Intérêt de la prise en compte de propriétés auditives en annulation d'écho et débruitage

Valérie Turbin, Christophe Beaugeant, André Gilloire, Pascal Scalart

CNET DIH/CMC
Technopôle Anticipa
2, avenue Pierre Marzin
22307 LANNION Cedex
e-mail : {turbin, beaugean, gilloire, scalart}@lannion.cnet.fr

RÉSUMÉ

La communication avec un système "mains-libres" peut être considérablement altérée par le phénomène de l'écho acoustique et par le bruit ambiant. Aussi des traitements d'annulation d'écho et de débruitage s'avèrent indispensables pour assurer une communication de bonne qualité. Nous présentons une méthode de réduction de ces différentes perturbations à base de filtrage optimal, pouvant remplacer avantageusement les techniques classiques. L'introduction de propriétés auditives - s'appuyant sur un modèle de masquage hybride - dans le principe de filtrage permet de limiter les distorsions générées par celui-ci. Les résultats des simulations démontrent la pertinence de cette démarche, tout particulièrement pour le cas où la perturbation est un signal de parole.

ABSTRACT

Communication with an audio terminal device that operates in hands-free mode may be seriously impaired by the acoustic echo and the ambiant noise. Thus acoustic echo cancellation and noise reduction techniques are required to guarantee the good quality of the communication. In place of classical approaches, we propose a method to reduce these perturbations based on optimal filtering. The introduction of psychoacoustic criteria - which rely on a hybrid masking model - in the filtering technique enables to reduce the amount of distortion generated by the optimal filter. Results of simulations show that the use of human auditory properties in our perturbations reduction method yields effective distortion reduction when the perturbation is speech.

1 Introduction

Les services de communication mettent de plus en plus fréquemment en œuvre des techniques de prise de son de type "mains-libres", que ce soit dans un souci de confort et de sécurité (radiotéléphones de voiture), ou dans le but de conserver le caractère naturel de la communication (téléconférence). Une prise de son éloignée entraîne la superposition au signal de parole utile de perturbations propres à l'environnement, qui sont l'écho acoustique et le bruit ambiant, ce dernier étant particulièrement critique dans le cas des mobiles. La réduction de ces perturbations fait l'objet de nombreuses recherches et des solutions à base de filtrage optimal, ont été proposées dans des contextes radiotéléphonie mains-libres [1]. Cependant, si de telles solutions diminuent de manière notable les perturbations, elles introduisent des distorsions sur le signal de parole utile. Afin de minimiser ces distorsions, la prise en compte de propriétés auditives peut être associée aux traitements.

Dans cet article, nous présentons l'impact de l'intégration de propriétés psychoacoustiques dans certains traitements de type filtrage optimal, dont la description fait l'objet du second paragraphe. La prise en compte de propriétés auditives passe typiquement par l'utilisation d'un modèle de masquage simultané. Nous présentons le modèle retenu et discutons ce choix dans le troisième paragraphe. Finalement, nous montrons, dans deux contextes étudiés, la téléconférence et la radiotéléphonie mains-libres, que l'intervention de contraintes exploitant la perception auditive permet de réduire de manière notable la distorsion.

2 Principe du filtrage utilisé

L'estimation linéaire $\hat{S}_w(f)$ du signal de parole utile, $S(f)$, minimisant l'erreur quadratique moyenne $\mathrm{E}[(S(f)-\hat{S}_w(f))^2]$ dans le domaine fréquentiel est obtenue suivant le principe du filtrage de Wiener. Une expression générale de la réponse en fréquence du filtre est donnée par :

$$G(m,f) = \frac{RSP(m,f)}{1+RSP(m,f)} \qquad (1)$$

$RSP(m,f)$ désigne le rapport entre la densité spectrale de puissance (dsp) du signal utile et celle(s) de la (ou des) perturbation(s) de la trame m. Le calcul du rapport $RSP(m,f)$ est basé sur un principe exploité en débruitage [2] :

$$RSP(m,f) = \beta.\frac{\left|\hat{S}_W(m-1,f)\right|}{\hat{\gamma}_P(m,f)} + (1-\beta).RSP_{post}(m,f) \qquad (2)$$

Dans l'expression (2), $\hat{\gamma}_P(m,f)$ désigne une estimée de la dsp de la (des) perturbation(s) et $RSP_{post}(m,f)$ le Rapport Signal à Perturbation a posteriori défini par :

$$RSP_{post}(m,f) = \frac{\left|Y(m,f)\right|^2}{\hat{\gamma}_P(m,f)} - 1 \qquad (3)$$

où $Y(m,f)$ désigne la Transformée de Fourier à Court Terme (TFCT) de la trame *m* du signal microphonique. Le paramètre β peut être ajusté suivant les caractéristiques de la perturbation ou simplement fixé à une constante dont une valeur typique est 0,98.
L'inconvénient majeur de cette technique est de filtrer le signal utile qui peut alors subir des distorsions nuisant ainsi à la qualité de la restitution de ce signal.

3 Introduction de contraintes auditives

Un son en présence d'un autre peut devenir partiellement ou complètement inaudible : c'est ce que l'on appelle l'effet de masque [3]. Lorsque le signal utile masque la (ou les) perturbation(s), le traitement est inutile. Ne pas effectuer ce dernier permet alors de limiter les dégradations du signal utile sans pour autant réduire les performances du filtrage. Nous nous limitons au masquage *simultané* intervenant dans le domaine fréquentiel. L'exploitation de celui-ci passe par le calcul de seuils de masquage déterminés à partir d'un modèle représentant le comportement de l'oreille interne à une excitation donnée et d'informations spectrales à court terme du signal masquant.

3.1 Choix du modèle de masquage

Nous nous sommes intéressés au modèle de Johnston [4] et à celui de la Norme ISO MPEG Psychoacoustic Model II [5] pour lesquels les différentes étapes nécessaires au calcul du seuil de masquage sont décrites succinctement sur la figure 1.

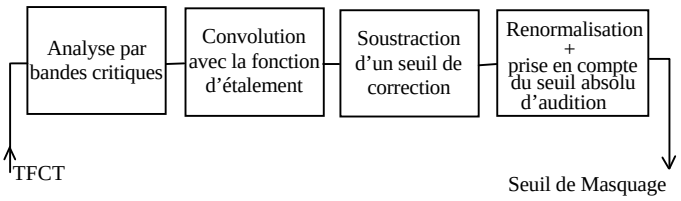

Figure 1. Etapes de calcul du seuil de masquage.

Le procédé de calcul est similaire pour les deux modèles qui se distinguent par l'application d'une règle de correction différente, dont une expression générale est:

$$correction(b) = \alpha(b)\, TMB(b) + (1-\alpha(b))\, BMT(b) \qquad (4)$$

où *b* désigne la fréquence en Bark, *TMB(b)* la valeur de correction à appliquer dans le cas d'une tonale masquant un bruit, *BMT(b)* la valeur de correction dans le cas d'un bruit masquant une tonale et $\alpha(b)$ l'indice de tonalité.

Les deux modèles préconisent un seuil de correction *BMT(b)* identique et constant de 5,5 dB. *TMB* est variable suivant la bande critique, et ses valeurs proposées dans le modèle ISO sont supérieures à celles du modèle de Johnston, en particulier aux basses fréquences. De plus, une méthode de calcul coûteuse est employée dans le modèle ISO pour obtenir un indice de tonalité variant avec la bande critique. A l'inverse, le modèle de Johnston propose une méthode nettement moins complexe de calcul d'un indice de tonalité constant sur toutes les bandes critiques.

Le modèle qui a été finalement retenu est un modèle "hybride" dans le sens où nous appliquons le modèle de Johnston dont les valeurs *TMB(b)* ont été remplacées par celles du modèle ISO. Ce choix permet d'améliorer les performances de masquage du modèle de Johnston dans les basses fréquences tout en conservant une complexité raisonnable.

3.2 Filtrage sous contraintes psychoacoustiques

Le seuil de masquage sur une trame *m*, *T*(*m*,*f*), est obtenu à partir de l'estimée $\hat{S}_W(m,f)$. Ce seuil permet de discriminer les composantes fréquentielles non masquées des composantes masquées qui correspondent au cas où les perturbations sont rendues inaudibles par la présence du signal utile. Pour ces composantes, le filtre est forcé à 1. L'estimation du signal de parole utile sous contraintes psychoacoustiques, $\hat{S}_{Wm}(m,f)$, est ainsi donnée par :

$$\hat{S}_{Wm}(m,f) = \begin{cases} U(m,f) & \text{si } \hat{\gamma}_P(m,f) \le T(m,f) \\ \hat{S}_W(m,f) & \text{sinon} \end{cases} \qquad (5)$$

où $U(m,f)$ désigne la TFCT du signal à filtrer par *G*. Cette méthode a déjà été proposée en débruitage [6].

4 Applications

Nous avons appliqué le principe décrit dans les paragraphes précédents à deux contextes, celui de la téléconférence et celui de la radiotéléphonie mains-libres. La nature des perturbations nous a amené à développer deux mises en œuvre pratiques prenant en compte les spécificités des contextes considérés.

4.1 Contexte de la téléconférence

La perturbation particulièrement critique dans ce contexte est l'écho acoustique. Pour traiter ce problème, une approche similaire à celle proposée dans [7] est utilisée au lieu d'une annulation d'écho classique (filtre adaptatif de grande taille identifiant la réponse impulsionnelle de couplage). Le système, schématisé sur la figure 2, est composé d'un annuleur d'écho de taille réduite (identifiant approximativement l'onde directe et les premières réflexions de la réponse de couplage) et d'un post-filtre *G*.

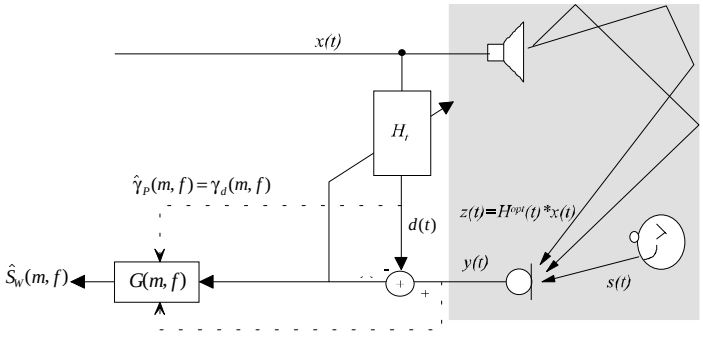

Figure 2. Principe du filtrage en contexte de téléconférence.

Le rôle du post-filtre est d'atténuer l'écho résiduel, c'est-à-dire la contribution de l'écho encore présente après l'annulation de l'écho. Différents types de post-filtrage ont été testés et permettent d'aboutir à de très bonnes performances en termes de réduction d'écho [8]. Une mise en oeuvre possible consiste à utiliser le principe décrit au premier paragraphe. Une estimée de la perturbation est obtenue en utilisant l'écho estimé par l'annuleur d'écho, soit :

$$\hat{\gamma}_p(m,f) \approx \hat{\gamma}_d(m,f) \qquad (6)$$

Notons qu'une autre possibilité consiste à utiliser une estimation de l'écho résiduel (effectuée à partir des signaux *e(t)* et *x(t)*) plutôt que l'estimation donnée par l'annuleur d'écho. Une spécificité du traitement consiste à filtrer non pas le signal microphonique mais le signal *e(t)*. La combinaison d'un annuleur d'écho de taille réduite et d'un post-filtre obéissant à ce principe permet d'obtenir de très bonnes performances en termes de réduction d'écho. Avec l'approche classique, un filtre adaptatif de plusieurs milliers de coefficients serait nécessaire pour obtenir des performances équivalentes. Mais, à la différence d'une annulation d'écho classique, cette méthode génère des distorsions sur le signal utile qui peuvent être limitées en associant les contraintes psychoacoustiques au post-filtrage données par (5).

4.2 Contexte des radiotéléphones mains-libres

Les traitements dans ce contexte se doivent de considérer deux perturbations distinctes : le bruit ambiant et l'écho. On obtient alors :

$$\hat{\gamma}_{PP}(m,f) = \hat{\gamma}_Z(m,f) + \hat{\gamma}_B(m,f) \qquad (7)$$

où $\hat{\gamma}_Z(m,f)$ (resp. $\hat{\gamma}_B(m,f)$) est l'estimée de la dsp de l'écho (resp. du bruit). Le filtrage, correspondant aux équations (1) à (3) et schématisé par la figure 3, cherche ainsi à annuler à la fois l'écho et le bruit.

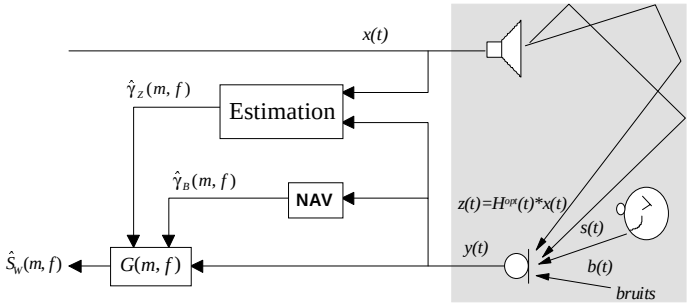

Figure 3. Principe du filtrage en contexte mobile.

Par rapport à une association de deux traitements séparés (un filtre pour l'annulation d'écho et un pour le débruitage)[9], le coût de calcul est fortement diminué mais la distorsion sur le signal utile est augmentée. Cette augmentation est en partie due au fort biais des estimateurs des perturbations, calcul de la dsp du bruit pendant les périodes de non-activité vocale (NAV), de celle de l'écho à partir des spectres des signaux microphonique et haut-parleur. L'introduction des propriétés de masquage en sortie de traitement (équation (5)) a pour but de limiter les distorsions et de se rapprocher des performances des systèmes à deux filtres avec un coût de calcul réduit.

5 Résultats expérimentaux

Les simulations ont été effectuées à l'aide de corpus de signaux permettant de reproduire des situations réalistes dans les deux contextes d'application. Pour la téléconférence, des conditions identiques à celles de [8] ont été choisies : salle acoustiquement traitée, rapport Signal à Echo de l'ordre de 3-5 dB, atténuation fournie par l'annuleur d'environ 12 dB laissant un écho résiduel audible et génant avant le post-filtrage. En contexte mobile, des conditions d'une voiture en mouvement ont été simulées avec un rapport Signal à Echo de l'ordre de 3-5 dB et un rapport Signal à Bruit du même ordre. Le but des simulations est de comparer en termes de distorsion le filtrage sous contraintes (résultat de l'équation (5)) au filtrage sans contraintes. Rappelons que le filtrage sans contraintes permet de rendre en sortie de système l'écho très peu perceptible et de maintenir, dans le cas des mobiles, un bruit de confort mais, parallèlement, génère des distorsions audibles. La distorsion apportée par un traitement est évaluée par la distance cepstrale entre le signal de parole utile et le même signal sur lequel a été effectuée une recopie du traitement. Des tests d'écoute ont également été utilisés pour qualifier subjectivement l'amélioration apportée.

Il s'avère que le masquage diminue peu la distorsion lorsque le perturbateur est un bruit ambiant comme l'illustre la figure 4. On retrouve ici des conclusions similaires à [10].

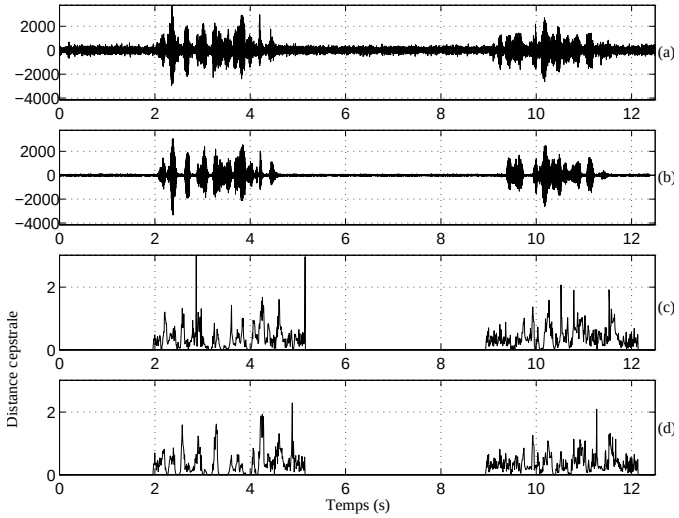

Figure 4. Résultats obtenus en débruitage (contexte mobile)
(a) Signal temporel bruité.
(b) Signal temporel débruité.
(c) Distance cepstrale, système sans masquage.
(d) Distance cepstrale, système avec masquage.

Par contre, dans le cas où le signal perturbateur est l'écho (contexte radiotéléphonie) ou bien l'écho résiduel (téléconférence), les améliorations apportées sont appréciables, comme l'illustrent les exemples des figures 5

et 6. Les tests d'écoute effectués confirment que la distorsion en sortie des systèmes avec masquage est très peu perceptible.

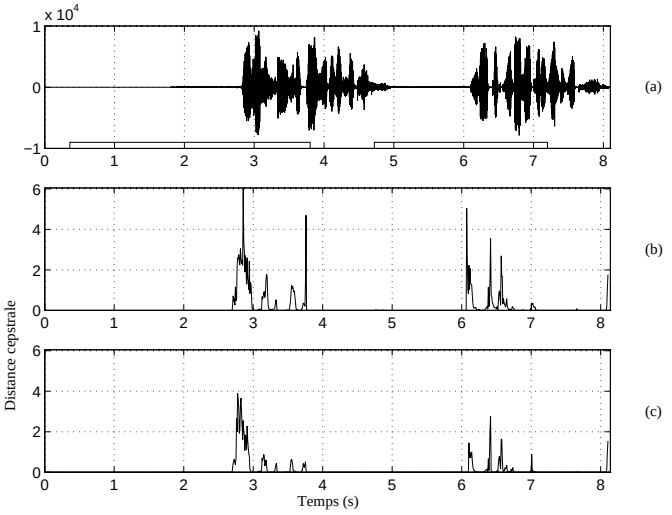

Figure 5. Résultats obtenus en annulation d'écho (contexte téléconférence)
(a) Signal de parole utile avec présence de l'écho signalée par la courbe en escalier.
(b) Distance cepstrale, système sans masquage.
(c) Distance cepstrale, système avec masquage.

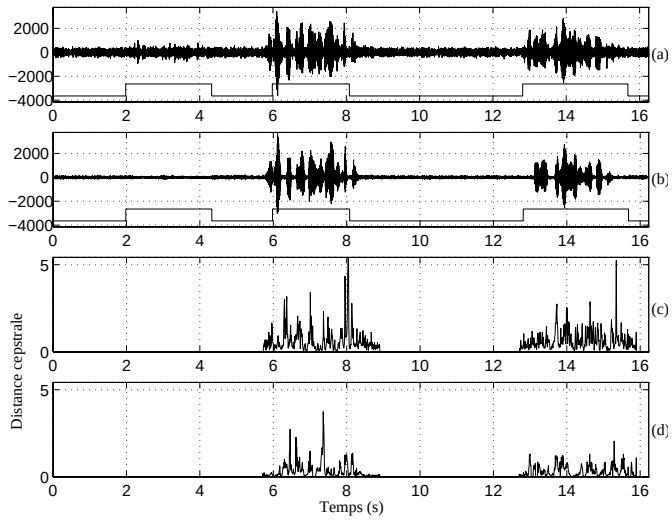

Figure 6. Résultats obtenus en annulation d'écho (contexte mobile)
(a) Signal temporel bruité avec présence de l'écho signalée par la courbe en escalier.
(b) Signal temporel débruité et débarrassé de l'écho.
(c) Distance cepstrale, système sans masquage.
(d) Distance cepstrale, système avec masquage.

L'amélioration apportée par les propriétés de masquage semble être liée à la nature spectrale des perturbateurs. En effet, si le spectre du perturbateur est de la même nature que celui du signal masquant, le filtrage de Wiener apportera une distorsion plus importante que si les spectres sont disjoints. Aussi, la réduction d'écho par le filtre décrit en 2 apportera une dégradation plus importante sur le signal utile que celle générée par la réduction de bruit. L'apport des propriétés de masquage sera donc plus appréciable dans le premier cas que dans le second, ce que retrouvent les résultats expérimentaux.

6 Conclusion

La prise en compte de propriétés auditives permet d'effectuer un filtrage sélectif limité aux fréquences où les perturbations ne sont pas masquées, ce qui amène une diminution de la distorsion du signal utile. La simplicité du modèle "hybride" proposé ici permet de plus d'envisager des applications peu coûteuses en temps de calcul. Les simulations réalisées montrent que les traitements intégrant des propriétés auditives permettent de réduire effectivement la distorsion et que le modèle utilisé est pertinent en particulier dans le cas de l'annulation de l'écho acoustique.

7 Références

[1] R. MARTIN, P. VARY, "*Combined acoustic echo control and noise reduction for hands-free telephony - State of the art and perspectives.*", Signal Processing VIII, Trieste, pp 1107-1110, 1996.

[2] Y. EPHRAIM, D. MALAH, "*Speech enhancement using optimal non-linear spectral amplitude estimation*", ICASSP'83, Boston, pp 1118-1121.

[3] E. ZWICKER, R. FELDTKELLER, "*Das Ohr als Narichtenempfänger*" ou "*Psychoacoustique. L'oreille récepteur d'information.*", Stuttgart, West Germany : Hirzel Verlag, 1967 (MASSON 1981, traduit de l'allemand par Christel SORIN).

[4] J. D. JOHNSTON, "*Transform coding of audio signals using perceptual noise criteria*", IEEE Journal on selected areas in communications, vol. 6, n°2, pp. 314-323, February 1988.

[5] Projet de Norme Internationale ISO 11172-3 MPEG Audio, Londres, Novembre 1992.

[6] D. TSOUKALAS, M. PARASKEVAS, J. MOURJOPOULOS, "*Speech enhancement using psychoacoustic criteria*", ICASSP'93, Minneapolis, pp. II.359-II.362.

[7] R. MARTIN, J. ALTENHONER, "Coupled adaptive filters for acoustic echo control and noise reduction", ICASSP'95, Detroit, USA, pp. 3043-3046.

[8] V. TURBIN, A. GILLOIRE, P. SCALART, "*Comparison of three post-filtering algorithms for residual acoustic echo reduction*", ICASSP'97, Munich, vol. 1, pp. 307-310.

[9] Y. GUELOU, A. BENAMAR, P. SCALART, "*Analysis of two structures for combined acoustic echo cancellation and noise reduction*", ICASSP'96, Atlanta, pp. 637-640.

[10] A. AKBARI AZIRANI, "*Réhaussement de la parole en ambiance bruitée. Application aux télécommunications mains-libres*", Thèse de doctorat de l'Université de Rennes I, 1995.