

# SEPTIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 28 MAI au 2 JUIN 1979

---

UN CORRELATEUR HYBRIDE POUR L'ANALYSE EN TEMPS REEL DU SIGNAL VOCAL

AN HYBRID CORRELATOR FOR REAL TIME SPEECH ANALYSIS

J. LEROUX, J. PRADO, J.C. DOMENGER

DEPARTEMENT SYSTEMES ET COMMUNICATIONS  
ENST. 46 rue Barrault, 75634 PARIS - Cedex 13

---

## RESUME

L'analyse des signaux, et en particulier de la parole, par prédiction linéaire nécessite le calcul de 11 à 13 coefficients de corrélation du signal à analyser en moins de 20 msec, de façon à calculer à partir de ceux ci les coefficients d'un filtre récursif permettant une synthèse.

Il est possible d'obtenir une estimation de cette corrélation par des techniques hybrides (multiplieurs digitaux analogiques, registres à décalage, filtres analogiques). Nous montrons ici la justification théorique de cette méthode et la façon dont elle a été implantée en pratique parallèlement à l'analyse d'autres traits acoustiques (amplitude du signal, voisement, période du fondamental) sous la forme d'un périphérique de microprocesseur (TMS 9900) pour effectuer une analyse - synthèse de la parole en temps réel utilisant des moyens informatiques réduits, ce corrélateur pour l'analyse et un filtre hybride programmable pour la synthèse.

## SUMMARY

Signal analysis by linear prediction has become a widely used method. For speech analysis, this method requires the computation of 11 or 13 correlation coefficients from which we can estimate the coefficients of a recursive filter.

This all pole filter is driven by a wide band source and generates synthetic speech.

The estimation of the correlation coefficients can be achieved by hybrid techniques (i.e. digital analog multipliers, shift registers, analog filters). In this paper we show the theoretical basis of the estimation and how it has been applied, with the analog measurement of other acoustical speech characteristics (amplitude, voicing, pitch frequency) as an external operator of a microprocessor (TMS 9900).

This realisation takes place in a real time speech processing system, which includes a small size computer, the hybrid correlator for analysis and a programmable hybrid lattice filter for the synthesis.



## INTRODUCTION

Les méthodes de prédiction linéaire (1. Markel 1976) permettent de calculer à partir des coefficients de corrélation d'un signal échantillonné les paramètres d'un filtre numérique stable dont la fonction de transfert est de la forme

$$F(z) = \frac{1}{\sum a_i z^{-i}}$$

Le signal synthétisé par ce filtre et une entrée à densité spectrale constante (impulsions, bruit blanc) a une densité spectrale très proche de celle du signal analysé.

Cette technique est bien adaptée à l'analyse et à la synthèse de la parole : le signal vocal est échantillonné toutes les 100  $\mu$ s, sa corrélation (11 ou 13 coefficients  $r_k$ ) est calculée toutes les 10 ou 20 ms sur 200 échantillons. On en déduit par un calcul numérique (1. Markel 1976) (2. Leroux 1978) les paramètres du filtre  $F(z)$  d'ordre 10 ou 12 qui sera utilisé pour la synthèse du signal pendant cette période de 20 ms.

Il faut parallèlement mesurer l'amplitude du signal, la fréquence du fondamental (intonation) de la voix, l'énergie du bruit pour les sons fricatifs, de façon à générer une entrée qui sera appliquée au filtre (Fig. 1).

L'implantation de cette méthode en temps réel nécessite des moyens de calcul assez puissants (300 000 multiplications par seconde). Au lieu d'utiliser des calculateurs numériques perfectionnés (3. Hofstetter 1977, 4. Wiggins 1978), nous avons tenté d'utiliser une technologie hybride, sous la forme de périphériques de microprocesseur (TMS 9900). Le filtre hybride programmable utilisé pour la synthèse (5. Gueguen 1975, 6. Leroux 1979) a d'abord été réalisé et donné des résultats satisfaisants. Nous avons ensuite mis au point une méthode permettant d'utiliser cette technologie hybride pour le calcul des coefficients de corrélation.

## ESTIMATION DES COEFFICIENTS DE CORRELATION

La méthode habituelle de calcul des coefficients de corrélation (7. Tretter 1976) sur un horizon temporel donné (0, N) est de multiplier le signal  $S_n$  par une fenêtre de Hamming  $H_n$  nulle en dehors de cet

horizon et valant

$$0 < n < N \quad H_n = \alpha - \beta \cos 2\pi \frac{n}{N}$$

$$\alpha = 0.54, \beta = 0.46$$

soit  $S_n^1 = H_n S_n$

L'estimation de la corrélation est alors

$$r_k = \frac{1}{N+1} \sum_{n=0}^N S_n^1 S_{n-k}^1$$

$$r_k = \frac{1}{N+1} \sum_{n=0}^N S_n S_{n-k} H_n H_{n-k}$$

Si le signal  $S_n S_{n-k}$  est passé dans un filtre de réponse impulsionnelle  $G_{m,k}$

$$G_{m,k} = H_{N-m} H_{N+k-m}$$

la sortie de ce filtre est à l'instant N le coefficient de corrélation  $r_k$ .

Comme on calcule peu de coefficients de corrélation ( $k \ll N$ ), on en déduit qu'en première approximation les filtres permettant d'obtenir les différents coefficients  $r_k$  sont des filtres passe bas identiques dont la réponse impulsionnelle est donnée par

$$G_{m,0} = (\alpha - \beta \cos 2\pi \frac{m}{N})^2$$

(L'écart maximum pour  $N = 200$ , entre les deux réponses  $G_{m,0}$  et  $G_{m,10}$  est de l'ordre de 1/100 de l'amplitude maximum).

La fenêtre de pondération  $H_n$  du signal temporel peut prendre plusieurs formes (8. Stanley 1975). Il existe de la même façon une certaine latitude pour choisir le filtre se rapprochant de  $G_{m,0}$  qui servira au calcul de  $r_k$ .

De plus, les coefficients de corrélation d'un signal de parole varient lentement et peuvent être considérés comme constant pendant des durées supérieures à N. Nous chercherons donc une approximation de  $G_{m,0}$  sous la forme d'un filtre passe bas respectant ses caractéristiques principales :

- réponse impulsionnelle quasi symétrique
- passant par un maximum pour  $m = N/2$
- croissant régulièrement de 0 au maximum pour m variant de 0 à N/2

- décroissant régulièrement vers 0 de  $N/2$  à  $N$
- pratiquement nulle en dehors de l'intervalle  $(0, N)$

et nous utiliserons ce filtre pour calculer l'estimation de  $r_k$  à partir du signal  $S_n S_{n-k}$ .

Un filtre aisément réalisable correspondant à ces spécifications est le filtre analogique de Bessel (9. Bildstein 1972) d'ordre 6 ayant une fréquence de coupure de 40 Hz, correspondant à  $N = 200$ .

Sa réponse impulsionnelle est proche de celle de  $G_{m,0}$  (Fig. 2) correspondant à la fenêtre de Hamming.

#### IMPLANTATION PRATIQUE

Pour utiliser ce filtre il faudra lui fournir une entrée valant  $S_n S_{n-k}$  sous une forme analogique. Ce dernier signal est obtenu par un multiplieur digital analogique : le signal à analyser est échantillonné ( $S_n$ ), il est aussi numérisé et retardé pour donner  $S_{n-k}$ . La valeur  $S_n$  est bloquée entre les instants d'échantillonnage. Le résultat du produit  $S_n S_{n-k}$  (sortie du multiplieur) est échantillonné et bloqué puis filtré par un des filtres de Bessel (Fig. 3). Les 11 sorties des filtres (pour  $k = 0, \dots, 10$ ) sont ensuite multiplexées, numérisées sur 12 e.b. et envoyées au microprocesseur (Fig. 4).

Parallèlement, la mesure de l'amplitude des hautes fréquences ( $>2000$  Hz) servira à déterminer l'énergie  $B$  du bruit nécessaire à la génération de certains sons. Un système analogique de détection de crêtes appliqué au signal filtré passe bas ( $<300$  Hz) permet de mesurer l'amplitude  $V$  et la période  $T$  des "impulsions" quasi périodiques provenant des vibrations des cordes vocales : après la détection d'un maximum du signal  $s(t)$  d'amplitude  $V_1$  à l'instant  $t_1$ , un signal "suiveur"  $u(t)$  est bloqué à la valeur  $V_1$  pendant le temps  $\tau$ . Il décroît ensuite exponentiellement avec une constante de temps  $\tau^1$ . Lorsqu'il rencontre le signal  $s(t)$ ,  $u(t)$  suit  $s(t)$  jusqu'à ce qu'il commence à décroître, à l'instant  $t_2$ . Son amplitude est alors  $V_2$ , le filtrage de  $u(t)$  et de la différence de temps  $(t_1 - t_2)$  (calculée par un intégrateur) donne une mesure analogique de  $V$  et de  $T$  (Fig. 5). Ces 3 données sont aussi multiplexées, numérisées et envoyées au microprocesseur. Elles serviront à reconstruire l'entrée du filtre pour la synthèse du signal vocal.

Remarques :

Cette méthode d'estimation de la corrélation n'est valable que pour des signaux stationnaires pendant des périodes assez longues par rapport au temps de propagation de groupe du filtre de Bessel. Il faut de plus que le signal ne comporte pas de fréquences proches de la fréquence de coupure du filtre (40 Hz) : le signal de parole est filtré passe haut ( $>300$  Hz) avant l'analyse. La précision de l'estimation est alors de  $5 \cdot 10^{-3}$  environ.

#### APPLICATIONS

Les coefficients de corrélation sont disponibles en temps réel et permettent à un microprocesseur TMS 9900 d'en déduire par un programme dont le temps d'exécution est de 15 ms les coefficients du filtre récursif utilisé à la synthèse de la parole (l'analyse synthèse de la parole est ainsi effectuée en temps réel).

Pour obtenir une meilleure qualité de synthèse, cette chaîne d'analyse est mise en parallèle avec un second microprocesseur qui a pour fonction l'analyse des basses fréquences, le corrélateur ne servant qu'à l'analyse des fréquences supérieures à 600 Hz. (10. Mourikis 1978).

Les données fournies par ce corrélateur en temps réel sont aussi utilisées dans les recherches sur la reconnaissance de la parole (11. Grenier, Miclet 1978) où elles permettent de diminuer de façon importante la durée des opérations à effectuer.

#### Références

1. JD Markel, Linear prediction of speech, Verlag-1976.
2. J Leroux, C Gueguen, A fixed point computation of partial correlation coefficients, IEEE Transactions on ASSP 23, Vol n°3 June 1977.
3. EH Hofstetter et AL, Microprocesseur realization of a linear prediction vocoder, IEEE on ASSP 25, Vol n°5 Octobre 1977.
4. R Wiggins et AL, Three chip system synthesizes human speech, Electronics Vol 151 n°18 August 31 1978.
5. C Gueguen et AL, Un synthétiseur à structure programmable, 6ième JEP du GALF, Toulouse 1975.



UN CORRELATEUR HYBRIDE POUR L'ANALYSE EN TEMPS REEL DU SIGNAL VOCAL  
AN HYBRID CORRELATOR FOR REAL TIME SPEECH ANALYSIS

6. J Leroux, L Miclet, Transcription orthographique phonémique et synthèse de parole, 10ième JEP du GALF, Grenoble 1979.
7. SA Tretter, Introduction to signal processing John Wiley & Sons 1976.
8. WD Stanley, Digital signal processing Reston Publishing Company 1975.
9. P Bildstein, Filtres Actifs, Editions Radio 1972.
10. X Mourikis et AL, Low bit rate high quality speech transmission, International Conference on information theory, system theory in digital communication, Berlin Septembre 1978.
11. Y Grenier, L Miclet, Reconnaissance des signaux par prédiction linéaire, Congrès AFCET IRIA, Chatenay Malabry, Février 1978.

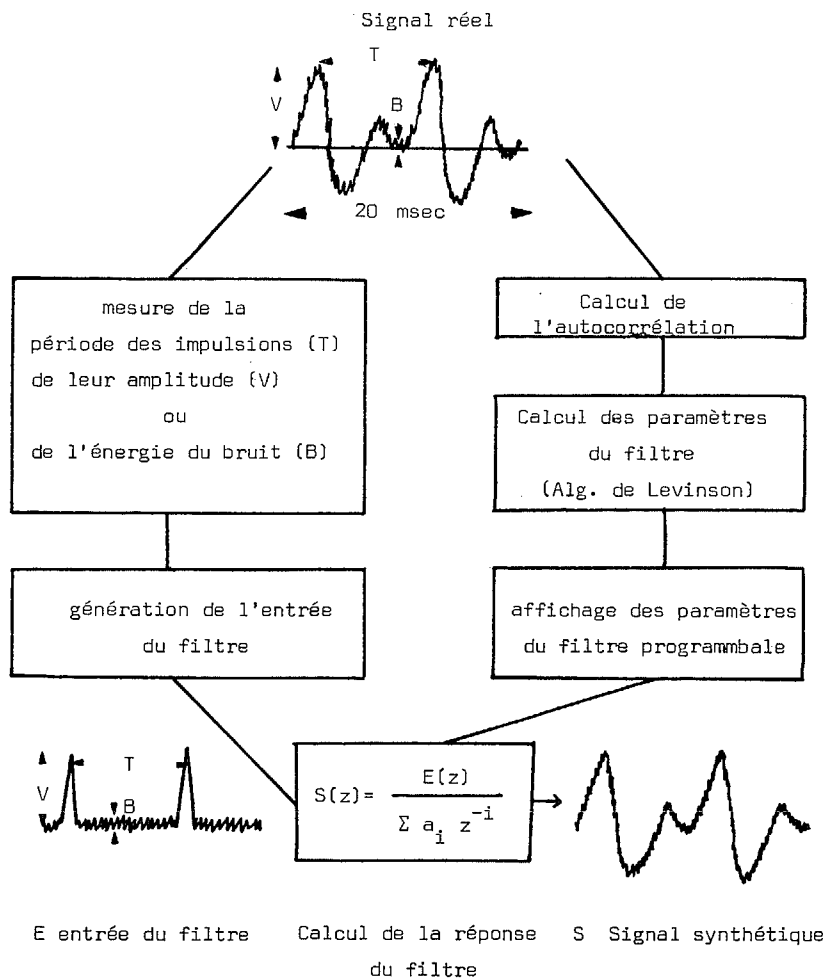


Figure 1 : Schéma de l'analyse par prédiction linéaire

UN CORRELATEUR HYBRIDE POUR L'ANALYSE EN TEMPS REEL DU SIGNAL VOCAL  
 AN HYBRID CORRELATOR FOR REAL TIME SPEECH ANALYSIS

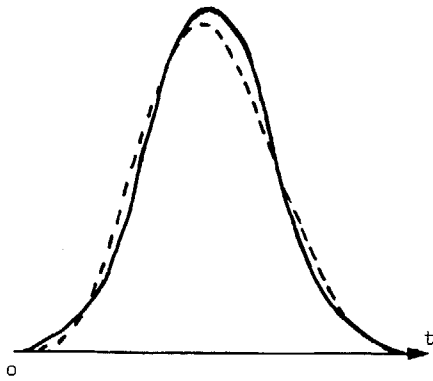


Figure 2 : réponses impulsionnelles  
 ----- du filtre de Bessel  
 ————— du filtre correspondant  
 à la fenêtre de Hamming

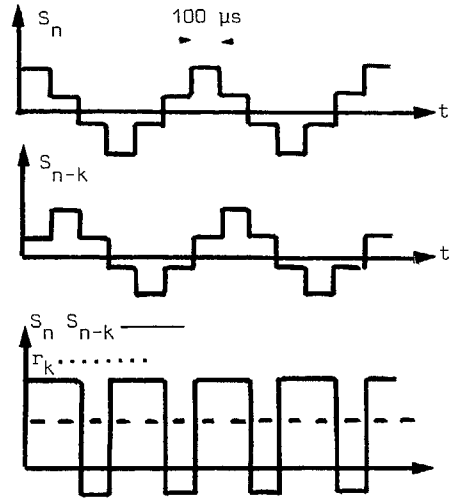
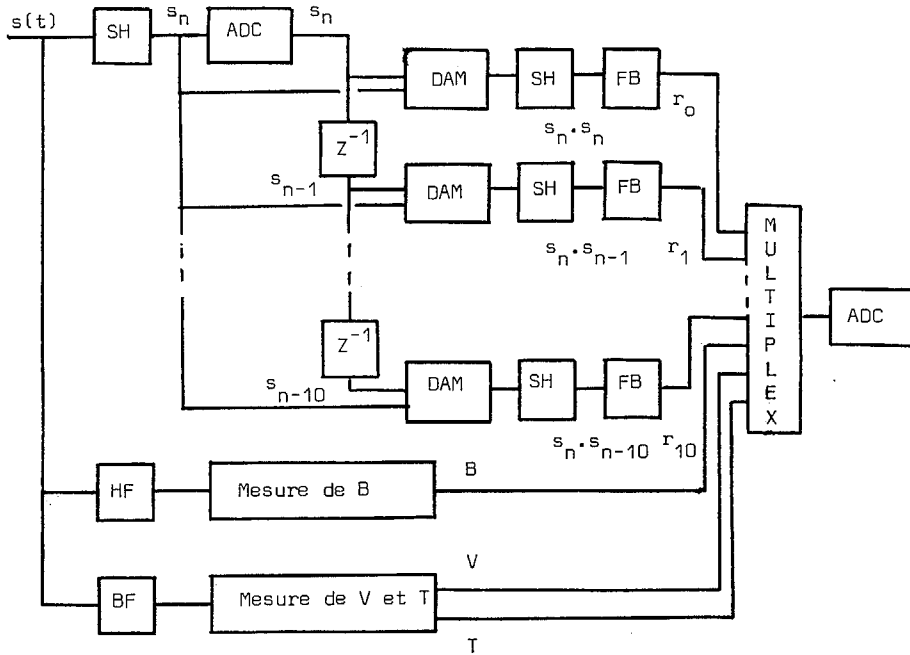


Figure 3 : Calcul hybride du  
 coefficient d'auto-  
 corrélation  $r_k$ .



- SH Echantillonneur bloqueur
- ADC Convertisseur analogique numérique
- DAM Multiplieur analogique numérique
- $Z^{-1}$  retard numérique
- FB Filtre de Bessel
- HF filtre passe haut > 2000 Hz
- BF filtre passe bas < 300 Hz

Figure 4 : Schéma du corrélateur hybride.



UN CORRELATEUR HYBRIDE POUR L'ANALYSE EN TEMPS REEL DU SIGNAL VOCAL  
AN HYBRID CORRELATOR FOR REAL TIME SPEECH ANALYSIS

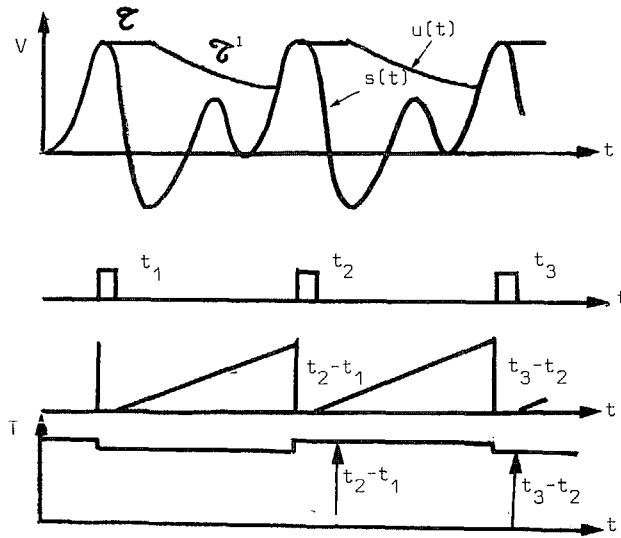


Figure 5 : Détection du Fondamental