

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT : Electronique

PROJET DE FIN D'ETUDES

SUJET

TRAITEMENT DU SIGNAL
DE LA PAROLE
MESURE DE LA FREQUENCE
FONDAMENTALE (Pitch)

Proposé par :
Mr B. BOUSSEKSOU

Etudié par :
A. ALLAQUI
F. RAGUEB

Dirigé par :
Mr B. BOUSSEKSOU



PROMOTION : Juin 1984

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

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

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT :

PROJET DE FIN D'ETUDES

SUJET

TRAITEMENT DU SIGNAL
DE LA PAROLE
MESURE DE LA FREQUENCE
FONDAMENTALE (Pitch)

Proposé par :
Mr B. BOUSSEKSOU

Etudié par :
A. ALLAOUI
F. RAGUEB

Dirigé par :
Mr B. BOUSSEKSOU



PROMOTION :



— REMERCIEMENTS —

- * Nous tenons à remercier notre promoteur Monsieur Bousséksou de nous avoir proposé le sujet et de nous avoir guidés tout au long de notre travail.
 - * Nous tenons à remercier également Mademoiselle Moussaoui pour l'aide précieuse qu'elle nous a fournie en nous procurant une documentation qui nous a été très utile.
 - * Nous sommes très sensible à l'aide précieuse que nous a apporté Monsieur ZERRARI.
 - * Nous remercions également tous ceux qui ont contribué à notre formation au sein de l'école nationale polytechnique.
-

DEDICACES

À la mémoire de ma grande-mère Barka.

À mon père, ma mère.

À mes frères et sœurs.

À toute ma famille.

À tous mes amis.

Je dédie ce travail.

- A. Abdelhamid -

À la mémoire de ma grande-mère ALdjia.

À mon père, ma mère pour leur sacrifice.

À mon oncle Amar et sa femme.

À mes frères et sœurs; cousins et cousines.

À tous (tes) mes amis(ies)

- R. Fouad -

SOMMAIRE

PAGES

Introduction	1
Chapitre 1: Production naturelle de la parole	3
1. Définition.	3
1.1. L'appareil vocal	3
1.1.1. Description	3
1.1.2. Fonctionnement	3
1.1.3. Types de sons	4
1.2. La parole phénomène physique.	7
1.2.1. Introduction.	7
1.2.2. Caractéristiques de la parole.	7
Chapitre 2: Détection du pitch "méthode spectrale"	9
2.1 Analyse spectrale.	9
1.1. Transformée de Fourier.	9
1.2. Série de Fourier.	9
1.3. Intégrale de Fourier.	10
1.4. Puissance, énergie des signaux	11
1.5. Spectre de puissance d'un signal	11
1.6. Traitement numérique.	11
1.6.1. Échantillonnage.	11
1.6.2. Transformée Discrète de Fourier "DFT"	12
1.6.3. Transformée Rapide de Fourier "TRF"	13
2.2. Traitement homomorphique des signaux.	15
2.2.1. Généralités.	15
2.2.2. Forme canonique.	16
2.2.3. Système homomorphique convolutif	16

2.2.4 . Notion de Cepstre.	18
2.3 . Estimation du pitch par la méthode Cepstrale.	19
2.3.1. Calcul numérique du Cepstre.	23
2.3.2. Conclusion.	26
Chapitre 3. Methodes de décalage.	27
3.1. Généralités.	27
3.2 . fonction d'autocorrelation.	27
3.2.1. propriétés.	28
3.3. fonction d'autocorrelation à court terme.	28
3.4. Estimation du pitch.	31
3.5. méthode des différences.	40
3.5.1. Conclusion.	41
Chapitre 4.	42
4.1. Introduction.	42
4.2. filtrage.	43
4.2.1. filtres actifs.	43
4.2.2. Principaux types de filtres.	43
4.2.3. fonction de Butterworth.	43
4.2.4. fonction de Tchebitcheff	45
4.2.5. Détermination pratique d'un filtre.	45
4.2.6. Calcul des éléments passifs:	46
4.2.7. Calcul du filtre passe haut "Ordre 4"	47
4.2.8. Calcul du filtre passe bas "Ordre 2"	48
4.2.9. filtre passe bande.	48
4.3. Compression.	52
4.3.1. fonctionnement.	53
4.3.2. le décompteur.	55

4.4.3. Calcul de la période d'horloge.	56
4.3.4. fonctionnement du décompteur.	56
4.4. Le détecteur de crêtes.	59
4.4.1. fonctionnement.	59
4.4.2. description et calcul du détecteur de crêtes.	62
4.4.3. Alimentation.	63.
4.4.4 Manipulation	71

Conclusion

Bibliographie.

TABLE DES FIGURES

	PAGES
figure. 1-a. coupe du conduit vocal	5
figure. 1-b. schematisation du conduit vocal	5
figure. 2-a production de son voisé	6
" 2-b production de son non voisé.	6
" 2.1 Forme canonique d'un système homomorphe	17
" 2.2 " " " " " " Convolutif	17
" 2.3 Forme canonique dans le domaine temporel.	17
" 2.a modèle de base pour la production de la parole.	21
" 2.b Cepstre.	21
" 1 fonction d'auto correlation "fenêtre rectangulaire"	33
" 2 " " " " " " Hamming"	33
" 3 " " " " " " pour différents échantillons.	35
" 4 échantillons inclus dans le calcul de la fct d'auto.	38
" 5 fonction d'auto correlation modifiée.	39
" 6 " " " " " " pour différents échantillons.	39
" 7 fonction de différences.	41
" 4.0 Courbe de réponse de Butterworth.	44
" 4.1 Bloc préfiltrage.	49
" 4.2 Courbe de réponse du filtre passe haut	50
" 4.3 " " " " " " bas	50
" 4.4 " " " " " " bloc préfiltrage.	51
" 4.5 schéma fonctionnel du compresseur.	52
" 4.6 le bloc compresseur.	58
" 4.7 synoptique du détecteur de crêtes.	60
" 4.8 bloc détecteur de crêtes.	64

fig 4.9. schema électrique de la maquette.	65.
fig 4.10. circuits intégrés utilisés.	67
fig 4.11. Circuit imprimé du bloc filtrage.	68
fig 4.12. Circuit imprimé du bloc compression	69
fig 4.13. Circuit imprimé du détecteur de crêtes.	70
fig 4.14. Oscillogrammes.	72
fig 4.15. " " "	73

INTRODUCTION

Depuis fort longtemps le domaine du traitement de signal de la parole a suscité d'énormes recherches notamment dans le domaine de l'extraction de la fréquence fondamentale, celle-ci étant qualitativement une caractéristique subjective de la voix à laquelle on attribue une valeur dans une gamme qui s'étend des voix dites basses aux voix dites hautes; c'est un phénomène quasi-périodique et ce sont non seulement la période et l'amplitude de l'onde source qui varient mais également sa forme.

C'est pour cela que l'opération de détection et de mesure du fondamental reste une chose difficile or l'analyse et la restitution de cette fréquence sont des facteurs déterminants pour l'obtention d'une parole de bonne qualité à l'aide de "vocodeur" c'est à dire que le naturel de la parole synthétique dépend beaucoup de la justesse et de la finesse de l'analyse du pitch.

C'est pour cela que de nombreuses méthodes ont été, sont et seront proposées et ces différents procédés évoluent dans le temps en fonction des progrès de la technologie.

Ainsi de nombreuses classifications sont possibles, elles seront toujours un peu arbitraires tant est

grand le foisonnement et l'interpénétration des différents procédés.

Ainsi dans notre travail on a opté pour une classification qui consiste en une étude des méthodes spectrales, décalage temporel et enfin les méthodes de détection d'événements.

CHAPITRE 1

PRODUCTION NATURELLE DE LA PAROLE

1/ Définition

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

1.1/ L'appareil vocal:

1.1.1/ Description:

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

1.1.2/ Fonctionnement:

Lorsque nous parlons, l'air venant de nos poumons passe dans le larynx qui contient les cordes vocales, dont les vibrations excitent la masse d'air du pharynx, des cavités buccale et nasale.

Ces organes en se modifiant créent des fréquences multiples ou harmoniques, de →

la fréquence de vibrations des cordes vocales. Elles se distinguent par leurs nombres, leurs rangs ainsi que leurs intensités.

Pour cela on donne sur la "figure 1-a" une coupe du conduit vocal et sur la "figure 1-b" une schématisation du conduit vocal.

1.1.3/ Types de sons

Vu le fonctionnement de l'appareil vocal on peut établir deux catégories de sons, selon qu'ils sont dûs

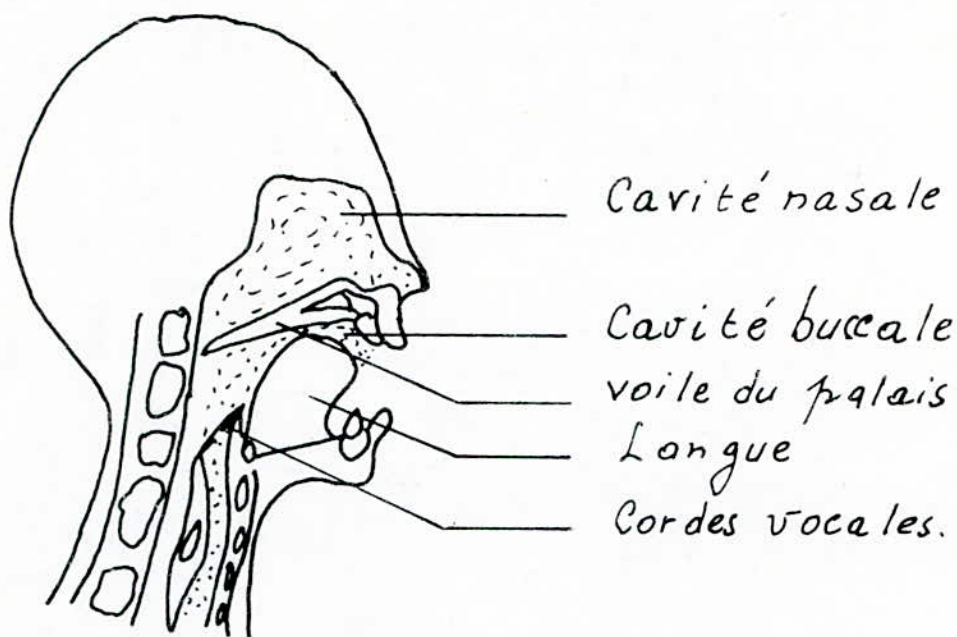
- A une vibration des cordes vocales, ce sont des "sons voisés" ou Sonores.
- A une génération de bruit à travers une constriction du conduit vocal, ce sont des sons "non voisés" ou Sourds.

Dans la première catégorie de sons, les variations du débit d'air au niveau des cordes vocales au cours du temps sont sensiblement triangulaires de période T_0 appelé aussi "pitch".

Du point de vue spectral

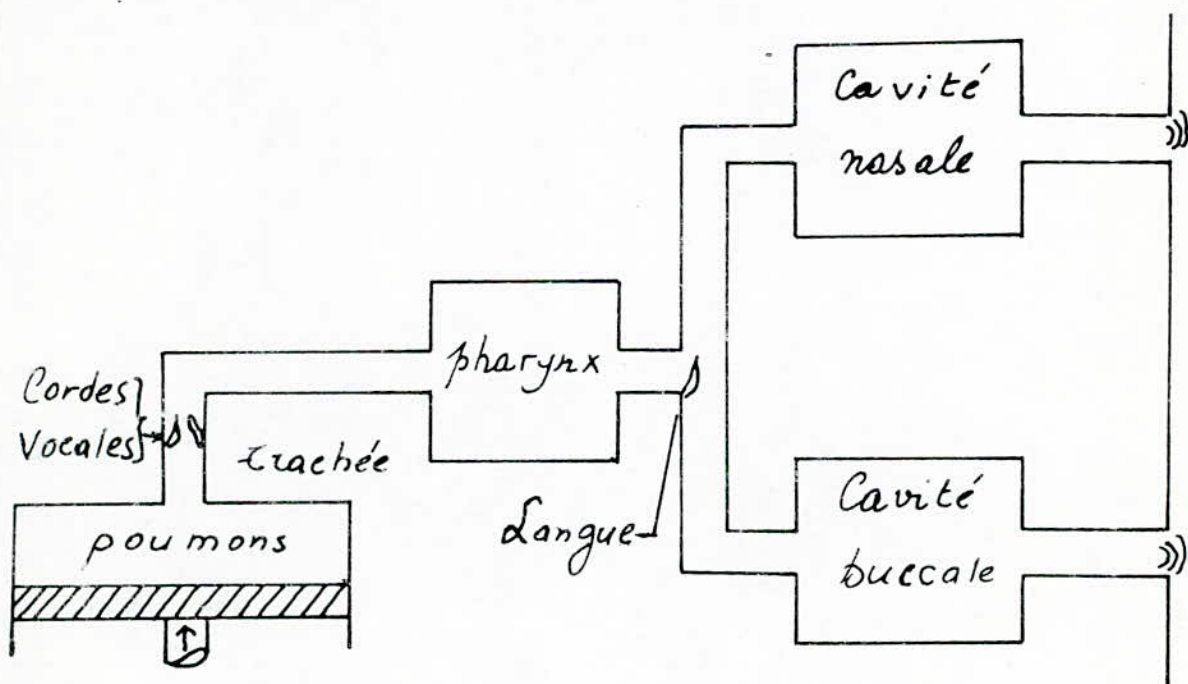
Ceci est représenté sur la "figure 2-a"

Dans le second cas où il y a absence de vibration des cordes vocales, le signal est a périodique. Son spectre est relativement uniforme sur une large bande de fréquence. - "figure 2-b"



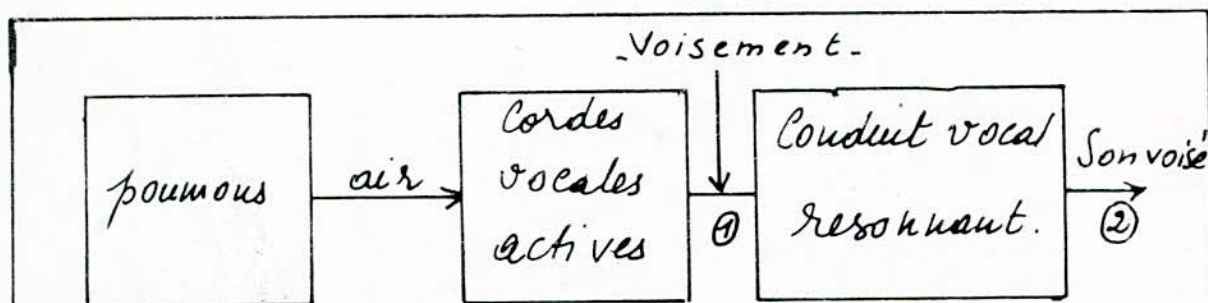
- Coupe du Conduit vocal -

- fig 1. a -

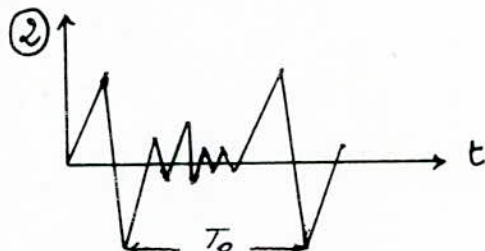
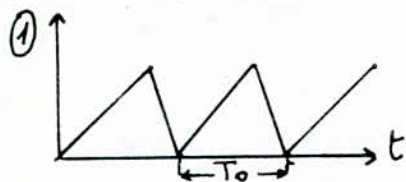


- Schématisation du Conduit vocal -

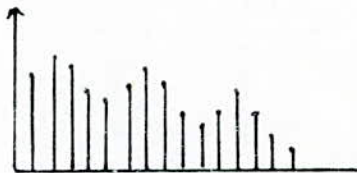
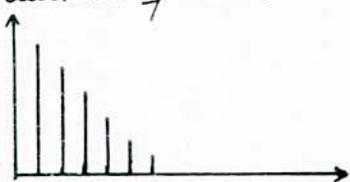
- fig 1. b -



Signaux en fonction du temps.

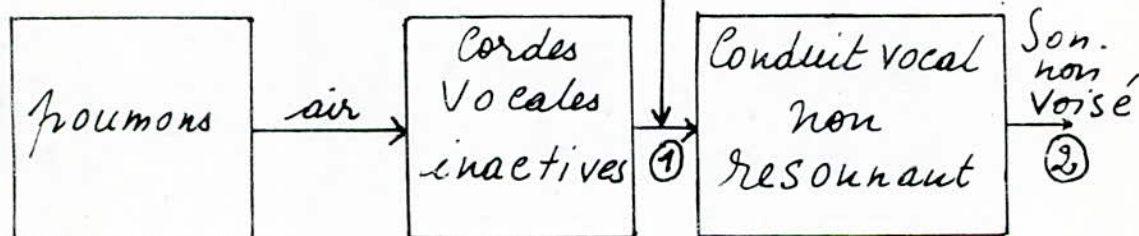


Signaux en fonction de la fréquence.

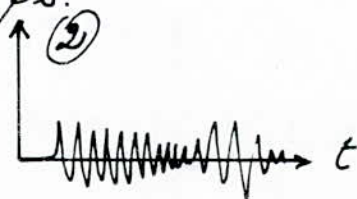
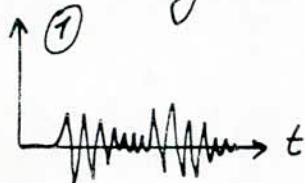


- fig 2 - a -

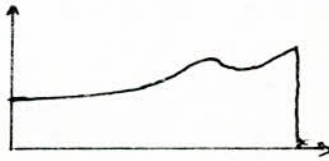
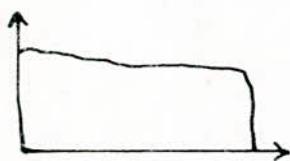
- Souffle -



Signaux en fonction du temps.



Signaux en fonction de la fréquence.



- fig - 2 - b -

1.2 / La parole, phénomène physique.

1.2.1 / Introduction.

Physiquement la voix est schématiquement caractérisée par deux paramètres, d'une part par la fréquence des cordes vocales que l'on appelle mélodie de la voix et d'autre part l'intensité de ces vibrations.

1.2.2 / Caractéristiques de la parole.

Les principales caractéristiques de la parole sont:

- Niveau =

Compte tenu des instants de silence l'énergie moyenne développée au cours d'une conversation est de l'ordre de $10 \mu W$.

Il y a cependant des différences selon les individus pour lesquels cette moyenne peut varier de $1 \mu W$ à $500 \mu W$.

- timbre =

Il peut être considéré comme un paramètre du son, il se caractérise par l'amplitude harmonique. Chaque harmonique a son amplitude propre, donc pour retrouver l'intégrité du timbre il faudra que chaque harmonique soit respectée.

- Bande passante =

Quand les fréquences entre 100 Hz et 10 KHz sont correctement rendues, on peut admettre qu'une

transmission de parole est fidèle.

- Dynamique =

On définit la dynamique d'un signal comme étant l'écart en (dB) entre le niveau maximum et le niveau minimum de ce signal.

$$D_s = (N_{\max} - N_{\min}) \text{ (dB)}$$

globalement l'énergie de la parole varie au plus entre $1/10$ et $1000 \mu\text{W}$, soit une dynamique de 40 dB.

- hauteur =

La fréquence la plus basse dans le signal parole est la fréquence fondamentale de vibrations des cordes vocales, elle peut varier de 90 Hz à 120 Hz pour une voix d'homme et jusqu'à 220 Hz pour une voix de femme.

CHAPITRE 2

DETECTION DU PITCH

METHODE SPECTRALE

2.1 Analyse Spectrale =

1.1/ Transformée de Fourier =

La transformée de Fourier constitue un outil d'analyse dans différents domaines notamment les télécommunications, la théorie des systèmes et les processus aléatoires. Un des domaines concerne le traitement des signaux et le calcul de leur spectre.

1.2/ Série de Fourier =

Soit une fonction périodique de période T

$$x(t) = x(t+T)$$

Elle peut être décomposée en une somme de sinusoides harmoniques.

$$x(t) = \frac{a_0}{2} + \sum \left(a_n \cos 2\pi n f_0 t + b_n \sin 2\pi n f_0 t \right)$$

avec :

$$a_n = \frac{2}{T} \int_{-T/2}^{+T/2} x(t) \cos 2\pi n f_0 t \, dt$$

$$b_n = \frac{2}{T} \int_{-T/2}^{+T/2} x(t) \sin 2\pi n f_0 t \, dt$$

le spectre de fréquence sera alors défini comme suit :

$$X(n f_0) = \frac{1}{2} (a_n - j b_n) = \frac{1}{T} \int_{-T/2}^{+T/2} x(t) e^{-j 2\pi n f_0 t} \, dt$$

Celui peut être décomposé en deux spectres :

a) Spectre d'amplitude :

$$|X(n f_0)| = \frac{1}{2} \sqrt{a_n^2 + b_n^2}$$

b/ Spectre de phases:

$$\varphi(nf_0) = \text{Arctg} \left(-\frac{b_n}{a_n} \right)$$

reciproquement:

$$\begin{cases} X(f) = \sum_{-\infty}^{+\infty} |X(nf_0)| e^{-j\varphi(nf_0)} \\ x(t) = \sum_{-\infty}^{+\infty} e^{j2\pi f_0 t} \frac{1}{T} \int_{-T/2}^{T/2} x(\sigma) e^{-j2\pi f_0 \sigma} d\sigma \end{cases}$$

Il est important de remarquer que le spectre d'une fonction périodique de période T se compose de raies dont l'écart minimal (sur axe fréquentiel) est $f_0 = 1/T_0$, c'est à dire que ce spectre serait discontinu, il n'existerait que pour les valeurs de fréquences multiples de f_0 .

1.3/ Intégrale de Fourier.

Du point de vue de la communication, la série de Fourier ne présente pas un grand intérêt vu qu'elle est appliquée à des signaux périodiques. Ainsi l'intégrale de Fourier permet l'analyse de n'importe quel signal (ie a-périodique) tout en éliminant le temps au profit de fréquence.

$$\begin{cases} x(t) = \int_{-\infty}^{+\infty} e^{j2\pi ft} df \int_{-\infty}^{+\infty} x(\sigma) e^{-2\pi jf\sigma} d\sigma \\ X(f) = \int_{-\infty}^{+\infty} x(t) e^{-2\pi jft} dt \end{cases}$$

On écrit généralement $x(t) \Leftrightarrow X(f)$

$X(f)$ est généralement complexe.

$$\begin{cases} \text{Re}[X(f)] = \int_{-\infty}^{+\infty} x(t) \cos 2\pi ft dt \\ \text{Im}[X(f)] = \int_{-\infty}^{+\infty} x(t) \sin 2\pi ft dt \end{cases}$$

Le spectre d'amplitude sera alors.

$$|X(f)| = \sqrt{(\operatorname{Re}[X(f)])^2 + (\operatorname{Im}[X(f)])^2}$$

Le spectre de phase.

$$\varphi(f) = \operatorname{Arctg} \left(-\frac{\operatorname{Im}[X(f)]}{\operatorname{Re}[X(f)]} \right)$$

1.4/ Puissance, énergie des signaux:

Dans une quelconque mesure, l'objet de celle-ci subit toujours un prélèvement d'énergie de la part du dispositif de mesure, ainsi cette notion de puissance est importante.

1.5/ Spectre de puissance d'un signal:

La puissance d'un signal en fonction de la fréquence est notée $S_{xx}(f)$

$$\text{Si } x(t) \rightleftharpoons X(f)$$

on aura alors:

$$S_{xx}(f) = |X(f)|^2$$

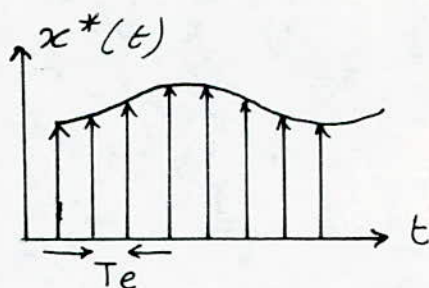
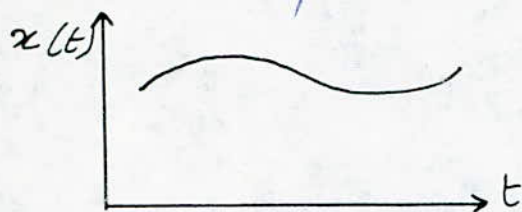
donc: $S_{xx}(f)$ est réelle positive.

1.6/ Traitement numérique:

1.6.1/ Échantillonnage:

Pour appliquer les méthodes de traitement du signal numérique sur un signal analogique, il faut représenter ce signal sous forme d'une séquence de nombres, ceci est obtenu par échantillonnage du signal analogique.

Théorème =



Le signal échantillonné $x^*(t)$ est obtenu à partir de la fonction $x(t)$ échantillonnée à la fréquence $f_e = 1/T_e$ par le train d'impulsions de Dirac $\delta(t-nT)$

$$x(nT) = x^*(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t-nT)$$

$$\text{où } \delta(t-nT) = \begin{cases} 1 & \text{si } t=nT \\ 0 & \text{si } t \neq nT \end{cases} ; n \in \mathbb{N}$$

Pour que le signal échantillonné soit fidèlement reconstitué il faut avoir: $f_e \geq 2f_c$, f_c représentant la fréquence maximale contenue dans le spectre du signal qui représente un spectre borné.

1.6.2 / Transformée de Fourier discrète (DFT):

De même que la transformée de Fourier fait correspondre une fonction à une fonction, la DFT fait correspondre une suite finie à une suite finie.

Si le signal $x(t)$ est représenté par la suite de N échantillons $\{X(n)\}$, $n = 0, 1, \dots, N-1$ la T.F.D est représentée elle aussi par une séquence de N valeurs.

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

La transformée de Fourier Discrète inverse (TFD⁻¹) s'écrit alors:

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

1.6.3/ La transformée rapide de Fourier (TFR):

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

En effet la TFD d'une séquence finie $\{x(n)\}$ a pour expression.

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

Si on pose:

$$w = e^{-j \frac{2\pi}{N}}$$

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

ou on a:

$$e^{-2\pi j} = 1 \longrightarrow w^N = 1$$

$$e^{-j\pi} = -1 \longrightarrow w^{N/2} = -1$$

La DFT⁻¹ s'écrit alors

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

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

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

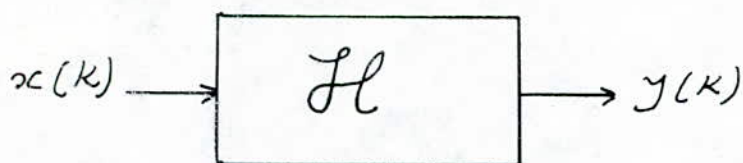
Ce dernier nécessite $N \log_2 N$ multiplications, alors que le calcul direct de la DFT pour N valeurs de $\{x(n)\}$ nécessite N^2 multiplications et $N(N-1)$ additions ce qui correspond à une quantité énorme de calcul si N devient grand.

2.2 / Traitement homomorphique des signaux =

2.2.1 / Généralités.

Les systèmes linéaires en traitement du signal présentent une grande importance, ils sont faciles à analyser et à caractériser et leur réalisation pratique est également relativement facile permettant ainsi plusieurs opérations de traitement. Par contre les systèmes non linéaires sont relativement difficile à analyser et à caractériser mathématiquement. Le traitement homomorphique tout en étant limité dans l'énorme variété des systèmes non linéaires a le mérite de se prêter facilement à une analyse mathématique et d'avoir des applications pratiques importantes. Celui-ci étant basé sur le principe de superposition généralisé.

Ainsi on peut schématiser un système homomorphique comme suit.



□ = Le symbole peut représenter, une addition, ou un produit de convolution. Ceci pour les signaux d'entrées.

○ = même chose que "□" mais pour les signaux de sortie.

2.2.2/ Forme canonique =

Tout système homomorphique peut être représenté par trois systèmes en série dont celui du milieu serait un système linéaire, la représentation schématique est sur la figure 2.1.

Les systèmes A_0 et A_0^{-1} sont appelés systèmes caractéristique de la classe particulière des systèmes homomorphiques \mathcal{H} .

2.2.3/ Système homomorphique Convolutif.

Dans des applications très variées il est très fréquent de se trouver devant le problème d'isoler une composante utile du signal ou de mesurer certaines de ses caractéristiques, lorsque celui-ci est combiné par convolution avec d'autres signaux.

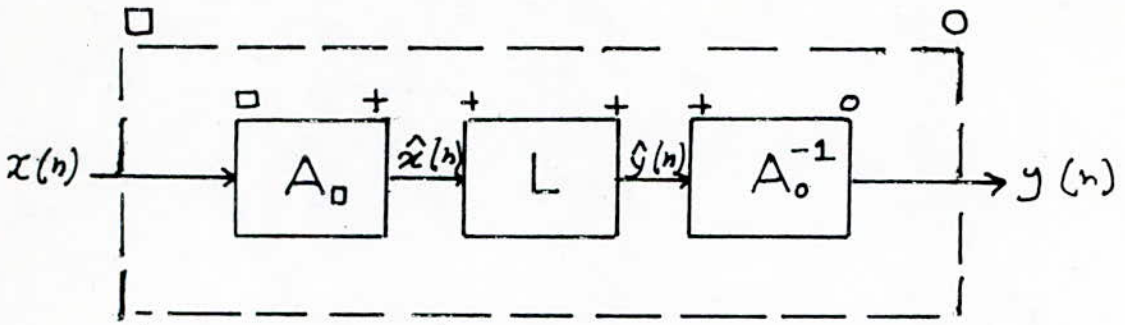
En supposant que les lois à l'entrée et à la sortie du système homomorphique sont identiques et que la loi de combinaison est le produit de convolution, le système sera représenté sous la forme canonique donnée par la figure 2.2

$$X(f) \Rightarrow x(k)$$

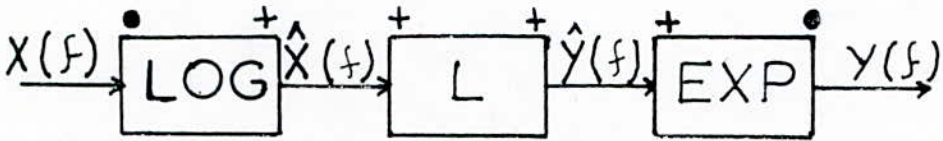
$$\text{et } x(k) = x_1(k) * x_2(k)$$

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

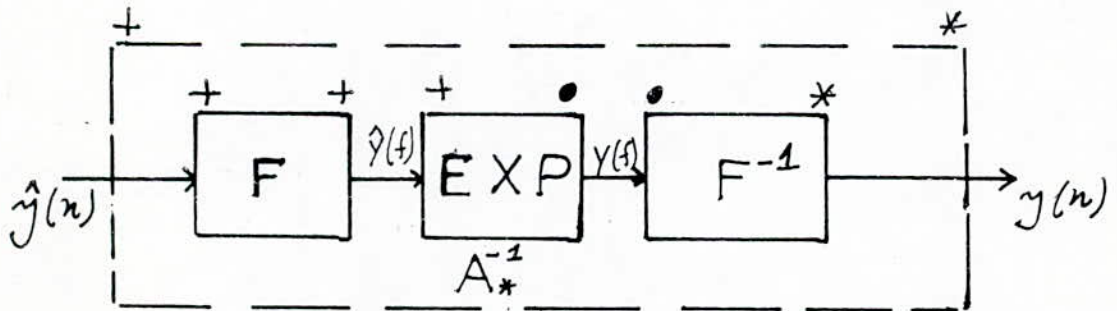
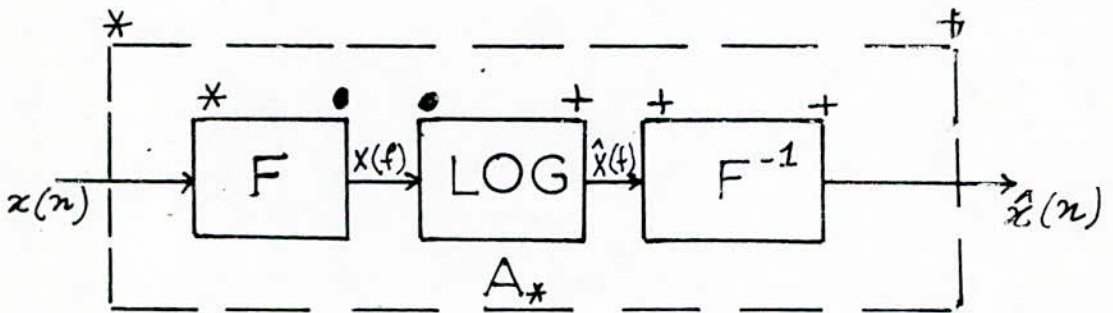
Une autre forme canonique peut-être envisagée non plus dans le domaine fréquentiel mais dans le domaine temporel, ce qui est représenté schématiquement sur la figure 2.3.



- figure 2.1 -



- figure 2.2 -



- figure 2.3 -

Une contrainte doit être posée sur le système caractéristique inverse A_*^{-1} ; en effet par définition on doit avoir :

$$A_*^{-1} [A_* (x(n))] = x(n)$$

2.2.4/ Notion de Cepstre =

Le cepstre existe sous différentes formes, mais toutes peuvent être considérées en réalité comme le spectre d'un spectre logarithmique. Cela veut dire qu'il peut servir à la détection de toute structure périodique dans le spectre, On montre également que les effets qui sont soumis à une convolution dans le signal temporel (multiplication dans le spectre) deviennent additifs dans le cepstre. (Chap 2.3).

2.3/ Estimation du pitch par la méthode Cepstrale:

Le nouveau modèle familier de la forme d'onde de la parole et l'analyse dépendante du temps peut être combiné avec la théorie des "systèmes homomorphiques".

On rappelle aussi que le modèle de production de la parole consiste en un système linéaire qui varie lentement dans le temps, excité par un train d'impulsions quasi-périodiques ou par un bruit, donc on peut dire qu'un segment de son voisé a été généré par l'excitation d'un système linéaire par un train d'impulsion périodique, de même un segment de son non voisé résultera de l'excitation d'un système linéaire mais cette fois-ci par un bruit.

Le schéma de la figure (2.a) nous donne un modèle de base pour la production d'un son voisé.

$s(t)$, représente le signal source de souffle périodique d'air.

L'effet du conduit vocal est spécifié par sa réponse impulsionnelle $h(t)$, donc $f(t)$ sera le produit de convolution de $s(t)$ et $h(t)$.

$$\begin{cases} f(t) = s(t) * h(t) \\ F(\omega) = S(\omega) \cdot H(\omega) \end{cases}$$

Ainsi le procédé consiste à prendre le logarithme du spectre de puissance quasi instantané $|F(\omega)|$ dans le but de séparer les effets de la source et du tractus

vocal, leurs spectres relatifs étants additifs dans le logarithme de $|F(\omega)|^2$

$$\begin{aligned}\text{Log} [|F(\omega)|^2] &= \text{Log} [|S(\omega)|^2 |H(\omega)|^2] \\ &= \text{Log} |S(\omega)|^2 + \text{Log} |H(\omega)|^2\end{aligned}$$

Ceci est expliqué aisément sur la figure (2. b)

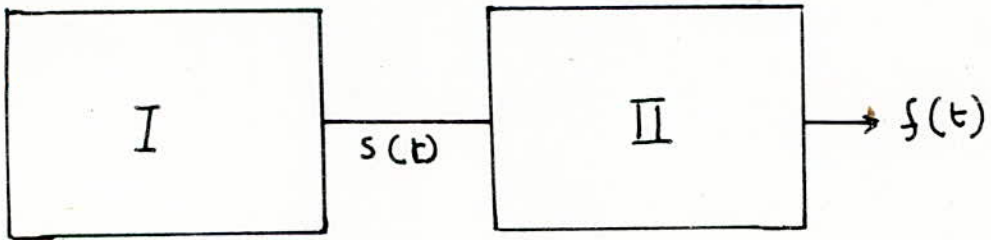
L'effet du tractus vocal est de produire des ondulations B.F dans le logarithme du spectre tandis que la périodicité de la source vocale se manifeste par des ondulations H.F dans le logarithme du spectre. Ainsi le pic se situant dans les basses fréquences serait égal au pitch du son voisé et l'information située dans les hautes fréquences nous renseigne sur la structure formantique du son.

Or les paramètres de la parole et en particulier le pitch varient avec le temps, pour cela il est préférable d'avoir une restriction du signal en de petits intervalles temporels.

Ainsi l'analyse se fera sur chaque intervalle qui seront obtenus en multipliant le signal par des fenêtres temporelles. Cette méthode est utilisée généralement pour voir si la portion du signal est voisée ou non.

$$g(t) = [s(t) * h(t)] w(t)$$

$w(t)$ est une fenêtre temporelle qui sera nulle pour $|t| > T_w$. Ainsi on peut représenter la source périodique

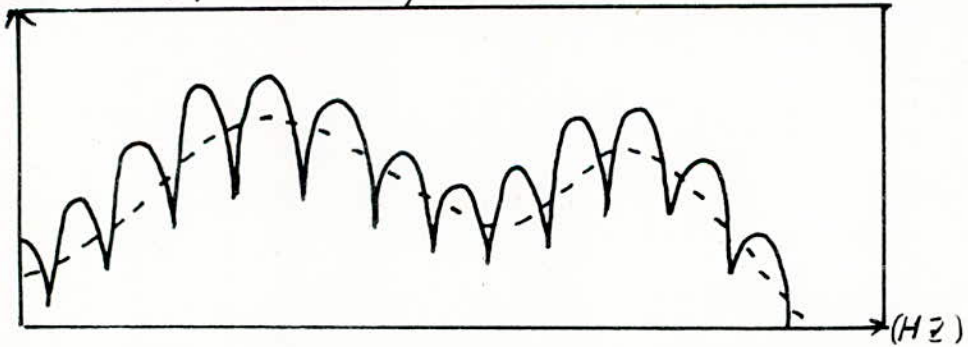


I → Source vocale

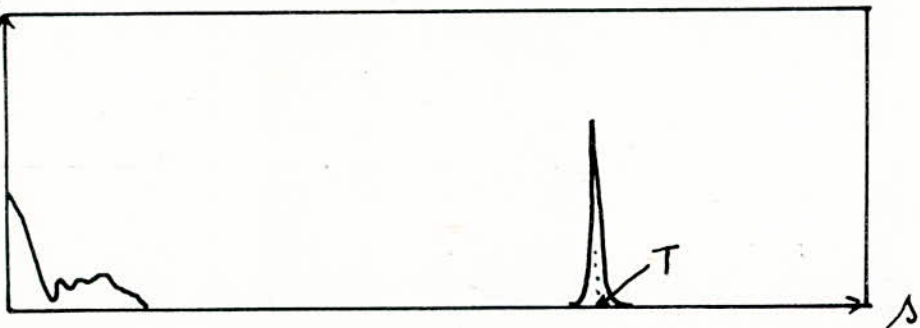
II → Conduit vocal

- figure 2.a -

logarithme du spectre de puissance d'un son voisé -



Cepstre



(Figure 2.b)

du signal $s(t)$ par la superposition d'une série infinie de signaux identiques $S_0(t)$ qui se répète à "T" sec.

$$s(t) = \sum_{n=-\infty}^{+\infty} S_0(t - nT) = S_0(t) * \sum_{n=-\infty}^{+\infty} \delta(t - nT)$$

d'où :

$$g(t) = \left\{ \left[S_0(t) * \sum_{n=-\infty}^{+\infty} \delta(t - nT) \right] * h(t) \right\} w(t)$$

La transformée de Fourier ou le spectre complexe $G(\omega)$ de $g(t)$ sera alors :

$$\begin{aligned} G(\omega) &= \left\{ \left[S_0(\omega) \sum_{n=-\infty}^{+\infty} \delta\left(\omega - n \frac{2\pi}{T}\right) \right] H(\omega) \right\} * W(\omega) \\ &= \left[\sum_{n=-\infty}^{+\infty} H(\omega) S_0(\omega) \delta\left(\omega - n \frac{2\pi}{T}\right) \right] * W(\omega) \end{aligned}$$

$$G(\omega) = \left[\sum_{n=-\infty}^{+\infty} H\left(n \frac{2\pi}{T}\right) S_0\left(n \frac{2\pi}{T}\right) \delta\left(\omega - n \frac{2\pi}{T}\right) \right] * W(\omega)$$

Où $S_0(\omega)$; $H(\omega)$; $W(\omega)$ sont les transformées de Fourier de $S_0(t)$; $h(t)$; $w(t)$.

Le signal de parole original $S(t) * h(t)$ n'a pas subi de fenêtrage alors le spectre complexe présentera une série infinie d'impulsions espacées de " $\frac{1}{T}$ " Hz et avec une amplitude " $H\left(n \frac{2\pi}{T}\right) S_0\left(n \frac{2\pi}{T}\right)$ ", et si le (spectre) signal $s(t) * h(t)$ est à bande limitée, le spectre complexe serait nul pour $|\omega| > \omega_{\max}$. Ainsi la multiplication du signal original par une fenêtre temporelle correspond à un produit de convolution dans

le domaine spectral.

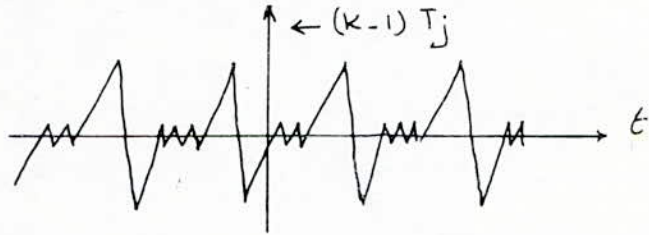
$$G(\omega) = [S(\omega) \cdot H(\omega)] * W(\omega)$$

Ainsi pour que le spectre soit nul pour $|\omega| > \omega_{\max}$ il faut que le spectre de $w(t)$ ne possède pratiquement pas de lobes secondaires, on a opté alors dans notre cas pour la fenêtre de "Hamming".

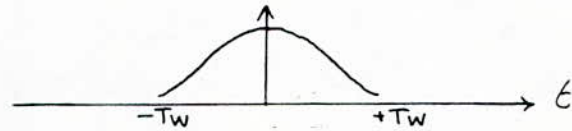
$$\begin{cases} w(t) = \alpha + (1-\alpha) \cos\left(\frac{\pi t}{T_w}\right); & |t| \leq T_w \\ = 0; & |t| > T_w \\ \alpha = 0.54 \end{cases}$$

Qui présente des lobes secondaires à -44 dB par rapport au lobe principal, donc on aura une atténuation considérable des lobes secondaires.

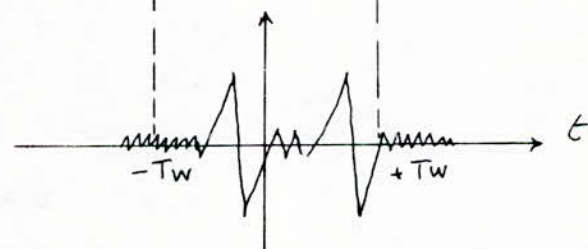
Signal parole
-1-



Fenêtre de Hamming
-2-



-1- x -2-



La figure précédente montre comment on peut obtenir le cepstre instantané d'un signal parole, les fenêtres ici sont décalées de T_j secondes.

2.3.1) Calcul numérique du Cepstre:

$f(t)$ étant un signal parole, sa transformée de Fourier

Sera alors: $F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt$

Si $f(t)$ est multipliée par une fenêtre temporelle $w(t)$ tel que $w(t) = 0$ pour $|t| > T_w$ on aura alors:

$$F(\omega) = \int_{-T_w}^{T_w} w(t) f(t) \cos(\omega t) dt - j \int_{-T_w}^{T_w} w(t) f(t) \sin(\omega t) dt$$

puisque la transformée de Fourier s'effectue sur une portion du signal on pourra appliquer le théorème d'échantillonnage. dans le domaine fréquentiel tout en respectant le critère de Nyquist; on aura ainsi:

$$\omega = m \Delta\omega, \quad \Delta\omega \leq \frac{2\pi}{2T_w}$$

En outre puisque le signal $f(t)$ est à bande limitée de 0 à $\omega_c/2\pi$; le paramètre temps peut être représenté par $t = l \Delta t$ avec $\Delta t \leq \frac{2\pi}{\omega_c}$

$F(\omega)$ prendra alors la forme suivante.

$$F(m \Delta\omega) = \Delta t \sum_{l=-L}^{L} w(l \Delta t) f(l \Delta t) \cos(l m \Delta t \Delta\omega) - j \Delta t \sum_{l=-L}^{L} w(l \Delta t) f(l \Delta t) \sin(l m \Delta t \Delta\omega)$$

avec $L = T_w / \Delta t$

L'équation précédente consiste à mémoriser $2L + 1$ échantillons du signal, qui sera considéré comme stationnaire vu qu'il a été multiplié par une fenêtre.

Ainsi le cepstre sera obtenu par le procédé suivant: Les parties réelles et imaginaires du spectre seront produites

par un analyseur de spectre, le spectre de puissance sera obtenu en prenant le module du spectre complexe. Le logarithme de ce spectre de puissance nous donne alors le cepstre.

Dans l'équation un seul spectre est défini parmi une série de spectres, si la fenêtre de Hamming se déplace chaque T_j secondes, le $k^{\text{ième}}$ spectre instantané $F_k(m)$ sera défini de la façon suivante:

$$F_k(m) = \sum_{e=-L}^L w(e) f[(k-1)K + e] \cos(e m \Delta t \Delta \omega) - j \sum_{e=-L}^L w(e) f[(k-1)K + e] \sin(e m \Delta t \Delta \omega)$$

$$L = \frac{T_w}{\Delta t} ; K = \frac{T_j}{\Delta t} ; m = 0, 1, \dots, \frac{\omega_c}{\Delta \omega}$$

Alors le $k^{\text{ième}}$ spectre de puissance instantané sera le suivant :

$$|F_k(m)|^2 = \left\{ \sum_{e=-L}^L w(e) f[(k-1)K + e] \cos(e m \Delta t \Delta \omega) \right\}^2 + \left\{ \sum_{e=-L}^L w(e) f[(k-1)K + e] \sin(e m \Delta t \Delta \omega) \right\}^2$$

$$\text{avec } \Delta \omega \leq \frac{2\pi}{4T_w}$$

Le $k^{\text{ième}}$ cepstre instantané $C_k(n)$ peut s'écrire alors:

$$C_k(n) = \sum_{m=0}^M \log |F_k(m)|^2 \cos(\Delta t \Delta \omega m n)$$

avec ; $\Delta\omega \leq \frac{2\pi}{4T_w}$; $M = \frac{\omega_c}{\Delta\omega}$; $n = 0, 1, \dots, N$

et $\Delta\tau \leq \frac{2\pi}{4\omega_c}$

2.3.9/ Conclusion :

Pour que la méthode cépstrale donne de bons résultats il faut que l'analyse soit faite avec une grande résolution. Ainsi pour des fréquences fondamentales faibles, et quand celles-ci ne sont pas suffisamment stables pendant la durée d'une fenêtre d'analyse, la méthode cépstrale est omise.

En outre le volume des calculs est très important et pour pouvoir analyser 2 secondes d'un signal parole échantillonné à $f_e = 10\text{KHz}$ avec une durée de fenêtrage égale à 30ms il faut 2 heures de temps.

Et malgré l'apparition des algorithmes "FFT" la durée d'analyse a été réduite seulement à 45 minutes, soit 8 fois moins que précédemment.

Enfin on peut dire que la méthode cépsrale est peu employée vue qu'elle n'est pas réalisable en temps réel.

CHAPITRE 3

METHODES DE DECALAGE

3.1) Généralités =

Ces méthodes sont destinées à exploiter la forte corrélation existante entre deux périodes fondamentales successives d'un signal voisé.

Parmi ces méthodes on trouve la fonction d'autocorrélation, l'AMDF (Average magnitude difference function) ou aussi méthode des différences.

3.2. / Fonction d'autocorrélation =

La fonction d'autocorrélation en temps discret est définie, par ce qui suit:

$$\Phi(k) = \sum_{m=-\infty}^{\infty} x(m) x(m+k) \quad (3.1)$$

Une autre définition de la fonction d'autocorrélation pour un signal aléatoire serait:

$$\Phi(k) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n) x(n+k) \quad (3.2)$$

Dans les deux cas précédents la représentation de la fonction d'autocorrélation d'un signal est un chemin assez commode pour tirer certaines propriétés de celui-ci.

par exemple si le signal est périodique, de période P échantillons, il va de soit que la fonction d'auto-correlation est aussi périodique et de même période que le signal.

$$\phi(k) = \phi(k+P) \quad (3.3)$$

3.2.1) propriétés =

- La fonction d'auto correlation est une fonction paire

$$\phi(k) = \phi(-k) \quad - a -$$

- Sa valeur maximale est atteinte pour $k=0$.

$$|\phi(k)| \leq \phi(0) \quad \forall k \quad - b -$$

- la quantité $\phi(0)$ représente une énergie.

$$E = \sum_{m=-\infty}^{+\infty} x^2(m) \quad - c -$$

3.3./ fonction d'auto correlation à court terme =

Si on considère l'équation 3.3 ainsi que les propriétés - a - et - b - on peut dire que pour un signal périodique, la fonction d'auto correlation atteint le maximum aux échantillons $0, \pm P, \pm 2P, \dots$.

Ainsi la période peut être estimée par le repérage du premier maximum de la fonction d'auto correlation.

Cette propriété de la fonction d'auto correlation constitue un outil solide pour l'estimation de la périodi-

-ité de n'importe quel signal y compris la parole.

Vu que la fonction d'auto corrélation contient beaucoup d'information en ce qui concerne la structure détaillée du signal, il serait extrêmement important de considérer une autre définition de la fonction, qui serait cette fois-ci à court terme, elle est définie de la façon suivante:

$$R_n(k) = \sum_{m=-\infty}^{+\infty} x(m) w(n-m) x(m+k) w(n-k-m) \quad (3.4)$$

Celle-ci peut être interprétée de la façon suivante. En premier lieu le segment du signal parole est sélectionné par l'intermédiaire d'un fenêtrage.

Vu que $R_n(k) = R_n(-k)$

$R_n(k)$ pourra s'écrire sous la forme suivante:

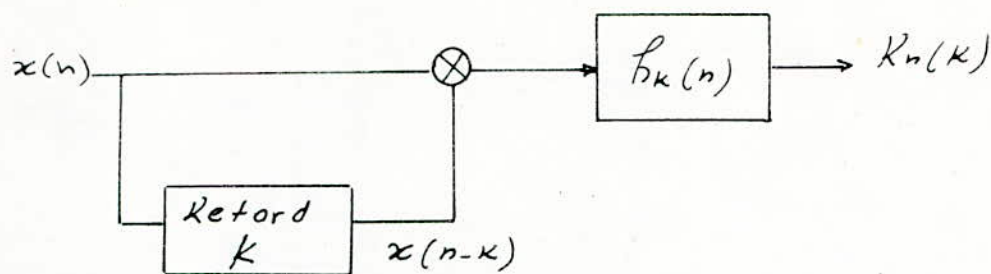
$$R_n(k) = R_n(-k) = \sum_{m=-\infty}^{\infty} x(m) x(m-k) [w(n-m) w(n+k-m)] \quad (3.5)$$

posons $h_k(n) = w(n) w(n+k)$

L'équation sera alors:

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m) x(m-k) h_k(n-m)$$

Donc $R_n(k)$ serait obtenue par le filtrage de la séquence $x(n) x(n-k)$ par un filtre de réponse impulsionnelle $h_k(n)$ comme l'indique la figure suivante



Le calcul de la fonction d'autocorrelation à court terme est généralement effectué à partir de l'équation (3.4) après l'avoir réécrite sous la forme suivante:

$$R_n(k) = \sum_{-\infty=m}^{+\infty} [x(n+m)w'(m)] [x(n+m+k)w'(k+m)] \quad (3.6)$$

ou $w'(n) = w(n)$

L'équation précédente montre aisément que l'origine des temps de la séquence d'entrée a été effectivement déplacé de "n" échantillons, et multipliée par une fenêtre w' en vue de sélectionner un court segment de parole.

Si la fenêtre w' est de durée finie, la séquence résultante $x(n+m)w'(n)$ sera de durée finie et l'équation prendra la forme suivante:

$$R_n(k) = \sum_{m=0}^{N-1-k} [x(n+m)w'(m)] [x(n+m+k)w'(m+k)] \quad (3.7)$$

On peut constater que pour calculer $R_n(k)$ par l'intermédiaire de l'équation précédente cela demandera N multiplications pour le calcul de $x(n+m)w'(m)$ et $(N-k)$ multiplications et additions pour calculer $x(n+m+k)w'(k+m)$.

3.4.1 Estimation du pitch =

Nous allons illustrer quelques exemples montrant la fonction d'autocorrelation d'un signal parole échantillonné à une fréquence de 10 KHz, tout en se basant sur l'équation (3.7) ;

$N = 401$; Cette fonction a été évaluée pour les retards $0 \leq k \leq 250$.

La figure (1) illustre trois exemples, les deux premiers concernant une voix voisée, le troisième un segment non voisé.

Pour le premier segment de voix voisée les pics se reproduisent approximativement à des multiples de 72 indiquant ainsi la période à 7,2 ms ceci veut dire que le pitch est voisin de 140 Hz.

Pour le second cas on voit que des effets similaires de périodicité seulement dans ce cas les pics de la fonction d'autocorrelation sont des multiples de 58 échantillons indiquant ainsi le pitch à 5,8 ms. Enfin pour le troisième et dernier cas concernant un segment de voix non voisée il n'y a pas de périodicité des pics de la fonction d'autocorrelation indiquant ainsi un manque de périodicité de la forme d'onde de la parole.

- N. B. -

On note que dans le cas précédent le signal a subi un fenêtrage du type rectangulaire et que la fonction d'autocorrelation a été normalisée à l'unité ; c'est à dire

$$R_h(0) = 1.$$

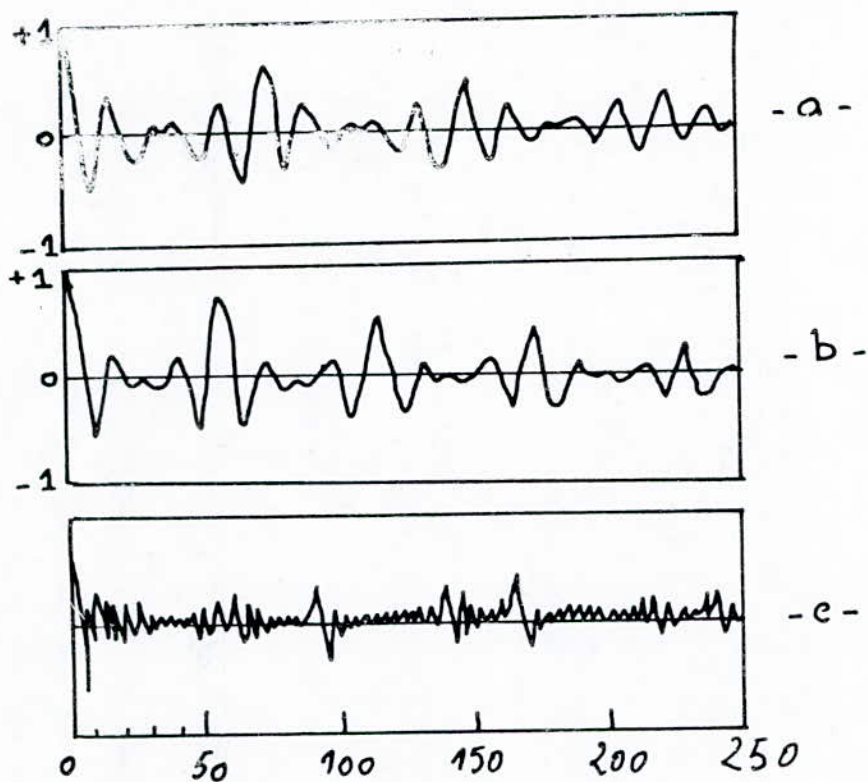
La figure -2- présente les mêmes exemples seulement ici on utilise un fenêtrage de Hamming.

Ainsi si on veut faire une comparaison avec les résultats obtenus avec une fenêtre rectangulaire, on peut dire que cette dernière donne beaucoup plus d'indications en ce qui concerne la périodicité.

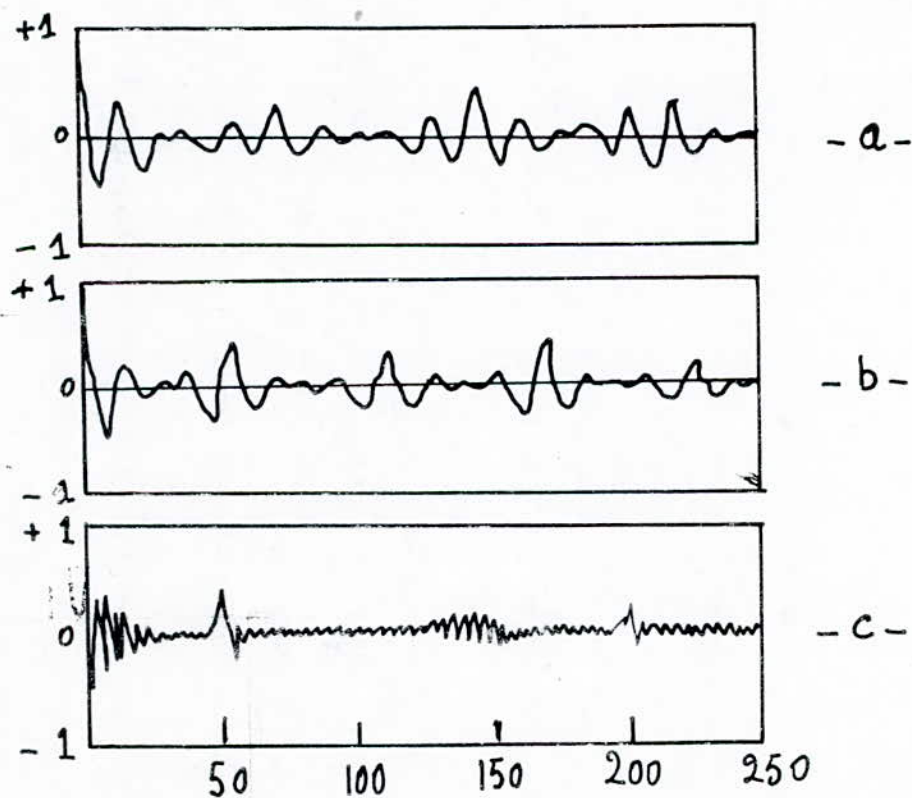
Ceci n'est pas surprenant vu que la fenêtre de Hamming possède une forme conique, donc elle va provoquer une atténuation du segment de signal parole donc un mauvais repérage des pics marquant la périodicité dans la fonction d'autocorrelation.

Les exemples des figures -1- et -2- étaient donnés pour un nombre d'échantillons $N = 401$. Une question pourrait se poser sur le nombre d'échantillons choisi et qui pourrait donner une bonne indication en ce qui concerne la périodicité; vu le changement des propriétés du signal parole, on devrait choisir " N " le plus petit possible. Ceci d'une part, et d'autre part pour avoir une assez bonne indication concernant la périodicité de la fonction d'autocorrelation, la fenêtre choisie doit avoir une durée d'au moins deux fois la période fondamentale de l'onde.

En effet vu la longueur finie du signal fenêtré, on aura de moins en moins de données contenues dans le calcul de $R_n(k)$ quand k augmente.



- figure 1 -



- figure 2 -

Ceci veut dire que la limite supérieure de la Somme (équation 3.7) conduit à la réduction de l'amplitude des pics de la corrélation quand K augmente.

Pour le cas d'un signal parole ceci est facilement illustré par la figure 3. qui montre l'effet d'une fenêtre rectangulaire de différentes longueurs.

Les lignes en pointillées représentent la fonction d'auto-corrélation de la fenêtre rectangulaire, celle-ci a pour équation.

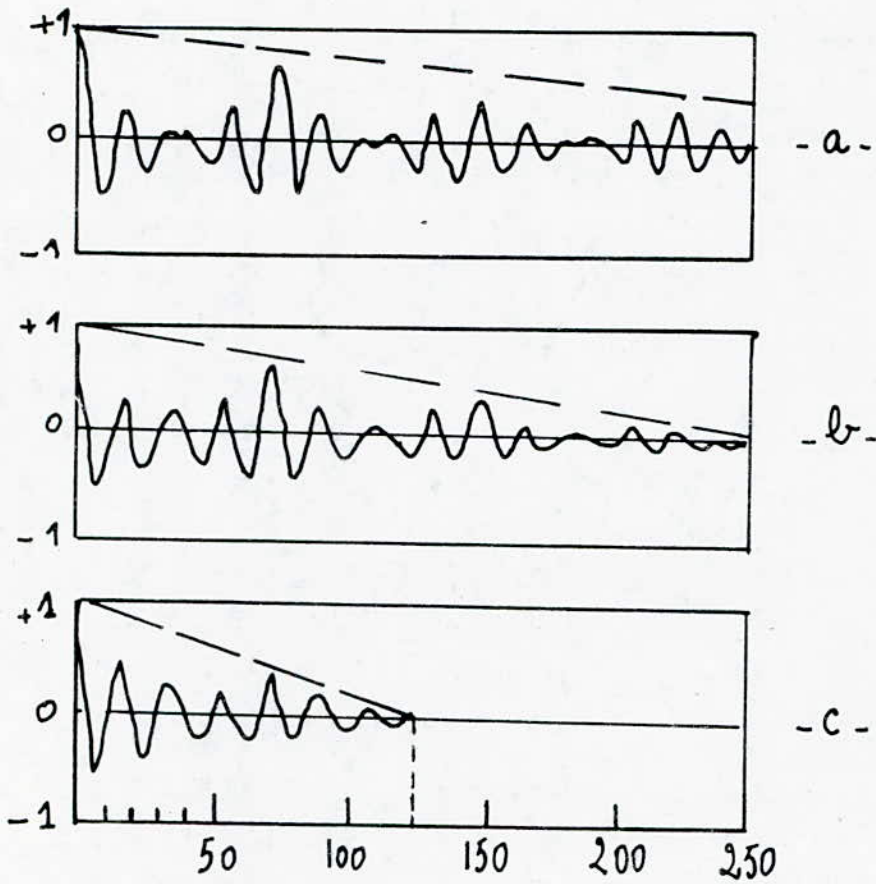
$$R(k) = 1 - \frac{|k|}{N} ; |k| < N \quad (3.8)$$

L'exemple illustré dans la figure 3. montre que les pics sont plus éloignés de la ligne "pointillée" dans le cas où $N=401$ que dans les deux derniers cas.

La figure 3.c. correspond à une fenêtre longue de 125 échantillons, puisque la période pour cet exemple est de 72 échantillons, on constate que même pas deux périodes fondamentales sont incluses dans la fenêtre, d'où un cas à éviter.

Une approche serait d'avoir une fenêtre suffisamment large pour loger la plus grande période fondamentale mais ceci conduit à une moyenne indésirable pour les périodes fondamentales de courte durée.

Enfin une autre solution utilisant des fenêtres de courtes durées tout en modifiant la fonction d'auto-corrélation, celle-ci est définie de la façon suivante.



- figure 3 -

fonction d'autocorrelation d'une voix voisée

-a- $N = 401$

-b- $N = 251$

-c- $N = 125$

$$\hat{R}_n(k) = \sum_{m=-\infty}^{+\infty} x(m) w_1(n-m) x(m+k) w_2(n-m+k) \quad (3.9)$$

qui peut s'écrire aussi sous la forme.

$$\hat{R}_n(k) = \sum_{m=-\infty}^{+\infty} x(n+m) w_1(m) x(n+m+k) \hat{w}_2(m+k) \quad (3.10)$$

avec $w_1(m) = w_1(-m)$

$$\hat{w}_2(m) = w_2(-m)$$

Pour remédier au problème de diminution due à la variabilité de la limite supérieure de l'équation (3.7) on peut choisir la fenêtre \hat{w}_2 et inclure les échantillons en dehors de l'intervalle non nul de la fenêtre \hat{w}_1 , ainsi on peut définir les deux fenêtres, de la façon suivante :

$$\begin{cases} w_1(m) = 1 & \text{pour } 0 \leq m \leq N-1 \\ & = 0 & \text{ailleurs} \\ \text{et} \\ w_2(m) = 1 & \text{pour } 0 \leq m \leq N-1+k \\ & = 0 & \text{ailleurs.} \end{cases}$$

k représente ici le plus grand retard.

Dou l'équation (3.10) pourra s'écrire sous la forme suivante :

$$\hat{R}_n(k) = \sum_{m=0}^{N-1} x(n+m) x(n+m+k) \quad (3.11)$$

$$0 < k \leq k$$

Ce qui veut dire que la moyenne est au dessus de N

échantillons, ce qui sont en dehors de l'intervalle $n \bar{a}$ $n+N-1$ sont inclus dans le calcul de l'équation (3.7) et (3.11); ceci étant illustré sur la figure-4.

Cette dernière montre les différents échantillons inclus dans le calcul de la fonction d'auto corrélation à court terme.

Plus précisément, la fig-4-a- montre le signal parole sous forme d'onde, la fig-4-b- montre un segment de N échantillons sélectionnés par l'intermédiaire d'une fenêtre rectangulaire; ce dernier serait utilisé pour les deux termes de l'équation (3.11).

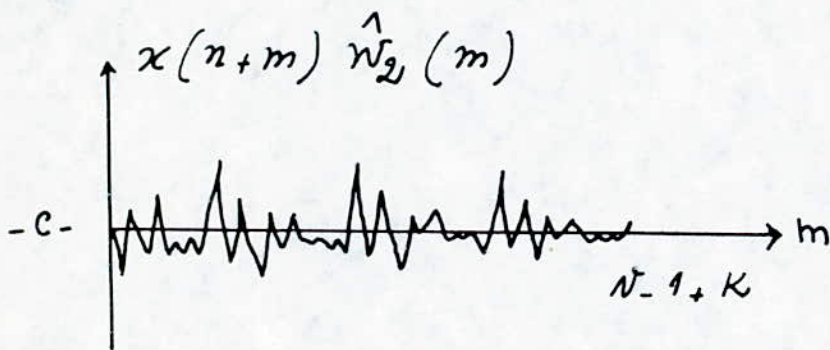
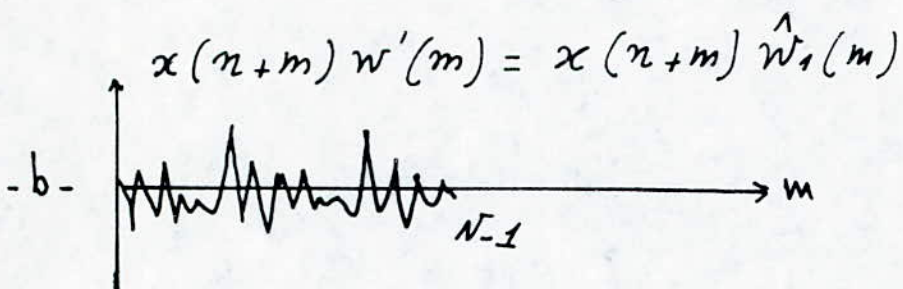
La fig-4-c- montre l'autre terme de l'équation (3.11) Noter ici que K échantillons sont inclus.

Ainsi l'équation (3.11) serait équivalente à une fonction d'auto corrélation à court terme modifiée, ou plus précisément "Cross-Correlation" de deux segments de parole différents mais de longueur finie, le premier $x(n+m) \hat{W}_1(m)$ et le second $x(n+m) \hat{W}_2(m)$.

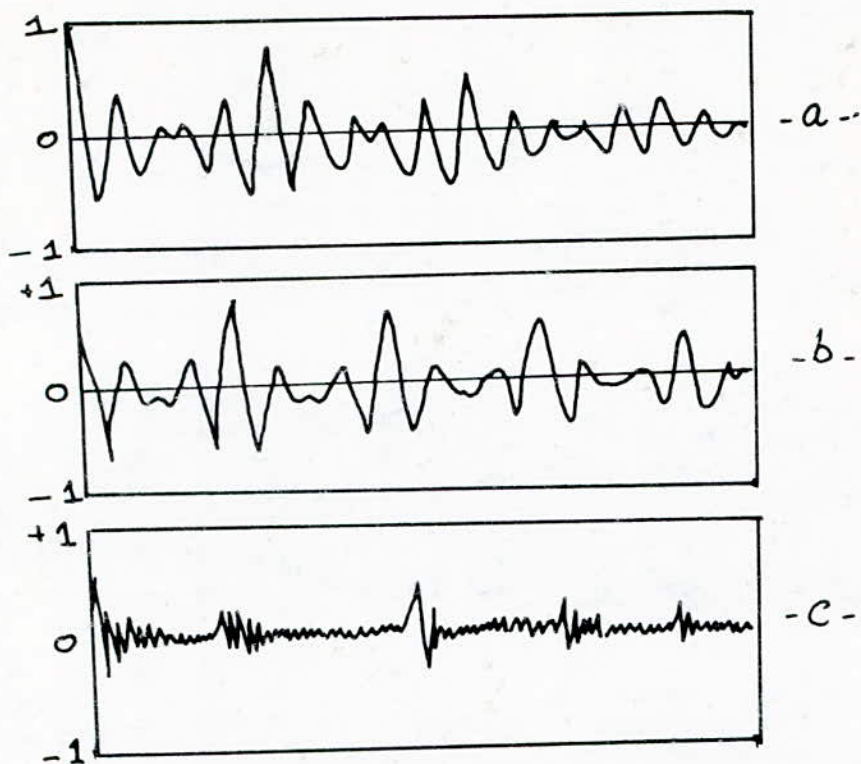
La figure -5- montre la fonction d'auto corrélation modifiée correspondante aux exemples de la fig-1-.

En effet pour $N=401$ les variations de la forme d'onde affectent l'effet fuselé aux niveaux des pics dans la fig-1-, Ainsi les deux figures semblent identiques

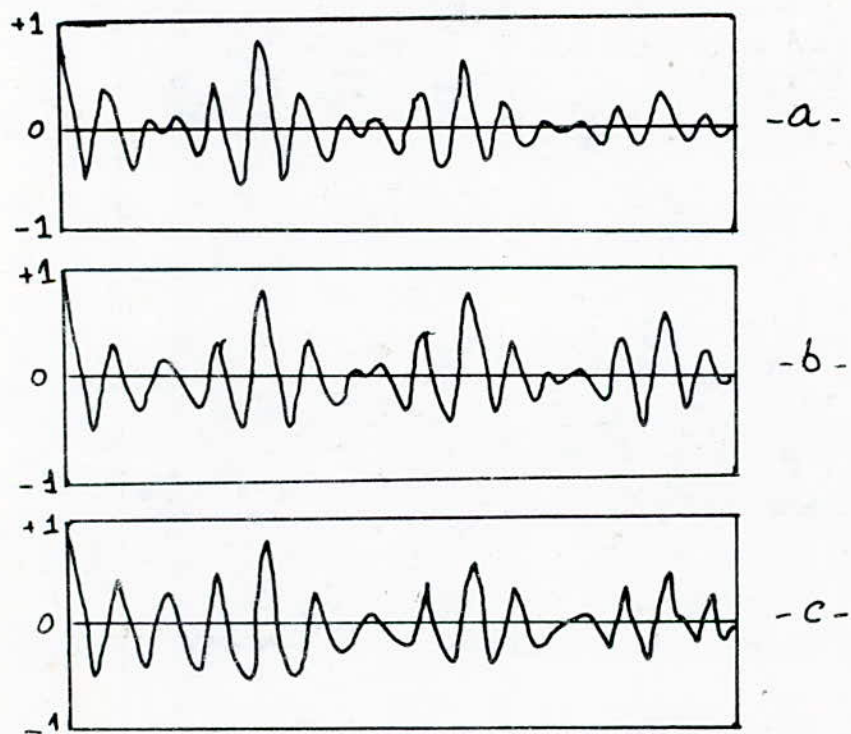
En faisant une comparaison entre la figure-6- et la figure-3-. La différence est beaucoup plus apparente pour les petites valeurs de N .



- figure 4 -



- figure 5 -



- figure 6 -

3.5.1 Méthode des différences = (AMDF).

La fonction d'autocorrelation nécessite un volume de calcul considérable, il existe une technique qui élimine la nécessité des multiplications, elle est basée sur l'idée que pour une séquence vraiment périodique de période P , la séquence suivante:

$$d(n) = x(n) - x(n-k)$$

serait nulle pour $k = 0 \pm P; \pm 2P \dots$

Pour un segment de voix voisée il est plus correct de dire que $d(n)$ sera très proche de zéro, mais n'est pas nulle.

Donc la fonction de différence de $d(n)$ sera voisine de zéro quand k sera égale à la période.

On définit aussi la fonction de différence à court terme de la façon suivante:

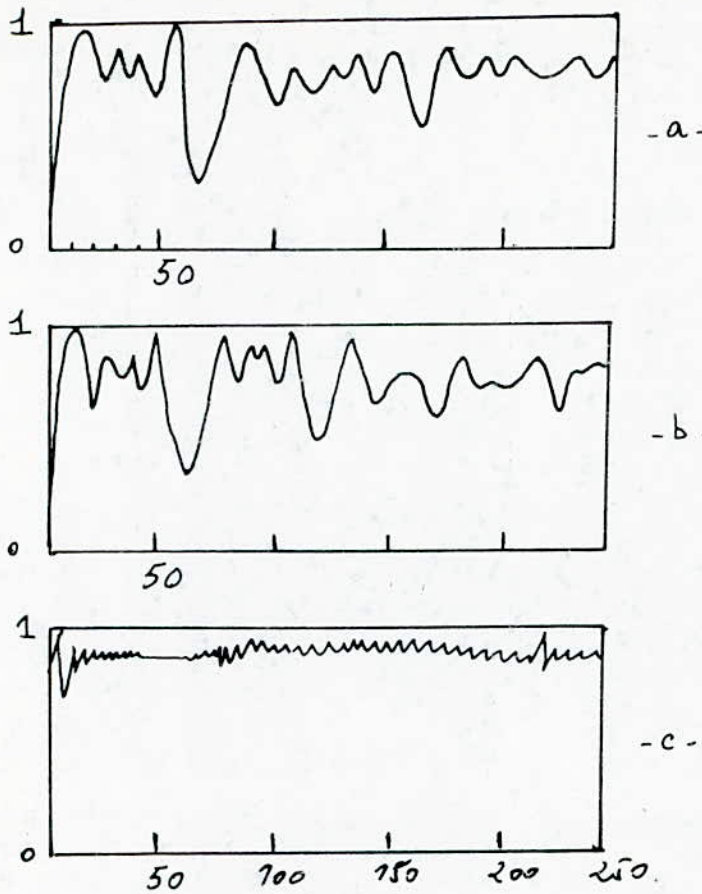
$$\gamma_n(k) = \sum_{m=-P}^{+\infty} \left| x(n+m)w_1(m) - x(n+m-k)w_2(m-k) \right|$$

L'estimation du pitch par la fonction de différence est analogue à celle de la fonction d'autocorrelation.

Seulement la fonction de différence est une méthode qui présente l'avantage de n'utiliser que des opérations de soustractions, additions, et valeur absolue; alors que la fonction d'autocorrelation la multiplication en plus des opérations utilisées dans la fonction de différence.

La figure suivante montre la fonction de différence

pour les segments de parole de la figure -1-



- fig 7 -

3.5.1 Conclusion =

Vu que la fonction de différence utilise moins d'opérations que la fonction d'auto-corrélation, on pourra dire que du point de vue traitement du signal de la parole et extraction du pitch, il est plus commode d'utiliser la méthode des différences vu que le temps de calcul est réduit.

CHAPITRE 4

REALISATION

4.1 Introduction =

Actuellement il existe plusieurs méthodes de mesure du pitch, deux d'entre-elles ont été développées dans les chapitres deux et trois.

La réalisation proposée est basée sur la détection des crêtes significatives de la périodicité. C'est un des principes les plus utilisés et les plus éprouvés. Cette technique consiste à mettre en évidence et à détecter les crêtes significatives de la périodicité du signal parole, celle-ci est précédée d'un préfiltrage et d'une normalisation des crêtes maximum.

Le filtrage est adapté à une large gamme de fréquences fondamentales, il est réalisé par la mise en cascade d'un filtre passe haut " $f_0 = 70\text{Hz}$ " avec une pente de 24dB/oct et d'un filtre passe bas " $f_0 = 400\text{Hz}$ " avec une pente de 12dB/oct .

Seulement notre maquette traitera rien que l'alternance positive du signal comme l'indique le schéma synoptique suivant.



4.2./ Filtrage =

Soit un phénomène physique dont le spectre s'étend sur un intervalle de fréquences quelconque, le filtrage consiste à privilégier les composantes du signal dont les fréquences sont situées dans certains domaines tout en affaiblissant les autres.

4.2.1/ Filtrés actifs =

Un filtre actif est constitué par l'association de réseaux passifs (R,C) et d'éléments amplificateurs.

À l'égard des filtres passifs, les filtres actifs montrent un certain nombre d'avantages dont les principaux sont les suivants :

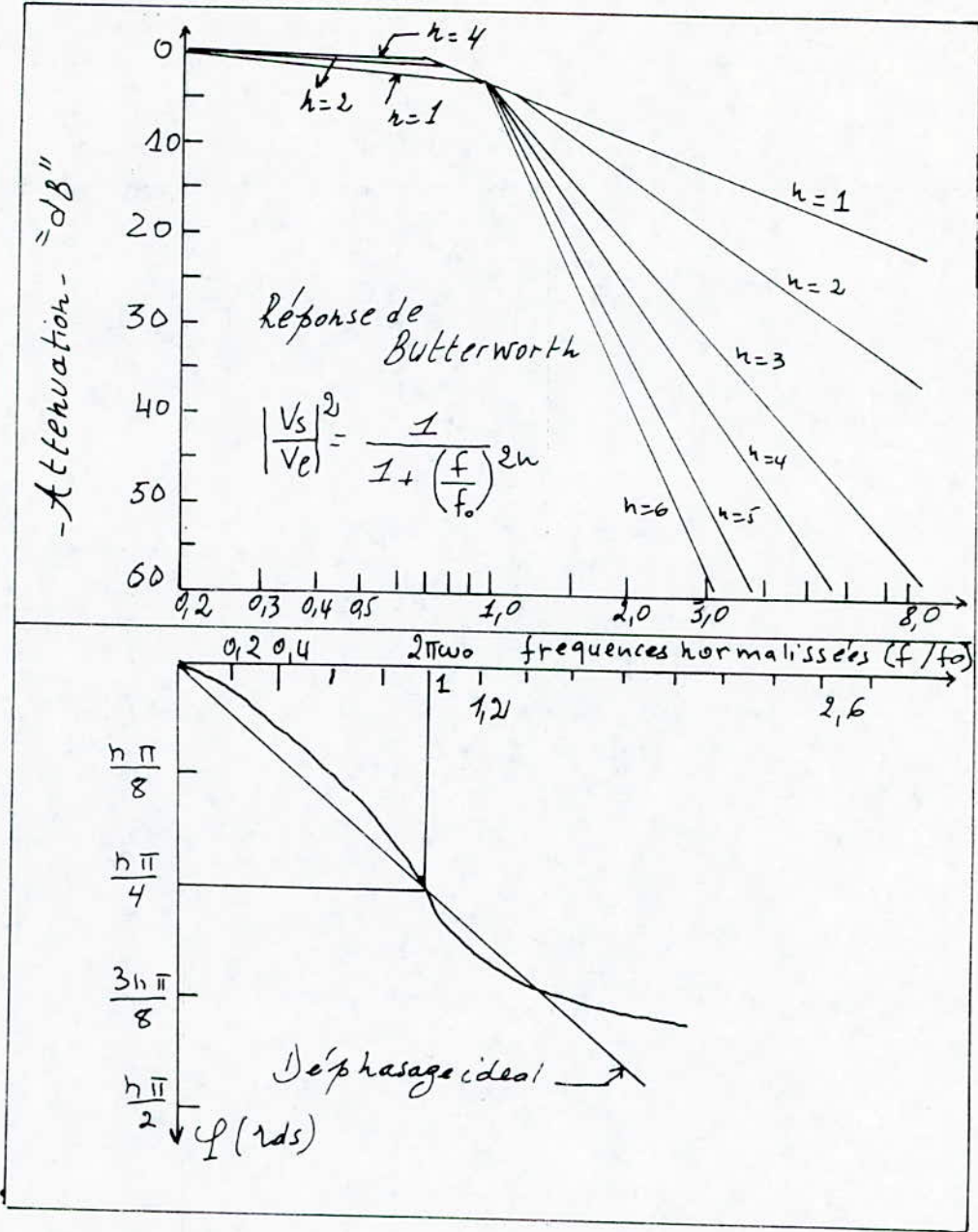
- Impédance d'entrée élevée ; impédance de sortie faible, ce qui permet de monter en cascade plusieurs filtres
- possibilité de travailler avec un certain gain.
- réglages continus de la fréquence d'accord du filtre par des moyens électriques simples.

4.2.2/ Principaux types de réponses de filtres =

Pour séparer des signaux de fréquences très voisines il va de soit que c'est la pente de décroissance de la caractéristique qui nous intéresse en premier lieu, on a alors le choix entre différents types de réponses ; parmi elles - celles de Butterworth, Tchebicheff - ...

4.2.3/ Fonction de Butterworth =

Figure 4.0



Ordre 4		Cellule 1		Cellule 2	
ondulation	chute en dB à 2f₀	m₁	q₁	m₂	q₂
0 dB	24	0,923	1,082	0,38	2,62

Ordre 2		Cellule 1	
ondulation	chute en dB à 2f₀	m	q
0 dB	12	0,707	1,414

- figure 4.0 -

La fonction de Butterworth est donnée par la formule suivante:

$$\left| \frac{V_s}{V_e} \right|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_0} \right)^{2n}}$$

$\omega_0 \rightarrow$ pulsation de coupure.

$n \rightarrow$ ordre du filtre.

La caractéristique principale de cette fonction réside dans le fait qu'elle présente une "plage" très linéaire au niveau de la bande passante. La chute du gain est assez régulière à partir de la fréquence " f_0 " avec une pente de 6ndB/oct

4.2.4) Fonction de Tchebitcheff =

celle-ci est donnée par la formule suivante.

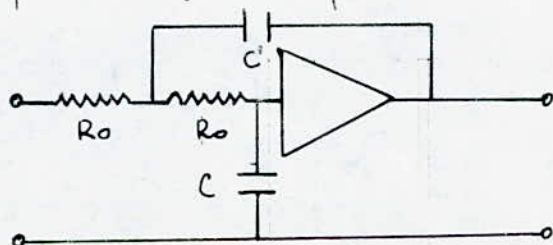
$$\left| \frac{V_s}{V_e} \right|^2 = \frac{1}{1 + \varepsilon^2 T_n^2 \left(\frac{\omega}{\omega_0} \right)}$$

ε représente ici un coefficient plus petit que 1 qui nous détermine l'ondulation au niveau de la bande passante.

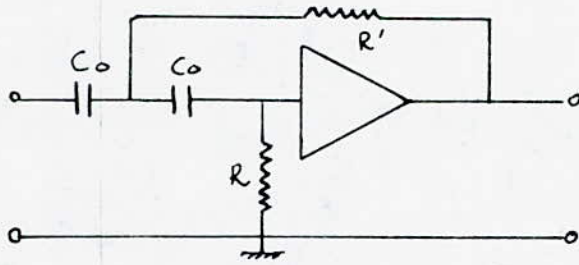
L'intérêt de la réponse de Tchebitcheff est sa rapide décroissance dans la partie coupe bande.

4.2.5) Détermination pratique d'un filtre =

Le schéma de principe d'un filtre passe bas d'ordre 2 est le suivant:



Pour le passage du filtre passe bas au filtre passe haut du même ordre on remplace les résistances par des capacités et les capacités par des résistances. On aura le schéma suivant.



4.2.6/ Calcul des éléments passifs:

Notre premier impératif est d'avoir une réponse plate au niveau de la bande passante; donc on a intérêt à choisir une réponse du type Butterworth, vu que celle de Tchebichef n'est utilisée que lorsqu'une coupure raide est voulue.

a) Filtre passe bas d'ordre 2.

$$\text{on a } \omega_0 = 2\pi f_0$$

La résistance est choisie selon la caractéristique de l'amplificateur opérationnel utilisé.

On fixe dans ce cas R_0 :

$$C_0 = \frac{1}{R_0 \omega_0}$$

pour avoir C et C' , on multiplie C_0 par les coefficients établis à l'aide d'un calculateur numérique pour les réponses de Butterworth; soit dans ce cas ci

$$m = 0,707 ; \quad q = 1,414$$

$$\text{D'où } C = m C_0 ; \quad C' = q C_0$$

b) Filtre passe haut d'ordre 2

Dans ce cas c'est la valeur de la capacité C_0 qui est fixée.

$$R_0 = \frac{1}{C_0 \omega_0}$$

Et pour avoir les résistances R et R' , on divise R_0 respectivement par les coefficients m et g .

$$R = \frac{R_0}{m} ; \quad R' = \frac{R_0}{g}$$

4.2.7) Calcul du filtre passe haut d'ordre 4 " $f_0 = 70\text{Hz}$ "

Ce filtre est réalisé par la mise en cascade de deux cellules passe haut du 2^{ème} ordre.

Le calcul des composants passifs se fera de la façon suivante.

$$f_0 = 70\text{Hz} ; \quad \omega_0 = 2\pi f_0 = 439,8 \text{ rd/s}$$

$$R_0 = \frac{1}{C_0 \omega_0} = 10,335 \text{ k}\Omega ; \quad \text{tout en fixant } C_0 \text{ à } 220\text{nF.}$$

- Cellule n°1:

Les coefficients m_1 et g_1 sont respectivement égaux à 0,9238 et 1,082.

$$R_1 = \frac{R_0}{m_1} ; \quad R'_1 = \frac{R_0}{g_1}$$

- Cellule n°2:

$$m_2 = 0,382 ; \quad g_2 = 2,613.$$

$$R_2 = \frac{R_0}{m_2} ; \quad R'_2 = \frac{R_0}{g_2}$$

Les valeurs des résistances normalisées donnant une réponse assez proche de la théorie sont en fin de chapitre.

4.2.8/ Calcul du filtre passe bas d'ordre 2 ; $f_0 = 400\text{Hz}$

Dans ce cas c'est la valeur de la résistance R_0 qui est fixée

$$f_0 = 400\text{Hz} \rightarrow \omega_0 = 2513,27 \text{ rd/s}$$

$$R_0 = 12,2 \text{ k}\Omega ; \text{ Comme } R_0 = \frac{1}{C_0 \omega_0} \rightarrow C_0 \approx 33,1 \text{ nF}$$

$$C_1 = m C_0$$

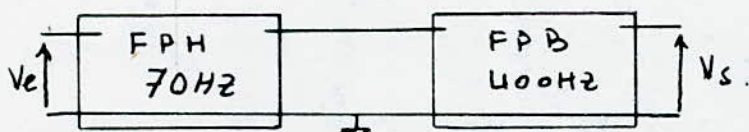
$$C_2 = q C_0$$

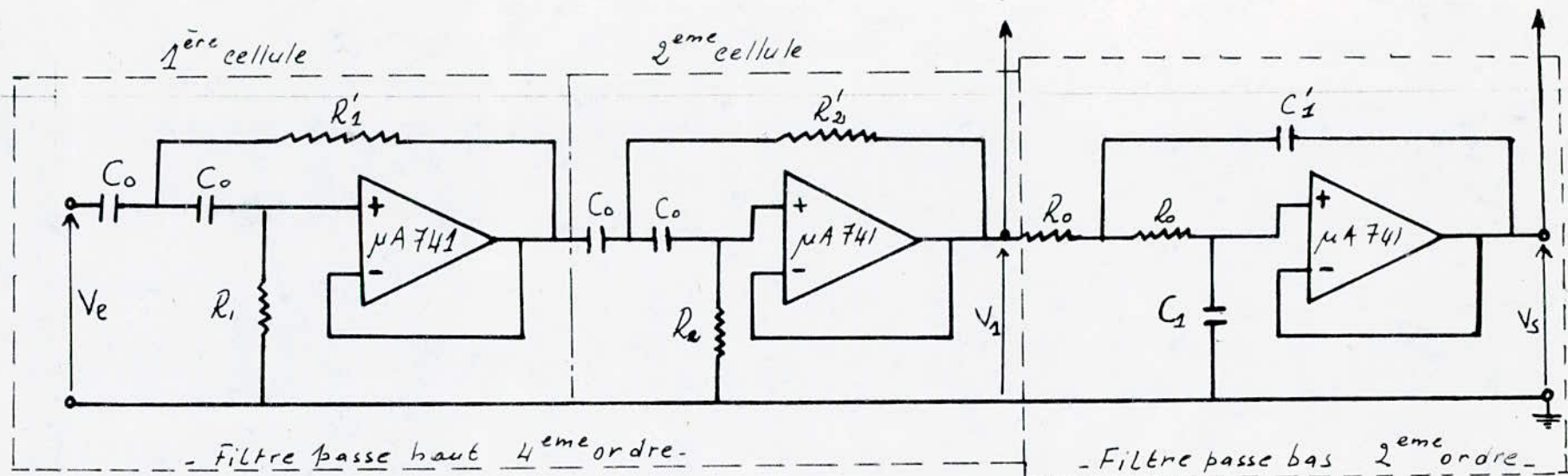
$$\text{dans ce cas } m = 0,707 ; q = 1,414$$

Les valeurs des capacités donnant une assez bonne réponse sont en fin du chapitre.

4.2.9/ Filtre passe bande:

Le filtre passe bande sera réalisé en disposant en cascade le filtre passe haut "70Hz" et du filtre passe bas "400Hz". Comme l'indique le schéma suivant.

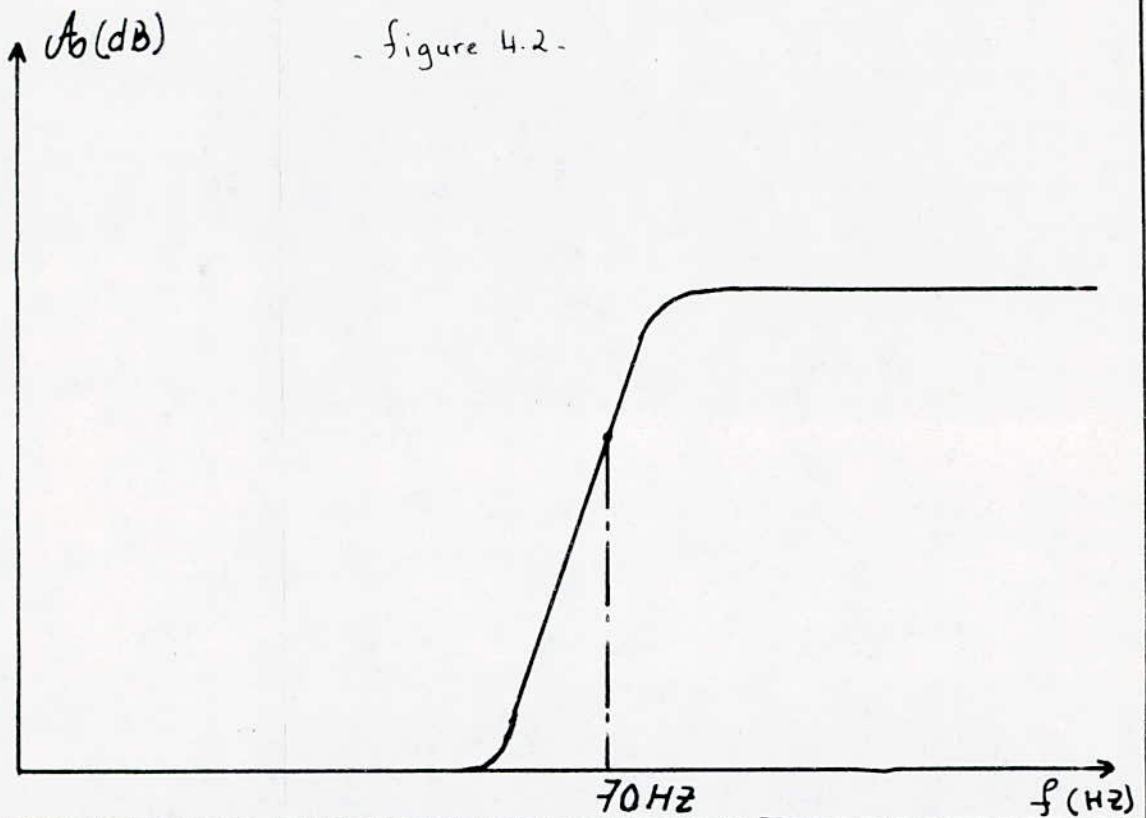




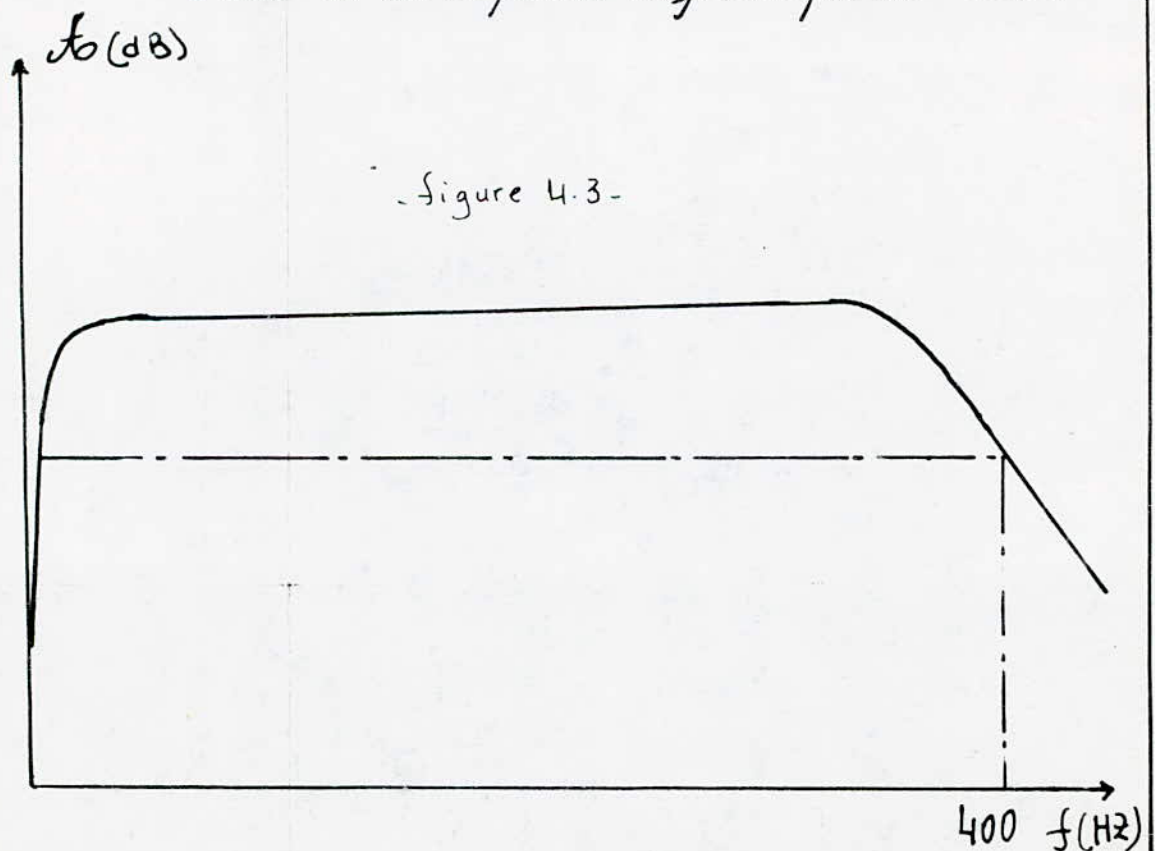
- Bloc préfiltrage -

figure 4.1

- Courbe de réponse du filtre passe haut -



- Courbe de réponse du filtre passe bas -



- Courbe de réponse du bloc préfiltrage -

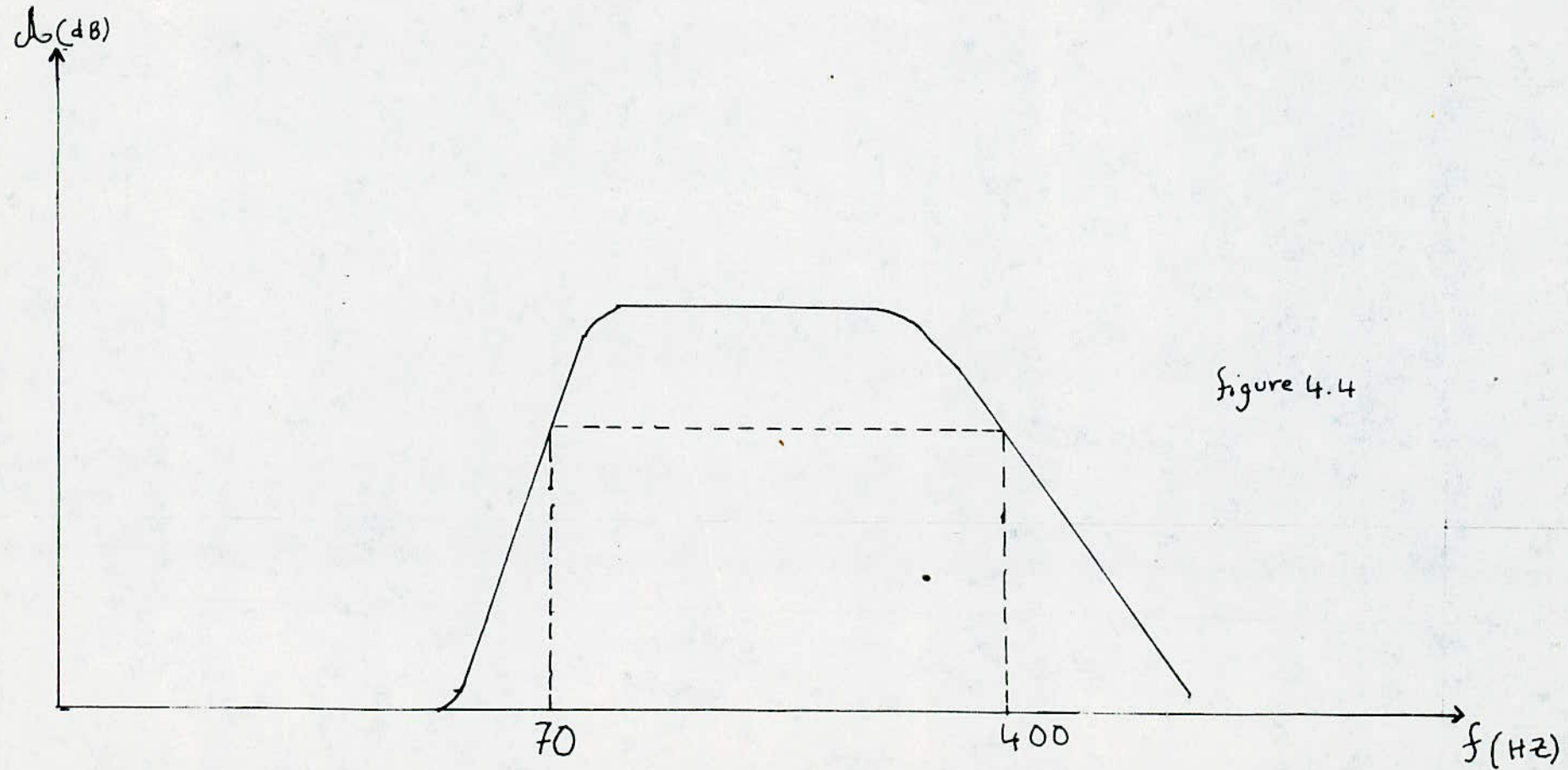


Figure 4.4

4.3/ Compression:

Après avoir effectué un renforcement de la fréquence fondamentale par l'intermédiaire du bloc préfiltrage, le signal va subir une compression de l'enveloppe des crêtes maximales.

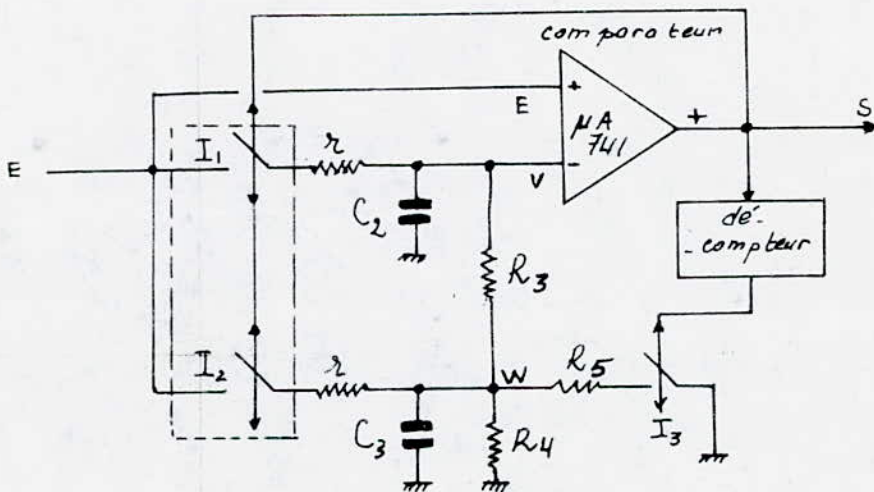
En effet si la détection des crêtes significatives est facile quand cette enveloppe varie peu ou pas, il n'en va pas de même dans le cas d'évolution très rapide de l'énergie du signal.

Le compresseur utilisé se caractérise par un temps de réaction instantané à la montée et un temps de réaction relativement long à la descente.

Cet effet peut être obtenu à l'aide d'une mémoire de crêtes qui commande la compression.

Cette valeur mémorisée décroît à peu près linéairement après un certain retard.

Le schéma fonctionnel du bloc compresseur et le suivant:



- fig 4.5 -

4.3.1. fonctionnement :

Le bloc compresseur fonctionne comme suit :

* Si la tension $E > V$ la Sortie du Comparateur "MA741" alimenté entre $V_{cc} = 5V$ et $-V_{cc} = 12V$ provoque la fermeture des interrupteurs I_1 et I_2 de sorte que les capacités C_2 et C_3 sont instantanément chargées.

Notons que la sortie du comparateur doit être compatible en logique TTL c'est à dire que le niveau logique 1 correspondra à $+5V$ et le niveau logique 0 correspondra à $0V$ ceci en vue d'assurer le fonctionnement du décompteur ainsi que les interrupteurs I_1, I_2, I_3 qui sont des circuits intégrés CMOS "CD4016".

Ceci est facilement réalisable en plaçant une diode à la Sortie du Comparateur, en effet cette dernière ne va laisser passer que l'alternance positive du signal, l'alternance négative étant éliminée.

* Si la tension $E < V$, la Sortie du Comparateur tombe à zéro on aura ainsi une fonction perte de mémoire.

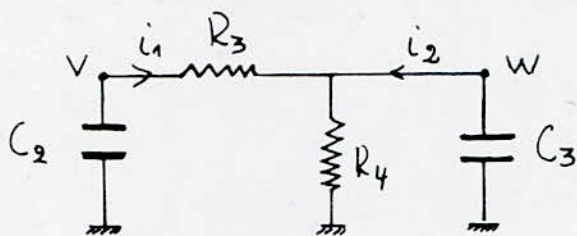
pendant ce temps les interrupteurs I_1 et I_2 sont ouverts et la fuite du circuit de mémoire erêtes se fait suivant un profil qui minimise la distorsion du signal compressé "ie une décroissance du 2^{ème} ordre.

$$-V(t) = \frac{V - W}{R_3 C_2} t + \text{cste}$$

Au début de la décharge, $t=0$ on a $V(0) = V$
d'où

$$V(t) = \frac{V - W}{R_3 C_2} T - V$$

$$\boxed{V(t) = V - (V - W) A} \quad ; \quad A = \frac{T}{R_3 C_2}$$



$$V = R_3 i_1 + R_4 (i_1 + i_2) = (R_3 + R_4) i_1 + R_4 i_2$$

$$V = \left(\frac{R_3 + R_4}{R_3} \right) (V - W) - R_4 C_3 \frac{dW}{dt}$$

$$\text{car } i_1 = \frac{V - W}{R_3} \quad ; \quad i_2 = -C_3 \frac{dW}{dt}$$

$$\left[\frac{V}{R_4 C_3} - \left(\frac{R_3 + R_4}{R_3} \right) \left(\frac{V - W}{R_4 C_3} \right) \right] dt = -dW$$

Les conditions initiales étant les mêmes que précédemment :

$$W(t) = W - \left[V - \left(\frac{R_3 + R_4}{R_3} \right) (V - W) \right] \frac{T}{R_4 C_3}$$

$$W(t) = W - \left(W (R_3 + R_4) - V R_4 \right) \frac{T}{R_3 R_4 C_3}$$

$$\text{posons } B = (R_3 + R_4) \quad ; \quad R_2 = C \quad \text{et} \quad X = \frac{T}{R_3 R_4 C_3}$$

$$\boxed{W(t) = W - (BW - CV) X}$$

Dans le cas d'une chute de l'énergie du signal, si la tension E reste inférieure à V au delà de 12 ms le décompteur entraîne la fermeture de l'interrupteur I_3 et les constantes B, C, X deviennent alors:

$$B = R_3 + (R_4 \parallel R_5)$$

$$C = R_4 \parallel R_5$$

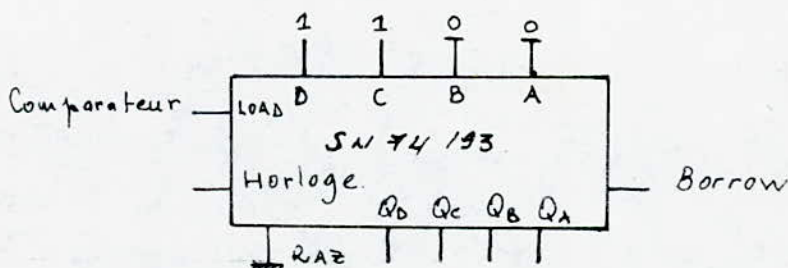
$$X = \frac{T}{R_3 \cdot (R_4 \parallel R_5) \cdot C_3}$$

Ce qui provoque une perte de mémoire plus rapide afin de mieux suivre les crêtes.

4.3.2 | le décompteur.

Étant donné que l'on veut qu'à partir de 12 ms le décompteur doit générer une impulsion de commande qui va pouvoir fermer l'interrupteur I_3 , et rendre ainsi plus rapide la décharge du circuit de mémoire crêtes, on utilisera alors un compteur du type SN 74193 qui est un compteur hexadécimal programmable.

Le schéma du système de décomptage est le suivant.



La sortie du comparateur attaque l'entrée "LOAD"
Ainsi l'autorisation de décomptage se fera quand "LOAD"

Sera à zéro.

Etant donné que l'on veut qu'à partir de 12ms le décompteur commandera la fermeture de l'interrupteur I_3 , On choisira alors la période T de l'horloge " $T=1ms$ " et les entrées D, C, B, A seront programmées comme l'indique la figure précédente.

4.3.3/ Calcul de la période d'horloge =

Le schéma de l'horloge est donné sur la figure 2.2. La formule donnant la période T de cet astable est :

$$T = C \ln 2 (R_A + 2R_B)$$

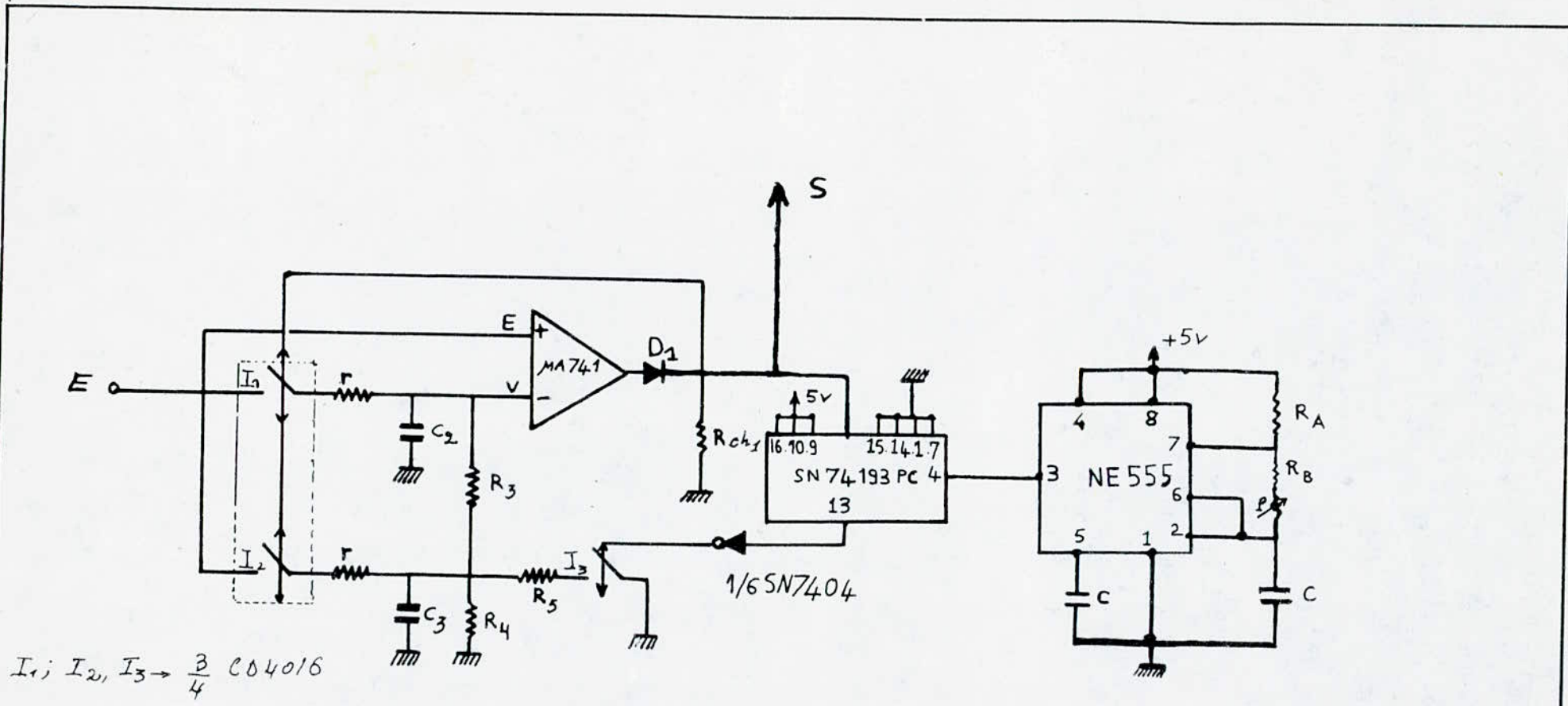
L'ajustable P en série avec la résistance R_B permet de régler le rapport cyclique à 0,5.

4.3.4/ Fonctionnement du décompteur.

Comme on l'a déjà dit dès que l'entrée "LOAD" passe de 1 à 0, c'est à dire dès que la tension E devient inférieure à V , il y a autorisation de décomptage.

D'après les caractéristiques du SN 74LS193, la sortie "Borrow" se met à l'état 1 dès le début de décomptage et dès que les sorties Q_D, Q_C, Q_B, Q_A atteignent la valeur affichée sur les entrées D, C, B, A; cette sortie "Borrow" tombe à zéro formant un creneau de durée égale à celle de l'impulsion d'horloge.

Ainsi en plaçant un inverseur sur la sortie "Borrow" on aura l'état zéro pendant la durée de 12ms, ensuite elle passe à l'état 1; donc par l'intermédiaire de cette sortie on pourra commander la fermeture de la porte analogique I₃.



LE 13LOC COMPRESSEUR

figure 4.6

4.4/ Le détecteur de crêtes.

C'est le traitement proprement dit, il consiste essentiellement en un repérage des crêtes qui marquent la périodicité; le schéma de principe est donné par la figure - 4.7 -

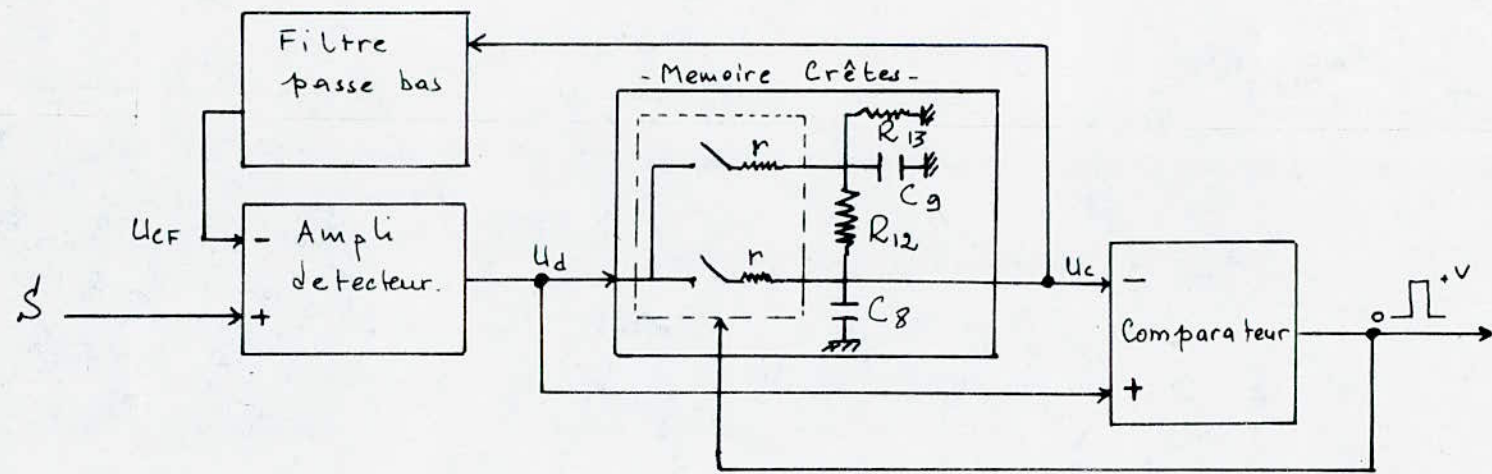
4.4.1/ Fonctionnement:

La sortie S attaque un amplificateur détecteur simple alternance dont le signal de sortie U_d est connecté à un circuit de mémoire crêtes analogue à celui du bloc compresseur qui délivre un signal d'enveloppe de crêtes; U_c et U_d sont connectées aux entrées négatives et positives d'un comparateur, quand la sortie impulsion du comparateur est positive les interrupteurs du circuit de mémoire crêtes sont fermés.

Etant que la tension U_d augmente elle est légèrement supérieure à la tension U_c du condensateur C_8 , la sortie du comparateur est donc positive, et les interrupteurs sont fermés. Mais quand la tension U_d diminue, la sortie du comparateur devient nulle provoquant l'ouverture des interrupteurs, Les capacités C_8 et C_9 se déchargent alors à la masse par l'intermédiaire des résistances R_{12} , R_{13} .

La décroissance de U_c obéit à l'équation différentielle suivante.

$$\frac{d^2 U_c}{dt^2} [R_{12} R_{13} C_8 C_9] + \frac{dU_c}{dt} [R_{12} C_8 + R_{13} C_9 + R_{13} C_8] + U_c = 0$$



- Schema Synoptique du detecteur de crêtes -

- figure -4.7 -

posons $\theta_1 = R_{12} C_8$; $\theta_2 = R_{13} C_9$; $\theta_3 = R_{13} C_8$

l'équation précédente prendra la forme suivante.

$$\frac{d^2 U_c}{dt^2} (\theta_1 \cdot \theta_2) + \frac{d U_c}{dt} (\theta_1 + \theta_2 + \theta_3) + U_c = 0$$

La solution est de la forme.

$$U_c(t) = K_1 e^{x_1 t} + K_2 e^{x_2 t}$$

x_1 ; x_2 sont les racines réelles de l'équation caractéristique :

$$\theta_1 \theta_2 x^2 + (\theta_1 + \theta_2 + \theta_3) x + 1 = 0$$

Les conditions initiales à $t=0$ de l'ouverture des interrupteurs sont :

$$U_c(0) = U_{c\max} \text{ et } U_c'(0) = 0$$

Ceci nous permet de calculer K_1 et K_2 on trouve alors.

$$U_c(t) = \frac{x_2}{x_2 - x_1} U_{c\max} e^{x_1 t} + \frac{x_1}{x_1 - x_2} U_{c\max} e^{x_2 t}$$

C'est la somme de deux décharges exponentielles. La décharge résultante a comme propriété intéressante de démarrer avec une pente nulle, ce qui permet d'éviter la détection d'une crête voisine de la crête d'amplitude maximale et d'amplitude légèrement inférieure. Le signal U_c est retardé dans un filtre passe bas, est appliqué en contre réaction sur l'entrée négative de l'amplificateur détecteur dans le but d'affaiblir le signal U_c après la détection d'une crête significative. Cet affaiblissement va en diminuant au rythme de la décroi-

- ssance de U_c jusqu'à la détection de la nouvelle crête significative.

En plus du retard qu'il introduit, ce filtre permet l'affaiblissement des composantes hautes fréquences engendrées par les discontinuités aux instants où les capacités de la mémoire de crêtes cessent de se décharger.

4.4.2/ description et calcul du détecteur de crêtes.

La fonction de transfert du filtre passe bas de la figure. 4.8. a pour expression.

$$F(p) = \frac{1}{R_{14} R_{15} C_{10} C_{11}} \frac{1}{p^2 + \left(\frac{1}{R_{14} C_{10}} + \frac{1}{R_{15} C_{10}} + \frac{1}{R_{14} C_{10} R_{15} C_{11}} \right)}$$

Les résistances et les capacités R_{14} , R_{15} ; C_{11} ; C_{10} sont choisies de telle façon que le retard $\tau = \tau_1 + \tau_2$

$$\tau = R_{14} C_{10} + R_{15} C_{11} \text{ soit égal à } 1 \text{ ms.}$$

La cellule (R_{10} ; C_7) atténue légèrement les composantes hautes fréquences engendrées par les discontinuités qui peuvent se produire dans le signal à la sortie de l'amplificateur détecteur simple alternance.

Le filtre 1 atténue les hautes fréquences ainsi que l'élimination de la composante continue par la capacité C_6 .

Le principe de la mémoire de crêtes est analogue à celui du bloc compresseur.

Ainsi le détecteur de crêtes émet périodiquement des paquets d'impulsions, le front descendant du signal de

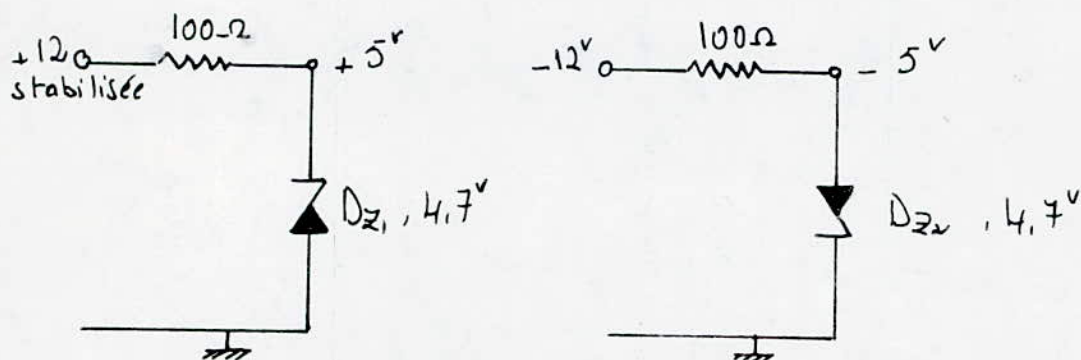
Sortie impulsion marque le sommet de la crête significative de périodicité.

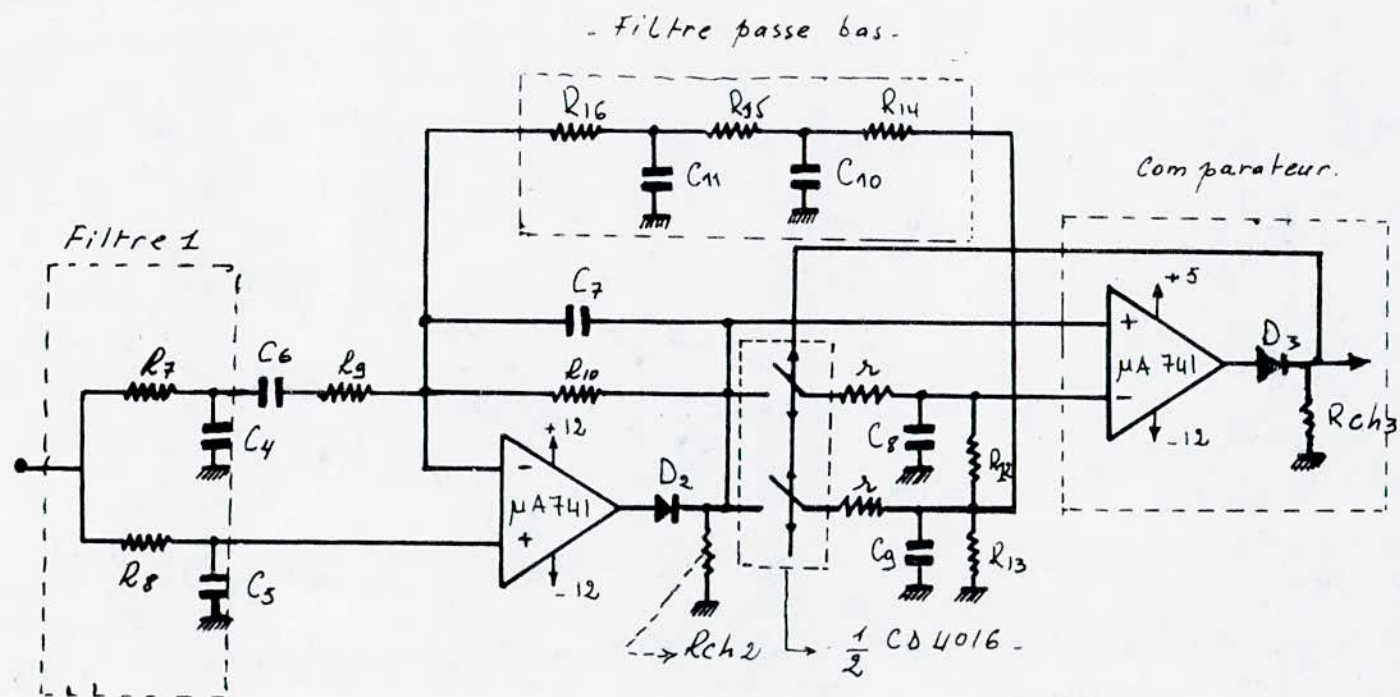
Alors la période T_0 du signal sera le temps entre séparant deux fronts descendants relatifs aux crêtes significatives.

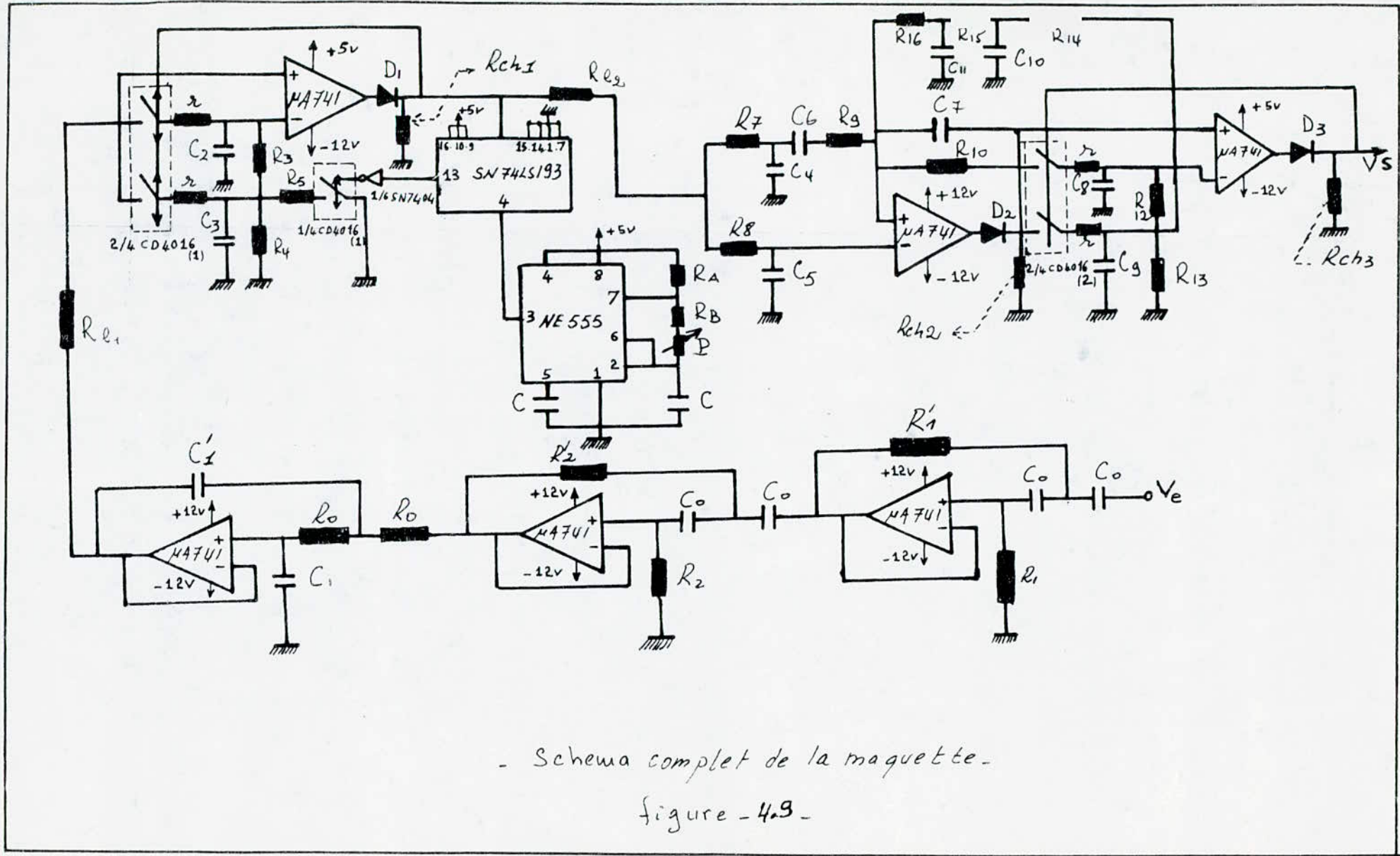
4.4.3/

Etant donné que les circuits (SN74LS193, SN7404), CD4016 nécessitent respectivement $+5\text{V}$ et $(+5\text{V}; -5\text{V})$ comme alimentations, et vue leur faible consommation, on pourrait à partir de $+12\text{V}, -12\text{V}$ avoir une tension stabilisée $(+5\text{V}, -5\text{V})$ par l'intermédiaire de diodes zeners.

Le schéma sera le suivant:







- Schema complet de la maquette -
 figure - 4.9 -

Valeurs des Composants.

1. Bloc filtrage =

$$R_1 = 15 \text{ k}\Omega \pm 5\%$$

$$R'_1 = 10 \text{ k}\Omega \pm 5\%$$

$$C_0 = 0,22 \mu\text{F}$$

$$R_2 = 22 \text{ k}\Omega$$

$$R'_2 = 3,9 \text{ k}\Omega$$

$$R_0 = 12,2 \text{ k}\Omega$$

$$C_1 = 22 \text{ nF}$$

$$C'_1 = 22 \text{ nF}$$

2. Bloc Compresseur.

$$r = 68 \Omega$$

$$R_3 = 36,3 \text{ k}\Omega$$

$$R_4 = 47 \text{ k}\Omega$$

$$R_{ch1} = 15 \text{ k}\Omega$$

$$R_5 = 10 \text{ k}\Omega$$

$$C_2 = 0,22 \mu\text{F}$$

$$C_3 = 0,22 \mu\text{F}$$

$$R_A = 1 \text{ k}\Omega$$

$$R_B = 6,8 \text{ k}\Omega$$

$$C = 0,1 \mu\text{F}$$

$$P = 10 \text{ k}\Omega$$

$$R_{e1} = 180 \Omega$$

Bloc detecteur de crêtes

$$R_7 = 12 \text{ k}\Omega$$

$$R_8 = 10 \text{ k}\Omega$$

$$C_4 = 33 \text{ nF}$$

$$C_5 = 39 \text{ nF}$$

$$C_6 = 0,22 \mu\text{F}$$

$$R_9 = 47 \text{ k}\Omega$$

$$C_7 = 4,7 \text{ nF}$$

$$R_{10} = 30 \text{ k}\Omega$$

$$r = 68 \Omega$$

$$R_{14} = 10 \text{ k}\Omega$$

$$R_{15} = 10 \text{ k}\Omega$$

$$R_{16} = 22 \text{ k}\Omega$$

$$C_{10} = 47 \text{ nF}$$

$$C_{11} = 47 \text{ nF}$$

$$C_8 = 0,22 \mu\text{F}$$

$$C_9 = 0,22 \mu\text{F}$$

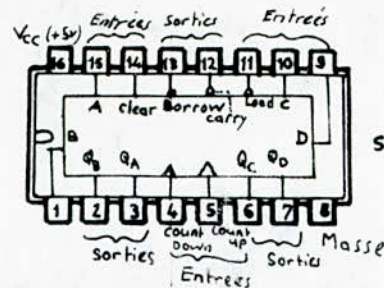
$$R_{12} = 33 \text{ k}\Omega$$

$$R_{13} = 6,8 \text{ k}\Omega$$

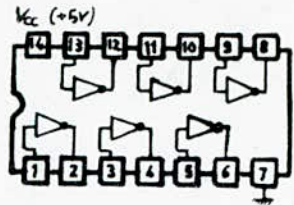
$$R_{ch2} = 10 \text{ k}\Omega$$

$$R_{ch3} = 10 \text{ k}\Omega$$

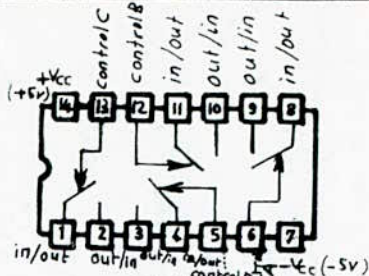
$$R_{e2} = 15 \text{ k}\Omega$$



SN74LS193

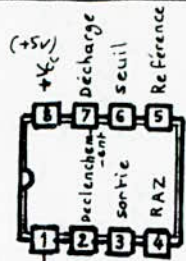


SN7404

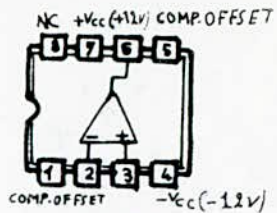


CD 4016

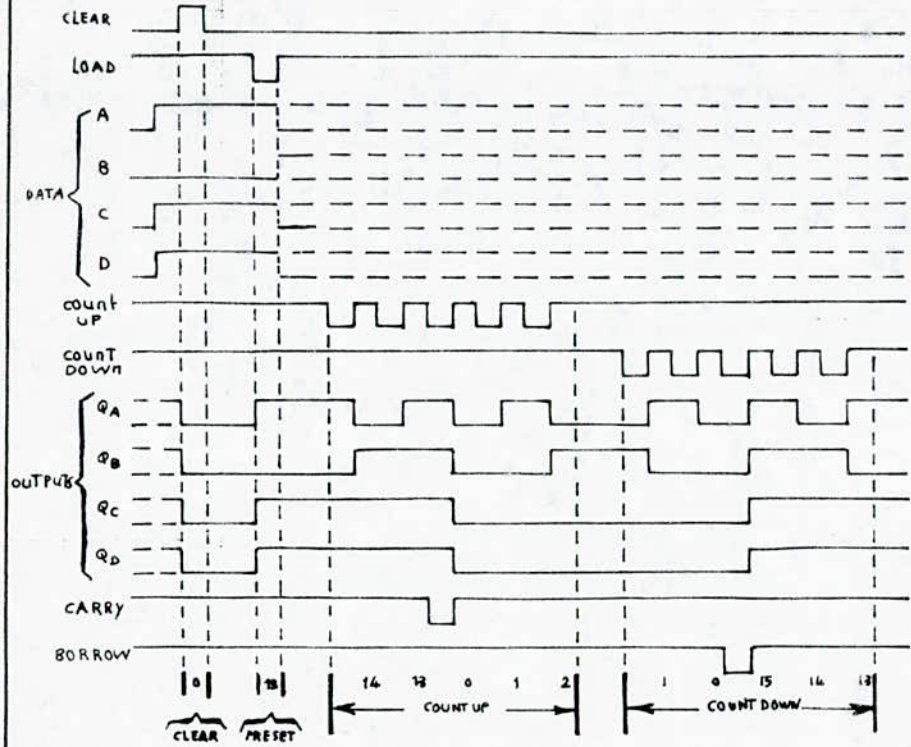
$R_{on} = 300 \Omega$



NE555



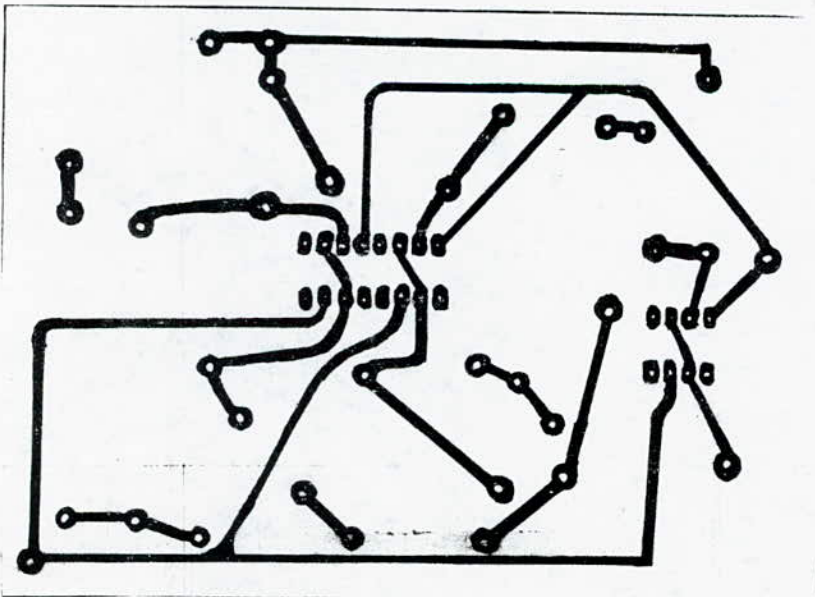
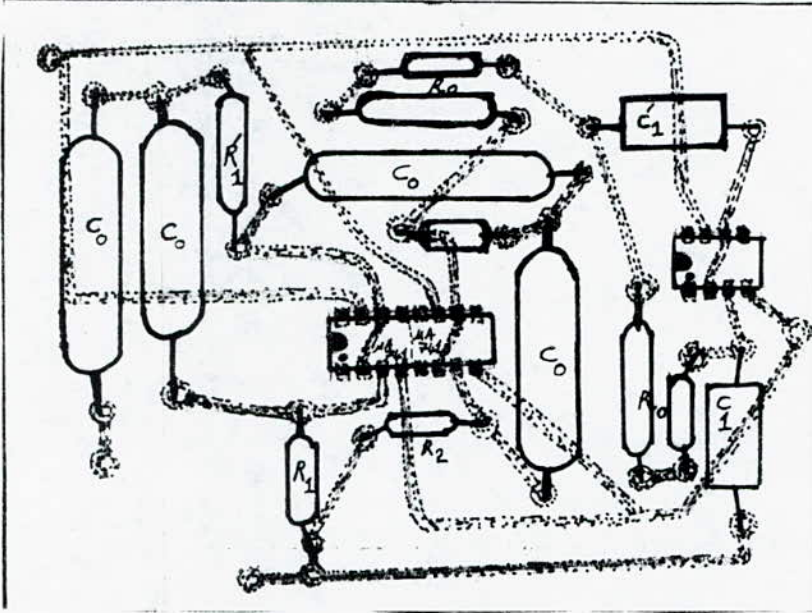
MA741

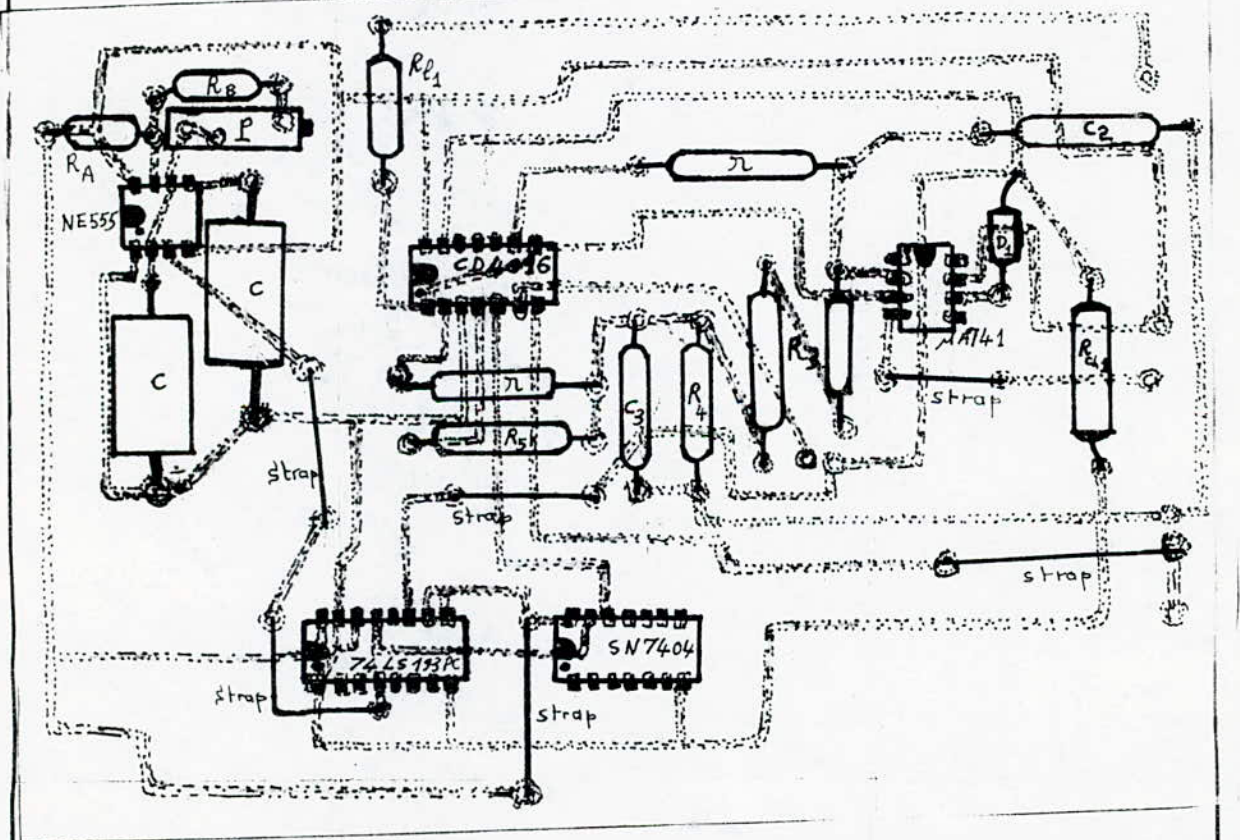
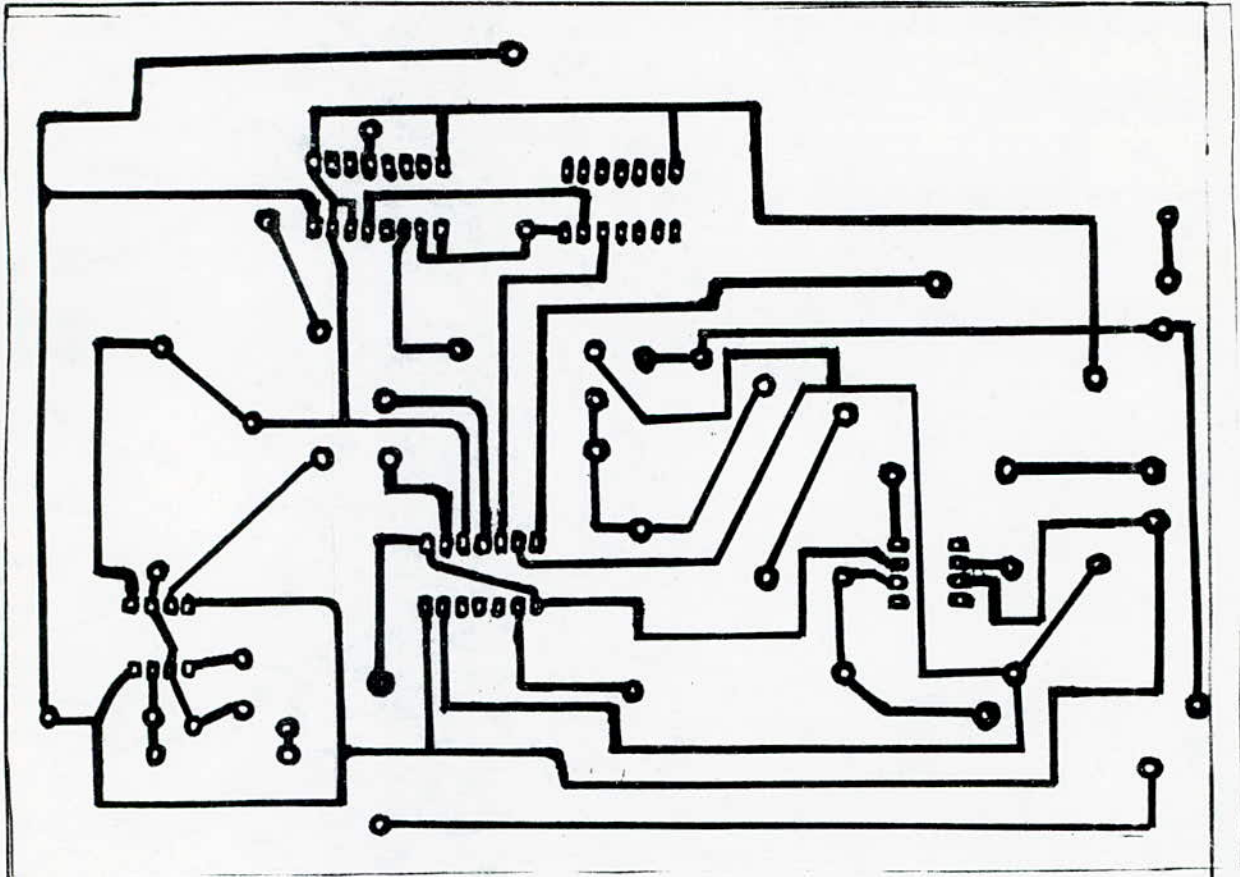


- D₁ → 1N 4001
- D₂ → 1N 4004
- D₃ → 1N 4004
- 2DZ → ZY 88 4,7V

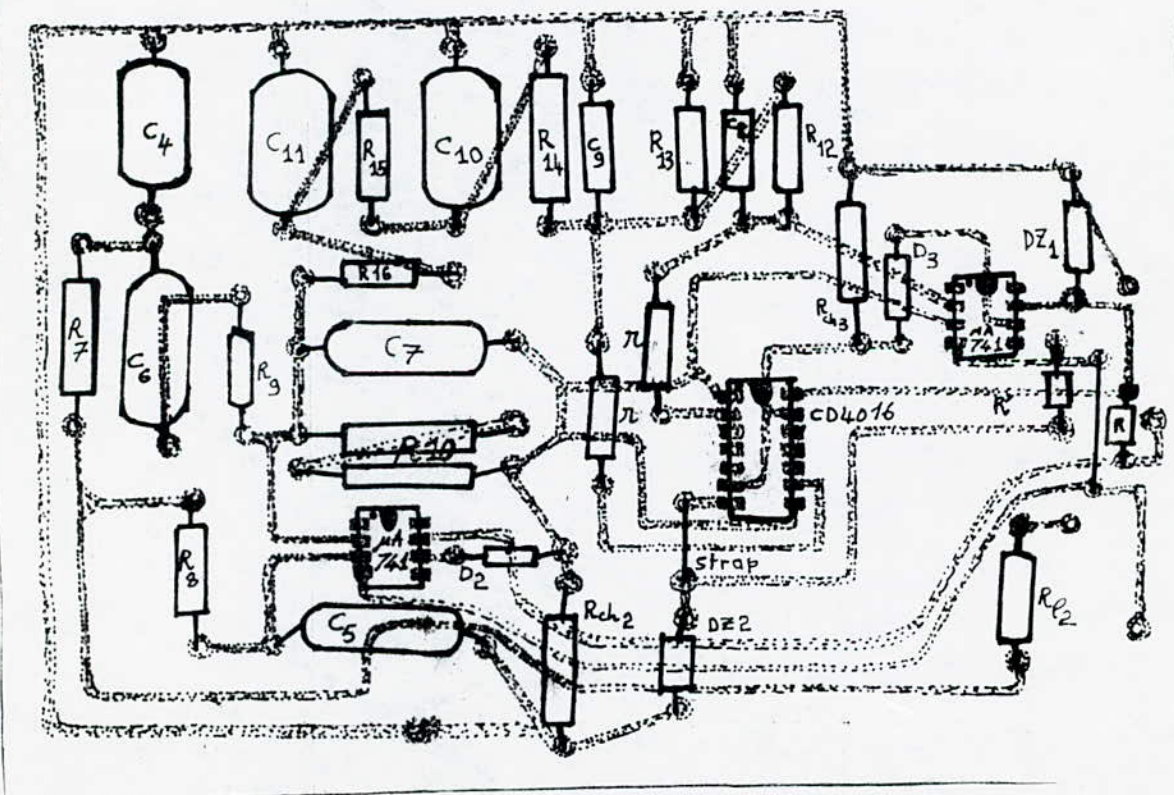
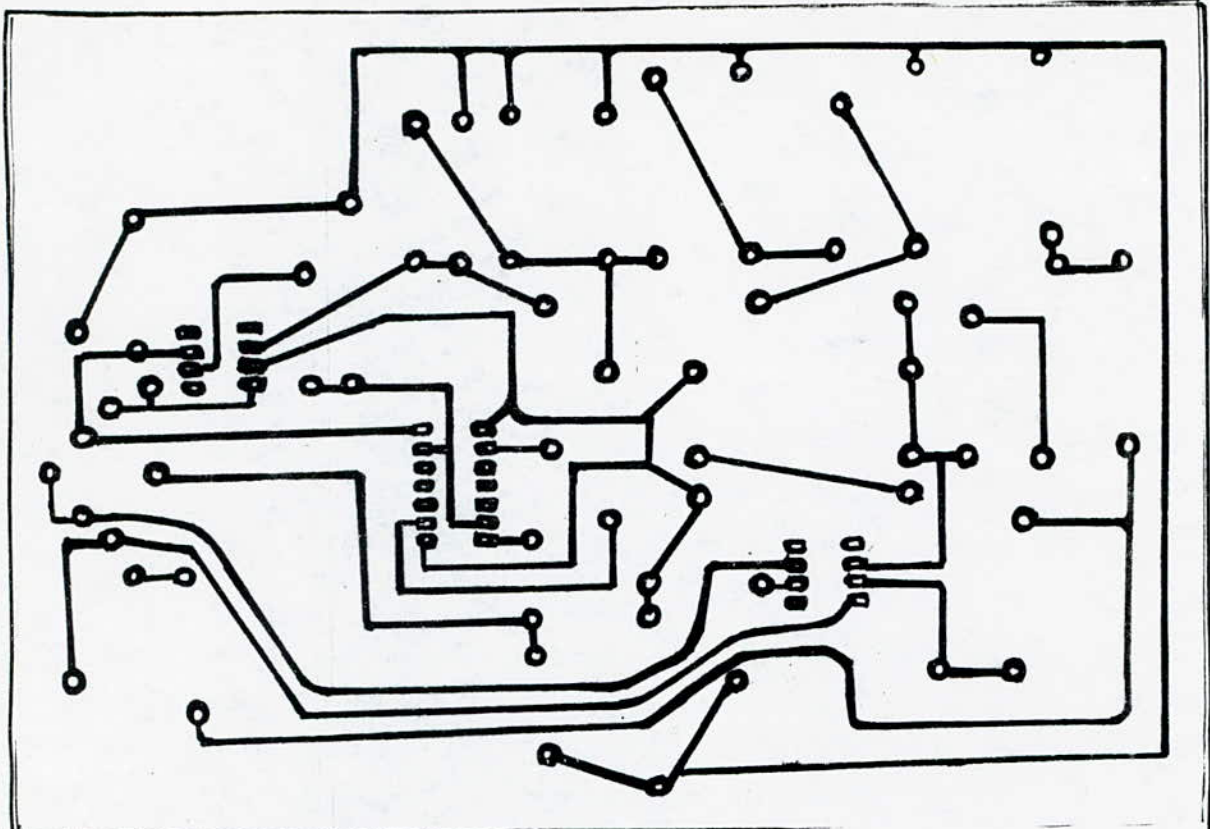
figure 4. 10 -

- Circuit imprimé du bloc filtrage -
- figure





- figure - Circuit imprimé du bloc compresseur.



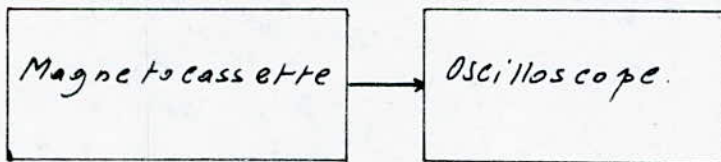
- figure - Circuit imprimé "détecteur de crêtes"

RESULTATS

4.4.4/ Manipulation =

Pour une bonne détection du fondamentale il faut une bonne détection de voisement, or notre maquette ne possède pas cette détection, donc notre manipulation sera limitée à la "mesure" du fondamental de quelques voix voisées par exemples des voyelles.

4.4.5/ Relevé des oscillogrammes =

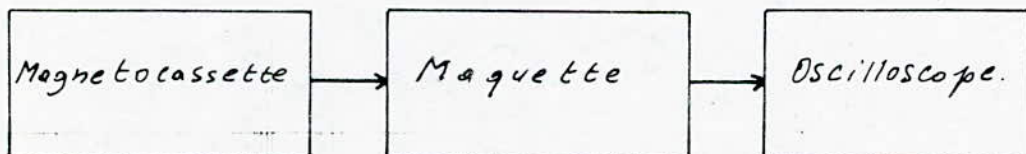


pour cela on a relevé pour trois locuteurs différents dont une femme les oscillogrammes des figures 4.14 et 4.15.

Les différentes voyelles sont ; [a] ; [o] ; [i] ; [œ].
B et C étant des hommes ; C une femme.

4.4.6/ Mesure du fondamental.

La maquette est maintenant placée entre la platine magnetocassette et l'oscilloscope. Comme l'indique le schéma suivant.

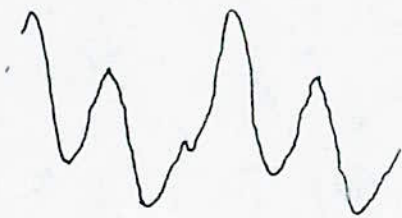


La sortie de la maquette "détecteur de crêtes" émet périodiquement des creneaux, la fréquence fondamentale



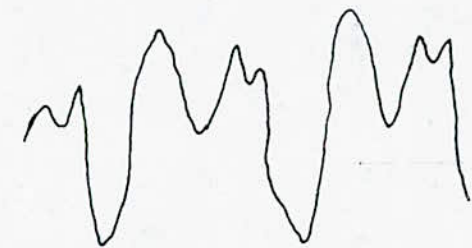
Locuteur. C

[o]



Locuteur. C

[i]



Locuteur. C

[œ]



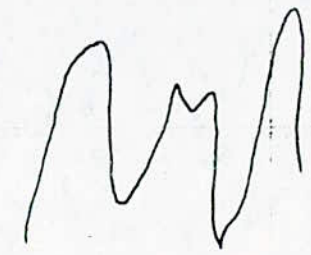
Locuteur. B

[o]



Locuteur. B

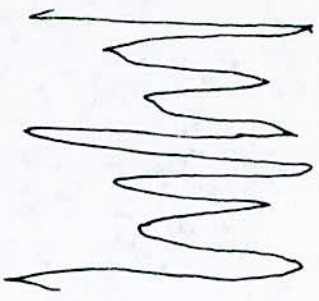
[i]



Locuteur. B.

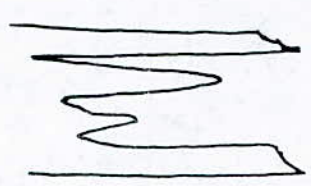
[œ]

figure 4.14



locuteur B.

[a]



locuteur A

[a]



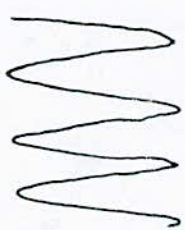
locuteur C.

[a]



locuteur A

[œ]



locut

[i]



locuteur A

[o]

figure 4.15

est calculée en mesurant le temps séparant deux fronts descendants qui représentent les crêtes significatives.

Les valeurs des différentes fréquences fondamentales sont regroupées dans les tableaux suivants.

- Locuteur B -

Voyelles	[a]	[o]	[i]	[œ]
F ₀ (Hz)	200	175	180	180

- Locuteur C -

Voyelles	[a]	[o]	[i]	[œ]
F ₀ (Hz)	260	227	238	220

- Locuteur A -

Voyelles	[a]	[o]	[i]	[œ]
F ₀ (Hz)	330	300	350	300

une constatation pourrait se faire, en ce qui concerne la variabilité interlocuteurs, En effet on voit une différence nette entre les fréquences des différents locuteurs, on pourrait noter aussi que les fréquences du locuteur A sont nettement supérieures à celles des locuteurs B et C, ce qui est évident vu que "A" est une femme.

En outre on pourrait ajouter que certains paramètres relatifs aux crêtes sont pris en compte.

- Les crêtes retenues sont celles dont l'amplitude a augmenté le plus par rapport aux crêtes précédentes.

- Les crêtes retenues sont celles dont la pente est la plus importante. Ce sont elles qui émergent le plus avant que l'écoulement ne leur soit retranché.

CONCLUSION

Ce travail nous a permis d'élargir nos connaissances dans le domaine du traitement du signal de la parole qui est actuellement un nouveau foyer de recherches.

L'étude des différentes méthodes de traitement et d'extraction du fondamental nous a mené à en choisir une des plus éprouvées.

Malgré cela la mesure du pitch reste une opération délicate, c'est pour cela que notre réalisation n'a porté que sur l'alternance positive du signal.

Elle pourra certainement donner de meilleurs résultats si elle était précédée d'un détecteur de voisement ainsi qu'un traitement similaire de l'alternance négative du signal, et moyennant certaines contraintes elle pourra fonctionner pour des recherches en phonétique, en reconnaissance et en synthèse.

Comme perspectives d'avenir et étant donné le progrès enregistré dans le domaine de l'informatique une version numérique du détecteur analogique pourrait être réalisée.

- BIBLIOGRAPHIE -

- J.S. Lienard. "Les processus de la communication parlée" - 1977
- A.M. NOLL. "Cepstrum pitch determination" J.A.S.A. (VOL 41) - 1967
- RABINER/SCHAFER. "Digital processing of speech signals" prentice Hall - 1980
- R. Descout. "Les techniques de synthèse de la parole" 1982
- CNET - Lannion.
- I. EL. MALAWANY. "Synthèse et évaluation comparative de melograpbes numériques" - CNET - 1978
Note technique NT, DAS, SST/10
- ZURCHER. (J.F) "Detection et mesure du fondamental.
6^{eme}. journées d'études sur la parole. "Toulouse"
1977.
- AUMIAUX. Logique binaire et ordinateurs.