

Étude théorique et expérimentale de réducteurs de bruit Application au débruitage de la parole

Theoretical Study of Noise Reduction Filters Application to Speech Enhancement

par Régine LE BOUQUIN-JEANNÈS, Gérard FAUCON

Laboratoire de Traitement du Signal et de l'Image
Bât. 22, Campus de Beaulieu, Université de Rennes I
F-35042 Rennes cedex

résumé et mots clés

Le problème traité est celui de l'estimation d'un signal perturbé par un bruit additif lorsque l'on dispose de deux observations chacune composée d'un signal et d'un bruit additif. On se place dans le cas où les signaux utiles sur chaque voie sont déduits d'un même signal par filtrage linéaire et les bruits complètement décorrélés. Nous cherchons à évaluer l'apport de la deuxième observation pour une meilleure estimation du signal par rapport au cas où une seule observation est disponible. Le filtrage de Wiener vectoriel ainsi que deux structures sous-optimales dites PIS (Prétraitement + Identification entre Signaux) proposées pour l'amélioration de la parole bruitée sont présentés. Une étude théorique (en supposant optimaux les filtres utilisés) est d'abord menée : ces différents systèmes ainsi que le filtrage de Wiener mono-voie sont comparés entre eux en termes de distorsion, bruit résiduel et erreur globale sur le signal à estimer. Un classement de ces méthodes est donné en fonction des valeurs relatives des rapports signal à bruit sur chaque voie. Nous montrons l'intérêt des deux structures PIS par rapport au filtrage de Wiener mono-voie lorsque le rapport signal à bruit de l'observation sur laquelle est présent le signal à estimer est faible. Après une présentation des différents algorithmes effectivement utilisés dans le cadre du débruitage de la parole, nous les appliquons à des signaux réels enregistrés dans une voiture. Une comparaison des algorithmes est faite suivant des critères objectifs et des tests d'écoute. Les résultats indiquent une supériorité des deux méthodes PIS par rapport aux filtrages de Wiener mono-voie et bi-voie.

Estimation, Débruitage de la parole.

abstract and key words

This paper deals with the estimation of a speech signal disturbed by acoustical additive noises when two noisy observations are available. The useful signals received on two microphones are issued from the same speech signal and noises are assumed to be decorrelated. Our concern is the improvement due to the second channel in comparison with the estimation obtained using only one channel. The vectorial Wiener filtering and two structures, called PIS (Preprocessing + Signal Identification), are presented. A theoretical study based on optimal filters is first presented. The different systems as well as the mono-channel Wiener filtering are compared in terms of distortion, residual noise and global error. The interest of the proposed structures is proved for low input signal to noise ratios on the first channel. The four algorithms — mono-channel and two-channel Wiener filterings and the two structures PIS — are tested on real noisy signals recorded in a moving car. They are compared using objective and subjective measures. Results confirm the improvement brought by our methods on real signals.

Estimation, Speech enhancement.

1. introduction

Avec l'essor des radiocommunications mobiles et le développement des terminaux mains-libres, les techniques de réduction de bruit trouvent aujourd'hui de plus en plus leur place. Dans ce papier, nous présentons de manière détaillée une comparaison théorique entre les filtrages de Wiener (mono-voie et bi-voie) et une technique de réduction de bruit, appelée PIS (Prétraitement puis Identification entre Signaux), que nous avons proposée dans [Faucou], et dont nous donnons ici deux versions, PISA et PISB. L'objectif est de savoir si cette nouvelle structure permet, au prix d'un coût de calcul faible, de réduire le bruit aussi bien ou sinon mieux qu'en utilisant un filtre de Wiener. Après la présentation des différentes structures dans le paragraphe 2, nous évaluons leurs performances théoriques respectives en 3. Pour terminer, nous indiquons la mise en œuvre de nos méthodes et donnons des résultats à la fois objectifs et subjectifs obtenus à partir d'enregistrements réels.

2. méthodes présentées

Nous présentons ici les différentes méthodes étudiées pour l'estimation d'un signal s_1 . Celui-ci est perturbé par un bruit additif n_1 tel que l'observation disponible s'écrit : $x_1 = s_1 + n_1$. Dans le cas où l'on dispose uniquement de cette voie, la meilleure estimée de s_1 au sens du minimum de l'erreur quadratique moyenne est donnée par le filtre de Wiener dont le gain en fréquence est :

$$H_1 = \frac{\gamma_{s_1}}{\gamma_{s_1} + \gamma_{n_1}} \quad (1)$$

où γ_u représente la densité spectrale de puissance (dsp) d'un signal u . La motivation de cette étude est de savoir si la disponibilité d'une seconde voie sur laquelle est présente une observation bruitée x_2 ($x_2 = s_2 + n_2$) permet d'améliorer l'estimation du signal s_1 . On considère que les signaux s_1 et s_2 sont déduits d'un même signal source par filtrage linéaire, les spectres de ces signaux sont donc liés par une fonction de transfert et l'on écrit : $S_2 = C \cdot S_1$. Par contre, les deux bruits sont supposés décorrelés et donc la densité interspectrale $\gamma_{n_1 n_2}$ est nulle. De telles situations peuvent se rencontrer dans les applications telles que le débruitage de la parole avec deux microphones. L'application qui nous concerne est celle de la téléphonie mobile où le combiné téléphonique est remplacé par un poste mains-libres. Un signal source s issu de la bouche du locuteur situé à l'intérieur d'un véhicule donne naissance à deux signaux s_1 et s_2 sur chaque microphone, les fonctions de transfert entre s et s_1 ou entre s et s_2 étant liées à l'habitacle de la voiture. Par contre, le bruit est diffus et dans le cas où les sources de bruit sont réparties sur une sphère, la cohérence

entre les bruits n_1 et n_2 reçus par les microphones est de la forme $\sin x/x$ où x est proportionnel à la fréquence et à la distance entre microphones. Si cette distance est suffisante, cette cohérence sera faible excepté en basses fréquences. Les composantes de bruit corrélées peuvent être fortement réduites par un simple filtrage passe-haut sans altérer le signal de parole. Avec ces hypothèses, on obtient :

$$\gamma_{x_1 x_2} = \gamma_{s_1 s_2} \quad (2)$$

$$\gamma_{s_1 s_2} = C^* \cdot \gamma_{s_1} \quad (3)$$

$$\gamma_{s_2} = |C|^2 \cdot \gamma_{s_1}. \quad (4)$$

A partir des observations x_1 et x_2 , la meilleure estimation du signal s_1 au sens de l'erreur quadratique moyenne minimale est donnée par le filtrage de Wiener vectoriel $W = [W_1 W_2]^T$ [Feder]. Les bruits étant décorrelés, l'estimation du spectre de s_1 , notée \hat{S}_1 , est donnée par :

$$\hat{S}_1 = W^T \cdot X \quad (5)$$

où :

$$X = [X_1 X_2]^T \quad (6)$$

et

$$W_1 = \frac{1}{\Delta} \gamma_{s_1} \gamma_{n_2}, \quad (7)$$

$$W_2 = \frac{1}{\Delta} C^* \gamma_{s_1} \gamma_{n_1}, \quad (8)$$

$$\Delta = |C|^2 \gamma_{s_1} \gamma_{n_1} + \gamma_{n_2} \gamma_{s_1} + \gamma_{n_1} \gamma_{n_2}. \quad (9)$$

La mise en place de ce filtrage de Wiener vectoriel nécessite la connaissance des densités spectrales de puissance de s_1 , s_2 , n_1 et n_2 . Dans la pratique, celles-ci doivent être estimées à partir des observations reçues en présence ou en absence de parole. Les dsp des bruits sont apprises en absence de parole, les bruits étant supposés stationnaires. Les autres grandeurs, estimées en présence de signal, sont généralement estimées sur un intervalle de temps court pour tenir compte de sa non stationnarité. Chacune de ces dsp possède une variance, celle des bruits étant plus faible. Le filtrage de Wiener qui dépend d'un grand nombre de grandeurs estimées possède alors une variance importante et se traduit par un traitement peu efficace. Pour pallier cet inconvénient, une structure ISAB (Identification entre Signaux puis Annulation de Bruit) réduisant le nombre d'opérations a été proposée [Tazi], dans laquelle une identification entre signaux est réalisée en absence de bruit afin de déterminer la fonction de transfert entre s_1 et s_2 . En présence de signal bruité, l'observation x_2 est appliquée à un filtre dont la fonction de transfert est celle trouvée précédemment de telle façon que la composante signal issue de ce filtre soit identique à s_1 . En soustrayant l'observation x_2 filtrée de l'observation x_1 , on obtient un bruit utilisé comme référence pour donner par filtrage une estimation du bruit perturbateur n_1 . La contrainte d'un tel système est d'identifier parfaitement la fonction de transfert entre s_1 et s_2 et que celle-ci reste stationnaire en absence et en présence de bruit.

Ceci nous a conduit à développer une nouvelle structure où les opérations de débruitage et d'identification de la fonction de transfert sont effectuées dans un ordre opposé. Nous obtenons la structure PIS (Prétraitement puis Identification entre Signaux) représentée figure 1. La première étape consiste à effectuer une réduction de bruit sur chaque voie. Nous obtenons deux nouvelles grandeurs s'_1 et s'_2 ayant un meilleur rapport signal à bruit (RSB) et qui représentent respectivement des estimées de s_1 et s_2 . La deuxième étape consiste à estimer la fonction de transfert entre les deux grandeurs préfiltrées puis à filtrer s'_2 par la fonction de transfert mesurée pour obtenir une seconde estimée du signal s_1 . Cette estimée est ajoutée à s'_1 et, après division par deux, on obtient l'estimée finale de s_1 . Cette méthode est dénommée PISA. Une variante consiste à identifier la fonction de transfert entre s'_2 et l'observation x_1 , puis à filtrer s'_2 par cette fonction de transfert, les autres opérations restant inchangées. On obtient alors la structure PISB.

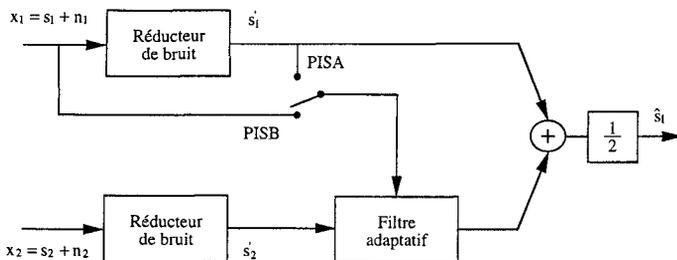


Figure 1. – Schéma de la structure PIS.

Pour la méthode PISA, s'_2 est filtré pour obtenir une estimation du signal contenu dans s'_1 . Celui-ci représente cependant une version distordue du signal s_1 . Tandis que dans la structure PISB, s'_2 est filtré pour obtenir une estimation de s_1 (contenu dans x_1) mais le bruit n_1 présent dans x_1 perturbe davantage l'estimation du filtre. On désignera dans la suite par WM la méthode de Wiener mono-voie et par WB le filtre de Wiener bi-voie.

3. étude théorique des performances

Dans cette étude théorique, les deux méthodes PISA et PISB utilisent un filtrage de Wiener pour effectuer la réduction de bruit sur chaque observation, l'étape d'identification étant également effectuée au sens de Wiener.

Nous évaluons les performances des quatre méthodes présentées en calculant un taux de distorsion, un facteur de réduction de bruit et un gain global. Le calcul des performances est fait à chaque fréquence f et pour plus de clarté, dans les formules données ci-après, la variable f sera sous-entendue. Pour l'étude théorique des performances, nous supposons que les filtres utilisés sont

optimaux. Nous ne prenons pas en compte les erreurs dues au temps d'estimation fini, au temps de convergence, au nombre fini de bits. Les filtres sont déduits des observations présentes et sont appliqués aux signaux seuls pour déterminer la distorsion apportée au signal à estimer s_1 . En appelant s_{1f} la composante signal obtenue en sortie de traitement, la distorsion s'écrit : $\varepsilon_1 = s_1 - s_{1f}$. On calcule la densité spectrale de puissance γ_{ε_1} puis le taux de distorsion D :

$$D = \frac{\gamma_{\varepsilon_1}}{\gamma_{s_1}}. \quad (10)$$

De même, les traitements peuvent être appliqués aux bruits seuls pour connaître le bruit résiduel de sortie n_r et donc sa dsp γ_{n_r} . On en déduit le facteur de réduction de bruit R :

$$R = \frac{\gamma_{n_1}}{\gamma_{n_r}}. \quad (11)$$

Finalement, on évalue l'erreur globale sur le signal à estimer. Celle-ci est à comparer à la grandeur perturbatrice présente dans l'observation x_1 , à savoir n_1 . Soit ε l'erreur globale : $\varepsilon = \varepsilon_1 + n_r$, on a :

$$\gamma_{\varepsilon} = \gamma_{\varepsilon_1} + \gamma_{n_r} = D \cdot \gamma_{s_1} + \frac{\gamma_{n_1}}{R}. \quad (12)$$

Soit G le gain global défini par :

$$G = \frac{\gamma_{n_1}}{\gamma_{\varepsilon}}, \quad (13)$$

on a alors :

$$\frac{1}{G} = D \cdot \frac{\gamma_{s_1}}{\gamma_{n_1}} + \frac{1}{R}. \quad (14)$$

Ces différents termes D , R et G dépendent des RSB de chaque observation x_1 et x_2 notés respectivement :

$$\rho_1 = \frac{\gamma_{s_1}}{\gamma_{n_1}}, \quad \rho_2 = \frac{\gamma_{s_2}}{\gamma_{n_2}}. \quad (15)$$

Pour la comparaison, on suppose que le signal s_1 existe toujours ($\rho_1 > 0$). Quant au signal s_2 , s'il est nul, les filtrages de Wiener mono-voie et bi-voie sont identiques; pour les structures PIS, le filtre d'identification est nul et le signal estimé est deux fois plus faible que celui obtenu par les deux structures de Wiener mono-voie et bi-voie. Dans cette étude, on supposera que les rapports ρ_1 et ρ_2 ne sont jamais nuls. D_i , R_i , G_i ($i = 1, 2, 3, 4$) représentent respectivement le taux de distorsion, le facteur de réduction de bruit et le gain; l'indice 1 se rapporte à la méthode WM, l'indice 2 à la méthode PISA, l'indice 3 à la méthode PISB et l'indice 4 au filtre WB.

3.1. calcul des taux de distorsion

Rappelons que le signal à estimer est le signal s_1 .

3.1.1. Pour la méthode WM, la distorsion sur le signal de sortie s'écrit :

$$\varepsilon_1 = s_1(1 - H_1) \quad (16)$$

Réducteurs de bruit

où

$$H_1 = \frac{\gamma_{s_1}}{\gamma_{x_1}} = \frac{\rho_1}{1 + \rho_1}. \quad (17)$$

Avec les notations choisies précédemment, le taux de distorsion encouru par le signal vaut :

$$D_1 = \frac{1}{(1 + \rho_1)^2}. \quad (18)$$

3.1.2. En ce qui concerne la méthode PISA, le filtrage appliqué à la voie x_1 est identique au filtrage Wiener mono-voie et est donné par H_1 . La réduction de bruit sur la voie x_2 est effectuée par le filtre H_2 avec :

$$H_2 = \frac{\gamma_{s_2}}{\gamma_{x_2}} = \frac{\rho_2}{1 + \rho_2}. \quad (19)$$

L'estimation du signal contenu dans s'_1 à partir de s'_2 est effectuée par le filtre de Wiener H_3 :

$$H_3 = \frac{\gamma_{s'_1 s'_2}}{\gamma_{s'_2}} = \frac{H_1}{C}. \quad (20)$$

L'erreur sur le signal vaut :

$$\varepsilon_1 = s_1 \left[1 - \frac{1}{2} (H_1 + C H_2 H_3) \right]. \quad (21)$$

En remarquant que :

$$\gamma_{n_2} \rho_2 = |C|^2 \gamma_{n_1} \rho_1 \quad (22)$$

le calcul du taux de distorsion, noté D_2 , nous conduit à :

$$D_2 = \frac{(2 + \rho_1 + 2\rho_2)^2}{4(1 + \rho_1)^2(1 + \rho_2)^2}. \quad (23)$$

3.1.3. Pour la méthode PISB, on retrouve les filtres H_1 et H_2 définis précédemment.

L'estimation du signal contenu dans x_1 à partir de s'_2 est effectuée par le filtre de Wiener H'_3 avec :

$$H'_3 = \frac{\gamma_{x_1 s'_2}}{\gamma_{s'_2}} = \frac{1}{C}. \quad (24)$$

L'erreur sur le signal s'écrit :

$$\varepsilon_1 = s_1 \left[1 - \frac{1}{2} (H_1 + C H_2 H'_3) \right] \quad (25)$$

et après calculs, nous obtenons le taux de distorsion D_3 :

$$D_3 = \frac{(2 + \rho_1 + \rho_2)^2}{4(1 + \rho_1)^2(1 + \rho_2)^2}. \quad (26)$$

3.1.4. Pour terminer, nous évaluons le taux de distorsion lorsque l'on utilise le filtre de Wiener vectoriel optimal. Pour la partie signal, on peut écrire :

$$\hat{s}_1 = (W_1 + C W_2) s_1 \quad (27)$$

où W_1 et W_2 ont été définis dans les équations (6), (7) et (8).

L'erreur sur le signal vaut alors :

$$\varepsilon_1 = (1 - W_1 - C W_2) s_1. \quad (28)$$

Finalement, pour le taux de distorsion D_4 , il vient :

$$D_4 = \frac{1}{(1 + \rho_1 + \rho_2)^2}. \quad (29)$$

3.1.5. comparaison des taux de distorsion

La comparaison des 4 taux de distorsion effectuée en supposant ρ_1 et ρ_2 strictement positifs peut s'exprimer suivant l'inégalité entre ρ_1 et ρ_2 :

$$\begin{aligned} D_4 < D_3 \leq D_1 < D_2 & \text{ si } \rho_1 \leq \rho_2 \\ D_4 < D_1 < D_3 < D_2 & \text{ si } \rho_1 > \rho_2. \end{aligned} \quad (30)$$

On constate que, quels que soient ρ_1 et ρ_2 , la méthode WB apporte la plus faible distorsion et la méthode PISA la plus grande. On remarque également que la méthode PISB distord moins que la méthode WM lorsque que le RSB de la voie 2 est plus fort que celui de la voie 1 compensant ainsi un mauvais choix de la voie où l'on cherche à estimer le signal. Ceci montre l'intérêt de cette méthode PISB lorsque l'on dispose d'une voie auxiliaire présentant un meilleur RSB que la voie où s_1 est présent.

3.2. calcul des facteurs de réduction de bruit

Pour chacune des méthodes, nous désirons connaître la réduction de bruit apportée vis-à-vis du bruit perturbateur n_1 . Pour les méthodes opérant sur deux capteurs, le bruit résiduel provient des bruits n_1 et n_2 . Il est plus facile de calculer l'inverse du facteur de réduction de bruit :

$$\frac{1}{R} = \frac{\gamma_{n_r}}{\gamma_{n_1}}, \quad (31)$$

où n_r représente le bruit résiduel en sortie.

3.2.1. Pour le filtrage de Wiener opérant sur la voie x_1 seule, le bruit résiduel est donné par :

$$n_r = H_1 n_1 \quad (32)$$

et on obtient :

$$\frac{1}{R_1} = \frac{\rho_1^2}{(1 + \rho_1)^2}. \quad (33)$$

3.2.2. Dans la méthode PISA, le bruit résiduel est donné par :

$$n_r = H_1 n_1 + H_2 H_3 n_2. \quad (34)$$

La densité spectrale du bruit de sortie vaut :

$$\gamma_{n_r} = |H_1|^2 \gamma_{n_1} + |H_2 H_3|^2 \gamma_{n_2} \quad (35)$$

et après calculs, on obtient :

$$\frac{1}{R_2} = \frac{\rho_1^2(1 + \rho_2^2 + \rho_1 \rho_2 + 2\rho_2)}{4(1 + \rho_1)^2(1 + \rho_2)^2}. \quad (36)$$

3.2.3. En ce qui concerne la méthode PISB, le bruit résiduel s'écrit comme précédemment :

$$n_r = H_1 n_1 + H_2 H_3' n_2. \quad (37)$$

L'inverse du facteur de réduction de bruit devient :

$$\frac{1}{R_3} = \frac{1}{4} \left(\frac{\rho_1^2}{(1 + \rho_1)^2} + \frac{\rho_1 \rho_2}{(1 + \rho_2)^2} \right). \quad (38)$$

3.2.4. Finalement, la réduction de bruit est évaluée pour le filtre de Wiener vectoriel optimal. Le bruit résiduel est donné par :

$$n_r = W_1 n_1 + W_2 n_2. \quad (39)$$

À partir de l'expression des filtres donnée dans les équations (7) à (9), on obtient :

$$\frac{1}{R_4} = \frac{\rho_1^2 + \rho_1 \rho_2}{(1 + \rho_1 + \rho_2)^2}. \quad (40)$$

3.2.5. comparaison des facteurs de réduction de bruit

Nous comparons d'abord les différentes méthodes à la méthode WM.

Si l'on compare les méthodes WM et PISA, les calculs montrent que :

$$R_1 \geq R_2 \text{ si } \rho_1 \geq 12 \text{ et } \rho_{2a} \leq \rho_2 \leq \rho_{2b}. \quad (41)$$

ρ_{2a} et ρ_{2b} sont des valeurs critiques définies par :

$$\rho_{2a} = \frac{\rho_1 - 6 - \sqrt{\rho_1^2 - 12\rho_1}}{6} \quad (42)$$

et

$$\rho_{2b} = \frac{\rho_1 - 6 + \sqrt{\rho_1^2 - 12\rho_1}}{6}. \quad (43)$$

Si l'on compare maintenant les méthodes WM et PISB, on montre que :

$$R_1 \geq R_3 \text{ si } (\rho_1 \leq 0,1 \text{ ou } \rho_1 \geq 9,9) \text{ et } \rho_{2c} \leq \rho_2 \leq \rho_{2d}. \quad (44)$$

Les valeurs critiques ρ_{2c} et ρ_{2d} sont données par :

$$\rho_{2c} = \frac{\rho_1^2 - 4\rho_1 + 1 - \sqrt{(\rho_1^2 - 4\rho_1 + 1)^2 - 36\rho_1^2}}{6\rho_1} \quad (45)$$

et

$$\rho_{2d} = \frac{\rho_1^2 - 4\rho_1 + 1 + \sqrt{(\rho_1^2 - 4\rho_1 + 1)^2 - 36\rho_1^2}}{6\rho_1}. \quad (46)$$

Finalement, la comparaison des méthodes WM et WB donne :

$$R_4 \geq R_1 \text{ si } \rho_2 \geq \rho_{2e} = \frac{1 - \rho_1^2}{\rho_1}. \quad (47)$$

Nous avons tracé ces différentes valeurs critiques (en dB) en fonction de ρ_1 (en dB) afin de délimiter les régions du plan (ρ_1, ρ_2) où une méthode réduit mieux le bruit qu'une autre (figures 2 à 4). Les méthodes PIS apparaissent plus performantes que le filtrage de Wiener mono-voie pour un RSB sur la première voie inférieur à 10,8 dB pour la méthode PISA (figure 2) et compris entre -9,96 dB et +9,96 dB pour la méthode PISB (figure 3). La figure 4 montre que pour un RSB de cette voie supérieur à 0 dB, le filtrage de Wiener bi-voie réduit mieux le bruit que le filtrage de Wiener mono-voie. Si ce RSB devient inférieur à 0 dB, le RSB de la voie auxiliaire doit être supérieur à la valeur critique donnée par la

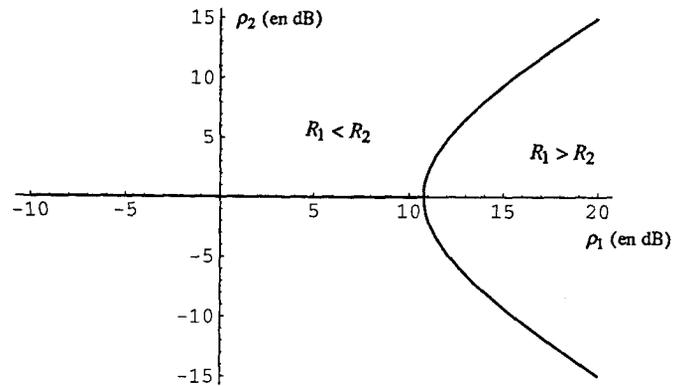


Figure 2. – Régions du plan (ρ_1, ρ_2) où $R_1 > R_2, R_1 < R_2$.

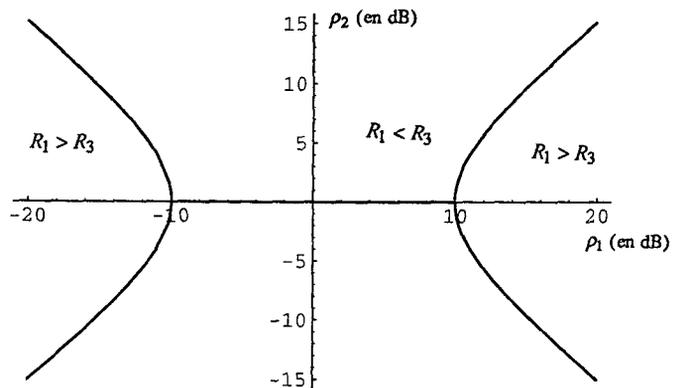


Figure 3. – Régions du plan (ρ_1, ρ_2) où $R_1 > R_3, R_1 < R_3$.

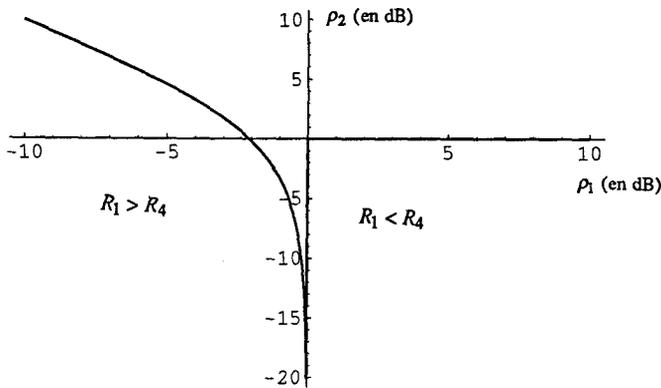


Figure 4. – Régions du plan (ρ_1, ρ_2) où $R_1 > R_4, R_1 < R_4$.

courbe pour que le filtrage de Wiener bi-voie soit préférable en terme de réduction de bruit.

En ce qui concerne les deux méthodes PIS, on a :

$$R_2 > R_3, \forall \rho_1, \rho_2. \quad (48)$$

Par contre, il s'avère difficile de comparer la méthode WB aux méthodes PISA et PISB à partir des expressions des réductions de bruit correspondantes, les équations auxquelles on aboutit étant de degré 4 en ρ_1 et ρ_2 . Les courbes correspondant à $R_2 = R_4$ puis à $R_3 = R_4$ ont été obtenues avec le logiciel Mathematica permettant de résoudre ce type d'équations. Les figures 5 et 6 délimitent les régions du plan (ρ_1, ρ_2) où les méthodes PISA et PISB réduisent mieux le bruit que le filtrage de Wiener bi-voie. L'examen de ces figures montre que les deux méthodes PIS réduisent davantage le bruit que le filtrage de Wiener bi-voie dans la région définie par des RSB sur les voies 1 et 2 inférieurs à 9 dB.

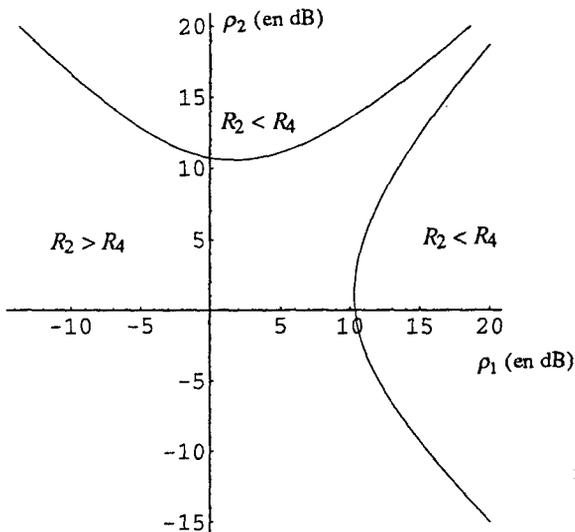


Figure 5. – Régions du plan (ρ_1, ρ_2) où $R_2 > R_4, R_2 < R_4$.

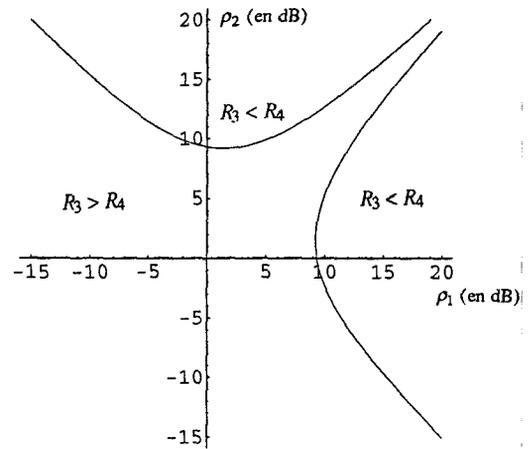


Figure 6. – Régions du plan (ρ_1, ρ_2) où $R_3 > R_4, R_3 < R_4$.

3.3. calcul des erreurs globales

Les paragraphes 3.1. et 3.2. conduisent tout naturellement à l'estimation de l'erreur globale pour les quatre méthodes étudiées. Selon l'équation (14), on a pour chaque méthode l'écriture de l'inverse du gain global G_i :

$$\frac{1}{G_i} = \rho_1 \cdot D_i + \frac{1}{R_i}. \quad (49)$$

La méthode est d'autant plus performante que le gain G_i est élevé. D'après les calculs précédents, nous obtenons pour les quatre méthodes étudiées :

$$\frac{1}{G_1} = \frac{\rho_1}{(1 + \rho_1)}, \quad (50)$$

$$\frac{1}{G_2} = \frac{\rho_1(2 + \rho_1 + 2\rho_2)^2 + \rho_1^2(1 + \rho_2^2 + \rho_1\rho_2 + 2\rho_2)}{4(1 + \rho_1)^2(1 + \rho_2)^2}, \quad (51)$$

$$\frac{1}{G_3} = \frac{\rho_1(2 + \rho_1 + \rho_2)^2 + \rho_1^2(1 + \rho_2)^2 + \rho_1\rho_2(1 + \rho_1)^2}{4(1 + \rho_1)^2(1 + \rho_2)^2}, \quad (52)$$

$$\frac{1}{G_4} = \frac{\rho_1}{1 + \rho_1 + \rho_2}. \quad (53)$$

Il va sans dire que la méthode WB présente un gain global supérieur à celui des trois autres méthodes quels que soient les rapports signal à bruit d'entrée.

En ce qui concerne la comparaison des méthodes PISA et PISB, elle se réduit à la simple expression :

$$G_3 > G_2 \quad \forall \rho_1, \rho_2. \quad (54)$$

La méthode PISB apparaît plus performante que PISA, cependant cette dernière réduit davantage le bruit. Quant à la comparaison des gains des méthodes PISA et PISB vis-à-vis de la méthode

WM, il s'avère nécessaire d'introduire deux grandeurs critiques ρ_{2f} et ρ_{2g} :

$$\rho_{2f} = \frac{\rho_1 + 1}{3}, \quad (55)$$

$$\rho_{2g} = \frac{\rho_1}{3}, \quad (56)$$

telles que :

$$\text{si } \rho_2 \geq \rho_{2f} \quad G_2 \geq G_1 \quad (57)$$

et

$$\text{si } \rho_2 \geq \rho_{2g} \quad G_3 \geq G_1. \quad (58)$$

La figure 7 donne les différentes régions du plan (ρ_1, ρ_2) où les gains G_1, G_2, G_3 sont comparés. Pour toute combinaison (ρ_1, ρ_2) située au-dessus de la courbe (a), les deux méthodes PIS sont supérieures au filtrage de Wiener mono-voie tandis que pour les combinaisons (ρ_1, ρ_2) situées en dessous de la courbe (b), le filtrage de Wiener mono-voie est plus performant. Pour évaluer la différence des gains obtenus par les méthodes WM, PISA et PISB, nous avons également représenté (figures 8 et 9) les valeurs de $G_1/G_2, G_1/G_3$ (en dB) en fonction des RSB. Pour des rapports signal à bruit sur la première voie faibles, les deux méthodes PIS

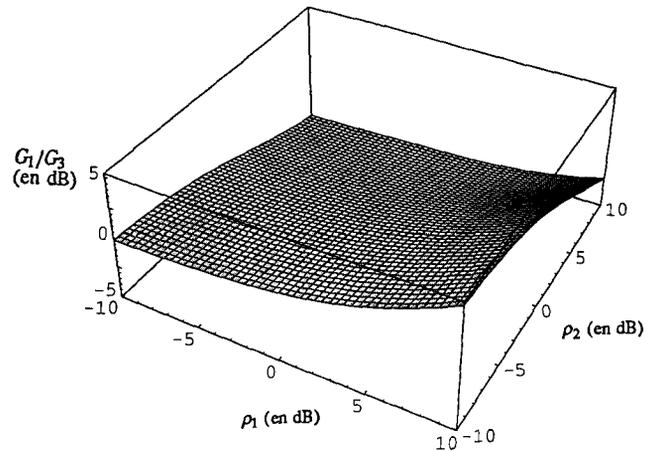


Figure 9. – Tracé de G_1/G_3 (en dB).

sont équivalentes voire plus performantes que le filtrage de Wiener mono-voie appliqué sur l'observation x_1 . Cette conclusion montre l'intérêt de ces deux méthodes et justifie leur utilisation pour le débruitage de la parole.

4. mise en œuvre des méthodes et résultats

Nous appliquons maintenant les différentes méthodes à des signaux réels enregistrés dans une voiture. Nous indiquons tout d'abord comment ces algorithmes sont mis en œuvre, puis nous présentons quelques résultats objectifs et subjectifs. Pour les quatre méthodes, le traitement est fréquentiel et est effectué par bloc de 256 échantillons, avec un recouvrement de 50%. Pour apprendre les caractéristiques de bruit, il est nécessaire de connaître les instants de début et fin d'activité vocale. Ceux-ci ont été déterminés par un étiquetage manuel sur la parole pure.

Ephraïm et Malah proposent un algorithme pour l'estimation de parole dans le cas mono-voie basé sur l'estimation de l'amplitude de chaque composante spectrale [Ephraïm]. Cet algorithme réduit mieux le bruit musical [Cappé] que la soustraction spectrale proposée par [Berouti]. Pour des rapports signal à bruit élevés, celui-ci se simplifie et conduit à un filtrage de Wiener prenant en compte l'incertitude sur la présence du signal. Il est utilisé comme filtre de Wiener mono-voie et également comme filtre appliqué à chacune des observations pour les deux méthodes PIS. Nous rappelons tout d'abord l'algorithme simplifié d'Ephraïm et Malah. Nous donnons ici une présentation générale où l'observation notée x est composée d'un signal à estimer s et d'un bruit perturbateur n . Soient $S(f, m), N(f, m), X(f, m)$ les composantes spectrales respectives à la fréquence f du signal s , du bruit n , et de l'observation x pour le bloc m :

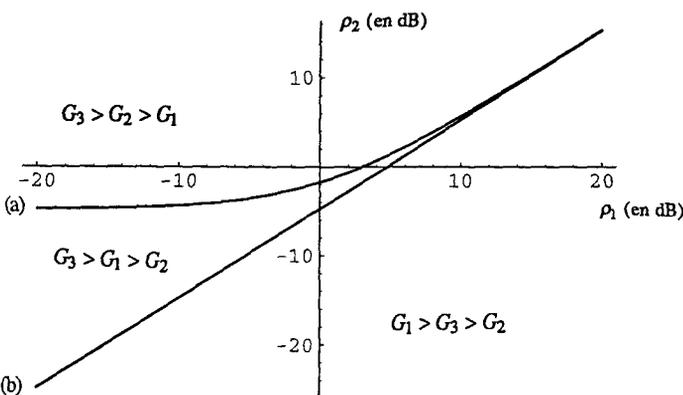


Figure 7. – Régions du plan (ρ_1, ρ_2) où $G_3 > G_2 > G_1, G_3 > G_1 > G_2, G_1 > G_3 > G_2$.

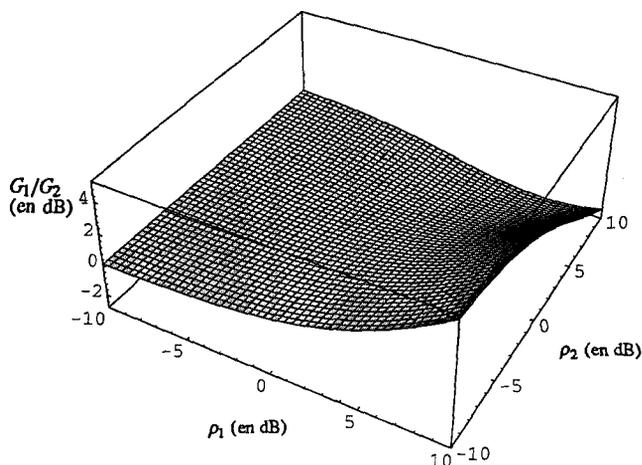


Figure 8. – Tracé de G_1/G_2 (en dB).

Réducteurs de bruit

$$\begin{aligned} S(f, m) &= A(f, m) \exp j\theta_s(f, m), \\ N(f, m) &= D(f, m) \exp j\theta_n(f, m), \\ X(f, m) &= R(f, m) \exp j\theta_x(f, m). \end{aligned} \quad (59)$$

Définissons les rapports signal à bruit a priori et a posteriori respectivement par :

$$\xi(f, m) = \frac{E[A^2(f, m)]}{E[|N(f, m)|^2]} \quad (60)$$

$$\eta(f, m) = \frac{R^2(f, m)}{E[|N(f, m)|^2]}. \quad (61)$$

$E[|N(f, m)|^2]$ ne peut être directement estimé et sera remplacé par la dsp du bruit $\gamma_n(f)$ estimée en absence de parole en supposant celui-ci stationnaire. L'estimateur $\hat{A}(f, m)$ qui minimise l'erreur quadratique moyenne sur l'amplitude du signal s est donné par :

$$\hat{A}(f, m) = T_1(f, m) T_2(f, m) R(f, m) \quad (62)$$

où $T_1(f, m)$ représente simplement, dans le cas de fort RSB, le filtre de Wiener défini par :

$$T_1(f, m) = \frac{\xi(f, m)}{1 + \xi(f, m)} \quad (63)$$

et $T_2(f, m)$ prend en compte l'incertitude sur la présence du signal :

$$T_2(f, m) = \frac{\Lambda(f, m)}{1 + \Lambda(f, m)}, \quad (64)$$

$\Lambda(f, m)$ est le rapport de vraisemblance généralisé qui, en supposant un modèle gaussien pour les composantes spectrales, est donné par :

$$\Lambda(f, m) = \frac{1 - q(f, m)}{q(f, m)} \cdot \frac{\exp\left(\frac{\xi(f, m)}{1 + \xi(f, m)} \eta(f, m)\right)}{1 + \xi(f, m)}. \quad (65)$$

$q(f, m)$ représente la probabilité d'absence du signal. Comme dans [Ephraïm], nous avons fixé cette probabilité $q(f, m)$ à une valeur constante égale à 0.2.

Un problème important est l'estimation du RSB a priori de chaque composante spectrale à l'intérieur de chaque bloc traité. La meilleure approche, selon [Ephraïm], est donnée par :

$$\hat{\xi}(f, m) = \alpha \frac{\hat{A}^2(f, m-1)}{\gamma_n(f)} + (1 - \alpha) P[\eta(f, m) - 1], \quad (66)$$

avec :

$$P[u] = \begin{cases} u & \text{si } u \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (67)$$

où α est un facteur de pondération (choisi égal à 0.98) et $P[u]$ est un opérateur pour assurer que le second terme de (66) est positif. $\hat{A}(f, m-1)$ représente l'estimée du signal obtenue au

bloc précédent à la fréquence f . La valeur du RSB a posteriori $\eta(f, m)$ est donnée à chaque bloc m par le rapport de l'estimée du module du spectre de l'observation à la dsp du bruit apprise $\gamma_n(f)$. Celle-ci est obtenue par un moyennage des dsp de bruit sur les 10 derniers blocs adjacents de 32 ms chacun avant l'apparition de la parole :

$$\gamma_n(f) = \frac{1}{10} \sum_{k=1}^{10} |N(f, k)|^2. \quad (68)$$

Nous donnons maintenant les algorithmes utilisés dans les deux méthodes PIS. Les étages de réduction de bruit appliqués à chaque observation sont donnés par l'algorithme d'Ephraïm et Malah présenté précédemment. Soient s'_1 et s'_2 les sorties de ces deux étages. Le filtre appliqué à s'_2 est calculé soit à partir de s'_1 et s'_2 soit à partir de x_1 et s'_2 . Pour la méthode PISA, ce filtre s'écrit :

$$H_3(f, m) = \frac{\hat{\gamma}_{s'_1 s'_2}(f, m)}{\hat{\gamma}_{s'_2 s'_2}(f, m)} \quad (69)$$

et pour la méthode PISB, ce filtre est donné par :

$$H'_3(f, m) = \frac{\hat{\gamma}_{x_1 s'_2}(f, m)}{\hat{\gamma}_{s'_2 s'_2}(f, m)} \quad (70)$$

Les densités spectrales sont calculées de façon récursive :

$$\hat{\gamma}_{uv}(f, m) = \lambda \hat{\gamma}_{uv}(f, m-1) + (1 - \lambda) U(f, m) V^*(f, m), \quad (71)$$

où λ est un facteur d'oubli. $\hat{\gamma}_{s'_1 s'_2}(f, m)$, $\hat{\gamma}_{s'_2 s'_2}(f, m)$ et $\hat{\gamma}_{x_1 s'_2}(f, m)$ sont obtenues avec $\lambda = 0.7$.

Le filtrage de Wiener vectoriel est donné par les équations (5) à (9). Il peut être réalisé directement à partir de celles-ci. D'autres expressions des filtres W_1 et W_2 peuvent également être obtenues en utilisant le fait que les signaux et les bruits sont additifs et indépendants. Nous avons choisi de tester le filtre de Wiener à partir de filtres W_1 et W_2 ne nécessitant pas d'opérations de soustraction pour obtenir les dsp des signaux s_1 et s_2 . Les expressions proposées pour les filtres W_1 et W_2 utilisent seulement les grandeurs directement accessibles : les dsp des observations, la densité interspectrale entre observations ainsi que les dsp des bruits apprises pendant les périodes de silence. W_1 et W_2 sont donnés ainsi :

$$W_1(f, m) = 1 - \frac{\gamma_{n_1}(f) \gamma_{x_2 x_2}(f, m)}{\Delta(f, m)} \quad (72)$$

$$W_2(f, m) = \frac{\gamma_{x_1 x_2}(f, m) \gamma_{n_1}(f)}{\Delta(f, m)} \quad (73)$$

où

$$\Delta(f, m) = \gamma_{x_1 x_1}(f, m) \gamma_{n_2}(f) + \gamma_{n_1}(f) \gamma_{x_2 x_2}(f, m) - \gamma_{n_1}(f) \gamma_{n_2}(f) \quad (74)$$

Les différentes densités spectrales $\gamma_{x_1 x_1}(f, m)$, $\gamma_{x_2 x_2}(f, m)$, $\gamma_{x_1 x_2}(f, m)$ sont estimées par l'équation (71) avec un facteur d'oubli égal à 0.5. Les dsp des bruits $\gamma_{n_1}(f)$ et $\gamma_{n_2}(f)$ sont déterminées comme dans l'algorithme d'Ephraïm et Malah par moyennage sur les 10 derniers blocs précédant le début de l'activité vocale.

Pour évaluer les performances réelles des filtrages de Wiener mono-voie et bi-voie ainsi que des méthodes PIS, des enregistrements ont été faits dans une Renault 18. Le premier microphone, duquel on doit extraire le signal de parole à transmettre, est situé sur le tableau de bord en face du conducteur, tandis que le second microphone, utilisé pour les méthodes PIS, est placé sur le doublage intérieur de pied avant gauche (montant situé entre le pare-brise et la portière avant gauche). La distance entre les deux microphones est environ 40 cm. Les mesures objectives que nous avons retenues sont le gain segmental sur le rapport signal à bruit (75) et la distance cepstrale (76). Mieux corrélé avec les tests d'écoute que le RSB conventionnel, le rapport signal à bruit segmental RSB_{seg} se note :

$$RSB_{seg} = \frac{1}{L} \sum_{k=0}^{L-1} 10 \log_{10} \left(\frac{\sum_{i=0}^{N-1} s^2(i + Nk)}{\sum_{i=0}^{N-1} (s(i + Nk) - s'(i + Nk))^2} \right), \quad (75)$$

s est le signal de parole pur et s' représente soit le signal bruité à l'entrée du réducteur (calcul du rapport en entrée) soit le signal débruité en sortie (calcul du rapport de sortie). Les signaux sont divisés en L blocs de N échantillons chacun. La quantité RSB_{seg} est obtenue en moyennant le RSB exprimé en décibels sur tous les blocs. Dans notre application, la valeur de N est fixée à 256 (ce qui correspond à une tranche de signal de 32 ms pour un signal échantillonné à 8 kHz). Le gain G est naturellement calculé en soustrayant le RSB_{seg} d'entrée au RSB_{seg} de sortie.

La distance cepstrale est une version tronquée de la norme L_2 de la différence des logarithmes des spectres. Elle est définie au bloc k comme suit :

$$d_{cep} = \sum_{n=1}^{2p} (c(n, k) - c'(n, k))^2 \quad (76)$$

où $c(n, k)$ et $c'(n, k)$ sont les coefficients cepstraux correspondant à s et s' respectivement; l'ordre p du modèle est fixé à 8. La distance cepstrale finale d_{cep} est obtenue par moyennage sur les différents blocs.

Dans les équations (75) et (76) le signal de parole pur à considérer sera le signal s_1 et dans l'équation (76), seule la distance entre signal pur et signal estimé sera évaluée. Les méthodes seront d'autant plus performantes que le gain G est élevé et d_{cep} faible.

Pour nos expérimentations, les signaux de parole et de bruit ont été enregistrés séparément. Le bruit est stationnaire et correspond à la voiture roulant à 130 km/h. Chaque fichier de parole est constitué de deux phrases prononcées par 4 locuteurs différents (1 locuteur français, 1 locuteur anglais, 1 locutrice française, 1 locutrice anglaise), chaque locuteur prononçant 5 couples de phrases. Les signaux de parole et les bruits sont ajoutés de manière à obtenir des rapports signal à bruit de 0 dB et 3 dB calculés sur la durée du signal utile. Les valeurs du gain et de la distance cepstrale sont moyennées sur les périodes d'activité vocale des 20 fichiers construits.

Cette évaluation objective est complétée par des tests subjectifs d'écoute conduits sur les mêmes signaux. Les signaux obtenus par les 4 traitements ont été enregistrés 3 fois sur une cassette DAT de façon aléatoire. La durée du test réalisé en deux temps est d'environ 40 minutes. Nous avons demandé à 15 auditeurs de juger la qualité des signaux en donnant une note comprise entre 1 et 5 selon l'échelle suivante :

- 1 : inacceptable
- 2 : pauvre ou insuffisant
- 3 : médiocre
- 4 : bon
- 5 : excellent.

Les tableaux 1 et 2 donnent, pour des rapports signal à bruit d'entrée calculés sur la durée de l'activité vocale respectivement égaux à 0 dB et 3 dB, les valeurs moyennes du gain, de la distance cepstrale et de la note d'écoute. Les résultats objectifs indiquent la supériorité des méthodes PIS par rapport aux filtrages de Wiener mono-voie et bi-voie. Notons que le filtrage de Wiener bi-voie semble n'apporter sur le plan objectif aucune amélioration. Les méthodes PISA et PISB présentent des mesures objectives voisines. Les mesures subjectives confirment la supériorité des méthodes PIS par rapport aux filtrages de Wiener. Le filtrage de Wiener bi-voie possède un bruit résiduel intermittent et non stationnaire. Le filtrage de Wiener mono-voie ne fait pas apparaître de bruit musical mais un bruit résiduel subsiste. Pour les méthodes PIS, le bruit résiduel est réduit et devient quasiment inexistant pour PISA, au détriment cependant d'une distorsion sur le signal plus importante pour un RSB de 0 dB. Le test d'écoute indique cependant une légère préférence pour la méthode PISA.

Tableau 1 : Evaluation des performances pour un RSB d'entrée égal à 0 dB.

	G (en dB)	dcep	note subjective
méthode WM	12,83	0,5352	2,75
méthode PISA	13,06	0,5749	3,53
méthode PISB	13,03	0,5547	3,41
méthode WB	0,75	0,7220	1,51

Tableau 2 : Evaluation des performances pour un RSB d'entrée égal à 3 dB.

	G (en dB)	dcep	note subjective
méthode WM	11,73	0,4156	3,72
méthode PISA	12,09	0,3818	4,50
méthode PISB	12,01	0,3499	4,28
méthode WB	0,77	0,5847	1,97

5. conclusion

Dans le cadre du débruitage de la parole, deux méthodes PIS ont été proposées. Une étude théorique permet de comparer ces deux méthodes aux filtrages de Wiener mono-voie et bi-voie. Si ce dernier donne les meilleures performances théoriques en terme de gain, dans la pratique, lorsque l'on travaille sur des signaux non stationnaires, celui-ci ne s'avère pas intéressant. Sur le plan théorique, les deux méthodes PIS sont équivalentes ou plus performantes que le filtrage de Wiener mono-voie, vis-à-vis du facteur de réduction de bruit et du gain, pour des rapports signal à bruit de la première voie faibles. Des tests pratiques montrent l'intérêt des deux méthodes PIS au prix d'un accroissement de complexité tout à fait acceptable vis-à-vis d'un filtrage de Wiener mono-voie.

LES AUTEURS

Régine LE BOUQUIN



Régine Le Bouquin Jeannès a obtenu son doctorat de l'Université de Rennes 1 en Traitement du Signal et Télécommunications en 1991 et est Maître de Conférences à l'Université de Rennes 1 depuis 1992. Ses travaux de recherche sont effectués au Laboratoire de Traitement du Signal et de l'Image et portent d'une part sur le traitement de la parole, et plus particulièrement sur les thèmes de la réduction de bruit et de l'annulation d'écho acoustique dans le contexte des télécommunications mains-libres, et d'autre part sur l'analyse de signaux stéréoélectroencéphalographiques orientée vers les méthodes temps-fréquence.

BIBLIOGRAPHIE

- [Faucon] G. Faucon, R. Le Bouquin, S. Tazi Mezalek, « Proposal and Comparison of Two Methods for Noise Reduction on Speech Signals in a Car », *22nd International Symposium on Automotive Technology and Automation*, Florence, May 14-18 1990, pp. 501-508.
- [Feder] M. Feder, A.V. Oppenheim, E. Weinstein, « Maximum Likelihood Noise Cancellation Using the EM Algorithm », *IEEE Trans. on ASSP*, vol. ASSP-37, n°2, February 1989, pp. 204-216.
- [Tazi] S. Tazi Mezalek, « Algorithmes pour le Débruitage de la Parole », *Thèse de l'Université de Rennes 1*, 25 Septembre 1990.
- [Ephraim] 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.
- [Cappé] O. Cappé, « Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor », *IEEE Trans. on Speech and Audio Processing*, vol. 2, n°2, April 1994, pp. 345-349.
- [Berouti] M. Berouti, R. Schwartz and J. Makhoul, « Enhancement of Speech Corrupted by Acoustic Noise », *ICASSP*, April 1979, pp. 208-211.

Manuscrit reçu le 27 Février 1995.

Gérard FAUCON



Gérard Faucon est Professeur à l'Université de Rennes 1 depuis 1991. Son travail de recherche se déroule au Laboratoire de Traitement du Signal et de l'Image et porte notamment sur le filtrage adaptatif, la réduction de bruit, le rehaussement de la parole dans les télécommunications mains-libres et l'étude des signaux stéréoélectroencéphalographiques.