

# DIXIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

875



NICE du 20 au 24 MAI 1985

---

## AMELIORATION DES PERFORMANCES DU CODAGE LPC A EXCITATION MULTI-IMPULSIONNELLE

Olivier PASSIEN      Jean-Paul LEFEVRE

CIT-ALCATEL, Laboratoire Traitement de Parole  
10bis, Rue L. Lormand, 78320 La Verrière, FRANCE

---

### RESUME

Depuis l'introduction par B.S. Atal et J.R. Remde du concept d'excitation multi-impulsionnelle pour le codage LPC, les études sur ce sujet se sont développées de façon considérable et suscitent de nombreuses variantes. Dans cette communication, nous proposons une technique originale introduisant un prédicteur de pitch à long terme dans la boucle de recherche des impulsions. Nous présentons cette approche comme une généralisation de la technique originale. La prise en compte d'hypothèses réalistes nous permet ensuite de simplifier les calculs de façon significative. Puis cette nouvelle technique est combinée avec une méthode assurant une estimation optimale de l'ensemble des amplitudes des impulsions. Finalement, des résultats expérimentaux obtenus dans des conditions typiques sont présentés. La qualité de la parole restituée et la complexité de mise en oeuvre des techniques proposées sont commentées.

### SUMMARY

Since the presentation of multi-pulse excitation concept for LPC coders, by Atal and Remde, many different analysis techniques have been proposed to derive the excitation waveform. In this paper, we introduce an original approach based on the insertion of a long term pitch predictor in the pulse search procedure. This approach can be viewed as a generalization of the original method. By assuming realistic hypothesis, we subsequently reduce computational loads by a significant amount. In addition, this new technique is combined with a procedure which compute again the jointly optimal set of amplitudes after a sequential estimation of the pulses. Finally, results of experimental evaluations for typical configurations, with respect to implementation complexity as well as speech quality, are given.



## I - INTRODUCTION

La conception de codeurs de parole restituant une bonne qualité pour des débits inférieurs à 16 Kbit/s présente un intérêt grandissant avec le développement d'applications nécessitant l'utilisation de canaux de transmission à faible largeur de bande, telles par exemple la messagerie vocale ou encore la radiotéléphonie mobile.

Il est maintenant bien connu qu'avec les techniques classiques de type "vocodeur" modélisant le mécanisme de production, la qualité de la parole ne peut guère être supérieure à la qualité synthétique, même avec un accroissement important du débit. On peut pour expliquer, du moins en partie, cette limitation, invoquer le principe même des vocodeurs, et en particulier la modélisation extrêmement simplifiée de la source d'excitation. D'autre part, nous savons également que les performances des codeurs par forme d'ondes (par exemple de type MIC différentiel) s'effondrent rapidement pour des débits inférieurs à 16 Kbit/s. En effet ces codeurs, s'appuient surtout sur des propriétés de corrélation et de stationnarité à court terme du signal de parole, propriétés qui deviennent inefficaces pour un codage de bonne qualité à débit moyen. Par conséquent, pour obtenir une parole de bonne qualité tout en conservant un débit raisonnable, la mise en oeuvre de techniques hybrides, utilisant de façon complémentaire les avantages des techniques par forme d'ondes et ceux des techniques de type "vocodeur", semble une solution particulièrement intéressante.

Généralement, dans les codeurs hybrides, l'information relative au filtre représentatif du conduit vocal est calculée puis transmise, d'une manière identique à celle utilisée dans les vocodeurs. Par contre, la mesure de la fréquence fondamentale est évitée, et la source d'excitation est dérivée directement du signal résiduel qui doit donc être transmis sous une forme plus ou moins compacte. Il convient de noter que la non-nécessité d'une détection précise de la fréquence fondamentale permet à ce type de codeur d'être relativement insensible au bruit ambiant. Parmi les très nombreuses associations de dispositifs respectant le concept de codage hybride, le codage LPC à excitation multi-impulsionnelle apparaît comme l'une des approches les plus prometteuses. Dans un tel codeur on cherche une représentation relativement fidèle de la source vocale (excitation) que l'on modélise par un petit nombre d'impulsions, typiquement une dizaine par fenêtre de dix millisecondes. Le processus d'analyse permet de choisir les positions et les amplitudes de ces impulsions de façon à minimiser un critère d'erreur perceptuel.

Cette technique, proposée à l'origine par Atal et Remde [1], s'est développée de façon considérable et a suscité de nombreuses variantes, qui pour la plupart cherchent soit à réduire le volume de calculs mis en jeu, soit à faciliter le codage ultérieur des impulsions. Dans cette communication, nous proposons une technique originale introduisant un prédicteur à long terme dans la boucle de recherche des impulsions. Ce prédicteur va influencer le signal représentatif des impulsions, lequel est réactualisé de façon récursive après la mise en place de chaque impulsion. Cette approche qui peut être vue comme une généralisation de la technique originale conduit à des performances sensiblement accrues tout en ne nécessitant pas un surcoût excessif de complexité compte tenu de l'introduction de certaines simplifications.

## II - STRUCTURE GENERALISEE INTRODUISANT UN PREDICTEUR A LONG TERME

## II-1 Mise en équations

Rappelons que nous cherchons à trouver une suite d'impulsions pouvant se substituer au signal résiduel obtenu par les techniques classiques de prédiction linéaire. Au niveau du récepteur, un tel train d'impulsions pourra être avantageusement utilisé comme excitation du filtre LPC de synthèse  $H(z)$ . Lors de l'analyse, les positions et les amplitudes des impulsions sont déterminées de façon à minimiser l'énergie d'une erreur pondérée entre le signal original et le signal synthétisé. Tel que précédemment suggéré [1], le filtre de pondération, qui a pour but de masquer le bruit de codage en le focalisant dans les zones formantiques, peut être défini par :

$$W(z) = H(cz) / H(z) \quad (1)$$

où  $c$  est un nombre fractionnaire compris entre 0 et 1

D'autre part, plusieurs auteurs [2,3], ont déjà remarqué que pour des signaux de parole voisée, la suite des impulsions laisse apparaître une forte corrélation avec la période fondamentale. Ce résultat n'est pas à proprement parler surprenant, compte-tenu des propriétés bien connues de stationnarité à court terme du signal de parole voisée, parole caractérisée par une structure formantique et un spectre de raies.

Cette corrélation significative entre impulsions d'une période fondamentale et de la suivante nous a conduit à envisager l'intégration d'un prédicteur à long terme dans la boucle de calcul de l'excitation [4]. L'effet d'un tel prédicteur sera le suivant : si la période fondamentale correspond à  $M$  échantillons, nous pourrons, à partir d'une impulsion d'amplitude  $A$  placée en  $k$ , directement déduire une série d'impulsions positionnées en  $k$ ,  $k+M$ ,  $k+2M$ , ...,  $k+IM$  et ayant pour amplitudes  $A$ ,  $bA$ ,  $b^2A$ , ...,  $b^IA$  ;  $b$  étant le gain du prédicteur à long terme.

La figure 1 ci-dessous représente l'ensemble du système. Il est évident que la branche inférieure n'est qu'une forme de représentation du signal original.

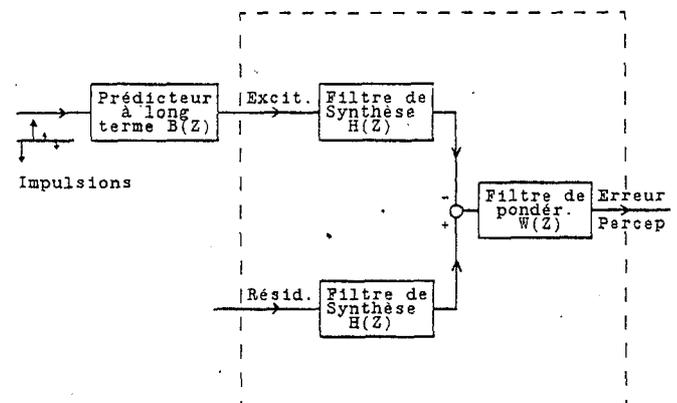


Figure 1. Modélisation de la structure généralisée

Tel qu'exposé précédemment, nous allons chercher à déterminer la position d'une impulsion, sur une fenêtre de  $N$  échantillons, en minimisant l'énergie de l'erreur perceptuelle.



AMELIORATION DES PERFORMANCES  
DU CODAGE LPC A EXCITATION MULTI-IMPULSIONNELLE

En suivant la figure 1, nous pouvons exprimer l'erreur perceptuelle,  $\{e^{[q]}(k)\}$ , après que  $q$  impulsions aient déjà été placées sur cette fenêtre en  $m(1) \dots m(q)$ .

$$e^{[q]}(k) = \sum_{n=-\infty}^k r(n) \cdot h'(k-n) - \sum_{i=1}^q \left( \sum_{p=-\infty}^n A(i) \cdot d(p, m(i)) \right) \cdot w(n-p) \cdot h'(k-n) \quad (2)$$

où  $\{r(n)\}$  est le résiduel,  $A(i)$  l'amplitude de l'impulsion placée en  $m(i)$ ,  $\{h'(n)\}$  la réponse impulsionnelle pondérée du filtre de synthèse (c'est à dire la réponse impulsionnelle de  $H(z)$  passée dans  $W(z)$ ),  $d$  le symbole de Kronecker et  $\{w(p)\}$  la réponse impulsionnelle du filtre de prédiction à long terme.

$$w(p) = \begin{cases} b^s & \text{si } p = sM \text{ pour } s = 0, \dots, \infty \\ 0 & \text{sinon} \end{cases} \quad (3)$$

Il convient de noter que dans (2), les indices  $i \leq 0$  correspondent aux impulsions des fenêtres précédentes prises en remontant le temps.

L'équation (2) peut s'exprimer plus simplement, en ne gardant que les termes non nuls, sous la forme :

$$e^{[q]}(k) = \sum_{n=-\infty}^k r(n) \cdot h'(k-n) - \sum_{i=1}^q A(i) \cdot \sum_{s=0}^{\lfloor (k-i)/M \rfloor} b^s \cdot h'(k-m(i)-sM) \quad (4)$$

où  $\lfloor (k-i)/M \rfloor$  représente la partie entière de  $[k-m(i)]/M$  et vaut zéro si  $M=0$ .

Soit  $m(q+1)$  la position optimale pour l'impulsion  $[q+1]$ , et  $A(q+1)$  son amplitude.

L'erreur perceptuelle avec cette impulsion supplémentaire s'écrit :

$$e^{[q+1]}(k) = e^{[q]}(k) - A(q+1) \cdot \sum_{s=0}^{\lfloor (k-m(q+1))/M \rfloor} b^s \cdot h'(k-m(q+1)-sM) \quad (5)$$

Après avoir introduit les notations suivantes :

$$t^{[q]}(k) = \sum_{n=1}^N e^{[q]}(n) \cdot h'(n-k) \quad (6)$$

$$tt^{[q]}(k) = \sum_{i=0}^{\lfloor k \rfloor} b^i \cdot t^{[q]}(k+iM) \quad (7)$$

$$\Phi(k, k', M) = \sum_{i=0}^{\lfloor k \rfloor} \sum_{j=0}^{\lfloor k' \rfloor} b^{i+j} \cdot \left[ \sum_{n=1}^N h'(n-k-iM) \cdot h'(n-k'-jM) \right] \quad (8)$$

où  $f(k)$  représente la partie entière de  $(N-k)/M$  et vaut zéro si  $M=0$ , nous pouvons exprimer l'énergie par

$$E^{[q+1]} = E^{[q]} - 2 \cdot A(q+1) \cdot tt^{[q]}(m(q+1)) + A^2(q+1) \cdot \Phi(m(q+1), m(q+1), M) \quad (9)$$

Le minimum de  $E^{[q+1]}$  est obtenu lorsque la dérivée par rapport à  $A(q+1)$  de l'expression (9) ci-dessus s'annule, c'est à dire pour :

$$A(q+1) = tt^{[q]}(m(q+1)) / \Phi(m(q+1), m(q+1), M) \quad (10)$$

En reportant cette valeur dans (9) on déduit immédiatement la position optimale  $m(q+1)$  de l'impulsion  $[q+1]$ . Cette position correspond à l'indice  $k$  maximisant le signal suivant :

$$ZZ^{[q]}(k) = [tt^{[q]}(k)]^2 / \Phi(k, k, M) \quad (11)$$

Dans le cas où une impulsion aurait préalablement été positionnée en  $m(q+1)$ , l'amplitude calculée  $A(q+1)$  est ajoutée à l'ancienne valeur.

Pour déterminer la position optimale de l'impulsion suivante, il faudra estimer  $\{ZZ^{[q+1]}(k)\}$ , qui peut être déduit du pas précédent. En effet on peut montrer que  $\{tt^{[q+1]}(k)\}$  peut être déduit de  $\{tt^{[q]}(k)\}$  par la relation récursive suivante :

$$tt^{[q+1]}(k) = tt^{[q]}(k) - A(q+1) \cdot \Phi(m(q+1), k, M) \quad (12)$$

Un tel processus récursif est initialisé par la connaissance de  $\{tt^{[0]}(k)\}$  qui va être obtenu à partir de (7). Pour cela, il est nécessaire d'évaluer  $\{t^{[0]}(k)\}$ .

Le développement de (6) conduit à :

$$t^{[0]}(k) = \sum_{n=1}^N \left[ \sum_{l=1}^n R(l) \cdot h'(n-l) \right] \cdot h'(n-k) \quad (13)$$

en posant

$$R(l) = r(l) - \sum_{i=1}^l A(i) \cdot \left[ \sum_{s=0}^{\infty} d(l-sM, m(i)) \cdot b^s \right] \quad (14)$$

où  $0(1)$  est le minimum de zéro et de 1.

La séquence d'opérations est arrêtée après un certain nombre  $Q$  d'itérations, fixé à l'avance.

Le calcul de la fonction d'excitation est finalement complété par un ajustement final de l'amplitude de toutes les impulsions. Cette réactualisation finale consiste à remplacer toutes les amplitudes  $A(i)$  calculées par :

$$A(i) + tt^{[q]}(m(i)) / \Phi(m(i), m(i), M) \text{ pour } i = 1, \dots, Q \quad (15)$$

Comme indiqué sur la figure 1, lors de la synthèse, le train des  $Q$  impulsions déterminé lors de l'analyse est passé à travers le prédicteur à long terme, afin de générer l'excitation du filtre de synthèse.

Il convient de remarquer qu'à partir des équations ci-dessus relatives à la structure généralisée introduisant un prédicteur à long terme, il est aisé de retrouver les équations de la technique originale [1] en forçant l'indice  $s$  à la valeur nulle dans les équations (4) à (14). Rappelons simplement que cette technique consiste à ne considérer que la portion de la figure 1 entourée de pointillés.

## II-2 Simplifications

A partir de la technique d'analyse rigoureuse développée dans le paragraphe précédent, nous allons proposer un algorithme modifié ne nécessitant plus la connaissance explicite de l'erreur perceptuelle qui apparaît sous forme développée (somme sur  $l$ ) dans (13). Au contraire, nous n'aurons plus désormais besoin que de la connaissance du signal résiduel et de la réponse impulsionnelle pondérée du filtre de synthèse.

Les simplifications introduites sont tout d'abord basées sur certaines propriétés de la réponse impulsionnelle pondérée du filtre de synthèse. Cette réponse impulsionnelle pondérée  $h'(n)$  correspond, comme nous l'avons déjà signalé, à la mise en cascade des fonctions de transfert  $H(z)$  et  $W(z)$ . En raison de la forme très spécifique de  $W(z)$ ,  $h'(n)$  n'est rien d'autre que la réponse impulsionnelle du filtre de synthèse  $h(n)$  pondérée par une fenêtre exponentielle :

$$h'(n) = h(n) \cdot c^n \quad (16)$$

La causalité du prédicteur linéaire entraîne immédiatement :

$$h'(n) = 0 \text{ pour } n < 0 \quad (17)$$

Toutefois, le point le plus important concerne l'affaiblissement très rapide de l'enveloppe de  $h'(n)$  puisque  $c$  est une constante inférieure à 1. (En pratique,  $c$  est choisi entre 0,7 et 0,9). Par conséquent nous pouvons raisonnablement écrire :

$$h'(n) = 0 \text{ pour } n < K \text{ avec } K \ll N \quad (18)$$

En utilisant cette hypothèse nous pouvons considérablement simplifier les expressions de  $\Phi(k, k', M)$  et de  $\{t^{[q]}(k)\}$  (équations 8 et 13). En effet, en



utilisant (17) et (18) nous pouvons écrire après quelques lignes intermédiaires :

$$\sum_{n=k}^N h'(n-k-M) \cdot h'(n-k'-jM) \approx \sum_{t=0}^{K-u} h'(t) \cdot h'(t+u) \quad (19)$$

$$\text{avec } u = |k+iM - (k'+jM)| \text{ et } u \leq K \quad (20)$$

Si nous définissons :

$$\Psi(k) = \sum_{t=k}^K h'(t-k) \cdot h'(t) \quad (21)$$

Alors l'équation (8) se réduit à la forme :

$$\hat{\Phi}(k, k', M) = \sum_{i=0}^{\hat{K}(k)} \sum_{j=0}^{\hat{K}(k')} b^{i+j} \cdot \Psi(u) \quad (22)$$

$$\text{avec } u = |k+iM - (k'+jM)| \leq K \quad (23)$$

D'une manière analogue, en faisant successivement les changements de variables :

$u = n-k$  puis  $l = k+u-v$ , et finalement en remplaçant la borne  $l = \infty$  par  $l = K$  on obtient :

$$t^{(0)}(k) = \sum_{u=0}^K \left[ \sum_{i=0}^K R(k+u-1) \cdot h'(1) \right] \cdot h'(u) \quad (24)$$

En explicitant tous les termes de cette double sommation et en les recombinaut deux à deux  $\{t^{(0)}(k)\}$  s'exprime finalement par :

$$t^{(0)}(k) = R(k) \cdot \Psi(0) + \sum_{i=1}^K [R(k+i) + R(k-1)] \cdot \Psi(i) \quad (25)$$

Cette expression, tout comme (22), peut être avantageusement normalisée en définissant une variable  $C(i)$ , que nous appellerons coefficient d'autocorrélation normalisée, comme étant le rapport  $\Psi(i) / \Psi(0)$ . (Par définition  $C(0)=1$ ). En particulier (25) peut alors être réécrit sous la forme :

$$t^{(0)}(k) = R(k) + \sum_{i=1}^K [R(k+i) + R(k-1)] \cdot C(i) \quad (26)$$

Nous envisageons maintenant l'hypothèse supplémentaire  $K < M$ , hypothèse que nous avons pu pratiquement valider. Cette hypothèse est particulièrement intéressante lorsque l'on détermine la position optimale d'une impulsion. En effet, nous avons montré que cette position correspond à l'indice  $k$  maximisant le signal  $ZZ^{(0)}(k)$  défini par la relation (11). Considérons le dénominateur de cette relation après normalisation. En utilisant (22) et (23) nous pouvons le réécrire :

$$\hat{\Phi}(k, k', M) = \sum_{i=0}^{\hat{K}(k)} \sum_{j=0}^{\hat{K}(k')} b^{i+j} \cdot C(|iM-jM|) \quad (27)$$

Avec l'hypothèse  $K < M$ , pour respecter la condition exprimée en (23), c'est à dire  $|iM-jM| \leq K$ , il nous faut obligatoirement avoir  $i=j$ . D'où :

$$\hat{\Phi}(k, k, M) = \sum_{i=0}^{\hat{K}(k)} b^{2i} \cdot C(0) = \sum_{i=0}^{\hat{K}(k)} b^{2i} \quad (28)$$

Cette nouvelle relation facilite évidemment les calculs mis en oeuvre dans (11), ainsi que la détermination de l'amplitude de l'impulsion après sa localisation, estimée par l'intermédiaire de (10).

### II-3 Optimisation de l'ensemble des amplitudes

Jusqu'à maintenant les positions des impulsions et leurs amplitudes ont été déterminées de façon complètement séquentielle, sans aucune ré-estimation de l'amplitude des impulsions déjà connues. Or, une telle approche séquentielle peut conduire à une certaine dégradation. En effet, et plus particulièrement si les impulsions sont irrégulièrement espacées, les calculs tendent à devenir imprécis lorsque l'on optimise les amplitudes une à une ; chaque impulsion étant déterminée par une itération. Le plus souvent, on ne sait compenser cette imprécision que par la mise en place d'impulsions supplémentaires. Il convient de noter que ces dégradations, dues essentiellement aux corrélations mal prises en compte entre impulsions voisines,

apparaissent fréquemment avec notre structure où est introduit un prédicteur à long terme, puisque l'on manipule en fait des trains d'impulsions. Or de tels problèmes peuvent être atténués dans une proportion importante si l'on conserve optimales les amplitudes des impulsions précédemment positionnées durant la recherche d'une impulsion supplémentaire.

A partir de l'expression de l'énergie de l'erreur perceptuelle déduite de (2), si l'on connaît les positions de  $q$  impulsions, il est possible d'optimiser globalement les amplitudes de façon à minimiser cette énergie. En annulant les dérivées partielles de l'énergie par rapport aux amplitudes  $A(i)$ , on obtient la solution en résolvant le système linéaire suivant :

$$G * A = T^{(0)} \quad (29)$$

$$\text{avec } A^T = [A(1), \dots, A(q)] \quad (30)$$

$$T^{(0)T} = [t^{(0)}(m(1)), \dots, t^{(0)}(m(q))] \quad (31)$$

et le terme générique de la matrice  $G$  défini par :

$$G(i, j) = \hat{\Phi}(m(i), m(j), M) \quad (32)$$

les  $m(i)$  étant réordonnés en une suite strictement croissante.

Il est intéressant de noter ici que le terme additionnel introduit dans la réactualisation finale simple telle que nous l'avons présentée en (15) correspond à une matrice  $G$  dont les seuls termes non nuls sont ceux de la diagonale et à un vecteur  $T^{(0)}$ .

Pratiquement, la réoptimisation globale des amplitudes peut être envisagée en suivant plusieurs stratégies. Tout d'abord, les amplitudes peuvent être optimisées globalement à l'aide de (29) une seule fois, après que les  $Q$  impulsions aient été toutes déterminées par l'approche séquentielle. Au contraire, il est également possible de réoptimiser à chaque itération l'ensemble des amplitudes des impulsions alors connues. Une solution intermédiaire consiste à réoptimiser globalement les amplitudes dès que quelques impulsions, estimées par l'approche séquentielle, sont connues. Ces différentes alternatives conduisent, comme nous le montrerons dans nos résultats à une complexité de mise en oeuvre et à des performances relativement différentes.

### III - EVALUATION DE LA QUALITE ET DE LA COMPLEXITE

#### III-1 Conditions expérimentales

Afin d'apprécier, pour les différentes possibilités décrites dans le chapitre précédent, la qualité de parole restituée ainsi que la complexité de mise en oeuvre des algorithmes, une évaluation systématique a été réalisée. Dans tous les essais, les conditions de départ ont été conservées identiques, en vue de permettre des comparaisons significatives. Ces conditions nous ont semblé être un choix raisonnable compte-tenu des débits de codage envisagés.

En particulier, nous avons toujours utilisé un filtre classique de prédiction linéaire tout-pôles, comprenant 16 pôles. Les paramètres de prédiction sont calculés sur des fenêtres de 20 millisecondes, sans recouvrement, à l'aide de la méthode d'autocorrélation. Préalablement à l'analyse, le signal est pré-accentué par l'intermédiaire d'un coefficient égal à 0,8.

Le processus d'analyse de l'excitation multi-impulsionnelle fonctionne également sur des intervalles successifs de 20 millisecondes, le facteur  $c$  ayant pour valeur 0,9. De plus, pour tous les résultats présentés,  $K$  vaut 16 et la fenêtre de recherche des impulsions est

étendue légèrement de part et d'autre de l'intervalle considéré, afin de minimiser l'influence d'éventuels effets de bords.

### III-2 Evaluation de la qualité de la parole

Pour les différentes approches envisagées, la qualité de parole restituée a fait l'objet d'évaluations expérimentales réalisées à partir de simulations informatiques des algorithmes. Les résultats présentés dans cette communication proviennent d'essais menés à partir de deux banques de parole naturelle. La première contient deux phrases de test prononcées successivement par quatre locuteurs masculins et trois féminins, pour une durée totale de 42 secondes. La seconde a été constituée à partir de la suite des voyelles françaises /a,e,i,o,u/ prononcées par un locuteur et une locutrice ayant volontairement modulé l'évolution de leur fréquence fondamentale (durée totale : 16 secondes). Les échantillons de parole ont été filtrés pour limiter leur contenu à la bande téléphonique, puis échantillonnés à 8 kHz. Une évaluation subjective de la qualité a été réalisée de manière uniquement informelle grâce à des tests d'écoute. Pour les mesures objectives, nous avons eu recours au calcul des rapports signal sur bruit (SNR) et signal sur bruit segmental (SNRseg), ce dernier étant estimé sur des intervalles de 16 millisecondes. Dans la table 1, ci-après, nous présentons les résultats obtenus pour ces mesures objectives. Ces valeurs correspondent sauf indication contraire, à un nombre d'impulsions  $Q$ , par période de 20 millisecondes, égal à 20.

### III.3 Evaluation de la complexité

Pour mieux appréhender la complexité de mise en oeuvre des algorithmes, nous avons évalué le nombre d'opérations arithmétiques nécessaires pour chaque approche envisagée. Les résultats obtenus, également présentés dans la table 1, sont exprimés en nombre de divisions et en nombre "d'opérations élémentaires", par fenêtre d'analyse. Nous définissons une "opération élémentaire" comme la somme d'une addition et d'une multiplication. Une telle représentation trouve sa justification dans la structure particulière de la majorité des processeurs spécialisés de traitement du signal. Il est intéressant de préciser que pour obtenir les résultats présentés, nous avons supposé toutes les opérations de filtrage réalisées à l'aide d'une structure en treillis. Or, une telle technique, qui présente par ailleurs d'incontestables avantages, s'avère deux fois plus coûteuse en calcul qu'un filtrage direct à partir des prédicteurs. Dans la table 1, nous n'avons pas pris en compte les opérations nécessaires pour la détection des maximum, ces opérations étant les mêmes dans tous les cas, ni la charge de calculs requise pour l'estimation de la période fondamentale  $M$  et du gain  $b$  du prédicteur à long terme.

### III.4 Discussion et Conclusion

A partir de l'évaluation subjective et des résultats objectifs regroupés dans la table 1, nous pouvons préciser les points suivants :

- L'introduction d'un prédicteur à long terme, lié à la fréquence fondamentale, dans la boucle de recherche des impulsions améliore sensiblement les rapports signal sur bruit (ligne 3 comparée à ligne 1). Subjectivement, la qualité perçue paraît bien meilleure. Par rapport à la méthode originale, notre approche se traduit surtout par une amélioration de l'agrément des voix féminines, et plus généralement des voix à fréquence fondamentale élevée, ce qui peut être expliqué par une évolution plus lisse, puisque filtrée, des variations de la fréquence fondamentale.

- L'addition de la procédure d'optimisation de l'ensemble des amplitudes des impulsions améliore encore les rapports signal sur bruit (lignes 4 et 5). Il est intéressant de noter l'efficacité de cette procédure combinée avec notre structure généralisée pour seulement 20 impulsions par fenêtre, due à la manipulation de trains d'impulsions. Au contraire, pour la technique sans prédicteur à long terme, nous avons déjà relevé [4] la nécessité d'un grand nombre d'impulsions pour justifier la réoptimisation globale des amplitudes.

- Les tests d'écoute confirment les résultats ci-dessus. Toutefois, d'un point de vue perceptuel, nous avons noté que la majeure partie de l'amélioration de la qualité est déjà obtenue avec seulement une optimisation globale des amplitudes après la dernière itération ; l'avantage de cette solution, étant évidemment sa plus faible complexité.

- Les résultats des lignes 6 à 11 de la table 1, obtenus à partir d'essais avec des signaux essentiellement voisés, sont mieux corrélés avec les évaluations subjectives. En effet, l'oreille est surtout impressionnée par la qualité des sons voisés, tandis qu'au contraire les rapports signal sur bruit sont essentiellement dégradés, dans un tel système de codage, par la restitution moins bonne des zones non-voisées et de faible énergie. Ainsi, objectivement, l'introduction du prédicteur à long terme conduit à des performances proches de celles de la technique originale (sans prédicteur), lorsque cette dernière met en oeuvre un nombre double d'impulsions par fenêtre d'analyse (lignes 7 et 8). L'évaluation subjective a parfaitement confirmé ce résultat.

- Pour les lignes 9 à 11, seules les portions voisées du signal ont été considérées. Ainsi nous sommes en mesure de quantifier objectivement l'amélioration subjective apportée par le prédicteur à long terme. Le gain obtenu, proche de 8 dB, est considérable. Rappelons en effet, qu'une différence de 8 dB sur des rapports signal sur bruit correspond, environ à 1,5 bit de quantification MIC.

- Finalement, il est intéressant de noter que les performances accrues de la structure généralisée, présentée dans cette communication, peuvent être obtenues moyennant un faible nombre d'opérations supplémentaires (si nous supposons connue la période fondamentale).

### REFERENCES

- [1] B.S. Atal, J.R. Remde, "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, Paris, France, Mai 1982, pp. 614-617.
- [2] P. Kroon, Ed.F. Depreterre, "Experimental Evaluation of Different Approaches to the Multi-Pulse Coder", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, San-Diego, Calif. Mars 1984, pp. 10.4.1-10.4.4.
- [3] S. Singhal, B.S. Atal, "Improving Performance of Multi-Pulse LPC Coders at Low Bit Rates", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, San-Diego, Calif., Mars 1984, pp. 1.3.1-1.3.4.
- [4] J.P. Lefèvre, O. Passien, "Efficient Algorithms for Obtaining Multipulse Excitation for LPC Coders", Proc. Int. Conf. on Acoustics, Speech and Signal Processing, Tampa, Floride, Mars 1985, Article 25.6.



AMELIORATION DES PERFORMANCES  
DU CODAGE LPC A EXCITATION MULTI-IMPULSIONNELLE

Algorithme Testé	Résultats Objectifs (en dB)			Nombre d'Opérations par Fenêtre d'Analyse	
	Corpus Utilisé	SNR	SNRseg	Opération Élémentaire	Division
1/ Technique originale (sans prédicteur à long terme)	1	15,7	14,6	12 000	16
2/ Technique originale (40 impulsions par fenêtre)	1	19,3	17,8	12 700	16
3/ Structure généralisée (avec prédicteur à long terme)	1	16,9	15,1	16 000	20
4/ Structure généralisée + Optimisation globale des amplitudes après la dernière itération	1	18,0	16,1	21 900	40
5/ Structure généralisée + Optimisation globale des amplitudes après chaque itération	1	18,5	16,7	56 000	230
6/ Idem 1	2	13,2	13,1		
7/ Idem 2	2	18,1	17,5		
8/ Idem 5	2	18,9	18,7		
9/ Idem 6	2	14,8	14,7		
10/ Idem 7	2	19,5	19,6		
11/ Idem 8	2	22,2	23,3		

Table 1. Résultats