

Réduction de bruit et traitements paramétriques de la parole en large bande destinés à améliorer la compréhension des sujets malentendants

A. Goyé¹, J. Wable¹, G. Guilmin², B. Frachet¹

¹MEDICOM / CRISTAL, ORL - Avicenne, 125, rue de Stalingrad, 93009 Bobigny Cedex

alain.goye@avc.ap-hop-paris.fr, jocelyne.wable@avc.ap-hop-paris.fr, bruno.frachet@avc.ap-hop-paris.fr

²THALES Communications, 66, rue du fosse blanc - BP 156, 92231 Gennevilliers Cedex

gwenael.guilmin@fr.thalesgroup.com

Résumé – Nous tentons de pallier les déficiences des aides auditives actuelles pour certaines surdités, par une plus grande prise en compte de la structure particulière du signal de parole. Pour cela nous voulons intégrer à une aide auditive des traitements paramétriques en bande élargie (0-8000 Hz). L'appareil doit savoir isoler les signaux qui bénéficieront de ces traitements. Il s'agit exclusivement des signaux de parole suffisamment peu bruités. Cette étude vise donc à déterminer les limites d'applicabilité des traitements paramétriques, après passage du signal dans un réducteur de bruit. Le réducteur de bruit basé sur un filtrage de Wiener est intégré au codeur. Les performances sont testées par l'évaluation, pour des normo-entendants et des malentendants : (1) de la capacité auditive (CA), via la mesure de la capacité de discrimination des indices acoustico-phonétiques de la parole en situation de bruit, (2) du confort auditif, par un test de préférence sur des phrases traitées et non traitées, prononcées dans diverses situations sonores significatives. Alors que la réduction de bruit isolée peut améliorer la CA chez certains normo-entendants uniquement, le codage paramétrique détériore la CA pour l'ensemble des sujets, ainsi que la qualité de la parole perçue par les malentendants.

Abstract – To better compensate the weaknesses of current hearing aids for some hearing losses, we try to take more into account the specificities of the speech signal. Therefore, we wish to integrate into a hearing aid wideband (0-8000 Hz) parametric processing. The device will need to be able to extract the signals which can benefit from this processing. Only speech signals with limited noise may be selected. Hence this study aims at determining the limits of applicability of such parametric processing, after having the signal pre-processed by a noise reduction system. This noise reduction system, which is based on Wiener filtering, is incorporated in the coder. Performance is tested through evaluation, for normal-hearing and hearing-impaired listeners: (1) of auditory capacity (CA), with the measurement of their capacity to discriminate acoustico-phonetic features of speech in noise. (2) of auditory comfort, with a preference test on processed and unprocessed sentences, uttered in several significant sound situations. While noise reduction alone may improve CA but only for some normal-hearing subjects, parametric coding deteriorates CA for the whole set of subjects, as well as it harms the perceived quality of speech for hearing-impaired people.

1. Contexte : le projet AUDIVOCAL

Les aides auditives proposées actuellement ne satisfont pas une grande partie des surdités “ sévères ” ou “ moyennes ”. Ne pouvant pallier que partiellement les distorsions liées aux déficiences perceptives, il est souhaitable de prendre plus en compte la structure particulière du signal de parole. Le projet AUDIVOCAL a pour but d'intégrer les nouvelles technologies vocales numériques dans les aides auditives pour en augmenter les performances en termes d'intelligibilité et d'agrément d'écoute.

Les techniques de traitement paramétrique sont largement utilisées pour la compression de parole en téléphonie. Judicieusement appliquées dans un système d'aide auditive, elles pourraient améliorer l'intelligibilité de la parole ; cependant elles sont spécifiques au signal vocal. Le projet AUDIVOCAL prévoit donc une classification automatique des situations pour n'activer ces nouveaux traitements

paramétriques que lorsqu'ils sont réellement applicables, c'est-à-dire en présence de parole suffisamment peu bruitée.

La présente étude vise à déterminer cette limite d'applicabilité des traitements paramétriques, après passage du signal dans un réducteur de bruit.

2. Procédés de traitement

2.1 Réduction de bruit

Une grande partie des méthodes existantes aborde le problème de la réduction de bruit sous son aspect fréquentiel afin de profiter des caractéristiques de ce domaine d'étude pour appliquer de nombreuses méthodes numériques. Parallèlement aux nombreuses variantes de la soustraction spectrale, les méthodes s'inspirant d'un filtrage optimal se sont développées depuis les travaux réalisés par Wiener dès

1949 [1][2]. Celles-ci prennent en compte souvent la statistique du signal de parole [3].

Ces méthodes appliquent des modifications sur la transformée de Fourier court terme du signal bruité, en utilisant les informations *a priori* disponibles sur le bruit.

Soient la transformée de Fourier court-terme $Y(f)$ du signal observé $y(n)$, et $S(f)$ et $B(f)$ les transformées de Fourier court-terme de $s(n)$, le signal vocal utile, et $b(n)$ le bruit ambiant additif. A partir des statistiques du signal de parole, il est possible de déterminer un filtre optimal $H(f)$, soit dans le domaine temporel soit dans le domaine fréquentiel, afin d'obtenir une estimée du signal utile :

$$\hat{S}(f) = H(f).Y(f) \quad (1)$$

2.1.1 Méthode de réduction de bruit combinée

Le filtrage de Wiener intégrant un terme d'incertitude sur la présence du signal de parole donne des résultats prometteurs sur l'estimation des paramètres du vocodeur développé et sur l'intelligibilité du signal codé [4]. Seule l'estimation de l'enveloppe spectrale ne semble pas optimale.

Ce filtre est obtenu en minimisant l'erreur quadratique moyenne sur les amplitudes, ce qui revient à calculer l'espérance conditionnelle :

$$|\hat{S}(f)| = E\{S(f) | Y(f), H_1\} \cdot p(H_1 | Y(f)) \quad (2)$$

Le filtrage de Wiener basé sur la modélisation auto-régressive du signal de parole (filtre de Lim) [5], qui s'attache principalement à la structure formantique du signal, obtient en revanche des résultats prometteurs au niveau de l'enveloppe spectrale du signal.

Le signal de parole est alors supposé suivre un modèle auto-régressif (AR) où chaque échantillon du signal de parole dépend des échantillons précédents et d'une excitation :

$$s(n) = \sum_{k=1}^P \alpha_k \cdot s(n-k) + g \cdot u(n) \quad (3)$$

La limitation sur le coût de calcul global du système nous empêche d'utiliser deux systèmes parallèles et de combiner les paramètres pour lesquels chacun est le plus performant. Nous nous proposons donc d'étudier le comportement d'un filtre de pré-traitement combinant linéairement ces deux algorithmes, en faisant varier le facteur de combinaison en fonction de la densité spectrale de puissance du bruit estimé. Cette méthode ne nécessite alors qu'un seul niveau de reconstruction [6].

2.1.2 Application du système

Le traitement combiné découpe le signal en trames de 180 échantillons, soit 11,25 ms avec une fréquence d'échantillonnage de $F_e = 16$ kHz, et travaille dans le domaine spectral sur 512 points, en utilisant la technique du *zero-padding*, pour avoir une seule trame de retard à la reconstruction par overlap-add.

La Figure 1 présente le résultat obtenu par la réduction de bruit proposée, sur un signal original bruité avec un rapport

signal à bruit de +10dB. La structure harmonique est accentuée (raies horizontales) et la structure formantique est restituée.

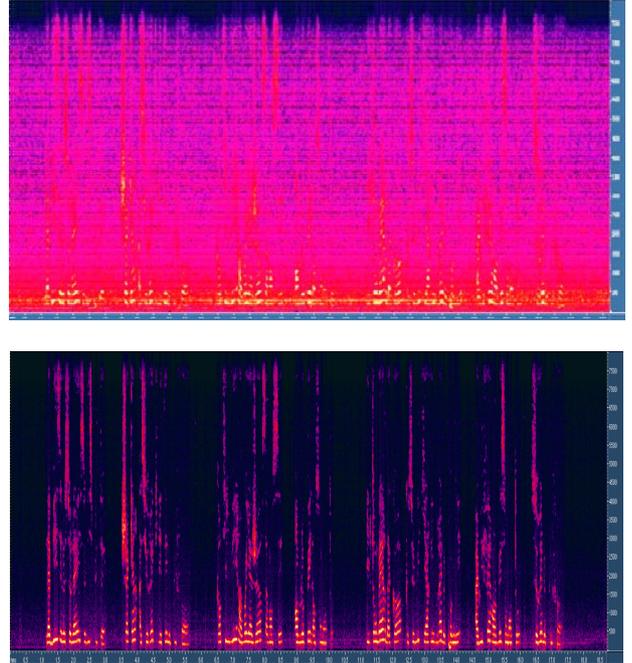


FIG. 1 : Résultat de la réduction du bruit ambiant : signal original large bande (en haut) et signal débruité (en bas).

2.2 Codage en bande élargie

L'étude bibliographique sur le codage de parole large bande (ou *Wide Band* : WB), c'est-à-dire le passage de la bande téléphonique 300-3400 Hz à la bande 50-7000 Hz, nous a permis de proposer différentes configurations algorithmiques pour le développement d'un tel système.

Notre choix s'est porté sur un développement pleine bande (*full band*). Cette technique utilise la méthode de base, dite paramétrique, du codeur HSX (*Harmonic and Stochastic Excitation*) développé par THALES Communications [7], en l'étendant à la bande élargie. La bande basse peut être considérée comme voisée ou non voisée, alors que la bande haute est non voisée. La fréquence de séparation F_c (ou « fréquence de coupure ») entre ces deux bandes peut avoir différentes positions pré-définies inférieures à 4 kHz.

La longueur des trames de traitement a été conservée à 22,5 ms (soit un doublement du nombre d'échantillons par rapport au codeur bande étroite).

Les premiers résultats subjectifs obtenus par ce système montrent une réelle amélioration de l'intelligibilité de la voix et également de la qualité par la restitution d'un son beaucoup plus naturel.

La Figure 2 montre le spectrogramme d'un signal de parole original large bande, puis vocodé en bande élargie. En bande étroite, certains sons perdent une grande partie de leur énergie (les chuintement et les sifflement tels que les « s », « ch », ...) : l'intelligibilité du signal est par conséquent réduite en bande étroite. En bande élargie ces sons, très importants pour

l'intelligibilité, sont conservés, ce qui permet une meilleure restitution du signal.

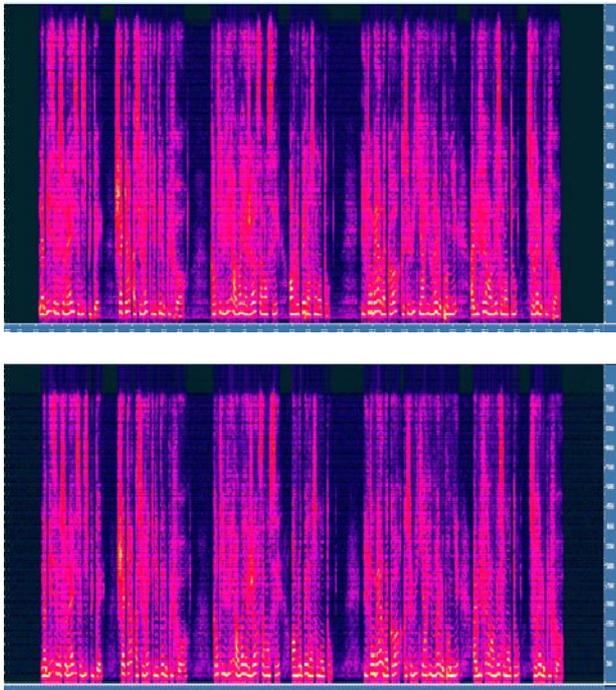


FIG. 2 : Résultat de l'analyse paramétrique : signal original large bande (en haut) et signal après analyse paramétrique en bande élargie (en bas)

3. Méthodologie des tests

Les tests sont effectués par 10 sujets normo-entendants et par 25 malentendants volontaires, d'âge et de sexe indifférent, porteurs d'une déficience auditive à composante perceptive importante donc susceptibles de bénéficier des traitements proposés. La déficience auditive est évaluée par audiométrie tonale et vocale.

Dans tous les tests le niveau sonore global est réglé au niveau de confort maximum du sujet.

3.1 Capacité auditive

Nous évaluons d'abord la capacité auditive (CA), à travers les capacités de discrimination des indices acoustico-phonétiques de la parole, en situation de bruit de type multilocuteur.

Nous comparons trois conditions : signal non traité, réduction de bruit seule, et réduction de bruit suivie du codage paramétrique, et cela pour différentes valeurs du rapport signal à bruit (RSB), afin de connaître les limites de son domaine d'application. On choisit quatre valeurs de RSB allant de 0 à 7,5 dB par pas de 2,5 dB.

Le test voyelle-consonne-voyelle (VCV) consiste à faire identifier au sujet une consonne présentée entre deux voyelles, parmi une liste fermée de consonnes (ou phonèmes consonantiques) possibles. Nous nous limitons à 16 consonnes, encadrées par la voyelle « a », soit sur les stimuli « aba, acha, etc. ». Ce test est réalisé en monaural, pour les

quatre valeurs de RSB, et 3 conditions de traitement, soit au total 12 conditions de test.

Le niveau d'écoute est préalablement ajusté au niveau de confort maximal durant une phase préliminaire d'habituation. Le test dure de 10 à 20 minutes suivant les sujets.

3.2 Qualité perceptive

Nous évaluons par ailleurs la qualité perceptive de phrases présentées alternativement dans les trois conditions : signal non traité, réduction de bruit seule, et réduction de bruit suivie du codage paramétrique.

On emploie des phrases de construction simple, enregistrées par un locuteur de sexe masculin.

Les conditions de bruit testées, pour être le plus possible représentatives des conditions difficiles d'utilisation de l'aide auditive, consistent en un bruit stationnaire à spectre de parole, et un bruit de parole multilocuteur.

Différentes valeurs de rapport signal-à-bruit sont choisies de sorte que les données permettent d'estimer le domaine d'efficacité optimale de la réduction de bruit : de -2.5 dB à 7.5 dB par pas de 2.5 dB.

Le test consiste à faire écouter, pour chaque type de bruit, une même phrase présentée, dans les trois conditions, présentées de manière aléatoire. Le sujet doit indiquer quel est l'item le plus agréable à entendre. Chaque sujet évalue dix phrases.

4. Résultats des tests

4.1 Capacité auditive

4.1.1 Pour les 10 normo-entendants

Sur le total des réponses justes, toutes consonnes confondues, on constate (voir Figure 3) que la CA des normo-entendants semble améliorée dans certains cas par application de la réduction de bruit. Cette amélioration n'est significative que pour une valeur de RSB, soit 2,5 dB. En revanche, par rapport à la situation où seule la réduction de bruit est appliquée, la superposition du codage paramétrique détériore les performances des sujets, d'une manière significative dans 3 cas sur 4 (pour les RSB compris entre 0 et +5 dB).

4.1.2 Pour les 25 malentendants

Sur le total des réponses justes, toutes consonnes confondues, on observe que la CA des malentendants semble dans 3 cas sur 4 détériorée par application de la réduction de bruit. Cependant ce résultat n'est jamais significatif. En revanche, de même que pour les normo-entendants, le codage paramétrique détériore la capacité de discrimination des consonnes par rapport aux deux autres situations, d'une manière significative dans 3 cas sur 4 (pour les RSB compris entre 0 et +5 dB).

Les données collectées ne permettent pas d'extraire des résultats statistiquement significatifs sur des catégories sonores plus précises (par phonème ou par trait acoustique) non plus que sur des sujets individuels.

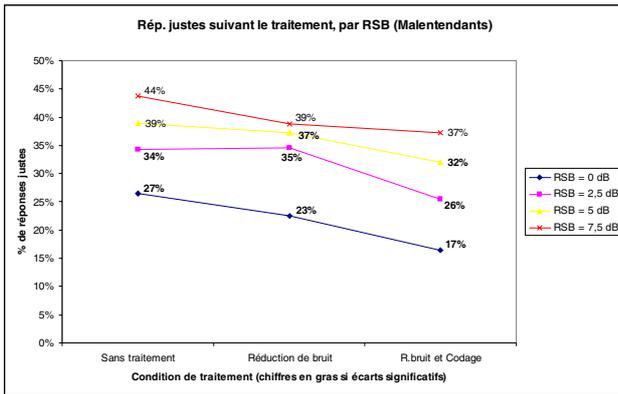
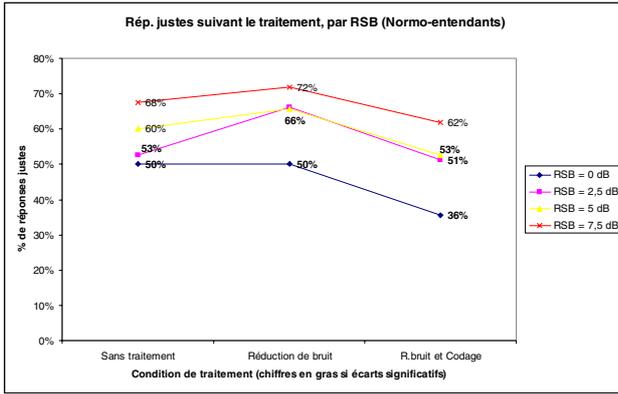


FIG. 3 : Pourcentages d'identification suivant le niveau relatif de bruit et la condition. Pour des normo-entendants (en haut), la réduction de bruit peut améliorer la CA, mais le codage annule et souvent inverse cet effet. Pour des malentendants (en bas), la réduction de bruit n'améliore pas la CA, et le codage la détériore dans presque tous les cas

4.2 Qualité perceptive

Les résultats globaux, toutes conditions de bruit confondues, portant sur 21 sujets sont les suivants (voir Figure 4) :

- Préférence pour phrases bruitées : 43 % des cas ;
- Préférence pour phrases débruitées : 38 % des cas ;
- Préférence pour phrases débruitées + codées : 19 % des cas.

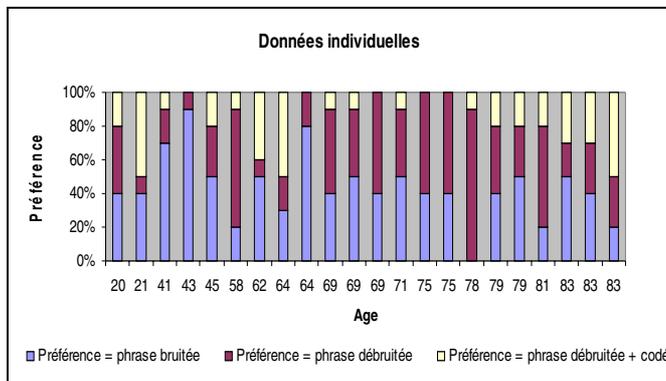


FIG. 4 : Données individuelles du test de confort

On note une forte variabilité inter-individuelle. La condition 'phrases débruitées et codées' est moins souvent préférée que les autres. Cependant si l'on cherche à analyser l'effet du débruitage, en regroupant les données des deux conditions débruitées on aurait 43 % vs 57 %. La comparaison des conditions 'bruitées' et 'débruitées sans codage' dans une expérience ultérieure doit permettre de préciser ce point.

En ce qui concerne l'effet du niveau du bruit il semblerait exister une tendance à l'accroissement du bénéfice de la réduction de bruit avec l'élévation du niveau relatif du bruit. Ceci devrait être confirmé ou infirmé par l'augmentation du nombre de sujets.

5. Conclusion

D'après ces résultats, et dans les conditions que nous avons testées, il apparaît donc que :

1. la réduction de bruit seule peut améliorer la CA des normo-entendants (de + 2,4 dB en moyenne), mais semble détériorer la CA des malentendants (de - 1,1 dB en moyenne). Les résultats suggèrent par ailleurs qu'elle peut améliorer la qualité perçue. Cette hypothèse nécessite d'être confirmée par la comparaison des préférences de signaux bruités et « débruités », hors codage.
2. le codage paramétrique, précédé de la réduction de bruit, détériore toujours la CA : -2,9 dB pour les normo-entendants, -3,5 dB pour les malentendants. Il détériore également la qualité perçue de la parole.

Références

- [1] M. Berouti, R. Schwartz, et J. Makhoul, *Enhancement of Speech Corrupted by Acoustic Noise*, ICASSP, pp. 208-211, 2-4 Avril 1979.
- [2] S.F. Boll, *Suppression of Acoustic Noise in Speech Using Spectral Subtraction*, IEEE Trans. on ASSP, vol. ASSP-27, n°2, pp. 113-120, Avril 1979.
- [3] Y. Ephraim, et D. Malah, *Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator*, IEEE Trans. on ASSP, vol. ASSP-32, n°6, pp. 1109-1121, Décembre 1984.
- [4] A. Akbari Azirani, R. Le Bouquin, et G. Faucon, *Speech Enhancement Using a Wiener Filtering under Signal Presence Uncertainty*, EUSIPCO, pp. 971-974, 1996.
- [5] LIM J.L., and OPPENHEIM A.V., *Enhancement and Bandwidth Compression of Noisy Speech*, Proceedings of the IEEE, vol. 67, n°12, pp. 1586-1604, Décembre 1979.
- [6] G. Guilmin, R. Le Bouquin et P. Gournay, *Study of the Influence of Noise Pre-Processing on the Performance of a Low Bit Rate Parametric Speech Coder*, EUROSPEECH, Budapest, Hongrie, pp. 2367-2370, Septembre 1999.
- [7] C. Laflamme, R. Salami, R. Matmti, et J-P. Adoul, *Harmonic-Stochastic Excitation (HSX) Speech Coding below 4kbts/s*, ICASSP, pp. 204-207, 1996.