

Amélioration de l'estimation  
du spectre de parole  
par suppression d'impulsions  
dans le résidu

Improved ARMA spectrum estimation of speech signals

by impuls suppression in the residue



Gang FENG

Institut de la Communication Parlée, UA 368, 46, avenue Félix-Viallet, 38031 GRENOBLE Cedex

Diplôme d'Ingénieur en Radio-Électronique de l'Institut Polytechnique de Huazhong en Chine (1982). Thèse de Doctorat en Traitement du Signal et Automatique de l'Institut National Polytechnique de Grenoble (1986). Ses recherches ont porté sur la modélisation articulatoire-acoustique et sur le traitement du signal de parole, en particulier pour les voyelles nasales. Depuis 1982 a contribué à l'équipement en traitement du signal et en synthèse de l'Institut de Phonétique de Grenoble (Institut de la Communication Parlée, CNRS UA n° 368) en collaboration avec le Centre d'Études des Phénomènes Aléatoires et Géophysiques (CNRS UA n° 346). Membre du GRECO Systèmes Adaptatifs en Robotique, Traitement du signal et Automatique depuis 1985. Actuellement poursuit une recherche post-doctorale sur le codage du signal de parole à débits réduits, dans le département CMC du CNET Lannion A.



Jean-Louis LACOUME

8, allée du Passeur, 38620 GIERES

Ancien élève de l'École Normale Supérieure, Agrégé de Physique (1964), Docteur ès Sciences (1969), Professeur à l'Institut National Polytechnique de Grenoble, Directeur du CEPHAG. Domaine de recherche : Géophysique : étude de l'environnement ionisé de la terre par sondage électromagnétique, Traitement du Signal : Analyse spectrale et interspectrale, applications en Instrumentation, Acoustique Sous-Marine, Géophysique.

RÉSUMÉ

Les méthodes récursives d'estimation des paramètres d'un modèle ARMA sont *a priori* attractives pour l'analyse spectrale du signal de parole. Mais en l'absence d'information sur le signal d'excitation, l'application de ces méthodes ne peut être directe. Après avoir mis en évidence le mécanisme qui limite la précision de l'estimation dans ce cas, nous présenterons une solution fondée sur la détection et la suppression d'impulsions dans le résidu. Cette approche permet d'améliorer l'estimation des paramètres dans le cas où le résidu est proche d'un train d'impulsions. Nous examinerons enfin l'utilité de cette méthode pour l'analyse spectrale du signal de parole.

MOTS CLÉS

Analyse spectrale, signal de parole, modèle ARMA, estimation, algorithme récursif.

**SUMMARY**

*Recursive methods for ARMA parameter estimation is attractive for speech spectrum analysis. These methods can not be applied directly to speech analysis where the input signal (the excitation source of the vocal tract) is unknown. In this paper, after clarifying the mechanism, in the case of unknown input signal, that limits the estimation precision, we describe a solution based on the impuls detection and suppression in the residue. This approach can improve the parameter estimation accuracy mainly for the case where the residue is close to an impuls train. It has been shown that this method improves the spectrum estimation of speech signals.*

**KEY WORDS**

*Spectrum analysis, speech signal, ARMA model, estimation, recursive algorithm.*

**TABLE DES MATIÈRES****1. Introduction****2. Le modèle**

- 2.1. Moindres carrés récurrents
- 2.2. Introduction du signal d'entrée

**3. Le problème de la précision de l'estimation****4. La suppression d'impulsions dans le résidu**

- 4.1. Excitation par un bruit blanc
- 4.2. Excitation par un train d'impulsions
- 4.3. Excitation par une forme d'onde complexe
- 4.4. Analyse d'un signal de parole naturel

**5. Conclusion****Remerciements****Bibliographie****1. Introduction**

Le modèle AR (autorégressif) a prouvé son efficacité dans l'analyse spectrale du signal de parole. La raison en est que la parole est, dans la plupart des cas, produite par l'excitation des résonances acoustiques du conduit vocal, et que celles-ci correspondent bien à une fonction de transfert tout pôle.

Mais la production de la parole est en réalité bien plus complexe. La source d'excitation glottale et le couplage source/conduit peuvent de toute évidence modifier la fonction de transfert du conduit vocal [6]. Cette dernière peut elle-même présenter une structure complexe, notamment une ou plusieurs paires pôle-zéro, dans le cas où plusieurs cavités sont couplées acoustiquement [4, 5]. C'est pourquoi on peut obser-

ver dans le spectre du signal de parole des creux marqués, des variations brusques ou des parties plates, qui sont difficiles à modéliser par un simple filtre tout-pôle d'ordre limité [8].

Afin d'améliorer la modélisation du spectre de parole, on a été amené à introduire les modèles ARMA [2, 3, 12]. Depuis quelques années, de nombreux algorithmes récurrents d'estimation des paramètres d'un modèle ARMA ont été développés [10, 14]. Ils permettent avantageusement de suivre les variations des paramètres, d'en avoir une évaluation précise et d'estimer simultanément les pôles et les zéros. Il paraît donc intéressant de les appliquer à l'analyse spectrale du signal de parole.

Cette application ne peut cependant pas être directe. Rappelons que ces algorithmes d'identification sont normalement conçus pour estimer les paramètres d'un système inconnu, ou modèle, soumis à un signal d'excitation connu. Or pour la parole, le signal d'entrée ne nous est pas accessible !

Nous voulons montrer dans cet article, que lorsqu'on ne connaît pas le signal d'excitation, la précision de l'estimation des paramètres est très limitée. Il est clair que l'ajustement des paramètres du modèle dépend essentiellement du signal résiduel. Dans le cas où le signal d'entrée est inconnu, celui-ci se retrouve dans le signal résiduel et participe donc à l'erreur de prédiction. Cette présence du signal d'entrée inconnu dans le résidu perturbe le mécanisme d'ajustement, et diminue fortement la précision de l'estimation. Ce point sera repris de façon détaillée plus bas.

Après cet examen de l'influence du signal d'entrée inconnu, nous présenterons une approche de la détection et de la suppression de l'excitation impulsionnelle dans le résidu. On élimine ainsi l'influence du signal d'entrée inconnu dans le résidu, augmentant par là-même la précision de l'estimation. L'efficacité de cette méthode dépend de la structure du résidu. Dans le cas d'une excitation par un train d'impulsions, la méthode est très efficace. Même lorsque l'excitation est plus complexe et lorsqu'il existe une interaction entre l'excitation et le système, cette méthode reste efficace à condition que le résidu soit proche d'un train d'impulsions. Nous examinerons enfin l'utilité de cette méthode pour le signal de parole.

## 2. Le modèle

### 2.1. MOINDRES CARRÉS RÉCURSIFS

Parmi les diverses formulations des algorithmes récurrents, beaucoup peuvent se développer à partir de la méthode des moindres carrés récurrents. Nous la rappelons ici brièvement.

La relation entre l'entrée  $u_k$  et la sortie  $v_k$  d'un système dynamique du type ARMA ( $p, q$ ) peut se décrire par :

$$(1) \quad v_k = - \sum_{i=1}^p \alpha_i v_{k-i} + \sum_{i=0}^q \beta_i u_{k-i}$$

Supposons que la sortie soit bruitée par  $w_k$  (en général il s'agit d'un bruit blanc gaussien) :

$$(2) \quad y_k = v_k + w_k$$

On peut donc écrire pour  $y_k$  l'équation :

$$(3) \quad y_k = - \sum_{i=1}^p \alpha_i y_{k-i} + \sum_{i=0}^q \beta_i u_{k-i} + \sum_{i=1}^p \alpha_i w_{k-i} + w_k$$

Nous cherchons à estimer les paramètres du système ( $\alpha_i$  et  $\beta_i$ ) à partir de l'observation de l'entrée  $u_k$  et de la sortie  $y_k$ . Pour réaliser cette estimation on utilise un modèle ARMA de référence dont les paramètres sont  $a_i$  et  $b_i$ , et l'on compare la sortie de ce modèle à l'instant  $k$ ,  $\hat{y}_k$  à la sortie observée  $y_k$ . La sortie  $\hat{y}_k$  obtenue à partir des entrées et des sorties aux instants précédents ( $u_n$  et  $y_n, n < k$ ) est un prédicteur linéaire de  $y_k$  représenté par l'équation :

$$(4) \quad \hat{y}_k = - \sum_{i=1}^p a_i y_{k-i} + \sum_{i=0}^q b_i u_{k-i}$$

Aucun terme de bruit n'apparaît dans cette équation car le bruit est inconnu. L'erreur de prédiction est donc (fig. 1) :

$$(5) \quad \varepsilon_k = y_k - \hat{y}_k = y_k + \sum_{i=1}^p a_i y_{k-i} - \sum_{i=0}^q b_i u_{k-i}$$

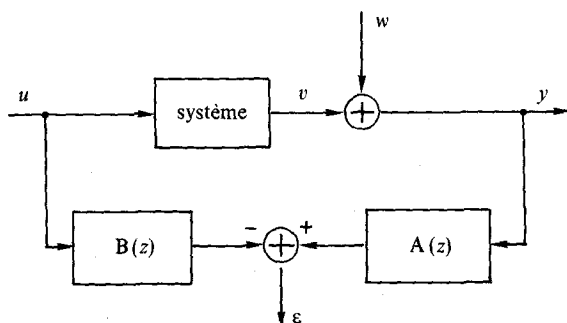


Fig. 1. - Identification d'un système dynamique par la méthode des moindres carrés. Le modèle ARMA est constitué par deux filtres :

$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i} \quad \text{et} \quad B(z) = \sum_{i=0}^q b_i z^{-i}$$

dont les paramètres  $a_i$  et  $b_i$  sont ajustés de telle façon que l'erreur de prédiction  $\varepsilon$  soit minimale.

En définissant :

$$(6) \quad \theta = [a_1, \dots, a_p, b_0, \dots, b_q]^T$$

$$(7) \quad \phi_k = [-y_{k-1}, \dots, -y_{k-p}, u_k, \dots, u_{k-q}]^T$$

on peut écrire les équations sous forme matricielle :

$$(8) \quad \hat{y}_k = \phi_k^T \theta$$

$$(9) \quad \varepsilon_k = y_k - \phi_k^T \theta$$

Dans le cas des moindres carrés en « batch », on acquiert les données sur un intervalle  $N$  ( $N > 2p$ ). On peut ainsi obtenir  $N$  équations linéaires pour les paramètres  $a_i$  et  $b_i$  avec  $k \in [p+1, p+N]$  et alors :

$$(10) \quad \varepsilon = y - \Phi \theta$$

avec

$$(11) \quad \varepsilon = [\varepsilon_{p+1}, \dots, \varepsilon_{p+N}]^T$$

$$(12) \quad y = [y_{p+1}, \dots, y_{p+N}]^T$$

et

$$(13) \quad \Phi = [\phi_{p+1}, \dots, \phi_{p+N}]^T$$

En minimisant l'erreur quadratique :

$$(14) \quad J = \sum_{k=p+1}^{p+N} \varepsilon_k^2 = \varepsilon^T \varepsilon$$

par

$$(15) \quad \left. \frac{\partial J}{\partial \theta} \right|_{\theta = \hat{\theta}} = 0$$

on obtient :

$$(16) \quad \hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T y$$

C'est la solution bien connue de la méthode des moindres carrés, appliquée initialement aux systèmes dynamiques par Kalman [7].

On peut résoudre de façon récurrente les équations des moindres carrés à l'aide du lemme d'inversion de matrice [1]. Supposons qu'à l'instant  $k$  on connaisse déjà  $P_k = (\Phi^T \Phi)^{-1}$  et  $\hat{\theta}_{k-1}$ , pour les données suivantes  $y_k$  et  $u_k$ , l'utilisation des équations récurrentes suivantes permet d'obtenir directement une nouvelle estimation des paramètres :

$$(17) \quad \hat{\theta}_k = \hat{\theta}_{k-1} + \frac{P_k \phi_k}{1 + \phi_k^T P_k \phi_k} (y_k - \phi_k^T \hat{\theta}_{k-1})$$

$$(18) \quad P_{k+1} = P_k - \frac{P_k \phi_k \phi_k^T P_k}{1 + \phi_k^T P_k \phi_k}$$

Pour initialiser cet algorithme à l'instant  $k=1$ , on peut poser  $P_1 = cI$  et  $\theta_0 = 0$  si  $c = Cte$  assez grande.

### 2.2. INTRODUCTION DU SIGNAL D'ENTRÉE

Dans la méthode des moindres carrés, si l'on ne connaît pas l'entrée  $u$ , il n'est plus possible d'estimer

correctement les paramètres (notamment les  $b_i$ ). Pour contourner cette difficulté, on peut songer à utiliser le résidu comme estimation de l'excitation [12]. La démarche précise est la suivante : supposons que les paramètres soient correctement estimés; alors le résidu sera un bruit blanc. Ce dernier pouvant servir d'entrée « équivalente », l'estimation des paramètres  $b_i$  est donc possible ( fig. 2).

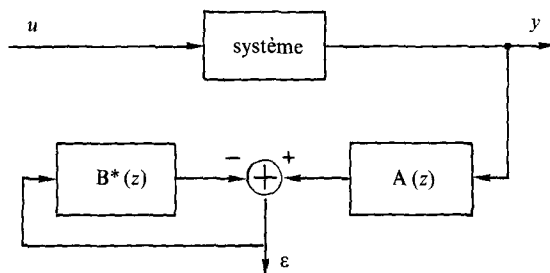


Fig. 2. — Estimation des paramètres lorsque le signal d'entrée  $u$  est inconnu. On utilise dans ce cas le résidu comme estimation du signal d'entrée. Notons

$$B^*(z) = \sum_{i=1}^q b_i z^{-i}.$$

A partir de ce principe, on peut construire un prédictor dont l'équation du résidu est la suivante :

$$(19) \quad \varepsilon_k = y_k + \sum_{i=1}^p a_i y_{k-i} - \sum_{i=1}^q b_i \varepsilon_{k-i}$$

Notons que l'entrée inconnue  $u$  est remplacée par le résidu  $\varepsilon$ . On réécrit maintenant cette équation de la façon suivante :

$$(20) \quad \varepsilon_k = - \sum_{i=1}^q b_i \varepsilon_{k-i} + \left( y_k + \sum_{i=1}^p a_i y_{k-i} \right)$$

Ceci révèle que le résidu est la sortie d'un filtre autorégressif excité par le signal issu du filtre  $A(z)$ . Ce filtre autorégressif est défini par :

$$(21) \quad \frac{1}{B(z)} = \frac{1}{1+B^*(z)} = \frac{1}{1 + \sum_{i=1}^q b_i z^{-i}}$$

Ainsi, le résidu correspond au signal filtré par les filtres  $A(z)$  et  $1/B(z)$  ( fig. 3), équivalents au filtre unique  $A(z)/B(z)$ . La minimisation de  $\varepsilon$  correspond donc au blanchissement du spectre du signal de sortie

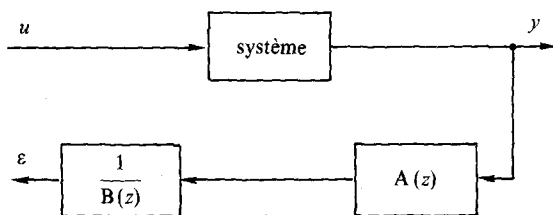


Fig. 3. — Modèle équivalent à la figure 2.

par le filtre  $A(z)/B(z)$ . Autrement dit, on modélise le signal par le filtre  $B(z)/A(z)$ . Si  $B^*(z)=0$ , c'est-à-dire si  $B(z)=1$ , on retrouve l'analyse par le modèle AR classique.

### 3. Le problème de la précision de l'estimation

L'utilisation du résidu comme entrée estimée permet d'obtenir un ensemble de coefficients  $a_i$  et  $b_i$ . Mais sans information sur le signal d'entrée, même dans le cas où celui-ci est simple, il n'est pas *a priori* évident que l'estimation des paramètres soit toujours de bonne qualité.

Dans la méthode des moindres carrés récurrents, l'ajustement des paramètres du modèle, c'est-à-dire la mise à jour du vecteur  $\hat{\theta}$  [cf. équation (17)] dépend de façon critique du signal résiduel  $\varepsilon$ , qui est en fait l'erreur de prédiction. Ce mécanisme conduit à une bonne précision de l'estimation car l'erreur de prédiction, reflétant la différence entre le modèle et le système, est en principe toujours minimisable. Elle peut même attendre une valeur nulle s'il n'y a pas de bruit et si le modèle est parfaitement correct.

En revanche, dans le cas où on utilise le résidu comme entrée estimée, le rôle du résidu est complètement différent. Même si les paramètres sont parfaitement estimés, le résidu ne peut être nul : il sera de l'ordre de grandeur du signal d'entrée (cf. fig. 3). Dans ce cas le résidu comprend non seulement l'erreur de prédiction proprement dite, mais également le signal d'entrée inconnu.

Du point de vue du mécanisme d'ajustement qui ne doit prendre en compte que l'erreur de prédiction, cette présence du signal d'entrée dans le résidu est indésirable. D'une part, le signal d'entrée, considéré comme « blanc » et d'énergie déterminée, est non minimisable et d'autre part, il est impossible pour le mécanisme d'ajustement de le distinguer de l'erreur de prédiction. En toute rigueur, l'ajustement devrait être effectué à partir du signal  $\varepsilon_k - u_k$ .

Par conséquent, lorsque les paramètres sont encore loin d'une estimation optimale, le résidu se compose principalement de l'erreur de prédiction et le mécanisme est efficace pour l'ajustement du modèle. Mais lorsque l'estimation est presque parfaite, la part de l'erreur due à la différence entre modèle et système devient minuscule et l'ajustement n'est plus efficace. De ce fait, la précision de l'estimation est limitée.

De toute évidence, nous pouvons prévoir que cette imprécision de l'estimation se manifesterait plus clairement dans les zones où l'énergie est relativement faible et où la structure du spectre est relativement complexe comme, par exemple, au voisinage des paires pôle-zéro.

A titre de test, nous avons étudié un système dont la fonction de transfert possède deux pôles et une paire pôle-zéro, c'est-à-dire un système ARMA (6,2) ( fig. 4). La faible distance du pôle et du zéro de la paire donne une zone spectrale difficile à estimer.

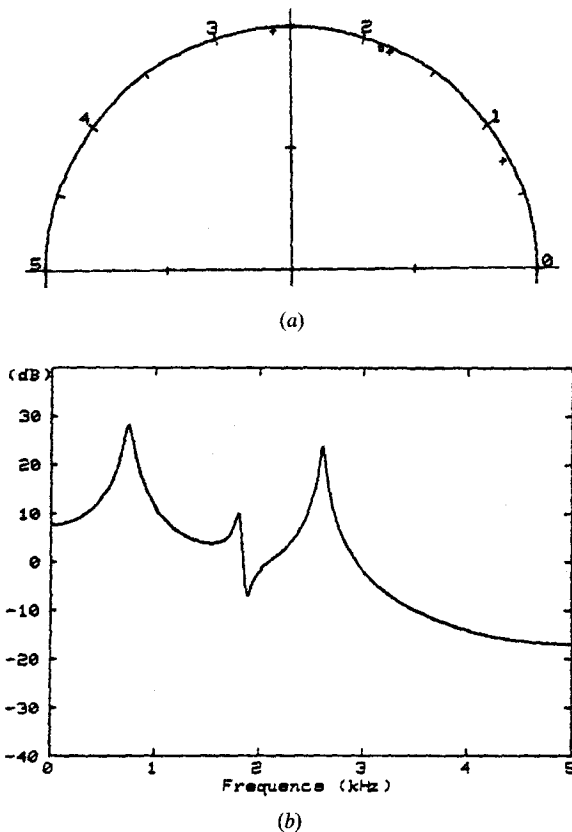


Fig. 4. — Un système ARMA (6,2) pour la génération des signaux de test. (a) Répartition des pôles et des zéros. (b) Fonction de transfert correspondante. Notons que la paire pôle-zéro correspond à une zone de fréquence où l'énergie est relativement faible par rapport à celle correspondant aux pôles seuls.

Les signaux de test ont été produits avec deux types d'excitation : un train d'impulsions de période égale à 6,4 ms (64 échantillons pour une fréquence d'échantillonnage de 10 kHz) et un bruit blanc gaussien. Dans un premier temps, l'ordre du modèle était le même que celui du système, c'est-à-dire 6 pour la partie AR et 2 pour la partie MA.

Nous présentons les résultats de simulation aux figures 5 (excitation impulsionnelle) et 6 (excitation par bruit blanc). En (a) on trouve les fonctions de transfert calculées à partir des paramètres estimés au cours de l'analyse, et tracées chaque cinq pas de récursion. Les parties (b), (c) et (d) présentent de haut en bas le signal d'entrée, le signal de sortie du système et le signal résiduel.

Nous pouvons constater tout d'abord que dans le cas de l'excitation impulsionnelle, l'estimation des deux pôles principaux ne pose pas de problème, à l'inverse de ce que l'on constate pour la paire pôle-zéro (fig. 5 a).

Pour la première période (64 échantillons) de l'analyse, l'estimation de la partie MA est nulle. Ceci est dû au fait que l'information relative à la partie MA n'existe que dans les  $q$  points suivants l'excitation impulsionnelle. Or dans notre analyse, l'algorithme

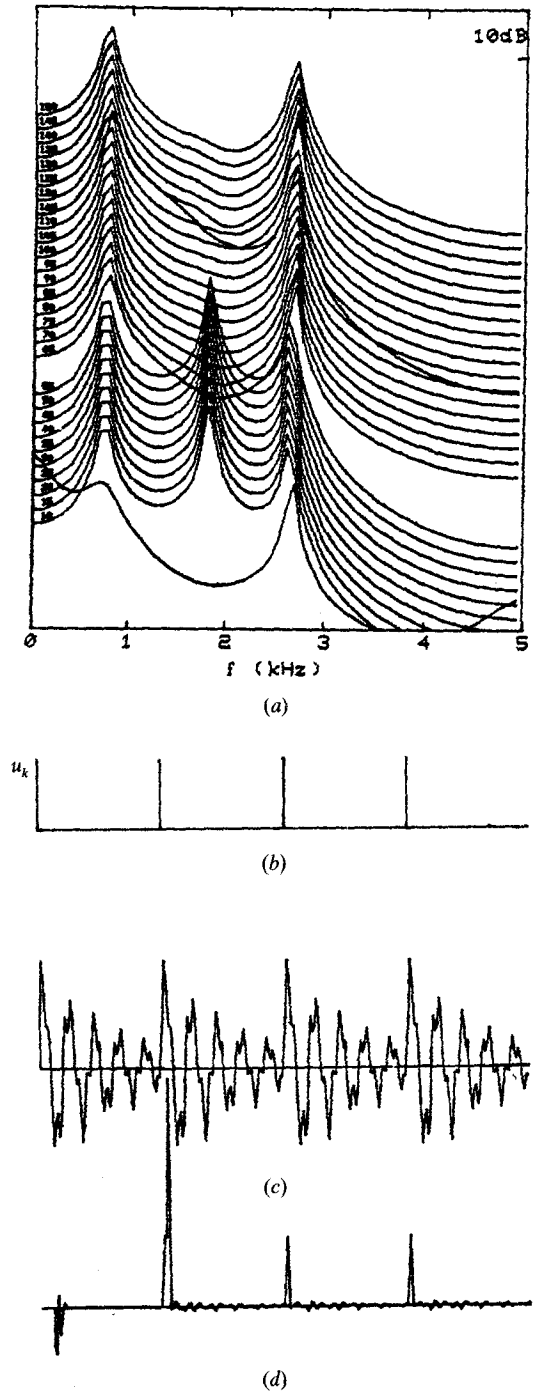


Fig. 5. — Résultats de l'estimation. Cas d'une excitation par un train d'impulsions. (a) Fonctions de transfert estimées, tracées tous les cinq pas de calcul. (b) Signal d'excitation ( $k \in [0, 255]$ ). (c) Signal de sortie du système. (d) Signal résiduel.

ne démarre que quelques points après le début de la fenêtre comprenant la première impulsion d'excitation. La raison de ce décalage est qu'il est nécessaire de remplir le vecteur de données.

Au début de la deuxième période, on constate l'apparition d'une impulsion relativement grande dans le signal résiduel due à la présence de l'excitation

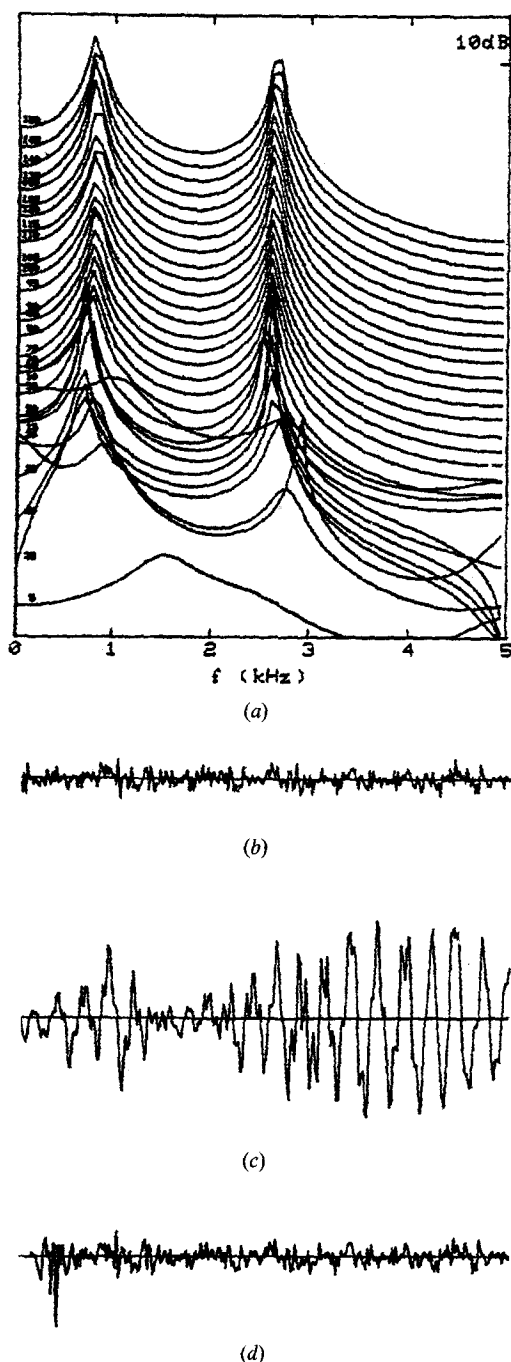


Fig. 6. — Résultats de l'estimation. Cas d'une excitation par un bruit blanc. (a), (b), (c), (d) : mêmes légendes que figure 5.

(fig. 5d). Cette « fausse » erreur de prédiction perturbe l'ajustement des paramètres. Par conséquent, l'estimation de la paire pôle-zéro n'est pas bonne.

On constate en outre que l'excitation par un bruit blanc donne des résultats plus médiocres que l'excitation par un train d'impulsions (fig. 6a). Sur les fonctions de transfert estimées, on ne voit pas apparaître un quelconque indice de la paire pôle-zéro. De plus la position des deux pôles fluctue légèrement au cours du temps.

Le signal d'entrée et le signal résiduel sont très ressemblants sauf pour les quelques dizaines de points initiaux (fig. 6b, 6d). Comme dans le cas de l'excitation impulsionnelle, c'est la présence du signal d'entrée dans le résidu qui diminue la précision de l'estimation. Mais la situation est pire dans ce cas du fait que la perturbation est permanente.

#### 4. La suppression d'impulsions dans le résidu

Nous avons vu que le résidu, contenant le signal d'entrée, agit de façon défavorable sur les performances de l'ajustement des paramètres. Malheureusement, sans information sur le signal d'entrée, la séparation dans le résidu de l'excitation et de l'erreur de prédiction est impossible.

Signalons toutefois que nous avons le choix entre deux alternatives pour traiter chaque échantillon du résidu. Soit on le considère entièrement comme l'erreur de prédiction et on l'introduit dans le mécanisme d'ajustement; soit on considère le résidu entièrement comme excitation et dans ce cas on ne l'introduit évidemment pas dans le mécanisme d'ajustement. Aucune des deux solutions n'est optimale. Cependant suivant la proportion des deux composantes du résidu, l'une des alternatives peut se révéler mieux adaptée.

Nous avons déjà vu, dans la section 3, que le choix constant de la première alternative peut provoquer une forte perturbation de l'ajustement et ainsi conduire à une mauvaise estimation. De même, on ne peut pas s'en tenir partout à la deuxième alternative, puisqu'on a besoin de l'erreur de prédiction pour l'ajustement. Une question se pose alors : est-il possible de passer à certains instants de la première solution à la seconde pour éviter une forte perturbation de l'ajustement ? La réponse est différente selon le type d'excitation employé.

##### 4.1. EXCITATION PAR UN BRUIT BLANC

Malgré les propriétés statistiquement simples d'un bruit blanc, il est impossible de prédire à un instant donné la valeur d'un échantillon. Ainsi, dans le cas d'une excitation par un bruit blanc, on n'a pas de moyen d'identifier le signal d'entrée dans le résidu et, par conséquent, il n'est pas possible de passer de la première alternative à la seconde.

##### 4.2. EXCITATION PAR UN TRAIN D'IMPULSIONS

Dans le cas d'un train d'impulsions, le résidu augmente brusquement à l'instant de l'excitation et la valeur atteinte est en général bien supérieure à l'écart type de l'erreur de prédiction sur une période. De toute évidence, à cet instant précis, la deuxième solution est beaucoup plus raisonnable que la première. Par ailleurs, les caractéristiques simples de l'excitation permettent de la détecter facilement dans le résidu.

Ainsi, la démarche suivante est envisageable : détecter d'abord l'instant de l'excitation impulsionnelle et passer à cet instant de la première alternative à la deuxième pour l'ajustement des paramètres.

La réalisation de cette démarche est présentée schématiquement à la figure 7. Compte tenu de l'amplitude de l'impulsion dans le résidu, un simple seuil suffit à l'identifier. A l'instant de cette détection, l'erreur de prédiction étant par définition nulle (puisque le résidu est considéré comme excitation), il n'y a donc pas d'ajustement des paramètres. Notons que l'actualisation de  $P_k$  ne sera pas influencée car elle est indépendante de l'erreur de prédiction.

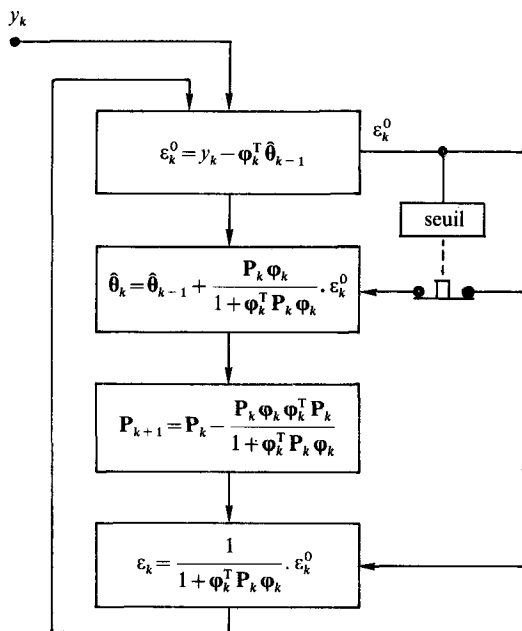


Fig. 7. — Schéma de l'algorithme proposé pour réaliser la suppression d'impulsions. Ici,

$$\hat{\theta}_k = [a_1, \dots, a_p, b_1, \dots, b_q]^T,$$

$$\Phi_k = [-y_{k-1}, \dots, -y_{k-p}, \varepsilon_{k-1}, \dots, \varepsilon_{k-q}],$$

$$\varepsilon_k^0 = y_k - \Phi_k^T \hat{\theta}_{k-1},$$

l'erreur *a priori*, et

$$\varepsilon_k = y_k - \Phi_k^T \hat{\theta}_k,$$

l'erreur *a posteriori* [13].

Du fait qu'on supprime l'impulsion du résidu à l'instant de l'excitation, on appellera simplement cette technique approche par suppression d'impulsions.

La figure 8 montre un résultat de simulation. Le signal de test est le même que celui utilisé dans la section précédente. L'influence du signal d'entrée étant éliminée, l'estimation atteint maintenant une bonne précision, notamment en ce qui concerne la paire pôle-zéro (fig. 8 a). Le résidu est très faible entre deux impulsions d'excitation successives (fig. 8 d).

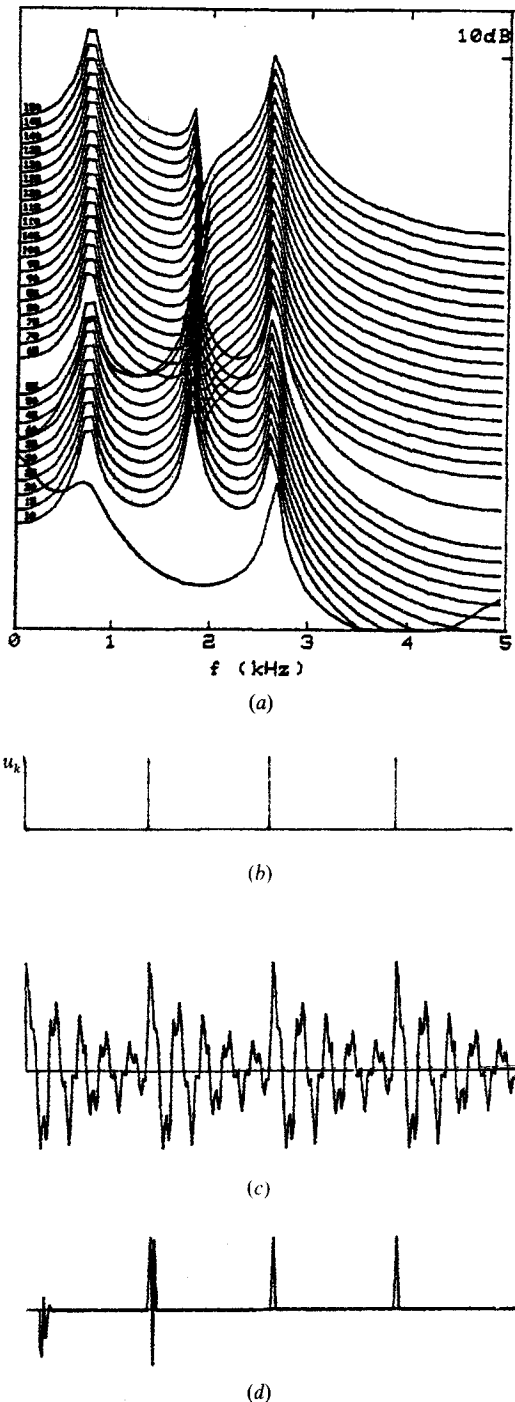


Fig. 8. — Résultats de l'estimation avec suppression d'impulsions. (a), (b), (c), (d) : mêmes légendes que figure 5.

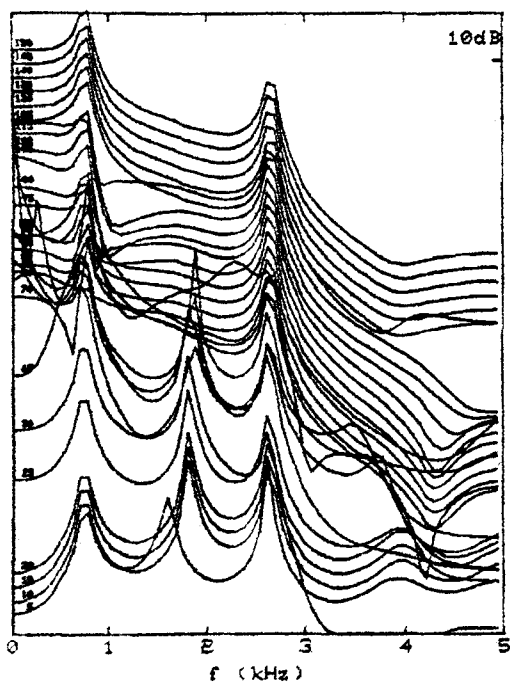
#### 4. 3. EXCITATION PAR UNE FORME D'ONDE COMPLEXE

Dans le cas plus général du signal de parole, l'excitation périodique présente une forme d'onde complexe. Il existe de plus une interaction entre l'excitation et le système. Cette dernière modifie de façon non linéaire la fonction de transfert du conduit vocal au cours d'une période.

La question se pose donc : l'approche utilisant la suppression d'impulsions reste-elle valable dans ce cas ?

Cette question revient à s'interroger sur la possibilité de trouver un filtre  $A(z)/B(z)$  pour lequel le résidu est proche d'un train d'impulsions. On sait que c'est

impossible avec l'analyse LPC (modèle AR) classique. Le filtre statique reflète uniquement les caractéristiques moyennes du signal dans la fenêtre d'analyse. De plus, il ne se rend pas compte des phases du spectre du signal. La forme du résidu ne sera donc pas celle d'un train d'impulsions. Pour ce type de résidu, l'approche par suppression d'impulsions est inapplicable.



(a)



(b)

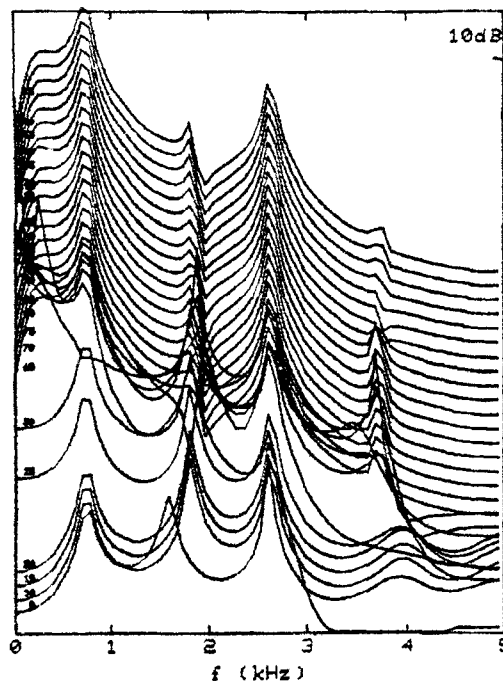


(c)



(d)

Fig. 9. — Résultats de l'estimation sans suppression d'impulsions. Cas d'une excitation par une forme d'onde glottale. Les bandes passantes de la fonction de transfert du système sont variables pendant la phase d'ouverture des cordes vocales [5]. (a), (b), (c), (d) : mêmes légendes que figure 5.



(a)



(b)



(c)



(d)

Fig. 10. — Résultats de l'estimation avec suppression d'impulsions. Mêmes excitation et système que figure 9. (a), (b), (c), (d) : mêmes légendes que figure 5.



Dans l'algorithme récursif, le modèle est plus souple : la mise à jour pas à pas des paramètres du filtre permet de suivre les variations au sein d'une période du signal; le filtre possède des zéros; les pôles ne sont pas contraints de se placer à l'intérieur du cercle unité. Nous pouvons espérer que le résidu sera plus proche d'un train d'impulsions et que la méthode décrite précédemment pourra être appliquée.

Un premier test a été réalisé sur un signal synthétique incorporant l'influence de la source d'excitation et du couplage source/conduit. Nous avons utilisé une forme d'excitation typique de la source glottale. Pour simuler le couplage, nous avons fait varier les bandes passantes des résonances proportionnellement à l'ouverture de la glotte [5].

En ce qui concerne le choix de l'ordre du modèle, il est évident que ce dernier doit être supérieur à l'ordre de la partie du système lié au conduit vocal, car le modèle doit pouvoir représenter également la source glottale. Dans notre signal synthétique, le modèle du conduit vocal comprend six pôles et deux zéros. En supposant qu'il faille au moins deux pôles pour modéliser la source, ainsi que deux pôles pour la variation spectrale, l'ordre de la partie AR a été fixé à 10, et l'ordre de la partie MA à 6.

Nous présentons les résultats obtenus sans suppression d'impulsions à la figure 9 et avec cette suppression à la figure 10.

Les principaux enseignements qui se dégagent sont les suivants :

La méthode par suppression d'impulsions dans le résidu semble efficace : on constate que la paire pôle-zéro est bien estimée (fig. 10 a). Comme prévu, le résidu ressemble bien à un train d'impulsions (fig. 10 d). Ceci résulte à l'évidence de la souplesse du modèle qui représente correctement l'ensemble du système (conduit vocal, source et couplage).

Par contre, sans cette suppression, la paire pôle-zéro ne peut être estimée (fig. 9 a). La perturbation du mécanisme d'ajustement interdit une bonne estimation des paramètres. Plusieurs valeurs élevées apparaissent pour chacune des périodes dans le signal résiduel (fig. 9 d).

#### 4. 4. ANALYSE D'UN SIGNAL DE PAROLE NATUREL

Nous avons effectué de nombreuses analyses sur des signaux naturels. Les résultats semblent malheureusement moins satisfaisants que dans le cas de signaux synthétiques. Nous présentons ici un exemple typique. Il s'agit d'une voyelle nasale [ã] prononcée par un locuteur masculin. 30 ms de signal au centre de cette voyelle ont été utilisées pour l'analyse. L'ordre du modèle a été fixé à 12 pour la partie AR et à 6 pour la partie MA.

Les résultats sont présentés aux figures 11 (sans suppression d'impulsions) et 12 (avec suppression).

On peut constater qu'une légère amélioration est obtenue par la suppression d'impulsions. Les fonctions de transfert estimées révèlent plus nettement un zéro vers

2,5 kHz, dès la deuxième période (fig. 12 a, d). Sur le signal résiduel (fig. 12 c), on voit également qu'après quelques valeurs élevées dans cette deuxième période, le résidu diminue par rapport à ce que l'on observe sans suppression d'impulsions (fig. 11 c).

Mais, l'effet de la suppression d'impulsions est moins net que précédemment : le résidu ne tend pas vraiment vers un train d'impulsions (fig. 12 c). Ceci limite évidemment l'accroissement de la précision de l'estimation.

Ces améliorations moins manifestes peuvent s'expliquer par les raisons suivantes :

L'hypothèse principale de cette approche étant que le résidu doit être proche d'un train d'impulsions, l'adéquation du modèle et du système doit être très bonne. Dans le cas de signal de parole, un filtre ARMA semble insuffisant pour satisfaire cette exigence. L'efficacité de cette méthode décroît rapidement lorsque le modèle s'éloigne du système.

Le signal synthétique, bien que prenant en compte l'influence de la source glottale et du couplage, est encore trop simple par rapport à la complexité de la parole naturelle, où la forme de l'onde glottale est très variable et où le couplage s'effectue de manière complexe. De plus le conduit vocal ne peut être véritablement représenté par un filtre ARMA, car il existe des facteurs souvent négligés et difficiles à modéliser précisément comme, par exemple, le rayonnement aux lèvres.

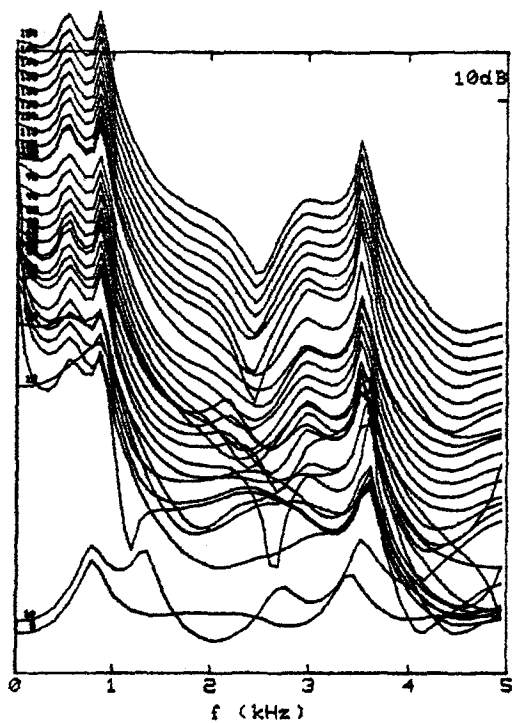
## 5. Conclusion

Nous avons étudié la précision de l'estimation dans la modélisation du spectre de parole par un modèle ARMA récursif, notamment les limites de celle-ci lorsqu'on effectue l'estimation sans information sur le signal d'excitation.

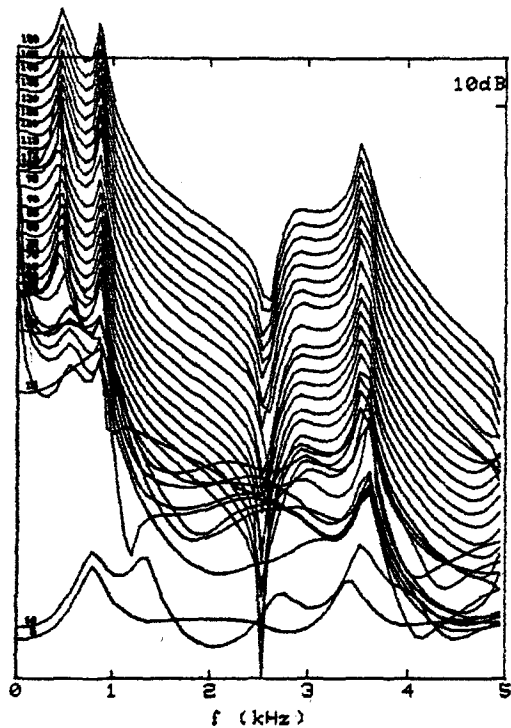
Après avoir montré que la présence d'une excitation inconnue dans le résidu perturbe l'ajustement des paramètres et limite la précision de l'estimation, nous avons proposé une méthode par suppression d'impulsions pour éliminer l'influence du signal d'excitation dans le résidu. Cette méthode a prouvé son efficacité dans le cas où le système est véritablement excité par un train d'impulsions. Pour la parole synthétisée incorporant une excitation glottale et le couplage source/conduit vocal, cette méthode reste valable. En fait, dans tous les cas où le signal résiduel est proche d'un train d'impulsions, on peut utiliser cette méthode pour améliorer l'estimation des paramètres.

L'efficacité de cette approche semble limitée pour la parole naturelle. Ceci est dû au fait que l'adéquation du modèle au signal de parole est loin d'être parfaite pour en assurer une analyse très fine. La complexité du signal de parole semble exiger une modélisation encore plus élaborée qui tient compte des effets non linéaires.

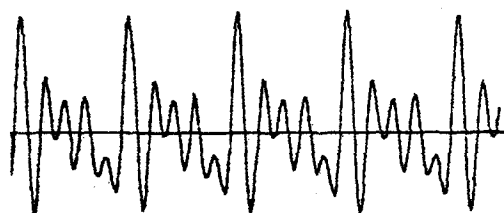
*Manuscrit reçu le 18 janvier 1986*



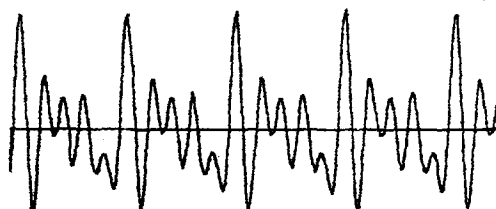
(a)



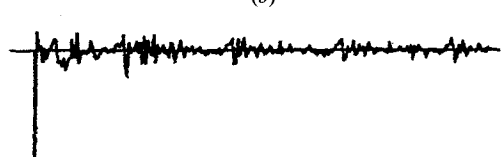
(a)



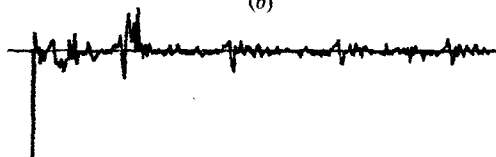
(b)



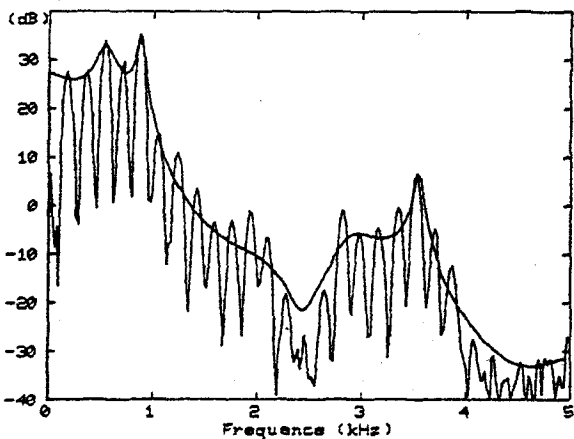
(b)



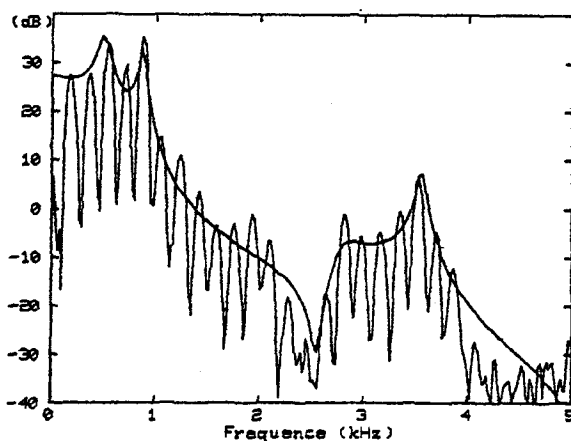
(c)



(c)



(d)



(d)

Fig. 11. — Résultats de l'estimation sans suppression d'impulsions. Cas d'un signal de parole naturelle : voyelle [a] prononcée par un locuteur masculin. (a) Spectre estimé, tracé tous les cinq pas de calcul. (b) Signal de parole (25,6 ms). (c) Signal résiduel. (d) Spectre estimé au 140<sup>e</sup> pas de calcul; et spectre FFT du signal.

Fig. 12. — Résultats de l'estimation avec suppression d'impulsions. Même signal que figure 11. (a), (b), (c), (d) : mêmes légendes que figure 11.

## Remerciements

Tous nos remerciements à M<sup>mes</sup> C. Latombe, D. Pascal, MM. F. Lonchamp, S. Maeda pour leurs discussions fructueuses et leurs contributions à la rédaction de cet article.

## BIBLIOGRAPHIE

- [1] K. J. ASTROM et P. EYKHOFF, System identification. A survey, *Automatica*, 7, 1971, p. 123-162.
- [2] B. S. ATAL et M. R. SCHROEDER, Linear prediction analysis of speech based on a pole-zero representation, *J. Acoust. Soc. Am.*, 64, n° 5, 1978, p. 1310-1318.
- [3] J. A. CADZOW, High performance spectral estimation. A new ARMA method, *IEEE Trans. on ASSP*, ASSP-28, n° 5, 1980, p. 524-529.
- [4] G. FANT, *Acoustic theory of speech production*, 1960, Mouton The Hague.
- [5] G. FENG, Modélisation acoustique et traitement du signal de parole. Le cas des voyelles nasales, *Thèse de Docteur*, INP, Grenoble (arrêté ministériel du 5 juillet 1984), 1986.
- [6] J. L. FLANAGAN, *Speech analysis synthesis and perception*, 2<sup>e</sup> éd., 1972, Springer-Verlag, Berlin, Heidelberg, New York.
- [7] R. E. KALMAN, Design of a self-optimizing control system, *Trans. AMSE*, 80, 1958, p. 468-478.
- [8] J. L. LACOUME, *Nouvelle méthodes d'analyse spectrale*, à paraître dans *Méthodes et moyens de traitement du signal. Application aux mesures physiques*, 4<sup>e</sup> éd., J. MAX, éd.
- [9] J. L. LACOUME, C. LATOMBE, N. MARTIN, M. GHARBI et R. LIDIN, AR and ARMA models in spectral and multispectral analysis, *IEEE ASSP Spectrum Estimation Workshop II*, Florida, nov. 1983.
- [10] I. D. LANDAU, Unbiased recursive identification using model reference adaptative techniques, *IEEE Trans. on AC*, AC-21, n° 2, 1976, p. 194-202.
- [11] Y. MIYANAGA, N. MIKI, N. NAGAI et K. HATORI, A speech analysis algorithm which eliminates the influence of pitch using the model reference adaptative system, *IEEE Trans. on ASSP*, ASSP-30, n° 1, 1982, p. 88-96.
- [12] H. MORIKAWA et H. FUJISAKI, Adaptative analysis of speech based on on a pole-zero representation, *IEEE Trans. on ASSP*, ASSP-30, n° 1, 1982, p. 77-87.
- [13] V. PANUSKA, An adaptative least squares identification method, *Proc. IEEE Symp. Adaptative Process.*, 1969.
- [14] T. SODERSTROM, L. LJUNG et I. GUSTAVSSON, A theoretical analysis of recursive identification methods, *Automatica*, 14, 1978, p. 231-244.