



DEUXIÈME COLLOQUE
SUR LE TRAITEMENT DU SIGNAL
ET SES APPLICATIONS
NICE - 5 AU 10 MAI 1969

45/ 1

Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

J.M. PERSON (C.N.E.T.-LANNION)

- RESUME -

Le but de cette communication est d'attirer l'attention sur l'intérêt que présentent certaines techniques modernes dans le traitement du signal de parole. Ces techniques sont essentiellement basées sur les méthodes de convolution que la compression de temps (DELTIC) permet d'effectuer en temps réel. On donne le principe et quelques détails de réalisation d'un autocorrélateur destiné à la recherche de la fréquence fondamentale des sons vocalisés, ainsi qu'un aperçu sur d'autres développements possibles. On relève également certaines limitations, prévisibles compte tenu de l'origine des signaux traités.

- SUMMARY -

The purpose of this paper is to draw the attention on some interesting methods of signal processing, when used in speech research. These methods have their main foundations in convolution calculations that time compression techniques allow to do in real time. Principles of operation and some details of construction are described for a self-correlation computer intended to detect and measure the pitch of voiced sounds.



Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

Some brief ideas are also given about other purposes apparatus. Some predictable limits are noted, due to the origin of the processed signals.



DEUXIÈME COLLOQUE
SUR LE TRAITEMENT DU SIGNAL
ET SES APPLICATIONS
NICE - 5 AU 10 MAI 1969

45/5

Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

J.M. PERSON (C.N.E.T.-LANNION)

I - INTRODUCTION

Les techniques de traitement du signal utilisées désormais couramment en RADAR et SONAR* peuvent s'avérer bénéfiques dans le domaine des signaux de parole.

Toutefois, l'emploi ne s'en est encore pas beaucoup répandu. A cela on peut trouver deux raisons : le fait que la plupart des travaux étaient du domaine militaire classifié, d'une part, mais aussi la difficulté pour les spécialistes à abandonner des outils de recherche auxquels ils sont bien habitués, comme le sonographe, ou bien auxquels ils commencent à peine à l'être : filtres miniatures, générateurs divers, convertisseurs analogiques-numériques vocoders et enfin calculateurs de dimensions variées.

Or, il semble que les techniques précitées, surtout combinées avec les procédés de compression de temps (DELTIC), fournissent à bon compte des solutions à des problèmes anciens, et donnent en outre la possibilité d'effectuer des traitements en temps réel.

*Calculs de convolutions divers : corrélation, filtrage adapté, compression de temps.



Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

L'exposé qui va suivre n'a d'autre but que d'indiquer quelques unes des nombreuses possibilités offertes. De plus, l'emploi de composants modernes (registres à MOS par exemple), procure une grande souplesse d'analyse.

II-RAPPELS SUR LES SIGNAUX DE PAROLE

Outre les variations de niveau qui leur sont propres, ils sont essentiellement caractérisés par :

- 1/ - La structure fine du spectre à court terme (fenêtre de 15 à 20 ms) : raies harmoniques d'une fréquence fondamentale dans le cas des sons vocalisés (voyelles), bandes dans le cas des sons non vocalisés (plosives, fricatives : "s" "ch" "f"), mixte ("z", "j", "v" : fricatives vocalisées).
- 2/ - L'allure globale du spectre (ou enveloppe), c'est à-dire la répartition des extrema d'énergie (formants), qui dépend du locuteur, de son environnement (réverbération, bruits) et surtout du phonème émis.

Dans pratiquement toutes les analyses des signaux vocaux, la détection et la mesure précise de la fréquence fondamentale présentent un intérêt important, parfois primordial.

Aucune des multiples méthodes envisagées et exploitées actuellement ne donne véritablement satisfaction (Détection des pointes, statistiques des passages par zéro, "clampings", filtrages, etc...).

Des publications récentes (SCHROEDER & NOLL : communication au 5e CIA - Sept 65 ; SOBOLEV : journal d'acoustique soviétique : janv. 69) et quelques expériences pratiquées en laboratoire



Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

au C.N.E.T., semblent prouver que les méthodes basées sur l'inter ou l'autocorrélation de signaux sont dignes d'un grand intérêt. Mais les expériences mentionnées ne semblent pas avoir franchi le stade des simulations sur ordinateur, c'est-à-dire à une échelle où il est encore délicat de vouloir généraliser les résultats. (SOBOLEV donne toutefois un échantillonnage très important et intéressant).

III-PRINCIPE DE FONCTIONNEMENT D'UN CORRELATEUR EN TEMPS REEL ADAPTE AU PROBLEME

Soient deux mémoires à circulation M_1 et M_2 (figure n° 1), ayant le même nombre d'adresses : $n+1$, une adresse étant susceptible de contenir un échantillon de signal.

Soient d'autre part $S(t_0)$, $S(t_0+\theta)$, la suite des échantillons représentatifs d'un signal $S(t)$.

La mémoire M_2 contient en permanence les $(n+1)$ échantillons $S(t_0) \dots S(t_0+n\theta)$.

La mémoire M_1 contient $n+1$ échantillons $S(t_0+\tau)$ $\dots S(t_0+\tau+n\theta)$, renouvelés en permanence par le remplacement, tous les $n\theta$, du plus ancien par le plus récent

A un moment, supposons que les échantillons soient rangés dans l'ordre indiqué sur la figure n° 1, à partir de cet instant, et jusqu'à la n° impulsion d'horloge suivante, le résultat fourni par l'opérateur figurant la multiplication et la sommation sera :

$$\sum_{n=0}^1 S(t_0+i\theta) \cdot S(t_0+i\theta+\tau) = S(t_0+n\theta) \cdot S(t_0+\tau+n\theta) + \\ S(t_0+(n-1)\theta) \cdot S(t_0+\tau+(n-1)\theta) + \dots$$



45/8

Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

Soit l'approximation, à un facteur multiplicateur constant près, d'une valeur de la fonction d'autocorrélation:

$$g(\tau) = \frac{1}{n\theta} \int_{t_0}^{t_0+n\theta} S(t).S(t+\tau)dt$$

On obtient donc, en un tour de mémoire, un point de la fonction d'autocorrélation.

Si δ est la période d'horloge, on a $\theta = n\delta$. Au bout d'un cycle, τ est augmenté de θ ($\tau = k\theta$, avec k entier), jusqu'à ce que $\tau = n\theta$. On a alors obtenu, dans l'intervalle $0 < \tau < n\theta$, des échantillons représentatifs de la fonction $g(\tau)$. En fait, on a obtenu une fonction $g_{t_0}(\tau)$. Pour obtenir $g_{(t_0+n\theta)}(\tau)$, on renouvelle entièrement, le contenu de la mémoire M_1 , en y transférant celui de la mémoire M_2 , en un cycle de durée θ .

Les opérations reprennent, pour obtenir la nouvelle fonction $g(\tau)$ dans l'intervalle $t_0+n\theta$, $t_0+2n\theta$.

IV-DETAILS DE LA REALISATION

1/ - Dans le problème particulier étudié, on s'efforce de partir d'une fonction $S(t)$ binaire, conservant autant que possible le caractère de périodicité du signal d'où elle est extraite, ce qui apporte une grande simplicité, mais un codage à plus de deux niveaux n'introduirait aucune difficulté théorique.

La qualité du résultat (détection et mesure du fondamental dépend évidemment de la façon dont ce signal est obtenu, mais les développements nécessaires à ce genre de considération sortiraient du cadre de cet exposé.



Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

2/ - Dans le cas des sons vocalisés, les fonctions $g(\tau)$ ont l'allure indiquée sur la figure n°2, comme l'on pouvait s'y attendre, connaissant la forme des fonctions de corrélation de signaux binaires périodiques. La période fondamentale T et ses multiples sont matérialisées par des pointes très fines. La décision et la mesure s'opèrent d'après l'examen automatique de ces zones, lors de leur apparition : par comparaison à des seuils, qui dépendent eux-mêmes de τ . La figure n° 3 donne le schéma de principe plus complet du système.

3/ - La fonction $g(\tau)$ doit être évaluée pour des valeurs de $\tau > T_{\max}$.

D'autre part, l'erreur de quantification étant θ celle-ci doit être suffisamment faible pour demeurer acceptable. L'erreur relative tolérable dans les applications se situe aux alentours de $1/4$ ton, soit 3%. Or cette erreur sera maximum pour $T_{\min}$.

Pour une conversation normale (hommes et femmes), on a $2\text{ms} < T < 16\text{ms}$.

On peut prendre $n=250$; ce qui donne $\theta = 64\mu\text{S}$;

$\frac{\theta}{T_{\min.}} = 3,2\%$; $\delta = 0,256\mu\text{S}$; soit une fréquence d'horloge de 4 MHz.

Ces chiffres indiquent que les R.D. à MOS constituent le type de mémoire adéquat, avec ou sans mise en parallèle d'un certain nombre.

4/ - Le résultat est amélioré si, au lieu de la fonction $g(\tau)$, on ne calcule que la somme des k premiers termes ($k = \frac{T}{\theta}$) de l'expression qui fournit $g(\tau)$, soit :



45/10

Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

$$\sum_{l=1}^k S(t_0+n\theta) \cdot S(t_0+l+n\theta) + \dots + S(t_0+(n-k)\theta) \cdot S(t_0+n\theta).$$

Ceci permet en effet d'éliminer l'influence du bruit, et d'obtenir une décision avec un retard moindre.

V-AUTRES EXEMPLES D'APPLICATION

Les deux paragraphes précédents donnent les grandes lignes d'une solution intéressante au problème précis que constitue la recherche et la mesure du caractère vocalisé des sons. La simplicité du matériel mis en oeuvre conduit à envisager la possibilité de résoudre d'autres questions plus complexes :

- Par des intercorrélations entre signaux, il est vraisemblablement possible de distinguer, dans le cas où le spectre a une structure mixte (raies et bandes), les zones de celui-ci où il y a prépondérance des raies ou des bandes.
- Des appareils effectuant en temps réel une analyse spectrale du type hétérodyne ont déjà été réalisés et même commercialisés. Il n'en existe cependant aucun dont les caractéristiques soient véritablement bien adaptées au traitement de la parole. Par le développement des techniques précitées, on peut prévoir la réalisation de "vocoders" de type nouveau, rappelant les vocoders classiques connus, à canaux, à formants ou à corrélation, ces appareils ayant les applications habituelles que l'on en attend : analyse, synthèse de parole, transmissions à bande comprimée, reconnaissance automatique.



Techniques modernes de traitement des signaux en temps réel.

45/11

Applications au domaine de la parole.

- La possibilité d'effectuer en temps réel des produits de convolution fort complexes peut donner naissance à des méthodes originales de recherche pour la reconnaissance automatique : méthodes dérivées du filtrage adapté des signaux, bien connu en détection.

VI- LIMITATIONS PREVISIBLES DU DOMAINE D'APPLICATION DES TECHNIQUES NOUVELLES

Dans les années qui vont suivre, l'emploi judicieux des techniques mentionnées, basées essentiellement sur la compression de temps, peut provoquer un bond spectaculaire dans le domaine du traitement et de l'analyse des signaux de parole, comme ce fut le cas dans le domaine de la détection aérienne ou sous marine.

Il faut toutefois se garder d'un enthousiasme excessif, car le problème présente un degré supérieur de complexité, dont les remarques suivantes donneront un aperçu.

- 1/ - En détection, les signaux traités ont un caractère aléatoire résultant de la présence de bruit et surtout de l'interaction de signaux certains avec des "opérateurs" (milieux de propagation, objectifs) aléatoires, ceux auxquels on porte de l'intérêt.

Dans l'étude des signaux de parole, on s'intéresse également aux "opérateurs" qui déterminent la "forme acoustique". (Cavités bucco-nasales, et autres phénomènes articulatoires) mais on n'est pas maître de la source de puissance sonore. Cette source présente elle-même des caractères aléatoires ; de plus, il existe probablement une interaction dans une proportion



45/12

Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

encore mal définie, entre cette source et les "opérateurs", ce qui n'est absolument pas le cas en détection.

2/ - L'aspect "reconnaissance des formes" a jusqu'ici été passé sous silence ; or les techniques envisagées ne donnent absolument pas la clé pour la définition et la simulation par des machines de processus d'apprentissage, ou "auto-adaptatifs" comparables à ceux du cerveau.

C'est cependant là un aspect fondamental, sur lequel les techniques connues ont déjà buté.



Techniques modernes de traitement des signaux en temps réel.

45/13

Applications au domaine de la parole.

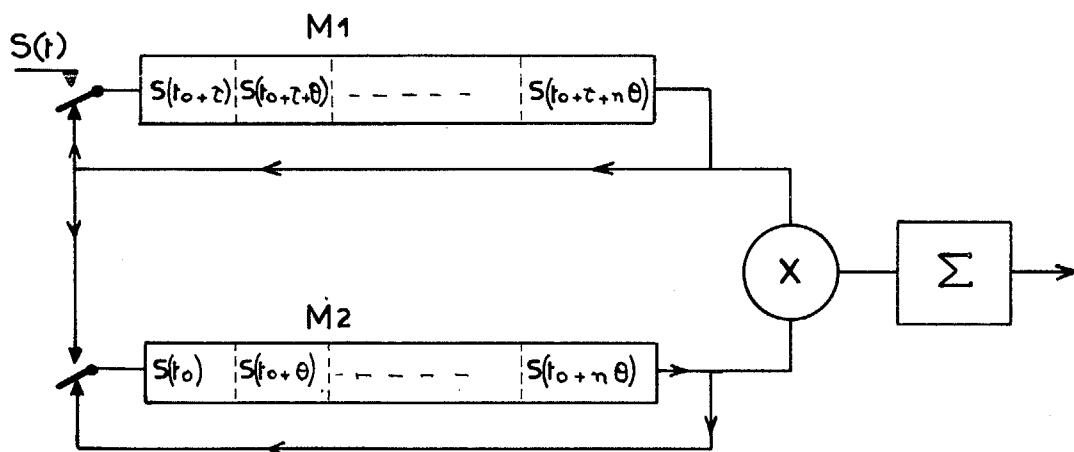


Figure n°1



45/14

Techniques modernes de traitement des signaux en temps réel.

Applications au domaine de la parole.

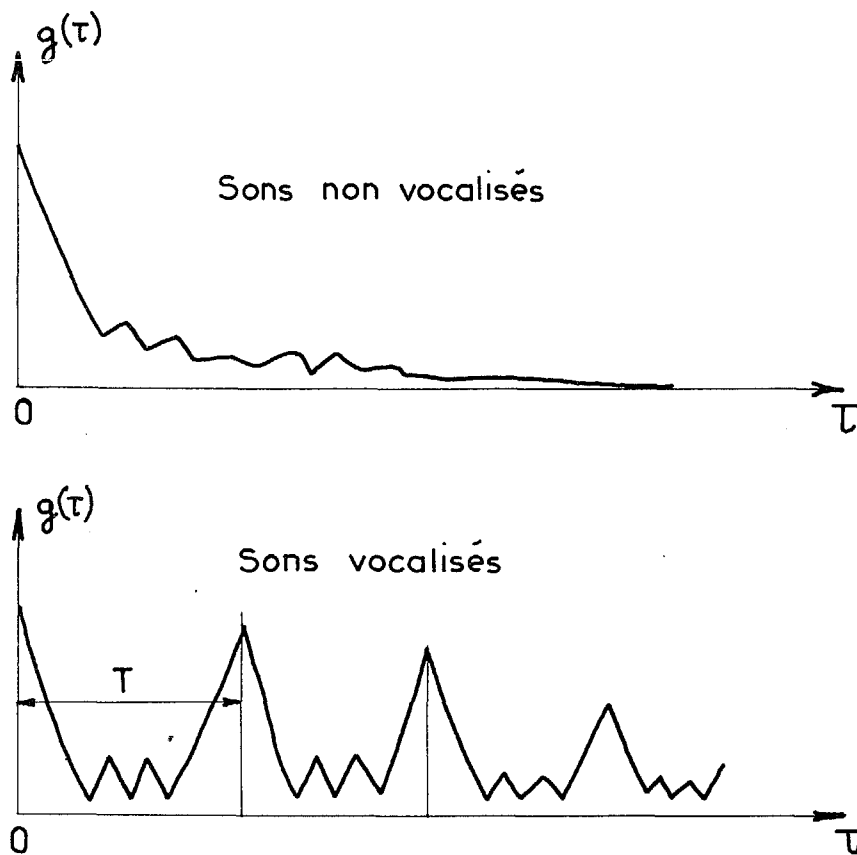
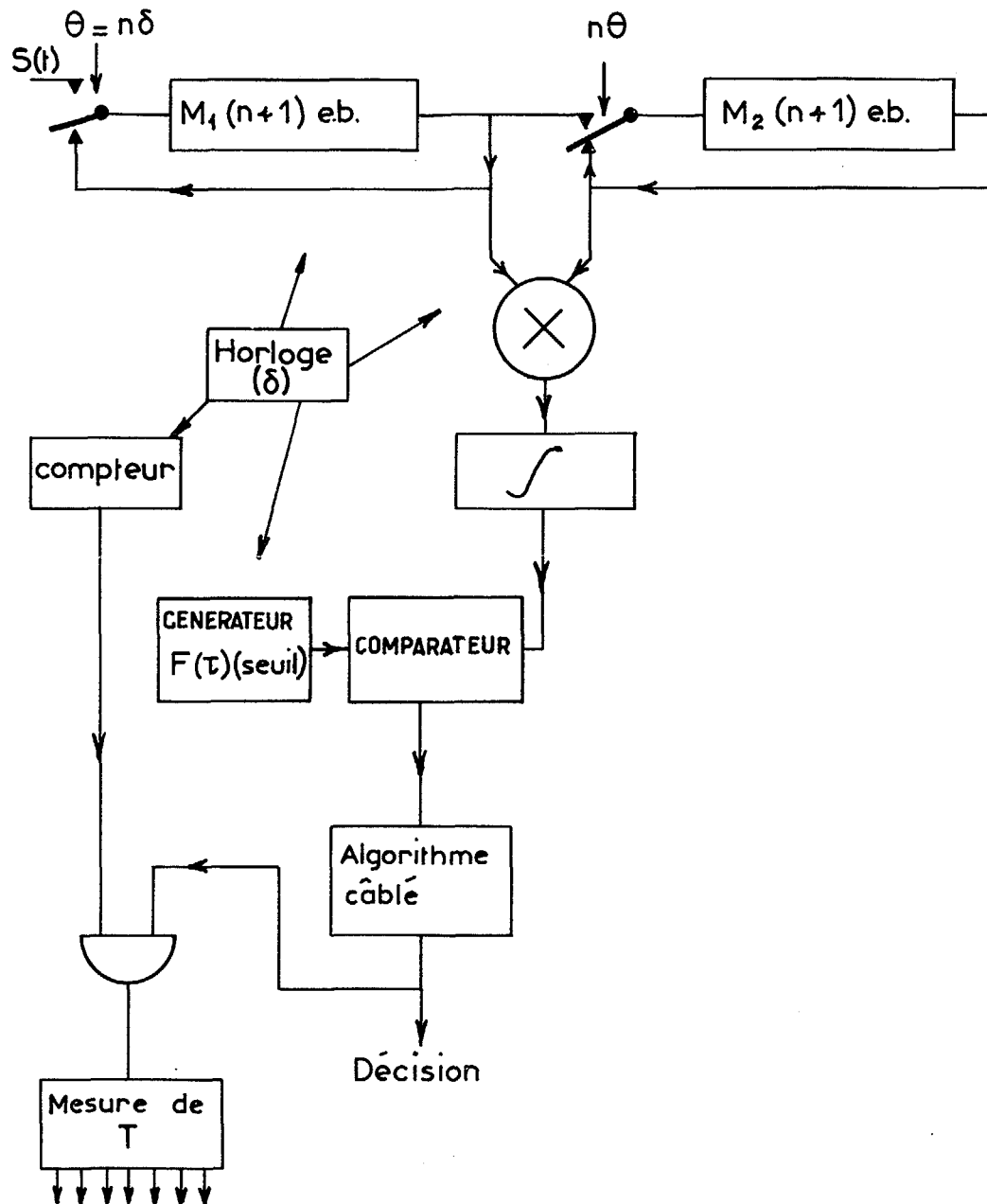


Figure 2



Techniques modernes de traitement des signaux en 45/15
temps réel.

Applications au domaine de la parole.



— Figure 3 —