MODÈLES ADAPTATIFS POUR LA DÉTECTION AUTOMATIQUE DE RUPTURES DANS LE SIGNAL DE PAROLE

N. ACHAB et G. FENG

Institut de la Communication Parlée
U.R.A.- CNRS N° 368 ; I.N.P.G.- Université Stendhal
BP 25X Domaine Universitaire - 38040 Grenoble Cedex, France

RÉSUMÉ

La segmentation de la parole basée sur la détection automatique de ruptures conduit à des résultats peu dépendants du locuteur. Le test de divergence présente toutefois une dissymétrie qui entraîne des omissions de détection. La méthode *forward-backward*, qui consiste à segmenter le signal dans les deux sens, résout ce problème mais empêche toute application en temps réel. Dans une étude précédente, nous avons proposé l'utilisation des modèles adaptatifs permettant de réduire considérablement la dissymétrie du test. Nous présentons dans cette communication un nouvel algorithme de segmentation qui met en œuvre à la fois le test du rapport de vraisemblance et la modélisation adaptative. La procédure de détection est simplifiée et l'algorithme peut parfaitement être utilisé dans les applications en ligne. L'évaluation montre des résultats tout à fait satisfaisants.

ABSTRACT

Speech segmentation based on automatic detection of abrupt changes allows speaker-independant results. However, the asymmetrical behavior of the divergence test often causes failures in transition detection. The forward-backward method, which consistes in segmenting signals in two directions, offers a good sulution to the problem but makes also on-line segmentation impossible. In a previous study, we have proposed the use of adaptive models which can efficiently reduce the asymmetrical behavior of the test. We present in this communication an on-line segmentation algorithm using both the likelihood ratio test and adaptive models. The detection procedure is simple and efficient. Evaluation has shown very satisfactory results, compared with classical methods.

I. INTRODUCTION

La segmentation du signal de parole en parties homogènes peut être considérée comme un problème de détection de changements (ou ruptures) dans ses caractéristiques spectrales. La surveillance des changements de paramètres des modèles autorégressifs pour la détection séquentielle de ces ruptures, conduit à des méthodes très efficaces et relativement faciles à mettre en œuvre.

Pour réaliser une segmentation de la parole peu dépendante du locuteur, André-Obrecht a appliqué la détection automatique de ruptures basée sur une nouvelle méthode statistique, développée par Basseville et Benveniste [4], qui consiste à utiliser deux modèles autorégressifs. Cela permet de prendre en compte aussi bien l'information globale (modèle long terme), que l'information locale (modèle court terme) contenue dans le signal

Deux distances peuvent être utilisées pour comparer les modèles et décider d'une rupture. Elles sont de type somme cumulée et basées sur le test d'hypothèses suivantes :

H_0 : il n'y a pas de rupture, les observations suivent le modèle long terme ;

H_1 : il existe une rupture à un instant r. Jusqu'à cet instant les observations suivent le modèle long terme, et au delà (jusqu'à l'instant courant), elles suivent le modèle court terme.

Ces deux distances sont le rapport de vraisemblance *a posteriori* entre les deux lois associées aux modèles et le test de divergence. Dans le cas simple des résidus gaussiens les expressions respectives des incréments des deux tests sont :

- Test de vraisemblance :

$$T_{1,n} = \frac{1}{2} \text{Log} \frac{\sigma_0^2}{\sigma_1^2} + \frac{(e_n^0)^2}{2\sigma_0^2} - \frac{(e_n^1)^2}{2\sigma_1^2} \qquad (1)$$

- Test de divergence :

$$T_{2,n} = \frac{1}{2} \left(2 \frac{e_n^0 e_n^1}{\sigma_1^2} - \left(1 + \frac{\sigma_0^2}{\sigma_1^2}\right) \frac{(e_n^0)^2}{\sigma_0^2} + \left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)\right) \qquad (2)$$

$e_n^i, i = 0,1$: résidus des modèles long et court terme ;

$\sigma_i^2, i = 0,1$: variance des résidus.

La somme cumulée des incréments de ces deux tests oscille autour de zéro avant une rupture, et elle a une dérive négative après la rupture. La règle d'arrêt de Hinkley permet de réduire le retard à la détection, et de mieux estimer l'instant de rupture. Elle consiste à ajouter à chaque incrément un biais fixé *a priori*, l'instant de rupture correspond alors au maximum local de la statistique.

Le test de divergence n'est pas symétrique : il peut réagir lors d'une transition entre deux zones stationnaires consécutives, et il peut rester insensible lorsque leur position est inversée. Cela entraîne des omissions de détection. Pour palier à cet inconvénient André-Obrecht a proposé une solution dans laquelle le signal est segmenté dans les deux sens (méthode *forward-backward*) [2][3]. Cette solution a permis de résoudre le problème de dissymétrie mais elle empêche toute application en temps réel.

Dans l'algorithme de André-Obrecht, le modèle long terme est identifié séquentiellement par la méthode de Burg sur une fenêtre croissante, et le modèle court terme par la méthode d'autocorrélation sur une fenêtre glissante (20 ms).

Nous avons mis en évidence, dans une précédente étude [1,5], la relation étroite existant entre la modélisation du signal et la dissymétrie du test de divergence. Avec une modélisation adéquate : le modèle long terme identifié séquentiellement par la méthode de Burg avec un coefficient d'oubli et le modèle court terme identifié séquentiellement par l'algorithme *prewindow* [6], la dissymétrie du test de divergence est réduite dans une grande proportion. La détection de ruptures peut donc se faire sans avoir recours à la méthode *backward*.

La procédure de détection adoptée est la suivante : aux variations spectrales importantes correspondent toujours des changements de pente de la statistique, calculée de manière continue (sans réinitialisation des modèles après les ruptures). Pour détecter ces changements, la statistique est d'abord linéarisée afin de masquer les faibles variations, une rupture est décidée si la différence entre les pentes de deux segments consécutifs est supérieure à un seuil.

L'évaluation de cet algorithme sur un corpus de logatomes et des signaux de parole continue a fournit des résultats satisfaisants : les changements spectraux sont correctement détectés, les variations de pente de la statistique étant très nettes.

Dans cette communication, nous présentons un nouvel algorithme de segmentation de la parole, qui met en œuvre à la fois le test de vraisemblance et la modélisation adaptative. Cette approche conduit à une procédure de détection simple et efficace, ce qui permet d'envisager des applications en ligne. Dans la suite, après un développement de cette méthode, nous donnons les résultats qui illustrent les performances de l'algorithme que nous avons mis au point.

II. CONTRIBUTION DES DIFFÉRENTS TERMES DANS LA DÉTECTION

Le test de divergence (expression 2) est constitué du rapport de vraisemblance *a posteriori* entre les deux lois associées aux modèles long et court terme (RV, expression 1), corrigé par la divergence de Kullback (DIV) [7] :

$$T_{2,n} = RV - DIV$$

$$DIV = -\frac{1}{2} + \frac{1}{2}\left(\log\frac{\sigma_1^2}{\sigma_0^2} - \frac{1}{2\sigma_1^2}\left((e_0 - e_1)^2 + \sigma_0^2\right)\right) \qquad (3)$$

Pour analyser le comportement de la statistique nous avons étudié le rôle de chacun de ces deux termes : RV et DIV. Leur contribution dans la détection est montrée à la figure 1. De haut en bas sont représentés :

(a) - le signal de parole : il s'agit de la transition [ini] dans le logatome [nini]. Les deux transitions [i] à [n] et [n] à [i] sont presque symétriques d'un point de vue acoustique. Mais le test de divergence réagit très différemment pour ces deux transitions : cela illustre sa dissymétrie.

(b) - le rapport de vraisemblance RV.

(c) - le terme de RV qui apparait dans l'expression finale du test de divergence $T_{2,n}$ après simplification, soit :

$$\frac{1}{2}\frac{e_0^2}{\sigma_0^2} \qquad (4)$$

(d) et (e) Les termes de "DIV" qui restent dans l'expression finale du test de divergence après simplification, soit :

$$\frac{e_0 e_1}{\sigma_1^2} - \frac{e_0^2}{2\sigma_1^2} \qquad (5) ;$$

- et :

$$\frac{1}{2}\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right). \qquad (6)$$

(f) la statistique (la somme cumulée de $T_{2,n}$).

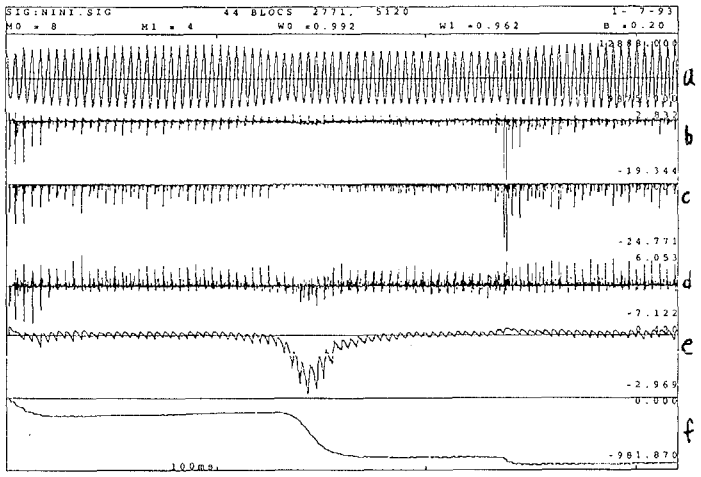

Figure 1. *Éléments servant au calcul de la statistique du test de divergence autour des transitions [i]-[n]-[i].*

Sur la figure, nous pouvons constater que :

- C'est le rapport des deux variances (terme 6) qui provoque le changement de pente de la statistique au niveau de la transition [i] à [n]. En effet, la variance du modèle court terme décroît plus rapidement que celle du modèle long terme, ce dernier suivant plus lentement que le premier les variations rapides du signal. De part et d'autre de cette transition, les deux variances sont du même ordre de grandeur. Cela explique le "creux" dans la courbe (e) et se traduit par le changement de pente de la statistique.

- La contribution à la détection de la transition [n] à [i] par le terme (4) est nettement mise en évidence par les courbes (b) et (c). En effet, le résidu du modèle long terme augmente brusquement, et plus rapidement que sa variance au niveau de la transition considérée. Le rapport, relativement élevé, de ces deux grandeurs provoque un changement de pente dans la statistique. Ce terme représente en fait, à une constante près, la statistique de test utilisée dans les techniques classiques de segmentation mettant en œuvre un seul modèle AR [8].

III. UTILISATION DU TEST DE VRAISEMBLANCE

L'analyse précédente montre que les termes dont la contribution est la plus importante pour la détection des deux transitions [i] à [n] et [n] à [i], sont les suivants :

$$\frac{\sigma_0^2}{\sigma_1^2} \text{ et } \frac{1}{2}\frac{e_0^2}{\sigma_0^2}.$$

On note qu'ils sont également présents dans la statistique du test du rapport de vraisemblance $T_{1,n}$. Cela signifie que l'on peut l'utiliser seul pour la détection de ruptures dans le signal de parole. Pour mettre en évidence ce fait, nous montrons à la figure 2 la statistique du rapport de vraisemblance (a), la divergence de Kullback (b) et la statistique du test de divergence (d).

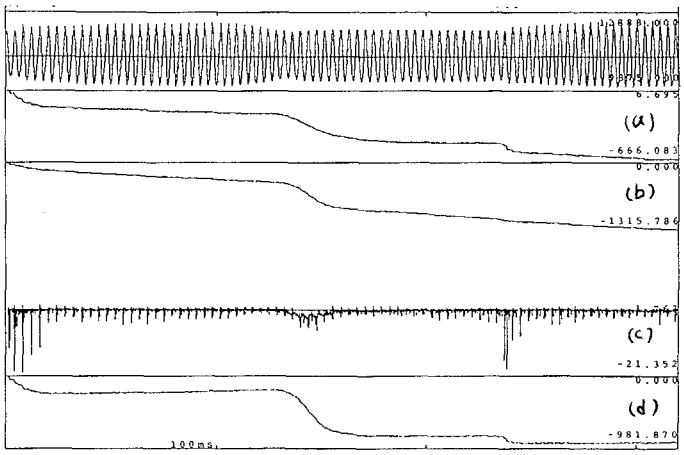

Figure 2. *Comparaison du test de vraisemblance et du test de divergence.*

Nous pouvons constater que la statistique du test de vraisemblance seule (a) réagit tout aussi bien que celle du test de divergence (d) aux deux transitions du signal.

Par ailleurs, l'évaluation de l'algorithme de André-Obrecht ainsi que celle de l'algorithme proposé dans [1] et [5], nous a permis de mettre en évidence un autre phénomène : dans certaines zones de signal la statistique du test de divergence a une forte dérive négative. Par conséquent, plusieurs ruptures successives peuvent être détectées juste après la période d'initialisation, dans l'algorithme de André-Obrecht. Ce défaut se traduit dans l'algorithme utilisant les modèles adaptatifs par des réactions inégales de la statistique aux changements spectraux. À certaines de ces variations correspondent des changements de pente importants de la statistique, tandis que pour d'autres ils le sont moindre. Nous pouvons constater sur la figure 2 une forte chute dans la divergence de Kullback (courbe b), au moment de la première transition, par rapport à celle du test de vraisemblance (courbe a). Ce qui fait que la statistique (courbe d) réagit avec des pentes très différentes aux deux transitions.

Ce phénomène se produit moins souvent dans le test de vraisemblance. Comme il permet de détecter tout aussi bien que le test de divergence les ruptures du signal, nous nous sommes proposés de l'utiliser dans notre algorithme de segmentation.

IV. ALGORITHME DE SEGMENTATION PROPOSÉ

Nous avons développé un algorithme de segmentation en ligne basé sur l'utilisation du test de vraisemblance et une modélisation adaptative. Le modèle long terme est identifié sur une fenêtre croissante à l'aide de l'algorithme de Burg. L'ordre retenu est égal à 8. Plusieurs méthodes d'identification pour le modèle court terme ont été comparées. Les meilleurs résultats sont obtenus en utilisant celle de *prewindow*, avec un coefficient d'oubli égal à 0,962. L'ordre de ce modèle a été fixé à 4.

Pour la procédure de détection, nous avons adopté la méthode classique qui consiste à comparer la chute de la statistique par rapport à son maximum local et décider d'une rupture quand un seuil fixé au préalable est franchi. La réinitialisation des modèles après chaque rupture est dans ce cas nécessaire.

Un détecteur de voisement permet d'adapter le seuil de détection et le biais à ajouter à la statistique. Leurs valeurs sont réactualisées toutes les 4 ms.

V. RÉSULTATS ET DISCUSSION

Nous avons évalué notre algorithme de segmentation sur un large corpus multi-locuteur (logatomes et parole continue). La méthode proposée permet de détecter la quasi tolalité des ruptures dont la détection exigeait l'emploi de la méthode *backward*.

Aux figures 3 et 4, nous montrons un exemple de résultats obtenus respectivement avec l'algorithme de André-Obrecht et avec l'algorithme proposé. Ces résultats sont très proches : tous les changements spectraux importants sont correctement détectés. On peut noter que toutes les ruptures rattrapées par la méthode *backward* sont détectées directement en utilisant le test de vraisemblance et les modèles adaptatifs. Cela signifie que cet algorithme peut être utilisé dans des applications en temps réel.

Si le problème d'omissions est résolu, celui de surdétections ne l'est que partiellement. Une sursegmentation a souvent lieu pour les occlusives voisées. Certes cela ne constitue pas un inconvénient majeur puisque la notion de sursegmentation est relative, mais des études sont en cours pour améliorer ce point.

CONCLUSION

Le test de divergence utilisé dans la segmentation statistique du signal de parole est dissymétrique. Les solutions apportées par ailleurs à ce problème ne sont pas enviageables pour des applications en temps réel. C'est à ce problème que nous avons essayé d'apporter une solution.

Dans cette communication, nous montrons que le test de vraisemblance peut être avantageusement utilisé dans la détection automatique de ruptures dans le signal de parole. Il faut néanmoins souligner l'apport des modèles adaptifs sans lesquels les omissions seraient inévitables. Un algorithme de segmentation en ligne basé sur ce principe a été développé. La procédure de détection est très simple et efficace.

Les résultats obtenus sont très satisfaisants : les ruptures sont correctement détectées sans utiliser la méthode *backward*, ce qui permet d'envisager son application en temps réel.

RÉFÉRENCES

[1] Achab N. & Feng G. (1991), *Modèles adaptatifs dans la détection automatique de ruptures en vue du codage à débit variable*. Séminaire GCP, Le Mans 3-4 juin 1991, 96-103.

[2] André-Obrecht R. (1985), *Segmentation automatique du signal de parole*. Thèse de 3ème cycle, Univ. de Rennes I.

[3] André-Obrecht R. (1988), *A new statistical approach for the segmentation of continuous speech signals*. IEEE Trans. ASSP, vol. 36, n°1, 29-40.

[4] Basseville M. & Benveniste. A. (1983), *Sequential detection of abrupt changes in spectral characteristics of digital signals*. IEEE Trans. Inform. Theory, vol. IT-29, n°5, 708-723.

[5] Feng G., Achab N. & Combescure P., (1991) *On-line speech segmentation using adaptive models: application to variable speech coding*. Eurospeech 1991, Genova, Italy, vol. 2, 705-708.

[6] Friedlander B. (1982), Lattice filters for adaptive processing. Proceedings of the IEEE, vol. 70, n°8, 829-867.

[7] Kullback S. (1959), "*Information theory and statistics*". John Wiley & sons Inc., New-York.

[8] Seguen J. & Sanderson A.C., (1980), *Detcting change in a time serie*. IEEE Trans. Inform. Theory, vol. IT-26, N°2, 249-254.

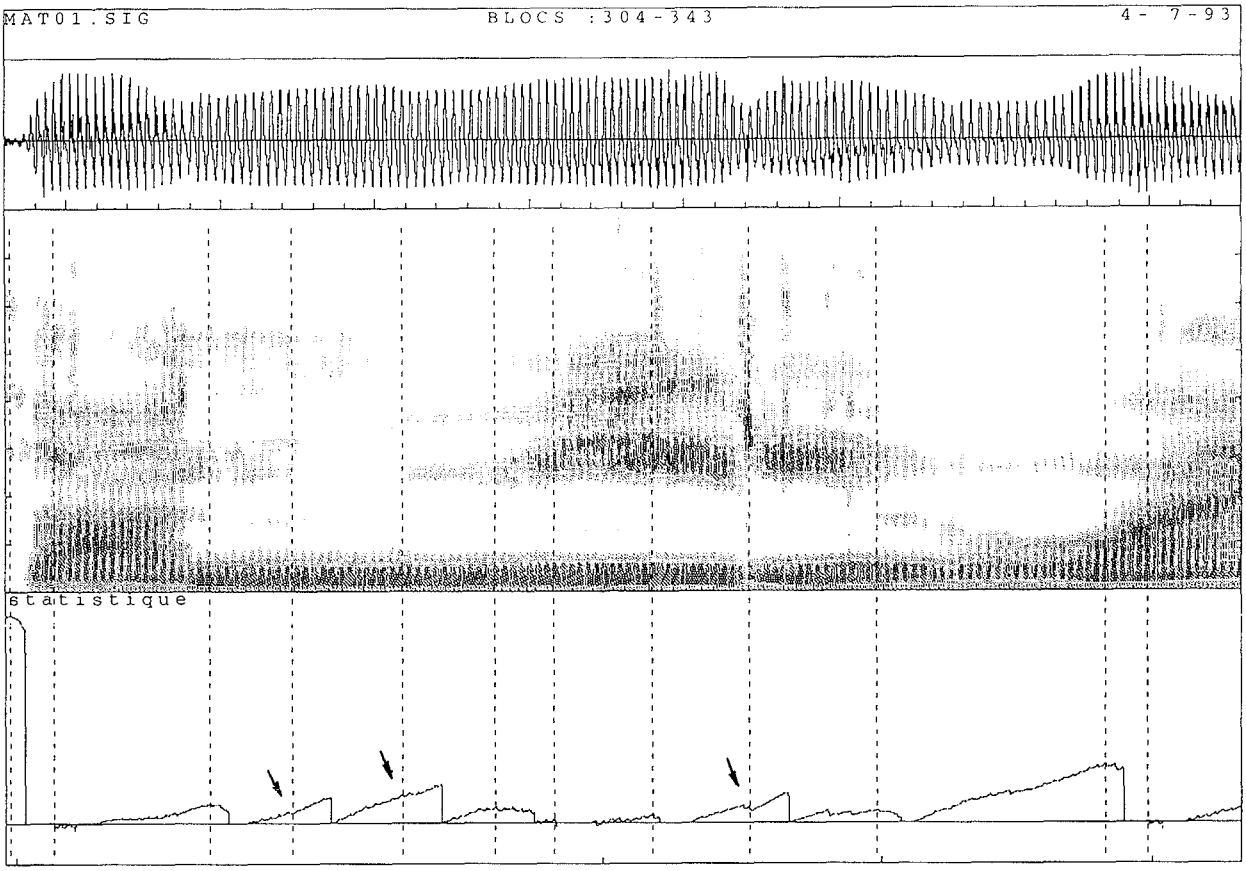

Figure 3. *Segmentation d'un signal de parole continue : [ã n µ] extrait de la phrase " Annie s'ennuie loin ...". Les changements spectraux sont tous correctement détectés. Les ruptures indiquées par les flèches sont détectées par la procédure backward.*

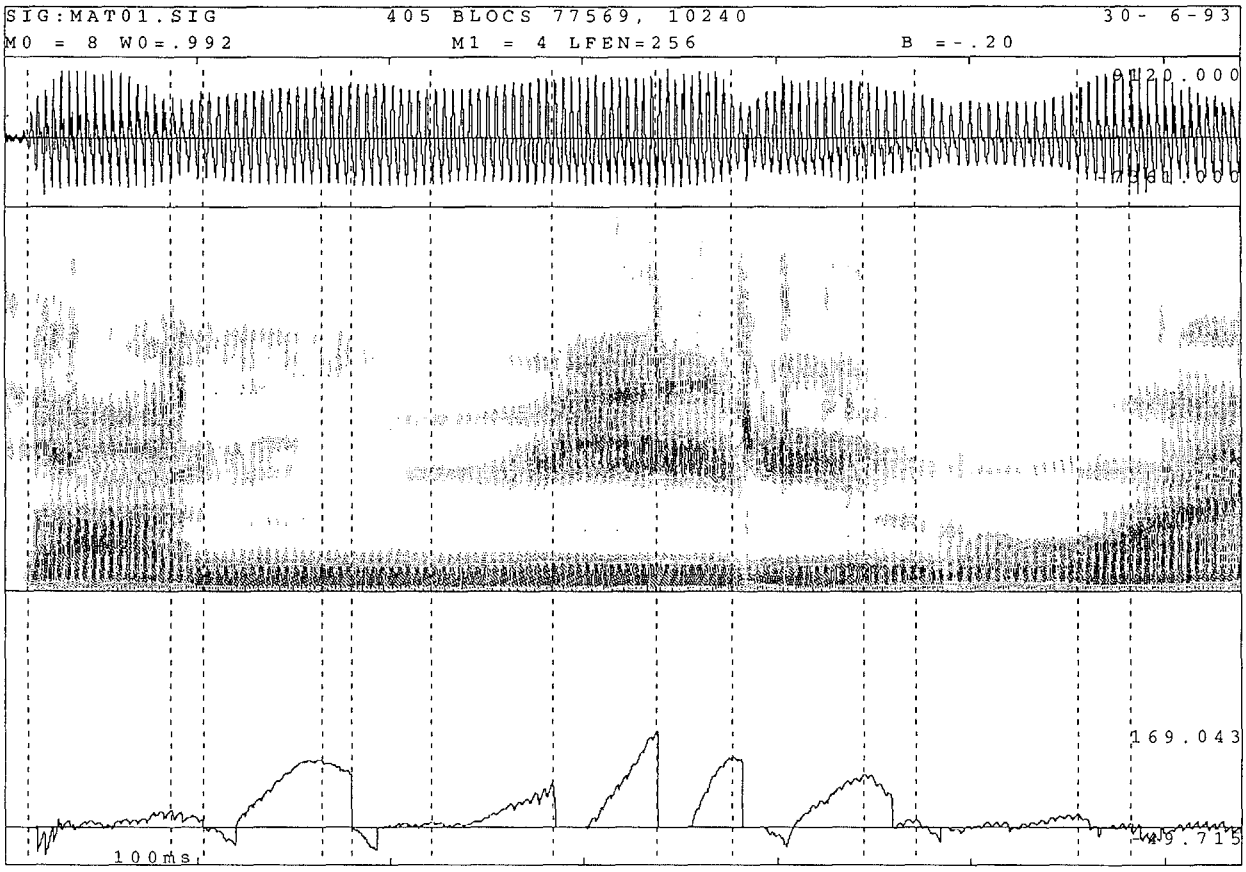

Figure 4. *Résultats de la segmentation du même signal que ci-dessus par l'algorithme proposé (test de vraisemblance + modèles adaptatifs). Toutes les ruptures sont détectées directement sans faire appel à la procédure backward.*