# GAIN OPTIMAL POUR LA DEREVERBERATION DE LA PAROLE

Banbang HIDAYAT, Jean-Pascal JULLIEN, André GILLOIRE

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

## Résumé

En utilisant un système d'analyse-synthèse par transformée de Fourier à court terme, un signal est synthétisé à partir de deux signaux captés par des microphones proches. Le but est d'éliminer au maximum la réverbération produite par la salle et captée par les microphones. La fonction de cohérence est utilisée pour segmenter les signaux et son intérêt pour corriger l'amplitude spectrale des signaux captés est discuté.

## Abstract

Using a short-time Fourier transform analysis-synthesis system, an output signal is derived from two input ones picked up by closely spaced microphones. The aim is to eliminate as much as possible the room reverberation picked-up by the microphones. The coherence function is used to segment signals and its usefullness to correct the picked-up signals magnitudes is discussed.

## Problème de la réverbération

Les nouveaux services des Télécommunications, comme la téléconférence ou les postes mains-libres, utilisent des terminaux audio qui sont sensibles à l'environnement acoustique. Le bruit ambiant et l'effet de salle détériorent la qualité du signal transmis, idéalement le seul son direct du locuteur. Le tableau 1 montre les résultats de tests subjectifs dont le but était d'évaluer la gêne apportée par l'effet de salle [CCITT]. Cet effet de salle est caractérisé par le critère K qui représente le rapport énergétique de la réverbération de la salle captée par le(s) microphone(s) du terminal sur le son direct capté par le même terminal. K croît avec le temps de réverbération de la salle TR et la distance entre la source et les microphones. Plusieurs configurations étaient testées avec différentes valeurs de K; on observe une augmentation importante de la gêne lorsque l'effet de salle augmente.

| K (dB) | 0 | -3 | -6 | -9 | -12 | -15 |
|---|---|---|---|---|---|---|
| pas gênant % | 7,1 | 14 | 23 | 41 | 71 | 89 |
| peu gênant % | 14,2 | 25 | 46,4 | 46,4 | 23,2 | 9 |
| gênant % | 42,8 | 48,2 | 30,6 | 12,6 | 5,8 | 2 |
| très gênant % | 35,8 | 12,8 | 0 | 0 | 0 | 0 |

**Tableau 1**: Répartition des jugements de «pas gênant» à «très gênant» en fonction de K, le rapport énergétique du son réverbéré sur le son direct.

## Algorithmes de déréverbération

Parmi les solutions proposées dans la littérature, plusieurs algorithmes [Allen, 1], [Bloom] utilisent un système d'analyse-synthèse par Transformée de Fourier à Court Terme (TFCT) [Allen, 2], calculée sur les signaux captés par deux microphones proches (figure 1).

h(m) : fenêtre glissante indicée par m = sR
X(m,k), Y(m,k) : $k^{ième}$ raies des TFCTs de x(t) et y(t) fenêtrés par h(m)
G = f ( X, Y ) : gain calculé à partir de X et Y
pour tout (m,k) : Z = G( aX + bY )
z(t) : signal traité, synthétisé à partir des transformées de Fourier inverses des Z(m,k)

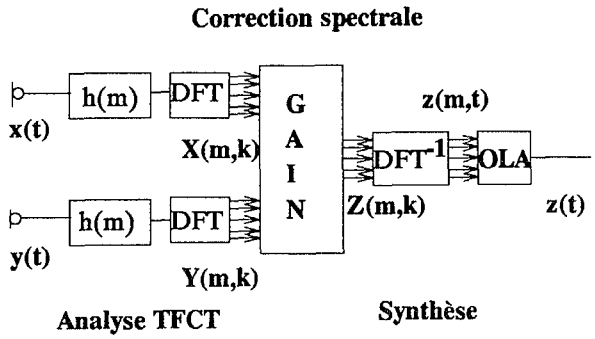


**Figure 1** : Schéma général d'un algorithme de déréverbération.

Le signal traité est synthétisé par la méthode OLA [Allen, 2] à partir des TFCTs, X et Y, des signaux microphoniques x(t) et y(t). Si G est égal à 1, ces procédés d'analyse-synthèse assurent une reconstruction parfaite du signal (z = x ou z = y). Idéalement, G doit être égal à 1 lorsque le signal ne contient que du son direct, il doit être nul lorsqu'il n'y a que du son réverbéré. Une valeur intermédiaire entre 0 et 1 doit permettre de diminuer l'effet du son réverbéré lorsqu'il est superposé au son direct. Ainsi, ces traitements agissent essentiellement sur l'amplitude des signaux captés en essayant de synthétiser un signal dont les TFCTs aient les mêmes amplitudes que les TFCTs du signal de son direct. On appellera Gain»idéal» le gain qui permet d'obtenir l'amplitude des TFCTS du son direct. Suivant cette définition, on trouve :

$G^2_{idéal} = 1 / (1 + K)$ (1)

Le gain calculé est variable dans le temps et indépendant pour chaque raie de la TFCT, pour tenir compte des variations fortes de l'énergie du signal de parole en fonction de la fréquence et du temps.

## Fonction de cohérence à court terme

L'intérêt de cette fonction est de donner une estimation de K dont dépend le gain idéal. En notant $C_{ab}(f)$ la densité de l'interspectre entre les signaux a et b, l'amplitude de la Fonction de Cohérence est définie de la façon suivante :

$C(f) = |C_{xy}(f)| / (C_{xx}(f)C_{yy}(f))^{1/2}$ (2)

MSC (magnitude squared coherence) = $C^2$

On montre aisément que

si y = x + b (b: bruit additif décorrélé),

alors : $MSC = 1 / (1 + C_{bb}/C_{xx})$ (3)

Dans le cas où la réverbération est considérée comme un bruit décorrélé et x comme le son direct de référence, on obtient directement $G_{idéal} = C$. Les fonctions de gain proposées dans la littérature sont toujours proches de C [Allen, 1], [Bloom]. En pratique, les estimations de C(f) sont obtenues par moyennage sur les TFCTs avec des fenêtres exponentielles de l'ordre de la durée de stationnarité de la parole. Le tableau 2 présente la corrélation entre le gain idéal et l'estimation de C pour différentes valeurs de lissage [Hidayat] (seuls les passages où le son direct est présent ont été sélectionnés pour effectuer la corrélation); 64 ms est une valeur satisfaisante pour toutes les fréquences.

| Lissage (ms) | 8 | 16 | 32 | 64 | 128 |
|---|---|---|---|---|---|
| 250 Hz | 0,602 | 0,611 | 0,615 | 0,725 | 0,565 |
| 500 Hz | 0,647 | 0,655 | 0,655 | 0,756 | 0,614 |
| 1 kHz | 0,647 | 0,655 | 0,655 | 0,651 | 0,663 |
| 2 kHz | 0,635 | 0,621 | 0,614 | 0,669 | 0,637 |
| 3 kHz | 0,728 | 0,717 | 0,709 | 0,697 | 0,752 |
| 4 kHz | 0,595 | 0,609 | 0,629 | 0,643 | 0,692 |

**Tableau 2** : Corrélation entre le gain idéal et l'estimation de la fonction de cohérence à court terme pour différentes valeurs de lissage; phrase de 2s; TR = 1.7s; fenêtres de 32 ms; recouvrement de 75%.

On observe aussi que C n'est pas une estimation parfaite du gain optimal; plusieurs raisons peuvent expliquer cela. D'une part, l'estimation de C comporte un biais et une variance non négligeable [Carter] qui obligent à prendre un taux minimal de recouvrement des fenêtres d'analyse: ce taux doit dépasser 50%. D'autre part et surtout, la valeur théorique de C(f), lorsque le son direct est absent et que le champ réverbéré est parfaitement diffus, n'est pas nulle mais limitée inférieurement par la fonction sinc(kd) où k est le nombre d'onde et d la distance entre les microphones [Bendat]. La figure 2 montre les valeurs moyennes de C obtenues avec des micros placés dans une salle très réverbérante. Cette raison explique pourquoi les traitements proposés ne fonctionnent pas de façon satisfaisante dans les basses fréquences lorsque le champ réverbéré est important [Hidayat].

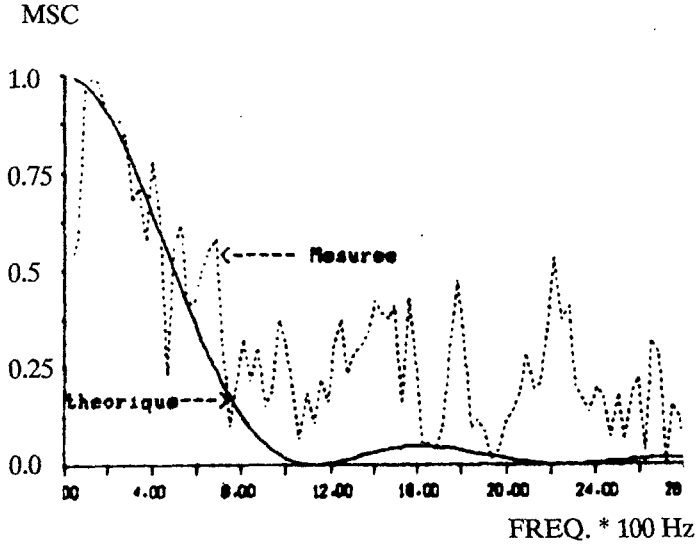


**Figure 2** : Valeur théorique (champ diffus) de la MSC, comparée à sa valeur moyenne observée en fonction de la fréquence; phrase de 2s; TR = 1.7s.

Il reste une dernière raison pour expliquer que C n'est pas une estimation parfaite de $G_{idéal}$ : le processus de réverbération est un filtrage linéaire et l'hypothèse de décorrélation entre le son direct et le champ réverbéré n'est pas correcte; seules sont

décorrélées les réflexions qui arrivent suffisament tardivement pour être en dehors de la durée de stationnarité du signal de parole. Le tableau 3 compare les valeurs de K mesurées dans la salle et celles estimées à partir de MSC par l'équation (3) où $K = C_{bb} / C_{xx}$. Ainsi, les valeurs estimées de K s'approchent du critère C80 qui représente le rapport des premières réflexions (onde directe + premières réflexions avant 80 ms) sur les réflexions tardives (réflexions après 80 ms).

| dB | Dir/Rev | C80 | K estimé |
|---|---|---|---|
| 1-250 Hz | -6,8 | 1,4 | 1,8 |
| 1-1 kHz | -2 | 2,2 | 0,9 |
| 2-250 Hz | -0,8 | 6,5 | 4,1 |
| 2-1 kHz | -2,5 | 3,6 | -1,4 |
| 3-250 Hz | -1,6 | 5 | 2,9 |
| 3-1 kHz | 1,1 | 7,1 | 3,3 |
| 4-250 Hz | 10 | 14 | 12 |
| 4-1 kHz | 1,9 | 21 | 2,5 |

**Tableau 3** : Comparaison des valeurs estimées de K avec les critères Dir/Rev (=1/K) et C80 (voir texte); 4 configurations de salle; 2 fréquences de la TFCT.

## Segmentation fréquentielle

Les observations précédentes ont amené à scinder le problème de la déréverbération en deux parties : d'une part, une segmentation temporelle pour chaque raie du spectre afin de sélectionner les parties où le son direct est présent; d'autre part, l'application d'un gain qui diminue l'influence du champ réverbéré lorsqu'il se mêle au son direct. Ainsi, une règle simple d'atténuation maximale du signal est-elle appliquée lorsque le son direct n'est pas détecté (un lissage du gain est toutefois nécessaire pour éviter les effets négatifs des variations brutales).

Dans les algorithmes précédents [Allen, 1], [Bloom], la valeur de C est l'indicateur principal pour calculer le gain G; Bloom propose une fonction de gain fortement non linéaire par rapport au MSC, à la limite cela revient à chercher un seuil absolu sur C pour détecter la présence du son direct. Cette démarche n'a pas abouti et aucun seuil satisfaisant n'a été trouvé. Par contre, le taux de croissance du MSC s'est avéré un excellent détecteur de l'apparition du son direct; pour décider de la disparition de celui-ci, un critère de décroissance énergétique du signal a été testé [Hidayat] . L'algorithme complet de segmentation a été évalué à l'aide des critères utilisés pour tester les détecteurs de parole. Les résultats sont présentés dans le tableau 4.

| | 250 Hz | 500 Hz | 1 kHz | 2 kHz |
|---|---|---|---|---|
| PNDP % | | 4 | 4 | 13 |
| PNDS % | 20 | 12 | 9 | 14 |

**Tableau 4** : pourcentage du son direct non détecté, PNDP, et pourcentage de champ réverbéré seul non détecté, PNDS; phrase de 2s; TR = 1.7s.

On peut considérer que le PNDP, correspondant au son direct non détecté, reste à un niveau acceptable; la valeur plus élevée du PNDP à la fréquence 2kHz traduit la difficulté de préserver les consonnes brèves lorsqu'elles sont noyées dans une réverbération importante. Par contre, il reste des plages importantes où l'algorithme ne sait pas détecter l'absence du son direct, cela réduit d'autant sa capacité à éliminer le son réverbéré.

## Gain optimal

En supposant que l'algorithme effectue une segmentation idéale (dérivée du signal anéchoïque de référence), plusieurs fonctions de gain ont été testées pour corriger l'amplitude du signal lorsque le son direct est détecté. Le tableau 5 compare les performances de ces différentes fonctions en donnant les corrélations obtenues avec le gain idéal.

| | G = 1 | G = C | G = MSC |
|---|---|---|---|
| 250 Hz | 0,653 | 0,725 | 0,723 |
| 500 Hz | 0,603 | 0,756 | 0,714 |
| 1 kHz | 0,68 | 0,65 | 0,609 |
| 2 kHz | 0,73 | 0,669 | 0,508 |
| 3 kHz | 0,799 | 0,697 | 0,573 |
| 4 kHz | 0,685 | 0,643 | 0,552 |

**Tableau 5** : Corrélations de 3 fonctions de gain avec le gain idéal; gain constant (1), fonction de cohérence (C), carré de la fonction de cohérence (MSC); phrase de 2s, TR = 1.7s.

Le résultat principal est que le gain constant est le meilleur candidat pour les fréquences au dessus de 1 kHz; au dessous de cette fréquence, la fonction de gain G = C est légèrement meilleure. En effet, la segmentation fréquentielle élimine déjà les portions du signal où l'on a que du son réverbéré, il ne reste donc que les portions où le son direct se mélange avec les premières réflexions et, comme il a été vu plus haut, la fonction de cohérence ne peut pas séparer le son direct des premières réflexions (non décorrélation entre ces deux signaux à l'horizon des 64 ms).

## Reconstruction de la phase

On a vu dans les paragraphes précédents que le travail effectué a consisté à chercher une fonction de gain qui agit uniquement sur l'amplitude du signal; aucune correction de la phase n'a été appliquée. On sait pourtant que la réverbération, qui peut être simulée par une combinaison de retards et de déphaseurs, agit fortement sur la phase du signal. A l'écoute, de forts effets de phasing sont perçus sur le signal traité, même lorsque l'on applique le gain idéal. L'algorithme LSEE-MTFCTM [Griffin] permet de reconstruire la phase du signal traité à partir des amplitudes de ses TFCTs. La figure 3 montre le rapport signal à bruit obtenu lorsque la différence entre le signal traité et le signal anéchoïque est considérée comme du bruit. Ce rapport atteint des valeurs de l'ordre de 20 dB à partir d'une cinquantaine d'itérations : le résultat est donc excellent mais au prix d'une charge de calcul importante.

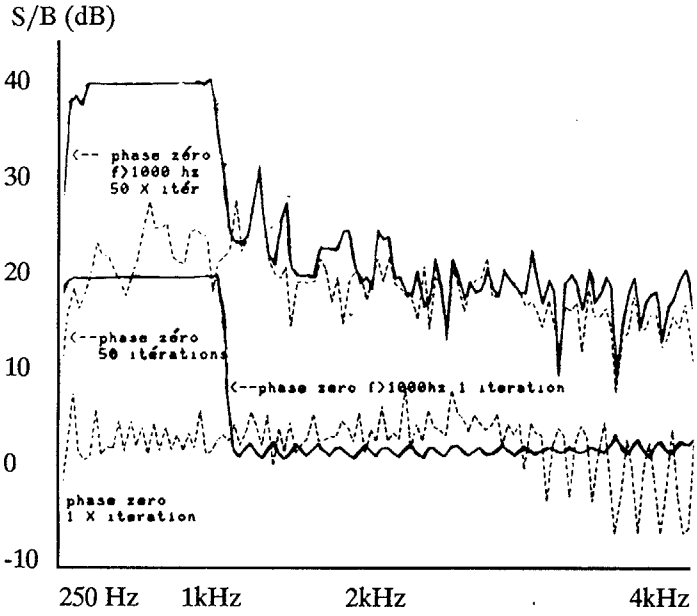


**Figure 3** : Efficacité de la reconstruction de la phase en terme de rapport signal à bruit, en fonction de la fréquence; le signal traité est comparé au signal anéchoïque de référence; phase initiale nulle ou partiellement nulle; 1 ou 50 itérations.

## Conclusion

La fonction de cohérence ne permet pas de corriger effectivement l'amplitude du signal réverbéré, probablement à cause de la nature du champ réverbéré. Au contraire, elle permet d'estimer, sur les signaux captés dans la salle, l'importance des premières réflexions par rapport aux réflexions tardives; à ce titre elle pourrait être un estimateur correct de l'intelligibilité dans une salle mais cela reste à vérifier. Elle permet par contre de détecter précisément l'apparition du son direct même lorsqu'il est noyé dans la réverbération. La qualité du déréverbérateur est essentiellement liée à la qualité de la segmentation fréquentielle; le critère énergétique de détection d'arrêt du son direct doit être amélioré pour réaliser une bonne segmentation. La reconstruction de la phase reste une tâche coûteuse, non réalisable en temps réel, mais très efficace.

## Références bibliographiques

**[Allen, 1]** J.B. Allen, D.A. Berkley, J. Blauert, «Multimicrophone signal processing technique to remove reverberation from speech signals», JASA vol. 62, n°4, Oct. 1977, pp. 912-915.

**[Allen, 2]** J.B.Allen, L.R.Rabiner, «A unified approach to short time Fourier analysis and synthesis», Proc. IEEE vol. 65 n°11, Nov. 1977, pp. 1558-1564.

**[Bendat]** J.S.Bendat, A.G.Piersol, «Engineering applications of correlation and spectral analysis», John Wiley and Sons 1980.

**[Bloom]** J.P.Bloom, G.D.Cain, «Evaluation of two input speech dereverberation techniques», ICASSP 1982, Paris, pp. 164-167.

**[Carter]** C.C.Carter, «Coherence and time delay estimation», Proc. IEEE vol. 75, n°2, Feb. 1987, pp. 236-255.

**[CCITT]** CCITT, «Delayed contribution n°FR5 (WP XII/2), Geneva 28-30 April 1987.

**[Griffin]** D.W.Griffin, J.S.Lim, «Signal estimation from modified short-time Fourier transform», IEEE trans. ASSP, vol. 32, n°2, April 1984, pp. 236-243.

**[Hidayat]** B.Hidayat, «traitements fréquentiels à la prise de son multi-microphones en vue de la déréverbération du signal de parole», Thèse de l'Université de Rennes I, Janvier 1989