

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT

d'électronique

المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

PROJET DE FIN D'ETUDES

S U J E T

**LES PRINCIPALES METHODES
D'ANALYSE DE LA PAROLE**

Proposé par :

M. GUERTI

Etudié par :

**NOUR . F. Z
SELLAM . M**

Dirigé par :

M. GUERTI

PROMOTION : **JUIN 86**

-oOoOo- REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE -oOoOo-

- MINISTERE DE L'ENSEIGNEMENT ET DE LA RECHERCHE SCIENTIFIQUE -

المدرسة الوطنية المتعددة التقنيات
المكتبة — BIBLIOTHEQUE
Ecole Nationale Polytechnique

-o- ECOLE NATIONALE POLYTECHNIQUE D'ALGER -o-

DEPARTEMENT D'ELECTRONIQUE

PROJET DE FIN D'ETUDES
-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-

S U J E T

LES PRINCIPALES METHODES
D'ANALYSE DE LA PAROLE

Proposé par :
Mlle M. GUERTI

Etudié par :
NOUR F.Z
SELLAM M.

Dirigé par :
Mlle M. GUERTI

PROMOTION juin- 86-

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

D E D I C A C E S
-O-O-O-O-O-O-O-O-

A MES PARENTS .

A MES PROFESSEURS QUE J'AI BEAUCOUP ESTIME :

MESSAOUDI Louisa , MOUDJEB Samira , DJEMAA Sara , Guerti
Mhania .

A MES AMIES :

HAMSOUI Leila , HAMAD Faiza , AIT-AMEUR Nylida , DJAFER Traki ,
DJOUHRI Naima , NOUIOUAT Nazek , et M'HARI Sofia .

NOUR F.Z.L
-O-O-O-O-

A MES PARENTS .

A TOUTE MA FAMILLE .

A TOUS (TES) MES AMIS (ES) .

A LA MEMOIRE DE MES GRANGS PARENTS MATERNELS ET PATERNELS .

A LA MEMOIRE DE MON FRERE CHERIF .

A LA MEMOIRE DE MON COUSIN DOMINIQUE .

S. MOURAD .
-O-O-O-O-

R E M E R C I E M E N T S

Nous tenons à remercier vivement notre promoteur **Mlle M. GUERTI** pour toute l'aide qu'elle nous a apportée tout le long de l'élaboration de ce travail .

Nous sommes particulièrement reconnaissants à **Mme L.HAMAMI** et à **Mlle M. GHEZLI** pour l'aide précieuse qu'elle nous ont apportée en programmation .

Les responsables de l'I L P (Institut de Linguistique et de Phonétique) nous ont permis de visiter leur laboratoire , nous tenons à les remercier .

Nous remercions également **Mlle SOUKEUR Radia** , **Mlle T.DJAFER** ainsi que tous ceux qui ont participé de près ou de loin à la réalisation de ce travail .

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

T A B L E D E S M A T I E R E S
=====

INTRODUCTION

CHAPITRE I : PRODUCTION NATURELLE DE LA PAROLE

- I.1 L'appareil phonatoire
- I.2 Etude des sons de la parole .

CHAPITRE II : ANALYSE DE LA PAROLE

- II. 1 Application du traitement du signal au domaine de la parole .
- II. 2 Le filtrage
- II. 3 Les différentes techniques d'analyse de la parole .

CHAPITRE III : LA SONAGRAPHIE

- III 1. Les premiers appareils analyseurs de la parole .
- III 2. La sonographie
- III 3. Le sonographe
- III 4. Utilisation du sonographe
- III 5. Avantages et inconvénients du sonographe .
- III 6. Le sonographe et les autres appareils d'analyse

CHAPITRE IV : LA PREDICTION LINEAIRE

- IV.1. Les différents critères utilisés en prédiction linéaire
- IV.2. Les différentes méthodes de la prédiction linéaire .
- IV.3. Détermination des coefficients de réflexion à l'aide de la méthode du filtre en treillis suivant L_1

:
suite ...

CHAPITRE IV.4. Comparaison des méthodes L_1 et L_2 .

IV.5. Programmation .

CONCLUSION

BIBLIOGRAPHIE

INTRODUCTION

=====

Le développement de l'électronique , de l'enregistrement et des télécommunications , fait que la parole est devenue - ou redevenue - un moyen privilégié de communication entre les hommes . Les récents progrès de celle-ci font que la civilisation évolue sous une forme de plus en plus verbale .

Ceci n'est pas le fruit du hasard , mais l'aboutissement de longues recherches . En effet , l'intérêt porté à la parole n'est pas nouveau . Depuis très longtemps , l'homme s'est intéressé à ce phénomène qui lui permet de s'exprimer , de communiquer , en un mot : de vivre en société .

La communication parlée étant un domaine pluridisciplinaire puisqu'il intéresse aussi bien l'électronique , que l'informatique , la phonétique , la physiologie , la phonologie , etc , le progrès dans ce cas , reste tributaire de l'apport de chaque discipline .

Si des résultats encourageants ont été obtenus en synthèse de la parole , il n'en est pas de même pour la reconnaissance automatique qui soulève encore de grands problèmes .

En tout cas , les résultats obtenus laissent augurer de larges espoirs pour plusieurs applications telles que , l'aide aux handicapés , la commande verbale de machines-outils dans l'industrie , la réalisation de systèmes de " traduction automatique " d'une langue humaine en une autre , etc.

Pour pouvoir effectuer de la synthèse de la parole , ou de la reconnaissance , une analyse préalable est nécessaire .

Depuis un demi-siècle , les techniques d'analyse se sont singulièrement perfectionnées et leur abord nécessite une préparation de type mathématique et physique .

C'est ainsi que diverses méthodes ont été développées . Parmi les plus puissantes on distingue la prédiction linéaire et la sonographie . Celle-ci a été d'un apport remarquable pour le développement des recherches dans le domaine de la parole . En effet , elle a introduit une nouvelle représentation du phénomène sonore : la forme globale , ce qui est plus proche de la réalité physiologique (cf Chap. III) .

La prédiction linéaire est également d'une grande importance Elle est souvent utilisée suivant un critère bien connu , celui des moindres carrés (appelé aussi L_2) .

Pourtant un autre critère - celui de la valeur absolue de l'erreur minimale (dit aussi L_1) - a été décrit bien avant . Ce dernier n'est généralement pas utilisé du fait que les méthodes correspondantes présentent de nombreux inconvénients .

Cependant , si les premiers algorithmes L_1 étaient trop compliqués par rapport à ceux de L_2 , il n'en est plus de même aujourd'hui . En effet , grâce au développement de la programmation , des algorithmes L_1 puissants ont pu être développés . Ainsi un nouvel algorithme alliant la robustesse du critère L_1 à la stabilité de la méthode de Burg a été proposé par E. Denoël de Phillips Research Laboratory de Bruxelles et J.-P. Solvay du Centre de Recherche de Grenoble .

Grâce à ces qualités , il acquiert des performances comparables à celles des méthodes L_2 classiques .

La présente étude comporte quatre chapitres :

- Le premier consiste en une brève étude de la " production naturelle de la parole " ainsi que celle des sons du langage .
- Le deuxième chapitre aborde d'une façon générale les méthodes d'analyse .
- Le troisième traite des différents appareils d'analyse de la parole et principalement du sonagraphe .
- Le dernier chapitre est consacré à la prédiction linéaire et plus particulièrement à son utilisation suivant le critère de la valeur absolue de l'erreur minimale , ce qui est illustré par la détermination de coefficients de corrélation partielle par l'algorithme cité précédemment . Pour cela , nous avons écrit un programme donné en fin de chapitre . Ce programme a été réalisé sur mini-ordinateur VAX .

Dénominations	Symbole phonétique	Mot-clé	Observations	
Consonnes fricatives : (constrictives)	[f]	fameux	« sourdes », ou non-voisées	
	[s]	saucisson		
	[ʃ]	chat		
	Consonnes nasales : (occlusives nasales)	[v]	yert	« sonores », ou voisées
		[z]	zèbre	
		[ʒ]	janvier	
Consonnes nasales : (occlusives nasales)	[m]	menthe		
	[n]	Nantes		
Consonnes liquides :	[ɲ]	agneau		
	[l]	saïon	souvent appelée « latérale »	
Consonnes liquides :	[r]	bugeau	nombreuses variantes en français ; les principales sont notées [r] , [R] , [ʁ]	
Consonnes plosives : (occlusives orales)	[p]	pari	« sourdes », ou non-voisées	
	[t]	baïcau		
	[k]	égart		
	Consonnes plosives : (occlusives orales)	[b]	barbare	« sonores », ou voisées
		[d]	baïaud	
		[g]	langue	
Voyelles orales :	[i]	lit		
	[e]	été		
	[ɛ]	marais		
	[y]	l'ursule		
	[œ]	peur	son voisin de la voyelle neutre théorique [ɜ]	
	[ɐ]	petit	son voisin de [œ], mais souvent plus court, ou à prononciation facultative.	
	Voyelles orales :	[ɛ]	jeu	Cette distinction tend à disparaître au profit d'un A moyen
		[a]	patte	
		[ɑ]	pâte	
		[ɔ]	soi	
		[o]	sauie	
		[u]	bijou	
Voyelles nasales :	[ɛ̃]	brin	Cette distinction tend à disparaître, notamment à Paris au bénéfice du seul [ɛ̃]	
	[ɔ̃]	brun		
	[ɑ̃]	chant		
	[ɔ̃]	bonjour		
Semi-voyelles :	[j]	paille	appelées quelquefois semi-consonnes	
	[w]	lui		
	[w]	Louis		

Figure 1.8 : Classification des sons du français

(d'après LIENARD)

CHAPITRE I

PRODUCTION NATURELLE DE LA PAROLE

Le phénomène de la production de la parole suscite aujourd'hui, plus que jamais, l'intérêt des chercheurs. En effet, une meilleure connaissance de ce phénomène est devenue nécessaire pour pouvoir avancer dans les secteurs de la synthèse et de la reconnaissance de la parole. Une telle connaissance exige l'étude de plus en plus approfondie de l'appareil phonatoire, des caractéristiques acoustiques du signal de la parole ainsi que celle des différents sons du langage.

I.1. L'APPAREIL PHONATOIRE .

I.1.1. ANATOMIE

L'appareil phonatoire (fig. I.1) est essentiellement constitué des organes suivants :

- +) Les poumons qui constituent les principaux organes de l'appareil phonatoire et qui jouent ici le rôle de générateur d'air .
- +) La trachée artère qui comporte à son extrémité supérieure le larynx . Celui-ci supporte deux muscles : " les cordes vocales " séparées par un espace appelé " glotte " et qui varie au rythme de leurs vibrations .
- +) Le conduit vocal qui se compose de deux parties :
 - 1) La partie orale qui est l'ensemble des cavités pharyngale et buccale , de forme et de volume extrêmement variables .
 - 2) La partie nasale qui se compose de deux cavités fixes : " les fosses nasales " .

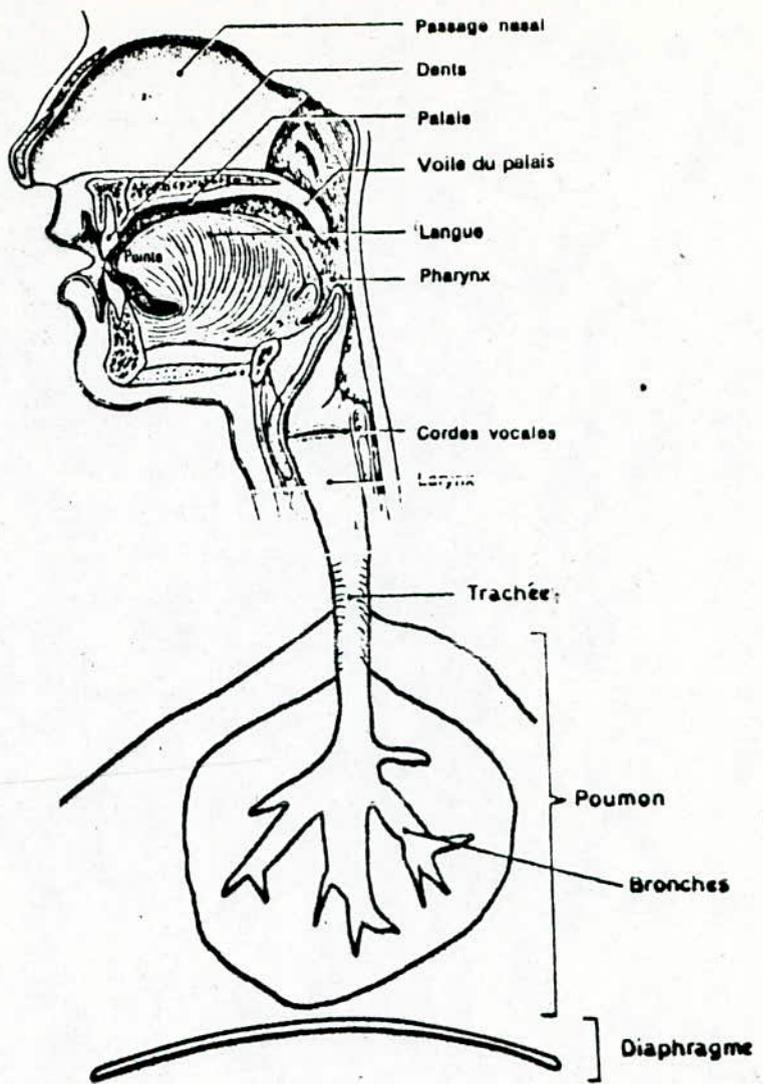


FIG. (I.1) : appareil phonatoire

Les deux parties orale et nasale sont connectées par l'intermédiaire d'un muscle mobile appelé (voile du palais) ou **velum** , qui permet le couplage acoustique entre elles (fig.I.2.

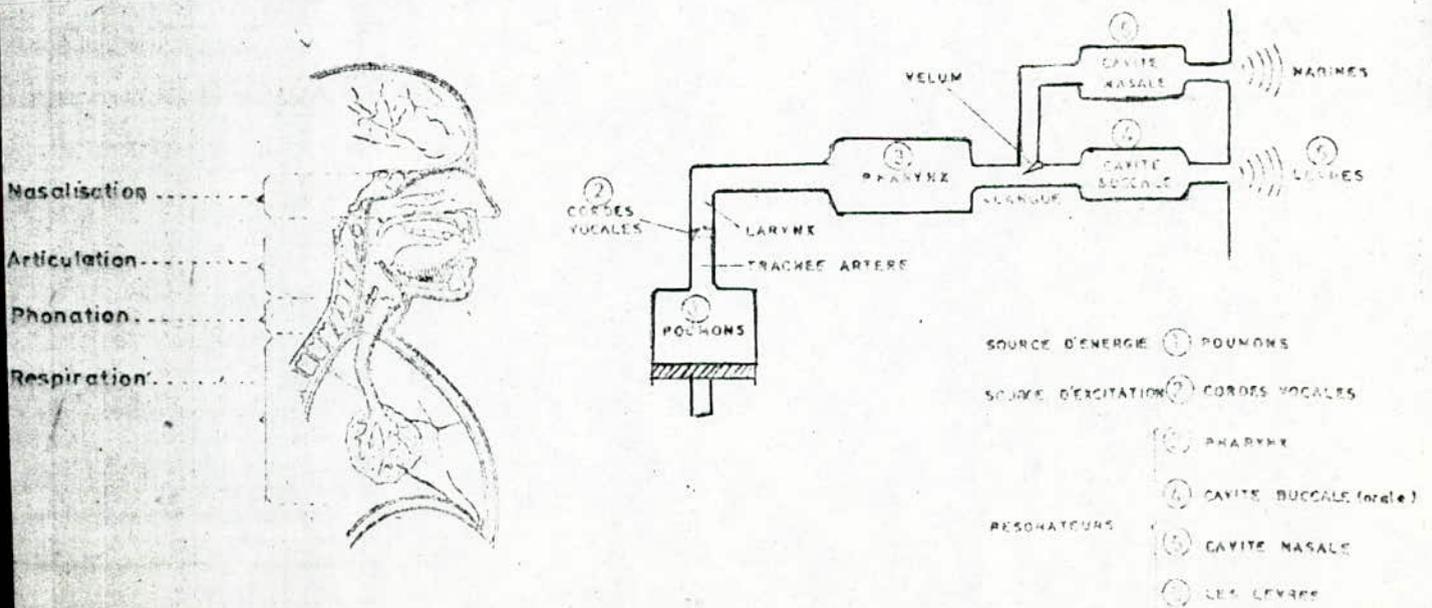


FIG. (I. 2)

FIG. (I.2) L'appareil phonatoire en tant que système acoustique .

I.1.2. PRINCIPE DE FONCTIONNEMENT

Le processus de la production de la parole par le système phonatoire peut être décrit de la manière suivante :

*) EMISSION DU SON AU NIVEAU DU LARYNX .

.....

Si pour la respiration , la glotte est ouverte , il n'en est pas de même pendant la phonation .

En effet , à la fin de l'inspiration , les cordes vocales s'accolent et affrontent leur bord interne sur toute leur longueur . Ceci constitue un obstacle , qui s'oppose au passage de l'air expiratoire , créant ainsi une pression sous-glottique (ou intra-trachéale) (fig. I.3) .

Lorsque celle-ci excède la tension élastique des cordes vocales , ces dernières s'écartent , laissant ainsi passer un flux d'air . Ceci abaisse la pression intra-trachéale et permet à l'élasticité des cordes de rétablir la position initiale en fermeture - de la glotte .

L'air pulmonaire est refoulé par le diaphragme de façon continue . Ceci permet d'accroître de nouveau la pression sous-glottique et le cycle décrit recommence alors . Ainsi prend naissance une vibration sonore d'origine laryngée , constituant un signal périodique , dont la période correspond au fondamental " FO " (ou pitch) .

Lorsque la partie inférieure de la glotte est ouverte en laissant passer l'air , on obtient alors une voix chuchotée et dans ce cas , la vibration des cordes vocales n'a pas lieu . Elle est remplacée par un bruit dû au flux d'air venant de la glotte (fig. I.3) .

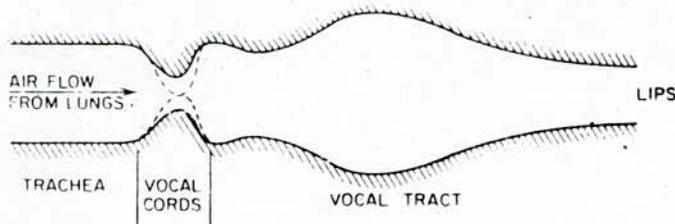


Fig. 3.28 Schematic representation of the vocal system.

FIG. (I.3)

+) LES RESONATEURS .

.....

: Le son laryngé produit va se propager dans les différentes zones supraglottiques , constituées par les cavités pharyngale , nasale , et labiale , appelées résonateurs . Ce son étant riche en harmoniques , il pourra être filtré suivant les résonances propres des différentes cavités .

Ainsi , le pavillon buccal avec les lèvres permet diverses modulations . Avec l'adjonction de la cavité nasale , on obtient la nasalisation (quand le velum est abaissé) . Mais c'est surtout la place de la langue par rapport aux parois de la bouche qui donnera les plus grandes variétés aux sons à émettre .

I.2. ETUDE DES SONS DE LA PAROLE .

I.2.1. DEFINITION DE LA PAROLE

Le son est une vibration complexe de l'air (ou de tout autre milieu élastique) .

La parole est constituée par des assemblages de sons complexes , combinés de manière à transmettre un message . On distingue :

+) Les sons voisés :

.....

Dans ce cas , la vibration contient dans son mouvement , le mélange d'un certain nombre de vibrations simples (sinusoïdales) dites harmoniques dont les fréquences d'oscillation sont des multiples entiers de la fréquence de répétition (appelée fréquence fondamentale) de la vibration complexe .

+) Les sons non voisés ou bruit :

.....

Celui-ci est présent dans la voix chuchotée et les consonnes et se caractérise par le fait qu'il ne comporte pas d'harmoniques mais seulement des vibrations partielles (dites partiels) de fréquences quelconques .

I.2.2. CARACTERISTIQUES ACOUSTIQUES DE LA PAROLE .

Le signal de la parole est caractérisé par sa fréquence fondamentale " FO " .

Elle est de 100 à 150 HZ en moyenne , pour les hommes , de 200 à 300 HZ pour les femmes , et au-delà , pour les enfants (MALMBERG , B. , année 1979) .

+) L'analyse acoustique de la parole permet de déterminer les fréquences et les amplitudes des harmoniques (ou partiels) ainsi que le timbre du son .

+) Timbre du son :

On appelle ainsi , la qualité globale d'un son , qui le distingue d'un autre , de même fréquence fondamentale et de même intensité .

Le timbre dépend de la stabilité plus ou moins grande de la fréquence fondamentale . Par ailleurs , il est lié à la richesse en harmoniques donc à la position des formants dans l'échelle des fréquences .

+) Notion de formant .

Celle-ci est souvent floue . Pourtant , les formants sont à l'heure actuelle , le seul élément qui permet de caractériser les voyelles .

On peut définir les formants comme étant les maxima de la courbe de réponse en fréquence du conduit vocal (fig. I.4) .

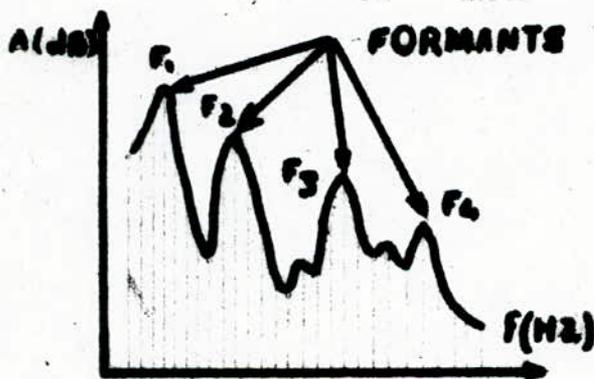
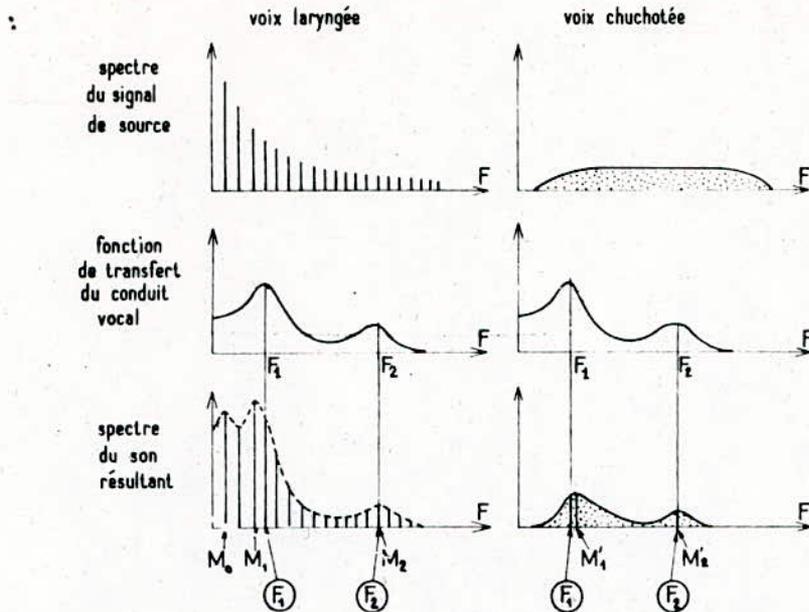


Fig. (I.4) : spectre vocalique présentant quatre resonances formantiques .

On peut noter qu'ils ne correspondent pas toujours aux maxima du spectre de la parole . Ceci est montré par la figure (I.5) où l'on a représenté deux exemple : l'un pour

un signal laryngé , donc quasi-périodique et l'autre pour un bruit (voix chuchotée) .



Formation du spectre d'une voyelle. Le conduit vocal a le même profil, donc la même fonction de transfert dans les deux cas. La source est soit un signal périodique provenant des cordes vocales (voix laryngée), soit un bruit faible à large bande provenant de l'écoulement de l'air à travers la glotte (voix chuchotée). Les maximums du spectre final ne coïncident pas exactement avec les formants.

fig. (I. 5) formation du spectre d'une voyelle .

+) Sons et formants .

.Sons à un formant :

De tels sons , qui n'ont qu'une bande de fréquence dominante , ne recouvrent qu'une gamme restreinte de sons naturels .

On peut citer comme **exemples**:

- le cas où le formant est grave et où l'on a alors une impression de ronflement .
- celui où il est aigu et c'est alors un sifflement .

.Sons à deux formants :

ceux-ci sont très répandus dans la nature . C'est le cas des voyelles en particulier .

+) Théorie formantique des timbres .

C'est la généralité des sons à deux formants qui a conduit à émettre cette théorie . Celle-ci est basée sur les deux hypothèses suivantes :

- On admet que les sons ont deux zones formantiques caractéristiques .
- Le timbre dépend de l'emplacement et des zones dans l'échelle des fréquences .

Cette théorie a été fructueuse . On a pu ainsi constater, en dessinant le diagramme formantique des voyelles (en abscisse la fréquence moyenne du formant 1 , en ordonnée , celle du formant 2) que les points représentatifs de toutes les voyelles possibles se groupaient le long des deux côtes d'un vague triangle dit triangle de Delattre (fig. I.6) .

Le timbre particulier des voyelles est en général déterminé par deux formants : le premier se situe entre 200 et 700 hz , le deuxième , entre 700 et 2500 hz (CUREAU, J. et VULETIC, B. , année 1976) .

Le premier formant indique l'aperture (i.e. ouverture de la bouche) : plus il est haut , plus le son est plus ouvert . Le second donne la hauteur du son (équivalent psycho-acoustique de la grandeur physique : fréquence) : plus ce formant est élevé , plus le son est aigu .

Cette théorie est cependant insuffisante puisqu'elle néglige les sons à plus de deux formants , qui restent mal étudiés .

Par ailleurs , une classification des consonnes d'après les mêmes critères acoustiques ne peut être aussi systématique que pour les voyelles . Les consonnes présentent elles aussi des fréquences plus fortes que les autres , cependant , elles sont mieux caractérisées par les transitions des formants que par ceux-ci eux-mêmes .

I.2.3. CRITERES DE CLASSIFICATION DES SONS DE LA PAROLE .

Les principaux critères sont les suivants :

I.2.3.1. L'opposition sourde-sonore .

.....

Selon que les cordes vocales vibrent (i. e. sont actives) ou non , lors de la production d'un son , on parle dans un cas de son sonore (ou voisé ou encore vocalisé) et dans l'autre de son sourd (ou non voisé) (fig.I.7 et I.8) .

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

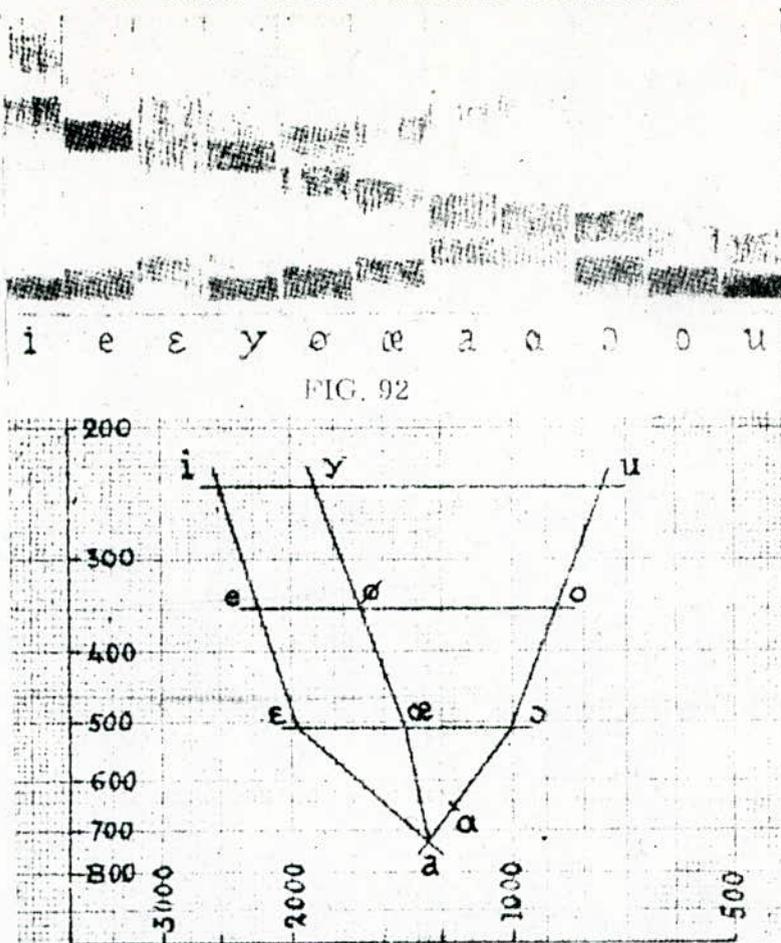


FIG. (I.6) Le triangle vocalique de P. DELATTE .

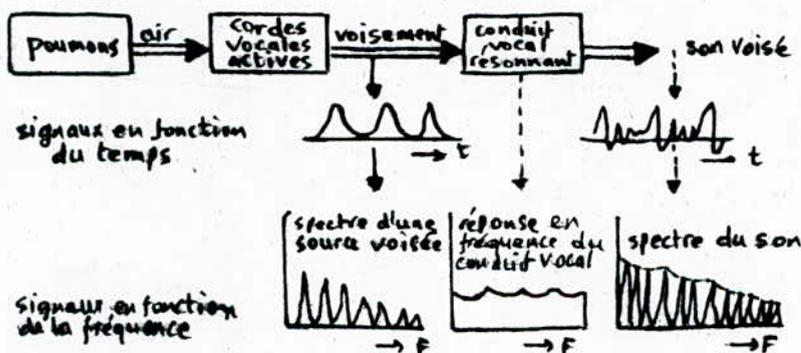


FIG. (I. 7) production d'un son voisé.

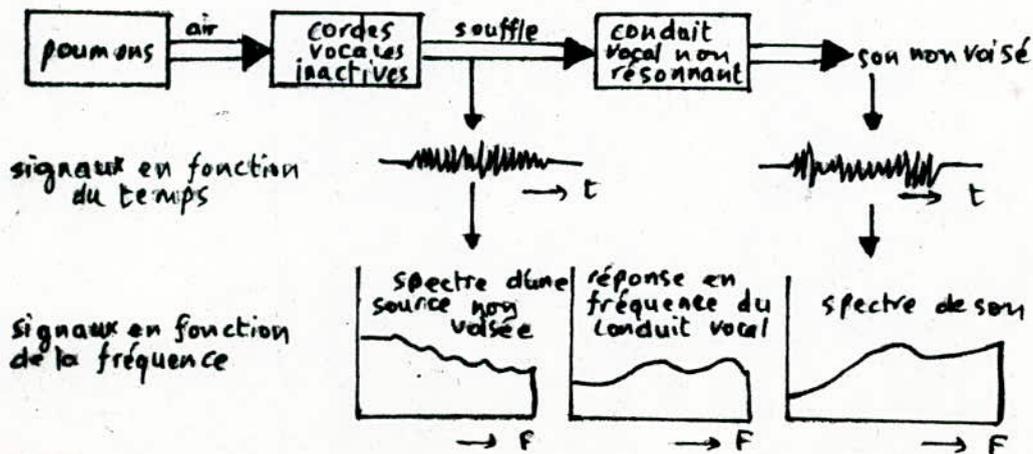


FIG. (I.8) production d'un son non voisé.

I.2.3.2. L'opposition orale-nasale .
.....

Si le voile du palais est abaissé , on parle de sons **oraux** , si ce n'est pas le cas , il s'agit alors de sons **nasaux** .

I.2.3.3. Le lieu d'articulation .
.....

Ce point représente l'endroit de la constriction maximale du conduit vocal (i. e. le lieu où le plus grand rapprochement des organes articulatoires est fait) . Nous distinguons ainsi les sons bilabiaux , palataux , etc (fig.I.10) .

I.2.3.4. Le mode d'articulation .
.....

Ce critère tient compte de l'aperture . On distingue dans ce cas : les sons ouverts (voyelles) et les sons fermés (consonnes) ou de façon plus précise : les voyelles ouvertes , semi-ouvertes , semi-fermées et fermées et parmi les consonnes : les semi-consonnes (dites aussi semi-voyelles) , liquides , nasales , fricatives , affriquées , occlusives (fig.I. 6 et fig. I.9) .

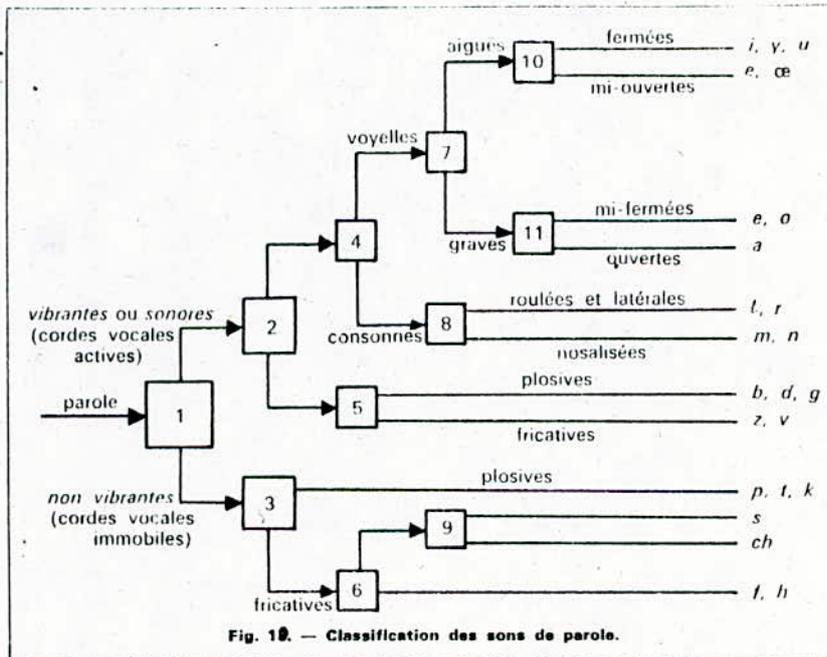


Fig. 10. - Classification des sons de parole.

FIG; (I.9) classification des sons de parole .

I.2.4. LES DIFFERENTS TYPES DE SONS DU LANGAGE .

Les sons de la parole peuvent être classés (fig.I.9 et fig.I. 10) d'après les critères précédents en deux grandes classes :

I.2.4.1. Les voyelles .

.....

Elles se caractérisent par :

- Un passage libre de l'air .
- la vibration des cordes vocales et l'absence de bruit audible .

On distingue :

- Les voyelles orales ex : [a] , [u] .
- Les voyelles nasales ex : [ã] , [õ] .

I.2.4.2. Les consonnes .

.....

Les consonnes sont caractérisées par une constriction ou une fermeture (soit momentanée , soit complète) du passage de l'air et par la présence de bruit .

On distingue :

- Les occlusives .

Elles sont dites aussi " plosives " (ou encore momentanées) . Elles sont caractérisées par le fait que la phase la plus importante de leur formation est l'occlusion momentanée du passage de l'air (occlusion complète suivie d'une ouverture brusque) .

- Les fricatives (ou constrictives) .

Ces consonnes appelées aussi spirantes , sont caractérisées par un retrécissement du passage de l'air qui produit un bruit de friction en passant par la mince ouverture formée par l'organe articulant .

- Les liquides et les semi-voyelles .

Leurs caractéristiques se rapprochent des voyelles orales .

- Les consonnes nasales .

Elles sont caractérisées par le rôle privilégié du conduit nasal dans leur formation .

- Les affriquées .

Elles sont dites aussi mi-occlusives . Ce sont des occlusives dentales suivies de façon très immédiate par un son sifflant ou chuintant très furtif . ex : [ts] , [tʃ] .

CONSONNES

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

CHAPITRE II

LES DIFFERENTES TECHNIQUES D'ANALYSE

DE LA PAROLE

La parole est un signal très redondant . C'est ainsi qu'il véhicule non seulement l'information propre au message à transmettre mais également celle relative au locuteur (timbre , hauteur de la voix , ...) .

Afin de pouvoir faire de la synthèse ou de la reconnaissance de la parole , il est nécessaire d'en faire l'analyse . Celle-ci consiste en l'extraction , à partir du signal de la parole , des paramètres pertinents (fréquence fondamentale , formants , ...) qui le caractérisent au mieux .

Cette opération permet , tout en conservant l'intelligibilité du message , de diminuer la redondance du signal , ce qui présente l'intérêt de diminuer la durée et le coût de son traitement . Par ailleurs , elle fait appel de plus en plus à des systèmes numériques qui ont tendance à remplacer ou à compléter les systèmes analogiques , puisqu'ils présentent l'avantage d'être très rapides et de permettre l'exécution de fonctions diverses sans nécessiter de modification de composant mais seulement celles de programme .

Plusieurs méthodes analogiques et numériques d'analyse de la parole , ont été décrites . On les classe habituellement en deux grandes classes :

- méthodes spectrales .
- méthodes temporelles .

II -1- Application du traitement du signal au domaine de la parole.

Les diverses méthodes d'analyse de la parole font largement appel aux notions de traitement du signal , telles que l'analyse de Fourier , le fenêtrage , le filtrage , ...

L'application de ces " outils " au signal de la parole souffre cependant de certaines limitations . Celles-ci sont dues à la non-stationnarité et à la pseudo-périodicité du signal considéré .

En effet , la phonation , même dans le cas de sons voisés , n'est que pseudo-périodique . De plus , les changements de l'articulation au cours du temps font que le signal de la parole ne soit pas stationnaire .

Toutefois , les phénomènes précédents sont à évolution lente , ce qui permet d'appliquer les lois du traitement du signal sur des intervalles de temps assez courts , de l'ordre de 10 à 25 ms .

II-2- Le filtrage .

Etant donné l'importance de l'opération de filtrage dans les méthodes d'analyse , il est utile d'en rappeler ici quelques notions principales .

II-2.1- Introduction .

La transmission intégrale d'une information est très onéreuse . Aussi , est-il toujours intéressant de chercher à extraire la partie pertinente de l'information et à éliminer la partie non désirable (bruit) . Ceci peut être réalisé grâce à un circuit appelé filtre .

II-2.2- Définition .

Etant donné un phénomène physique dont le spectre s'étend sur un intervalle de fréquence quelconque , le filtre a pour rôle de renforcer certaines composantes du signal tout en affaiblissant les autres .

II-2.3- Les filtres analogiques .

Dans le domaine analogique , le rôle d'un filtre est de sélectionner parmi toutes les composantes spectrales du signal considéré , celles qui remplissent les conditions imposées :

- fréquence inférieure à une limite " f_c " , pour les " passe - bas " .
- fréquence supérieure à une limite " f_c " , pour les " passe-haut " .
- fréquence comprise dans un certain intervalle " $f_c \pm Df$ " pour les " passe-bande " .

Ces différents filtres sont caractérisés par la courbe représentant les variations de la fonction de transfert " T " en fonction de la fréquence " f " , appelée courbe de réponse du filtre (fig. II.1) .

II. 4. Les filtres numériques .

Un filtre numérique est un système numérique qui , recevant à son entrée , les échantillons quantifiés $\{x(n)\}$ d'un signal $x(t)$ délivre à sa sortie une suite de nombres $\{y(n)\}$

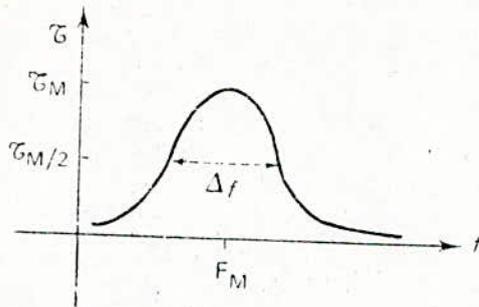


FIG. II.1 Courbe de réponse d'un filtre. Il s'agit ici d'un filtre passe-bande. f est la fréquence, σ la transmittance, Δf la bande passante.

Fig. II.1

On peut donc le représenter par un algorithme .

A partir des nombres représentant les amplitudes des échantillons successifs $x(n)$, on obtient dans le cas des filtres linéaires, la relation suivante, entre les séquences

$$y(n) = \sum_{k=0}^M a_k y(n-k) + \sum_{k=0}^M b_k x(n-k) \quad (\text{II.1})$$

où " M " est entier

et a_k et b_k : les coefficients servant à la "construction" du filtre numérique .

Les filtres de ce type qui font intervenir les valeurs filtrées antérieures (i. e. filtres pour lesquels, au moins, l'un des coefficients " a_k " est non nul) sont dits filtres récurrents .

On appelle fonction de transfert d'un filtre numérique, la quantité définie par :

$$H(z) = \frac{Y(z)}{X(z)} = \frac{N(z)}{D(z)} \quad (\text{II.2})$$

Les racines de l'équation $N(z) = 0$ constituent les zéros de la fonction de transfert et celles de $D(z) = 0$ en sont les pôles .

Les filtres numériques jouent un rôle très important dans le traitement des signaux, plus particulièrement dans le domaine de la parole .

Ils ont été développés dans le but de simuler les filtres analogiques (qu'ils remplacent dans bien des cas) sur calcul-

lateur . Par ailleurs , ils présentent de nombreux avantages par rapport à ces derniers . On peut citer :

- plus grande précision .
- souplesse d'utilisation .
- très grande variété .

Les applications des filtres numériques englobent des domaines divers : parole , télécommunications , etc...

II.3 - Les différentes méthodes d'analyse de la parole .

II.3.1. Analyse spectrale .

L'objectif de ce type d'analyse est la connaissance de la composition spectrale du signal de la parole .

Les méthodes utilisées dans ce cas peuvent être regroupées en :

- méthodes analogiques (analyse instrumentale) .
- méthodes numériques (analyse numérique) .

II.3.1.1. Analyse analogique .

Les méthodes utilisées ici sont essentiellement l'analyse par la transformée de Fourier (T. F) et l'analyse instrumentale .

A- Analyse de Fourier .

La T.F. pour un signal $s(t)$ est donnée par la relation suivante :

$$S(\omega) = \int_{-\infty}^{+\infty} s(t) \cdot e^{-j\omega t} dt . \quad (\text{II} . 3)$$

De même que les autres méthodes du traitement du signal , l'analyse par T.F. ne peut être appliquée qu'à des " tranches " successives du signal de la parole .

C'est pour cela que l'on désigne également cette méthode par l'appellation " analyse spectrale à court-terme " .

On obtient ainsi une suite de spectres dits " instantanés " . L'ensemble de ceux-ci représente le spectre " évolutif " du signal .

Parmi les systèmes physiques qui permettent la réalisation de l'analyse spectrale à court-terme du signal de la parole , on distingue :

- les analyseurs à filtres , tels que le spectrographe

- un équipement spécialisé , appelé :
transformateur de Fourier .

B- Analyse instrumentale .

Dans cette analyse , on distingue :

a- Les analyseurs à plusieurs filtres en parallèle .

Ce type d'analyseur présente l'avantage de permettre une analyse pratiquement instantanée du signal de la parole . Ceci est possible , puisqu'on peut dans ce cas , appliquer le signal $s(t)$ simultanément à un grand nombre de filtres fixes en fréquences et montés en parallèle .

Chacun d'eux analysera la bande de fréquence correspondant à sa propre bande passante .

Parmi ces systèmes , on peut citer l'analyseur du vocoder à canaux , constitué d'un banc de 10 filtres couvrant la bande de 0 à 3000 HZ .

b- analyseur à filtre unique .

Le principe de base de ce genre de système est l'exploration de toute l'étendue du spectre utile du signal de la parole , à l'aide d'un filtre unique .

L'exemple-type de cette catégorie d'analyseurs est le sonographe (cf chap. III) constituant l'outil d'analyse du synthétiseur à formants .

II-3.1.2. Analyse numérique .

La numérisation du signal et sa mise dans la mémoire d'ordinateur (fig. II.2) faciliteront considérablement son traitement ultérieur .

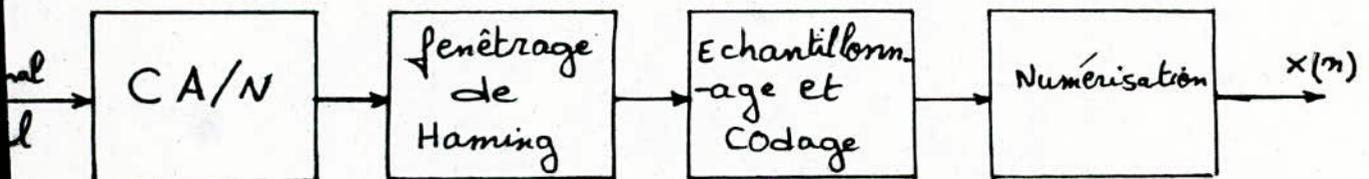


Fig. II. 2.

Le signal vocal est d'abord converti en signal numérique. Ensuite, on lui applique le fenêtrage (fenêtre de Hamming) afin de pouvoir le considérer comme étant stationnaire. Enfin, il doit être échantillonné puis codé avant sa mise en mémoire.

Parmi les méthodes d'analyse numérique, on peut citer celle qui repose sur la transformée de Fourier discrète (T.F.D.) et la méthode cepstrale.

A- Transformée de Fourier discrète .

L'analyse de Fourier vue précédemment peut être réalisée sur ordinateur grâce à la notion de T.F.D.

L'objectif de cette méthode est la recherche d'une approximation de la transformée de Fourier (T.F.) qui pourrait être traitée par ordinateur. Ceci peut être réalisé dans des délais très courts grâce à l'algorithme appelé transformation de Fourier rapide (F.F.T.).

La T.F.D. d'une séquence de valeurs $x(n)$ peut être définie comme étant une suite de valeurs telles que :

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j 2\pi n k / N}, \quad k = 0, 1, \dots, N-1 \quad \text{II.4}$$

avec " N " période du signal.

Elle peut aussi s'écrire sous forme matricielle :

$$\begin{bmatrix} X_0 \\ \vdots \\ X_{N-1} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & W & \dots & W^{N-1} \\ \vdots & W^2 & \dots & \vdots \\ 1 & W^{N-1} & \dots & W^{(N-1)^2} \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{N-1} \end{bmatrix} \quad (\text{II.5})$$

où $W = \exp(-j 2\pi / N)$

B- Analyse cepstrale .

Soit $|s(f)|$ le spectre en amplitude d'un signal de parole et $|s(f)|^2$ son spectre de puissance.

Pour un son voisé, le spectre de puissance peut se mettre sous la forme :

$$|S(f)|^2 = |G(f)|^2 \times |C(f)|^2 \quad (\text{II.6})$$

où $|G(f)|$ est le spectre de la source glottale et la réponse en fréquence du conduit vocal.

En prenant le logarithme de l'expression précédente , il vient :

$$\begin{aligned} \log |s(f)|^2 &= \log [|G(f)|^2 \times |C(f)|^2] \\ &= \log [|G(f)|^2] + \log [|C(f)|^2] . \quad (\text{II.7}) \end{aligned}$$

En lui appliquant la T.F. , on obtient

$$\text{T.F.} [\log |s(f)|^2] = \text{T.F.} [\log |G(f)|^2] + \text{T.F.} [\log |C(f)|^2] . \quad (\text{II.8})$$

Enfin , en élevant au carré l'expression

$$\text{T.F.} [\log |s(f)|^2] ;$$

nous obtenons le cepstre ou cepstrum . Celui-ci peut donc être défini comme étant le spectre de puissance du logarithme du spectre de puissance . (fig. II.3)

En pratique, on utilise une approximation pour éviter le calcul intégral : on calcule l'inverse de la transformée de Fourier discrète (TFDI) du logarithme de la TFD du signal d'entrée

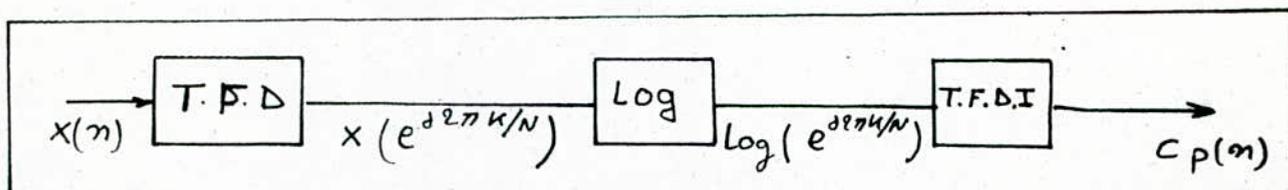


Fig II.3

Le cepstre est une fonction d'une variable appelée quérence , ayant la dimension d'un temps .

Le principe de la méthode cepstrale , comme l'indique l'expression (II.7) est de séparer les effets de la source d'excitation et la réponse du conduit vocal .

Cette séparation est bien effectuée , ce qui représente l'avantage de cette méthode , utilisée en reconnaissance de la parole , pour la détermination du pitch notamment .

Cependant , elle reste peu employée , du fait que le volume des calculs nécessaires est trop important .

II-3.2. Analyse temporelle .

Dans la production de la parole , il existe des phénomènes

évoluant rapidement dans le temps . C'est le cas par exemple de la fermeture brusque du conduit vocal lors de la production d'une plosive .

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

Les principales méthodes de cette analyse sont :

- la méthode de la fonction d'autocorrélation ,
- la méthode de passage par zéro ,
- la méthode de la prédiction linéaire .

II.3.2.1. Méthode des passages par zéro .

Le principe de cette méthode est qu'on ne s'intéresse qu'aux changements de signe du signal $s(t)$, considéré . De ce fait , la méthode présente une grande simplicité de mise en oeuvre et une grande rapidité puisqu'elle n'exige pas d'opérations arithmétiques compliquées .

Cette méthode permet de localiser les fréquences des premiers formants ; elle permet aussi la mesure de la fréquence fondamentale " F₀ " après filtrage passe-bas .

Son inconvénient majeur réside dans la perte de l'information sur l'amplitude du signal du fait qu'on ne s'intéresse qu'à ses changements de signe .

II.3.2.2. Méthode de la fonction d'autocorrélation .

Pour que la méthode d'autocorrélation puisse être utilisée , le signal doit être découpé en fenêtres temporelles de 10 à 20 ms ; ces fenêtres doivent être ni trop courtes ni trop longues afin qu'elles puissent permettre la mesure des paramètres .

Cette méthode , appliquée à un signal $s(t)$, fait appel au calcul de la fonction d'autocorrélation définie par :

$$g(\tau) = \int_{-\infty}^{+\infty} s(t) \cdot s(t - \tau) \, dt \quad (\text{II} \cdot 9)$$

où " τ " représente un retard .

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

$$P(f) = \int_{-\infty}^{+\infty} g(\tau) \cdot e^{-2\pi j f \tau} \, d\tau \quad (\text{II} \cdot 10)$$

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

$$P_T = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P(f) df \quad (\text{II.11})$$

La fonction d'autocorrélation peut être générée à partir du signal analogique par des lignes à retard, des **sommateurs** ou des **modulateurs**. Elle peut aussi l'être par des procédés numériques à partir du signal échantillonné et ce à l'aide de **calculateurs** ou grâce à un matériel spécialisé (**autocorrélateurs**).

Cette méthode permet de mesurer la fréquence fondamentale " F₀ ". Elle présente l'avantage d'être simple, cependant elle exige des calculs trop longs.

II.3.2.3. Méthode de la prédiction linéaire .

La méthode d'analyse par prédiction linéaire appelée aussi technique du codage prédictif linéaire (L.P.C.) peut être considérée aussi bien comme une méthode temporelle que comme une méthode spectrale, ce qui fait qu'elle est soumise aux mêmes limitations que les autres méthodes .

Le principe de cette méthode était connu depuis longtemps, mais ce n'est qu'en 1966 (travaux de F. Itakura S. Saito) et surtout en 1972 (algorithmes détaillés par Markel) qu'elle se développa .

Cette méthode a acquis une place très importante . Ceci est dû à ce :

- qu'elle soit fondée sur un modèle simple (filtre numérique récuratif) de production de la parole, constituant une bonne approximation du système phonatoire .
- qu'elle permette une estimation très précise des paramètres de la parole tels que : les formants, la fréquence fondamentale, etc...
- que le calcul de ces paramètres soit relativement rapide .

A- Principe de la méthode L.P.C .

Le principe de cette méthode repose sur l'hypothèse selon laquelle un échantillon du signal de parole S_n (n : nombre entier) de période d'échantillonnage T, est prédit approximativement par une somme pondérée linéairement, d'un nombre " P " d'échantillons le précédant immédiatement .

Le signal prédit s'écrit :

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

où " P " est l'ordre du prédicteur
et les " a_k " : les coefficients du prédicteur

L'analyse (fig. III.4) effectuée par prédiction linéaire permet d'extraire à partir du signal de la parole, les paramètres pertinents qui serviront à la synthèse .

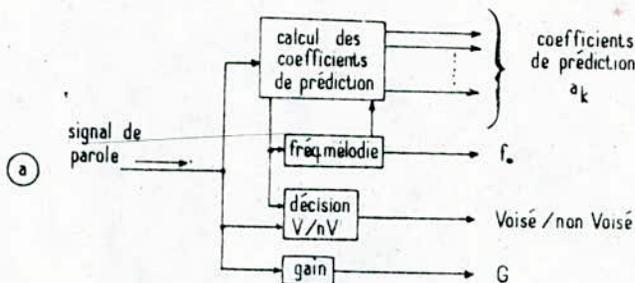


Fig. III.4

A- Calcul des coefficients prédicteurs .

Le problème en prédiction linéaire consiste à déterminer les " p " coefficients " a " de prédiction, de telle sorte que l'estimation du signal soit optimale (i.e. la meilleure possible) .
Cela revient à obtenir une optimisation de la relation (II.12).

Il existe plusieurs méthodes pour calculer les " p " coefficients . Elles sont souvent dites optimales au sens d'un certain critère qui est souvent celui des moindres carrés et parfois celui de la valeur absolue de l'erreur minimale .

C- Détermination du nombre des coefficients prédicteurs .

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

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

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

CHAPITRE III

LA SONAGRAPHIE

Il est possible grâce aux appareils de l'électroacoustique , d'enregistrer les sons , de les visualiser et de les analyser .

Le premier appareil qui a permis de donner , pour la parole , une représentation graphique des variations d'amplitude des vibrations sonores en fonction du temps est le KYMOGRAPHE .

Par la suite , il y eut l'OSCILLOGRAPHE (notamment le modèle cathodique) qui a apporté plus d'exactitude pour ce même tracé . On peut citer aussi l'ENREGISTREUR DE HAUTEUR et le BATHYMETRE .

Cependant , ces appareils ne permettent pas d'avoir une vue globale du phénomène sonore . Il fallut attendre l'avènement du SONAGRAPHE pour que des progrès plus nets soient réalisés dans le domaine de la parole .

III -1- Les premiers appareils analyseurs de la parole .

III.1.1. Le Kymographe .

Cet appareil permet de faire une analyse de la parole , en donnant une image des variations de la pression de l'air en fonction du temps .

Divers perfectionnements lui ont été apportés , notamment l'adjonction d'un régulateur électrique .

En reliant un magnétophone au kymographe , on obtient pour les vibrations de la parole - préalablement transformées en signaux électriques - un tracé kymographique appelé kymogramme (fig. III.1) .

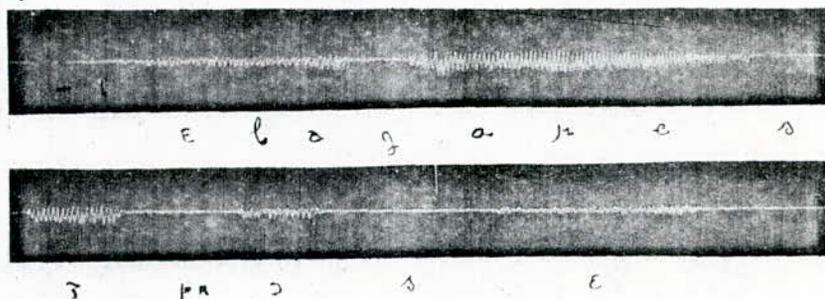


Fig. III. 1 : Kymogramme de la phrase : "elle a gagné son procès " (d'après Mettas , G. , 1971) .

Différents procédés peuvent être employés pour obtenir à partir de celui-ci , la courbe des variations de hauteur de la parole .

L'un d'entre eux peut être résumé de la manière suivante :

Chaque phénomène (i.e. plus petite unité distinctive du langage) prononcé devant le kymographe se traduit par une série de vibrations plus ou moins régulières , qui s'inscrivent, à l'aide d'un stylet , sur le papier du cylindre enrégistreur de l'appareil .

Il est alors possible de délimiter les sons , de mesurer , pour chaque voyelle , la distance qui sépare deux vibrations , c'est-à-dire la période et à partir des chiffres obtenus de calculer le fréquence (il suffit de connaître la vitesse " V " de rotation du cylindre et la longueur " L " de chaque période et l'on a alors $F = V/L$) . On peut ainsi dresser , sur papier millimétré , la courbe de ces vibrations , en portant la fréquence en ordonnée et la durée en abscisse (voir fig. III.21) .

Le kymographe a rendu de grands services dans les débuts de la phonétique expérimentale . Sa valeur en tant qu'appareil de recherche a diminué avec la création de moyens techniques plus perfectionnés tel que l'oscillographe .

Notons que le Kymographe peut être considéré comme un modèle primitif d'oscillographe mécanique .

III -1.2. OSCILLOGRAPHES .

Ces appareils permettent d'enregistrer les variations d'amplitude des variations d'un signal en fonction du temps . On peut les diviser en deux groupes :

- les oscillographes mécaniques .
- les oscillographes cathodiques .

III -1.2.1. Oscillographes mécaniques .

Il en existe divers modèles . On peut citer :

a) - Le mingographe

C'est un appareil constitué essentiellement d'un galvanomètre , enrégistreur et pour qui le stylet est remplacé par un fin jet d'encre .

Ce système présente l'avantage de réduire l'inertie du galvanomètre , ce qui permet d'enregistrer les fréquences jusqu'à 800 HZ (Mettas , O. , 1971) .

Par ailleurs , le mingographe a un autre intérêt : Celui d'être d'un maniement assez facile et d'un prix de revient des tracés peu élevé .

b - L'oscillographe à enregistrement direct par ultra-violet

Le stylet inscripteur est remplacé par un faisceau lumineux qui est réfléchi par un miroir oscillant, solidaire d'un galvanomètre.

L'inertie du système est, dans ce cas, considérablement réduite, ce qui permet d'étendre la gamme de fréquences jusqu'à 8000 Hz (Mettas, O., 1971).

L'inertie est donc très faible pour ces systèmes; mais l'idéal est de la supprimer entièrement, d'où l'apparition des **oscillographes** cathodiques.

III - 1.2.2. Oscillographes cathodiques.

Ces appareils qui commencèrent à se répandre dans les laboratoires, vers 1930, comprennent comme élément essentiel, un tube cathodique.

Dans ces systèmes, les inscripteurs mécaniques sont remplacés par un faisceau d'électrons projetés dans le vide à grande vitesse et venant percuter un écran fluorescent.

Si le phénomène à étudier est périodique, un dispositif de synchronisation permet d'observer la courbe toujours au même endroit de l'écran.

Cette fixité facilite l'étude visuelle de la courbe et permet de la photographier.

En faisant défiler un film à vitesse convenable, celui-ci va enregistrer les déplacements du spot en fonction du temps, c'est-à-dire, donner un tracé de l'oscillogramme (fig. III.2)

Fig. III.2 : oscillogramme de la phrase : " elle a gagné son procès " (d'après Mettas, O., 1971).

L'avantage majeur de l'oscillographe cathodique est qu'il présente une gamme de fréquences très grande : 0 à 10 MHz (Mettas, O., 1971).

Cet appareil donne un tracé précis et détaillé des vibrations de la parole en fonction du temps et permet ainsi l'étude de chacune d'elles. On peut ainsi, à partir de cette

courbe , obtenir le spectre de la parole , en analysant période par période pour en extraire les différents harmoniques .

Si le système est théoriquement parfait , il a cependant , quelques inconvénients d'ordre pratique : le délai de révélation du film et le prix . Mais le défaut principal de la méthode oscillographique n'est pas , quoiqu'il en soit , le prix , il est d'ordre psychologique (cf paragraphe III-2.2.) .

III.1.2.3. Oscillographe et sons de la parole .

Les voyelles donnent des oscillogrammes extrêmement complexes de formes toujours différentes , que ce soit pour des voyelles différentes ou même pour une même voyelle si elle n'est pas prononcée par le même locuteur ou si elle est énoncée sur des fréquences différentes (fig. III. 3) .

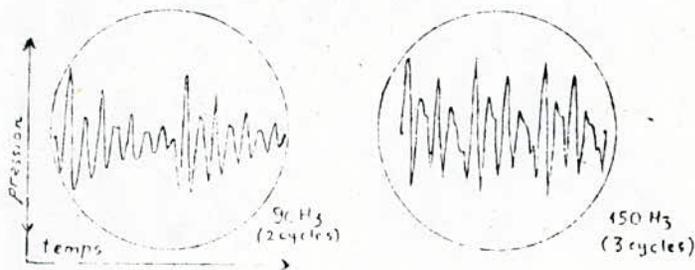


FIG. III.3 : aspect d'un oscillogramme vocalique : la voyelle a (d'après Emerit , E. , 1977) .

Les voyelles , du fait de leur instabilité fréquentielle , sont difficiles à observer , à moins d'utiliser sur l'oscillographe , un système de synchronisation spécial qui se règle sur les impulsions laryngées .

Sur les oscillogrammes , on remarque une différence essentielle : la voyelle parlée (fig. III.4) se reconnaît à sa quasi-périodicité (les cycles observés ne sont pas parfaitement identiques , ce qui fait que le phénomène considéré n'est pas exactement périodique) .



[a]

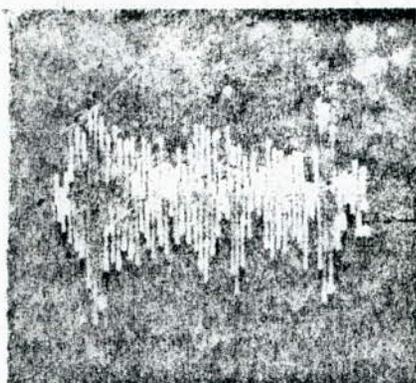


[ã]

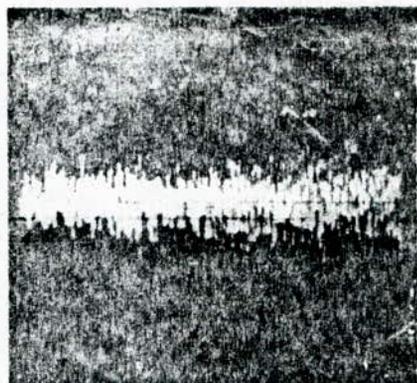
FIG. III.4 : oscillogramme de deux voyelles (d'après Emerit , E. , 1977) .

Par contre , pour les consonnes (fig. III.5) , on a diverses situations :

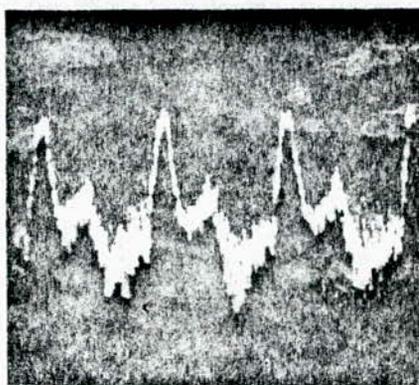
- Spirantes sourdes (ex. [ʃ]) : il n'y a aucune périodicité dans ce cas , puisqu'il ne s'agit que d'un bruit .
- Spirantes **sonores** (ex. [z] , [v]) : il y a une certaine périodicité . Le phénomène dans ce cas , est un bruit modulé par des impulsions laryngées (phénomènes semi-périodiques) .
- occlusives (sourdes ou sonores) : Ce sont des phénomènes quasi-instantanés et de ce fait , inobservables directement sur un oscillographe .



[ʃ]



[s]



[z]

Fig. III.5 : oscillogrammes de deux spirantes sourdes et d'une spirante sonore (d'après Emerit , E. , 1977) .

III -1.3. Enregistreur de hauteur .

C'est un appareil qui permet d'obtenir la courbe des variations de fréquences du fondamental (fig. III.6) .

Créé en 1927 par Grützmacher et Loser , il a été mis au point par G.Faht .

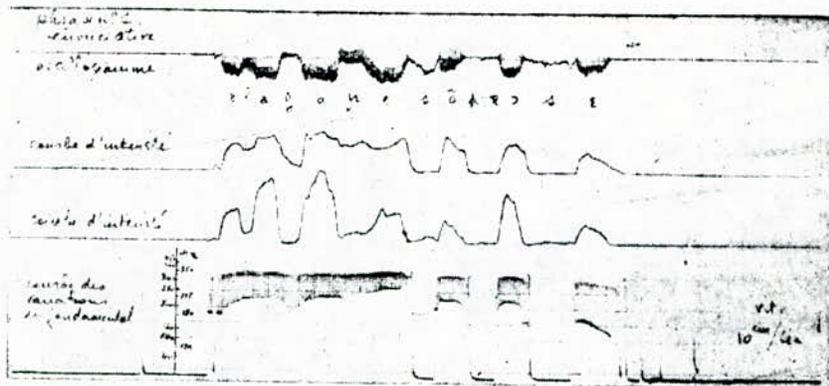


Fig. 25.



Fig. III. 6 : courbes des variations du fondamental de la phrase " elle a gagné son ~~procès~~ procès ", (d'après Mettas , Q. , 1971) .

Le signal sonore à analyser est transformé en oscillations électriques à l'aide d'un microphone . Après différents filtres , le signal original apparaît sous forme d'une courbe sinusoïdale dont la fréquence est celle du fondamental de la voix . Après un traitement approprié de la ligne sinusoïdale du fondamental , on obtient la courbe des variations de fréquence de celui-ci .

La gamme de fréquence de l'enregistreur de hauteur s'étend de 50 à 500 HZ (Mettas , O. , 1971) .

Par ailleurs , cet appareil présente un certain nombre d'avantages :

- c'est un appareil peu encombrant
- son maniement est relativement simple
- la courbe de variation du fondamental s'obtient assez rapidement .

Notons enfin , que la précision des mesures de la courbe de variations du fondamental est l'ordre de 98 % pour les voix de femmes et de 95 % pour celles d'hommes (Mettas , O., 1971) :

IV _1.4. Enregistreur logarithmique de niveau (bathymètre) .

Le bathymètre est un enregistreur qui permet de représenter sous forme logarithmique , les variations de l'intensité d'un signal sonore en fonction du temps et d'en mesurer le niveau en décibels (fig. III. 7) .

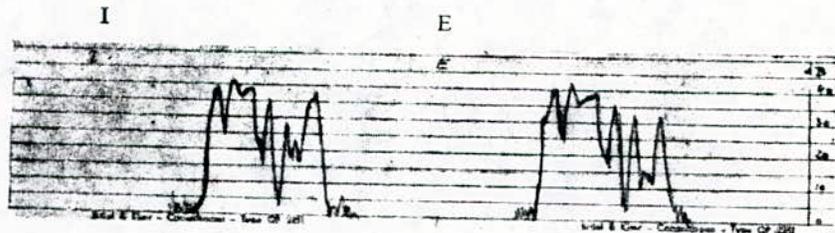


Fig. III. 7 : Courbes de niveau sonore de la phrase " elle a gagné son procès " (d'après Mettas, O., 1971)
I : phrase interrogative, E : phrase énonciative .

Le signal sonore transformé en oscillations électriques , par l'intermédiaire d'un micro , est transmis à un amplificateur de puissance , dont le gain en tension est commandé par un potentiomètre à curseur . A ce dernier est couplé un stylet qui se déplace sur un papier ciré .

Afin d'obtenir une représentation plus caractéristique des phrases , un filtre a été introduit entre le signal sonore et le bathymètre . Ainsi il a été établi , pour chaque bande de fréquence , la courbe de niveau de la phrase (interrogative et énonciative) [fig. III.8] .

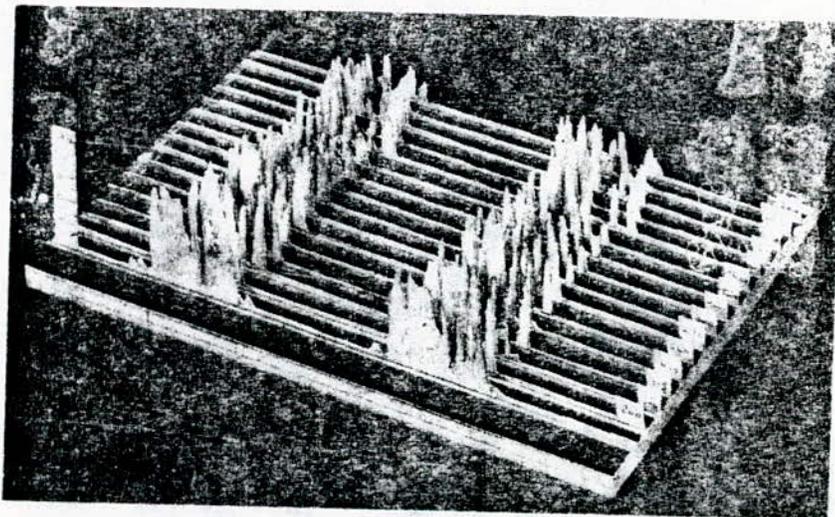


Fig. III.8 / : Courbes de niveau pour chaque bande , (d'après Mettas , O. , 1971) .

III.2. La sonographie .

Cette analyse est basée sur l'utilisation d'un appareil appelé sonographe (qui est une des réalisations commerciales) du spectrographe acoustique .

III.2.1. Notion de spectre du son .

L'onde sonore se propage dans l'air par la variation des molécules d'air qui se déplacent de telle façon à faire varier la pression sur nos oreilles . La forme de ces variations en fonction du temps constitue la forme d'onde . Fourier a montré que , si compliquée soit-elle , toute forme d'onde périodique pouvait être décomposée en un certain nombre d'oscillations régulières . La répartition des amplitudes de chacune des oscillations en fonction de leur fréquence correspond à ce qu'on appelle le spectre du son .

IV.2.2. Représentation physique du son -Notion d'objet sonore .

Le son peut être représenté par ses trois dimensions : amplitude , fréquence et durée .

Du point de vue physique , le langage apparaît alors comme un ensemble d'éléments sonores - sons quasi-périodiques et bruit qui évoluent continuellement dans le temps .

Nous avons vu lors de l'analyse du son par la méthode oscillographique que le problème était plutôt d'ordre psychologique .

En effet , il ne s'agit pas tellement de visualiser la totalité des vibrations individuelles puisque celles-ci ne sont pas perçues comme des phénomènes isolés ; au contraire, l'oreille les appréhende comme un tout continu possédant certaines propriétés globales auxquelles notre cerveau accède et qu'il importe par conséquent de concrétiser et d'exprimer .

On conçoit actuellement que , d'un point de vue théorique , un son du langage se présente comme une suite "d'objets sonores"

On appelle ainsi : la représentation , non plus de la pression en fonction du temps , mais celle des intensités des différentes fréquences , c'est-à-dire , le spectre du phénomène sonore qui évolue , en fonction du temps . Un objet sonore est donc un modèle des variations du spectre instantané en fonction du temps . Il se présente comme une sorte de montagne , avec des pointes et des creux dans les trois dimensions : amplitude , fréquence , temps (voir fig. III.8) .

Cette représentation s'est imposée comme étant la plus proche de la réalité psycho-physiologique de la communication verbale .

En fait , un même objet sonore peut être représenté de deux façons .

- Le physicien prend en considération les grandeurs physiques qui interviennent et qu'il mesure : l'intensité des variations de pression , la fréquence et le temps .

- Le psychologue , par contre , sait que les récepteurs psycho-physiques de l'être humain font subir à ces grandeurs une transformation (on parle d'anamorphose) selon des lois assez compliquées dont une première approximation est donnée par la règle de Fechner , selon laquelle la sensation (i.e. la grandeur perçue) est proportionnelle au logarithme de l'excitation (i. e. la grandeur physique correspondante)
 [fig. III.9] .

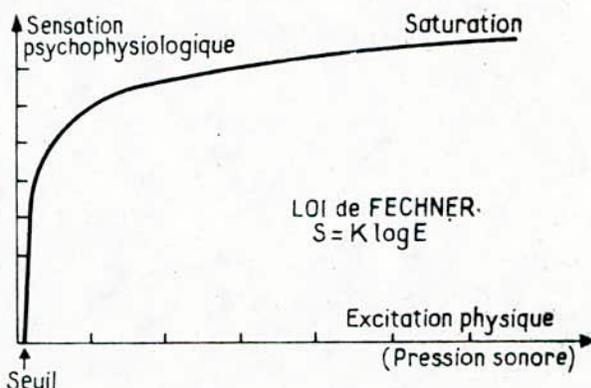


FIG. 14. — La loi de Fechner est une des lois les plus générales de la psycho-physiologie. Elle prévoit en gros que la sensation, c'est-à-dire le phénomène qui agit au niveau cérébral, est proportionnelle au logarithme de l'excitation physique : intensité du son, fréquence, durée, etc. Elle est à la base de la notation en décibels.

Fig.III.9 : La loi de Fechner (d'après Moles et Vallencien)

On est ainsi conduit à une anamorphose doublement logarithmique , qui remplace la grandeur " fréquence " par la "hauteur " (dont l'unité est l'octave) et l' " amplitude " (ou "intensité ") par le " niveau sonore " exprimé en dB .

Notons que d'après Koenig , " une bonne approximation de l'échelle subjective " pour des fréquences au-dessous 1000 HZ est une échelle linéaire .

"L'objet sonore " apparaît dans l'espace comme un trièdre (fig. III.10) .

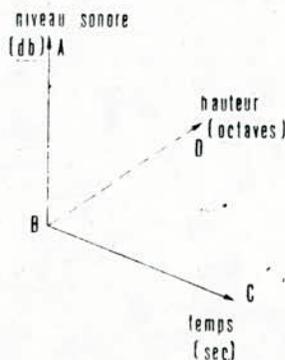


Fig. III.10. représentation de l'objet sonore (d'après Mettas , O. , 1971) .

Mais la représentation en trois dimensions n'est pas d'un maniement commode . Aussi adopte-t-on des représentations à deux dimensions , obtenues par des projections sur des plans de référence :

- la projection la plus simple est celle faite sur le plan ABC ; elle permet d'obtenir le diagramme des niveaux sonores en fonction du temps , relevé grâce au bathymètre
- Celle effectuée sur le plan ABD , permet de relever à différents intervalles de temps , le spectre instantané , c'est-à-dire , l'amplitude des composantes de Fourier en fonction de la fréquence .
- Enfin , la plus importante : c'est la projection sur le plan ABC (ou plan mélodique) . C'est probablement celle qui a donné ~~xxxx~~ la plus grande impulsion à la phonétique depuis 1945 .

On obtient ainsi à partir du plan fréquence-temps et en noircissant ce plan , de façon proportionnelle à l'intensité de la composante fréquence , ~~xxxx~~ instantanée , une représentation appelée sonagramme . Cette appellation est due aux laboratoires Bell .

Dans le sonagramme , les crêtes de l'objet sonore apparaissent comme des droites noires dont une étude approfondie faite dans toutes les langues , a conduit à un véritable alphabet sonographique .

IV. 3. Le sonographe

Le sonographe (visible speech) inventé pendant la guerre par les laboratoires Bell , a trouvé son utilisation par la suite dans l'analyse de la parole . C'est grâce à cet appareil que l'on obtient le sonagramme , ce qui permet de visualiser l'évolution au cours du temps , des fréquences composantes d'un signal complexe .

L'immense succès du système sonographique montre son importance . Il est devenu l'outil principal de l'analyse phonétique car il exprime la forme reconnaissable , celle que notre oreille perçoit , sur un diagramme de 30 cm de longueur étalant 2,4 s , c'est-à-dire , une courte phrase d'une vingtaine de phonèmes .

IV.3.1. Description de l'appareil .

Le sonographe (fig. III.11) est constitué de quatre parties essentielles :

- Un système d'enregistrement-lecture , comportant un disque magnétique , une tête d'enregistrement et une tête de lecture . La durée maximale d'enregistrement est de 2,4 s .

- Un filtre de fréquence variable
- Un amplificateur
- Un dispositif de marquage composé d'un cylindre métallique sur lequel est enroulé une feuille de papier spécial et un bâti supportant une vis sans fin le long de laquelle se déplace un stylet au cours de l'analyse .

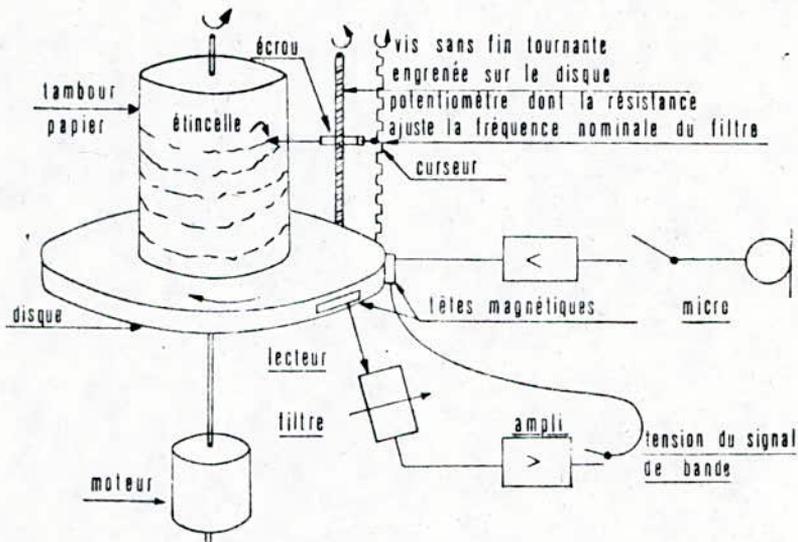


Fig. 20.
Sonographe

Fig. III.11. Le sonographe (d'après Mettas , O. , 1971) .

IV.3.2 . Fonctionnement .

Le sonographe était à l'origine composé d'un ensemble de filtres . Actuellement , il comporte un seul filtre dont la fréquence centrale (i.e. une fréquence qu'il laisse d'une façon privilégiée par rapport à toutes les autres) est variable .

Le signal de la parole à analyser est enregistré sur la périphérie du disque magnétique , soit directement à l'aide d'un microphone , soit par l'intermédiaire d'un magnétophone . Ce signal est transmis à un filtre dont la bande passante (i.e. bande de fréquences délimitée par deux fréquences de coupure, l'une inférieure et l'autre supérieure à la fréquence centrale) est constante , mais dont la fréquence centrale se déplace par pas d'environ 20 HZ , sur une gamme de 85 à 8000 HZ .

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

Cette analyse peut être décrite de la façon suivante: (fig.III.12)

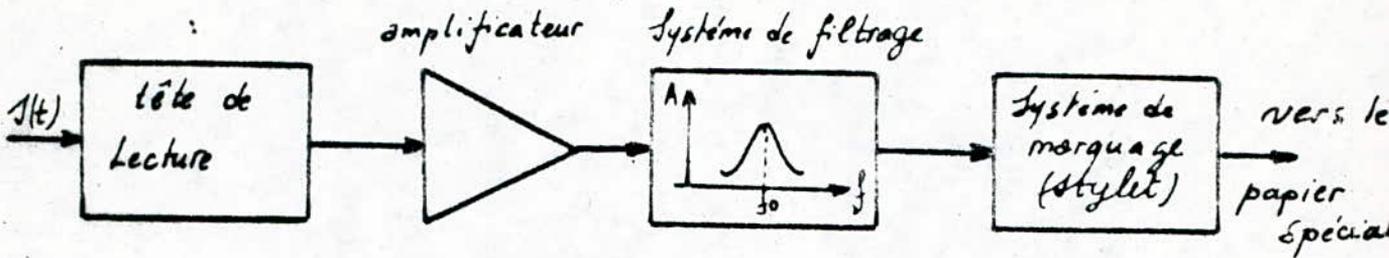


Fig. III.12 : Cheminement du signal "dans le sonographe ".

Le signal recueilli par la tête de lecture est acheminé au filtre . La tension électrique qui sort de celui-ci est amplifiée et devient suffisante pour alimenter une petite étincelle entre le stylet et le cylindre . La feuille de papier montée sur celui-ci , noircit sous l'influence de l'étincelle qui l'atteint et ceci est fonction de la tension de sortie du filtre . En fin d'analyse , on obtient sur le papier un sonogramme .

Notons que l'appareil est équipé de deux filtres passe-bande : l'un à bande étroite (45 HZ) , l'autre à bande large (300 HZ) .

IV.4. Utilisation du sonographe .

IV.4.1. Le sonographe en tant que spectrographe .

Bien que l'oscillogramme contienne toute l'information de la parole , il est pratiquement illisible , car il nous présente les harmoniques du son , inextricablement mêlés . De plus , il fournit une indication sans valeur auditive : la forme exacte de l'onde sonore .

Pour qu'un enregistrement sonore devienne " lisible " , il faut que tous les harmoniques (initialement mêlés dans le son) soient rangés par ordre de fréquence , sur un graphique appelé spectre sonore .

Ce dernier demeure inchangé , quelle que soit la forme que peut prendre le son sur l'écran de l'oscilloscope . Ceci montre l'énorme simplification qu'il apporte dans l'identification du son .

Un appareil relativement simple permet de tracer automatiquement de tels spectres : c'est le spectrographe . Pour le réaliser on fait appel à des filtres dont la bande passante est assez étroite et constante à toutes les fréquences .

Si l'on se limite à des analyses allant jusqu'à 4000 HZ ,

Il faudrait un minimum de 80 filtres , mais si l'on désire des analyses plus fines , il est alors nécessaire de disposer de pas moins de 160 filtres .

Cependant , une telle réalisation ne paraît pas raisonnable . Aussi a-t-on été amenés à envisager un autre moyen pour tracer les spectres sonores : il s'agit de l'utilisation du sonographe (en tant que spectrographe particulier) .

En parlant devant son microphone , on peut enregistrer par exemple , une voyelle . On déclenche l'analyse de celle-ci et l'on voit alors se dessiner le spectre sonore sur le cylindre .

On voit tout de suite l'énorme avantage que procure cet appareil par rapport à l'oscillographe , quand on sait que l'analyse par ce dernier demandait des jours entiers pour l'extraction de spectre sonore à partir de l'oscillogramme .

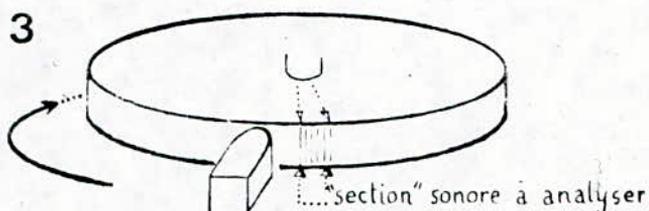


FIG. III. 13 (d'après Emerit , E. ; 1977) .

C'est à partir de " sections sonores " (fig. III. 13) que le sonographe permet de tracer plusieurs spectres de sons prononcés les uns après les autres . Ainsi , on observe par exemple , pour la voyelle [e] chantée sur quatre tons différents : 100 , 125 , 150 , et 200 HZ (fig. III. 14.) que : le spectre du bas montre des harmoniques rapprochés , celui du haut , des raies harmoniques éloignées . Par conséquent , un spectre sonore montre des raies d'autant plus écartées que la voix est haute .

Il s'ensuit que la structure harmonique est différente dans tous les cas , comme pour les oscillogrammes .

Cependant , il y a quelques chose de commun à tous ces spectres : il s'agit de l'enveloppe spectrale qu'on retrouve presque semblable pour une même voyelle , quelles que soient les différences de tons .

III. 4.2. Le sonogramme en bande étroite .

L'inconvénient majeur du spectre sonore est qu'il ne permet d'analyser qu'une " section sonore " très courte (une fraction de seconde) . On a remédié à ce défaut en adoptant le système dit du " sonogramme " en bande passante étroite .

C'est une sorte de spectre sonore mais dont les deux

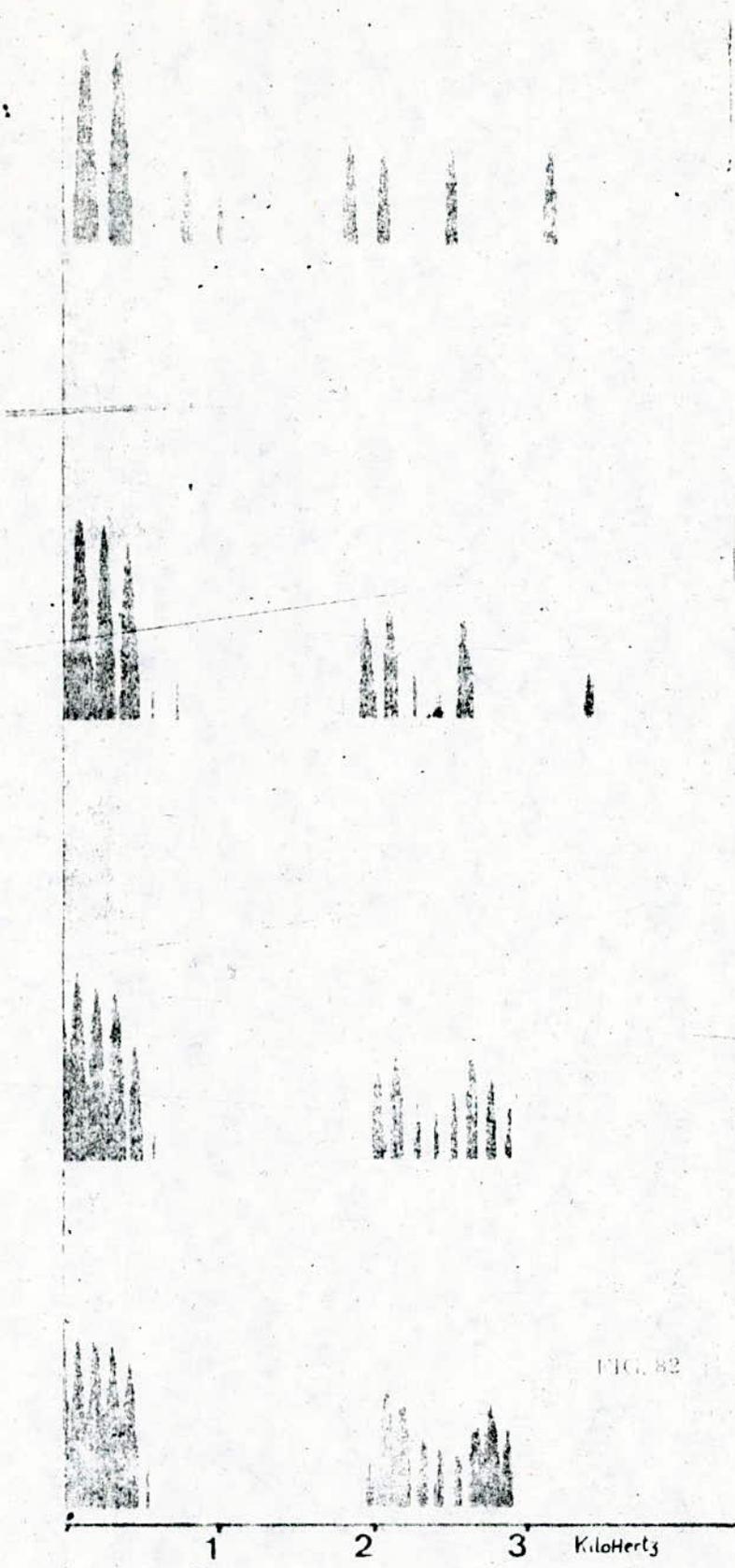


FIG. III. 14. (d'après Emerit , E. , 1977) .

dimensions sont les fréquences des harmoniques et le temps pendant lequel se déroule le discours (jusqu'à 2,4 s).

La troisième dimension, l'amplitude relative est rendue comme nous l'avons déjà vu par un noircissement plus ou moins grand du papier : plus une ligne harmonique est noire, plus l'harmonique est intense.

III. 4.2.1. Obtention

Lors de l'exploration du spectre, le filtre à bande étroite, extrait une bande de fréquence de 45 HZ, qui est toujours inférieure à la distance entre deux harmoniques.

L'image ainsi obtenue sera donc une résolution du spectre en ses différents harmoniques.

III.4.2.2. Informations apportées par le sonagramme en bande étroite.

Ce sonagramme (fig.III. 15) donne un aperçu global intéressant d'une séquence parlée.

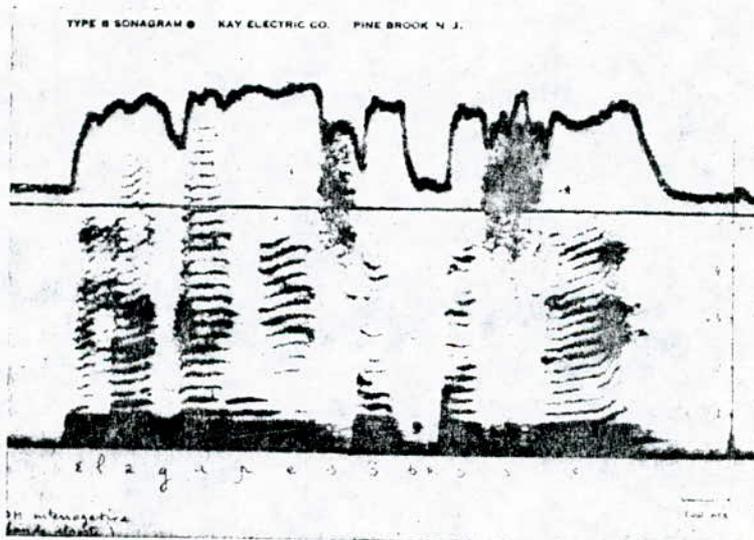


Fig. III. 15. Sonagramme en bande étroite de la séquence " elle a gagné son procès " (d'après Mettas, O., 1971).

Les informations apportées sont les suivantes :

a) le fondamental :

Il se dessine en ligne continue au bas de l'échelle des fréquences.

b) Les harmoniques :

- On remarque tout d'abord la nature des voyelles , qui se présentent sous forme de raies parallèles dans le sens vertical .
- la distance entre deux harmoniques qui se suivent , correspond théoriquement à la fréquence du fondamental de la voix .
- ceux des harmoniques qui ont été renforcés dans les cavités phonatoires apparaissent sous forme de lignes plus marquées (harmoniques renforcés dans certaines zones de fréquences) .

c) Les partiels :

Les consonnes spirantes sourdes (ex. [s]) se distinguent par l'absence d'harmoniques . Le sonagramme ne montre pas de raies , mais seulement un fouillis de partiels (vibrations partielles simples de fréquences quelconques) .

d) Phonèmes , syllabes , logatomes (fig. III. 16) .

- Les phonèmes : constituent les plus petites unités distinctives du langage , ils sont bien délimités .
- Les syllabes : groupes de phonèmes dont la délimitation est conventionnelle . Cette notion reste controversée .
- les logatomes : Ils correspondent à des associations d'une consonne et d'une voyelle attenante . Chaque consonne provoque en effet sur les voyelles attenantes des mouvements particuliers de formants appelés " transitions de formants " .

e) les formants :

Il s'agit de renforcements d'harmoniques dans certaines zones de fréquences particuliers à chaque voyelle.

On distingue (fig. III.16) au moins les quatre premiers formants (F1 , F2 , F3 , F4) .

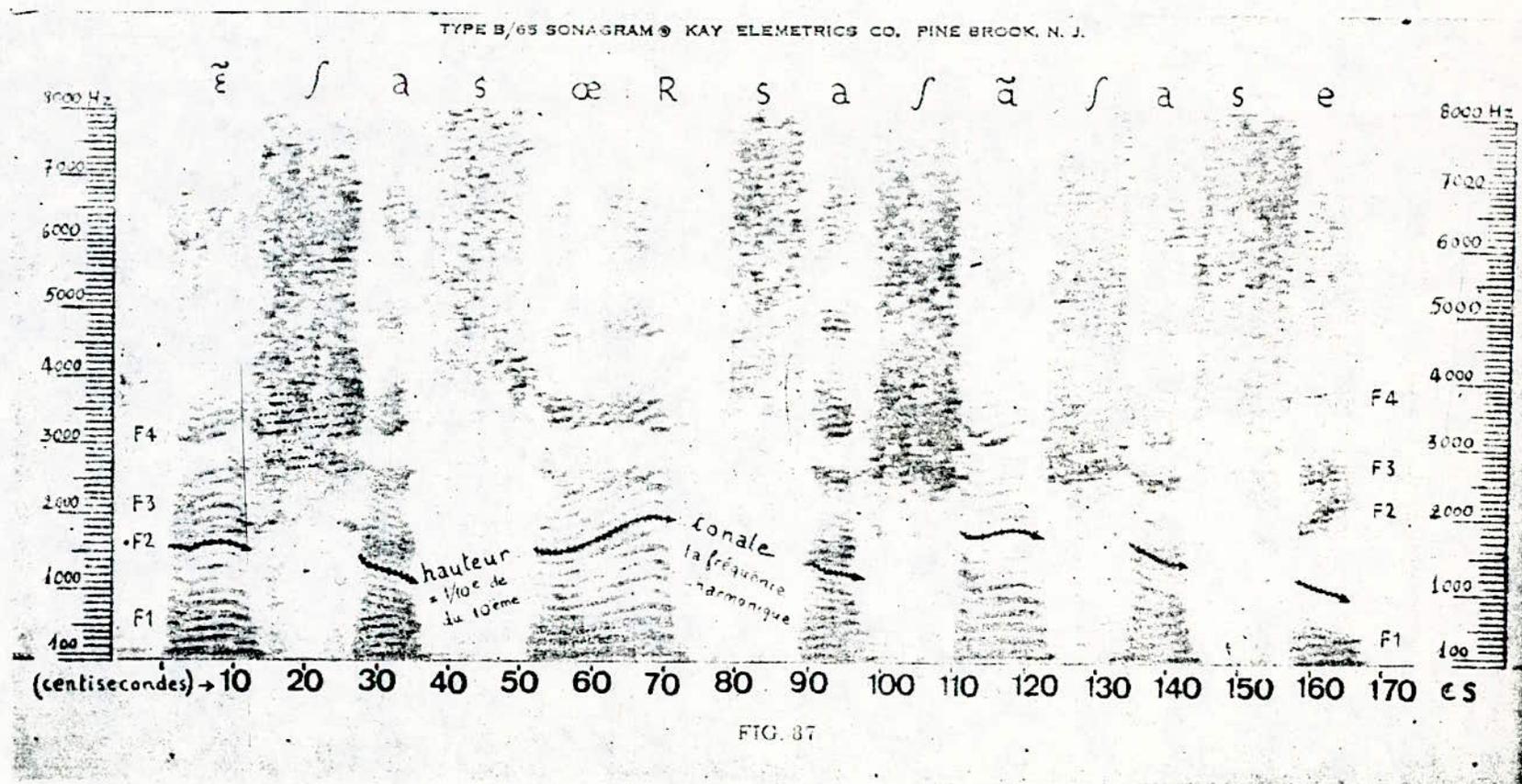
IV. 4.2.3. Défauts de la sonographie en bande passante étroite.

La sonographie en bande passante étroite apporte donc une quantité appréciable d'informations .

Cependant , elle présente deux défauts majeurs :

- La nette représentation des harmoniques gêne con-

FIG. IV. 16. (d'après Emerit , E. , 1977) .



sidérablement la lecture de l'aspect réel des formants .

— Le filtre d'analyse est trop sélectif (il présente une constante de temps de $1/45$ s) . Ceci confère au

C'est pour toutes ces raisons , que ce type d'analyse est généralement délaissé au profit d'un autre, moins sélectif (sonographie en large bande passante) qui présente de façon plus satisfaisante (i. e. de manière " lisible ") l'aspect des formants , autrement dit une analyse qui privilégie la forme des phonèmes au détriment des détails .

III. 4.3. La sonographie en large bande passante

III. 4.3.1. Obtention .

Cette analyse est effectuée avec le même sonographe que la précédente . La seule différence réside dans l'emploi d'un filtre dont la bande passante est de 300 HZ . Ceci confère au filtre une constante de temps très brève ($1/300$ s)

III .4.3.2. Informations apportées par cette analyse .

Dans ce cas , les formants sont mieux dessinés , plus " lisibles " (fig. III.17) mais on ne distingue plus les harmoniques . Toutefois , les renseignements sur les formants sont plus importants .

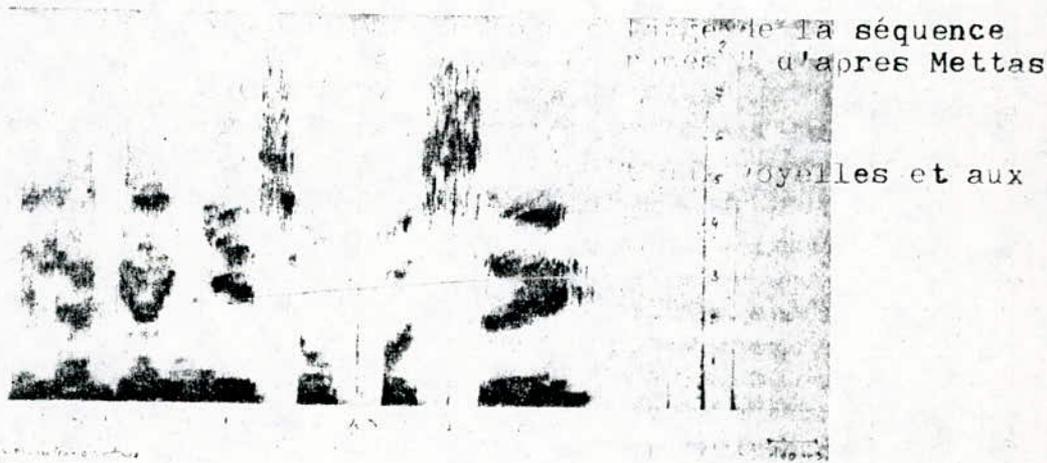


Fig. III. 17 : sonogramme en bande large de la séquence " elle a gagné son procès " d'après Mettas (O. , 1971) .

Les informations apportées se rapportent aux voyelles et aux consonnes :

A- Les voyelles .

- Les voyelles orales .

Elles se différencient principalement par leurs formants. Chacune d'elles apparaît avec une image qui lui est propre ; son timbre est déterminé par la position des deux premiers formants (les plus bas sur l'échelle des fréquences) . On peut se contenter de ne considérer qu'eux pour établir une première classification acoustique des voyelles, comme l'a fait P. Delattre (cf. chap. I) d'après la fréquence moyenne de leurs deux premiers formants F1 et F2 .

- Les voyelles nasales .

Rares sont les langues qui comportent des voyelles nasales. Celles-ci ont donné lieu à bien des controverses (fig. III.18)

Au départ , on a pensé à un " formant nasalisateur " en hautes fréquences . En effet , on avait remarqué que certaines voyelles nasalisées comportaient un faible formants autour de 4500 HZ . Ceci est le cas lors du passage de [ɔ] à [ɔ̃] par exemple .

Cependant lorsqu' on passe de la prononciation de [a] à celle de [ã] , on ne trouve aucune trace de ~~fréquence~~ renforcement , ni à cette fréquence , ni aux fréquences voisines .

En fait ce n'est pas aux hautes fréquences qu'il faut chercher la perception de la nasalité . Au contraire , il faudrait voir plutôt du côté des basses fréquences.

Ainsi , on remarque que si on affaiblit certaines bandes de fréquences dans la structure formantique d'une voyelle , on la nasalise .

Si on considère simultanément les sonagrammes des voyelles orales et de leurs correspondantes nasales , l'identification certaine est pratiquement impossible du fait que les fréquences des formants sont pratiquement les mêmes . Elles ne diffèrent que par les amplitudes relatives de leurs formants . C'est ainsi que l'on observe , par exemple dans le passage de la prononciation de [ɔ] à celle de [ɔ̃] , que la nasalisation a affaibli F₂ . En fait , le sonagramme est incapable de rendre compte de façon exacte de ce phénomène , ce qui amène à recourir à une autre méthode : le spectre de bande (fig. III. 19) .

Ce dernier est un graphique montrant les amplitudes relatives créées par les harmoniques dans les bandes de fréquences correspondant aux bandes passantes des filtres .

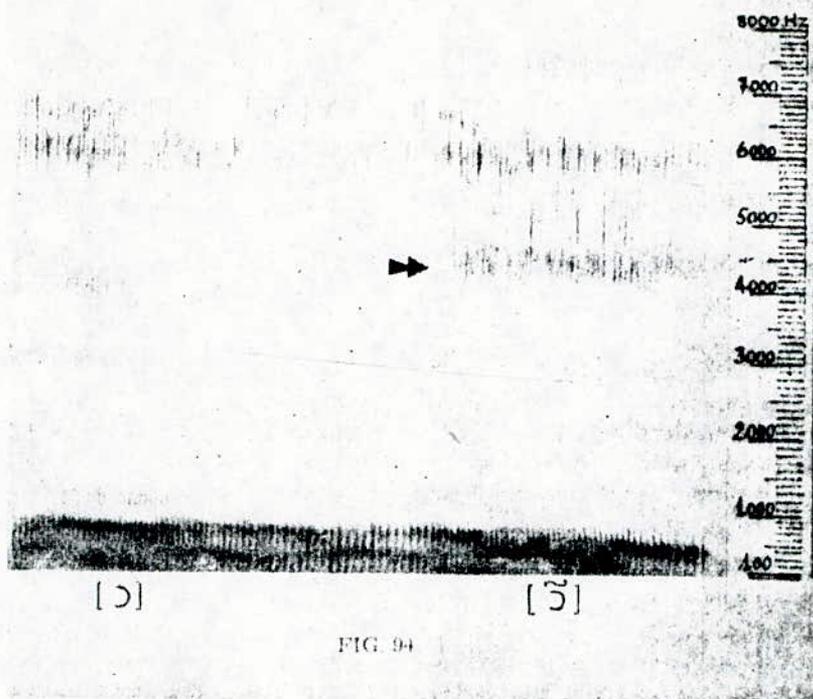


Fig. 94 : La flèche montre l'apparition d'un prétendu « formant nasalisateur » du genre de celui trouvé par M. DURAND, et dont l'inefficacité est maintenant prouvée.

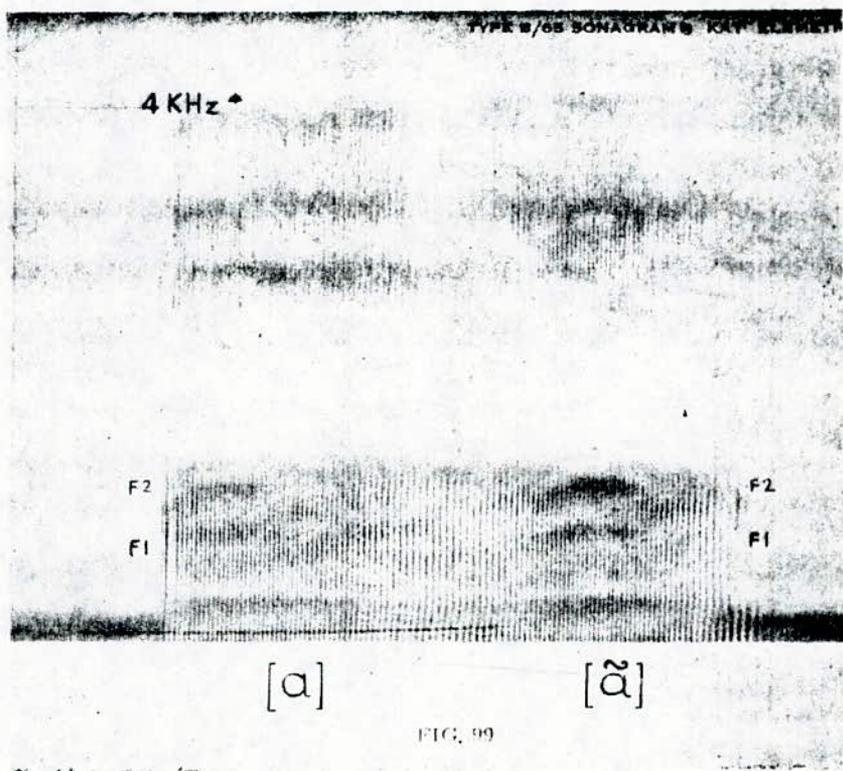


FIG. III. 18 (d'après Emerit , E. , 1977

Grâce à ce spectre , la mesure des amplitudes relatives devient alors très précise .

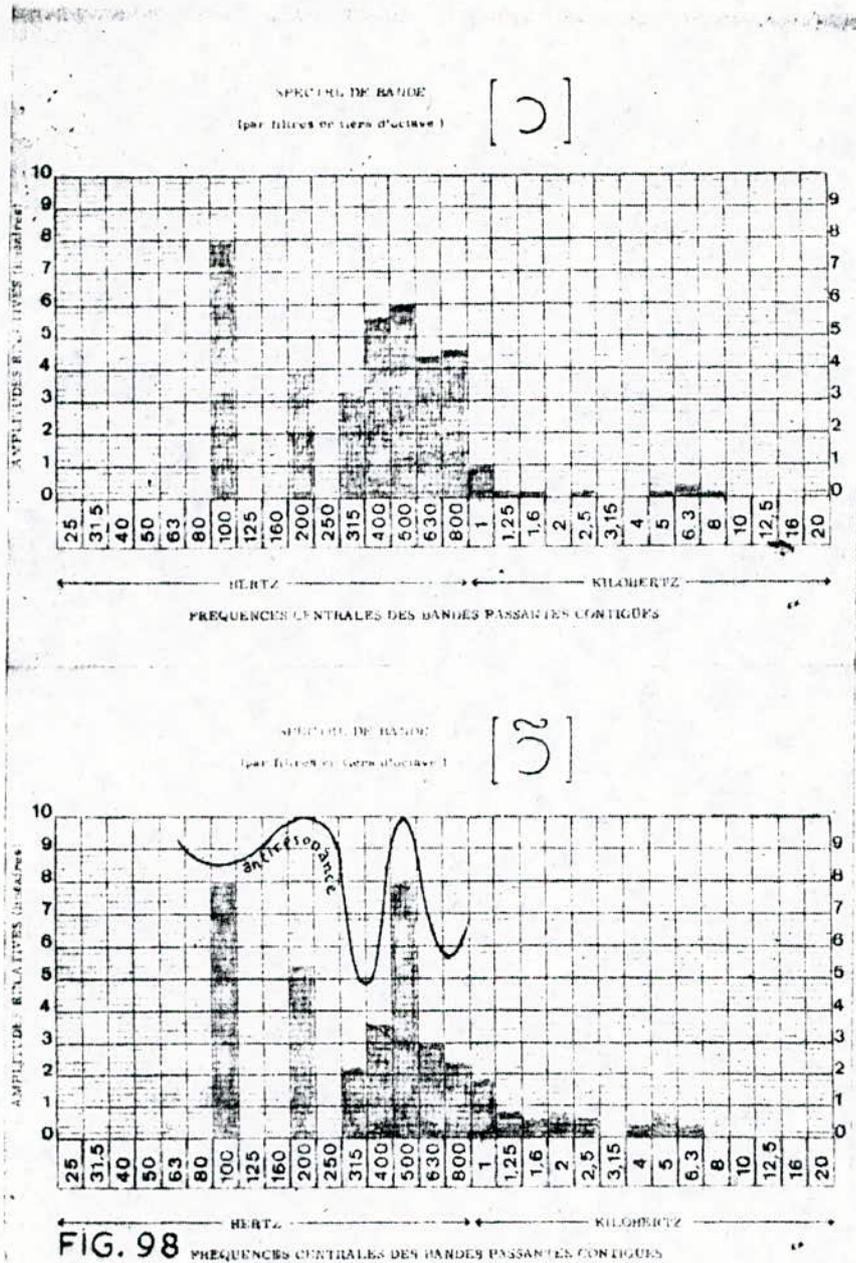


Fig. III.19 (d'après Emerit , E. , 1977) .

B- Les consonnes .

Une consonne est un phonème qui ne peut s'employer qu'en étant accompagné d'une voyelle , constituant ainsi l'ensemble appelé logatome .

Dans celui-ci , les formants de la voyelle , sont notablement perturbés par la proximité de la consonne . Ces perturbations sont dites " transitions des formants " et elles sont essentielles pour la perception des consonnes .

Ces perturbations sont dites " transitions des formants " et elles sont essentielles pour la perception des consonnes .

Sur le sonagramme (fig. III.20) , on observe l'exemple de "transition de formants " pour les logatomes formés sur la voyelle [a] dont les formants F1-F2-F3 en l'absence de consonne , sont de 750 - 1350 et 2500 HZ .

On remarque alors que ces formants n'ont pas leurs valeurs habituelles , lorsque la voyelle est précédée ou suivie d'une consonne . Ainsi dans le cas de [da] , les fréquences deviennent 200 - 1800 et 2700 à 2800 HZ pour les trois formants respectivement .

Par ailleurs , les transitions convergent toutes vers une fréquence unique caractéristique de chaque consonne : le "locus". Dans l'exemple de [ga] , celui-ci est pour F2 - F3 , d'environ 2500 HZ .

TYPE B/65 SONAGRAMS KAY ELECTRONICS CO. FINE BROOK, N. J.

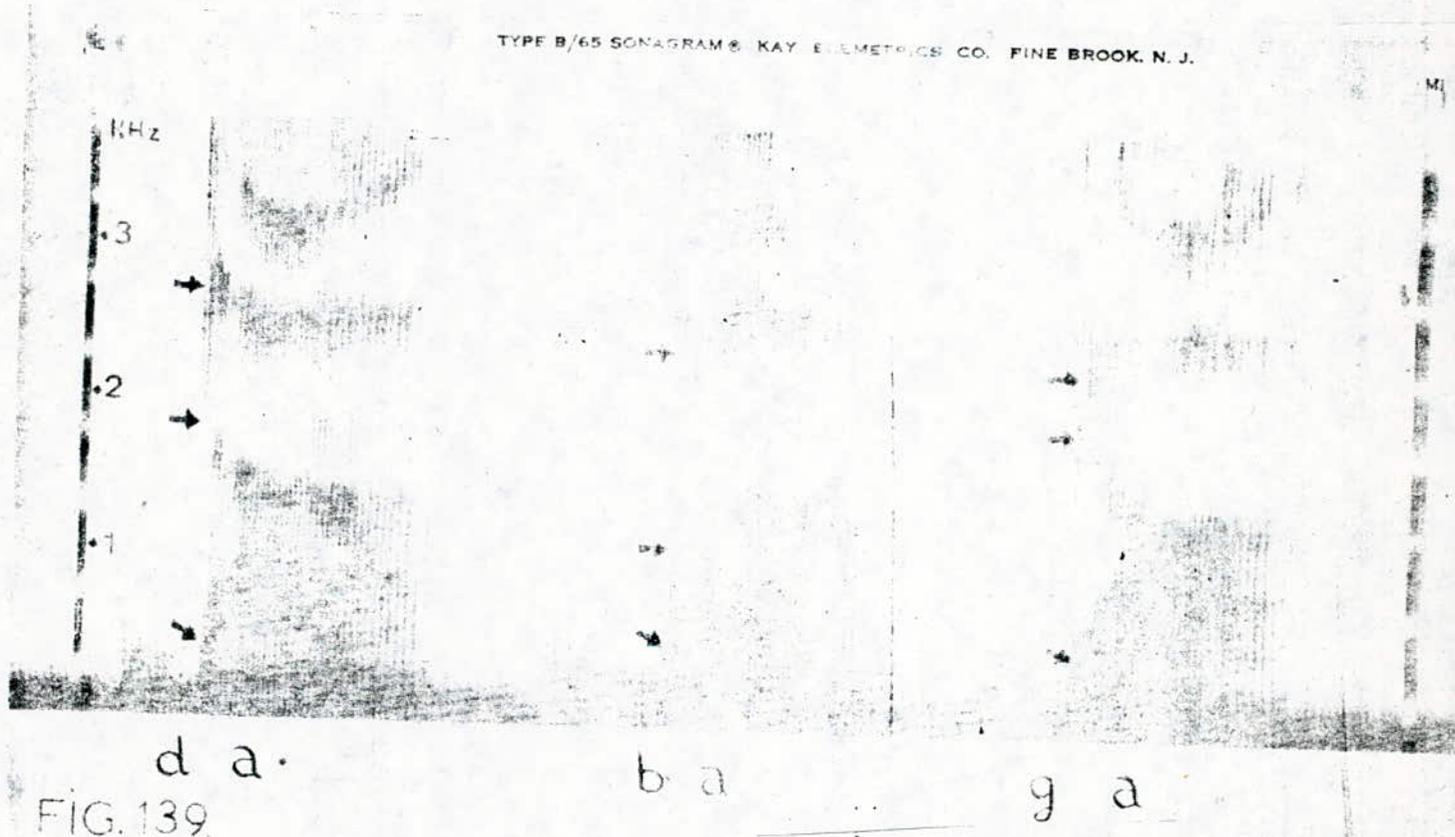


FIG. 139

Fig. III. 20 (d'après Emerit , E. , 1977) .

III.5. Avantages et inconvénients du sonographe .

En plus , de son prix relativement élevé , le sonographe présente un certain nombre d'inconvénients :

- la représentation du phénomène sonore n'est pas immédiate .
- l'intensité n'est pas indiquée avec une grande précision .
- la durée d'une phrase à analyser ne peut dépasser 2,4 s (ce qui correspond à une vingtaine de phonèmes) .

Cependant , ces quelques inconvénients sont relativement faibles , en comparaison des qualités que possède cet appareillage , et dont on peut citer :

- le sonagramme donne une image de la réalité physiologique . Ainsi , lorsqu'on prononce une suite de phonèmes , tels que [da] , par exemple , on passe d'un son consonantique [d] à un son vocalique [a] . Or les organes articulatoires ne se trouvent pas immédiatement dans la position requise pour l'émission de la voyelle (la forme et le volume des cavités phonatoires ne se modifient que progressivement) . Ceci se traduit sur le sonagramme par les " transitions de formants " (cf paragr. précédent) .
- On reproche souvent au sonographe , son manque de précision , mais cet appareil a l'avantage de présenter une forme , au lieu de mesures . Celle-ci fournit une vue globale du phénomène à chaque instant . Ainsi , il est possible de comparer immédiatement , pour une phrase interrogative et la même phrase énonciative par exemple (fig. III. 21) , non seulement les variations du fondamental , mais aussi le mouvement des formants , la durée des phonèmes , etc .

III.6- Le sonographe et les autres appareils d'analyse .

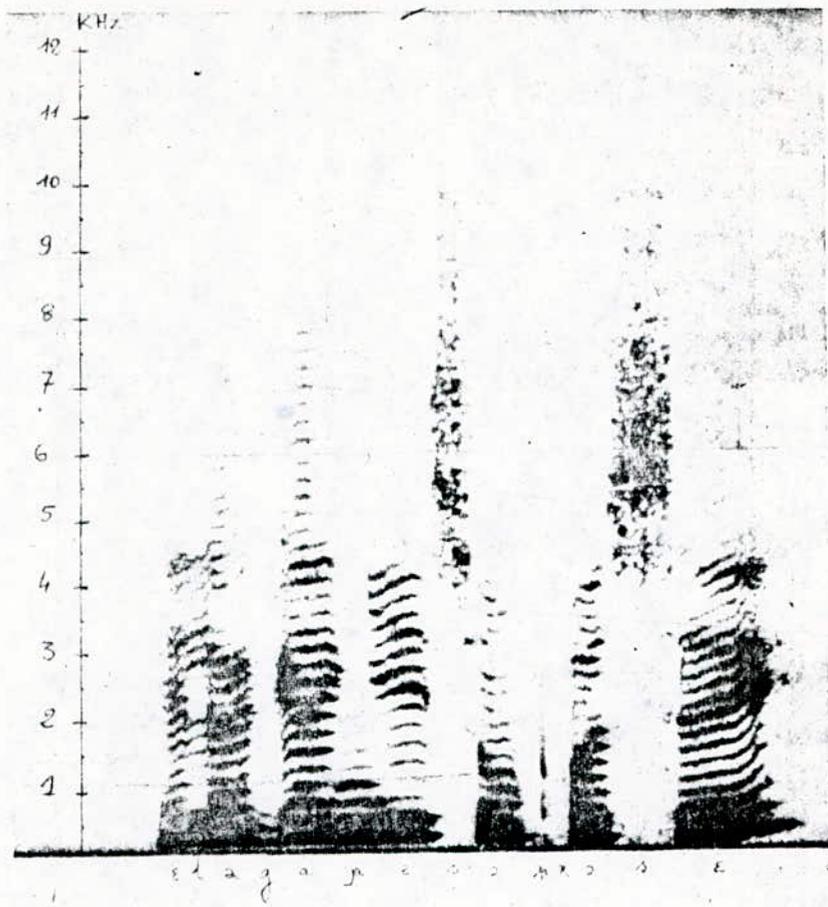
Parmi les appareils d'analyse de la parole , nous avons vu que l'oscillographe remplaçait avantageusement le kymographe . La totalité de l'information que l'on peut retirer du phénomène sonore est en principe contenue dans un oscillogramme bien fait . Cependant , ces données ne sont pas immédiates et il y a parfois un travail assez long à faire pour les découvrir .

D'autres appareils permettent de représenter un aspect particulier du phénomène sonore . C'est le cas , par exemple , de l'enregistreur de hauteur , qui permet d'obtenir la courbe de variations du fondamental . Mais le fait d'isoler un aspect du phénomène étudié constitue une limitation .

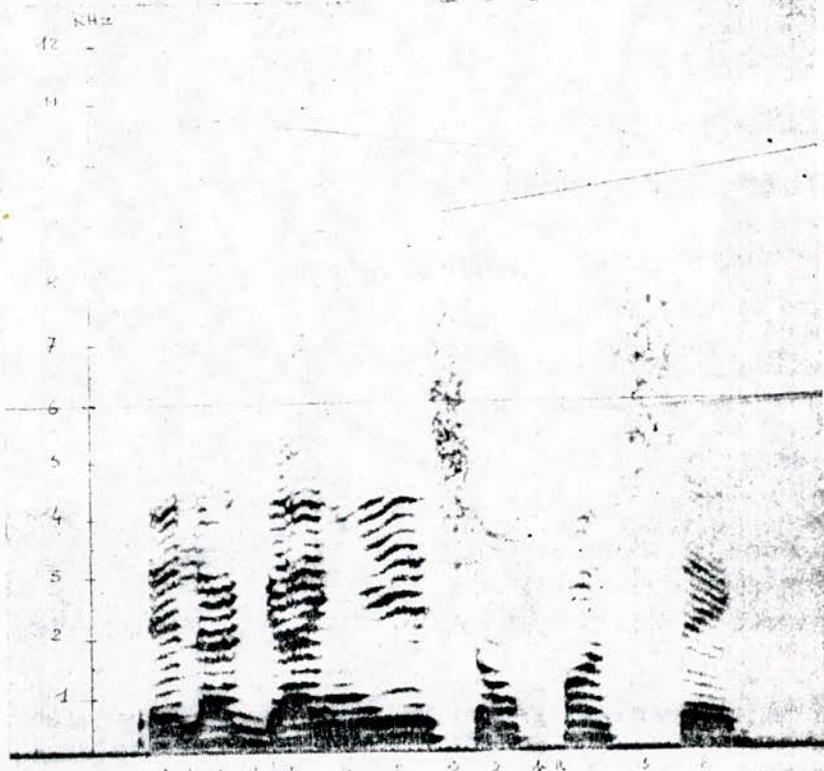
Le sonographe n'est pas très précis , mais il présente l'avantage de fournir une image globale du phénomène sonore , ce qui est très important (cf paragr. précédent) .

Conclusion .

L'ensemble des appareils que nous venons de voir , montre la richesse des moyens d'étude du domaine de la parole .



interrogative 100ms 1 seconde



evocative 100ms 1 seconde

Fig. III. 21 .

En effet , bien des changements ont eu lieu depuis que l'abbé Rousselot ouvrait la voie à la phonétique expérimentale .

La représentation de la parole sous son aspect le plus détaillé , n'est plus , comme au début du siècle , le principal objet d'intérêt : on s'attache également à l'aspect perceptif du langage . Cette orientation nouvelle des recherches , conjuguée avec l'évolution des techniques instrumentales fait que les laboratoires d'analyse de la parole voient leur appareillage s'accroître de plus en plus .

L'un des appareils les plus importants est le sonographe , qui a déjà fait avancer considérablement les recherches et contribué à enrichir nos connaissances dans le domaine du langage

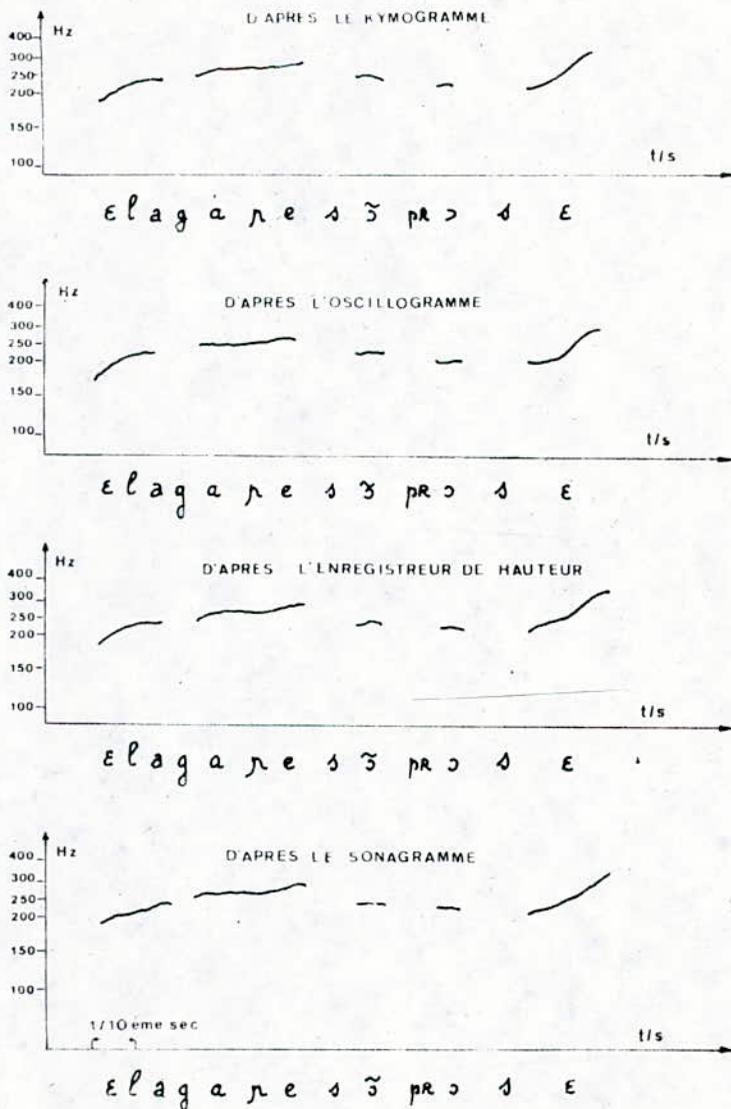


Fig. 30.
Courbes de variations du fondamental

CHAPITRE IV

LA PREDICTION LINEAIRE

L'analyse par prédiction linéaire constitue une méthode robuste et précise pour l'estimation des paramètres pertinents du signal. Dans le domaine de la parole, cette analyse regroupe diverses méthodes (on parle parfois de formulations).

On distingue notamment :

- la méthode de covariance
- la méthode d'autocorrélation
- la formulation du filtre en treillis.

Souvent, il suffit d'étudier les méthodes citées ci-dessus, toutes les autres leur sont équivalentes.

IV.1. Les différents critères utilisés en prédiction linéaire.

Les techniques de prédiction linéaire sont souvent utilisées dans le but de diminuer la redondance du signal. Ce dernier peut alors être décrit par un ensemble limité de paramètres. Pour déterminer ceux-ci, on peut utiliser différents critères. Les méthodes classiques font appel surtout au critère de l'erreur quadratique totale minimale (appelé aussi critère des moindres carrés ou encore critère L_2). Cependant, il s'avère parfois très intéressant d'utiliser un autre critère : celui de la valeur absolue de l'erreur minimale (dit aussi critère L_1) comme nous le verrons au cours de ce chapitre.

IV.1.1. Le critère L_2 .

L'erreur de prédiction entre l'échantillon du signal de parole original et l'échantillon prédit est définie par :

$$e(j) = s(j) - \hat{s}(j) = s(j) - \sum_{i=1}^P a_i s(j-i) \quad (\text{IV. 1})$$

soit E l'erreur quadratique totale :

$$E = \sum_j e^2(j) = \sum_j \left[s(j) - \sum_{i=1}^P a_i s(j-i) \right]^2 \quad (\text{IV. 2})$$

L'application du critère L_2 consiste en la minimisation de E , ce qui revient à annuler les dérivées partielles par rapport à chaque coefficient prédictif a_i :

$$\frac{\partial E}{\partial a_i} = 0 \quad 1 \leq i \leq P \quad : \quad (\text{IV. 3})$$

Il en résulte le système d'équations :

$$\sum_j S(j) S(j-k) = \sum_j \sum_{i=1}^P \alpha(i) S(j-i) S(j-k) \quad (IV. 4)$$

$1 \leq k \leq P, \quad 1 \leq i \leq P$

d'où :

$$\sum_j S(j) S(j-k) = \sum_{i=1}^P \alpha(i) \sum_j S(j-i) S(j-k) \quad (IV. 5)$$

Les coefficients " $\alpha(i)$ " constituent les " p " inconnues du système de " p " équations linéaires (IV. 5) appelées " équations normales " .

IV. 1.2. Le critère L_1 .

IV.1.2.1.- Historique .

Ce critère a été introduit par Boscovitch en 1757 (i. e. bien avant les premiers rapports sur les méthodes relatives au critère des moindres carrés) . En ce moment , l'application de ce critère était basée sur deux contraintes (d'une part , l'erreur totale devait être nulle et d'autre part , l'erreur absolue devait être minimale) .

En 1887 , Edgeworth suggéra de ne retenir que la seconde contrainte ; cependant , les algorithmes qu'il proposa étaient trop compliqués par rapport à ceux relatifs à L_2 . Il fallut attendre le développement de la programmation linéaire , pour qu'apparaissent des algorithmes L_1 puissants .

IV.1.2.2.- Définition .

Soit $S(j)$ le signal
 j : nombre entier .
 et $U(a)$ une approximation linéaire de $S(j)$
 (" a " étant le vecteur des coefficients de prédiction

L'application du critère L_1 consiste en la minimisation de la quantité :

$$L_1 = \sum_j |S(j) - U(a)| = \sum_j |S(j) - \sum_{i=1}^P \alpha(i) a(i)| \quad (IV. 6)$$

$1 \leq i \leq P$

(" i " : ordre du prédicteur) .

Dans le cas de la parole , $U(i)$ a une forme particulière :

$$U(i) = S(j-i)$$

il vient alors :

$$L_1 = \sum_j |S(j) - \sum_{i=1}^P \alpha(i) S(j-i)| \quad (IV. 7)$$

IV. 1.2.3. Propriétés .

L'estimateur L_1 présente une caractéristique importante : la robustesse . Celle-ci peut aisément être mise en évidence dans le cas simple $p=1$ et $u_1(j) = 1$, $j = 1, \dots, m$

Dans ce cas , le signal est approximé par une constante " a " .
La fonction L_1 peut alors s'écrire :

$$L_1 = \sum_{j=1}^m |s(j) - a| \quad (IV. 8)$$

La valeur optimale de " a " constitue la médiane des échantillons $s(j)$, i. e., une valeur de " a " telle que le nombre d'échantillons $s(j)$ plus petits que " a " est égal au nombre d'échantillons $s(j)$ plus grands que " a " .

Pour le même problème d'estimation mais avec la fonction L_2 , la solution pour " a " est la moyenne arithmétique des échantillons $s(j)$.

Si certaines des observations $s(j)$ renferment du bruit , alors des erreurs vont affecter la médiane et la moyenne arithmétique des échantillons .

Tant qu'un échantillon reste d'un même côté de la médiane celle-ci ne change pas . Ceci n'est pas le cas pour la moyenne arithmétique . Plus précisément , il a été montré que l'estimateur de la valeur absolue minimale est l'estimateur du maximum de vraisemblance et , par conséquent , il est asymptotiquement efficace lorsque les erreurs suivent une distribution de Laplace en double exponentielle . Cette propriété reste valable pour les systèmes d'ordre supérieur .

Le critère de la valeur absolue minimale est ainsi moins sensible aux erreurs extrêmes que le critère des moindres carrés ; l'estimateur L_1 est dit robuste .

L'exemple simple décrit précédemment révèle aussi une caractéristique moins désirable de l'estimateur L_1 . On peut voir facilement d'après (IV.8) que pour un même nombre " m " d'échantillons , la valeur optimale de " a " se situe n'importe où entre les deux échantillons $s(j)$ centraux . Ainsi , la minimisation de (IV. 8) conduit souvent à une infinité de solutions

IV. 2. Les principales méthodes de la prédiction linéaire .

IV. 2.1. Autocorrélation et covariance .

Les principales méthodes de la L P C suivant le critère L_2 sont la méthode exacte (non utilisée , cependant dans le domaine de la parole) , celle du filtre en treillis , et surtout la méthode de covariance et celle d'autocorrélation .

Dans le paragraphe précédent, nous avons vu que les coefficients " α_k " sont obtenus par la résolution des " équations normales " .

$$\text{Soit } C(k, i) = \sum_j S(j-i) S(j-k) \quad (\text{IV. 9})$$

Nous pouvons écrire alors, de façon plus compacte :

$$\sum_{i=1}^P \alpha_i C(k, i) = C(k, 0) \quad 1 \leq k \leq P \quad (\text{IV. 10})$$

Pour déterminer les " α_i " , il faut résoudre l'équation (IV. 10) .

Jusqu'ici les bornes de sommation n'ont pas été indiquées de façon explicite dans les équations (IV. 9) et (IV. 10)

Nous allons voir dans ce qui suit, les méthodes d'analyse qui les définiront .

IV. 2.1.1. La méthode de la covariance .

Cette technique suppose le signal non stationnaire à l'intérieur de l'intervalle d'analyse .

Elle est basée sur les hypothèses suivantes :

a - le signal est défini pour " $N+P$ " échantillons consécutifs, avec " N " entier .

b - un échantillon du signal de parole peut être prédit à l'aide des " P " échantillons précédents .

c - l'hypothèse (b) est valable pour les " N " échantillons consécutifs .

d - l'erreur quadratique totale entre le signal et sa valeur prédite est minimisée pour l'ensemble des " N " échantillons consécutifs .

D'après les hypothèses précédentes, on peut écrire :

$$\sum_{j=0}^{N-1} e^2(j) = E(j) \quad (\text{IV. 11})$$

et par suite :

$$C(k, i) = \sum_{j=0}^{N-1} S(j-k) S(j-i) \quad \begin{matrix} 0 \leq i \leq P \\ 1 \leq k \leq P \end{matrix} \quad (\text{IV. 12})$$

En changeant l'indice de sommation, nous pouvons exprimer $C(k, i)$ de la façon suivante :

$$C(k, i) = \sum_{j=k}^{N-1-k} S(j) S(j+k-i) \quad (\text{IV. 13})$$

Le système d'équations à résoudre s'écrit :

$$\sum_{i=1}^P a_i c(k, i) = c(k, 0) \quad \begin{matrix} 1 \leq k \leq P \\ 0 \leq i \leq P \end{matrix} \quad (\text{IV. 14})$$

ou, sous forme matricielle :

$$\begin{bmatrix} c(1,1) & \dots & c(1,P) \\ \vdots & & \vdots \\ c(P,1) & \dots & c(P,P) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} c(1,0) \\ \vdots \\ c(P,0) \end{bmatrix} \quad (\text{IV. 15})$$

Du fait que $c(k, i) = c(i, k)$, la matrice $P \times P$ de covariance est symétrique.

L'équation (IV. 15) peut être résolue efficacement grâce à l'algorithme de cholesky (dit aussi algorithme de la racine carrée).

IV. 2.1.2. La méthode d'autocorrélation .

Cette méthode considère que le signal est stationnaire dans un intervalle de temps fini. Aussi, son étude requiert-elle l'utilisation d'une fenêtre permettant d'obtenir un signal $S_N(j)$ nul à l'extérieur de l'intervalle de " \wedge " échantillons délimité par la fenêtre considérée. Cette dernière est souvent celle de Hamming (fig. IV. 1)

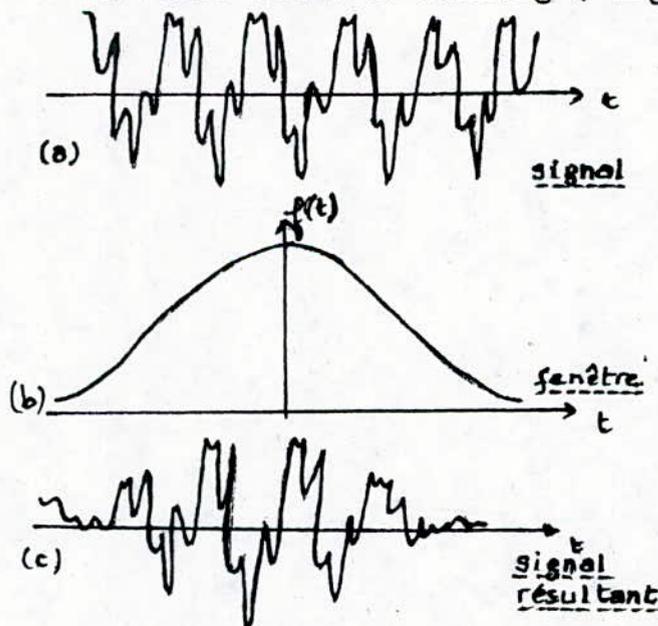


Fig. IV. 1 : Effet d'une fenêtre de Hamming (fig.1).

Cette technique repose sur les hypothèses suivantes :

a - le signal est nul à l'extérieur de l'intervalle $0 \leq n \leq N$.

$$S_N(j) = S(j) \times W(j)$$

(IV. 16)

$W(j)$ étant la fenêtre utilisée .

b - chaque échantillon peut être prédit par les " p " échantillons précédents et ceci pour tout le temps $j \in]-\infty, +\infty[$

Si $S(j)$ est différent de zéro seulement pour $0 \leq j \leq N-1$, alors l'erreur de prédiction correspondante $e(j)$ pour le prédicteur d'ordre " p " sera différent de zéro dans l'intervalle $0 \leq j \leq N+P+1$ et par suite, on peut écrire :

$$E(j) = \sum_{j=-\infty}^{+\infty} e(j) = \sum_{j=0}^{N+P+1} e^e(j) \quad (\text{IV. 17})$$

Les bornes de l'expression de $C(k, i)$ dans (IV.9) sont identiques à celle de l'équation (IV.16), on a alors :

$$C(k, i) = \sum_{j=0}^{N+P+1} S(j-k) S(j-i) \quad \begin{matrix} 0 \leq i \leq P \\ 1 \leq k \leq P \end{matrix} \quad (\text{IV. 18})$$

ou encore :

$$C(k, i) = \sum_{j=0}^{N-1-(k-i)} S(j) S(j+k-i) \quad (\text{IV. 19})$$

Nous remarquons que $C(k, i)$ est identique à la fonction d'auto-corrélation définie pour les intervalles de temps très courts et évaluée pour $(k-i)$:

$$C(k, i) = R(k-i) \quad (\text{IV. 20})$$

où

$$R(l) = \sum_{j=0}^{N-1-l} S(j) S(j+l) \quad (\text{IV. 21})$$

et puisque $R(l)$ est une fonction paire (c.e), $R(-l) = R(l)$, il vient :

$$R(l) = \sum_{j=0}^{N-1-|l|} S(j) S(j+|l|) \quad \begin{matrix} 1 \leq k \leq P \\ 0 \leq i \leq P \end{matrix} \quad (\text{IV. 22})$$

et $C(k, i) = R(k-i)$

et par suite, l'équation (IV. 10) s'écrit :

$$\sum_{i=1}^P a_i R(k-i) = R(k) \quad \begin{matrix} 1 \leq k \leq P \\ 0 \leq i \leq P \end{matrix} \quad (\text{IV. 23})$$

ce qui peut s'exprimer sous forme matricielle :

$$\begin{bmatrix} R(0) & R(1) & \dots & R(P) \\ R(1) & R(0) & & \\ \vdots & & \ddots & \\ R(P-1) & \dots & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} R(1) \\ \vdots \\ R(P) \end{bmatrix} \quad (\text{IV. 24})$$

La matrice " $R(p)$ " des valeurs d'autocorrélation est symétrique, définie positive et telle que les éléments situés de part et d'autres de la diagonale sont égaux.

Elle est par conséquent, de Toeplitz. Grâce à cette propriété, plusieurs procédures récursives ont été conçues pour résoudre le système d'équation (IV. 23) et la plus efficace est celle de l'algorithme de Durbin.

Notons que de la même manière que pour L_2 , on peut envisager les méthodes d'autocorrélation et de covariance, pour le critère L_4 .

Celles-ci sont basées sur des hypothèses similaires à celles vues pour L_2 et portent sur l'erreur définie dans (IV. 7).

IV. 2.2. Méthode du filtre en treillis .

IV.2.2.1 - généralités .

Dans le paragraphe précédent, nous avons vu deux méthodes qui ont été utilisées avec un grand succès dans le traitement de la parole. Dans ce qui suit, nous aborderons une autre méthode, appelée formulation du filtre en treillis. Nous en ferons l'étude selon le critère L_4 . Celui-ci est basé sur l'algorithme de Burg, qui sera détaillé plus loin.

La fonction de transfert du modèle de production du signal de la parole est :

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{c=1}^p a_c z^c} \quad (\text{IV. 25})$$

Le filtre $A(z)$ peut être donné par une structure en treillis (fig. IV. 2). Les équations de ce filtre sont :

$$f_i(j) = f_{i-1}(j) + \kappa_i b_{i-1}(j-1) \quad (\text{IV. 26})$$

$$b_i(j) = b_{i-1}(j-1) + \kappa_i f_{i-1}(j) \quad (\text{IV. 27})$$

$$f_0(j) = b_0(j) = s(j) \quad (\text{IV. 28})$$

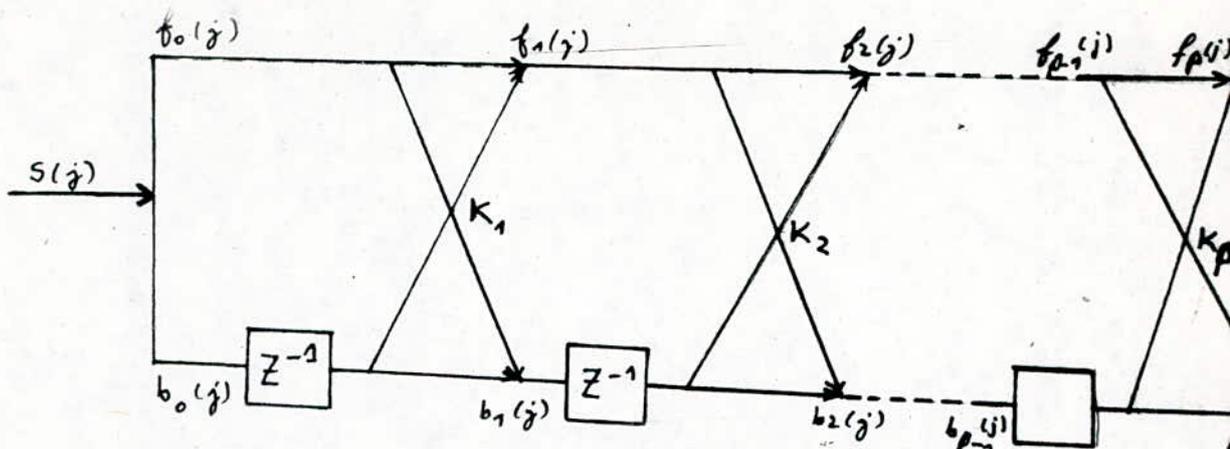


Fig. IV. 2 : filtre inverse en treillis .

Les équations (IV. 26) et (IV. 27) expriment respectivement l'erreur précédente de prédiction à l'ordre " i " et l'erreur suivante de prédiction pour le même ordre .

L'équation (IV. 28) exprime l'utilisation d'un prédicteur d'ordre zéro (ou la non utilisation de prédicteur) .

IV. 2.2.2. L'algorithme de Burg .

Parmi les procédures de la méthode du filtre en treillis , on distingue en particulier , celle de Burg , qui est la plus efficace pour cette formulation .

L'algorithme de Burg fait partie d'un ensemble de méthodes qui toutes utilisent le critère des moindres carrés . Il semble que l'on n'ait pas accordé suffisamment d'attention aux structures en treillis avec le critère de la valeur absolue ; pourtant il y a eu une suggestion dans ce sens , par Claerbout et Muir , depuis 1973 .

Un nouvel algorithme Burg L 4 qui ne présente pas les défauts de ses prédécesseurs est apparu et fera l'objet d'une étude et d'une application au paragraphe suivant .

IV. 3. Détermination des coefficients de réflexion à l'aide de la méthode du filtre en treillis suivant L₁ .

Les coefficients " k_i " qui apparaissent (fig. IV. 2) sont appelés coefficients de réflexion . Ils sont dits aussi coefficients de corrélation partielle ou encore " parcor " .

Cette appellation est justifiée par le fait que la relation donnant " k_i " (cf. éq. IV. 46) indique le degré de corrélation existante entre l'erreur précédente et l'erreur suivante .

IV. 3.1. position du problème .

L'algorithme de Burg avec le critère L₁ choisit pour la cellule d'ordre " i " , le coefficient de réflexion qui minimise la somme des valeurs absolues des erreurs précédentes et suivantes de prédiction , à la sortie de cette cellule .

$$L_1(\kappa_i) = \sum_j |f_i(j)| + |b_i(j)| \quad (\text{IV. 29})$$

En substituant (IV. 26) et (IV. 27) dans (IV. 29) , il vient :

$$L_1(\kappa_i) = \sum_j |f_{i-1}(j) + \kappa_i b_{i-1}(j-1)| + \sum |b_{i-1}(j-1) + \kappa_i f_{i-1}(j)| \quad (\text{IV. 30})$$

En supprimant , pour chaque cellule , l'indice " i " , pour plus de simplicité , le problème peut alors être reformulé de la façon suivante :

$$L_1 = \sum_j |b(j-1)| \left| \left(\frac{f(j)}{b(j-1)} + \kappa \right) \right| + |f(j)| \left| \left(\frac{b(j-1)}{f(j)} + \kappa \right) \right| \quad (\text{IV. 31})$$

ou encore :

$$L_1 = \sum_j w(j) |g(j) - \kappa| \quad (\text{IV. 32})$$

Le coefficient de réflexion " K " qui minimise L_1 est la médiane (cf IV. 3.2) des fractions :

$$g(j) = - \frac{f(j)}{b(j-1)} \text{ ou } \left(- \frac{b(j-1)}{f(j)} \right) \quad (\text{IV. 33})$$

avec leurs points respectifs :

$$w(j) = |b(j-1)| \text{ ou } |f(j)| \quad (\text{IV. 34})$$

Notons que dans le cas du critère étudié (L_1) , il est aisé de prouver par l'absurde que la magnitude des coefficients de réflexion " κ_i " est inférieure à l'unité et par conséquent le filtre modèle considéré est ainsi toujours stable .

IV. 3.2. Détermination des " κ_i " par la recherche de la médiane .

Dans cette étude , on utilise les notions de bloc de parole analysé (désigné par " N ") , de cellule (" I ") et d'échantillon " J " . Afin de tenir compte , de tous ces paramètres , nous reformulons les équations (IV. 26) , (IV. 27) , (IV. 28) , (IV. 33) et (IV. 34) de la façon suivante :

$$f(N, I, J) = f(N, I-1, J) + \kappa(I, N) b(N, I-1, J-1) \quad (\text{IV. 35})$$

$$b(N, I, J) = b(N, I-1, J-1) + \kappa(I, N) f(N, I-1, J) \quad (\text{IV. 36})$$

$$f(N, 0, J) = b(N, 0, J) = S(N, J) \quad (\text{IV. 37})$$

$$Q(N, I, J) = \frac{-f(N, I, J)}{b(N, I, J-1)} \quad (\text{IV. 38})$$

$$W(N, I, J) = |f(N, I, J)| \quad (\text{IV. 39})$$

IV. 3.2.1. Recherche de la médiane .

La valeur qui minimise la fonction

$$L_1 = \sum_j W(N, I, J) |Q(N, I, J) - a| \quad (\text{IV. 40})$$

est dite médiane $K(N, I, J)$ [cf. IV. 1.2.2] des observations $Q(N, I, J)$ que pondèrent les $W(N, I, J)$. Ce problème a été résolu pour la première fois par Laplace, et la solution $K(N, I, J)$, satisfait la condition de Laplace :

$$\sum_j W(N, I, J) \geq \left| \sum_{Q(N, I, J) > q} W(N, I, J) - \sum_{Q(N, I, J) < q} W(N, I, J) \right| \quad (\text{IV. 41})$$

(avec $q = K(N, I, J)$)

En d'autres termes, la distribution des poids sur l'un des côtés de la médiane doit être aussi proche que possible de l'équilibre. Quand l'inégalité (IV. 41) est stricte, la solution est unique.

La condition de Laplace peut aussi être réécrite de la façon suivantes :

$$\sum_j W(N, I, j) - \sum_{Q(N, I, j) < q} W(N, I, j) - \sum_{Q(N, I, j) = q} W(N, I, j) - 2 \sum_{Q(N, I, j) > q} W(N, I, j) > 0 \quad (\text{IV. 42})$$

$$\sum_j W(N, I, j) - \sum_{Q(N, I, j) < q} W(N, I, j) - \sum_{Q(N, I, j) = q} W(N, I, j) - 2 \sum_{Q(N, I, j) > q} W(N, I, j) \leq 0 \quad (\text{IV. 43})$$

(IV. 42) et (IV. 43)

Ces expressions suggèrent une méthode pour trouver la médiane :

Sommer les poids $W(N, I, J)$ pour les observations $Q(N, I, J)$ qui croissent jusqu'à ce que la somme partielle devienne égale ou supérieure à la moitié de la somme totale.

Quand un estimateur "qo" de la solution est disponible, la recherche de la médiane est plus facile. Par exemple, dans le cas où "q" est plus grand que "qo", les conditions (IV. 42) et (IV. 43) sont équivalentes aux

$$\sum_j W(N, I, J) < \frac{1}{2} \sum_j W(N, I, J) \quad (\text{IV. 42})$$

$$\sum_j W(N, I, J) \geq \frac{1}{2} \sum_j W(N, I, J) \quad (\text{IV. 43})$$

Les trois premiers termes de ces inégalités sont constants et peuvent être calculés une fois pour toutes. La médiane est alors obtenue, en soustrayant deux fois les poids $w(N, I, J)$, obtenus en prenant les observations $q(N, I, J)$ dans un ordre croissant, et en commençant par "q₀", jusqu'à ce que le résultat soit négatif ou nul, si l'on considère la condition (IV. 45).

IV.3.2.2. Commentaire .

Les principaux pas de l'algorithme utilisé (dont l'organigramme et le programme sont donnés en annexe) consistent en l'évaluation des fractions $q(N, I, J)$, leur triage et la sommation des poids $w(N, I, J)$ avec les fractions $q(N, I, J)$ croissant jusqu'à ce que la somme partielle des poids devienne supérieure ou égale à la moitié de la somme totale.

Le coefficient de réflexion " $K(N, I)$ ", autrement dit la solution du problème, prend alors la valeur de la fraction $q(N, I, J)$, la plus proche de l'équilibre des poids $w(N, I, J)$.

Notons que l'on dispose de deux simplifications intéressantes dans notre cas.

Premièrement, puisque l'algorithme de Burg produit toujours un coefficient de réflexion compris entre -1 et 1, il est inutile de considérer individuellement les fractions $q(N, I, J)$ au-delà de ces limites. Pour chaque échantillon il suffit d'évaluer une seule des deux fractions $-f(N, I, J)$ et $-b(N, I, J-1) / f(N, I, J)$.

Deuxièmement, la recherche de la médiane est rendue rapide par la donnée d'un estimateur initial (cf IV. 3.2.1) constitué par le coefficient de réflexion du même ordre provenant d'une analyse de parole effectuée auparavant, soit $K(N-1, I)$. Comme les coefficients de réflexion ne varient pas brusquement (du fait de l'inertie du système phonatoire), cette modification limite le nombre de fractions $q(N, I, J)$ qui doivent être triées.

Pour cela, nous avons utilisé pour le calcul de $K(N, I)$ l'équation :

$$K(N, I) = \frac{2 \sum_{j=0}^{N-1} [f(N, I-1, j) \cdot b(N, I-1, j-1)]}{\sum_{j=0}^{N-1} [f(N, I-1, j)]^2 + \sum_{j=0}^{N-1} [b(N, I-1, j-1)]^2} \quad (\text{IV. 46})$$

(Cette équation a été établie pour l'algorithme de Burg utilisant le critère L_2).

La démarche suivie pour la détermination des coefficients de réflexion a été la suivante :

On détermine d'abord $f(1,0,J)$ et $b(1,0,J)$, i.e., les valeurs à l'ordre " I " du premier bloc de parole analysé, grâce à l'équation (IV. 37). On reporte ensuite dans l'équation (IV. 46), ce qui détermine $K(1,1)$. On passe alors au calcul de $f(1,1,J)$ et $b(1,1,J-1)$ grâce aux équations (IV. 35) et (IV. 36) ce qui permettra de déterminer $K(1,2)$ et ainsi de suite, par récurrence, on estime tous les $K(1,I)$ du premier bloc, ce qui nous permettra ensuite d'appliquer l'algorithme décrit dans le paragraphe précédent.

IV.4. Comparaison des principales méthodes L_1 et L_2 .

On peut établir une comparaison entre ces méthodes selon différents points de vue. Parmi ceux-ci, on peut citer la quantité de calculs, le fenêtrage et la stabilité du prédicteur (fig. IV. 3) :

IV.4.1. étude selon la quantité de calculs.

- la méthode d'autocorrélation L_2 est plus rapide que celle de L_1 .
- les méthodes de covariance et de Burg L_1 sont plus rapides que leurs correspondantes de L_2 .

IV.4.1. étude selon le fenêtrage.

- pour la méthode d'autocorrélation, la fenêtre est nécessaire aussi bien dans le cas L_1 que dans celui de L_2 .
- pour la méthode de covariance, elle est inutile dans les deux cas.
- pour la formulation de Burg, elle est conseillée pour le cas L_2 , et elle est tout simplement inutile pour celui de L_1 .

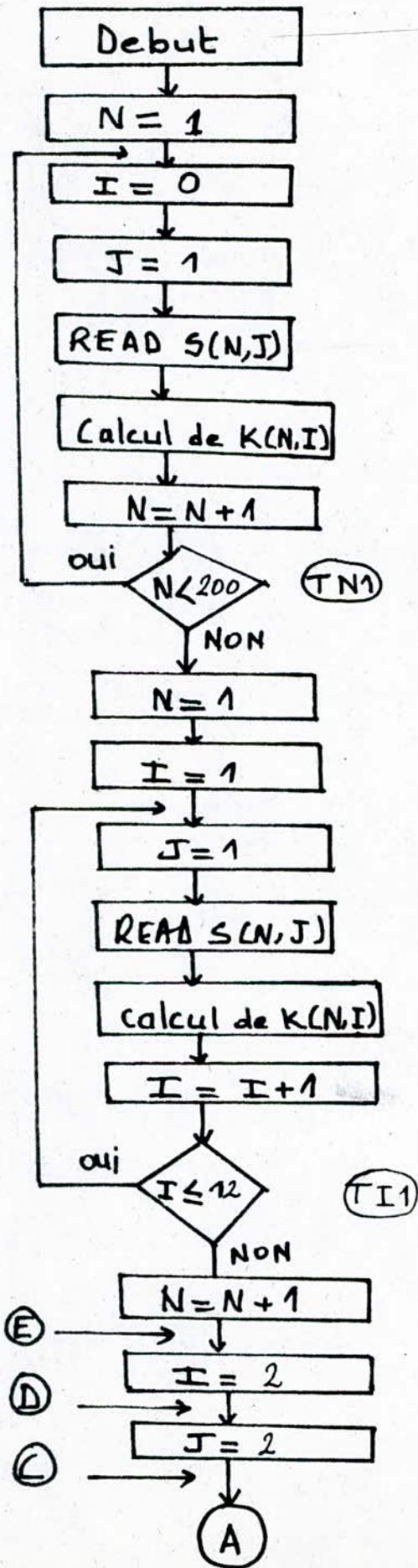
IV.4.3. étude suivant la stabilité du prédicteur.

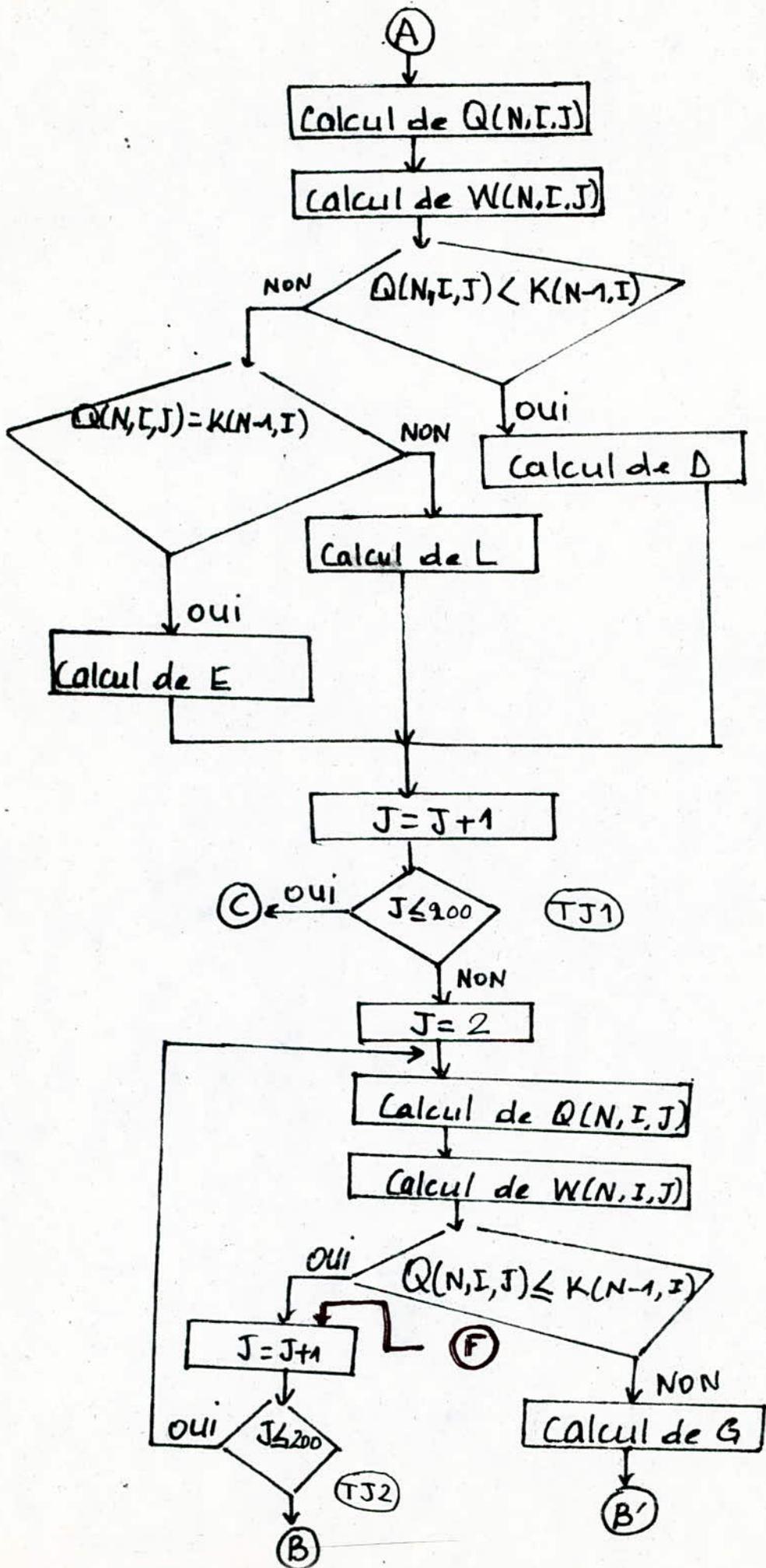
- pour la méthode de covariance, la stabilité n'est garantie ni pour L_1 ni pour L_2 .
- pour la méthode de corrélation, elle est garantie pour le cas L_2 mais pas pour L_1 .
- pour la formulation de Burg, la stabilité est garantie dans les deux cas.

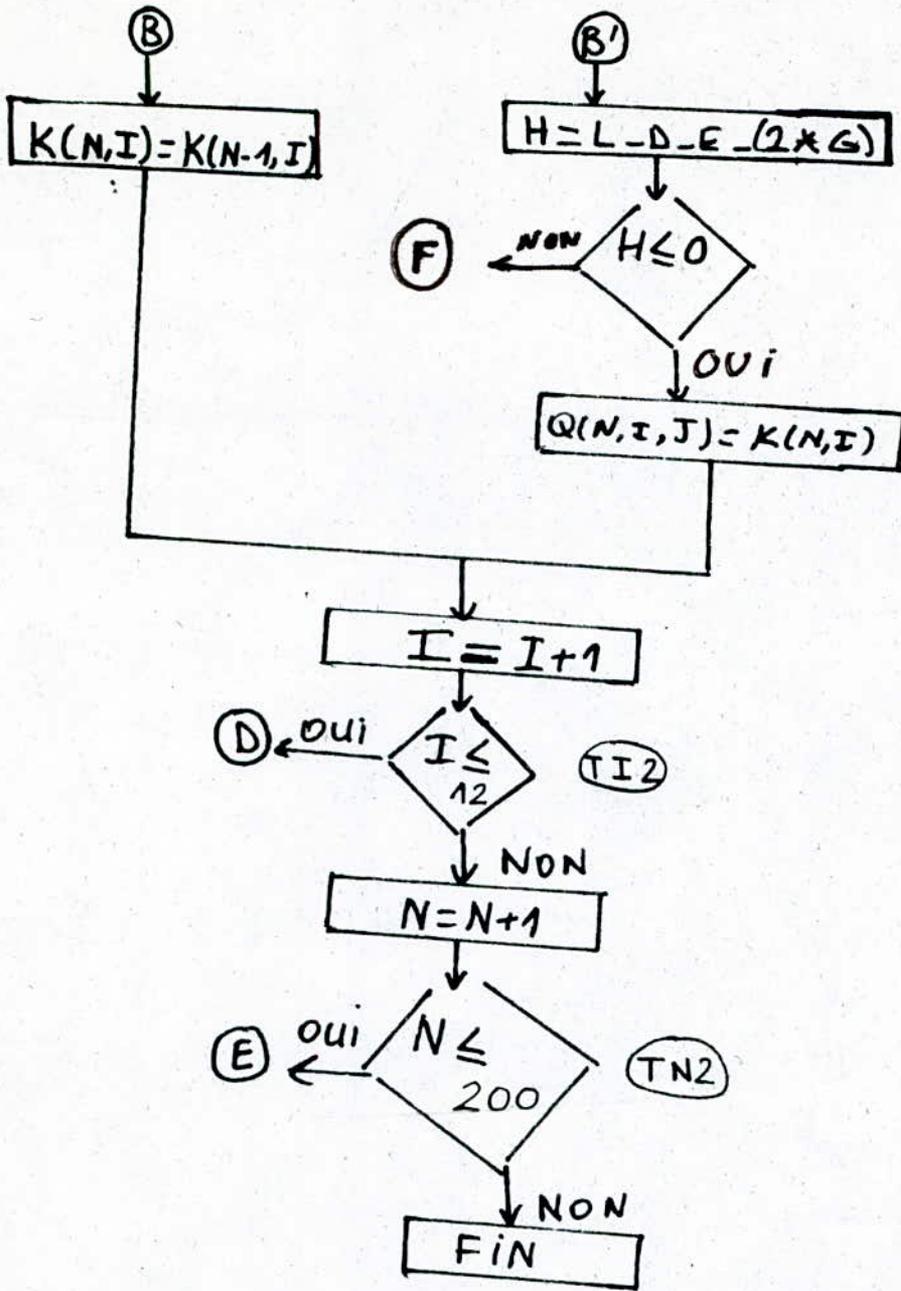
	Autocorrélation		covariance		Burg	
	L2	L1	L2	L1	L2	L1
quantité de calculs	$mp + p^2$	$mp^2 + p^3$	$mp + \frac{p^3}{6}$	mp^2	$5mp$	$2mp$
fenêtrage	Nécessaire		inutile		conseillé	inutil
stabilité	garantie	non garantie	non garantie		garantie	

Fig.(IV.3): Comparaison des méthodes L₁ et L₂.
 ("p" = ordre du filtre
 "m" = nombre d'échantillons)

organigramme







C O N C L U S I O N

Ce travail consistait en l'étude des principales méthodes d'analyse de la parole .

Pour cela , il était nécessaire de faire au préalable l'étude de l'appareil phonatoire et des différents sons de la parole . Puis nous avons tenté une approche générale des méthodes d'analyse .

En abordant ensuite l'analyse par les différents appareils, nous nous sommes rendu compte de la place privilégiée qu'occupe parmi eux le sonographe . Celui-ci présente , en effet des avantages (représentation plus proche de la réalité , perçue , possibilité de comparaison du comportement de différents paramètres pour une phrase donnée) qui l'emportent largement sur son manque de précision . D'ailleurs , rien n'empêche pour un phénomène donné , de commencer par la mise en place des grandes lignes de l'analyse , au moyen du sonographe , puis d'affiner les résultats à l'aide d'un appareil plus précis . On peut citer comme exemple , l'étude de la nasalité . On a remarqué que le sonogramme donnait plutôt un résultat qualitatif et que par conséquent il fallait compléter ceci par un moyen plus précis ce qui donne le spectre de bande , qui est très satisfaisant à ce point de vue (cf chap. III) .

Enfin , nous avons étudié une autre méthode de grande importance : la prédiction linéaire , que nous avons illustrée par la recherche de coefficients de réflexion à l'aide d'un algorithme de Burg selon le critère L_1 (ou algorithme Burg L_1) .

Dans cette application , nous avons rencontré les problèmes suivants :

- 1 - la détermination des valeurs initiales des coefficients de réflexion (valeurs du premier bloc de parole considéré) .

- 2 - le cas où le test TJ_2 (cf organigr.) ne fournit pas de valeur pour la médiane .

Pour cela nous avons proposé :

- 1- la détermination des " k_i " à l'aide de l'expression donnant ces paramètres dans l'algorithme de Burg L_2 conformément à l'équation (IV. 46).
- 2- Quand le test TJ_2 ne fournit pas de valeur de la médiane , nous avons retenu , pour celle-ci, sa valeur du même ordre

du bloc immédiatement précédent .

Ceci est justifié par le fait que les coefficients de réflexion ne varient pas brusquement .

Par ailleurs , nous avons écrit un programme pour cette application .

A l'utilisation sur mini-ordinateur VAX , dont la capacité est de 2,5 MO , il s'est avéré que l'espace-mémoire était insuffisant pour permettre les calculs demandés par le programme (pour lequel il faut réserver 2 MO de zone-mémoire) du fait que le mini-ordinateur est utilisé en multi-poste ce qui réduit considérablement l'espace-mémoire par poste .

Nous avons alors pensé à réduire le nombre de blocs à analyser de 200 à 3 , celui des cellules de 12 à 4 et enfin celui des échantillons à 140 .

Cette application montre l'intérêt de l'algorithme étudié . En effet , si les procédures L₁ anciennes se heurtaient à des problèmes de stabilité et de quantités de calcul , ceci n'est pas le cas pour ce nouvel algorithme qui montre une efficacité comparable à celle des méthodes " L₂ " classiques .

Nous espérons que cet algorithme soit le prélude à un regain d'intérêt pour le critère L₁ .

Comme première perspective , on peut envisager l'utilisation de cet algorithme à la détermination du pitch . Ceci paraît très intéressant dans la mesure où de nombreuses méthodes habituellement utilisées sont souvent onéreuses .

Enfin , ce travail a été pour nous riche d'enseignements surtout qu'il nous a permis d'approcher des domaines aussi intéressants les uns que les autres .

```

PROGRAM CALCUL (INPUT,OUTPUT);
LABEL 10,20,30,40,50,80;
TYPE LISTE=ARRAY[1..5,0..4,1..140] OF REAL;
VAR
  S          :ARRAY[1..3,1..140] OF REAL;
  K          :ARRAY[1..3,1..4] OF REAL;
  F,B,Q,W: LISTE;
  N,I,J,X1,N1,I1: INTEGER;
  TJ,T2,T3,T4,A1,A2,A3,A11,A22,A33,A,Z,M,Y,C,U,L,D,E,G,H:REAL;

```

```

BEGIN
  FOR N:=1 TO 3 DO
  BEGIN
    FOR J:=1 TO 140 DO
    BEGIN
      WRITE('DONNEZ LES VALEURS DE S( ',N, ', ',J, ' )');
      READ(S[N,J]);
    END;
  END;

```

```

  FOR N:=1 TO 3 DO
  BEGIN
    FOR J:=2 TO 140 DO
    BEGIN
      TJ:=S[N,J];
      F[N,0,J]:=TJ;
      B[N,0,J]:=TJ;
      END;
      A1:=0;
      A2:=0;
      A3:=0;
      FOR J:=2 TO 140 DO
      BEGIN
        A11:=(F[N,0,J])*(B[N,0,J-1]);
        A1:=A1+A11;
        END;
      FOR J:=2 TO 140 DO
      BEGIN
        A22:=(F[N,0,J])**2;
        A2:=A2+A22;
        END;
      FOR J:=2 TO 140 DO
      BEGIN
        A33:=(B[N,0,J-1])**2;
        A3:=A3+A33;
        END;
      T2:=(2*A1)/(A2+A3);
      K[N,1]:=T2;
    END;
  END;

```

N:=1;

FOR I:=1 TO 4 DO
BEGIN

FOR J:=2 TO 140 DO
BEGIN

K[N,I]:=T3;

X1:=I-1;

F[N,I,J]:=(F[N,X1,J])+(T3)*(B[N,X1,J-1]);
B[N,I,J]:=(B[N,X1,J-1])+(T3)*(F[N,X1,J]);

END;

I1:=I+1;

A:=0;

M:=0;

C:=0;

FOR J:=2 TO 140 DO
BEGIN

Z:=(F[N,I,J])*(B[N,I,J-1]);
A:=A+Z;

END;

FOR J:=2 TO 140 DO
BEGIN

Y:=(F[N,I,J])**2;
M:=M+Y;

END;

FOR J:=2 TO 140 DO
BEGIN

U:=(B[N,I,J-1])**2;
C:=C+U;

END;

T4:=(2*A)/(C+M);

K[N,I1]:=T4;

END;

```

FOR N:=2 TO 3 DO
BEGIN
  30: I:=1;
  N1:=N-1;
D:=0;
E:=0;
L:=0;
G:=0;
  40 : FOR J:= 2 TO 140 DO
  BEGIN
    10 : Q[N,I,J]:= -F[N,I,J]/B[N,I,J-1];
    W[N,I,J]:=ABS(F[N,I,J]);
    IF Q[N,I,J]<K[N1,I] THEN
      D:=D+W[N,I,J]
    ELSE BEGIN
      IF Q[N,I,J]=K[N1,I] THEN
        E:=E+W[N,I,J]
      ELSE
        L:=L+W[N,I,J]
      END;
    END;

    J:=2;

  20: Q[N,I,J]:=-F[N,I,J]/B[N,I,J-1];
    W[N,I,J]:=ABS(F[N,I,J]);
    IF Q[N,I,J]<=K[N1,I] THEN GOTO 80
    ELSE
    BEGIN
      G:=G+W[N,I,J];
      H:=L-D-E-(2*G);
      IF H<=0 THEN
      BEGIN
        Q[N,I,J]:=K[N,I];
        END
      ELSE GOTO 80
    END;
    80: J:=J+1;
    IF J<=140 THEN GOTO 20
    ELSE BEGIN
      K[N,I]:=K[N1,I];
    I:=I+1;
    IF I<=4 THEN GOTO 40
    ELSE GOTO 50
  END;
  50:END
END.

```

IL SUFFIT D'AVOIR TOUTES LES VALEURS DE S(N,J) POUR
 ATTEINDRE LE BUT DE NOTRE PROGRAMME .

SELLAM et NOUR
 ~~~~~

## B I B L I O G R A P H I E

---

- AUVRAY , J. " Electronique des signaux échantillonnés  
et numériques " DUNOD , 1979 .
- CUREAU , J. , VULETIC , B. " Enseignement de la prononciation "  
DIDIER , 1976
- EMERIT , E. " Cours de phonétique acoustique "  
SNED , 1977 .
- DENOEL , E. , SOLVAY, J.P. " Linear prediction of speech with  
a least absolute error criterion "  
IEEE Trans. on acoustics , speech and  
signal processing , vol. AssP-33 ,  
N° 6 , dec. 85 .
- GUERTI , M. " Contribution à la synthèse de la parole  
en arabe standard "  
Thèse de magister , 1984.
- GUIBERT , J. " La parole. Compréhension <sup>en</sup> et synthèse par  
les ordinateurs "  
P.U.F , 1979
- LIENARD , J.S " Les processus de la communication parlée.  
Introduction à l'analyse et à la synthèse  
de la parole " - MASSON , 1977 -
- MALMBERG , B. " La phonétique "  
Que sais-je . P.U.F , 1975 .
- MATRAS , J.J " Techniques de l'ingénieur " ( Electronique )  
article E. 2602 .
- METTAS , O. " Les techniques de la phonétique expérimentale  
et l'intonation "  
Presses universitaires de Bruxelles , 1971 .
- MOLES , A. , VALLENCIEN , B. " Phonétique et phonation "  
MASSON , 1966 .
- RABINER , L.R. , SCHAFER , R.W. " Digital processing of  
speech signals "  
PRENTICE HALL , 1978 .