

## DEBRUITAGE PAR ANTENNE ACOUSTIQUE

Min XU, Yves GRENIER

Ecole Nat. Sup. des Télécommunications - Dépt SIGNAL, CNRS-URA 820  
46 rue Barrault, 75634 PARIS CEDEX 13, FRANCE.

### RESUME

Nous proposons pour le débruitage une antenne acoustique fonctionnant avec une référence bruit seul. Le principe consiste à réaliser séparément l'estimation et l'annulation du bruit. Le filtrage adaptatif monodimensionnel est utilisé pour estimer le bruit de chaque capteur. La formation de voie annule le bruit estimé avec beaucoup de souplesse. Les réalisations fréquentielle et temporelle de l'antenne sont présentées en détails. Ses performances ont été analysées en comparaison avec des techniques existantes. L'étude montre que l'antenne proposée a une meilleure performance réelle que le débruitage par soustraction classique. Dans un milieu réverbérant, l'atténuation sur le bruit est également plus efficace que celle de l'antenne de Frost.

### SUMMARY

We propose an acoustical antenna which operates with a noise-only reference to cancel noise. It separately estimates and cancels the noise. Monodimensional adaptive filtering is used to estimate the noise for each sensor. The estimate of the noise is then canceled by beamforming techniques. The frequency and temporal realizations of the antenna are detailed. Its performances are analyzed by comparison with classical techniques. It is shown that the real performance is better in the proposed antenna than in classical noise cancellation which subtracts the noise estimate. In a reverberant room, the cancellation is also more efficient than that of Frost's antenna.

### 1. INTRODUCTION

Dans des applications comme la prise de son, le débruitage est un problème préoccupant. Les techniques adaptatives existant qui sont destinées à résoudre ce problème perdent souvent leur performances en raison de la complexité des circonstances réelles. Lorsque l'on dispose d'une référence du bruit, le filtrage adaptatif permet, en filtrant la référence, d'estimer le bruit et de l'annuler par soustraction [1]. L'inconvénient de cette technique est que la performance est assez sensible aux erreurs de l'estimation du bruit. Une autre technique consiste à utiliser le principe de l'antenne de Frost [2]. Elle est capable d'atténuer le bruit en connaissant seulement la position de la source. Dans un milieu réverbérant, l'atténuation obtenue sur le bruit est malheureusement insuffisante. Nous proposons, en combinant les deux techniques, une antenne acoustique fonctionnant avec une référence. Nous allons montrer qu'une telle antenne est performante même en milieu réverbérant et qu'elle est moins sensible à l'erreur d'estimation du bruit.

Les traitements effectués par l'antenne en question sont présentés en détails dans le paragraphe 2. Une discussion sur les propriétés de cette méthode est faite dans le paragraphe 3 en comparaison avec l'annulation par soustraction d'une part, et avec l'antenne de Frost d'autre part. Le dernier paragraphe présente quelques résultats de simulations permettant

d'illustrer ses performances.

### 2. PRESENTATION DE L'ANTENNE

On utilise un réseau de  $N$  capteurs distribués dans l'espace, qui fournit un signal vectoriel  $X$  de  $N$  composantes. Supposons que les distances  $l_i$  entre la source et les capteurs soient connues. Il suffit pour cela de connaître leur position. Supposons également que l'on dispose d'une référence bruit seul  $b'$  qui est un signal scalaire.

L'antenne que nous proposons est composée de deux étages (voir la figure 1). Le premier a pour objectif d'estimer le bruit  $b$  qui est présent dans le signal capté en exploitant  $X$  et  $b'$ . Le deuxième étage est la formation de voie qui permet d'obtenir un estimateur du signal au moyen d'un filtrage sur  $X$ . L'optimisation des filtres s'effectue à l'aide de l'estimateur du bruit  $\hat{b}$ , obtenu dans l'étage précédent, et des distances entre la source et les capteurs. Elle permet d'annuler le bruit et d'améliorer, en réduisant la réverbération, la qualité du signal. La réalisation fréquentielle des traitements est détaillée dans 2.1 et 2.2. On présentera brièvement la réalisation temporelle dans 2.3.

#### 2.1 Estimation du bruit

A l'aide de la Transformation de Fourier à court terme, les signaux sont d'abord convertis par fenêtre dans le



domaine fréquentiel. On estime le bruit fréquence par fréquence. Prenons une fréquence donnée  $m$  dans la  $p$ -ième fenêtre. On considère que le signal capté  $X_p(m)$  est égal à la somme du bruit  $B_p(m)$  et de  $X'_p(m)$ , où  $X'_p(m)$  est composé du signal et des bruits divers, comme la réverbération du signal:

$$X_p(m) = B_p(m) + X'_p(m).$$

La référence  $b'$  de  $B$  est supposée décorrélée avec  $X'$ . On sait dans ce cas que de nombreux algorithmes permettent l'estimation linéaire du bruit. Il suffit pour cela de préciser la structure du modèle linéaire  $L$  qui transforme  $b'$  en  $\hat{B}_p(m)$ :  $\hat{B}_p(m) = L(b')$ , en sachant qu'il représente le système qui lie  $b'$  à  $B_p(m)$ .

Le modèle classique est un gain complexe  $H(m)$  qui représente la valeur de la fonction de transfert du système:

$$\hat{B}_p(m) = H(m) b'_p(m).$$

Dans la plupart des cas, ce modèle est satisfaisant quand on peut utiliser des fenêtres de longueur supérieure à celle de la réponse impulsionnelle. Lorsque cette condition n'est pas remplie, il faut employer des modèles plus complexes. L'annulation d'écho acoustique en audio-conférence est un tel exemple. Le système (canal acoustique) a une réponse très longue, et le retard interdit l'utilisation des fenêtres longues. On propose le modèle, dit "à double convolution", qui a la structure suivante [3]:

$$\hat{B}_p(m) = \sum_k \sum_n H'_n(k, m) b'_{p-n}(k)$$

où les  $H'_n(k, m)$  représentent les valeurs des fonctions de transfert à court terme dans une bande centrée en  $m$ .

## 2.2 Formation de voie

La formation de voie se réalise avec un filtrage sur chaque composante du signal capté  $X_p(m)$  et une sommation des signaux filtrés. Notons  $A_p(m)$  la valeur à la fréquence  $m$  de la fonction de transfert des filtres. La sortie est alors égale au produit scalaire de  $A_p$  et de  $X_p$ . On la considère comme un estimateur du signal:

$$\hat{s}_p(m) = A_p^{*T}(m) X_p(m).$$

### 2.2.1 optimisation du filtre $A_p$

Supposons que  $X_p(m)$  s'écrit sous la forme suivante:

$$X_p(m) = s_p(m)D(m) + B_p(m) + W(m)$$

avec:

$s_p(m)$  : signal utile (scalaire)

$W(m)$  : bruit indépendant de  $B(m)$

$D(m)$  : vecteur positionnel de la source

$$D(m) = [\dots, 1/l_i \exp(-j2\pi m l_i/c), \dots].$$

On propose d'utiliser le critère d'optimisation, sous contrainte, qui minimise le bruit estimé:

$$\text{Min}_{A_p(m)} |A_p^{*T}(m) (r\hat{B}_p(m) + W(m))|^2 \text{ avec } A_p^{*T}(m)D(m)=1$$

où  $r$  est un paramètre positif. Notons  $Q(m)$  la covariance de  $W(m)$ . Le filtre optimal correspondant est égal à:

$$A_p(m) = \frac{R^{-1}(m) D(m)}{D^{*T}(m) R^{-1}(m) D(m)}$$

avec  $R(m) = r^2 \hat{B}_p(m) \hat{B}_p(m)^{*T} + Q(m)$ .

On remarque que seuls  $r$  et  $Q(m)$  sont encore inconnus.  $r$  est un paramètre, pour lequel il convient de prendre une valeur grande devant 1.  $Q(m)$  peut être déterminé par une estimation préalable. Lorsqu'il s'agit de réverbération qui a un champ très diffus, on peut utiliser approximativement la matrice identité.

### 2.2.2 propriétés

On étudie ici quelques propriétés de cette méthode dans le cas où  $Q(m)=I$ .

#### 1. Quasi-orthogonalité

Il est facile de vérifier que le gain sur le bruit est une fonction homographique de  $r^2$ :

$$|A_p^{*T}(m) \hat{B}_p(m)| = \frac{\alpha}{1 + \beta r^2}$$

où  $\alpha$  et  $\beta$  dépendent de  $B_p(m)$  et de  $D(m)$ . Il décroît avec  $r$  et tend vers 0. Pour régler le gain, il suffit donc de choisir convenablement  $r$ . Pour les valeurs de  $r$  suffisamment grandes, le vecteur filtre est quasiment orthogonal au vecteur bruit. On remarque qu'il s'agit ici d'une orthogonalité spatiale, puisque les différentes composantes de  $B_p(m)$  correspondent au même bruit, capté en différentes positions.

#### 2. cône d'amplification

Pour étudier le comportement général de l'antenne, on utilise son gain relatif. On le définit pour un signal  $Y$  par rapport au signal utile:

$$g_r(Y) = (|A_p^{*T}(m)Y| \cdot ||D(m)||) / (|A_p^{*T}(m)D(m)| \cdot ||Y||)$$

Le volume défini par l'inégalité  $g_r(Y) > 1$  est un cône d'axe  $A_p(m)$  et de côté  $D(m)$ . On l'appelle "cône d'amplification". On peut démontrer que le filtre optimal  $A_p(m)$  fait un angle aigu avec  $D(m)$ , ce qui signifie que le cône d'amplification contient un faible volume de l'espace. Lorsque l'angle est petit, on peut réaliser une atténuation de tous les signaux par rapport au signal utile. Il est démontré également que l'angle croît avec  $r$ . Pour réaliser l'atténuation générale, il faut donc limiter la valeur de  $r$ .

#### 3. cas de bruits multiples

L'optimisation du filtre proposée se généralise sans difficulté au cas où il y a plusieurs bruits. L'étude sur la performance de l'antenne a montré que les atténuations des bruits obtenues sont quasiment indépendantes du nombre de bruits, lorsque ce dernier est inférieur au nombre de capteurs.

## 2.3 Réalisation temporelle

Nous venons de présenter l'antenne à deux étages dans le domaine fréquentiel. Il est également possible de la réaliser dans le domaine temporel. Dans ce cas nous conservons la structure à deux étages, mais nous introduirons des constantes de temps différentes, en travaillant sur un horizon de taille

croissante pour l'étage 1, et sur un horizon de taille fixée pour l'étage 2. Cette taille fixée peut s'obtenir soit dans un traitement par bloc ou par fenêtre glissante, soit par un traitement adaptatif à gain constant. On verra dans 3.1 que ceci peut être utile dans certains cas pour augmenter l'atténuation sur le bruit.

L'étage 1 est l'estimation du bruit à partir du signal capté et de la référence  $b'(t)$ . L'estimateur linéaire a pour forme:  $\hat{b} = h * b'$ , où  $h$  représente la réponse impulsionnelle du système du bruit. Les méthodes classiques de filtrage adaptatif sont applicables, en travaillant échantillon par échantillon.

Dans l'étage 2, si le filtrage du signal capté s'effectue par bloc de  $N$  échantillons, nous notons les fenêtres de découpage  $F_p = [pN, (p+1)N - 1]$ . A l'intérieur de chaque fenêtre, les coefficients du filtre sont ajustés suivant un critère d'optimisation local. Enfin, on filtre tout le bloc avec le filtre adapté. On note  $a_i^p$  la réponse impulsionnelle du filtre associé au  $i$ -ième capteur et  $d_i$  la transformée de Fourier inverse de la  $i$ -ième composante de  $D$ . Le résultat de la formation de voie s'écrit:  $\hat{s} = \sum_i a_i^p * x_i$ , soit avec les notations vectorielles:

$$A_p = \begin{bmatrix} a^p(0) \\ \vdots \\ a^p(q-1) \end{bmatrix}, \quad X(t) = \begin{bmatrix} x(t) \\ \vdots \\ x(t-q+1) \end{bmatrix}, \quad \hat{B}(t) = \begin{bmatrix} \hat{b}(t) \\ \vdots \\ \hat{b}(t-q+1) \end{bmatrix}$$

$$\hat{s}(t) = A_p^T X(t), \text{ pour } t \in F_p.$$

Le critère local utilisé est la minimisation sous contrainte de la moyenne quadratique du bruit, calculée sur la fenêtre:

$$\text{Min}_{A_p} \sum_{t \in F_p} (A_p^T \hat{B}(t))^2 \text{ avec } \sum_i a_i^p * d_i = \delta_\tau.$$

Avec les notations vectorielles, la contrainte s'écrit:  $A^T C = [0 \dots 0 \ 1 \ 0 \dots 0]$ , où  $C$  est une matrice qui regroupe les  $d_i$ . Si  $R$  est la matrice de covariance du vecteur  $\hat{B}(t)$  sous la fenêtre  $F_p$ , la solution optimale  $A_p$  est:

$$A_p = R^{-1} C (C^T R^{-1} C)^{-1} [0 \dots 0 \ 1 \ 0 \dots 0]^T$$

Une seconde manière d'adapter la formation de voie est de travailler échantillon par échantillon avec un algorithme de gradient sous contrainte qui peut être déduit de façon analogue à celui de Frost [2] en remplaçant  $X$  par  $\hat{B}$ :

$$\begin{cases} A(0) = F \\ \hat{b}_s(t) = A(t)^T \hat{B}(t) \\ \hat{s}(t) = A(t)^T X(t) \\ A(t+1) = P(A(t) - \mu \hat{b}_s(t) \hat{B}(t)) + F \end{cases}$$

avec:  $F = C (C^T C)^{-1} [0 \dots 0 \ 1 \ 0 \dots 0]^T$   $P = I - C (C^T C)^{-1} C^T$ .

### 3. COMPARAISON AVEC DES METHODES EXISTANTES

#### 3.1 Avec l'antenne de Frost

On rappelle que l'antenne de Frost minimise sous la même contrainte la quantité  $E((A^T X)^2)$  qui est l'énergie de la sortie de l'antenne, avec  $X = S + X''$ . Les points suivants la différencient de l'antenne à deux étages:

1. Lorsque  $X''$  contient uniquement le bruit  $B$  qui est décorrélié du signal  $S$ , le critère est équivalent à la minimisation du bruit à la sortie:  $E((A^T B)^2)$ . Par contre, quand  $X''$  contient une partie corrélée avec le signal, ce qui est le cas de la réverbération du signal, il n'y a plus d'équivalence. Dans la pratique, l'antenne risque d'atténuer fortement le signal. Par contre, l'antenne à deux étages minimise dans tous les cas le bruit estimé:  $E((A^T B)^2)$  en exploitant la référence bruit seul  $b'$ . L'annulation du signal sera désormais évitée.
2. Frost utilise un critère global sur toute la durée du signal. Au cas où la réponse impulsionnelle du système bruit  $h$  est très longue par rapport au filtre  $a$ , le bruit est perçu par l'antenne comme un grand nombre de bruits apparents qui sont indépendants entre eux [4]. Le filtre converge alors vers une solution qui atténue moyennement tous les bruits apparents, ce qui correspond à une atténuation totale médiocre. Dans l'antenne que nous venons de proposer, le deuxième étage utilise un critère local adaptatif qui permet d'annuler le bruit sur une échelle plus fine. Ainsi le premier étage suit le système du bruit, et le second étage suit les variations du bruit.

#### 3.2 Avec la soustraction

Après avoir estimé le bruit dans le premier étage, on aurait pu, au lieu d'effectuer la formation de voie, soustraire directement le bruit estimé:  $X(m) - \hat{B}(m)$ . C'est la méthode de débruitage classique qui a besoin d'un seul capteur (et d'une référence). Mais elle n'a pas la même robustesse aux erreurs de l'estimation du bruit. En effet, l'emploi de capteurs multiples en formation de voie permet d'exploiter les déphasages aléatoires des erreurs. Supposons que le bruit a été estimé avec une erreur de  $\Delta B(m)$ . Notons respectivement  $g_1$  et  $g_2$  les gains du rapport signal à bruit pour la soustraction et pour la formation de voie, on peut montrer que:

$$g_1 / g_2 \approx g_r(\Delta B(m))$$

où  $g_r$  est le gain relatif de l'antenne défini dans 2.2. Nous savons d'après la discussion faite que le cône d'amplification où  $g_r > 1$  est un faible volume dans l'espace. Si l'on suppose que l'erreur a une distribution uniforme dans l'espace, il y a une forte probabilité pour que  $g_r(\Delta B)$  soit inférieur à 1. Autrement dit, le gain du rapport signal à bruit obtenu en présence d'une erreur est probablement plus élevé en formation de voie qu'en soustraction.

Par ailleurs, la formation de voie est plus souple à utiliser: le gain (ou atténuation) sur le bruit est contrôlable avec un seul paramètre  $r$ ; il est possible de diminuer la sensibilité aux erreurs en réduisant avec  $r$  l'angle du cône d'amplification. Lorsque le débruitage est utilisé pour la prise de son dans un milieu réverbérant, la formation de voie a pour effet de déréverbérer le signal. L'expérience montre que la parole qui sort de l'antenne a une intelligibilité plus élevée.



### 3.3 Cas où la référence est corrélée avec le signal

Les trois méthodes discutées ont toutes pour hypothèse de base la décorrélation entre la référence  $b'$  et le signal  $s$ . Il est intéressant de savoir s'il y a une différence entre elles lorsqu'une corrélation se présente. D'après une analyse que nous avons faite sur l'erreur du filtre optimal, elles ont le même comportement asymptotique. En effet, pour toutes les trois, le signal estimé s'écrit en première approximation:

$$\hat{S}(m) = S(m) - \frac{E(S(m)B'^*(m))}{E(|B'(m)|^2)} B'(m)$$

Ceci est probablement dû au fait que ces méthodes utilisent des critères quadratiques qui ne permettent pas de discriminer les signaux corrélés entre eux.

## 4. SIMULATION

On a effectué les simulations pour l'annulation d'écho acoustique en audio-conférence. L'antenne acoustique est construite avec 9 microphones uniformément répartis sur un demi-cercle de diamètre 1,7m. Le signal et le bruit sont tous les deux de la parole échantillonnée à 16 kHz. Ils ont le même niveau d'énergie à l'entrée de l'antenne. Le système du bruit  $H$  est un canal acoustique de longueur 0,5s.

Les traitements sont réalisés dans le domaine fréquentiel. Le signal capté est converti par DFT sur 512 points avec un recouvrement de 3/4. Dans le premier étage, le système bruit est représenté par un modèle à double convolution. On applique pour chaque fréquence et pour chaque capteur le critère de moindre carré. Le paramètre  $r$  du second étage vaut 10. On a constaté que la convergence est atteinte au bout de 1 seconde. Les résultats montrés dans les figures 2 et 3 ont été obtenus à la convergence.

La figure 2 montre les spectres, moyennés sur 4s, du signal et du bruit obtenus à la sortie de l'antenne:

$$S_S(m) = \sum_p |A_p^{*T}(m) S_p(m)|^2, \quad S_B(m) = \sum_p |A_p^{*T}(m) B_p(m)|^2$$

Le bruit est inférieur au signal à chaque fréquence. L'écart est aux alentours de 25dB. Dans les tests auditifs, le bruit est complètement masqué par le signal.

Dans la figure 3, on a présenté les rapports signal à bruit obtenus avec la formation de voie et avec la soustraction. Ils sont définis respectivement par:

$$S_S(m) / S_B(m), \quad \left( \sum_p |S_p(m)|^2 \right) / \left( \sum_p |S_p(m) - \hat{S}_p(m)|^2 \right)$$

Le rapport est plus élevé pour la formation de voie (27dB contre 23dB). A l'écoute, l'effet de déréverbération de l'antenne est marqué par une impression de rapprochement du locuteur.

## REFERENCES

- [1] B. Widrow et al., "Adaptive Noise Cancelling: Principles and Applications", Proc. IEEE 63, 1975.
- [2] O.L. Frost III, "An algorithm for Linearly Constrained Adaptive Array Processing", Proc. of IEEE, Aug. 1972

- [3] M. XU and Y. GRENIER, "Time-Frequency Domain Adaptive Filters", ICASSP 89
- [4] M.M. Sondhi and G.W. Elko, "Adaptive Optimisation of Microphone Arrays under a Nonlinear Constraint", Proc. IEEE ICASSP 86, Tokyo

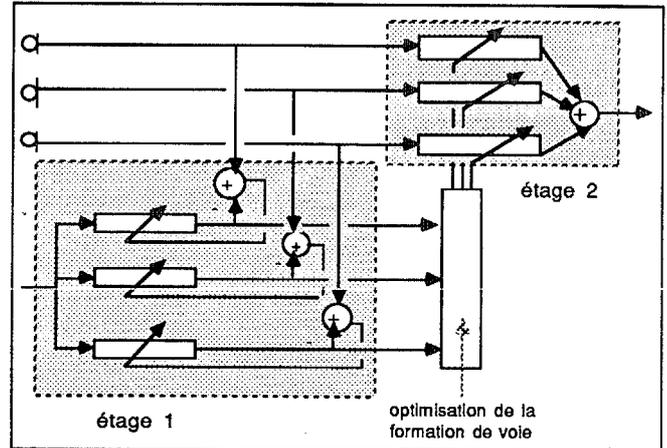


Figure 1: antenne à deux étages

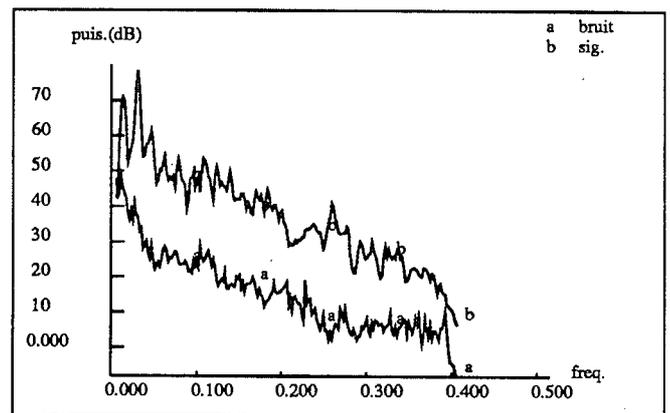


Figure 2: spectres moyens

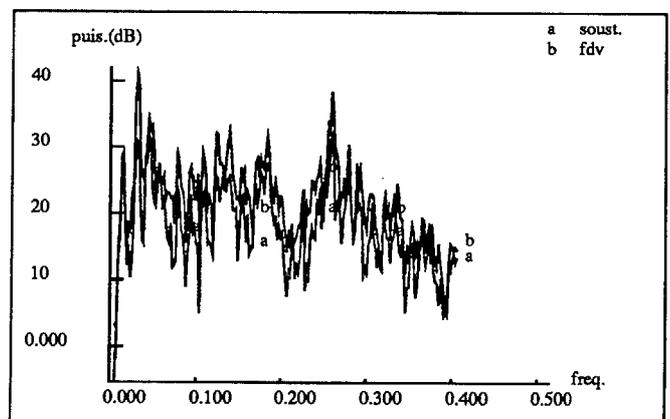


Figure 3: rapports signal à bruit obtenus par:

- a) soustraction (moy.=23dB)
- b) formation de voie (moy.=27dB)