

IDENTIFICATION DES CIBLES SPECTRALES PAR MODELES  
A EXCITATION MULTI-ECHELON

Alvaro DE LIMA-VEIGA, Yves GRENIER

Ecole Nationale des Télécommunications, Département Signal.  
46, rue Barrault, 75634 Paris, Cedex 13

### RÉSUMÉ

Récemment [1] un nouveau modèle pour représenter l'évolution temporelle des vecteurs d'analyse LPC a été proposé. Dans cette approche ceux-ci sont vus comme la sortie d'un modèle autorégressif vectoriel excité par une fonction constante par morceaux. Les parties constantes de la fonction d'excitation correspondent à des cibles spectrales vers lesquelles s'oriente à chaque instant la sortie du modèle. La fonction d'excitation détermine donc une segmentation du signal (donnée par les localisations des sauts) et caractérise les segments par les cibles spectrales qui leur sont associées. Cet article décrit une nouvelle méthode de localisation des sauts de la fonction d'excitation. Nous proposons une procédure qui place des paires de sauts symétriques et qui prend en compte l'ensemble de sauts déjà placés. La méthode est appliquée à un signal de parole issu de l'analyse d'un corpus constitué de noms de famille épelés en continu (extraits de la base de données BDSONS).

### SUMMARY

Recently [1] a new model to represent the evolution of the LPC analysis has been proposed. In this approach, they are seen as the output of a vector autoregressive model excited by a piece-wise constant function. The constant parts of the excitation function are associated to spectral targets to which the output of the model tends at each instant. The jumps in the excitation function determine a segmentation of the signal and characterize each segment by the corresponding spectral target. This article extends this previous work by describing a new jump location method. We place the jumps in the excitation function by couples and we take into account the jumps already localized. The method has been applied to a speech signal from the data base BDSONS.

## 1. INTRODUCTION

Les trajectoires décrites par les vecteurs de paramètres issus de l'analyse LPC ont une importance fondamentale dans le domaine du traitement de la parole. Non seulement elles contiennent une description de l'évolution temporelle du contenu spectral du signal de parole mais elles sont aussi directement connectées aux mouvements articulatoires exécutés par le conduit vocal. Ainsi, une représentation simple et acoustiquement significative de l'évolution temporelle des paramètres d'analyse aurait de nombreuses applications en transmission, synthèse et reconnaissance de la parole. Malgré son importance, ce problème a été l'objet d'un nombre assez réduit de travaux ([2],[3],[4]).

Récemment [1], nous avons proposé un nouveau modèle où les trajectoires des vecteurs d'analyse sont représentées comme la sortie d'un modèle autorégressif vectoriel excitée par une fonction (vectorielle) constante par morceaux. Dans cette approche, les parties constantes correspondent à des cibles articulatoires que devrait atteindre la sortie du modèle. En effet, si on excite un modèle stable avec une entrée constante, la sortie du modèle tendra vers la valeur de l'entrée à un facteur près, représenté par le gain du modèle. La présence d'un saut dans la fonction d'excitation, par contre, indique un changement de cible à atteindre et, par

conséquent, une réorientation des mouvements articulatoires. D'autre part, comme un saut dans la fonction d'excitation peut intervenir avant que la cible courante ne soit atteinte, on espère pouvoir traiter le problème de la coarticulation.

Cet article présente un rappel des éléments de ce modèle et propose une nouvelle méthode de localisation des sauts de la fonction d'excitation. Dans la section 2, nous présentons une description du modèle. Nous passons ensuite au problème de son identification qui se réalise en deux étapes. D'abord, on doit identifier le modèle autorégressif vectoriel. La section 3 décrit deux méthodes d'estimation et discute leur rôle dans l'algorithme. Ensuite, nous traitons le problème de l'estimation de l'excitation du modèle. Ceci est un double problème comprenant la localisation des sauts de la fonction d'excitation et l'estimation de leurs amplitudes. La section 4.1 décrit la nouvelle méthode (approchée) de localisation, la section 4.2 décrit une procédure pour améliorer cette solution et la section 4.3 fait un rappel du calcul des amplitudes des sauts.



$$J = \sum_{n=1}^N (\beta_n - \hat{\beta}_n)^T (\beta_n - \hat{\beta}_n) \quad (9)$$

où  $\hat{\beta}_n$  est la sortie du système défini dans les équations (1) et (2) quand il est excité par la fonction  $u_n$  estimée.

La minimisation de ce critère quand on considère un nombre  $L$  de sauts plus petit que la longueur  $N$  des trajectoires est un problème complexe car le nombre de combinaisons à tester peut être énorme. Pour réduire la complexité, les sauts sont rajoutés à la fonction d'excitation un à la fois, après quoi toutes les amplitudes doivent être réestimées. Le nombre de sauts sera déterminé implicitement, en imposant un seuil à l'erreur de reconstruction.

#### 4.1 Localisation des sauts - Méthode du placement double

Étant donné un ensemble de  $l$  localisations  $n_1, \dots, n_l$  et les amplitudes correspondantes  $v_{n_1}, \dots, v_{n_l}$ , nous voulons savoir où placer un nouveau saut. Dans un article récent [1], nous avons montré comment adapter la méthode multi-impulsionnelle d'Atal [6] pour accomplir cette tâche. Dans cette section, nous présentons une nouvelle méthode de localisation où on cherche à placer une paire de sauts à amplitudes symétriques. Pour ne pas tester toutes les combinaisons, on considère seulement les paires dont un des sauts coïncide avec un saut déjà existant. Ceci est aussi une manière de prendre en compte les localisations des sauts déjà placés dans la fonction d'excitation. Au départ, quand aucun saut n'a pas encore été placé, nous considérons que le deuxième saut est placé sur une localisation virtuelle  $N+1$ , après la fin des trajectoires. Ensuite, pour localiser chaque nouveau saut, on examine les paires de localisations dont la deuxième coïncide avec l'un des sauts déjà placés (la localisation virtuelle  $N+1$  comprise). Aussi, nous examinons uniquement des paires de localisations entre lesquelles il n'y a pas d'autres sauts.

La résolution exacte du problème de minimisation posé par ce placement double serait trop coûteux car le nombre de matrices à évaluer (et à retenir) augmente avec le carré du nombre  $N$  de vecteurs d'analyse. Nous considérons ici une approximation où l'ordre  $p$  du modèle AR est nul. La solution à ce problème simplifié est donnée par:

$$n_{l+1} = \max_S \left\{ \frac{1}{n_i - s} \sum_{n=s}^{n_i-1} (\beta_n^l)^T \sum_{n=s}^{n_i-1} \beta_n^l, n_i < s < n_{i+1}, i=0, \dots, l \right\} \quad (10)$$

où les limites des intervalles extrêmes sont notées, par simplicité,  $n_{l+1} = N+1$  et  $n_0 = 0$ .

#### 4.2 Remplacement des localisations

La méthode de localisation que nous proposons construit la fonction d'excitation en y rajoutant un saut à la fois. Il est donc évident que la localisation d'un saut ne peut pas prendre en compte les localisations de ceux qui n'ont pas encore été placés. Par conséquent, il serait utile de réviser l'ensemble des localisations une fois qu'il a été déterminé.

Soit  $S = \{ n_1, \dots, n_L \}$  l'ensemble des localisations et soit  $j, 1 \leq j \leq L$ , la localisation à réexaminer. Nous allons supposer par simplicité que l'ensemble de localisations est ordonné, c.a.d., que  $n_1 < n_2 < \dots < n_L$ . Alors, pour chaque

localisation, nous définissons un ensemble réduit de localisations:

$$S_j = S - \{ n_j \}$$

Nous sélectionnons une nouvelle localisation (qui peut éventuellement être la même que celle qui a été enlevée) en examinant les trajectoires résiduelles associées à l'ensemble réduit, ce qui produit un nouveau ensemble  $S$ . Ceci est répété pour  $j = 1, \dots, L$  permettant que chaque localisation puisse être revue.

#### 4.3 Estimation des amplitudes

Après l'inclusion de chaque élément dans l'ensemble des localisations on doit recalculer les amplitudes de tout l'ensemble. Une solution approchée qui préserve les caractéristiques locales des trajectoires et qui donne des résultats satisfaisants en pratique peut être obtenue en évaluant les amplitudes par rapport à une mesure localisée qui considère l'erreur sur le segment compris entre deux sauts uniquement, duquel on a enlevé l'influence des sauts précédents. Supposons que  $n_1 < n_2 < \dots < n_l$ . Alors, l'amplitude du  $j^{\text{ème}}$  saut sera calculée de façon à minimiser:

$$J_j = \sum_{n=n_j}^{n_{j+1}-1} (\beta_n^{j-1} - \hat{\beta}_n(n_j))^T (\beta_n^{j-1} - \hat{\beta}_n(n_j))$$

où  $n_{l+1} = N+1$ ,  $\hat{\beta}_n(n_j)$  est la réponse du système à une fonction avec un seul saut sur  $n_j$  et  $\beta_n^l$  est donné par l'équation (6). Après quelques manipulations, on obtient l'expression de l'amplitude optimale,

$$\hat{v}_{n_j} = [R_{N-n_{j+1}+n_j}]^{-1} \sum_{n=n_j}^{n_{j+1}-1} H_{n-n_j}^T \beta_n^{(j-1)}, R_s = \sum_{n=s}^N H_{n-s}^T H_{n-s}$$

Nottons que seules les amplitudes qui sont influencées par le nouveau saut seront recalculées, c.a.d., celles qui se trouvent après ou immédiatement avant le nouveau saut. Cela peut être aussi vérifié en observant que  $\beta_n^j$  est égal à  $\beta_n^{j-1}$  pour  $n < n_{j-1}$ .

### 5. RESUME DE LA METHODE

Initialement, la fonction d'excitation étant inconnue, on utilise la méthode approchée décrite dans la section 3.1 pour évaluer les matrices  $A_1, \dots, A_p$  indépendamment de  $u_n$ .

**Calcul de la fonction d'excitation  $u_n$ :** La fonction d'excitation est définie par l'ensemble des localisations des sauts  $n_1, \dots, n_L$  et leurs amplitudes respectives  $v_{n_1}, \dots, v_{n_L}$ . Ces deux éléments sont déterminés progressivement en rajoutant à la fonction d'excitation un saut à la fois. La localisation du saut se fait par l'application du critère exprimé dans l'équation (10). Après la détermination de chaque saut, les amplitudes sont calculées par la procédure décrite dans la section 4.3. Cela se fait jusqu'à ce que l'erreur de reconstruction donnée par l'équation (9) soit inférieure à un seuil prédéterminé. Finalement, avant de passer à la phase suivante, on applique la procédure de révision des localisations décrite dans la section 4.2.

**Affinement de la solution:** Après avoir déterminé une première fois la fonction d'excitation nous pouvons l'utiliser

## 2. DESCRIPTION DU MODELE

Soit  $\beta_n, n=1, \dots, N$  une suite de vecteurs contenant les paramètres obtenus par une analyse LPC d'ordre  $k$ . Nous représentons l'évolution temporelle du vecteur  $\beta_n$  comme la sortie d'un modèle autorégressif vectoriel, donnée par:

$$\beta_n = A_1 \beta_{n-1} + \dots + A_p \beta_{n-p} + u_n \quad (1)$$

où  $A_1, \dots, A_p$  sont des matrices de dimension  $k \times k$  et  $u_n, n=1, \dots, N$  est une fonction excitatrice vectorielle de dimension  $k$ .

L'utilisation de modèles autorégressifs vectoriels dans des contextes proches au nôtre a déjà été proposée par d'autres auteurs ([3], [4]). La nouveauté de notre approche repose sur la forme particulière que nous attribuons à l'excitation  $u_n$  qui est une fonction vectorielle constante par morceaux. Soit  $v_n$  l'amplitude d'un saut placé à l'instant  $n$ . Alors,  $u_n$  peut être exprimé par l'équation aux différences suivante:

$$u_n = u_{n-1} + v_n \quad (2)$$

où  $v_n$  est nul presque partout à l'exception des instants où il y a un changement de cible spectrale. Soit  $L$  le nombre de changements. Alors,  $v_n$  sera non-nulle seulement pour  $n = n_1, \dots, n_L$ .

Le modèle décrit par les équations (1) et (2) peut être exprimé directement en fonction des amplitudes des sauts  $v_{n_1}, \dots, v_{n_L}$ :

$$\beta_n = \sum_{i=1}^L H_{n-n_i} v_{n_i} \quad (3)$$

où  $H_n$  représente la réponse à l'échelon du système défini dans l'équation (1). Si les matrices  $A_1, \dots, A_p$  caractérisent un système stable alors  $H_n$  tend vers une valeur  $H_\infty$ , donnée par:

$$H_\infty = (I - A_1 - \dots - A_p)^{-1} \quad (4)$$

La valeur de la cible  $\beta$  vers laquelle tend la sortie du système quand celui-ci est excité par une entrée constante de valeur  $u$ , est donc donnée par:

$$\beta = H_\infty u \quad (5)$$

Nous définissons aussi  $\beta_n^l$  comme le vecteur des trajectoires originales soustrait de l'effet dû aux sauts 1 à  $l$ ;

$$\beta_n^l = \beta_n - \sum_{i=1}^l H_{n-n_i} v_{n_i} \quad (6)$$

où, bien entendu,  $\beta_n^0$  coïncide avec les trajectoires originales  $\beta_n$ .

Le modèle décrit dans les équations (1) et (2) est caractérisé par deux ensembles de valeurs. D'abord, il y a les coefficients matriciels  $A_1, \dots, A_p$ . Ces matrices sont inconnues et doivent être estimées. Ensuite, on doit déterminer la fonction d'excitation  $u_n$ , caractérisée par le nombre des sauts  $L$ , leurs localisations  $n_1, \dots, n_L$  et leurs amplitudes  $v_{n_1}, \dots, v_{n_L}$ . Leur détermination se fait par un algorithme itératif dont nous décrivons les éléments dans la suite.

## 3. IDENTIFICATION DU MODELE AR VECTORIEL

Nous examinons deux procédés d'identification. Nous considérons d'abord l'estimation de  $A_1, \dots, A_p$  indépendamment de l'entrée  $u_n$ . Il s'agit, bien sûr, d'une méthode approchée mais qui nous permettra d'estimer l'entrée  $u_n$  une première fois, laquelle sera utilisée dans la deuxième méthode (exacte) pour obtenir des nouvelles matrices  $A_1, \dots, A_p$ .

### 3.1 Méthode Approchée

Soit  $\bar{\beta}_n, n=1, \dots, N$  une suite de vecteurs contenant les trajectoires différenciées donné par  $\bar{\beta}_n = \beta_n - \beta_{n-1}$ . Alors, par les équations (1) et (2) nous avons:

$$\bar{\beta}_n = A_1 \bar{\beta}_{n-1} + \dots + A_p \bar{\beta}_{n-p} + v_n \quad (7)$$

L'expression ci-dessus dépend uniquement de  $v_n$  qui doit présenter un nombre relativement faible de valeurs non-nulles. L'estimation peut donc être effectuée indépendamment de  $v_n$  sans trop de problème. Les matrices  $A_1, \dots, A_p$  sont calculées de façon à minimiser:

$$\sum_{n=1}^N (\bar{\beta}_n - A_1 \bar{\beta}_{n-1} - \dots - A_p \bar{\beta}_{n-p})^T (\bar{\beta}_n - A_1 \bar{\beta}_{n-1} - \dots - A_p \bar{\beta}_{n-p})$$

ce qui est réalisé par une adaptation au cas vectoriel de l'algorithme de Levinson [5].

### 3.2 Méthode Exacte

Supposons connue l'entrée  $u_n$ . Alors, nous calculons les valeurs de  $A_1, \dots, A_p$  qui minimisent la trace de la matrice de covariance de l'erreur de prédiction, donnée par:

$$\sum_{n=1}^N (\beta_n - A_1 \beta_{n-1} - \dots - A_p \beta_{n-p} - u_n)^T (\beta_n - A_1 \beta_{n-1} - \dots - A_p \beta_{n-p} - u_n)$$

La minimisation de ce critère donne:

$$\begin{bmatrix} A_1, \dots, A_p \end{bmatrix} = \begin{bmatrix} U_1, \dots, U_p \end{bmatrix} \begin{bmatrix} B_0 & B_{-i} & B_{-p+1} \\ & \cdot & \\ & & B_{-i} \\ B_{p-1} & & B_0 \end{bmatrix}^{-1} \quad (8)$$

$$B_i = \begin{cases} \sum_{n=i}^{N-i} \beta_n \beta_{n+i}^T & \text{pour } i \geq 0 \\ \sum_{n=-i+1}^N \beta_n \beta_{n+i}^T & \text{pour } i < 0 \end{cases} \quad \text{et } U_i = \sum_{n=i+1}^N (\beta_n - u_n) \beta_{n-i}^T$$

Notons que seule la partie droite de l'équation (8) dépend de l'entrée  $u_n$ . Ainsi, une fois l'équation résolue pour une valeur particulière de l'entrée, on peut facilement obtenir de nouveaux modèles pour d'autres valeurs de l'entrée sans aucune inversion de matrice supplémentaire.

## 4. LOCALISATION/ESTIMATION DE SAUTS

Après avoir estimé les matrices  $A_1, \dots, A_p$  nous devons estimer l'entrée  $u_n$  de manière à minimiser une mesure de distance entre les trajectoires originales et reconstruites. La mesure que nous adoptons ici est l'erreur quadratique moyenne de reconstruction, donnée par:



pour recalculer les matrices  $A_1, \dots, A_p$  par la méthode exacte (eq.(8)). Alors, on recalcule les amplitudes et on applique la procédure de remplacement des localisations (section 4.2) et ainsi de suite jusqu'à convergence.

## 6. APPLICATION

La méthode a été appliquée à un signal de parole de la base de données BDSONS. Il s'agit d'un nom de famille épilé en continu /M-O-R-I-N/ échantillonné à 8KHz. Nous avons réalisé une analyse LPC d'ordre 12 sur 114 fenêtres de 20ms décalées de 10ms dont la figure 1 montre les 2 premiers coefficients LAR (log-area ratio). La même figure montre en pointillé les LAR obtenus par la sortie du système dans les équations (1) et (2) avec ordre 2 et excité par une fonction présentant 10 sauts. Le modèle a été identifié avec 5 itérations de la procédure de raffinement. L'erreur percentuelle est de 7.4% (par rapport à  $\sum_n \beta_n^T \beta_n$ ).

La figure 2.a montre les reliefs obtenues à partir des LAR originaux. Ensuite, nous montrons les cibles spectrales obtenues par l'équation (5) (fig. 2.b) et le relief reconstitué donné par la sortie du modèle (fig. 2.c). La représentation que nous utilisons est celle d'un sonogramme où les zones plus foncées correspondent à celles de plus grande énergie. L'erreur quadratique moyenne en dB prise sur les temps et sur les fréquences est de 3.3dB.

## 7. CONCLUSIONS

Cet article a décrit une nouvelle méthode de localisation de sauts dans la fonction d'excitation du modèle à cibles, proposé récemment [1]. Nous avons présenté une application sur un signal de parole qui met en évidence la possibilité d'utilisation de ce type de modèle dans plusieurs domaines du traitement de la parole. Nous envisageons surtout son utilisation en synthèse et reconnaissance, profitant ainsi de la description condensée donnée par les cibles spectrales obtenues par le modèle.

### REFERENCES:

- [1] DE LIMA-VEIGA A., GRENIER Y. "A Multi-step Excited Model for Speech Parameters Trajectoires". Proc. ICASSP, pp. 67-70, 1988.
- [2] ATAL B.S. "Efficient Coding of LPC Parameters by Temporal Decomposition". Proc. ICASSP-83, Boston, pp. 81-84, 1983.
- [3] GRENIER Y. "Utilisation de la Prédiction Linéaire en Reconnaissance et Adaptation au Locuteur". Rapport ENST-80008, 1980.
- [4] KAKUSHO O., YANAGIDA M. "Hierarchical AR Model for Time Varying Speech Signals". Proc. ICASSP-82, pp. 1295-1298, 1982.
- [5] WHITTLE P. "On the Fitting of Multivariate Autoregressions and the Approximate Canonical Factorization of a Spectral Density Matrix". Biometrika, vol. 50, pp. 129-134, 1963.
- [6] ATAL B.S., REMDE J.R. "A New Model of LPC Excitation for Producing Natural-sounding Speech at Low Bit Rates". Proc. ICASSP-82, pp. 614-617, 1982.

FIGURE 1: 2 LAR, ORIGINAL/RECONSTITUE

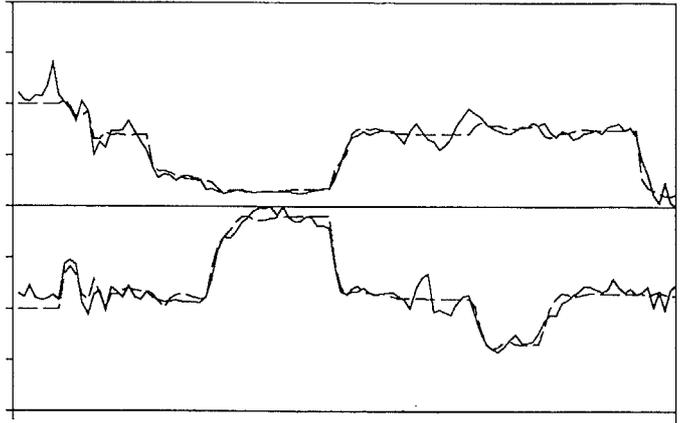
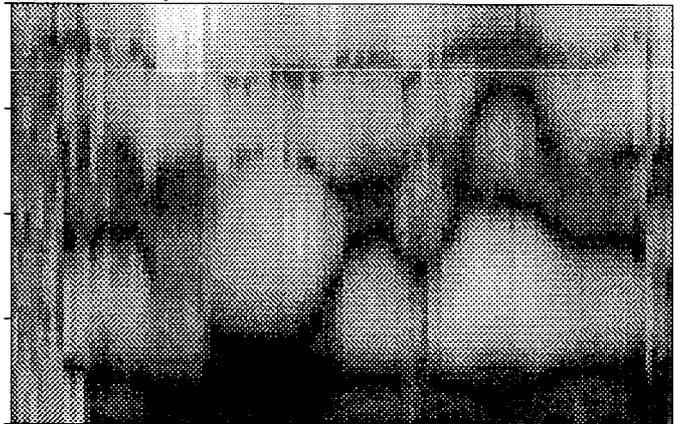
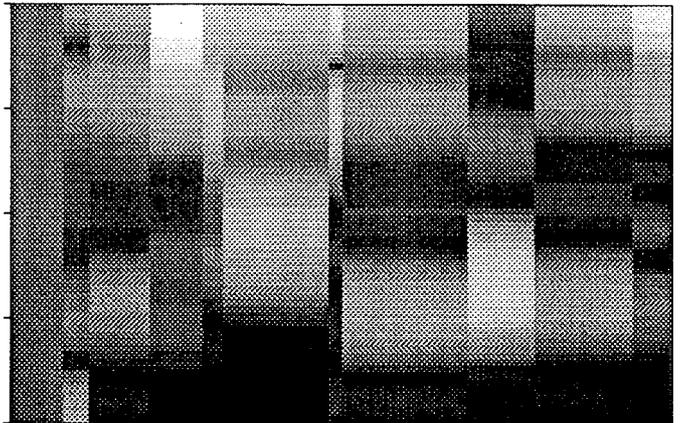


FIGURE 2:

(a) Relief original



(b) Spectres des cibles L=10



(c) Relief reconstitue, DIST. SPEC. 3.3dB

