

NEUVIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS



NICE du 16 au 20 MAI 1983

DETECTION DE LA PERIODE DU FONDAMENTAL A L'AIDE
DE LA FONCTION "AMDF".

E.AZIZ

J.MENEZ

G.BRUN

Laboratoire de Signaux et Systèmes-ERA 835 du CNRS-Université de Nice-41 Bld Napoléon III-06041-NICE Cedex

RESUME

RESUME

De nombreuses méthodes de détection de la période du fondamental ont été proposées à ce jour. La plupart d'entre elles mettent en oeuvre des algorithmes suffisamment simples pour pouvoir être implantés sur micro-calculateurs, avec un objectif de temps réel. Parmi ces méthodes, la méthode AMDF (Average Magnitude Difference Function) est sans doute la plus utilisée actuellement dans les systèmes de compression de la parole.

Or, ayant constaté expérimentalement que la normalisation de la fonction d'autocorrélation entraînait une amélioration sensible des performances de la méthode d'autocorrélation, nous avons alors voulu voir s'il en était de même pour la méthode AMDF. Les résultats expérimentaux n'étant pas suffisamment convaincants pour justifier l'accroissement sensible de la complexité engendrée par cette normalisation nous avons cherché à améliorer la méthode AMDF en appliquant à la fonction AMDF une technique de lissage élémentaire, et en introduisant un algorithme de poursuite de la mélodie : ce qui a été confirmé par les résultats expérimentaux.

Après un bref rappel des principales méthodes utilisées dans l'estimation de la période du fondamental, nous donnons une description des fonctions d'autocorrélation et d'AMDF normalisées. Cette description est suivie d'une étude comparative de performances des principales méthodes étudiées. Nous décrivons ensuite la technique de lissage et l'algorithme de poursuite utilisés. Cette communication s'achève par la présentation des résultats expérimentaux que nous avons obtenus.

SUMMARY

SUMMARY

Several methods for pitch detection have been proposed till now. Many of them use algorithms which are not complex and then well suitable to an implementation on microprocessors, with a real time objective. Among these methods, the AMDF (Average Magnitude Difference Function) is certainly the most used in speech processing systems. Well, after noticing, by experiment, that the normalization of the autocorrelation function have led to an appreciable improvement in performances of the autocorrelation method, we have attempted to see if it was likewise for the AMDF. The experimental results obtained are however not enough convincing to justify the sensitive increasing of the complexity generated by such normalization ; so we have tried to use an elementary smoothing technic and a tracking algorithm in order to improve the results given by the AMDF method : that's what the experimental results have confirmed.

After a brief review of the principal methods used for pitch detection, we give a description of the normalized autocorrelation and AMDF functions. This description is followed by a comparative study of the performances of the principal methods used. Then we describe the smoothing technic and the tracking algorithm we have proposed. This paper ends with the presentation of the experimental results that we have obtained.



I- INTRODUCTION

La période du fondamental du signal vocal est un paramètre qui intervient dans la réalisation de nombreux systèmes de compression numérique de la parole, en particulier dans celle de la plupart des vocodeurs [1] : d'où la multitude de travaux de recherche ayant pour cadre la détection de la mélodie et le nombre important de méthodes d'estimation de la période du fondamental proposées à ce jour ; méthodes qui, pour la plupart, ont été comparées tant au plan de leurs performances que de leur complexité de mise en oeuvre [2-5]. Certaines d'entre elles comme par exemple, la méthode AMDF, sont actuellement couramment utilisées, car leur simplicité relative leur permet de pouvoir être facilement implantées sur des microprocesseurs, en respectant les contraintes de temps réel. D'autres comme par exemple, la méthode du cepstre ou celle de la fonction d'autocorrélation normalisée, bien que souvent plus performantes, ont été quelque peu délaissées en raison de leur complexité de mise en oeuvre. Or les progrès technologiques récents en matière d'intégration à grande échelle et la baisse relative des coûts de production qu'ils ont entraînée, permettent d'envisager dès à présent la réalisation de méthodes de détection de plus en plus sophistiquées, mettant en oeuvre des algorithmes de plus en plus complexes, compatibles avec des objectifs temps réel. C'est ainsi que nous sommes intéressés à la méthode d'autocorrélation. Nous avons pu constater, expérimentalement que la normalisation de la fonction d'autocorrélation permettait d'accroître sensiblement les performances de cette méthode. Nous avons voulu savoir s'il était possible également d'obtenir une amélioration de la méthode AMDF [6] en normalisant la fonction AMDF. Les résultats obtenus dans ce cas n'étant pas convainquants, nous avons cherché à accroître les performances de cette méthode par utilisation de techniques de lissage et de poursuite. L'objet de cette communication est de présenter l'ensemble des résultats de cette étude.

II- LES PRINCIPALES METHODES DE DETECTION DE LA PERIODE DU FONDAMENTAL

Les méthodes de détection de la période du fondamental du signal vocal peuvent être classées en trois familles :

- Les méthodes qui utilisent les propriétés temporelles du signal de parole ;
- Les méthodes qui utilisent les propriétés spectrales ;
- Les méthodes qui utilisent à la fois les propriétés temporelles et spectrales.

II-1 LES METHODES TEMPORELLES :

Les méthodes qui utilisent les propriétés temporelles travaillent soit directement sur le signal vocal soit sur sa fonction d'autocorrélation. Parmi les méthodes temporelles les plus utilisées, on peut citer :

- a) la méthode d'autocorrélation dans laquelle la fonction d'autocorrélation est définie par :

$$A(J) = \sum_{n=0}^{N-1} S(n)S(n-J) \quad J = 0, 1, 2, \dots, J_{MAX}$$

où $S(n)$: représente le $n^{\text{ième}}$ échantillon du signal
 N : longueur de la suite $\{S(n)\}$ considérée.

La période du fondamental est obtenue par la localisation de l'indice J pour lequel la fonction $A(J)$ passe par un maximum.

Cette méthode est une des plus anciennes méthodes de détection de la mélodie ; elle a donné naissance à de grand nombre de variantes, telle que la méthode d'autocorrélation travaillant sur le signal échantillé, la méthode d'autocorrélation normalisée...

- b) La méthode AMDF (Average Magnitude Difference Function).

Cette méthode utilise la fonction définie par :

$$D(J) = \sum_{n=0}^{N-1} |S(n) - S(n-J)| \quad J = 0, 1, 2, \dots, J_{MAX}$$

Dans ce cas, la période du fondamental relative à la suite d'échantillons $\{S(n)\}$ est obtenue par la localisation de l'indice J pour lequel $D(J)$ prend un minimum. Ross et al. ont montré que, moyennant certaines hypothèses, il existe une relation entre les fonctions $A(J)$ et $D(J)$ précédemment définies.

II-2 LES METHODES SPECTRALES :

Parmi les méthodes spectrales utilisées pour estimer la période du fondamental, la méthode du Cepstre est sans doute la plus connue. Cette méthode dont le principe schématique est donné par la fig. 1, est performante mais complexe à mettre en oeuvre. L'estimation de la période du fondamental se fait par la localisation de l'indice m correspondant au maximum de la suite $\{C(m)\}$. Le maximum est également utilisé pour prendre la décision de voisement.

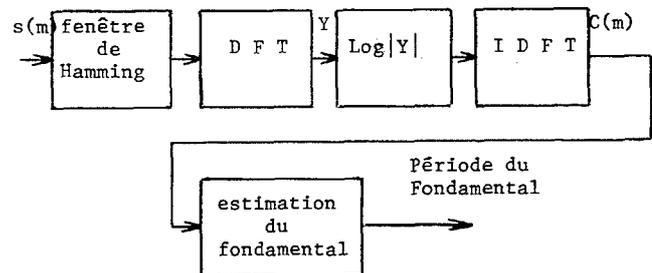


Fig. 1 - Méthode du Cepstre

II-3 LES METHODES MIXTES :

La méthode dite SIFT (Simplified Inverse Filtering Technique) est un exemple typique de méthodes dans lesquelles on exploite les propriétés temporelles et spectrales du signal vocal pour estimer la période du fondamental. Cette méthode utilise successivement le filtrage inverse simplifié du signal vocal et la fonction d'autocorrélation du signal de sortie relatif à ce filtre inverse fig. 2.

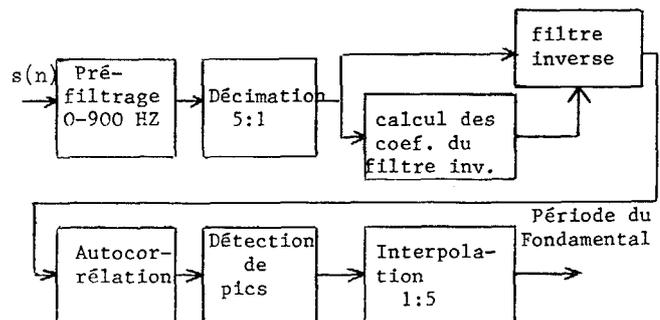


Fig. 2 - Méthode SIFT

III - ESTIMATION DE LA PERIODE DU FONDAMENTAL A L'AIDE DES FONCTIONS D'AUTOCORRELATION ET DE L'AMDF NORMALISEES.

III-1 NORMALISATION DE LA FONCTION D'AUTOCORRELATION

L'analyse de résultats expérimentaux obtenus par la méthode d'autocorrelation mettent en évidence que cette méthode conduit souvent à des résultats erronés de la période du fondamental, dans les zones où l'enveloppe de l'amplitude du signal vocal varie rapidement. C'est par exemple souvent le cas des transitions de voisé/non voisé ou non voisé/voisé. Ces erreurs sont alors essentiellement dues à la non normalisation de l'enveloppe du signal et disparaissent, pour le plupart, lorsque l'on remplace la fonction d'autocorrelation par la fonction d'autocorrélation définie par :

$$r(j) = \frac{\sum_{n=0}^{N-1} S(n)S(n-j)}{\sum_{n=0}^{N-1} S^2(n) \sum_{n=0}^{N-1} S^2(n-j)} \quad j = 0, 1, 2, \dots, J_{MAX}$$

L'estimation de la période du fondamental est obtenue par la localisation du maximum de la fonction $r(j)$.

III-2 NORMALISATION DE LA FONCTION AMDF

On propose de normaliser la fonction AMDF de la manière suivante :

$$F(j) = \sum_{n=0}^{N-1} \left| \frac{s(n)}{(\sum_{n=0}^{N-1} s^2(n))^{1/2}} - \frac{s(n-j)}{(\sum_{n=0}^{N-1} s^2(n-j))^{1/2}} \right|$$

Avec $j=0, 1, 2, 3, \dots, J_{max}$

L'estimation de la période du fondamental est obtenue par la localisation de l'indice J pour lequel $F(j)$ est minimale.

Si l'on considère l'inégalité suivante :

$$\sum |X(k)| \leq \sum X^2(k)$$

où $\{X(k)\}$ est une suite

on peut alors écrire que :

$$F(j) \leq \sum_{n=0}^{N-1} \left(\frac{S(n)}{(\sum_{n=0}^{N-1} S^2(n))^{1/2}} - \frac{S(n-j)}{(\sum_{n=0}^{N-1} S^2(n-j))^{1/2}} \right)^2 \Big)^{1/2}$$

En fait, on peut toujours trouver expérimentalement un facteur α_j , ne dépendant que du décalage j telle que l'on ait :

$$F(j) \sim \alpha_j \sum_{n=0}^{N-1} \left(\frac{S(n)}{(\sum_{n=0}^{N-1} S^2(n))^{1/2}} - \frac{S(n-j)}{(\sum_{n=0}^{N-1} S^2(n-j))^{1/2}} \right)^2 \Big)^{1/2}$$

avec α_j un facteur de proportionnalité déterminé et qui ne dépend que de l'indice j . Sachant que :

$$\sum_{n=0}^{N-1} \frac{S^2(n-j)}{\sum_{n=0}^{N-1} S^2(n-j)} = \sum_{n=0}^{N-1} \frac{S^2(n)}{\sum_{n=0}^{N-1} S^2(n)} = r(0) = 1$$

et en développant le terme entre crochets, on obtient :

$$F(j) \sim \sqrt{2} \alpha_j \sqrt{1-r(j)}$$

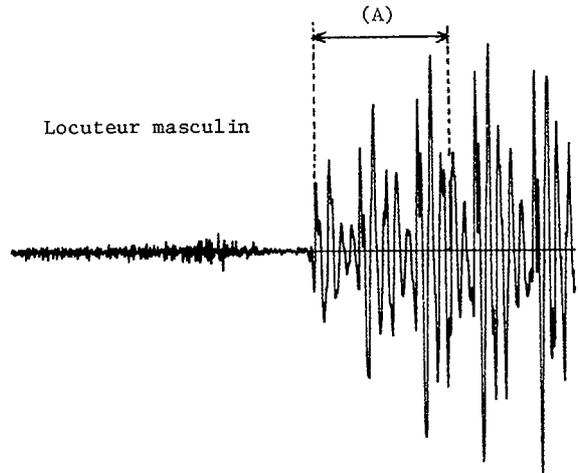
Cette relation montre que

- $F(j)$ est proportionnelle à $\sqrt{1-r(j)}$
- Les minima seront localisés aux points où $r(j)$ est maximale.

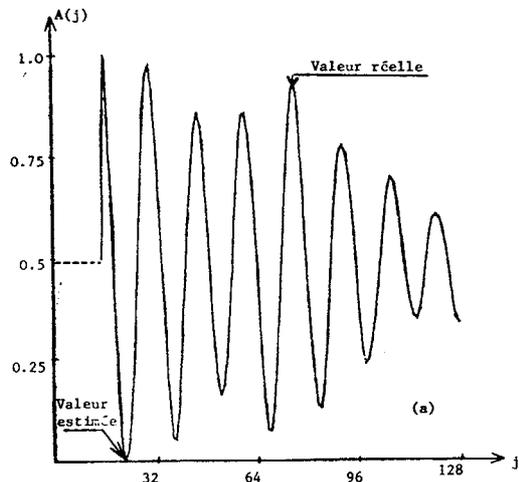
III-3 CONDITIONS D'EXPERIMENTATION

Afin de pouvoir comparer leurs performances, ces méthodes ont été expérimentées sur des signaux réels provenant d'un locuteur masculin et d'un locuteur féminin choisis en fonction des difficultés d'estimation du fondamental qu'ils présentent. Le signal vocal échantillonné à 8 kHz et numérisé sur 12 bits est tout d'abord partitionné en intervalles d'analyse de 256 échantillons. Pour chaque intervalle, les différentes fonctions (Autocorrelation, AMDF) sont obtenues par décalages successifs des 128 premiers échantillons sur les échantillons de l'intervalle.

La figure 3 donne deux exemples des différentes fonctions obtenues sur un locuteur masculin et un locuteur féminin. On indique pour chacune des courbes la valeur réelle de la période du fondamental et la valeur estimée.

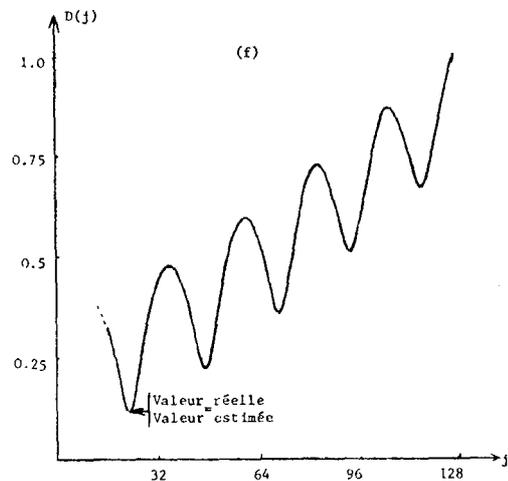
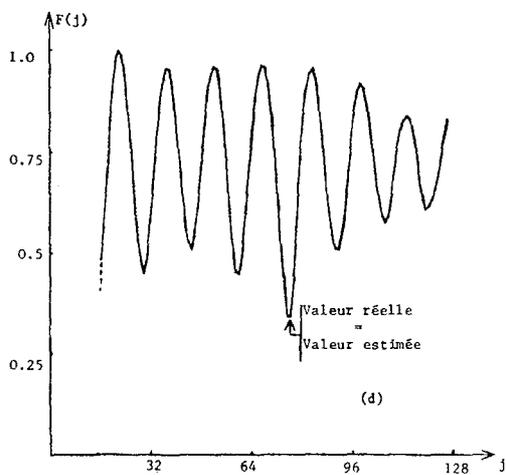
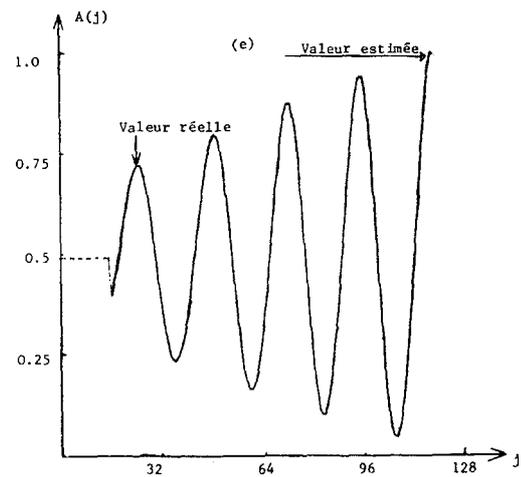
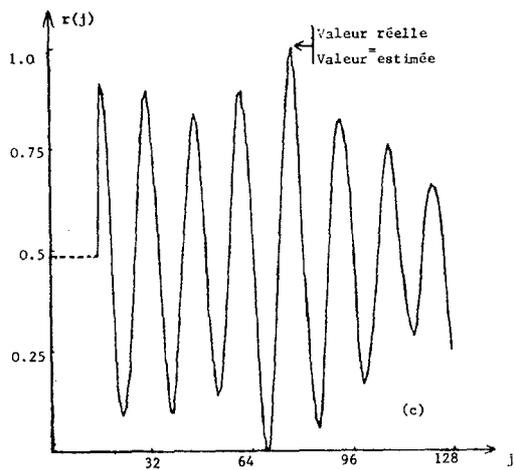
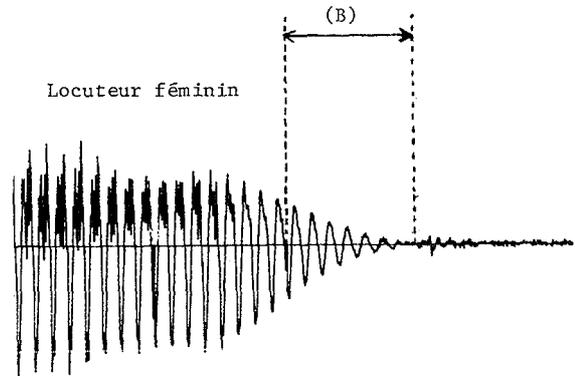
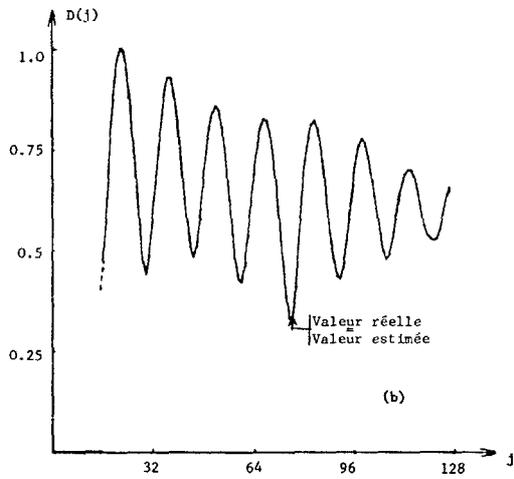


Locuteur masculin





Détection de la période du fondamental à l'aide de la fonction "AMDF"



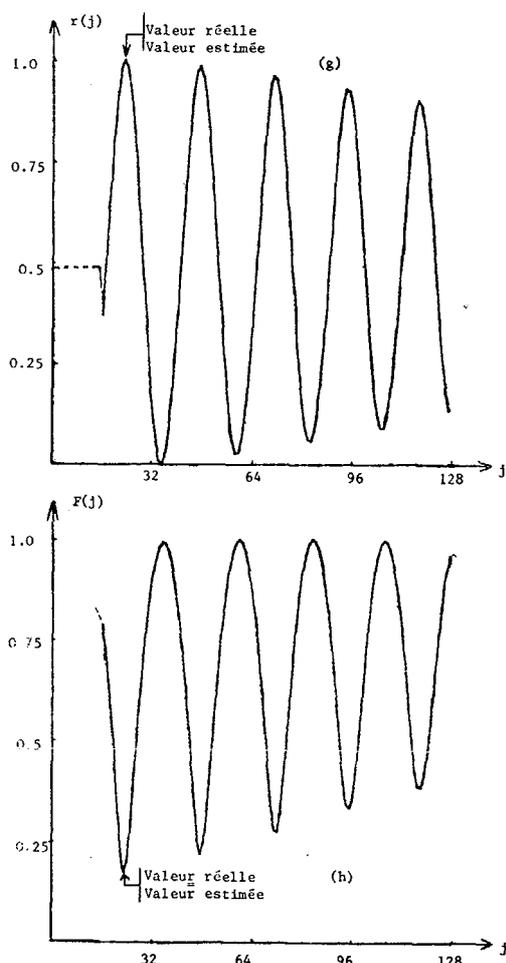


Fig.3-Exemples de courbes obtenues:

- A)-Fenêtre analysée,provenant d'un locuteur masculin;
- a)-b)-c)-d) sont respectivement les fonctions d'autocorrélation,AMDF,autocorrélation normalisée,AMDF normalisée,correspondant à la fenêtre analysée A.
- B)-Fenêtre analysée,provenant d'un locuteur féminin;
- e)-f)-g)-h) sont respectivement les fonctions d'autocorrélation,AMDF,autocorrélation normalisée,AMDF normalisée,correspondant à la fenêtre analysée B.

III-4 COMPARAISON DES RESULTATS EXPERIMENTAUX

Nous avons groupé dans le tableau I les pourcentages d'erreurs obtenus à l'aide chacune des méthodes étudiées :

Méthode	Locuteur Masculin	Locuteur Féminin	Pourcentage Global
	Pourcentage d'erreurs	Pourcentage d'erreurs	
Autocorrélation	9%	25%	17%
AMDF	5%	9%	7%
Autocorrélation normalisée	3%	10%	6,5%
AMDF normalisée	4%	11%	7,5%

Tableau I-Résultat expérimentaux obtenus par les différentes méthodes étudiées.

III-5 CONCLUSION

L'analyse des résultats présentés dans le tableau I montre que si la normalisation de la fonction d'autocorrélation améliore sensiblement les performances de l'autocorrélation, en contre partie elle n'améliore que fort peu les performances de l'AMDF. Cette normalisation engendre un accroissement de complexité de mise en oeuvre important de ces méthodes ; en conclusion, nous pouvons dire que la fonction AMDF sans normalisation offre de meilleur compromis performance-complexité.

IV- AMELIORATION DES PERFORMANCES PAR DES TECHNIQUES DE LISSAGE ET DE POURSUITE

IV-1- LISSAGE DES FONCTIONS D'AUTOCORRELATION ET D'AMDF.

Avant de procéder à l'estimation de la période du fondamental par localisation du maximum de la fonction d'autocorrélation ou du minimum de la fonction AMDF, on lisse les fonctions par filtrage passe-bas défini par :

$$s(j)=0,5f(J+1) + f(J) +0,5f(J-1)$$

où : s(j) représente le j^{ième} échantillon de la fonction lissée

et f(J) le J^{ième} échantillon de la fonction à lisser.

Le module de la réponse harmonique de ce filtre est donné par :

$$|H(e^{j\omega})| = 2 \cos^2 \frac{\omega}{2}$$

et est représenté sur la fig. 4.

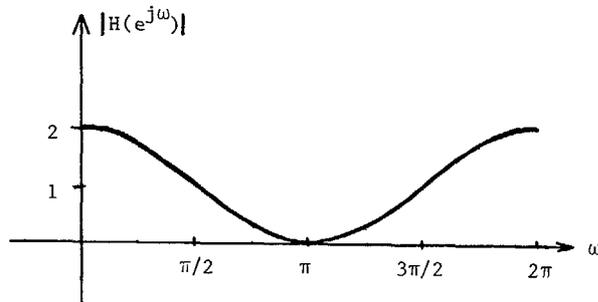


Fig4 -- Réponse harmonique du filtre de lissage.

IV-2 POURSUITE DE LA PERIODE DU FONDAMENTAL

IV-2-1- PRINCIPE

Expérimentalement, on constate que la période du fondamental du signal vocal varie en général lentement d'une période à l'autre lorsqu'on considère le signal d'un son voisé. Cette constatation permet d'envisager d'utiliser une technique de poursuite de ce paramètre, afin d'éliminer tout ou partie des erreurs d'estimation engendrés par les harmoniques. La procédure algorithmique de poursuite que nous proposons est la suivante : Pour chaque estimation de la période du fondamental on commence par déterminer une valeur prédite de cette période donnée par l'expression :

$$MP(j) = \frac{\sum_{i=1}^{j-1} \alpha^i M_i \rho_i}{\sum_{i=1}^{j-1} \alpha^i \rho_i}$$



DETECTION DE LA PERIODE DU FONDAMENTAL A L'AIDE
DE LA FONCTION "AMDF"

- où $MP(J)$: représente la $J^{i\text{ème}}$ valeur prédite de la période du fondamental.
 M_i : représente la $i^{i\text{ème}}$ valeur estimée de ce paramètre.
 α : coefficient d'amortissement permettant de rendre la mémoire du filtre constitué par le numérateur évanescence. La valeur de ce coefficient est de l'ordre de 0,8 à 0,9.
 ρ_i : coefficient d'autocorrélation normalisé correspondant au décalage M_i .

La valeur prédite est utilisée pour déterminer l'étendue de l'intervalle sur lequel il convient de calculer les points de la fonction d'autocorrélation ou de la fonction AMDF, pour en localiser le maximum (Autocorrélation) ou le minimum (AMDF). La localisation du maximum ou du minimum fournit la nouvelle valeur estimée de la période du fondamental. Cette étendue est donnée par :

$$J_{MIN} = 0,8 MP (j)$$

$$J_{MAX} = 1,2 MP (J)$$

L'initialisation de la procédure de poursuite est réalisée de la manière suivante :

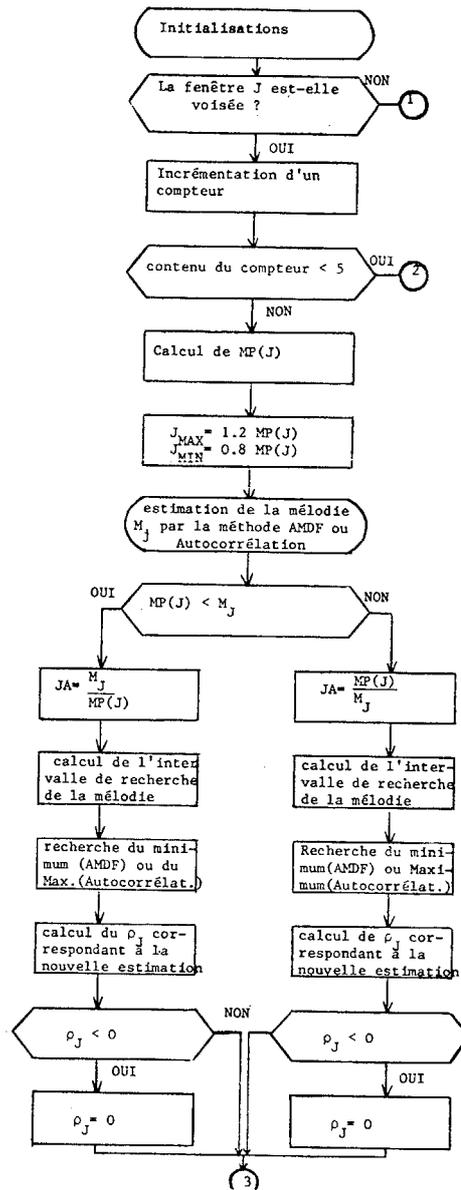
On démarre un compteur qui est incrémenté à chaque nouvelle estimation de la période du fondamental. Tant qu'il n'a pas atteint la valeur 6, l'estimation de cette période est obtenue par localisation du maximum (autocorrélation) ou du minimum (AMDF) sur toute l'étendue de recherche (128 points). Pour chaque valeur estimée M_j , on calcule le coefficient ρ_j correspondant. Lorsque le compteur atteint la valeur 6, on calcule la valeur prédite $MP(6)$. On en déduit alors l'étendue de l'intervalle de recherche du maximum (autocorrélation) ou du minimum (AMDF). Ce qui permet de déterminer la valeur estimée M_6 et le coefficient ρ_6 , d'où la nouvelle valeur prédite $MP(7)$ et ainsi de suite. Le compteur est remis à zéro à chaque détection de non voisement.

IV-2-2-Organigramme de poursuite de la mélodie.

Cet organigramme est donné par la figure 5.

V- COMPARAISON DES RESULTATS OBTENUS :

Afin de montrer l'amélioration éventuelle des performances apportée par le lissage et la poursuite, nous avons rassemblé dans le tableau II les résultats obtenus expérimentalement sur les mêmes phases et les mêmes locuteurs que précédemment.



Méthode	voix masculine		voix féminine		Pourcentage Global
	avec lissage	avec lissage + Poursuite	avec lissage	avec lissage + Poursuite	
Autocorrélation	9%	.95%	2.5%	1%	0.17%
AMDF	4%	.95%	5.5%	< 1%	<0.97%
Autocorrélation normalisée	3.5%	.95%	9,5%	<1%	<0,97%
AMDF normalisée	3%	0,25%	7,5%	2%	1,5%

Tableau II-Résultats obtenus par la technique de lissage et de poursuite.



DETECTION DE LA PERIODE DU FONDAMENTAL A L'AIDE
DE LA FONCTION "AMDF"

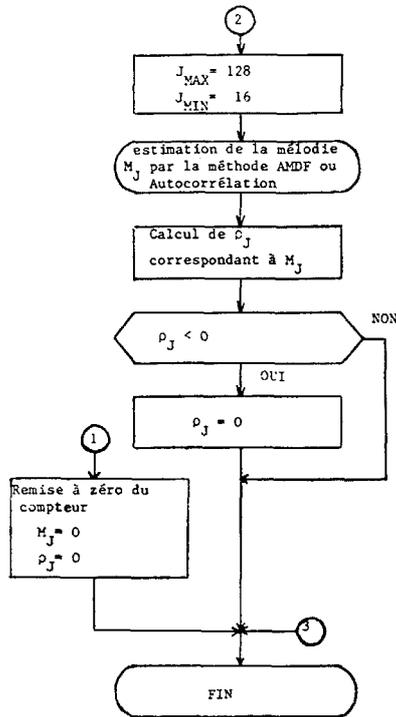


Fig.5- Organigramme de la poursuite du fondamental.

- [3] CAROL A. Mc CONEGAL, LAWRENCE R. RABINER, AARON ROSENBERG
IEEE - volume ASSP-25 n°3 June 1977.
"A Subjective Evaluation Of Pitch Detection Methods using LPC Synthesized Speech."
- [4] T.V. SREENIVAS - P.V.S. RAO
Signal processing 3 (1981) 277-284
North-Holland - Publishing company.
"Functional demarcation of pitch"
- [5] LAWRENCE R. RABINER - Michael J. CHENG
AARON E. ROSENBERG - CAROL A. Mc CONEGAL
IEEE Trans ASSP 24 n°5 oct 76
"A comparative performance study of several pitch detection algorithms".
- [6] ROSS M.J. - SHAFFER H.L. - COHEN A. - FREUDBERG R. MANLEY H.J.
IEEE Trans ASSP 1974-22 n°5
"Average Magnitude Difference Function Pitch Detection".

VI- CONCLUSION

Au vu de l'ensemble des résultats, on peut constater que :

- La normalisation de la fonction d'autocorrélation permet une amélioration importante des performances de la méthode d'autocorrélation, il en est pas de même pour la méthode AMDF. Au demeurant, il ne faut pas perdre de vue que cette normalisation entraîne un accroissement important du volume de calcul.
- Le lissage n'apporte pas de grandes améliorations au plan des résultats.
- La poursuite permet d'améliorer très sensiblement les performances des méthodes d'autocorrélation et d'AMDF.

Pour conclure, il apparaît que la méthode AMDF sans normalisation mais avec poursuite de la période du fondamental donne d'aussi bon résultats que ceux obtenus à l'aide de la méthode d'autocorrélation normalisée avec poursuite, mais à moindre complexité. Elle est sans nul doute celle qui offre le meilleur compromis performances - Complexité

VII - BIBLIOGRAPHIE

- [1] J. POTAGE - L'onde électrique 1981 - Volume 61 n° 8-9
"Les vocodeurs numérisation de la parole à des débits inférieurs à 5 Kbits/s".
- [2] LAWRENCE R. RABINER - IEEE Volume ASSP-25 n°1 Février 1977.
"On the use of autocorrelation Analysis for pitch detection".

