DIXIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 20 au 24 MAI 1985

CODAGE MULTI-IMPULSIONNEL POUR LA RESTITUTION DE PAROLE PAR MODELES EVOLUTIFS

M.C. OMNES-CHEVALIER, Y. GRENIER, G. CHOLLET

ECOLE NATIONALE SUPERIEURE DES TELECOMMUNICATIONS
46, rue Barrault 75634 PARIS CEDEX 13

RESUME

Les signaux de parole sont caractérisés par leurs non-stationnarités : les modèles évolutifs dont les coefficients sont approchés par des combinaisons linéaires finies sur une base de fonctions du temps, (les fonctions étant connues a priori) permettent de prendre en compte cette propriété. De tels modèles se révèlent utiles en synthèse de parole mais ils présentent un manque de stabilité ce qui est gênant pour cette application.

Dans ce papier, nous montrons comment assurer la stabilité du modèle évolutif lors de son estimation : ceci est réalisé parl'intermédiaire des fonctions d'aire logarithmiques. Nous décrivons ensuite le principe de la technique multi-impulsionnelle permettant de déterminer l'excitation du modèle évolutif et qui est semblable à celle proposée par Atal dans le cas de la modélisation stationnaire. Dans une dernière partie sont commentés les résultats obtenus par l'application de ces deux méthodes (modélisation évolutive et codage multi-impulsionnel) à la restitution du signal de parole.

SUMMARY

Speech signals are intrinsically non-stationary. Time-dependent models, with parameters approximated by a linear decomposition on a basis of orthogonal functions of time (known a priori), are well adapted to exploit this property - Depending on the parameters used, these models may become locally unstable and therefore need computationally inefficient control for speech synthesis applications.

The first part of this paper focusses on a technique using "Log Area Ratio" parameters to ensure stability of the model.

A second part describes a multipulse coding technique for the residual which is an extension of ATAL's technique to non-stationary models.

The last part is concerned with the application of multipulse excited non-stationary models to speech restitution.

CODAGE MULTI-IMPULSIONNEL POUR LA RESTITUTION DE PAROLE PAR MODELES EVOLUTIFS

Introduction

La qualité de la synthèse de parole par les méthodes paramétriques de traitement du signal est directement liée au modèle estimé et à la source d'excitation de celui-ci.

Or, une des principales caractéristiques du signal de parole est sa non-stationnarité. Les techniques classiques d'analyse paramétrique telles que la LPC (Linear Prediction Coding) prennent en compte cette propriété en considérant le signal comme quasi-stationnaire, c'est-à-dire stationnaire sous une fenêtre de courte durée (20 millisecondes) et estiment un modèle indépendant du temps à l'intérieur de cette fenêtre ; la fenêtre est ensuite décalée (d'un pas de 10 à 20 millisecondes) et un nouveau jeu de coefficients est déterminé.
Mais une telle méthode présente des inconvénients tels que la redondance des modèles si l'analyse s'effectue dans une zône stable du signal ; inversement, l'hypothèse de quasi-stationnarité n'est pas vérifiée si la fenêtre contient des évènements transitoires (tels que des sons plosifs, des consonnes vocaliques, des voyelles nasales ...) : l'estimation des modèles est alors biaisée, la fenêtre d'analyse étant trop longue.

Une alternative à cette méthode consiste à introduire les variations temporelles du signal dans un modèle à coefficients dépendant du temps. Un tel modèle, appelé "modèle évolutif", est basé sur l'hypothèse suivante : chacun des coefficients peut être approché par une combinaison linéaire d'une base de fonctions, choisie à priori et pondérée par des poids inconnus mais invariants. Le but de l'analyse est de déterminer ces coefficients de pondération.

Différentes techniques permettant de caractériser la source d'excitation d'un modèle stationnaire ont été utilisées ces dernières années : la plupart d'entre elles effectuent un codage du résidu ([1], [2]). Une des plus récentes est la méthode impulsionnelle proposée par Atal [3] : elle permet d'obtenir une restitution de parole d'une excellente qualité avec un coût de calcul peu important.

Le propos de ce papier est de présenter une procédure de codage multi-impulsionnel (inspirée de la technique précédente) pour la restitution de parole non pas par modèles stationnaires mais par modèles évolutifs.
Après un bref rappel du principe de la technique d'Atal, nous décrivons quelques aspects de l'analyse évolutive, aspects fondamentaux pour la restitution. Sera ensuite proposée la méthode de codage réalisée. Des exemples issus des résultats expérimentaux ainsi obtenus, permettront de juger la qualité des signaux de parole synthétisés.

1. CODAGE MULTI-IMPULSIONNEL

La technique multi-impulsionnelle a pour but de modéliser par une séquence d'impulsions le résidu obtenu à l'issue d'une méthode classique d'analyse telle que la LPC-10. En synthèse, l'excitation du modèle s'effectue en remplaçant le résidu par ce train d'impulsions dont il convient de déterminer les positions et les amplitudes. Il est important de noter qu'aucune hypothèse sur la nature du signal de parole (voisé ou non-voisé, valeur du fondamental ...) n'est posée. Ceci permet de s'affranchir des problèmes qui apparaissent dans les méthodes où l'excitation est calculée à partir d'une détection de voisement ou non-voisement : la synthèse obtenue par de tels procédés, même à débit élevé, manque de naturel (l'excitation est artificielle). Il existe, en effet, de nombreuses zônes pour lesquelles le voisement n'apparaît pas clairement. De plus, même lorsque le signal est nettement périodique, l'introduction d'un seul point d'excitation durant toute une période fondamentale est une hypothèse trop simplificatrice. Bien que l'excitation principale des sons voisés, soit située à la fermeture de la glotte, apparaissent également des excitations secondaires, en particulier durant l'ouverture. La principale difficulté réside alors dans l'estimation de ces impulsions : positions? amplitudes ?
La détermination séquentielle des impulsions qui a été proposée par Atal construit la source d'excitation du modèle et permet de résoudre un tel problème : elle est l'objet du paragraphe suivant.

1.1. Modèle d'excitation multi-impulsionnelle

La procédure de recherche des impulsions peut être représentée par le diagramme de la figure 1. L'estimation de la position et de l'amplitude de chaque impulsion s'effectue par une procédure d'analyse-synthèse : l'erreur entre le signal original y_t, observé sur l'intervalle [0,T], et le signal synthétique $\hat{y}_t$, réponse du modèle autorégressif de fonction de transfert H(z) à la séquence d'impulsions v_t est minimisée après pondération par un filtre dit "perceptuel".

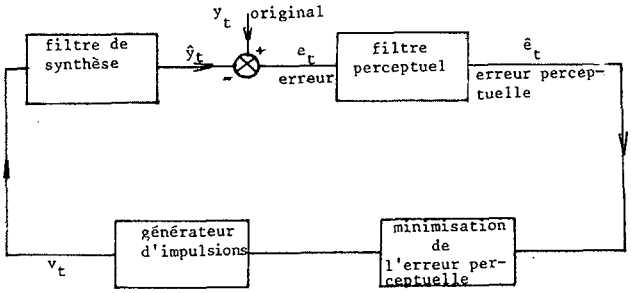


Figure 1 - Schéma de la procédure d'analyse-synthèse pour déterminer les positions et les amplitudes des impulsions de l'excitation multi-impulsionnelle.

L'intérêt du filtre perceptuel est d'obtenir un masquage spectral de l'erreur : ceci permet de tolérer une erreur plus grande dans les zônes formantiques, zônes où l'énergie du signal est importante, que dans celles interformantiques. C'est pourquoi Atal propose un filtre de pondération atténuant l'énergie au voisinage des formants et caractérisé par sa fonction de transfert W(z) (1) :

$$(1) \quad W(z) = \frac{H(\gamma z)}{H(z)}$$

Le paramètre γ est appelé "facteur perceptuel", ($0<\gamma<1$). Il permet de contrôler la pondération de l'erreur dans les zônes formantiques, le filtre W(z) variant de $W(z) = 1$ pour $\gamma = 1$ à $W(z) = \frac{1}{H(z)}$ pour $\gamma = 0$, γ doit être choisi en fonction de la pondération que l'on souhaite introduire dans le spectre de l'erreur. En pratique, la valeur optimale de ce coefficient est déterminée grâce à une série de tests auditifs (comparaison de la qualité de la restitution obtenue en faisant varier γ) ; les expériences montrent que le choix de γ peut être approximatif : pour une fréquence d'échantillonnage de 8kHZ, la valeur typique vaut environ $\gamma = 0.8$.

1.2. Placement séquentiel des impulsions

L'intérêt de la méthode d'Atal est de rendre linéaire un problème qui, a priori, ne l'était pas en déterminant les impulsions de manière séquentielle, procédure qui se déroule de la façon suivante : au début, le signal synthétique est engendré à partir de la mémoire du filtre de synthèse qui, en pratique, est un filtre autorégressif : on détermine alors l'erreur perceptuelle entre ce signal et l'original ; ceci permet de calculer la localisation et l'amplitude de la première impulsion. Une nouvelle erreur perceptuelle est

obtenue en retranchant la contribution de cette impulsion. La procédure d'estimation de nouvelles impulsions est ensuite réitérée jusqu'à ce que l'erreur perceptuelle soit, par exemple, inférieure à un certain seuil.

Les résultats expérimentaux [3] montrent que l'énergie de l'erreur diminue lorsque le nombre d'impulsions augmente. Toutefois au-delà d'une certaine limite (environ une impulsion toutes les millisecondes), la contribution de nouvelles impulsions est négligeable. Supposons que l'on souhaite M impulsions par fenêtre analysée. L'estimation de la position t_{m+1} et de l'amplitude W_{m+1} de la (m+1) ème, $1<m<M$, s'effectue à partir de la séquence v_t des m précédentes, v_t étant définie par :

$$(2)\quad v_t = \sum_{i=1}^{m} W(i) d_{t,t_i} \quad \text{avec} \quad \begin{cases} d_{t,t_i} = 1 \text{ si } t = t_i \\ = 0 \text{ ailleurs.} \end{cases}$$

L'erreur pondérée après placement de (m+1) impulsions, e_t^{m+1}, s'exprime en fonction de e_t^m par :

$$(3)\quad e_t^{m+1} = e_t^m - W_{m+1}\, h(t-t_{m+1})$$

h(t) désignant la Réponse Impulsionnelle à l'instant t du filtre de fonction de transfert $H(\gamma z)$;la nouvelle impulsion est déterminée afin que l'énergie

(4) $E(m+1)$ de e_t^{m+1} soit minimum.

$$(4)\quad E(m+1) = \sum_t (e_t^{m+1})^2$$

La minimisation de $E(m+1)$ par rapport à l'amplitude inconnue W_{m+1} permet de déterminer cette inconnue.

$$(5)\quad W_{m+1} = \frac{\alpha^{m+1}_{t_{m+1}}}{\Phi_{m+1,m+1}}$$

où . $\alpha^{m+1}_{t_{m+1}}$ désigne l'intercorrélation entre $\{e_t^m\}$ et

$\{h(t-t_{m+1})\}$

$$(6)\quad \alpha^{m+1}_{t_{m+1}} = \sum_t e_t^m\, h(t-t_{m+1})$$

. Φ_{m_1,m_2} représente l'intercorrélation des réponses impulsionnelles débutant aux instants t_{m_1} et t_{m_2}.

$$(7)\quad \Phi_{m_1,m_2} = \sum_t h(t-t_{m_1})h(t-t_{m_2})$$

L'énergie E(m+1), minimale par rapport à W_{m+1}, s'obtient en reportant (5) dans (3) :

$$(8)\quad E(m+1) = E(m) - \frac{(\alpha^{m+1}_{t_{m+1}})^2}{\Phi_{m+1,m+1}}$$

La (m+1) ème impulsion est alors positionnée en cherchant la valeur t_{m+1} pour laquelle l'énergie E(m+1) est minimum, ce qui revient à déterminer t_{m+1}.

de sorte que le rapport $\frac{(\alpha^{m+1}_{t_{m+1}})^2}{\Phi_{m+1,m+1}}$ soit maximum. Ce rapport, permettant de placer les impulsions, est souvent appelé "fonction de localisation". Par conséquent, la procédure consistant à estimer la (m+1) ème impulsion à partir des m précédentes est la suivante :

1) placement de l'impulsion afin de maximiser la fonction de localisation.

2) calcul de l'amplitude w_{m+1} et, éventuellement, réestimation de l'ensemble des impulsions positionnées précédemment.

Deux approches sont en effet, envisageables à cette étape :

a) les amplitudes des m-impulsions trouvées précédemment ne sont pas modifiées par l'introduction de la dernière dont l'amplitude est alors calculée à partir de (5).

b) on réestime, à chaque étape ou seulement lorsque toutes les impulsions ont été positionnées, l'ensemble des amplitudes en résolvant le système (9).

$$(9)\quad \underline{\Phi}\,\underline{W} = \underline{\alpha}$$

où $\underline{W}$ et $\underline{\alpha}$ désignent les vecteurs associés aux $\{W_j\}$ et $\{\alpha^{m+1}_{t_i}\}$ définis par (5) et (6), Φ la matrice des $\{\Phi_{i,j}\}$ déterminés par (7), $(1<i<m+1, 1<j<m+1)$

Remarque : la structure de la matrice $\underline{\Phi}$ des corrélations de la réponse impulsionnelle est fonction des bornes de sommation fixées dans (6) et (7). Si le signal est supposé nul à l'extérieur de la fenêtre analysée, la matrice a une structure de Toëplitz et la valeur prise par l'autocorrélation $\Phi_{i,i}$ $1<i<m+1$ est constante ce qui n'est pas le cas lorsque, au contraire, aucune hypothèse n'est faite en dehors de l'intervalle d'observation.

La qualité de la synthèse obtenue est satisfaisante pour un débit de 9.6 kbits/s. Mais en dessous de ce seuil, la détérioration devient importante : Singhal et Atal [11], Trancosco et al [12] proposent des améliorations de la méthode multiimpulsionnelle : utilisation d'une fenêtre de Hamming plutôt que d'une fenêtre triangulaire, diminution du nombre d'impulsions dans une période fondamentale, corrections de phase de la réponse impulsionnelle du filtre LPC ... .

2. MODELES A COEFFICIENTS DEPENDANT DU TEMPS

L'idée d'introduire les non-stationnarités du signal dans le modèle en décomposant les coefficients sur une base de fonctions a été émise initialement par Rao [4] et Mendel [5] au début des années 1970. Des travaux effectués par Liporace [6], Hall et al. [7], Grenier [8] ont ensuite permis de développer cette idée.

2.1. Structure transverse

Comme précédemment, supposons que le signal s_t soit observé sur l'intervalle de temps [0,T]. La relation définissant le modèle AR à coefficients dépendant du temps est donnée par (10) en fonction du résidu r_t

$$(10)\quad y_t + a_1(t-1)y_{t-1} + \ldots + a_p(t-p)y_{t-p} = r_t$$

L'hypothèse fondamentale est la suivante : chacun des coefficients $a_i(t)$ peut être décomposé linéairement (11) sur une base de fonctions $[f_o(t),\ldots,f_m(t)]$, de degré m, ces fonctions étant connues a priori et définies sur le même intervalle de temps [0,T] :

$$(11)\quad a_i(t) = \sum_{j=o}^{m} a_{ij} f_j(t)$$

notons Y_t la projection du signal y_t sur la base de fonction (12), le signe ' indiquant que le vecteur doit être transposé

$$(12)\quad Y_t = [f_o(t) \ \ldots \ f_m(t)]' y_t$$

une combinaison de (10), (11) et (12) permet de mettre en évidence la relation (13) :

$$(13)\quad y_t + [Y'_{t-1} \ \ldots \ Y'_{t-p}]\,\theta = r_t$$

où θ est un vecteur contenant les paramètres à déterminer

$$(14) \quad \theta=[a_{10}\cdots a_{1m}\, a_{20}\cdots a_{pm}]^{1}$$

(13) montre que la modélisation non-stationnaire d'un signal scalaire y_t a été réduite à la modélisation d'un signal vectoriel, non-stationnaire mais pour lequel le modèle à estimer est stationnaire.
Si la variance du résidu r_t est constante, le vecteur θ des paramètres à identifier peut facilement être déterminé en maximisant la vraisemblance $p(y_o,\ldots, y_t|\theta)$; cette démarche conduit aux équations de Yule-Walker (15) :

$$(15) \quad \sum_{t=p}^{T} \begin{bmatrix} \overline{Y}_{t-1} \\ \vdots \\ \overline{Y}_{t-p} \end{bmatrix} [Y'_{t-1}\cdots Y'_{t-p}]\theta = -\sum_{t=p}^{T} \begin{bmatrix} \overline{Y}_{t-1} \\ \vdots \\ \overline{Y}_{t-p} \end{bmatrix} y_t$$

2.2 Structure en treillis

Dans le cas stationnaire, la structure en treillis est souvent utilisée, à la place de la structure transverse. Dans le cadre de la modélisation non-stationnaire, le filtre est constitué par une cascade de p-cellules, chaque cellule ayant la même structure interne (figure 2), définie par (16).

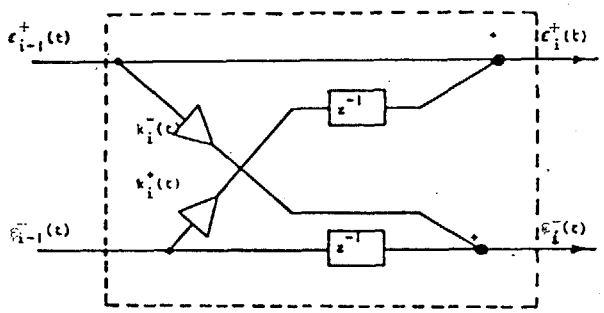


Figure 2 : structure de la ième cellule en treillis dans le cas non-stationnaire.

$\varepsilon^{+}_{i}(t)$ représente l'erreur de prédiction directe obtenue quand y_t est prédit à partir de $\{y_{t-1},\ldots, y_{t-i}\}$ tandis que $\varepsilon^{-}_{i}(t)$ désigne l'erreur de prédiction rétrograde obtenue par prédiction de y_{t-i} à partir de $\{y_t\cdots y_{t-i+1}\}$.

$$(16) \quad \begin{bmatrix} \varepsilon^{+}_{i}(t) \\ \varepsilon^{-}_{i}(t) \end{bmatrix} = \begin{bmatrix} 1 & k^{+}_{i}(t-1) \\ k_i(t) & 1 \end{bmatrix} \begin{bmatrix} \varepsilon^{+}_{i-1}(t) \\ \varepsilon^{-}_{i-1}(t-1) \end{bmatrix}$$

Une telle structure diffère essentiellement du cas stationnaire par la présence de deux coefficients de réflexion $k^{+}_{i}(t)$ et $k^{-}_{i}(t)$ et deux opérateurs 'retard' (au lieu d'un seul). L'hypothèse de décomposition des coefficients de réflexion sur la base de fonctions est analogue à celle de la structure transverse, les projections $\{k_{ij}\}$ $0<j<m$ étant calculées à la sortie de la ième cellule de sorte à minimiser les variances des erreurs de prédiction directes et rétrogrades.

L'utilisation de ces modèles évolutifs en restitution et à plus long terme en synthèse de parole est-elle valable ? Une difficulté de première évidence est le manque de stabilité des modèles ainsi estimés.

Prenons l'exemple de la structure transverse et désignons par $A_t(z)$ la transformée en z de la séquence $\{1,a_1(t),\ldots,a_p(t)\}$, les coefficients de cette séquence étant fixés à l'instant t. Dans le cadre de la restitution, nous nous intéressons à la stabilité locale du modèle autorégressif défini par (14). En effet, une telle stabilité assure, qu'à tout instant, les zéros de $A_t(z)$ restent intérieurs au cercle unité, ceci pour éviter les sauts d'énergie qu'induisent les excursions des zéros hors du cercle unité, et ont pour conséquence de dégrader la parole synthétisée.

Aucun fondement théorique n'assurant cette condition, une solution consiste à renvoyer à l'intérieur du cercle unité, lors de la phase de synthèse du signal, tous les zéros dont le module est plus grand que un. Mais une telle solution nécessite de nombreux calculs, ce qui a pour conséquence d'augmenter de manière importante le temps de restitution.

Une autre possibilité est d'imposer la stabilité du modèle, non pas lors de la restitution, mais dès l'analyse du signal. Mais avant de présenter la solution que nous avons mise en oeuvre, remarquons que l'estimation d'un filtre en treillis présente le même inconvénient que précédemment : même si les coefficients de réflexion $k_i(t)$ sont de module inférieur à 1 (condition de stabilité du filtre), leur approximation par une décomposition linéaire sur une base de fonctions n'est pas contrainte à appartenir à l'intervalle $]-1,+1[$. Une solution [9] consiste à forcer la stabilité du modèle au moyen d'une transformation non-linéaire des coefficients de réflexion, de sorte que l'intervalle de stabilité ne soit plus défini par $]-1,+1[$ mais soit étendu à $]-\infty,+\infty[$, le nouveau jeu de coefficients étant à estimer sur la même base de fonctions.

La fonction retenue transforme les $k_i(t)$, $1<i<p$, en Log Area Ratios ou fonctions d'aire logarithmiques définies par (17) :

$$(17) \quad \gamma_i(t) = \mathrm{Ln}\,\frac{1+k_i(t)}{1-k_i(t)} \quad \text{avec } \gamma_i(t)=\sum_{j=o}^{m} \gamma_{ij} f_j(t)$$

L'intérêt de déterminer les $\{\gamma_{ij}\}$ est d'assurer, en pratique, la stabilité du modèle estimé et donc de supprimer toute procédure de stabilisation.

2.3 LAR par l'intermédiaire d'un treillis

L'estimation la plus directe des $\{\gamma_{ij}\}$ consisterait à minimiser, à la sortie de la ième cellule $(1<i<p)$, la somme du carré des erreurs. Mais une telle procédure conduit à un problème non-linéaire et coûteux en calculs. Une autre solution est de résoudre le problème en deux étapes, chacune d'entre elles étant linéaire : dans un premier temps, le filtre en treillis est déterminé par les projections $\{k_{ij}\}$ de ses coefficients de réflexion sur la base de fonctions ; les trajectoires des LAR sont ensuite ajustées pour approcher, au sens des Moindres carrés, celles des $k_i(t)$.

Bien qu'une telle méthode soit approchée, elle permet toutefois d'obtenir des résultats satisfaisants ; le coût de calculs lors de la phase d'analyse est cependant plus important que pour l'estimation d'un modèle autorégressif tel que celui à structure transverse : il faut tout d'abord estimer un filtre en treillis et ensuite calculer les LAR à partir des coefficients de réflexion obtenus ; mais cet inconvénient est contrebalancé lors de la restitution du signal puisqu'il n'est plus nécessaire d'introduire une méthode de stabilisation.

3. CODAGE MULTI-IMPULSIONNEL POUR MODELES EVOLUTIFS

Déterminer la source d'excitation d'un modèle est une opération plus délicate dans le cas non-stationnaire, où la réponse impulsionnelle dépend de deux indices sur le temps, que dans le cas stationnaire. Le signal s_t réponse du filtre non-stationnaire à l'excitation ε_t s'écrivant :

$$s_t = \sum_{u=o}^{t} h_t(t-u)\,\varepsilon_{t-u}$$

où . t est défini sur l'intervalle de temps [0,T].
. $h_t(t-u)$ désigne la valeur à l'instant t de la réponse impulsionnelle débutant à l'instant (t-u).

La technique multi-impulsionnelle que nous avons mise en oeuvre dans le cadre des modèles évolutifs est semblable, dans, dans son principe, à celle d'Atal : position et amplitude des impulsions sont déterminées séquentiellement, l'ensemble des amplitudes pouvant être réestimé une fois les impulsions positionnées. Seule la fonction de localisation est différente, celle proposée par Atal entraînant, comme nous allons le montrer dans le paragraphe suivant, un nombre d'opérations trop élevé dans le cas de l'analyse non-stationnaire.

3.1 Fonction de localisation

Des calculs analogues au paragraphe 1.2 conduisent à estimer la position de la (m+1)ème impulsion en déterminant la valeur t_{m+1} pour laquelle le rapport :

$$(18) \qquad \frac{\left(\sum_{t=t_{m+1}}^{T} e_t^m \, h_t(t_{m+1})\right)^2}{\sum_{t=t_{m+1}}^{T} h_t^2(t_{m+1})} \quad \text{est maximum.}$$

Mais une telle démarche nécessite le calcul de la réponse impulsionnelle à chaque instant et pour chaque nouvelle position t_{m+1} ce qui conduit à un nombre d'opérations et un encombrement mémoire importants. Aussi avons-nous préféré déterminer, à partir des deux remarques suivantes, une autre fonction de localisation qui puisse nous permettre de placer les impulsions avec un coût de calculs moins importants.
La première remarque porte sur l'intercorrélation (19) entre le signal e_t^m et la réponse impulsionnelle du modèle autorégressif : l'intercorrélation à condition que cette réponse impulsionnelle soit à durée limitée, peut être obtenue par filtrage du signal à travers le modèle (20), la connaissance de la réponse impulsionnelle à chaque instant n'étant pas nécessaire :

$$(19) \qquad \alpha_{t_j}^{m+1} = \sum_{t=t_j}^{T} e_t^m \, h_t(t_j)$$

$$(20) \qquad \alpha_{t_j}^{m+1} = -\sum_{i=1}^{P} a_i(t_j)\alpha_{t_j+i}^{m+1} + e_{t_j}^m$$

L'intérêt de (20) est évident : le nombre d'opérations nécessaires au calcul de l'intercorrélation a nettement diminué, $\alpha_{t_j}^{m+1}$ s'obtenant par "filtrage inverse" du signal $e_{t_j}^m$ à travers le modèle autorégressif, les p-coefficients de prédiction étant, dans ce cas, tous calculés à l'instant t_j.
La deuxième remarque est liée à la présence, au dénominateur de la fonction de localisation (18), de l'autocorrélation de la Réponse impulsionnelle : cette autocorrélation peut être interprétée comme un facteur de pondération du carré de l'intercorrélation entre l'erreur perceptuelle et la réponse impulsionnelle ; la conséquence d'une telle pondération est de permettre une répartition relativement régulière des impulsions sur l'intervalle d'analyse.
Tirant parti de ces deux remarques, nous définissons la nouvelle fonction de localisation de la manière suivante : dans un premier temps, le signal issu du filtre perceptuel, excité par le signal s_t, est divisé par un facteur de poids qui peut être soit l'énergie du résidu obtenue après lissage, soit la variance de l'innovation du modèle. L'introduction d'un tel facteur a pour objectif de supprimer la pondération par l'autocorrélation de la réponse impulsionnelle ; le signal ainsi obtenu permet d'initialiser l'erreur perceptuelle e_t^o avant placement des impulsions.

La fonction de localisation de la (m+1) ème impulsion, $0<m<M$ où M est le nombre d'impulsions souhaité à l'intérieur de la fenêtre analysée, est ensuite déterminée par l'intercorrélation entre e_t^m et la réponse impulsionnelle, calculée à partir de (20).

La procédure d'estimation des impulsions est alors similaire à celle proposée par Atal :

a) calcul de la position de la (m+1) ème impulsion en déterminant la valeur t_{m+1} pour laquelle $|\alpha_{t_{m+1}}^{m+1}|$ est maximum.
b) calcul de l'amplitude W_{m+1} :

$$(21) \qquad W_{m+1} = \frac{\alpha_{t_{m+1}}^{m+1}}{\sum_{t=t_{m+1}}^{T} h_t^2(t_{m+1})}$$

c) correction de l'erreur perceptuelle :

$$(22) \qquad e_t^{m+1} = e_t^m - W_{m+1} \, h_t(t_{m+1})$$

L'algorithme est réitéré tant que le seuil choisi a priori, n'est pas atteint.

3.2 Optimisation des amplitudes

Dans l'algorithme décrit au paragraphe 3.1, nous avons supposé que l'introduction d'une nouvelle impulsion dans la séquence des précédentes ne modifiait pas l'amplitude de ces dernières, l'amplitude de chaque impulsion étant calculée de façon séquentielle à partir de (21).

La méthode présentée dans ce paragraphe consiste à réestimer de manière globale l'amplitude de l'ensemble des impulsions positionnées à l'aide de l'algorithme précédent.
Les amplitudes sont calculées en minimisant, à l'aide d'une méthode des moindres carrés à fenêtre glissante, l'énergie de l'erreur entre le signal original y_t et la réponse du modèle à la séquence d'impulsions dont on cherche les amplitudes ; ceci revient à résoudre au sens des moindres carrés, le système linéaire :

$$(22) \qquad Y = HW$$

où . Y désigne le vecteur du signal original
. H la matrice de la réponse impulsionnelle
. W le vecteur des impulsions

Dès lors, le principal point à fixer porte sur la définition de la fenêtre. Selon quels critères déterminer sa taille ? Comment la déplacer à l'intérieur de l'intervalle d'analyse ? ou question similaire : à partir de quel moment pourra-t-on affirmer que les amplitudes du début de la fenêtre sont correctement estimées et calculer les suivantes ?

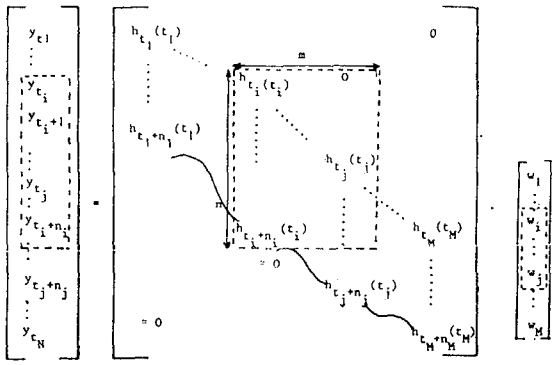
Le choix d'un critère s'avère donc nécessaire ; le principal paramètre qui intervient dans l'estimation des amplitudes étant la fonction de transfert du filtre, il apparaît judicieux d'estimer la longueur n de la fenêtre à partir de la durée de la réponse impulsionnelle du modèle : cette dernière devient comme précédemment, négligeable dès qu'elle est inférieure à un seuil fixé a priori. Ceci permet de déterminer le nombre d'échantillons n ainsi que le nombre d'impulsions r à l'intérieur de la fenêtre, n et m étant bien sûr variables d'une impulsion à la suivante.

Le système linéaire (22) s'écrit, alors, sous la forme matricielle représentée en (23).

L'estimation des amplitudes s'effectue de la manière suivante : on calcule l'amplitude des impulsions successives jusqu'à ce que la réponse impulsionnelle correspondant à la première impulsion de la fenêtre soit inférieure au seuil choisi.
Une fois ce seuil atteint, puisque l'influence de la réponse impulsionnelle correspondante est négligeable,

l'amplitude de la première impulsion est supposée correcte : on diminue alors la taille de la matrice en supprimant de la fenêtre cette impulsion et la réponse impulsionnelle associée ; l'amplitude d'une nouvelle impulsion est ensuite estimée en augmentant la taille de la fenêtre (introduction de cette dernière impulsion dans la fenêtre) puis en réitérant le processus.

(23)



4.4 Résultats

La figure 3 représente un exemple des signaux de parole obtenus à l'issue d'expériences mettant en oeuvre les méthodes précédemment décrites : estimation évolutive (paragraphe 2) du signal, excitation du modèle par la technique multi-impulsionnelle décrite au paragraphe 3.

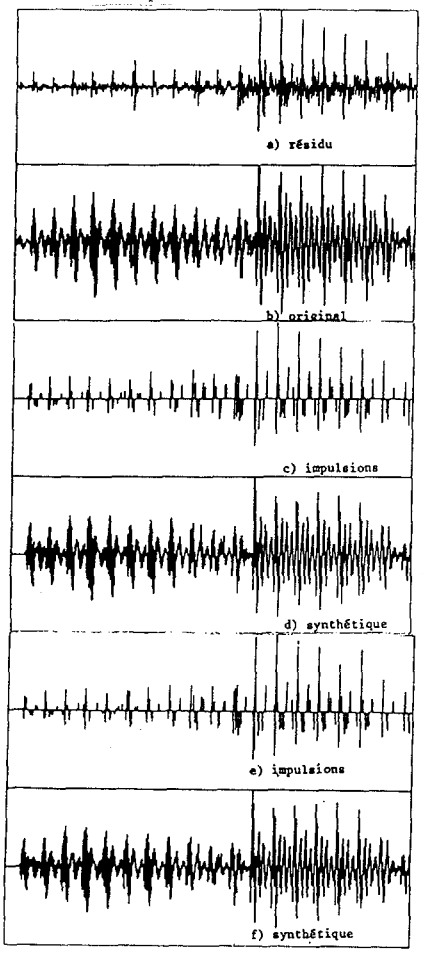


Figure 3 : exemples de signaux obtenus par analyse évolutive et codage multi-impulsionnel - durée du signal : 200 ms - a) résidu, b) original, c) séquence multi-impulsionnelle, d) signal synthétisé, e) séquence multi-impulsionnelle suivie d'une réestimation des amplitudes, f) signal synthétisé.

Le modèle évolutif estimé lors de la phase d'analyse est un modèle à structure transverse, d'ordre 12, chacun des prédicteurs étant approché par une combinaison linéaire sur une base de fonctions (dans le cas présent, il s'agit de la base de Fourier) de degré 5. La figure 3a) montre le résidu obtenu à l'issue de cette analyse.

La durée du signal étudié étant de 200 millisecondes, 200 impulsions ont été positionnées afin d'obtenir, en moyenne, une impulsion toutes les millisecondes.

CONCLUSION

Les résultats obtenus valident l'intérêt des modèles évolutifs en traitement de la parole, une telle technique permettant de paramétrer, de manière globale, l'évolution du signal sur le segment analysé.

Nous avons présenté le principe d'un algorithme pour estimer des fonctions d'aire logarithmiques à partir d'un filtre en treillis, ce qui permet d'améliorer le temps de restitution du signal. Ce sujet sera prochainement développé dans un article. La méthode multi-impulsionnelle étudiée pour de tels modèles montre qu'une réestimation globale des amplitudes après positionnement des impulsions n'est pas nécessaire, l'estimation séquentielle étant suffisante.

A l'issue de précédentes expériences, nous avons trouvé que les coefficients du modèle doivent être codés à raison d'environ 250 par seconde. Ceci est à comparer aux 500 coefficients par seconde que représente la LPC-10 (prédiction linéaire à l'ordre 10, un modèle toutes les 20 ms). Bien que le problème du codage de l'ensemble des paramètres du modèle évolutif reste encore à étudier, on peut espérer un gain d'un facteur 2 sur la LPC-10 avec une qualité analogue, si ce n'est meilleure ...

BIBLIOGRAPHIE

[1] C.K. UN, D.T. MAGILL, "The Residual-Excited Linear Prediction vocoder with Transmission Rate below 9.6 k bits/s", IEEE Trans. Comm., Vol. COM-23, pp 1466-1474, December 1975.

[2] V.R. VISWANATHAN, A.L. MIGGINS, W.H. RUSSEL, "Design of a robust baseband LPC coder for speech transmission over 9.6 kbits/s noisy channels" IEEE Trans, Vol. Com-30, n° 4, April 82, pp 663-673.

[3] B.S. ATAL, J.R. REMDE, "A new model of LPC excitation for producting natural-souding speech at low bit rates", IEEE ICASSP 82, pp 614-617, 1982.

[4] T.S. RAO, "The fitting of non-stationnary time-series models with time-dependent parameters", J. of the Royal Statist. Soc., Series B, Vol. 32, n°2, pp 312-322 , 1970.

[5] J.M. MENDEL, "A priori and a posteriori identification of time-varying parameters", 2nd Hawaï Int-Conf. on Syst. Sciences, pp 207-210, 1969.

[6] L.A. LIPORACE, "Linear estimation of non-stationary signals", J. Acoust. Soc. Amer., Vol. 58, n° 6, pp 1218-1295, 1975.

[7] M. HALL, A.V. OPPENHEIM, A. WILLSKY, "Time-varying parametric modelling of speech", Signal Processing, Vol. 5, n° 3, pp 267-285, 1983.

[8] Y. GRENIER, "Time-dependent ARMA modeling of non-stationary signals", IEEE Trans. on ASSP, Vol. 31, n° 4, pp 899-911, 1983.

[9] M.C. CHEVALIER, Y. GRENIER, "Stable time-varying autoregressive models through Log Area Ratios", IEEE ICASSP-85.

[10] M.C. CHEVALIER, G. CHOLLET, Y. GRENIER, "Speech analysis and restitution using time-dependent autoregressive models", IEEE ICASSP-85.

[11] S. SINGHAL, B.S. ATAL, "Improving performance of multi-pulse LPC Coders at low bit rates", Proc. ICASSP-84, pp 10-2-1 - 10-2-4.

[12] J.M. TRANCOSO, "A study on short-time phase and multi-pulse LPC", Proc. ICASSP-84, pp 10-3-1 - 10-3-4.