

# SYSTÈMES COMBINÉS D'ANNULATION D'ÉCHO ET DE RÉDUCTION DE BRUIT

Régine Le Bouquin Jeannès - Baghdad Ayad

Laboratoire de Traitement du Signal et de l'Image - Université de Rennes 1  
Bât. 22 - Campus de Beaulieu - 35042 RENNES CEDEX - FRANCE

## RÉSUMÉ

Ce papier traite de l'amélioration de la parole pour les télécommunications mains-libres. Les deux problèmes majeurs à résoudre sont la réduction de bruit et l'annulation d'écho. Deux structures de base sont généralement étudiées, qui se résument à la mise en cascade d'un réducteur de bruit et d'un annuleur d'écho. Notre objectif est de trouver de nouvelles structures permettant d'obtenir un signal utile peu distordu, pour des niveaux d'écho et de bruit résiduels acceptables. Deux structures sont proposées et évaluées dans le cadre de la téléphonie mobile.

## 1. INTRODUCTION

L'essor des télécommunications mains-libres stimule aujourd'hui les efforts de recherche et de développement de systèmes performants. Si ces télécommunications apportent un certain confort d'utilisation, voire dans certains cas une sécurité, elles doivent assurer la qualité du signal transmis malgré les perturbations. Les postes mains-libres sont souvent utilisés dans des environnements bruités et éloigner le combiné du locuteur diminue le rapport signal à bruit. De plus, le haut-parleur émettant le signal du locuteur lointain crée un écho sur le microphone. Un traitement doit donc être effectué sur le signal capté réduisant au mieux le bruit et l'écho acoustique, sans apporter de distorsion notable sur le signal utile. Les traitements classiques consistent à mettre en cascade chacun des traitements nécessaires, un annuleur d'écho et un réducteur de bruit [1,2]. Notre étude porte sur l'investigation de deux nouvelles structures combinant annulation d'écho et réduction de bruit. Celles-ci seront présentées dans le paragraphe suivant et leurs performances comparées à celles des structures de base dans le cadre des radio-mobiles pour divers rapports signal à bruit et écho à bruit.

## 2. SYSTÈMES COMBINÉS

### 2.1. Position du problème

Nous disposons d'un seul microphone et d'un haut-parleur. Le signal émis par le haut-parleur et provenant du locuteur lointain est noté  $z(t)$ . L'observation  $x(t)$  reçue par le microphone est la somme du signal de parole à transmettre  $s(t)$ , du bruit perturbateur  $b(t)$ , et d'un écho  $e(t)$  corrélé avec  $z(t)$ . Pour résoudre le problème de l'estimation du signal de parole, un annuleur d'écho et un réducteur de bruit s'imposent. Pour l'annuleur d'écho, une référence est disponible, qui est filtrée pour donner une estimée de l'écho  $e(t)$ .

## ABSTRACT

This paper deals with the problem of speech enhancement for hands-free communications, including two major problems: noise reduction and acoustic echo cancellation. Two basic structures are usually proposed, cascading a noise reduction system and an echo cancellation system. Our objective is to investigate new joint structures to get a near-end speech signal with a minimum distortion and low levels of echo and noise. Two structures are proposed. An objective evaluation of performance in a car environment is given.

Concernant le réducteur de bruit, aucune référence bruit seul n'est accessible, le filtre correspondant est alors déterminé à partir des séquences de l'observation  $x(t)$  où seul le bruit est présent. Les filtres respectifs et qui conditionnent les résultats expérimentaux sont brièvement décrits dans le paragraphe consacré à l'évaluation des performances.

### 2.2. Filtre optimal

Le filtre optimal au sens de l'erreur quadratique moyenne minimale et opérant dans le domaine fréquentiel s'écrit :

$$W(f) = \frac{\gamma_{ss}(f)}{\gamma_{ss}(f) + \gamma_{bb}(f)} \begin{bmatrix} 1 \\ -\frac{\gamma_{xz}(f)}{\gamma_{zz}(f)} \end{bmatrix}$$

Ce filtre est appliqué au vecteur  $Y(f) = [X(f) Z(f)]^T$  et le spectre du signal estimé vaut  $\hat{S}(f) = W^T(f)Y(f)$ .  $X(f)$ ,  $Z(f)$ ,  $\hat{S}(f)$  représentent respectivement les spectres des signaux  $x(t)$ ,  $z(t)$  et  $\hat{s}(t)$ .  $\gamma_{ss}(f)$ ,  $\gamma_{bb}(f)$  et  $\gamma_{zz}(f)$  sont les densités spectrales de puissance (dsp) de  $s(t)$ ,  $b(t)$  et  $z(t)$ .  $\gamma_{xz}(f)$  est la densité interspectrale de puissance entre les observations  $x(t)$  et  $z(t)$ . Ainsi, le filtre optimal est la mise en série d'un annuleur d'écho et d'un réducteur de bruit. La sortie de l'annuleur d'écho vaut idéalement  $s(t) + b(t)$ . Le réducteur de bruit est calculé ensuite et appliqué à cette sortie.

**Note :** dans la suite, sauf précision, le filtre réducteur de bruit est calculé à partir de son entrée.

### 2.3. Structures de base

La première structure de base, dénommée AEC+NR (Acoustic Echo Cancellation + Noise Reduction), correspond à la réalisation du traitement optimal précédemment cité (Figure 1). En pratique, le



filtre annuleur d'écho est adaptatif, le réducteur de bruit est appris sur un nombre limité de blocs et aucun des deux traitements n'est optimal. L'annuleur d'écho placé en tête est alors perturbé par le bruit présent ce qui se traduit par un écho résiduel important. D'autre part, les coefficients de l'annuleur d'écho étant perturbés, il apparaît nécessaire de poursuivre l'adaptation en présence de double parole.

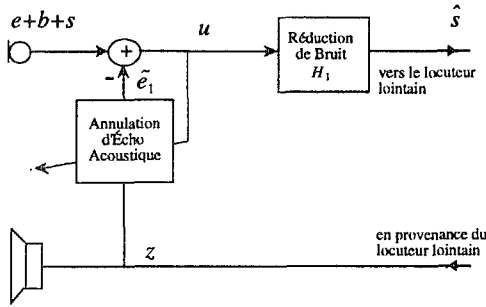


Figure 1. Structure AEC+NR

Afin de réduire l'influence du bruit sur l'annuleur d'écho, on peut effectuer la réduction de bruit avant l'annulation d'écho. Cette structure que nous dénommons NR+AEC (Noise Reduction + Acoustic Echo Cancellation) est représentée Figure 2. Dans la mesure où le bruit est réduit à l'entrée de l'annuleur d'écho, l'arrêt de l'adaptation en présence de double parole est envisageable. Cependant, que l'on arrête ou non l'adaptation, en présence d'écho, le bruit résiduel est plus important que dans la structure AEC+NR.

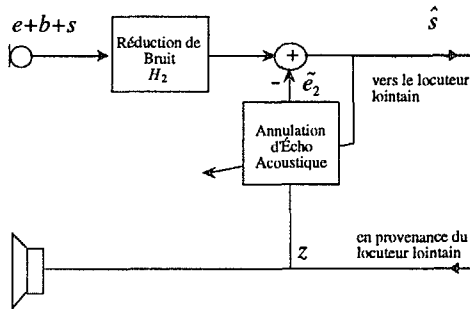


Figure 2. Structure NR+AEC

#### 2.4. Structures proposées

Dans la structure AEC+NR, la sortie de l'annuleur d'écho contient le signal de parole déformé, déformation qui est accentuée par le réducteur de bruit. En conséquence, dans la première structure nouvelle que nous proposons, c'est la dsp de l'observation  $x(t)$  qui est utilisée pour déterminer le réducteur de bruit (Figure 3). Cette structure est dénommée AEC//NR, les traitements d'annulation d'écho et de réduction de bruit étant calculés directement et simultanément à partir de l'observation.

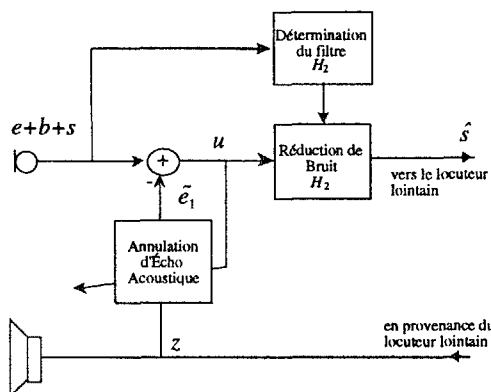


Figure 3. Structure AEC//NR

Les expériences menées sur les structures de base montrent que la puissance de l'erreur  $(e(t) - \tilde{e}_2(t))$  est plus faible que celle de  $(e(t) - \tilde{e}_1(t))$ . Ceci signifie que, malgré la distorsion apportée sur l'écho, il est préférable de commencer par réduire le bruit pour obtenir une meilleure estimation de l'écho. Cette remarque nous conduit à proposer une autre structure, dénommée structure AEC+2NR (Figure 4). Une réduction de bruit est d'abord réalisée sur l'observation par l'intermédiaire du filtre  $H_2$  avant d'effectuer une annulation d'écho fournissant un écho estimé  $\tilde{e}_2(t)$  alors soustrait de l'observation  $x(t)$ . On obtient la quantité  $v(t)$  constituée des grandeurs  $s(t)$ ,  $b(t)$  et d'un écho résiduel  $(e(t) - \tilde{e}_2(t))$  faible. Un réducteur de bruit est alors appliqué sur  $v(t)$  suivant le filtre  $H_3$ . La réduction de bruit effectuée sur l'observation permet de tendre vers la structure optimale.

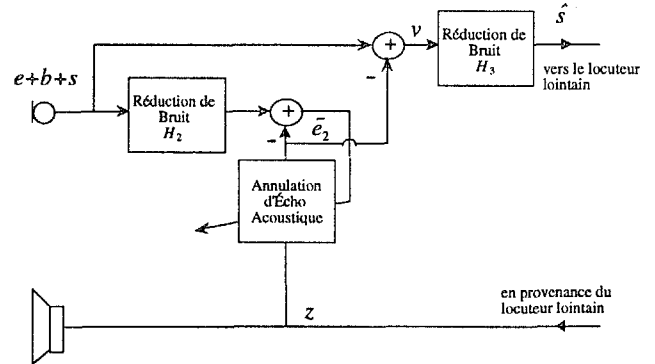


Figure 4. Structure AEC+2NR

### 3. MÉTHODOLOGIE D'ÉVALUATION

Les quatre structures présentées ont été testées dans le cadre de communications radio-mobiles. Pour construire la base de données, le signal de parole (locuteur proche), l'écho (locuteur lointain) et le bruit perturbateur sont enregistrés séparément afin de comparer les structures pour différents rapports signal à bruit  $S/B$  et écho à bruit  $E/B$ . Ces rapports sont respectivement calculés sur la longueur du signal et de l'écho. Le bruit est enregistré dans une voiture roulant à 130 km/h. Par simple combinaison, nous obtenons le signal composite donné Figure 9.c : la première séquence est un écho bruité ("Single Talk" : ST), la deuxième séquence comprend un signal de parole ajouté à l'écho bruité ("Double Talk" : DT). Le bruit existe en permanence et les périodes de silence permettent d'appréhender ses caractéristiques.

Les mesures objectives retenues sont les suivantes :

- la mesure "Echo Return Loss Enhancement"  $ERLE$  dans les modes ST et DT,

$$ERLE = \frac{1}{N} \sum_{k=1}^N 10 \log \frac{P_e(k)}{P_{e_r}(k)}$$

- le gain  $G$  en mode DT,

$$G = \frac{1}{N} \sum_{k=1}^N 10 \log \frac{P_{e+b}(k)}{P_{s-\hat{s}}(k)}$$

$P_y(k)$  représente la puissance de  $y$  calculée sur le bloc  $k$ ,  $N$  est le nombre de blocs sur lequel l'estimation est faite.  $\hat{s}$  est l'estimée finale du signal et  $e_r$  est l'écho résiduel obtenu comme suit :

- dans la structure AEC+NR,  $e_r$  est obtenu en filtrant  $(e - \tilde{e}_1)$  par  $H_1$ ,
- dans la structure NR+AEC,  $e_r$  est donné par la différence entre l'écho  $e$  filtré par  $H_2$  et l'écho estimé  $\tilde{e}_2$ ,

- dans la structure AEC//NR,  $e_r$  est obtenu en filtrant  $(e - \tilde{e}_1)$  par le réducteur de bruit  $H_2$ ,
  - dans la structure AEC+2NR,  $e_r$  est obtenu en filtrant la différence  $(e - \tilde{e}_2)$  par  $H_3$ .
- Notons que le gain  $G$  est une mesure pertinente qui prend en compte la distorsion sur la parole, le bruit et l'écho résiduels.

L'anneur d'écho acoustique est l'algorithme "Generalized Multi-Delay Filter" [3] ; celui-ci est un traitement adaptatif fréquentiel opérant par bloc. Ses deux caractéristiques sont : le partage de la réponse impulsionnelle en segments afin de contrôler le retard dû au traitement et l'introduction d'un paramètre qui contrôle la fréquence de mise à jour des coefficients.

Le réducteur de bruit est l'estimateur d'amplitude proposé par Ephraïm et Malah [4], pour lequel les composantes spectrales de bruit et de parole sont considérées comme des variables aléatoires indépendantes et gaussiennes. Le spectre de sortie du réducteur de bruit est obtenu en pondérant le spectre d'entrée par un gain qui inclut un filtrage de Wiener et une fonction prenant en compte la probabilité de présence du signal.

#### 4. RÉSULTATS

Dans un premier temps, nous avons choisi des rapports signal à bruit et écho à bruit identiques, ceux-ci prenant des valeurs entre -3 dB et +15 dB.

La Figure 5 représente les valeurs du  $ERLE$  obtenues en mode ST pour les quatre structures étudiées. Les valeurs les plus élevées sont obtenues pour la structure AEC+NR et la structure AEC+2NR, celle-ci étant la plus performante pour des rapports  $S/B$  et  $E/B$  faibles (inférieurs à 6 dB). La différence entre les valeurs du  $ERLE$  peut s'expliquer en considérant l'expression des gains des filtres réducteurs de bruit en les identifiant à des filtres de Wiener :

$$H_1(f) = \frac{\gamma_{uu}(f) - \gamma_{bb}(f)}{\gamma_{uu}(f)}, \quad H_2(f) = \frac{\gamma_{xx}(f) - \gamma_{bb}(f)}{\gamma_{xx}(f)},$$

$$H_3(f) = \frac{\gamma_{vv}(f) - \gamma_{bb}(f)}{\gamma_{vv}(f)}.$$

$\gamma_{uu}(f) < \gamma_{xx}(f)$  et  $\gamma_{vv}(f) < \gamma_{xx}(f)$  entraînent que les gains des filtres  $H_1$  et  $H_3$  sont plus faibles que le gain du filtre  $H_2$ .

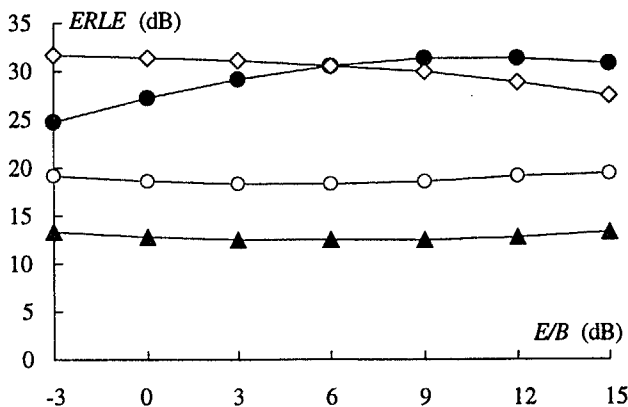


Figure 5.  $ERLE$  en mode ST  
 ● structure AEC+NR, ▲ structure NR+AEC,  
 ○ structure AEC//NR, ◇ structure AEC+2NR

Dans les structures AEC+NR et AEC//NR, l'écho estimé par l'anneur d'écho acoustique est le même, et l'écho résiduel est alors davantage réduit par la structure AEC+NR. Le gain du filtre  $H_3$  de la structure AEC+2NR est faible, ce qui explique également les valeurs importantes du  $ERLE$ . Les structures AEC+NR et AEC+2NR donnent des résultats comparables. Pour des rapports  $E/B$  faibles, l'anneur

d'écho de la structure AEC+NR est davantage perturbé, l'écho est donc moins réduit, le gain de  $H_3$  est plus faible que celui de  $H_1$  et la structure AEC+2NR donne un  $ERLE$  plus important. Par contre, pour des rapports  $E/B$  importants, le réducteur de bruit placé en tête distord l'écho, et ce dernier se trouve moins bien estimé. La structure AEC+NR est alors préférable. Les structures NR+AEC et AEC//NR utilisent le même réducteur de bruit. La Figure 5 indique cependant que la structure AEC//NR donne un  $ERLE$  plus important. Les Figures 6 à 8 correspondent à l'étude en mode DT. Les résultats les plus évocateurs obtenus (avec ou sans adaptation des coefficients de l'anneur d'écho) pour la mesure  $ERLE$  sont représentés Figure 6. La structure AEC+2NR donne la plus grande valeur du  $ERLE$ . Viennent ensuite les structures AEC+NR et AEC//NR. On retrouve le même classement qu'en mode ST si ce n'est que l'anneur d'écho de la structure AEC+NR étant perturbé par la présence de parole, le gain du filtre réducteur de bruit devient plus important rendant cette méthode moins performante.

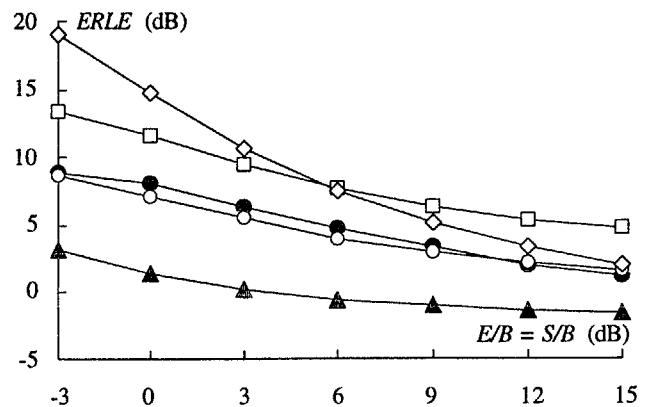
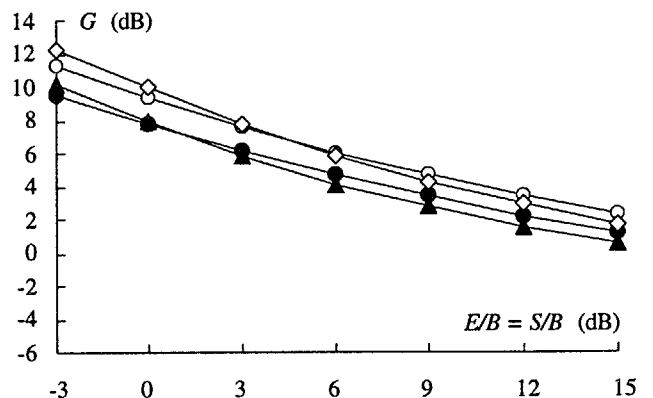
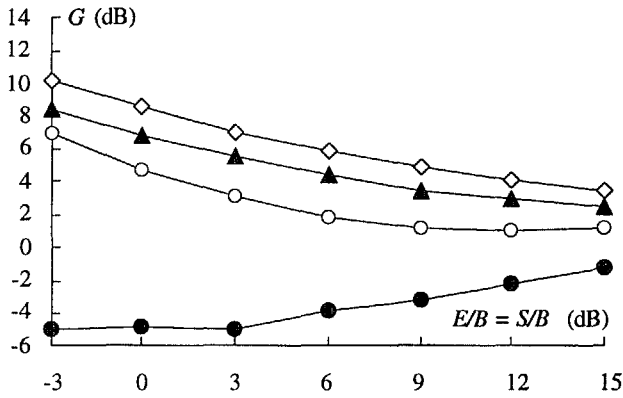


Figure 6.  $ERLE$  en mode DT  
 adaptation continue : ● structure AEC+NR, ▲ structure NR+AEC,  
 ○ structure AEC//NR, ◇ structure AEC+2NR  
 adaptation stoppée : □ structure AEC+2NR

Quant aux résultats sur le gain, il apparaît que, lorsque l'on poursuit l'adaptation, les deux structures nouvelles apportent un gain supplémentaire d'environ 2 dB par rapport aux structures de base (Figure 7.a). Quand on arrête l'adaptation (Figure 7.b), les structures AEC+NR et AEC//NR deviennent nettement moins performantes. Par contre, les structures NR+AEC et AEC+2NR obtiennent des scores plus élevés pour des rapports  $S/B$  et  $E/B$  supérieurs à 6 dB. Pour ces structures ayant un réducteur de bruit placé en tête, il est approprié d'arrêter l'adaptation pour des rapports  $S/B$  suffisants, la parole n'a alors aucune influence sur l'anneur d'écho. Sur la plage des rapports choisis, la structure AEC+2NR est celle donnant le gain le plus élevé, sans adaptation pour des rapports  $S/B$  et  $E/B$  supérieurs à 6 dB, avec adaptation sinon.



7.a. Gain en mode DT avec adaptation

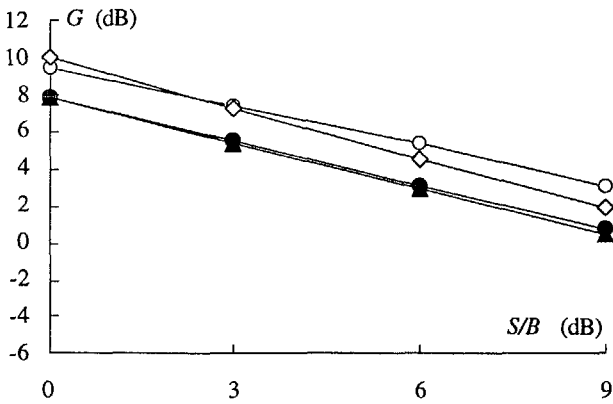


7.b. Gain en mode DT sans adaptation

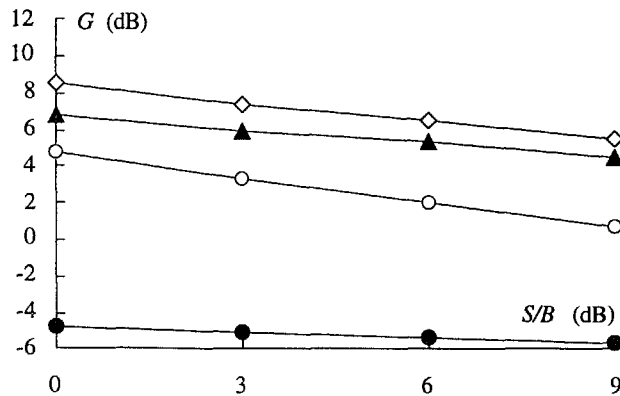
Figure 7. Gain pour  $E/B = S/B$

● structure AEC+NR, ▲ structure NR+AEC, ○ structure AEC//NR, ◇ structure AEC+2NR

Dans un deuxième temps, nous évaluons les performances uniquement en mode DT à rapport  $E/B$  fixé à 0 dB et à rapport  $S/B$  variable (de 0 à 9 dB). Lorsque l'adaptation est maintenue (Figure 8.a), les deux nouvelles structures AEC+2NR et AEC//NR donnent les plus grands gains. Si l'adaptation est arrêtée (Figure 8.b), les gains des structures AEC+NR et AEC//NR chutent. Par contre, les gains des structures NR+AEC et AEC+2NR deviennent plus importants pour des valeurs de  $S/B$  respectivement supérieures à environ 2 dB et 3 dB. Pour les structures où un réducteur de bruit est placé en tête, pour un rapport  $S/B$  suffisant, il est donc intéressant de stopper l'adaptation. La structure donnant le gain le plus fort sur l'ensemble des valeurs  $S/B$  choisies est la structure AEC+2NR, avec arrêt de l'adaptation pour un rapport  $S/B$  supérieur à 3 dB.



8.a. Gain en mode DT avec adaptation



8.b. Gain en mode DT sans adaptation

Figure 8. Gain pour  $E/B = 0$  dB et  $S/B$  variable

● structure AEC+NR, ▲ structure NR+AEC, ○ structure AEC//NR, ◇ structure AEC+2NR

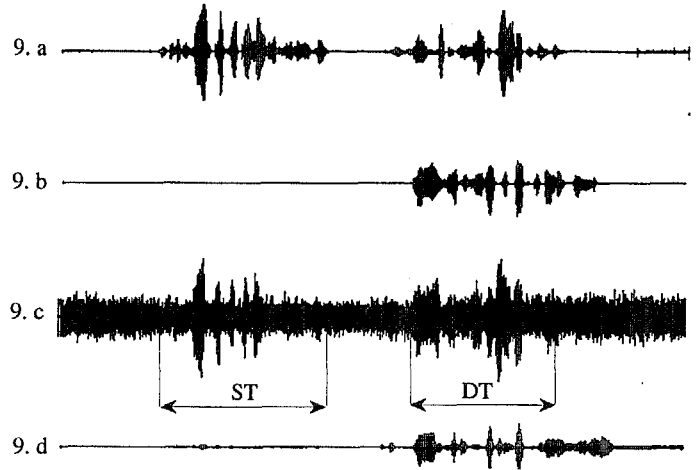


Figure 9. (a) écho, (b) signal pur, (c) signal composite, (d) signal estimé par la structure AEC+2NR

À titre d'exemple, nous présentons Figure 9 l'écho, le signal pur et le signal estimé par la structure AEC+2NR pour  $S/B = E/B = 0$  dB. En mode ST, l'écho est complètement supprimé et en mode DT, le signal utile paraît bien restitué.

### 5. CONCLUSION

Ce papier met en évidence les problèmes majeurs posés par les télécommunications mains-libres, à savoir l'annulation d'écho et la réduction de bruit. Les solutions de base consistent à placer en cascade les systèmes adaptés à chaque problème. Deux nouvelles méthodes sont ici proposées et comparées dans le cadre de la téléphonie mobile. Lorsque l'annuleur d'écho précède le réducteur de bruit, on obtient une bonne annulation d'écho en mode ST si le réducteur de bruit est calculé à partir de son entrée (AEC+NR) tandis que le gain obtenu en mode DT est plus important lorsque le réducteur de bruit est calculé à partir du signal microphone (AEC//NR), ceci se traduit notamment à l'écoute par une distorsion plus faible sur le signal. La structure utilisant deux réducteurs de bruit donne des performances intéressantes dans les deux modes. Pour cette dernière, l'adaptation de l'annuleur d'écho en mode DT doit être poursuivie pour de faibles rapports signal à bruit et peut être bloquée pour des rapports signal à bruit élevés. L'étude des performances doit être étendue à une gamme plus importante de rapports signal à bruit et écho à bruit ainsi qu'à d'autres types de bruit.

### REMERCIEMENTS

Les auteurs remercient Télécom Paris pour l'algorithme GMDF, Matra Communication (Paris) pour la base de données.

### RÉFÉRENCES

[1] R. MARTIN, J. ALTENHÖNER, *Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction*, Proc. ICASSP-95, pp. 3043-3046.  
 [2] H. YASUKAWA, *Acoustic Echo Canceller with Sub-band Noise Cancelling*, Electronics Letters, vol. 28, n°15, July 1992, pp. 1403-1404.  
 [3] E. MOULINES et al., *The Generalized Multidelay Adaptive Filters: Structures and Convergences Analysis*, IEEE Trans. on Signal Processing, vol. 43, n°1, January 1995, pp. 14-28.  
 [4] Y. EPHRAIM and D. MALAH, *Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator*, IEEE Trans. on ASSP, vol. ASSP-32, n°6, December 1984, pp. 1109-1121.