

ANTENNE ACOUSTIQUE OPTIMISÉE GÉOMÉTRIQUEMENT POUR LA PRISE DE SON EN TÉLÉCONFÉRENCE

Fadi KHALIL, Jean-Pascal JULLIEN, André GILLOIRE

Centre National d'études des Télécommunications
CNET/LAA/TSS/CMC, BP40 22301 LANNION Cedex FRANCE

RÉSUMÉ

ABSTRACT

Résumé. Dans les nouveaux services des télécommunications, la qualité des terminaux audio est affectée par l'environnement acoustique dont les effets sur la qualité de la communication (écho acoustique, réverbération de la salle, bruit) doivent être contrôlés. Nous proposons dans ce papier une nouvelle technique de formation de voie en champ proche permettant de minimiser ces dégradations dans le contexte des salles de téléconférence. Le principe consiste à définir un gabarit dans l'espace qui focalise au maximum l'antenne sur la source utile tout en minimisant les lobes secondaires. Quelques résultats de simulations et de mesures en salles réelles du CNET-LANNION sont présentés. Les résultats sont encourageants et confirment l'intérêt de la technique proposée.

Abstract. In new telecommunication services, the audio terminals quality is affected by the acoustical environment (acoustic feedback, room reverberation and background noise). In this paper we propose a new beamforming technique in near field which allows to minimize these degradations in the context of conference rooms. This new method consists of designing a spatial filter that focuses the antenna on the desired signal and minimizing the sidelobes simultaneously. Some computer simulation results and experimental measurements in conference rooms at CNET-LANNION are presented. These initial results are encouraging and confirm the validity of the proposed new technique.

1• Introduction

Les services de téléconférence permettent de mettre en relation deux ou plusieurs groupes distants qui, installés dans des salles spécialement aménagées, peuvent converser et tenir une réunion à distance. A la prise de son, le message du locuteur est modifié par la salle avant d'être transmis. Cette modification croît avec le temps de réverbération de la salle et la distance entre la source et les microphones. Une multitude de réflexions viennent s'ajouter au message du locuteur dégradant l'intelligibilité et donnant une impression d'éloignement du locuteur. A cette réverbération du signal utile s'ajoute l'effet dû au signal provenant de l'autre salle, qui est réinjecté dans les microphones. Un tel couplage est à l'origine du phénomène d'écho et des instabilités de la boucle électro-acoustique formée par l'interconnexion des salles (effets Larsen). Autre particularité de notre application, un niveau de bruit faible peut être considérablement amplifié par rapport à une prise de son rapprochée. On est ici en large bande (le spectre de la parole occupe plusieurs octaves) et en champ proche. Le traitement d'antennes pour notre application se pose donc en des termes différents de ceux rencontrés classiquement en radar ou en sonar [1].

Le but du traitement à la prise de son est de pouvoir transmettre sans distorsions les messages des locuteurs en minimisant l'effet de la salle. Nous présentons dans ce papier une formation de voie en champ proche purement géométrique et optimisée a priori, pour une antenne de microphones dédiée à la prise de son dans les salles de téléconférence. Le principe consiste à définir un gabarit dans l'espace qui focalise au maximum l'antenne sur la source utile tout en minimisant les lobes secondaires. L'originalité de cette méthode est que l'antenne est focalisée sur une région de l'espace et non pas sur une source ponctuelle ce qui modélise mieux le cas réel. Les calculs à effectuer sont simples et ne nécessitent pas de lourdes procédures d'inversion de matrices. La conception de l'antenne est faite indépendamment des signaux ce qui rend ses performances indépendantes d'une éventuelle cohérence entre les sources utiles et les brouilleurs.

2• Formulation de la réponse spatiale d'une antenne linéaire

L'antenne réceptrice est constituée de $N+1$ capteurs numérotés de $-N/2$ à $N/2$ espacés uniformément par d (voir figure 1). La distance entre la source et un capteur d'indice i



sera notée d_i . On se place dans l'hypothèse d'une propagation libre en champ proche. Les signaux enregistrés par les $N+1$ capteurs, séparément accessibles, sont filtrés puis sommés pour fournir le signal résultant. Ce signal est censé être le plus proche possible du signal envoyé par la source utile notée $s(t)$. Le signal $s(t)$ émis par la source est transformé par le milieu de propagation avant d'être capté (voir figure 1). Cette transformation, selon les hypothèses faites précédemment, s'écrit suivant le vecteur :

$$G_s(f, d_0) = \left[\frac{d_0}{d_{-N/2}} e^{j\psi_{-N/2}}, \dots, \frac{d_0}{d_i} e^{j\psi_i}, \dots, 1, \dots, \frac{d_0}{d_{N/2}} e^{j\psi_{N/2}} \right]^T \quad (\text{EQ 1})$$

avec $\psi_i = (2\pi f(d_i - d_0)) / c$ la phase entre les capteurs d'indice i et 0, d_0 la distance entre la source et le micro 0, f la fréquence de l'onde incidente et c la célérité de l'onde dans le milieu. $G_s(f, d_0)$ est connu sous le nom de vecteur source dont la dimension est égale au nombre de capteurs.

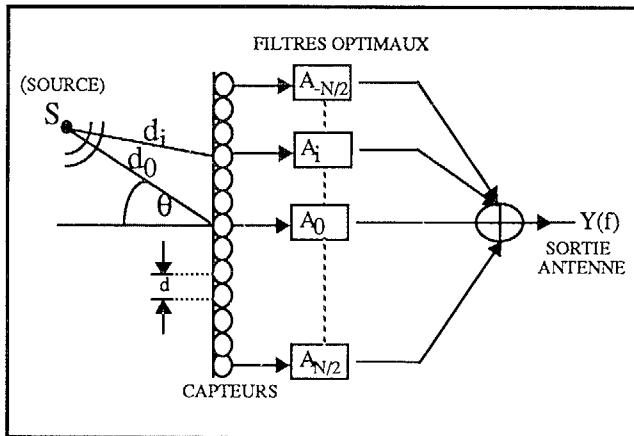


FIGURE 1 : Géométrie d'une antenne linéaire uniforme composée de $N+1$ capteurs. L'espacement entre capteurs est d . La source S est repérée par la distance d_i et l'angle polaire θ . Les filtres A_i permettent le pointage sur la source S . La sortie de l'antenne est la sommation des $N+1$ signaux derrière chaque capteur après filtrage.

La fonction de transfert du filtre derrière le capteur i sera notée $A_i(f)$. le spectre $X_i(f)$ du signal au capteur i normalisé par rapport au capteur 0 peut s'écrire d'après (EQ 1):

$$X_i(f) = \frac{d_0}{d_i} \cdot S(f) \cdot e^{j\psi_i} \quad (\text{EQ 2})$$

avec: $S(f) = \text{TF}[s(t)]$

La sortie totale du système notée $Y(f)$ peut s'exprimer en une sommation de signaux filtrés :

$$Y(f) = \sum_{i=-N/2}^{N/2} \frac{d_0}{d_i} \cdot S(f) \cdot e^{j\psi_i} \cdot A_i(f) \quad (\text{EQ 3})$$

$$Y(f) = S(f) \cdot H(f, s) \quad (\text{EQ 4})$$

avec :

$$H(f, s) = A^T(f) \cdot G_s(f, d_0) \quad (\text{EQ 5})$$

$H(f, s)$ ainsi défini est appelé la réponse de l'antenne à une fréquence f , qui dépend d'une part du vecteur source $G_s(f, d_0)$ déterminé à partir de la position géométrique de la source et des conditions de propagation, et d'autre part du vecteur filtrage $A(f)$ censé privilégier l'onde directe provenant de la source utile sous la contrainte d'absence de distorsion et atténuer les sources non désirables provenant des autres directions. Idéalement, en supposant que la source utile se trouve en s_0 , la réponse de l'antenne doit vérifier pour chaque fréquence f :

$$H(f, s) = \begin{cases} 1 & s = s_0 \\ 0 & \text{ailleurs} \end{cases}$$

Dans les cas réels, une certaine marge d'incertitude Δs sur les positions est souhaitée. D'où une nouvelle écriture pour $H(f, s)$:

$$H_{id}(f, s) = \begin{cases} 1 & s = s_0 \pm \Delta s \\ 0 & \text{ailleurs} \end{cases}$$

Notre but est d'optimiser les filtres $A_i(f)$ de façon à s'approcher le plus possible de cette réponse de l'antenne idéale définie ci-dessus.

3• Synthèse des filtres optimaux

L'expression de $H(f, s)$ donnée par (EQ 3) peut être simplifiée en se plaçant dans l'hypothèse du champ lointain. En changeant la numérotation des capteurs on obtient :

$$H(f, s) = \sum_{i=0}^N A_i(f) e^{-j2\pi f i \tau} \quad (\text{EQ 6})$$

avec : $\tau = d \cdot \sin(\theta) / c$

En considérant la fonction fenêtre:

$$\Lambda(i) = \begin{cases} 1 & 0 \leq i \leq N \\ 0 & \text{ailleurs} \end{cases}$$

et en rappelant que l'équation (EQ 6) est relative, pour une source s , à une fréquence f , on peut réécrire celle-ci comme:

$$H(f, s) = \sum_{i=-\infty}^{\infty} \Lambda(i) A_i(f) e^{-j2\pi f i \tau} \quad (\text{EQ 7})$$

Suivant la définition de la transformée de Fourier d'un signal aperiodique discret $x(k)$:

$$\text{TF}[x(k)] = \sum_{k=-\infty}^{\infty} x(k) e^{-j2\pi f k}$$

l'équation (EQ 7) peut être interprétée à une fréquence f comme une transformée de Fourier du signal discret $\Lambda(i)A_i(f)$ qu'on note $G(f\tau)$.

D'où:

$$H(f, s) = \text{TF}[\Lambda(i)A_i(f)] = G(f\tau) \quad (\text{EQ 8})$$

Avec cette double égalité on déduit d'une part que les coefficients des filtres $A_i(f)$ derrière chaque micro à une fréquence donnée peuvent être calculés par une Transformée de Fourier Inverse d'une fonction continue périodique de période unité et d'autre part la fonction $G(f\tau)$ n'est autre que la réponse de l'antenne idéale $H_{id}(f, s)$. Donc, suite à ce raisonnement, les fonctions $G(f\tau)$ seront définies avec la même allure que celle de la réponse idéale. Le signal discret $\Lambda(i)A_i(f)$ peut être calculé par Transformée de Fourier Inverse de $G(f\tau)$:

$$\Lambda(i)A_i(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} G(f\tau) e^{j2\pi f\tau} d(f\tau)$$

soit :

$$G(f\tau) = \begin{cases} 1 & -\frac{B}{2} \leq f\tau \leq \frac{B}{2} \\ 0 & \text{ailleurs} \end{cases}$$

on obtient :

$$\Lambda(i)A_i(f) = B \operatorname{sinc}(\pi i B)$$

Donc le vecteur filtre à une fréquence f n'est autre qu'un sinus cardinal échantillonné spatialement pour un pas égal à la distance entre deux capteurs et pondéré par une fenêtre de la taille de l'antenne.

Puisqu'on est en champ proche, l'antenne est corrigée en phase et en amplitude conformément à (EQ 3). Les simulations ont montré que l'approche champ lointain est approximativement valable lorsque les sources sont à une distance supérieure à deux fois la taille de l'antenne. Donc suivant la taille de l'antenne et la position de la source cette correction est nécessaire ou non.

4• Antenne logarithmique

A une fréquence donnée f le filtrage derrière chaque micro est donné par :

$$V_f(i) = \Lambda(i)A_i(f) = B \operatorname{sinc}(\pi i B) \quad (\text{EQ 9})$$

avec i l'indice du micro et B la bande en fréquence spatiale du lobe passant donnée par:

$$B = \frac{\Delta k \cdot d}{\lambda} \quad (\text{EQ 10})$$

Δk étant la largeur de la zone utile.

$V_f(i)$ peut être considéré comme un signal échantillonné spatialement par l'antenne. Pour un réseau uniforme cette période d'échantillonnage est égale à la distance entre deux capteurs. D'après le théorème de Shannon d/λ doit être inférieur à $1/2$ pour éviter le repliement spectral.

D'après (EQ 10) la distance d doit être choisie de façon à respecter Shannon pour la plus grande fréquence traitée, ce qui défavorise les basses fréquences. Par contre si on augmente la distance entre les capteurs, ce qui implique d'augmenter la taille de l'antenne, on améliore la résolution de l'antenne dans les basses fréquences, mais ceci introduit un repliement spectral

dans les hautes fréquences. D'où un compromis à faire.

Dans le but d'équilibrer l'antenne pour les différentes fréquences, on a donc étudié des antennes à espacement non uniforme des capteurs.

5• Outils d'évaluation

5.1 Gain moyen

Ce critère permet de calculer pour une source donnée le gain moyen de l'antenne sur la voix humaine en introduisant la loi de pondération spectrale A qui modélise la perception de l'oreille humaine. Pour les valeurs du spectre de la voix humaine, on a repris les résultats d'expériences faites au CNET-LANNION [5]. Ces valeurs sont groupées dans le tableau 1 avec celle de la loi A . Ce critère nous indique la discrimination spatiale de l'antenne.

5.2 Distorsion

La contrainte sans distorsion sur une zone de l'espace autour de la source utile est évaluée par un critère calculé à partir de la déviation spectrale du gain de l'antenne dans la direction de la source utile par bande d'octaves pondéré par l'énergie de la voix humaine et la loi A .

bande (Hz)	125	250	500	1000	2000	4000	8000
voix humaine	44.21	46	45.3	41	41	35	21.1
loi A	-10.7	-3.5	1.6	4.7	5.9	5.7	3.7

Tableau 1: Energie de la voix humaine et loi de pondération A (en dB par bandes d'octave).

6• Résultats de simulation

On a effectué des simulations pour une antenne linéaire à espacements logarithmiques composée de 11 capteurs, ayant une taille de 1.6 m. L'antenne globale est la réunion de 4 antennes uniformes à 5 capteurs optimisées pour chaque bande d'octave. Le filtre placé derrière chaque capteur est à réponse impulsionnelle finie avec 64 coefficients par filtre. La source utile est supposée en face de l'antenne ($\theta = 0$). La largeur du gabarit est de 10 degrés dans la bande 8 KHz et 20 degrés dans les autres bandes.

La discrimination spatiale et la largeur de la zone utile (lobe centrale) pour l'antenne considérée sont présentés sur la figure 2. La discrimination est satisfaisante, par contre le lobe passant aurait été souhaité plus étroit, mais vu la contrainte physique imposée par la taille de l'antenne, la sélectivité dans les basses fréquences (bande 125 Hz) est médiocre. La figure 3 correspond à la distorsion spectrale sur le signal de parole. Dans la zone utile autour de la source, la distorsion est négligeable pour l'antenne logarithmique. En résumé, les deux figures montrent la meilleure performance de l'antenne logarithmique vis-à-vis de l'antenne uniforme et de l'antenne retard-somme.

7• Résultats de mesures

Des mesures de réponses impulsionnelles ont été faites dans la salle d'audioconférence du CNET-LANNION. Cette salle a une géométrie parallélépipédique et un volume de $4.45 \times 4.8 \times 3.28 = 70 \text{ m}^3$. On a ainsi deux positions de locuteurs et une position du haut-parleur. Le locuteur L1 est considéré



comme orateur (source utile), L2 représente une autre position de L1 dans la zone utile pour évaluer la qualité de la prise du son lorsque le locuteur L1 bouge. Le haut parleur représente une source non désirable provenant d'une autre direction que celle de la direction utile. L'antenne est identique à celle simulée au paragraphe précédent. Elle a été placée verticalement de manière à se focaliser sur une zone circulaire d'axe le support de l'antenne. Le but des mesures est de tester la performance de l'antenne dans un contexte réel de prise du son. La mesure des réponses impulsionnelles entre les sources et les capteurs de l'antenne permet de calculer, d'après l'équation (EQ 5) le vecteur $G_s(f, d_0)$. Le vecteur $A(f)$ est calculé suivant la procédure décrite au paragraphe 3. Les calculs sont faits dans le domaine fréquentiel. On emploie un traitement sous forme de transformée de Fourier par blocs qui permet d'optimiser séparément pour chaque fréquence le diagramme de directivité de l'antenne. La référence est une prise de son à un seul capteur omnidirectionnel placé au milieu de l'antenne. Les résultats sont groupés dans le tableau 2. On voit bien l'amélioration apportée sur le rapport énergétique entre le son direct et le son réverbéré. L'atténuation sur l'onde directe du signal provenant du haut-parleur est de l'ordre de 16 dB ce qui confirme les résultats de simulation (voir figure 2). L'atténuation sur le signal global provenant du haut-parleur est de l'ordre de 11 dB.

locuteur 1	locuteur 2	brouilleur
K	K	A

Référence	+5.72	+3.23	-----
Antenne	+11.3	+7.0	+16.48

K = Rapport onde directe et champ réverbéré

A = Atténuation de l'onde directe du signal provenant du haut-parleur

Tableau 2 : Résultats de la mesure dans la salle d'audioconférence du CNET-LANNION (valeurs en dB).

8• Conclusion

Nous avons présenté une méthode d'optimisation géométrique d'une antenne focalisée sur une région de l'espace et non pas sur une source ponctuelle, ce qui modélise mieux le cas réel en prenant en compte les déplacements des locuteurs.

L'association des antennes uniformes optimisées pour chaque bande d'octave conduit à de meilleures performances que celles obtenus pour une antenne uniforme traitant toute la bande de fréquence. D'où l'intérêt des traitements par bandes d'octaves et l'utilisation d'antennes logarithmiques. La méthode présentée fournit des résultats satisfaisants pour l'application à l'audioconférence avec une complexité de calcul réduite. Cependant, dans un milieu très réverbérant, une amélioration de la directivité dans les basses fréquences doit être recherchée. Notre méthode est plus performante que la formation de voie retard-somme et plus robuste que les méthodes de hautes résolutions de type Capon dans le sens où elle est indépendante vis-à-vis des signaux.

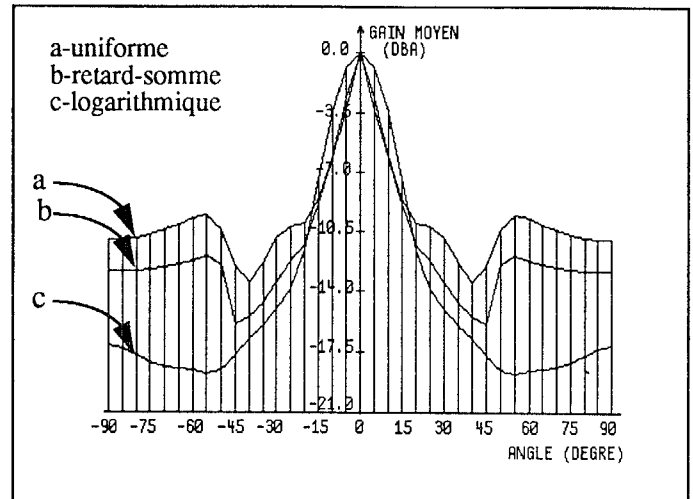


FIGURE 2 : Gain moyen de l'antenne sur la voix humaine en loi A.

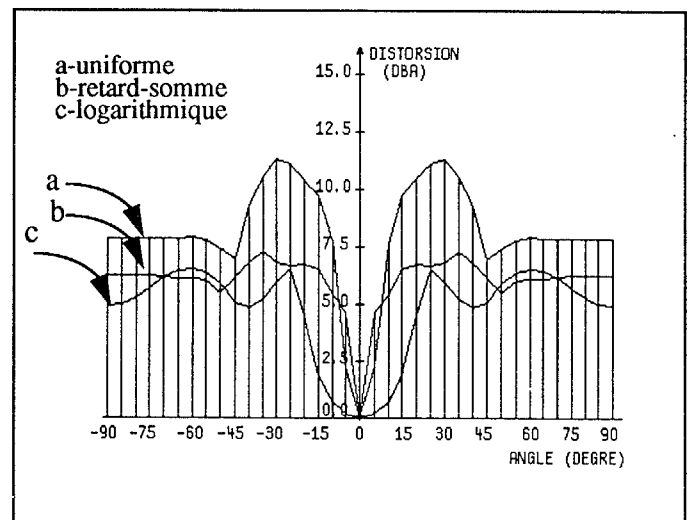


FIGURE3 : Distorsion sur la voix humaine en loi A.

RÉFÉRENCES :

- [1] A.Gilloire, J.P.Jullien, "L'Acoustique des Salles dans les Télécommunications", L'écho des Recherches, CNETFRANCE, N°127,1987.
- [2] A.J.Fenn, "Evaluation of Adaptive Phased Array Antenna Far-Field Nulling Performance in the Near-Field Region". IEEE Trans. on Antennas and Propagation, VOL38, N°2, Feb1990.
- [3] J.L.Flanagan, "Bandwidth Design for Speech-Seeking Microphone Arrays". Proc. of the IEEE, 1985.
- [4] O.L.Frost, "An Algorithm For Linearly Constrained Adaptive Array Processing". Proc. of the IEEE, Aug. 1972.
- [5] P.Lorand, G.Minier, "Spectre à Long Terme de la Parole et Téléphonométrie Objective", Colloque FASE 1975, PARIS-actes p134-143.
- [6] R.A.Monzingo and T.W.Miller, Introduction to Adaptive Arrays. NEW YORK:Wiley 1980.
- [7] Y.Grenier, M.Xu, J.Prado, D.Liebenguth, "une Antenne Acoustique Adaptative, Aspects Matériels et Logiciels",GRETSI 89, NICE..