Réhaussement de la parole par la séparation de sources dans un mélange convolutif

H. Nguyen Thi, J. Caelen, Christian Jutten

JOURNAL DE PHYSIQUE IV
Colloque C5, supplément au Journal de Physique III, Volume 4, mai 1994

Réhaussement de la parole par la séparation de sources dans un mélange convolutif

H.L. NGUYEN THI, J. CAELEN et Ch. JUTTEN*

ICP-INPG, 46 avenue Félix Viallet, 38031 Grenoble cedex, France
** TIRF-INPG, 46 avenue Félix Viallet, 38031 Grenoble cedex, France*

Abstract : In this paper, we proposed a approach to the two-microphone speech enhancement problem based on the Blind Separation of Sources for convolutive mixtures. The separation of signals can be achieved by a recursive structure of adaptive filters FIR. The filter coefficients must be updated by using Higher Order Statistics in order to insure the output independence. Various adaptive algorithms based on different independence criteria are proposed to estimate the filters. The experimental results have been obtained for different types of signals in various situations of mixtures. For instance, using two microphones, the algorithms are able to separate two speech signals without any knowledge of the signals.

1. INTRODUCTION

De nombreuses méthodes de rehaussement de la parole ont été proposées, mais ce problème reste très difficile. Les méthodes fondées sur deux ou plusieurs observations distinctes peuvent théoriquement fournir de meilleures performances [1]. Plusieurs algorithmes reposent sur l'hypothèse d'une référence de bruit seul. Cependant dans une situation réelle de prise de son par deux microphones, cette hypothèse n'est pas réaliste. Un signal observé à la sortie d'un microphone résulte d'un mélange convolutif inconnu de signaux primitifs eux-mêmes inconnus (sources). En généralisant l'algorithme de Hérault-Jutten [2], nous avions proposé une solution au rehaussement de la parole par la Séparation Aveugle de Sources dans le cas de mélanges convolutifs sans hypothèse a priori sur les signaux. Cependant, la convergence de l'algorithme n'est possible que si les densités de probabilité des sources sont paires [2]. Pour pallier cette limitation, un algorithme fondé sur des cumulants croisés d'ordre 4 a été étudié [4]. Les résultats expérimentaux des algorithmes ont été obtenus dans des mélanges simulés et enregistrés pour différents types des signaux : mélange du signal et du bruit aléatoire ou réel, mélange de deux signaux de la parole [3].

2. MODELE DE MELANGE CONVOLUTIF D'UNE PRISE DE SON

2.1. Modèle général

Considérons une situation de prise de son par deux microphones : une source de parole (locuteur) se situe près du 1er microphone, l'autre source de perturbation (un bruit ou un autre signal de la parole) est près du 2ème microphone. A la sortie des microphones, on observe une superposition des signaux primitifs inconnus selon un mélange inconnu (Fig.1). En général, c'est un mélange convolutif des signaux à large bande, qui dépend de la propagation des signaux dans le milieu, de la position des microphones et des sources, et des caractéristique de la salle. Ce modèle a été suggéré par Feder et al. [1]. Les équations en z des signaux du mélange s'écrivent alors :

$$Y_1(z) = A_{11}(z).X_1(z) + A_{12}(z).X_2(z) + B_1(z)$$
$$Y_2(z) = A_{21}(z).X_1(z) + A_{22}(z).X_2(z) + B_2(z) \quad (1)$$

où : X_1 et X_2 sont deux *"sources" inconnues* supposées *indépendantes*, Y_1 et Y_2 sont les deux signaux observés à la sortie des microphones, B_1 et B_2 sont des des erreurs de mesure, $A_{ij}(z)$ sont les fonctions de transfert des filtres linéaires.

2.2. Modèle convolutif simplifié

Nous proposons un modèle simplifié du mélange convolutif modélisé par les filtres RIF causaux, dans lequel un microphone est placé près du locuteur et l'autre près de la source de perturbation. On peut ainsi voir les filtres A_{11} et A_{22} comme des scalaires égaux à 1 et les erreurs $b_i(n)$ négligeables.

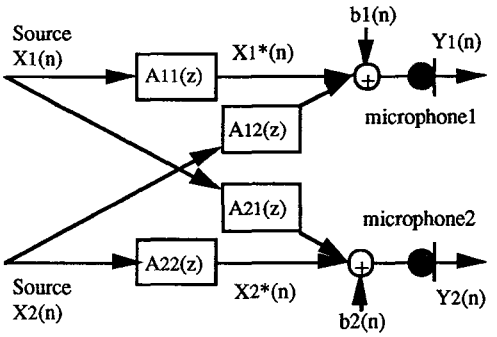

Fig. 1. Modèle général du mélange convolutif.

En supposant les filtres constitués de M coefficients, l'équation du mélange simplifié s'écrit :

$$Y_i(n) = X_i(n) + \sum_{k=0}^{M-1} a_{ij}(k).\, X_j(n-k) \quad \text{avec } i \neq j \text{ et } i,j \in [1, 2] \tag{2}$$

où les sources $X_i(n)$ et les filtres A_{ij} sont tous inconnus.

A partir des seuls signaux observés $Y_i(n)$, comment peut-on retrouver les sources $X_i(n)$?

3. SOLUTION DE LA SEPARATION DE SOURCES

3.1. Architecture et critère de séparation

En se basant sur le critère d'indépendance des signaux de sortie, le principe de la séparation de sources est très intéressant puisqu'aucune hypothèse spéciale sur des signaux n'est nécessaire. De nombreuses méthodes de séparation de sources dans le cas de mélanges instantanés ont été proposées. Dans le cas de mélanges convolutifs, une architecture de séparation des signaux à large bande est présentée dans la figure 2. L'équation des sorties s'écrit alors :

$$S_i(n) = Y_i(n) - \sum_{k=0}^{M-1} C_{ij}(k).\, S_j(n-k)\,, \tag{3}$$

et encore :

$$S_i(z) = \frac{(1 - C_{ij}.A_{ji}).X_i(z) + (A_{ij} - C_{ij}).X_j(z)}{(1 - C_{ij}(z).C_{ji}(z))} \tag{4}$$

Cette architecture peut fournir une solution de séparation de sources. En effet, considérons l'équation (4), si : $C_{ij}(z) = A_{ij}(z)$, (5) alors on obtient : $S_i(z) = X_i(z)$.

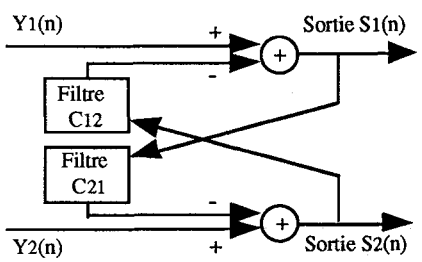

Fig. 2. Architecture de séparation

Par ailleurs, si les deux signaux S_1 et S_2 sont indépendants, alors tous les cumulants croisés de S_1 et S_2 sont égaux à zéro. Cependant, les sources X_i et les filtres A_{ij} sont inconnus : on ne peut pas trouver directement les filtres C_{ij}. Les coefficients des filtres peuvent être par exemple ajustés par un algorithme adaptatif en faisant appel à des statistiques d'ordre supérieur pour rendre les sorties indépendantes.

3.2. Algorithmes de séparation

A partir des critères d'indépendance des sorties, nous avons proposé des algorithmes de type "itération stochastique" [4], dans lesquels à chaque instant n, les coefficients k ($k \in [0, M-1]$) des filtres C_{ij} sont ajustés de façon à annuler un critère $\phi_{ij}(n,k)$ selon la règle :

$$c_{ij}(n+1,k) = c_{ij}(n,k) - \mu.\phi_{ij}(n,k) \tag{6}$$

où : μ, le gain d'adaptation, dépend du signe du critère au voisinage du zéro.

Les critères $\phi_{ij}(n,k)$ peuvent être fondés sur les moments ou les cumulants croisés des sorties. A la convergence de l'algorithme on obtient : $\phi_{ij}(n,k) \approx 0$, l'algorithme effectue donc une approximation de test d'indépendance des sorties. Les algorithmes suivants correspondent à différents critères.

3.2.1. Algorithme fondé sur les fonctions non-linéaires

Les coefficients des filtres sont ajustés par la règle suivante :

$$c_{ij}(n+1,k) = c_{ij}(n,k) - \mu.f(s_i(n)).g(s_j(n-k)) \qquad (7)$$

où : $\mu \geq 0$, f et g sont les fonctions non-linéaire impaires, et $s_i(n)$ est une estimation du signal centré de $S_i(n)$.

Dans le cas le plus simple, les fonctions sont : $f = (.)^3$ et $g = (.)$. Les coefficients $c_{ij}(n,k)$ sont donc ajustés en annulant les moments croisés d'ordre 4, $E\{s_i^3(n).s_j(n-k)\}$. Cependant, cet algorithme (7) ne converge que si la densité de probabilité (ddp) des sources est paire.

3.2.2. Annulation des cumulants croisés d'ordre 4

L'indépendance rigoureuse est obtenue lorsque tous les cumulants croisés de tous ordres sont nuls. Par exemple à l'ordre 4 , il existe trois cumulants croisés de deux signaux centrés $s_i(n)$ et $s_j(n)$:

$$\begin{aligned}
Cum_{31}(i,j,n) &= E\{s_i^3(n).s_j(n)\} - 3.E\{s_i^2(n)\}.E\{s_i(n).s_j(n)\} \\
Cum_{13}(i,j,n) &= E\{s_i(n).s_j^3(n)\} - 3.E\{s_j^2(n)\}.E\{s_i(n).s_j(n)\} \\
Cum_{22}(i,j,n) &= E\{s_i^2(n).s_j^2(n)\} - E\{s_i^2(n)\}.E\{s_j^2(n)\} - 2.(E\{s_i(n).s_j(n)\})^2
\end{aligned} \qquad (8)$$

où $E\{s_i^l(n).s_j^m(n)\}$ est le moment croisé d'ordre (l+m) à l'instant n des signaux s_i et s_j estimé par :

$$M_{lm}(i,j,n) = (1- a).M_{lm}(i,j,n-1) + a.s_i^l(n).s_j^m(n) \text{ , avec } a = cte. \qquad (9)$$

Dans le cas de mélanges convolutifs, on peut considérer comme critères approximatifs d'indépendance l'annulation des cumulants croisés $Cum_{31}(i,j,n,k)$ et $Cum_{13}(i,j,n,k)$. Cependant, expérimentalement, on observe que les cumulants peuvent s'annuler avec une pense soit positive soit négative. On propose donc la règle d'adaptation suivante qui tient compte du signe de la dérivée du critère :

$$c_{ij}(n+1,k) = c_{ij}(n,k) - \mu.sign\left(\frac{\partial Cum_{31}(i,j,n,k)}{\partial c_{ij}(n,k)}\right).Cum_{31}(i,j,n,k) \text{ , avec } \mu > 0 \qquad (10)$$

L'algorithme fondé sur les cumulants peut séparer les signaux de densité de probabilité quelconque, mais au prix d'une plus grande complexité.

3.3. Résultats expérimentaux

Mélanges simulés : Les signaux Y_1 et Y_2 de mélanges simulés sont calculés par l'équation (2), dans laquelle les deux sources X_1 et X_2 sont des signaux de la parole et des bruits aléatoires ou du bruit enregistré d'ambiance réelle, les deux filtres $A_{ij}(z)$ sont des filtres passe-bas de 10 coefficients. Les algorithmes ont été appliqués aux signaux Y_1 et Y_2, après une phase d'apprentissage d'environ 3000 échantillons, l'algorithme a convergé, les filtres C_{ij} s'approchent des filtres A_{ij}. Il fournit alors (Fig. 3) la sortie S_1 proche de la source X_1 (phrase "Le camp d'été s'est passé") et S_2 proche de la source de bruit X_2. Le RSB du signal estimé (opposé la diaphonie résiduelle) obtenu est environ de 15 à 20 dB (Fig. 4a), et les erreurs quadratiques paramétriques des filtres estimés C_{ij} par rapports aux filtres A_{ij} sont de 0,001 à 0,005 (Fig. 4b).

Mélanges réels enregistrés : Dans ce cas, les signaux des mélanges sont enregistrés par deux microphones dans une chambre sourde. La distance entre une source et un microphone est de 5 à 50 cm et la distance entre deux microphones est environ de 30 à 40 cm. Les résultats dans ce cas restent d'une qualité moyenne avec une diaphonie résiduelle moyenne estimée environ de - 8 à -10 dB.

4. IDENTIFICATION D'UN SYSTEME PHYSIQUE DE PRISE DE SON

Les résultats médiocres obtenus dans le cas de mélanges réels, peuvent être expliqués par un modèle très simple de mélange. Pour un modèle plus réaliste de mélange, les filtres $A_{ij}(z)$ doivent être de type RII, ce qui pose normalement des problèmes de stabilité et de complexité. Nous avons cherché à identifier les filtres

mis en jeu dans un système physique de prise de son "locuteur-microphone". Ce système se compose d'un haut-parleur placé devant d'un microphone à la distance de 5 à 50 cm. Dans une chambre sourde, on envoie, à l'entrée du haut-parleur, une séquence binaire pseudo-aléatoire u(n) qui modélise approximativement un bruit blanc discret "riche" en fréquence. En temps réel, on observe le signal v(n) à la sortie du microphone. La sortie mesurée est en général bruitée de manière aléatoire. L'ensemble du système est modélisé par un modèle ARMA de la forme :

$$v(n) = \frac{q^{-nk}.B(q^{-1})}{A(q^{-1})}.u(n) + e(n), \text{ avec } A(q^{-1}) = 1 + \sum_{k=1}^{na} a_k.q^{-k} \text{ et } B(q^{-1}) = \sum_{k=0}^{nb} b_k.q^{-k}, \quad (11)$$

où : e(n) est une erreur de prédiction.

Les résultats expérimentaux obtenus, en appliquant des méthodes d'identification basées sur le "blanchiment" de l'erreur de prédiction, montrent que le modèle ARMA avec un nombre de coefficients de l'ordre de 10 (na + nb) peut fournir de bonne identification. Avec un modèle MA (na = 0) comportant une cinquantaine de coefficients pour nb, on peut obtenir des résultats un peu inférieur.

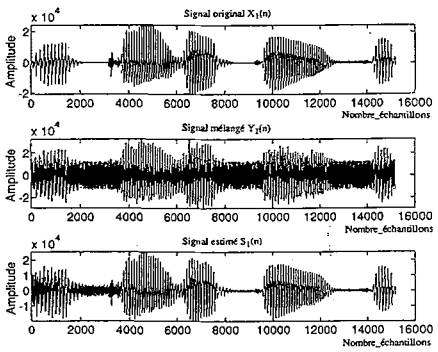

Fig. 3. Les signaux de la voie 1: signal original X_1, signal du mélange Y_1 et signal estimé S_1.

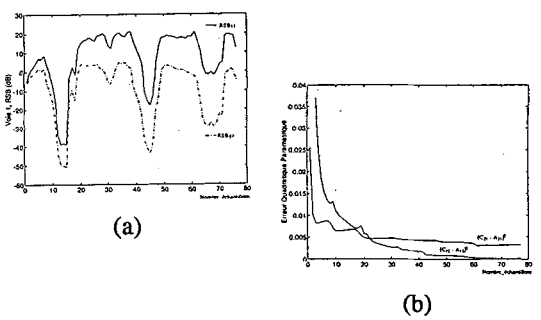

Fig. 4. Les courbes RSB des signaux Y_1 et S_1, et les erreurs paramétriques des filtres esimés C_{ij}.

5. CONCLUSION

Le rehaussement de la parole peut être atteint par un modèle de séparation des signaux. A partir de deux signaux observés à la sortie des microphones, la séparation de signaux non-stationnaires et à bande large peut être obtenue par une architecture récursive de deux filtres adaptatifs de type RIF pour rendre les sorties indépendantes. Le principe de la séparation de sources permet de rehausser la parole quelle que soit la nature de perturbation, en plus il est capable de restituer tous les signaux primitifs dans un mélange de prise de son. Dans tous les cas expérimentaux, de bons résultats ont été obtenus pour des mélanges simulés mais ils sont encore de qualité moyenne pour des mélanges réels. Cela montre que le modèle simplifié de mélange convolutif modélisé par des filtres RIF n'est pas suffisamment réaliste. En effet, les expériences d'identification du système de prise de son suggère que les hypothèses de mélange linéaire, de sources ponctuelles et du milieu sans écho doivent être remis en question. Par ailleurs, de nombreux points théoriques sont en cours d'étude: critères d'indépendance en mélange convolutif, convergence des algorithmes, stabilité des filtres, etc.

Références

[1]. FEDER M. , OPPENHEIM A. V. , WEINSTEIN E. , "Maximum likelihood noise cancellation using the EM algorithm", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-37, n° 2, February 1989.

[2]. JUTTEN C. , HERAULT J. , COMON P. , SOROUCHYARI E. "Blind separation of sources : Part I, Part II, Part III", Signal Processing, vol. 24, p. 1-29, 1991.

[3]. NGUYEN THI H. L. , JUTTEN C. , CAELEN J. "Speech Enhancement : Analysis and Comparison of methods on various real situations ", VI European Signal Processing Conference EUSIPCO-92 , Vol. I, p. 303-306, Brussels, August 24-27, 1992.

[4]. NGUYEN THI H. L. "Séparation aveugle de Sources à large bande dans un mélange convolutif : Application au rehaussement de la parole", Thèse doctorat de l'INP Grenoble, France, Janvier 1993.