniques. Ce son est appelé "son blanc".

Soit un résonateur de fréquence  $x$  dev résonance 250 Hz. A son entrée, on présente un son blanc de fréquence fondamentale 125 Hz (fig III.9). Alors, on obtient à la sortie :

- les harmoniques 1 et 3 avec leurs amplitudes sensiblement amorties.

- l'harmonique 2 (de fréquence 250 Hz) amplifié et les autres harmoniques sont presque totalement étouffés.

Les trois harmoniques  $x$  ainsi obtenus, par ce filtrage, forment un formant centré sur 250 Hz. On peut voir sa trace sur le sonogramme du son filtré (à l'extrême gauche de la fig III.10). Grâce à sa propriété de filtre passe-bande, le résonateur acoustique a engendré un formant. En présence d'un bruit blanc, le "résonateur-filtre" réagira de la même façon.

Le bruit blanc a un spectre uniforme et continu, il est composé non pas d'harmoniques mais d'une infinité de partiels.

Notre "résonateur filtre" transmettra alors la bande de bruit, formée de partiels de fréquences proches de sa propre fréquence de résonance. Nous obtenons alors un formant comparable à celui généré précédemment. Ceci explique le fait que la voix chuchotée, formée uniquement de bruit (sans vibration laryngienne), peut également donner naissance à des voyelles.





-le résonateur  $R_1$  constitué par la cavité pharyngale (en arrière du dos de la langue), qui génère le formant bas des voyelles  $F_1$ .

-le résonateur  $R_2$  constitué par la cavité buccale (en avant du bes de la langue), qui génère le formant haut des voyelles :  $F_2$ .

La disposition de  $R_1$  et  $R_2$ , dans le conduit vocal, fait qu'ils forment ensemble un résonateur double. Le couple  $(F_1, F_2)$ , représentant "les fréquences de résonance" de ce dernier, peut être modifié par la position de la langue, du voile du palais et des lèvres.

La mobilité de ces différents organes est responsable de la diversité acoustique des sons.

La langue joue un rôle prépondérant dans le résonateur double, pharyngal et buccal. Par ses diverses positions elle modifie les volumes des deux résonateurs et l'ouverture de sortie de  $R_2$ . Citons en exemple la production de [u] et [i] (fig III.12).

Ces considérations montrent l'étroitesse des liens entre les phénomènes articulatoires et acoustiques. D'ailleurs, le "trapèze articulatoire" (fig III.13) nous rappelle la forme du triangle de Delattre.

En outre, cette théorie des résonateurs permet d'expliquer aisément l'évolution de la structure formantique, d'une voyelle à l'autre.

#### III.4.6 Evolution de la structure formantique d'une voyelle à l'autre.

##### III.4.6.1 Les causes de fluctuations du premier formant $F_1$ :

Le formant  $F_1$  prend naissance dans la cavité pharyngale. La position de la langue détermine l'ouverture de sortie et le volume de celui-ci.

Ainsi la fréquence de  $F_1$  dépend essentiellement de l'abaissement plus ou moins grand de la langue. Plus la langue est basse plus  $F_1$  est aigu ( $[u]$ ,  $[o]$ ,  $[ɔ]$ ,  ~~$[ɔ]$~~   $[a]$ ), plus elle est haute, plus  $F_1$  est grave ( $[a]$ ,  $[ɛ]$ ,  $[e]$ ,  $[i]$ ).

Ceci apparaît est comparant la position de la langue (fig III.13) avec le formant  $F_1$  (fig III.4), pour deux voyelles consécutives.

#### III.4.6.2 Les causes des fluctuations du deuxième formant des voyelles non labialisées :

Pour ce formant, il est nécessaire de séparer les voyelles en deux classes :

-les voyelles compactes pour lesquelles la fréquence de  $F_2$  dépend essentiellement de l'ouverture des lèvres, plus celle-ci est grande, plus  $F_2$  est aigu.

-les voyelles médianes et diffuses pour lesquelles la fréquence de  $F_2$  est déterminée surtout par le volume de la cavité résonnante buccale; plus ce volume est petit, plus  $F_2$  est aigu.

#### III.4.6.3 Cas des voyelles labialisées :

Les voyelles labialisées ( $[œ]$ ,  $[ø]$ ,  $[y]$ ) proviennent de leurs correspondantes ( $[ɛ]$ ,  $[e]$ ,  $[i]$ ) non labialisées. Les unes et les autres sont exactement prononcées avec la même articulation pour chaque couple ( $[œ]$  -  $[ɛ]$  ;  $[ø]$  -  $[e]$  ;  $[y]$  -  $[i]$ ). La labialisation d'une voyelle consiste à arrondir les lèvres. Ceci diminue leur ouverture de sortie et par suite la fréquence du formant  $F_2$ . Il s'ensuit que chaque voyelle labialisée est moins aiguë que sa correspondante non labialisée (fig III.4). Quand au formant  $F_1$ , il ne varie pas car sa fréquence ne dépend que de l'articulation qui reste inchangée.



## III.4.6.4 Cas des voyelles nasales :

Les voyelles nasales courantes ( $[\tilde{o}]$ ,  $[\tilde{a}]$ ,  $[\tilde{e}]$ ) ont à peu de choses près les mêmes fréquences des formants que leurs correspondantes orales ( $[o]$ ,  $[a]$ ,  $[e]$ ). Elles n'en diffèrent que par les intensités relatives des formants  $F_1$  et  $F_2$  et quelques complications secondaires de leur structure formantique. Ces dernières sont dues à certains effets acoustiques engendrés par le passage d'une partie de l'air phonatoire par les fosses nasales.

Les modifications formantiques, subies par une voyelle orale pour se transformer en une nasale, peuvent être séparées en deux classes :

-les modifications sans valeurs acoustiques telles que : le regroupement des formants  $F_3$  et  $F_4$  dans certains cas, ou encore l'apparition d'un faible formant vers 4500 Hz, dans d'autres cas. Ces deux changements peuvent être aussi simultanés.

-Les modifications qui constituent les véritables paramètres acoustiques de la nasalité :

.l'effet d'antirésonance qui se manifeste par des affaiblissements de certaines bandes de fréquences de la structure formantique.

. le renforcement des composantes sonores qui se situent entre 200 et 300 Hz. Il est, en partie, responsable de l'impression de nasalité; on l'appelle le formant nasal (F.N).

Finalement, on retiendra surtout que les voyelles orales diffèrent de leurs correspondantes nasales par les amplitudes relatives de leurs formants.

### III. 22

#### III.4.6.5 La similitude formantique des voyelles parlées et chuchotées.

Une voyelle est parlée lorsque le signal d'excitation, à l'origine de sa production, est périodique (vibration laryngienne). Elle est chuchotée si celui-ci est un bruit de souffle (cordes vocales légèrement écartées et immobiles (chap I)).

Cependant, les spectres de ces deux signaux ne présentent qu'une seule différence notable : le premier discontinu, est formé d'harmoniques et le second continu est formé de partiels.

L'action du résonateur double (pharyngo-buccal) sur ces deux spectres sera la même car l'articulation reste inchangée. Il s'ensuit que deux voyelles correspondantes (chuchotée-parlée) ont des formants de même fréquence. Ils diffèrent surtout par leur composition, l'un étant formé d'harmoniques (voyelle parlée), l'autre d'une bande de bruit (voyelles chuchotées).

#### III.4.7 Caractéristiques spectrales des différentes consonnes.

Les consonnes ne peuvent être caractérisées par leurs formants comme les voyelles. Pour faciliter leur identification dans un sonagramme, on fait une étude sur leur composition spectrale.

##### III.4.7.1 Les consonnes fricatives (ou spirantes)

On peut les classer par couples ( $[f]/[v]$ ,  $[s]/[z]$ ,  $[ʃ]/[ʒ]$ ) dont le premier élément est un son sourd et le second un voisé. La composition spectrale de deux consonnes d'un même couple est pratiquement la même. La seule différence est l'apparition des "stries" verticales, représentant l'attaque glottale, dans le cas des voisés.



Le spectre de la paire [ʃ] / [ʒ] se distingue par une forte concentration de bruit, en bande assez large, centrée sur 4500 Hz environ. La composition de ce bruit est différente selon que la voyelle attenante est grave, médiane ou aiguë. Les résonateurs pharyngal et buccal favoriseront dans chaque cas, des formants de bruit situés dans des zones de fréquences différentes. Ceci s'explique par le fait que les propriétés de ces résonateurs dépendent de la voyelle attenante.

Pour la paire [s] / [z], sa composition spectrale présente une bande de bruit située surtout dans la région supérieure du spectre. La composition de ce bruit dépend de la voyelle attenante. Si cette dernière est grave, il comportera deux bandes assez écartées. Sinon, celles-ci se rapprochent au fur et à mesure que la voyelle attenante est aiguë. Ce phénomène s'explique de la même façon que pour [ʃ] / [ʒ].

Quant à la paire ([f] / [v]), sa composition spectrale est caractérisée par une bande de bruit fricatif, très large, englobant presque toutes les fréquences audibles. Elle ne présente pas de zone de concentration notable. Sa composition est indépendante de la voyelle attenante. Ceci est dû au fait que le bruit ne naît plus dans les cavités phonatrices mais au niveau des lèvres. Aussi, aucun phénomène de résonance ne le modifie.

#### III.4.7.2 Les consonnes plosives (ou occlusives)

Ce sont les consonnes sourdes [t], [p], [k], et leurs correspondantes voisées [d], [b], [g]. Elles sont caractérisées par :

-un silence total pendant la tenue des sourdes et presque total pendant celle des voisées. Ceci se traduit par une zone claire dans le spectre.



n -un bruit très bref (explosion) plus important et plus durable pour les sourdes que pour les sonores. Ce bruit est formé par des partiels à des fréquences très diverses. Il se manifestera donc, sur le sonagramme, par une mince barre verticale.

#### III.4.7.3 Les consonnes nasales [m] et [n]

Pour ces consonnes, le conduit vocal est fermé. Par contre, les fosses nasales ne sont plus obstruées par le velum. Elles agissent non pas comme un filtre résonnant mais comme un filtre coupe-bande. Leur effet détruit presque systématiquement tous les formants dont les fréquences seraient ~~formé~~ comprises entre 500 et 2000 Hz. C'est la raison pour laquelle on ne trouvera qu'un formant  $F_1$  de fréquence très basse et un formant  $F_3$ . Ce dernier est très souvent dédoublé par suite de complications acoustiques introduites par la nasalité.

#### III.4.7.4 Les consonnes liquides [l] et [r]

On les appelle "liquides" car elles résultent d'une modification profonde des caractéristiques des voyelles attenantes. Leurs spectres dépendent de celui de ces dernières.

Dans l'articulation de ces consonnes, la langue obstrue partiellement la cavité buccale. Il s'ensuit que le formant  $F_2$  est de faible intensité.

#### III.4.7.5 La consonne pharyngale [R]

Cette consonne très utilisée en français est fricative. Il faut la distinguer de celle liquide [r] ("r" roulé) utilisée dans bien d'autres langues.

Le spectre de [R] découle de celui de la voyelle attenante. Il se distingue par le fait que le premier formant de celui-ci disparaît presque complètement. Quand au second, il se transforme en une "tâche" de bruit hâché suivant un rythme régulier.

Ces hâchements sont dus au fait que la partie arrière de la langue vient s'accoler au palais suivant ce même rythme. La composante sonore du deuxième formant n'est ailleurs libérée que par-à-coups au moment des décollements.

#### III.4.7.6 Les semi-consonnes [w], [ɥ], [j]

Du point de vue acoustique et donc spectral, les semi-consonnes sont des phonèmes qui s'apparentent surtout aux voyelles. D'ailleurs, on les appelle aussi les semi-voyelles.

Leur caractère consonnantique est purement fonctionnel. Il vient du fait qu'elles s'emploient toujours avec au moins une voyelle attenante.

Exemples : [wi], [ɥi] et [jɔ]

Enfin, les semi-consonnes ont une durée très brève (1/2 à 1/4 de la voyelle attenante).

#### III.4.7.8 Les transitions de formants

Une consonne s'emploie toujours avec une (ou des) voyelle(s). L'ensemble acoustique consonne+voyelle, ou vis-versa, est appelé "un legatome". Ce dernier constitue un élément indivisible du langage.

DANS un legatome, les formants de la voyelle sont notablement perturbés par la proximité de la consonne. Ce sont ces perturbations caractéristiques que les appelle les "transitions de formants". Ces dernières s'expliquent par l'évolution des volumes et ouvertures de sortie des résonateurs de l'appareil phonatoire, lors du passage d'une consonne à une voyelle et vis-versa/

Aussi, les consonnes sont caractérisées par leurs propres spectres et les transitions de formants des voyelles avoisinantes. Ces transitions convergent toutes vers une fréquence unique caractéristique de chaque consonne : "le locus".

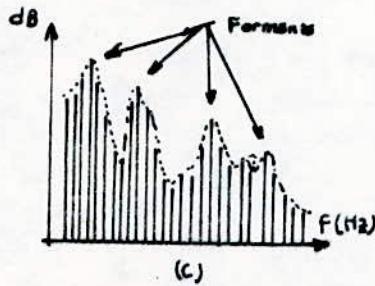
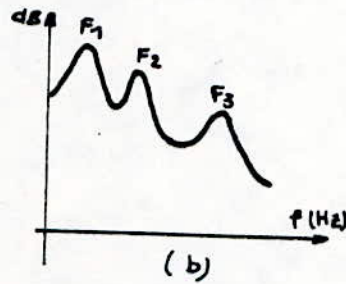
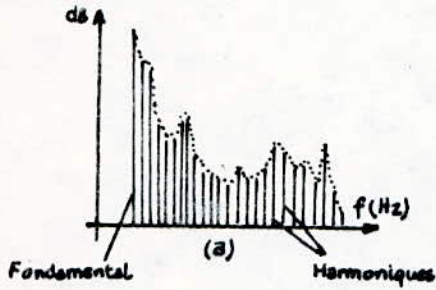


fig III. 1 :

Le signal sonore dont le spectre est en (a), traverse le conduit vocal de notre courbe de résonance en (b). On obtient le signal vocal en (c) les résonances observées sont les formants



FIGURE 1

Sonagramme d'une phrase prononcée par un homme.

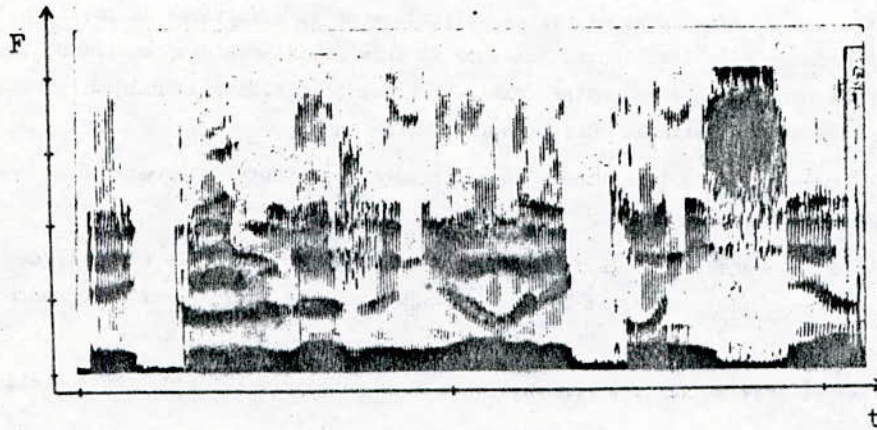


FIGURE 2

Sonagramme d'une phrase prononcée par une femme.

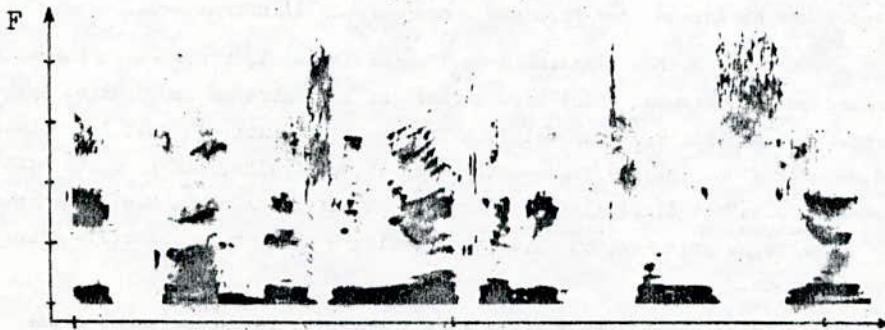


fig. III. 2 . Sonagrammes d'une même phrase<sup>t</sup>  
prononcée par un homme puis par une  
femme .

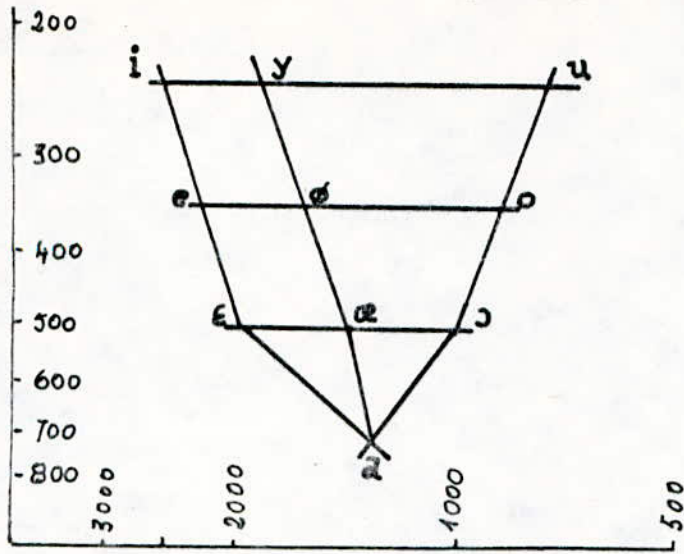


fig III. 3 Triangle vocalique de Delattre

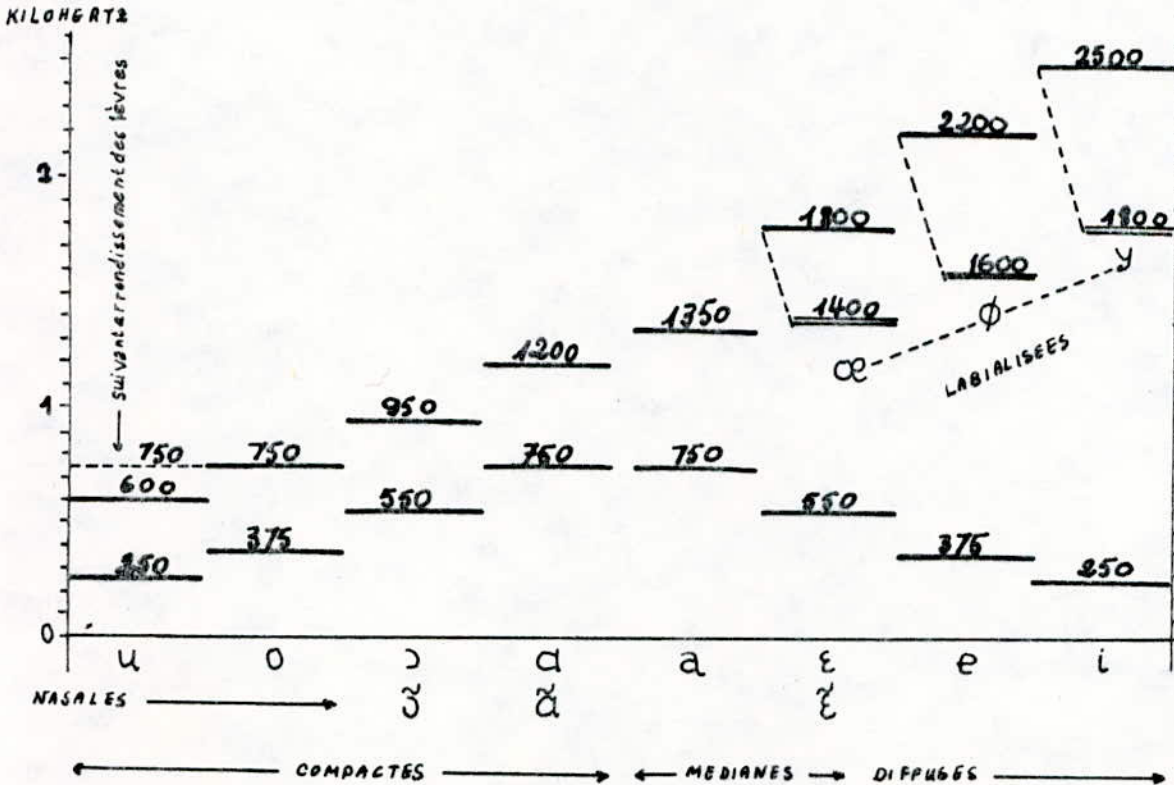


fig III. 4 Tableau de classification des voyelles selon l'ordre d'articulation.

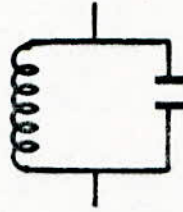


fig III.5 Circuit électrique résonant.

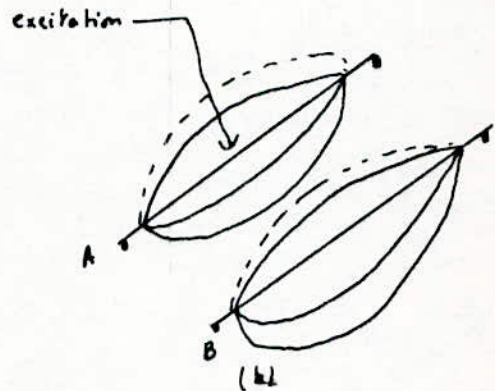
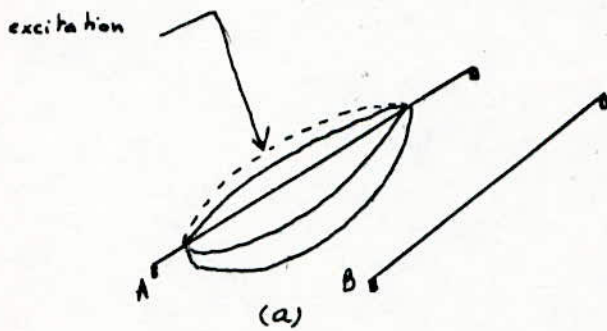


fig III.6 : Entrée en résonance d'un résonateur mécanique

- (a) Les deux cordes sont accordées sur des fréquences différentes.
- (b) Les deux cordes sont accordées sur la même fréquence.



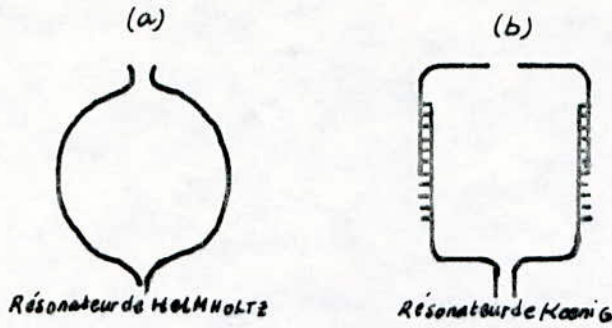


fig III.7 Les résonateurs acoustiques

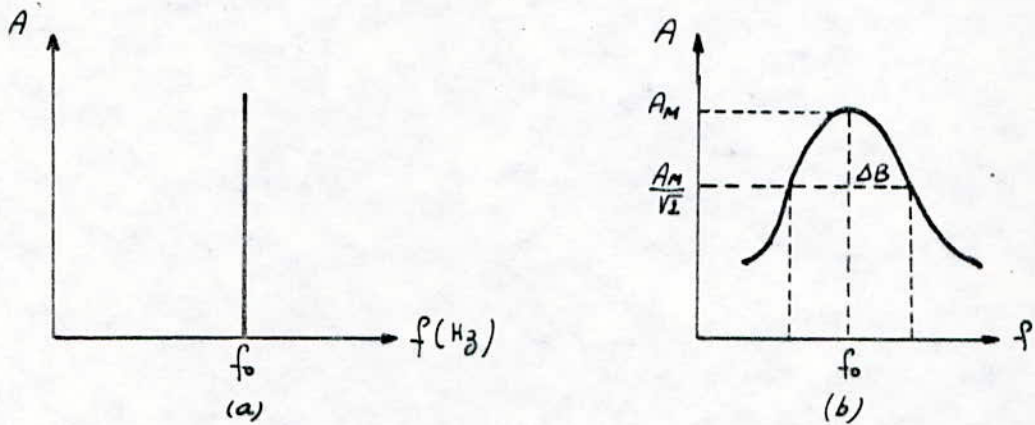


fig III.8 Réponse en fréquence d'un résonateur acoustique  
(a) parfait (b) non parfait.

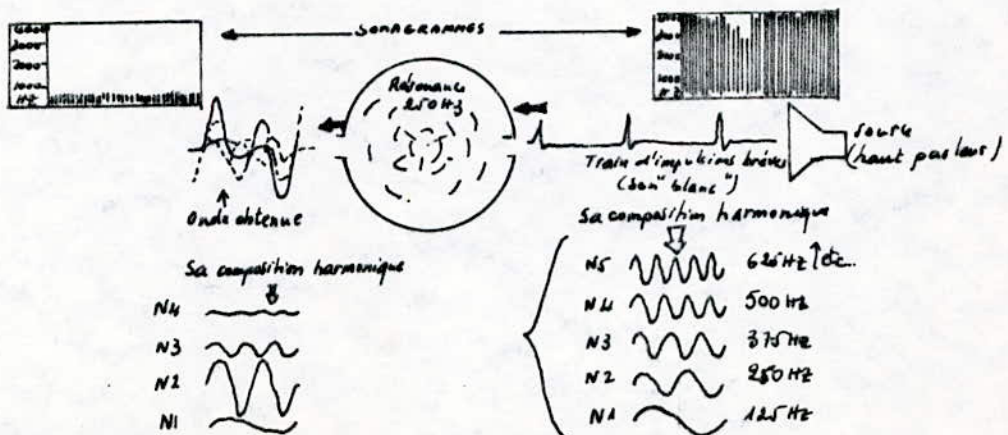


fig III.9 Génération d'un formant par action d'un résonateur acoustique sur un son blanc.

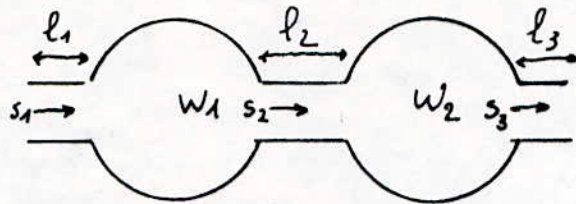


fig IV. 10 Résonateur double.

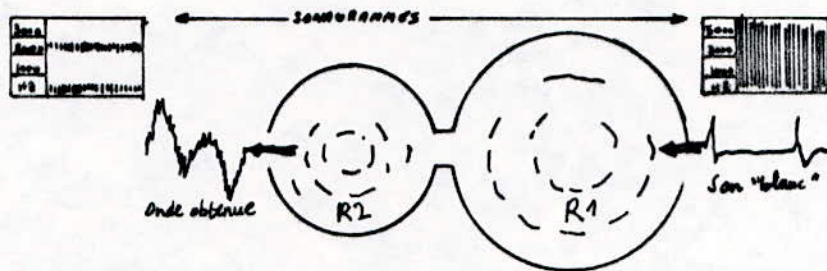


fig III. 11 Génération de deux formants par action d'un résonateur double (250 et 2500 Hz) sur un son blanc.

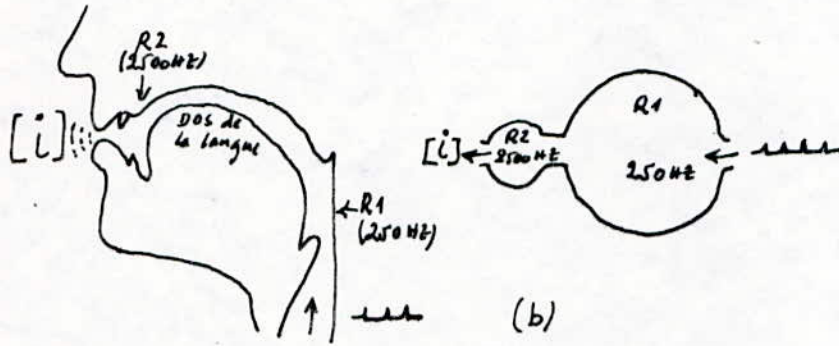
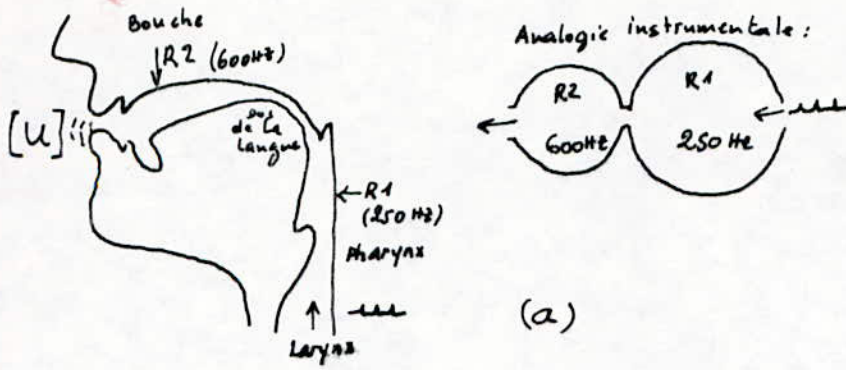


fig III. 12 Effet de la position de la langue dans le résonateur double du conduit vocal

(a) production de [u]      (b) production de [i]

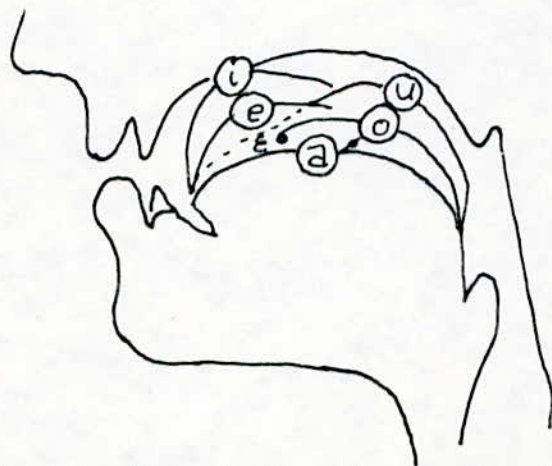


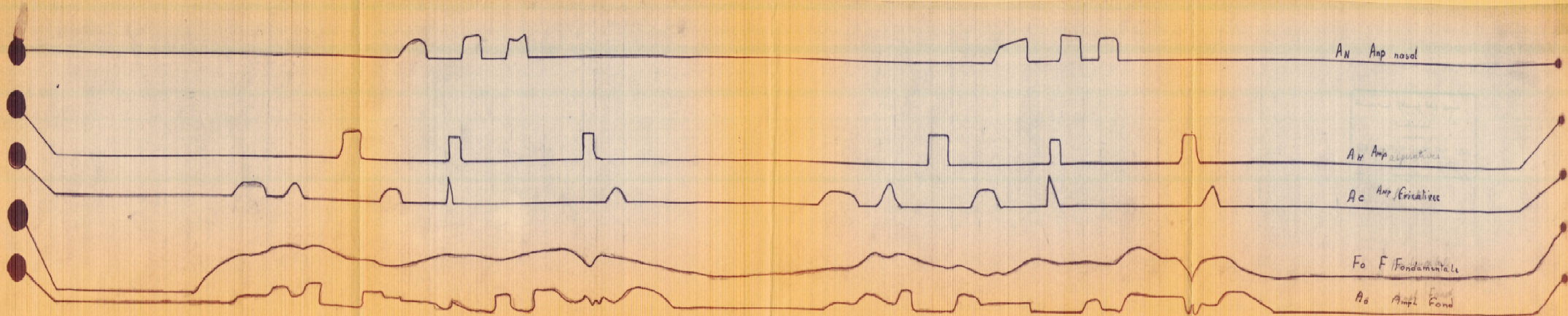
fig III. 13 : "Trapeze articulatoire"; positions du dos de la langue pour l'élocution de chaque voyelles.



Chapitre IV.0

SYNTHESE DE LA PAROLE

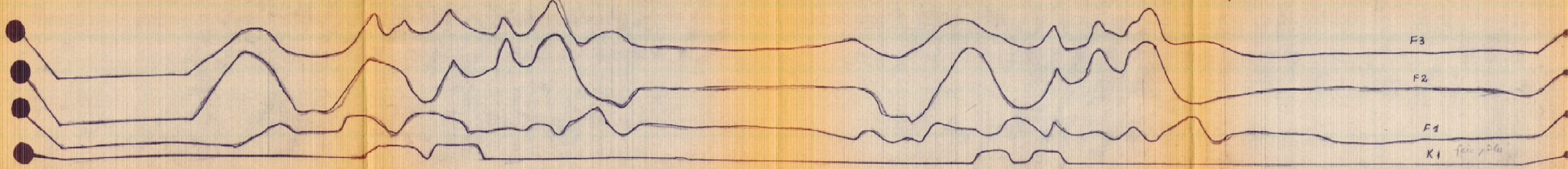
- IV.1 Historique
- IV.2 Technique de synthèse
- IV.3 Les différentes méthodes de synthèse
- IV.4 L'Electrical Vocal Analog(E.V.A.III).



3 əvu PR e z ā t m e m e j æ R v ø

3 əvu PR e z ā t m e m e j æ R v ø

بیت کتب  
کتاب  
ÉCOLE NATIONALE POLYTECHNIQUE  
BIBLIOTHÈQUE





## LA SYNTHÈSE DE LA PAROLE

On désigne par synthèse de la parole la production de messages parlés, soit par assemblage d'éléments phonétiques préenregistrés, soit par règles. Le qualificatif synthétique, loin de signifier que les éléments de base ont été recueillis sans le recours à un locuteur humain, traduit plutôt l'opération inverse de l'analyse. On pourrait apporter certaines modifications aux éléments de synthèse, le timbre ou l'accent du locuteur qui aura contribué à l'enregistrement n'en seront pas affectés.

Avant d'arriver à ce stade, la synthèse de la parole a traversé diverses étapes, suivant une certaine évolution. Dans ce chapitre, nous commençons par présenter brièvement les péripéties de cette évolution, puis les différentes techniques et méthodes de synthèse et enfin l'étude d'un exemple de synthétiseur formantique: l'E.V.A III qui constitue l'outil même de notre travail.

## IV.1 Historique

Depuis près de deux siècles, un grand nombre de chercheurs ont consacré leurs travaux à la simulation de la voix humaine. Les premières machines furent mécaniques, il faut dire que leur évolution a été trop lente, jusqu'à notre ère où les synthétiseurs électriques ont fait leur apparition. Un bref historique retraçant les grandes étapes traversées par le développement de ces machines parlantes ~~ixième~~ a été proposé sous forme de diagramme (M. GUERTI, 1984). Nous le reproduisons ci-après :




# HISTORIQUE

## Machines Parlantes


### Mécaniques

### Électrique

**KRÄTZENSTEIN (1782)**  
 À l'aide de résonateurs  
 acoustiques de sections  
 différentes, on pouvait  
 reproduire les voyelles



**VON. KEMPELEN (1791)**



Simulation acoustique des  
 organes phonatoires humains

**J. FABER (1835)**



Réalisation d'une machine  
 qui parle, chuchotte et chan-  
 te. Elle est dotée de cer-  
 taines particularités qui  
 permettaient une améliora-  
 tion de la qualité de la voix.

De nombreux essais d'imitation  
 des voyelles ont été faits  
 grâce à des procédés divers.

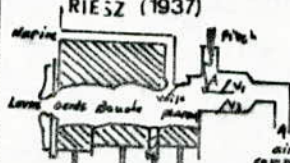
Les sirènes de  
 HENRI (1832)

Les diapasons de  
 HELMHOLTZ (1875)

Les tuyaux d'orgue de  
 MILLER (1916) et de  
 STUMPF (1926)

Les résonateurs de  
 caoutchouc de PAGOT  
 (1930)

**RIESZ (1937)**



Machine parlante de Macco (1939)  
 Dispositif simulant le con-  
 duit vocal.

**STEWART (1922)**  
 Utilisation de 2 circuits qui simulent les principaux résona-  
 teurs du conduit vocal.

**FLETCHER (1924)**  
 Synthétiseur grâce auquel on arrive à produire quelques mots à  
 l'état isolé

**K W WAGNER (1936)**  
 Réalisation d'un synthétiseur à formants pour les voyelles.

**DUDLEY, RIESZ et WATKINS (1939)**  
 Le Vo-ice DEMONSTRATOR (VODER)  
 permet la reconstitution d'une  
 parole continue à partir des tech-  
 niques électriques de transmis-  
 sion du signal.

**DUDLEY (1939)**  
 Le Vo-ice CODER (VOCODER)  
 a été construit pour la trans-  
 mission économique d'une conver-  
 sation téléphonique.

**BOBERT (1956)**  
 Vocodateur à divisions

**SCHROEDER et al. (1962)**  
 Vocodateur à auto corrélation

**MOWLEY et KLEIN (1963)**  
 Vocodateur à fonction orthogonale

**LINCOLN (1964), PLAVAGAN (1985)**  
 Vocodateur conservant la phase.

## IV.3

### IV.2 Les techniques de synthèse

A première vue, nous pouvons classer les différentes techniques de synthèse en trois grandes familles.

#### IV.2.1 Les techniques temporelles.

Le signal vocal est enregistré directement (sans numérisation), soit sur bande magnétique, soit sur un disque. Cette dernière possibilité facilite la recherche des éléments à assembler lors de la constitution du message. Quel que soit le support utilisé, la parole étant enregistrée sous forme analogique, la voix restituée peut être de très bonne qualité, du moins pour des mots isolés.

#### IV.2.2 Les techniques articulatoires

Elle consiste à simuler l'appareil phonatoire par une succession d'un grand nombre de tuyaux acoustiques ou de circuits électriques élémentaires. Chacun de ces éléments représente un tranche du conduit. La simulation est d'autant meilleure que le nombre d'éléments mis en série sera plus grand. Dans le cas de l'approche électrique, l'alimentation du circuit sera assurée par un signal analogue au signal acoustique issu de la glotte.

Ces techniques reproduisant assez fidèlement le fonctionnement de l'appareil phonatoire permettent d'avoir une bonne qualité de la parole synthétique.

#### IV.2.3 Les techniques spectrales

##### IV.2.3.1 Synthèse par prédiction linéaire

La prédiction linéaire cherche à exploiter la redondance existant dans la forme temporelle du signal de la parole. Elle est fondée sur l'équation (1.4).



#### IV.4

Le principe de la prédiction linéaire repose sur l'hypothèse que l'échantillon " $S_n$ " du signal à l'instant  $n$  peut être calculé à partir de la somme pondérée des valeurs passées  $S_{n-1}$ ,  $S_{n-2}$ , ... et éventuellement  $\delta_n$  qui ne diffère de 0 que lors de l'impulsion du pitch.

Le synthétiseur prédiction linéaire (fig IV.1) est commandée par les paramètres suivants :

- la période  $T_0$  du fondamental (s'il existe).
- un signal binaire relatif au choix de l'excitation (voisée ou non voisée).
- la valeur quadratique moyenne du signal de parole : échantillonnés.
- les  $p$  coefficients de prédiction fournis par l'analyse, considérés comme constants sur une période de 10 à 25 ms.

La commande binaire citée précédemment contrôle la commutation entre le générateur d'impulsions et le générateur de bruit blanc. L'amplificateur "G" ajuste l'amplitude du signal d'excitation. Un sommateur réalise la combinaison de la valeur  $\hat{S}_n$  du signal à l'instant  $nT$ , issu du prédicteur linéaire, avec l'excitation. On obtient ainsi le  $n$ -ième échantillon du signal synthétique qui attaquera un haut parleur à travers un filtre passe-bas.

#### IV.2.3.2 Les vocodeurs à canaux

Ils sont destinés à transmettre la parole avec un débit d'information réduit et se compose de deux parties (fig IV.2) :



-l'analyseur : la fonction d'analyse du spectre est effectuée à l'aide de canaux dont le nombre peut varier suivant les réalisations entre 10 et 20, chaque canal traite une bande de fréquence déterminée. Le signal de la parole issu d'un microphone est analysé au d'un banc de filtres passe-bandes contigus couvrant l'étendue spectrale de la voix. Chaque bande est transmise à un détecteur après quoi va repartir positive traverse un filtre passe-bas dont la fréquence de coupure est de 20 à 50 Hz; ainsi, il ne passera que les variations lentes du signal. La réaction obtenue représentera donc l'évolution au cours de l'élocution de la densité spectrale d'énergie relative à la bande passante du filtre.

Un dispositif supplémentaire permet de différencier les sons voisés des non voisés et par la même dans le premier cas de mesurer la fréquence du fondamental, ce dernier joint à l'enveloppe du spectre à court terme recueilli à la sortie des canaux, permettra de reconstituer le spectre.

Le synthétiseur: pour chaque canal, le signal basse fréquence issu du filtre d'analyse est multiplié par le signal d'excitation provenant du générateur d'impulsion ou du générateur de bruit selon que le signal est voisé ou non voisé. Il passera ensuite par un banc de filtres analogue à celui de l'analyseur. Enfin, l'addition des signaux recueillis à la sortie de chaque filtre donnera le signal final.

Les synthétiseurs à formants:

Partant du fait que l'oreille perçoit davantage les maxima (ou formants) du spectre à court terme que les minima, il est possible d'effectuer la synthèse en utilisant un nombre réduit de filtres associés à chacun des formants. Les appareils répondants à cette technique sont appelés: "les synthétiseurs à formants ou les synthétiseurs paramétriques".

Ils sont basés sur l'examen du mode de production de la p

## IV.6

Ils sont basés sur l'examen d'un mode de production de la parole par l'appareil vocal. Il s'agit donc d'une approche des caractéristiques acoustiques de la parole.

Les formants jouent un rôle de haute importance dans ce type de synthèse. Alors, il nous semble utile de donner certains détails, assez importants qui complètent l'étude de ces paramètres, faites au chapitre III.

### IV.2.3.3a Les formants

-Un formant est modélisable par la fonction de transfert (cf III.2) :

$$H(p) = \frac{\omega_c^2}{p^2 + bp + \omega_c^2} \quad (\text{IV.1})$$

.  $\omega_c$  : fréquence nominale de résonance

.  $b$  : bande passante de la courbe de résonance.

Si on pose :  $p = j\omega$ , le spectre énergétique est donné par :

$$|H(j\omega)|^2 = \frac{\omega_c^4}{[\omega^2 - (\omega_c^2 - b^2/2)]^2 + b^2(\omega_c^2 - b^2/4)} \quad (\text{IV.2})$$

Il passe par un maximum, lorsque  $\omega$  varie et atteint la valeur :  $(\omega_c^2 - b^2/2)^{1/2}$  pour laquelle le premier terme du dénominateur n'annule. Mais, généralement, la largeur de bande " $b$ " est petite devant la valeur de la fréquence centrale " $\omega_c$ ". Ainsi, on peut considérer avec une bonne approximation que le pic se situe à  $\omega = \omega_c$ , et son amplitude est alors égale à  $\omega_c/b$ .

Lorsque la fréquence est supérieure à celle correspondant au pic, l'énergie varie en  $1/\omega^4$ . Il s'ensuit que sur une octave le niveau d'énergie est divisé par le facteur 16. Cet amortissement s'exprime couramment en dB par :

$$10 \log 16 = 12 \text{ dB/octave.}$$



L'énergie minimale de la bande passante est à 3 dB au dessus du pic. Elle vaut alors :

$$\left| [H(j\omega)]_{3d} \right|^2 = \frac{1}{2} \left| [H(j\omega)]_{\max} \right|^2 = \frac{\omega_c^2}{2b} \quad (\text{IV.3})$$

et la valeur correspondante de la fréquence est :

$$\omega_{3d} = \omega_c \pm b/2 \quad (\text{IV.4})$$

La largeur de bande d'un formant est quasi-constante quelle que soit la fréquence du formant. Alors, le filtre électronique qui le réalise doit posséder nécessairement des propriétés inhabituelles. Car, en général, la largeur de bande varie comme la fréquence de résonance, de sorte que l'acuité de la résonance, qui est le rapport  $Q = \omega_c/b$  est constante. Donc, l'amplitude d'un formant croît avec sa fréquence de résonance. Cette dernière est dans un intervalle de fréquence appelé : "plage du formant".

Les plages des formants :

Les fréquences des trois premiers formants vocaux évoluent à l'intérieur de plages de fréquences bien déterminées. Ainsi, Faut (1960) a retenu les valeurs suivantes pour les voix masculines :

$$F_1 : 150 \text{ à } 850 \text{ Hz} ; F_2 : 500 \text{ à } 2500 \text{ Hz} ; F_3 : 1500 \text{ à } 3500 \text{ Hz}.$$

Il admet que ces valeurs doivent être majorées de 17% environ pour les voix féminines.

La détection automatique des formants à partir du spectre de la parole n'est pas simple. Or, pour commander un synthétiseur à formants pour étudier les problèmes de synthèse par règles, pour effectuer des études de reconnaissance de parole à partir des formants, il est nécessaire de pouvoir disposer rapidement, autant que possible en temps réel, de données sur les fréquences des formants.



Plusieurs procédés de détection automatique de ces paramètres ont été expérimentés par divers chercheurs :

-Calcul des moments spectraux (J.Suzuki , Y.Kadokawa,K.Nakata,1963).

-Détection des passages par zero du signal de parole (S.H.Chang,1956).

-Détection des maximums du spectre de la parole (J.L.Flanagan,1956;R.Carré,1966).

-Etude du spectre (M.R Schroeder,A.M.Noll 1965).

Tous ces procédés présentent des insuffisances. Les principales causes du mauvais fonctionnement sont les suivantes :

-la structure formantique n'est pas toujours bien définie (cas des spectres pauvres en harmoniques).

-dans les sons nasalisés, des formants particuliers apparaissent et s'intercallent entre les formants vocaux (apport de zéros et de pôles à la fin de transfert du conduit vocal par la nasalisation).

-les plages d'évolution des formants se chevauchent.

Aussi, pour concevoir un détecteur automatique de formants, il est indispensable d'étudier préalablement leurs évolutions. C'est ce qui permettra de déterminer les règles qui régiront le fonctionnement du détecteur. Des résultats satisfaisants ne peuvent être obtenus qu'après plusieurs analyses. Et, ils sont interprétés par référence aux phénomènes qui entrent en jeu lors de la production de la parole.

#### IV.2.3.3.b Les synthétiseurs à formants en série :

La façon la plus simple de réaliser un synthétiseur à formants est d'associer en série plusieurs filtres. On l'appelle alors : le synthétiseur à formants en série.

Dans ce type de machines, des circuits résonnants reproduisent les résonances formantiques dues au conduit vocal. Ils sont disposés en cascade et traversés successivement par le signal qui deviendra, à la sortie du système, de la parole synthétique.

##### 1°/ Description et principe de fonctionnement :

Le synthétiseur à formants en série (fig IV.3) est composé de deux sources d'excitation (voisée et non voisée) et d'un simulateur du conduit vocal.

Les deux sources d'excitation sont :

-pour un son voisé, un générateur d'impulsions dont le spectre se rapproche de celui du débit d'air au niveau des cordes vocales.

-pour un son non voisé, un générateur de bruit dont les caractéristiques n'ont pas une grande influence sur la qualité de la parole synthétique.

Le simulateur du conduit vocal, déterminant l'enveloppe du spectre, est constitué par trois canaux montés en parallèle :

-le canal des formants vocaux : on y trouve trois résonances variables, qui donneront naissance aux trois premiers formants  $F_1, F_2, F_3$ . La fréquence du quatrième formant est généralement fixe. On y trouve aussi un circuit de correction relatif aux résonances d'ordre plus élevés.



Quand un son est voisé, ce canal est attaqué par la source d'impulsions dont la fréquence de résonance est celle du fondamentale  $F_0$ .

Les bandes passantes formantiques sont le plus souvent invariables. Ceci est justifié par le fait que la largeur de bande des formants est quasiment constante et ne semble pas jouer un rôle important dans la perception de la parole.

La fonction de transfert d'un des circuits de formants pris isolément est donnée par la relation (IV.2). Pour un nombre "M" de tels circuits en cascade, la fonction de transfert globale sera le produit des "M" fonctions de transfert correspondantes. Elle aura pour expression :

$$H(p) = \prod_{n=1}^M \frac{\omega_n^2}{p^2 + b_n p + \omega_n^2} \quad (\text{IV.5})$$

si on pose :  $\tau_n = 1/\omega_n$  et  $2\xi_n = b_n$ , après avoir divisé le numérateur et le dénominateur de cette relation, on aboutit à la fonction de transfert du conduit vocal donnée par la relation (III.2).

-le canal de bruit :

il est destiné à la production des sons bruités (non voisés) et comporte trois circuits de formants de bruit  $B_1, B_2, B_3$ . Son signal d'excitation est fourni par le générateur de bruit.

Les sons bruités voisés sont générés par le canal des formants vocaux. Celui-ci est alors attaqué par les deux sources d'excitation en même temps. C'est la commande  $A_1$  qui permet "d'injecter" du bruit dans ce canal.



-Le canal de la nasalité:

Il contient les circuits de formants de la nasalité ( $N_1, N_2, N_3$ ) et celui traduisant l'effet d'antirésonance du conduit vocal  $N_0$ .

En effet, la mise en fonction du conduit nasal (par l'abaissement du voile du palais) fait apparaître des absorptions d'énergie au voisinage de certaines fréquences. Ces dernières sont alors appelées fréquences d'antirésonance. Si elles sont au nombre  $N$ , elles feront apparaître autant de zéros dans la fonction de transfert, qui devient alors:

$$H(p) = C \prod_{n=1}^N \frac{1}{p^2 + b_n p + \omega_n^2} \prod_{i=1}^N (p - Z_i) (p - Z_i^*) \quad (IV.6)$$

Le circuit "antiformant"  $N_0$  du canal de la nasalité crée un zéro dans la fonction de transfert globale de l'appareil ( $N = 1$ ).

-Le canal de bruit:

il est destiné à la production des sons bruités (non voisés) et comporte trois circuits de formants de bruit  $B_1, B_2, B_3$ , ; SON signal d'excitation est fourni par générateur de bruit.

Les sons bruités voisés sont générés par le canal des formants vocaux. Celui-ci est alors attaqué par les deux sources d'excitation en même temps. C'est la commande  $A_1$  qui permet "d'injecter" du bruit dans ce canal.

2°/ Les paramètres de commande :

Le synthétiseur est commandé à l'aide de 11 paramètres :

-les trois fréquences des formants  $F_1$ ,  $F_2$  et  $F_3$ .

-les trois fréquences des formants de bruit  $B_1$ ,  $B_2$ ,  $B_3$ .

-la fréquence du fondamental  $F_0$ .

-les quatre amplitudes  $A_0, A_B, A_N, A_1$  de l'excitation vocale (son voisé), du bruit, de la nasalité et du bruit injecté dans le canal vocal.

Ces paramètres sont à réactualiser toutes les 20 ms environ. Ils définissent l'ensemble des échantillons constituant la séquence de parole continue à synthétiser.

3°/ Inconvénient du synthétiseur à formants en série :

Comme les plages de formants se chevauchent, deux formants successifs peuvent avoir des fréquences très proches l'une de l'autre. A ce moment là, si les largeurs de bandes sont minimales, on atteint des acuités de résonance extrêmement élevées. Ces dernières constituent alors une charge trop importante pour les filtres analogiques. Il en découle une altération de ces derniers se traduisant par une dégradation de la qualité de la parole synthétisée. Heureusement, en pratique, ce risque d'amplification est très rare. De plus, on l'évite en choisissant des bandes passantes les plus larges possible et en implémentant le filtre de formants  $F_{1A}$  entre ceux des formants  $F_2$  et  $F_3$ . Ainsi, on atténue quelque peu le signal du formant  $F_2$  avant qu'il n'atteigne le circuit de  $F_3$ .

On peut donc dire que ce~~y~~ inconvé~~n~~ient est presque sur-  
monté. En outre, ce type de synthétiseur a l'avantage de  
donner une représentation assez exacte de propagation de  
l'onde sonore dans le conduit vocal.

4°/ Correction des formants élevés :

Le synthétiseur à formants permet une  
compression de bande assez appréciable puisqu'on ne consi-  
dère que les trois premiers formants. Pour ceux d'ordre  
supérieur, il est muni d'un circuit de correction. Celle-ci  
est simplifiée par le fait que la bande passante " $B_n$ "  
~~des~~ des formants augmente beaucoup moins vite que la  
fréquence " $\omega_n$ ". Donc, l'acuité ( $b_n / \omega_n$ ) est négligeable  
pour le formant d'ordre n.

L'effet d'un formant supérieur d'ordre n, donne une contri-  
bution (en dB) au spectre vpcal, du type :

$$(-10) \log (1 - \omega^2 / \omega_n^2)^2 + \frac{b_n^2 \cdot \omega^2}{\omega_n^2}$$

soit approximativement :

$$(-10) \log (1 - \frac{\omega^2}{\omega_n^2})^2$$

Après un développement limité, et en admettant que la  
fréquence des formants d'ordre supérieur a pour expres-  
sion approchée :

$$F_n = \frac{\omega_n}{2} = (2n-1) \cdot 500 \text{ Hz} \quad (\text{IV.8})$$

on obtient la contribution de tous les formants d'ordre  
supérieur à 3 en sommant les contributions de tous les  
formants; étant égale à 2 f on a :

$$20 \log e \cdot \sum_{n=4}^{+\infty} \frac{f^2}{(2n-1)^2 500^2}$$



Enfinement :  $20 \log e \cdot f^2/500 \left( \pi^2/8 - \sum_{n=1}^{\infty} 1/(2n-1)^2 \right) = 2,87 \cdot 10^{-6} f^2$   
(en dB)

Ainsi, la fonction sera d'autant plus élevée que le nombre de formants considéré est petit et que la hauteur du spectre correcte désirée est grande.

#### IV.2.3.3.c Les synthétiseurs à formants en parallèle :

Pour ce type de synthétiseurs, les circuits de formants sont disposés en parallèle (fig IV.4). Le dispositif d'excitation permet un choix entre signal périodique, dont on contrôle la fréquence, et le bruit. Tous les filtres sont alimentés par le même signal. Ils renforcent dans des bandes de fréquences différentes.

Dans le synthétiseur à formant en parallèle, les contrôles des différents circuits de formants sont indépendants les uns des autres. Ainsi, on peut simuler "correctement" les effets de l'effort vocal sur le spectre de la source globale; ce qui n'était pas le cas pour les synthétiseurs à formants en série. Le deuxième avantage de ce type de synthétiseurs est qu'ils permettent de reproduire des sons excités en un point quelconque du conduit vocal. Ceci est très important pour la synthèse des consonnes.

Cependant, ce qui permet ce double avantage est lui-même un inconvénient. En effet, le fait de spécifier l'amplitude et la fréquence de chaque formant requiert une quantité d'informations plus accrue que pour les synthétiseurs à formants en série.

#### IV.2.3.4 Comparaison des synthétiseurs à canaux et à formants :

En la faisant dans les domaines de l'analyse et de la synthèse, on aboutit au résultat suivant :

	Synthétiseurs à canaux	Synthétiseurs à formants
Analyse	-elle est simple et l'extraction des paramètres par les filtres d'analyse se fait en temps réel	-l'extraction des formants est assez complexe et ne se fait pas en temps réel
Synthèse	-la parole produite conserve souvent un timbre un peu métallique	-la parole produite a un "naturel" de bonne qualité

L'examen de ce tableau permet de conclure que : les synthétiseurs à canaux l'emportent par la simplicité de l'analyse et ceux à formants par le "naturel" de leur parole synthétique. Toutefois, lorsqu'il n'est pas nécessaire de faire l'analyse en temps réel; les synthétiseurs à canaux perdent dans une large mesure leur avantage. D'ailleurs si les synthétiseurs à formants n'ont pas encore pris le pas sur ceux à canaux, c'est bien à cause de la difficulté d'extraction automatique des formants.

#### IV.3. Les différentes méthodes de synthèse

La production de la parole synthétique, par ordinateur, se fait par assemblage d'éléments phonétiques préenregistrés. Ces derniers ont été analysés et mis en mémoire pour servir de commande au synthétiseur.



Pour la mise en oeuvre de celui-ci, il existe quatre méthodes de synthèse :

- synthèse par phrases
- synthèse par mots
- synthèse par règles
- synthèse par diphonèmes

#### IV.3.1 La synthèse par phrases :

Dans cette méthode, il ne s'agit pas d'une véritable synthèse. On reproduit plutôt des séquences de parole continues enregistrées préalablement. Cet enregistrement peut être analogique ou numérique. Des phrases telles que : "il n'y a pas d'abonné au numéro que vous avez demandé" sont enregistrés dans la mémoire d'un ordinateur. Puis, elles sont restituées en cas de besoin.

#### IV.3.2 La synthèse par mots :

Dans ce cas, la parole synthétique est obtenue à partir de mots préenregistrés et mémorisés sous forme comprimée. L'ensemble de tous les mots constitue le dictionnaire de base du système. Tout message dont les mots existent dans le dictionnaire est réalisé par la juxtaposition de ceci dans un ordre établi par lui-même, avec une intelligibilité très acceptable.

#### IV.3.3 La synthèse combinée : phrases et mots isolés :

Pour ce type de synthèse, le vocabulaire de base comprend :

- une partie fixe (phrase type) telle que :  
"votre compte indique..."
- une partie variable qui peut être un nombre.



L'ensemble donnera : "votre compteur indique N" (N est un nombre spécifié).

Jusque là, les méthodes citées ne peuvent convenir qu'à un vocabulaire restreint (la capacité mémoire est limitée). Pour remédier à cet inconvénient, des méthodes utilisant des éléments minimaux ont été conçues. La synthèse par règles répond à cet objectif.

#### IV.3.4 La synthèse par règles :

Elle consiste, en général, en l'introduction de règles de composition. Ces dernières permettent de définir les transitions d'un son de parole à un autre. Les études effectuées par les chercheurs de laboratoires HASKINS (1950) ont permis de déterminer l'ensemble de règle de transitions. En particulier, les transitions de formants (cf chap III) sont un indice perceptuel fondamental pour l'identification des consonnes.

#### IV.3.5 La synthèse par diphonèmes (diphones)

HARRIS (1953) constate qu'il n'est pas possible de produire la parole continue par simple juxtaposition d'éléments minimaux (équivalent aux phonèmes qui sont une entité abstraite n'ayant pas d'existence acoustique).

Exemple: la juxtaposition de deux sons [ʃ] et [a] enregistrés séparément ne donnera pas le mot: " [ʃ a] " (chat).

Cet état de chose est dû à l'absence des zones de transition possible entre les deux sons. Pour pallier à ce manque, il suffit de choisir des unités minimales contenant les zones stables des sons élémentaires et les zones de transitions spectrales entre phonèmes.

## IV.4 L'Electrical Vocal Analog (E.V.A III)

## IV.4.1 Description générale :

Le synthétiseur E.V.A III se présente sous forme d'un meuble électronique, composé de trois grandes unités, assurant chacune une fonction particulière :

1°/ Un tableau de bord pour assurer la commande (opérationnel manifold chassis)

2°/ Une table munie d'un dispositif de relecture et recevant la feuille de plastique qui porte la représentation paramétrique qui sert de programme à la synthèse (graphic playback unit).

3°/ Enfin, un châssis qui abrite les circuits électroniques reproduisant la voix à partir des informations issues de la représentation paramétrique.

## IV.4.2 Description et rôles des différentes unités :

## 1°/ Le Graphic Playback Unit :

Il se compose essentiellement de  
-a- une feuille de papier ou de celluloid quadrillée en inch et dixième d'inch (39x12 en inch) sur laquelle on représente à l'aide d'une encre conductrice conçue spécialement à cet effet, les dix paramètres suivants constituant les éléments significatifs de la séquence à reproduire :

1) la fréquence du fondamental et ses variations (quatrième ligne en partant du haut du tracé :  $F_0$ -Pitch).

2) les lignes d'évolution des quatre premiers formants ( $F_1, F_2, F_3, F_4$ ) (ligne 2 à 8 en partant du haut du tracé).



3) Les zones de fréquence ou encore bandes de bruit qui définissent les spectres des fricatives (ligne 12 :  $K_1$ )

4) L'amplitude de la voix pour tous les sons voisés.

5) L'amplitude de bruit des phonèmes de la chaîne parlée (troisième ligne :  $A_g$ ). Ces amplitudes sont relatives aux bandes de bruit de  $K_1$  et donc aux variations d'intensité des constrictives ainsi que des occlusives.

6) L'amplitude de bruit pour le système de filtres vocaliques (production de la voix chuchotée). (AH). Cette ligne sert également la production du bruit caractéristique de l'explosion des occlusives chuchotées.

7) L'amplitude des composantes nasales dont les fréquences peuvent être réglées manuellement (ligne 1 : AN).

Le positionnement des différentes lignes ainsi que leur numérotation sont donnés par le tableau ci-après :



Etalonnage du programme de SYNTHESE						
numéro de la ligne à partir du haut	para-mètre	MAXIMUM		MINIMUM		ECHELLE Nbre de cycles ou de volt par 1/10inch
		dist. de la ligne au bas de la feuille	fréq. ou intensité cores-pondante	dist. au bas de la feuille	fréq. ou intensité co-respon-dante	
12	K <sub>1</sub>	1,2	1700cps	0,8	3000cps	1000cps
11	F1	2,0	1000cps	1,0	0cps	100 cps
10	F2	3,5	2500cps	1,5	500 cps	100 cps
9	F3	4,5	3500cps	2,5	1500cps	100 cps
8	F4	5,5	4000cps	3,5	2500cps	100 cps
7	ligne	non	utilisée	-	-	-
6	ligne	non	utilisée	-	-	-
5	AO	7,2	0 dB	6,8	-50 dB	1 volt
4	FO	9,0	400 cps	7,0	0 cps	20 cps
3	AC	9,0	0 dB	8,5	-50 dB	1 volt
2	AH	9,7	0 dB	9,8	-50 dB	1 volt
1	AN	11,4	0 dB	11	-50 dB	1 volt

Ces données qu'il convient de respecter scrupuleusement si on veut obtenir de bons résultats permettent d'étalonner aisément le programme : à chaque ligne correspond une échelle précise graduée en Hz lorsqu'il s'agit de variations de fréquence (ligne K<sub>1</sub>, FO, F1, F2, F3, F4), en dB pour les variations d'intensité (ligne: AO, AC, AH, AN),

-b- le dispositif de relecture (chart reader):

il est formé d'un système mécanique constitué par un chariot (movable carriage) coulissant le long d'un tube d'acier et entraîné par un moteur logé sous la table, au moyen d'un ensemble de transmissions par câble et poulies. Un dispositif électrique composé de deux résistances bobinées suspendues à ce chariot et balayant le programme de synthèse lorsqu'elles sont en mouvement. Chacune de ces résistances <sup>couvre</sup> 5 lignes du programme sur lesquelles elles délivrent une tension variant d'un bout à l'autre du rouleau de 0 à 10 volts (celles d'en bas:  $K_1, F1, F2, F3, F4$ ; pour la première et celles d'un haut:  $A0, F0, AC, AH, AN$ , pour la seconde). Ainsi, chacune de ces résistances en contact avec les cinq lignes conductrices qu'elle couvre pourra être considérée comme un potentiomètre à cinq curseurs.

Chaque ligne du programme prélèvera alors à la manière d'un curseur de potentiomètre la tension qui correspondra à son point de contact avec la bobine. Cette tension sera fonction du niveau d'élevation du tracé conducteur par rapport à sa ligne de référence.

Toutes les tensions ainsi recueillies par les différentes lignes sont acheminées ensuite vers des circuits comparateurs où elles sont confrontées à des tensions de référence données par les f.e.m d'une série de piles (type Mallory Tr. 177-9, 8 volts).

Après cette opération, les tensions obtenues à la sortie des circuits comparateurs pour chaque ligne, ne peuvent prendre que les cinq valeurs entières comprises entre 0 et -4 volts représentant la position de repos et l'excursion maximale.



En résumé, le dispositif de lecture convertit les graphes des dix paramètres en un ensemble de signaux qui actionnent les circuits électroniques équivalents du synthétiseur de la parole.

-les fréquences et les amplitudes des composantes de la chaîne parlée portées en ordonnée sur le tracé sont rendues par les excursions latérales, les déplacements verticaux des lignes conductrices le long des résistances bobinées.

-le déroulement du discours dans le temps porté en abscisse, par le déplacement horizontal du chariot dont la vitesse (qui correspond par conséquent au débit) est normalement de  $2\frac{1}{2}$  inches par seconde et réglable à volonté par la rotation d'un bouton de commande.

2°/ Le châssis et les circuits électroniques équivalents (speech synthesizer)

Le châssis renferme un ensemble de circuits électroniques, qui, commandés par le programme de synthèse ont pour rôle de reconstituer la voix.

Les circuits électroniques sont alimentés par un secteur alternatif (50-60 Hzn, 115 volts) et leur consommation maximum est de cinq ampères.

L'appareil se compose pour l'essentiel :

- d'une alimentation stabilisée
- d'une série de trois circuits imprimés qui en constituent les pièces maîtresses et qui sont les équivalents du modèle humain
- d'une amplification de sortie avec un haut parleur qui transforme les oscillations électriques en vibrations sonores.



\*Le premier des trois circuits sert à l'excitation. Il se compose d'un générateur de bruits et joue pour les deux autres "équivalents" tantôt le rôle des cordes vocales et de la fermeture glottique initiale qui fait entrer en résonance le pavillon pharyngo-buccal lors de la production des phonèmes sonores, tantôt le rôle du souffle qui engendre sur son passage des bruits de friction (production de consonnes sourdes) et parfois même les deux à la fois : c'est le cas par exemple lorsque'il s'agit de produire une fricative sonore pour laquelle la composante laryngée vient se superposer à un spectre de bruits. Le générateur de friction est commandé par la ligne FO ou manuellement par la rotation d'un bouton : la fréquence est susceptible de varier de 20 à 400 Hz.

\*La deuxième carte se compose des circuits électroniques nécessaires à la production des nasales et des spectres de bruits. Le "murmure nasal" du segment occlusif des consonnes est fourni par un équivalent électrique du modèle humain sous la forme d'un spectre comprenant trois résonances et une anti-résonance, dont on peut faire varier manuellement les valeurs au moyen d'une série de commandes.

Les mêmes circuits peuvent servir à donner aux voyelles nasales leurs colorations particulières. L'amplitude des composantes nasales est déterminée dans tous les cas par les variations de la ligne AN.

Les circuits de formation des bruits qui figurent sur la même carte permettent d'engendrer toutes les fricatives au moyen d'une bande de fréquence; assez large pouvant se situer entre 3000 et 7000 Hz (fricative pôle K1), et d'un "creux" (fricative K0) dont la position peut varier entre 750 et 1750 Hz.

Rappelons à ce sujet que les maxima de densité spectrale se situent pour les sifflantes au-dessus de 4000 Hz, dans les régions supérieures du spectre et sensiblement plus bas pour les chuintantes (les deux zones généralement comprises entre 1500 et 4500 Hz). L'amplitude de ces bandes de bruits est déterminée, comme dit, par AC.

\* La troisième carte se compose des circuits nécessaires à la production des sons voisés. Ce "Vocal Analog" permet d'obtenir des spectres constitués de 5 zones formantiques dont les fréquences respectives sont déterminées par les lignes F1, F2, F3, F4 et une commande manuelle F5 dont les valeurs extrêmes sont 3000 et 5000 Hz. Tous les timbres vocaliques peuvent être engendrés. La largeur des différents formants est en moyenne 70 cps pour le premier, de 80 pour le suivant, de 100 cps pour le troisième, de 150 à 400 cps pour les deux derniers. Les amplitudes relatives des diverses composantes sont fixées par le constructeur et non susceptibles d'être modifiées. La carte comporte en plus un circuit KH (Hire Pôle Correction) dont le rôle est de restituer dans une certaine mesure les fréquences supérieures des spectres. Les caractéristiques sont les suivantes :

	Fréquence (Hz)	Largeur de bande
KH	idéale : réelle	idéale : réelle
	4 500 Hz : 4 700 Hz	4 500 Hz : 4 500 Hz

3°/ Le tableau de commande (opérationnel magnifold chassis)

Le "tableau de bord" de l'appareil contient sur le côté gauche toute une série de boutons destinés au réglage du niveau des lignes de référence des différents paramètres, ainsi qu'un indicateur de contrôle qui permet de réaliser les ajustements nécessaires.



Le côté droit comporte par contre tout un ensemble de leviers et de boutons de commande; ils permettent de faire fonctionner l'appareil soit en position automatique, les circuits étant actionnés par les signaux issus des différents paramètres, soit en position manuelle les circuits de synthèse étant commandés dans ce cas par la rotation de toute une série de boutons prévus à cet effet. Le bras de lecture peut être également desenclanché des systèmes de transmission et déplacé à la main.

#### IV.4.3 Principe de fonctionnement :

Le schéma d'ensemble de l'appareil est donné par la figure IV.5. son fonctionnement est aisé à comprendre :

Prenons par exemple le cas d'une voyelle orale. Il faudra pour produire ce phonème disposer d'impulsions à une certaine fréquence; ce rôle incombera à l'oscillateur de relaxation que commande F0. Il suffira ensuite d'ouvrir la porte A0 qui donnera accès aux circuits des voyelles orales lesquels imprimeront au signal une forme qui sera fonction des positions respectives des lignes F1, F2, F3, F4 du tracé et de la position de la commande manuelle F5. Lorsque'il s'agit de produire une voyelle nasale, on ouvre en plus au moyen de la ligne AN la porte qui donne accès aux circuits équivalents du conduit nasal de la deuxième carte. Le "murmure nasal" s'ajoute alors à la structure en formants qui donne à la voyelle son timbre propre et lui confère la coloration particulière des phonèmes qui se prononcent voile abaissé.

Supposons maintenant que nous ayons à reproduire une fricative sourde, la sifflante [s] par exemple. L'énergie sera fournie dans ce cas par le générateur de bruits et l'amplitude du signal sera déterminée par la ligne AC qui commande la porte donnant accès au circuit équivalent de la carte 2.



La délimitation en fréquence de la zone de bruits requise se fera au moyen de la ligne  $K_1$  qui commande ces circuits et qui peut varier entre 3000 et 7000 Hz dans le cas où nous nous sommes placés, cette ligne devra être à son maximum, le spectre du [ ] comportant des fréquences très élevées.

Le processus est le même lorsque'il s'agit d'engendrer une fricative sonore, mais on ajoute dans ce cas aux bruits la composante laryngée qui s'y superpose normalement, elle sera fournie par le générateur d'impulsions à la fréquence  $F_0$  désirée et son amplitude sera déterminée au moyen de la ligne  $A_0$  qui commande la porte donnant accès au "vocal analog". Pour les sourdes, cette porte doit être complètement fermée (position de repos de la ligne). Les composantes harmoniques étant généralement de faible intensité, il suffira de l'ouvrir très légèrement pour la formation des sonores sifflantes et chuintantes, d'avantage par contre pour des phonèmes qui, par leur constitution acoustique sont très proches des voyelles comme c'est le cas des liquides [l] et [R], très largement pour les semi-consonnes (j, y, w) dont les spectres sont très peu chargés de bruits. De l'ouverture de la porte commandée par  $A_0$  dépend en quelque sorte le degré de vocalisation des phonèmes, du signal issu de AC, le degré de friction ou de spirantisation. Les amplitudes respectives de ces signaux peuvent être dosées à volonté.

Les occlusives bilabiales (p; b; n) alvéo-dentales (t, d, n), vélaires (k, g) requièrent pour leur phase audible (explosion) de brusque variation d'énergie. Les "barres de bruits" produits lors de la rupture de l'occlusion par l'échappement très brusque de l'air phonatoire peuvent être rendus à la synthèse à la fois au moyen de la ligne  $AH$  qui commande la porte d'entrée du signal à spectre continu vers le vocal analog et qui

sert également à la production de la voix chuchotée et de la ligne AC qui commande la porte d'entrée du même signal vers les circuits formateurs des spectres des fricatives. Les transitions des formants (F2 en particulier) servent à la perception de leur lieu d'articulation. Les lignes AO et AN déterminent comme précédemment la sonorité et la nasalité des phonèmes.

#### IV.4.4 L'evagramme

La préparation du programme de synthèse (evagramme) se fait en trois temps : le premier consiste à enregistrer et à analyser la séquence, le second est nécessaire à la reproduction sur papier des différents paramètres, le dernier est réservé enfin aux "retouches" et à la mise au point définitive du tracé.

a-enregistrement et analyse, sonographique de la séquence :

La séquence que l'on veut produire artificiellement doit, de préférence, être enregistrée tout d'abord sur bande magnétique. La parole synthétique pourra être comparée ainsi en fin d'opération à la voix humaine et dans l'immédiat cet enregistrement servira à l'analyse sonographique de la séquence.

Le sonographe (cf chap II. ) ne peut analyser de toute manière que des "portions" de discours de deux secondes quatre dixième (2s 4/10); on est donc obligé de procéder par enregistrements successifs et de recoller les tracés obtenus. Ces tracés sont réalisés en filtrage large (300 cps) tout d'abord sur l'échelle normale (85-8000 Hz), puis sur l'échelle dilatée (85-4000 Hz). Sur cette même échelle, on réalise ensuite les sonagrammes en filtrage étroit en y faisant figurer au moyen du dispositif "amplitude display" les variations de l'intensité.



b- analyse des résultats sonographiques.  
et reproduction sonographique sur  
papier des différents paramètres :

Les tracés sonographiques, une fois réalisés, les phonèmes de la séquence à reproduire doivent être soigneusement délimités. Les mêmes divisions qui correspondent à la durée des sons et au déroulement du discours dans le temps sont reportées ensuite sur le programme de la synthèse au moyen de longs traits verticaux. A ces dispositions préliminaires, fait suite alors un travail beaucoup plus long qui consiste en un examen minutieux des tracés fournis par l'analyse sonographique et un recensement des caractères et indices acoustiques des différents phonèmes de la séquence considérée. Au cours de cette opération qui s'accompagne tout normalement d'une réflexion sur la genèse des sons et leur mode de production, les différents phonèmes sont analysés à travers la grille des paramètres qui commandent le synthétiseur de l'appareil. En remplissant les cases correspondants aux différents indices acoustiques des phonèmes, on obtient une "matrice de conversion" en quelque sorte qui permet de passer très aisément par la suite des données de l'analyse sonographique à la représentation paramétrique des sons du langage.

A partir de là, la réalisation du tracé devient assez aisée. Les lignes d'évolution des quatre premiers formants sont décalquées directement des sonagrammes en filtrage large (échelle dilatée-85-4000 Hz) sur un papier transparent ou une feuille de plastique coupée aux dimensions du tracé sonographique et reportés ensuite sur l'évagramme. Le tracé sonographique doit être nécessairement gradué ou étalonné du moins à l'aide d'un générateur de fréquence. L'opération qui vient d'être décrite peut soulever quelques difficultés, dans la mesure où



les transitions de formants n'apparaissent pas toujours avec une entière évidence, dans la mesure aussi où le bas des tracés est généralement trop noir et le haut trop pâle pour qu'on puisse discerner les structures avec une entière certitude. On peut remédier dans une certaine mesure à ces insuffisances en réduisant au moyen du "mark level" l'intensité de la trace lorsque le stylet est au début de sa course et en l'augmentant progressivement au fur et à mesure qu'il se lève.

La ligne mélodique s'obtient en décalquant sur le sonagramme réalisé en filtrage étroit (échelle dilatée, 85-4000 Hz) la ligne d'évolution du cinquième harmonique. Les variations de hauteur qui n'apparaissent pas suffisamment au bas des enregistrements sont ainsi multipliés par 5 et nettement visibles. La ligne qui correspond à la fréquence de 200 Hz sert généralement de référence pour les passages du sonagramme à l'évagramme. Ce tracé doit également être étalonné.

Les lignes AO et AC qui déterminent respectivement l'amplitude de la voix (voyelles et consonnes sonores) et l'amplitude des bruits (fricatives et occlusives sourdes et sonores) s'obtiennent à partir du graphe de l'évolution de l'intensité globale donnée par "l'amplitude display" de l'analyseur.

Les lignes  $K_1$  (fréquences des spectres de bruits AH (aspirations amplitudes) et AN (amplitude des composantes nasales) sont construites en fonction des données de l'analyse sonographique. La ligne AH se superpose en quelque sorte à la ligne AC pour la production de l'explosion des occlusives.

Tous les graphes sont réalisés d'abord au crayon, puis à l'encre spécialement destinée à cet usage. On vérifie, lorsque elle est sèche, au moyen d'un ohmmètre que les différentes lignes sont bien conductrices d'électricité. Le tracé est alors prêt pour la synthèse.

c-"Retouches." et mise au point définitive du tracé:

La parole artificielle produite par le synthétiseur est enregistrée à son tour sur bande magnétique pour être comparée à tout moment à la voix humaine. Elle est analysée aussi au sonographe pour pouvoir être confrontée à l'original.

Lorsqu'un passage ne donne pas entière satisfaction il est possible de l'améliorer. Il suffit de comparer pour cela l'enregistrement sonographique de la séquence artificielle à l'original, de déceler les éventuels écarts et de rectifier en conséquence le tracé des paramètres: l'encre s'enlève très facilement à l'aide d'un bouchon de coton imbibé d'Acétate de butyl.

Il est en effet possible de modifier la valeur des paramètres voire même de les supprimer comme on l'entend et de faire subir à la phrase de toutes les transformations que l'on désire. Il est possible aussi de modifier l'un après l'autre tous les paramètres et de voir dans quelle mesure la phrase demeure intelligible. Par sa très grande maniabilité, l'appareil se prête aux recherches les plus diverses et dans certaines débouche d'ailleurs sur des applications pratiques (prothèse pour sourds, matériel de phono-audiologie, répondeurs automatiques etc...).

Il convient tout autant aux recherches spécialisées de phonétique acoustique qu'aux études relatives à l'intelligibilité de la parole, à la perception ou encore à l'étude des faits d'accents et de mélodie. ... ..

A titre d'exemple d'évagramme, nous présentons ici celui de la phrase:

"Je vous présente mes meilleurs voeux!"



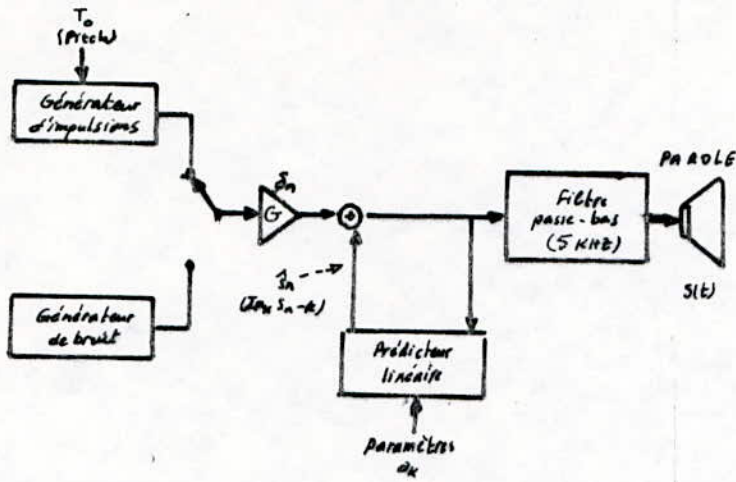


fig IV.1 Synthétiseur à prédiction linéaire.

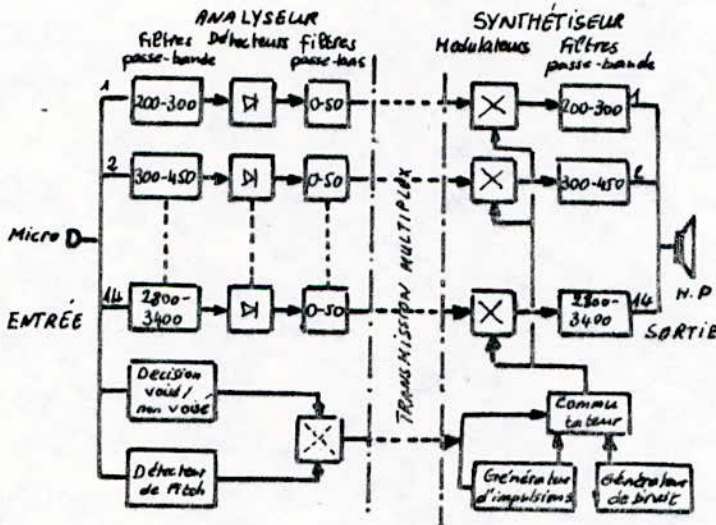


fig IV.2 Schéma d'un vocodeur à canaux.



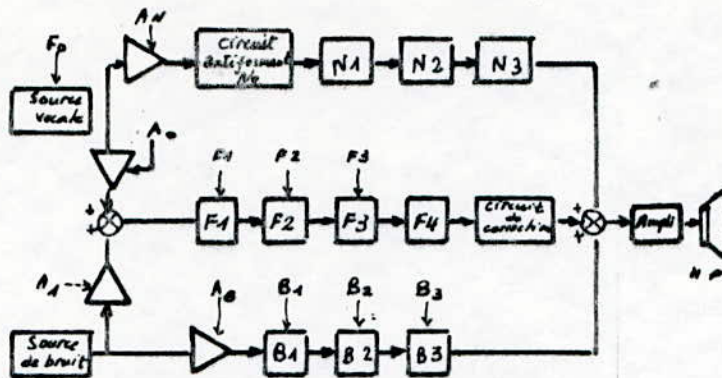


fig IV.3 : Schéma d'un synthétiseur à formants en série

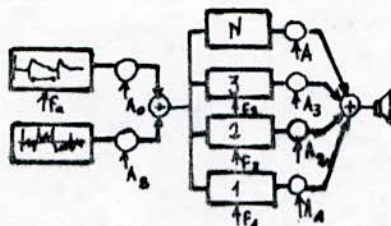


fig IV.4 : Synthétiseur à formants parallèle (d'après Lienard, 1977)

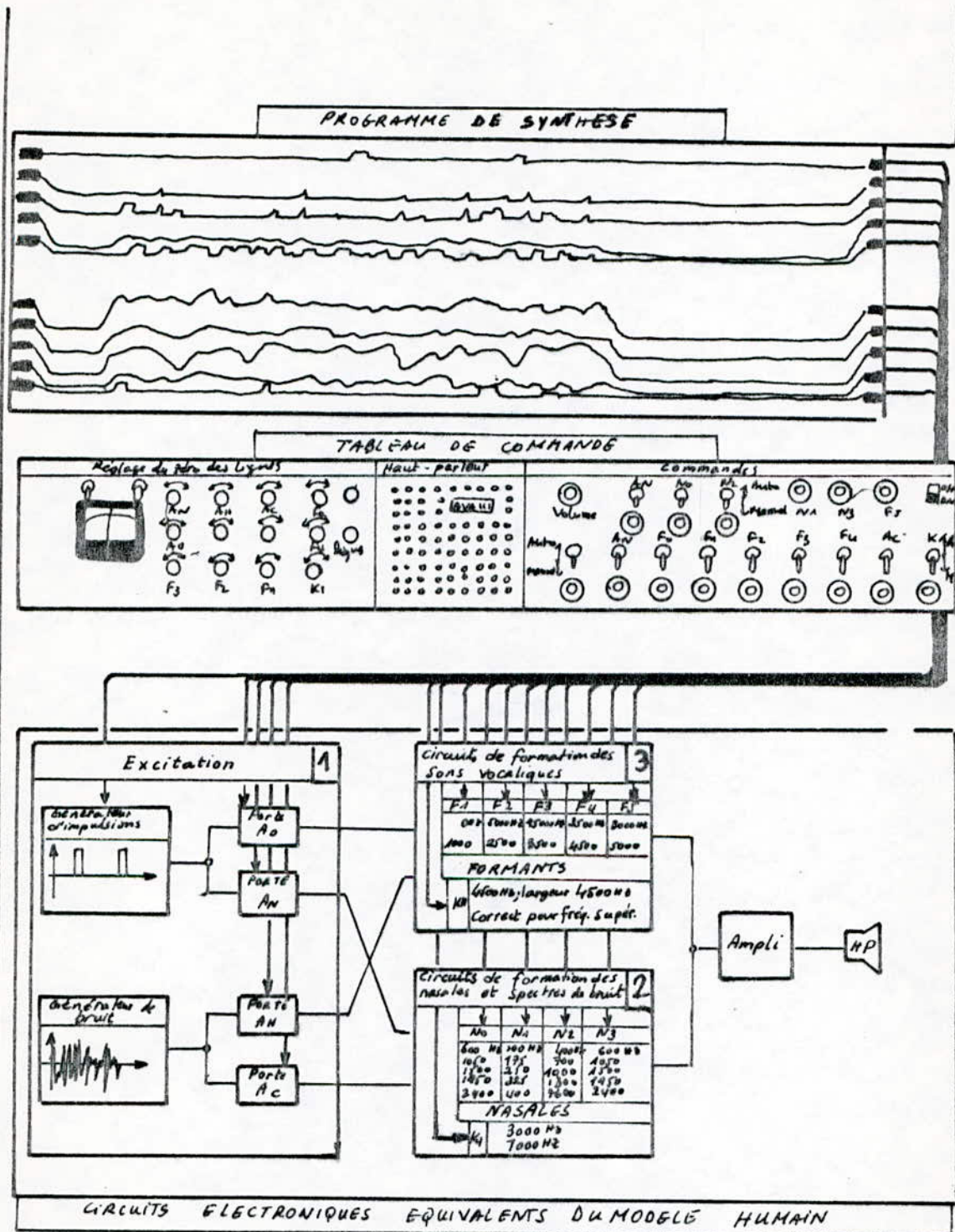


fig IV.5 Le synthétiseur EVA III



## C O N C L U S I O N

Le but de notre travail était d'étudier "la synthèse de la parole à l'aide de l'E.V.A III".

Pour parvenir à cette fin, nous avons traité successivement :

-la production de la parole d'une manière succincte mais exhaustive quant à l'objectif que nous nous sommes fixé.

-les formants qui sont les paramètres fondamentaux, d'un synthétiseur à formants.

-l'analyse de la parole (surtout dans le domaine fréquentiel) et le sonographe qui est l'analyseur utilisé dans ce projet.

-Enfin, la synthèse de la parole et l'E.V.A III qui est l'outil de notre travail.

Cette dernière partie nous a permis de constater que la préparation de l'évagramme demande beaucoup de temps et présente certaines difficultés :

-pour représenter les lignes de variation des différents paramètres, plusieurs analyses sont nécessaires (un sonogramme en bande courte : 45 Hz et deux sonogrammes en large bande : 300 Hz à échelle normale puis dilatée).

-un sonogramme est souvent très complexe. On ne peut pas savoir de prime abord, à quel point le schématiser en vue de la synthèse.

-le tracé de l'évagramme ne donne pas les résultats escomptés dès sa première version. En effet, il est souvent nécessaire de lui apporter certaines corrections. Celles-ci peuvent être effectuées après quelques essais.

Toutes ces complications font que l'E.V.A III est d'un abord difficile, il faut beaucoup de patience et plusieurs tentatives ratées pour arriver à un évagramme de bonne qualité : c'est à dire qui permet une intelligibilité parfaite de la parole synthétique. Hormis ce problème, l'E.V.A III rend de grands services en phonétique et toutes les sciences connexes. Néanmoins, cet appareil<sup>est</sup>, relativement ancien, si on considère l'étonnante rapidité des progrès techniques au cours de ces dernières années. En effet, plusieurs prototypes de synthétiseurs dont le fonctionnement est régi par l'ordinateur, ont été réalisés assez récemment. Actuellement, la synthèse ainsi que la reconnaissance de la parole se font par ordinateur.

Nous souhaitons que des recherches semblables se fassent dans notre pays.

Enfin, nous pouvons affirmer que le travail réalisé dans le cadre de ce projet nous a été très bénéfique. Il nous a permis d'acquérir des connaissances en traitement de la parole et d'améliorer notre organisation au travail.



## BIBLIOGRAPHIE

- Carré, R. (1971), "Contribution aux études sur l'analyse et la synthèse de la parole, rôle et importance déformants", Thèse de Docteur es Sciences physiques.
- Chrétien, L. (1964) "Elements d'acoustique" Chiron.
- Canare?, F. (1983) "Synthèse, reconnaissance de la parole"
- Ferreti, M. ( Editests.
- El-Malawani, J. (\*975) "Contribution au recherches sur la communication parlée, étude du vocodeur à prédiction linéaire, détermination de l'intervalle de fermeture de la glotte, détection de la mélodie. Extraction de la fonction d'aire du conduit vocal." Thèse de Docteur Ingénieur. Université Scientifique et médicale de Grenoble.
- Emerit, E. (1977) "Cours de Phonétique Acoustique "SNED.
- Guerti?, M. (1984) "Contribution à la synthèse de la parole en Arabe standard" Thèse de Magister.
- Guibert, J. (1979) "La parole compression et synthèse par les ordinnateurs". P.U.F.
- Guth, G. (1971) "El-Ellyanyyat" Revue Algérienne de linguistique  
Burg Staehler, D. -que .Volume I N°1. Pages 70 --57.
- Lienard, J.S. (1977) "Les processus de la communication parlée. Introduction à l'analyse et à la synthèse de la parole".
- m.Malemborg, B. ( 1975); "La phonétique "; Que sais-je. P.U.F.
- Max, J. (1981); "Methodes et Techniques de traitement du signal et application aux mesures phonétiques," Tome I, 3<sup>e</sup> Edition MASSON.

Mercier, G. (1977), "Analyse Acoustique et transcription  
phonétique du signal de parole", information  
générale. IG/IAS/SST/1.