

Influence sur les indices de localisation du beamforming et de la compression de dynamique dans les audioprothèses

Adrien LLAVE, Renaud SÉGUIER

CentraleSupélec, CNRS, IETR (Institut d'Électronique et de Télécommunications de Rennes)
UMR 6164, F-35000 Rennes, France

adrien.llave@centralesupelec.fr, renaud.seguier@centralesupelec.fr

Résumé – La Formation de Voie (FV) et la Compression de Dynamique (CD) sont utilisées dans les audioprothèses respectivement pour débruiter le signal d'entrée et compenser les pertes auditives de l'auditeur. Elles sont connues pour distordre les indices de localisation (IL) menant à une baisse des performances de localisation et de réception de la parole dans un environnement complexe. Depuis une décennie, des algorithmes de FV préservant les IL du bruit sont développés afin de résoudre ce problème. Cependant, la meilleure manière de les combiner avec le CD reste inconnue. Dans cette étude, nous nous concentrons sur le filtre de Wiener multicanal avec estimation partielle du bruit (MWF-N) et la meilleure façon de le combiner avec la CD. Nous montrons qu'il n'est pas trivial de les combiner sans risquer de distordre les IL du bruit ou de rendre le niveau d'écoute dangereux pour l'auditeur. Enfin, nous proposons une conception alternative incluant un nullformer permettant d'éviter la comodulation due à la CD.

Abstract – Beamforming and Dynamic Range Compression (DRC) are used in the hearing aids respectively to clean the input signal and to compensate the listener hearing loss. Both processings are known to distort the localizations cues such as the localization and speech reception performances decrease in complex environment. For a decade, noise localization cues perservation beamformers are developed in order to resolve this issue. However, the best way to combine them with the DRC stage remains unclear. In this study, we focus on the multichannel Wiener filter with partial noise estimation (MWF-N) and the best way to combine it with a DRC. We show that combining the MWF-N and the DRC is not straightforward without risking to distort the noise localization cues or making the listening level hazardous for the listener. Finally, we propose an alternative design including a nullformer avoiding the comodulation issue due to the DRC.

1 Introduction

1.1 Contexte

Depuis une décennie, des algorithmes de Formation de Voie (FV) [1, 2] et de Compression de Dynamique (CD) [3, 4, 5, 6] sont développés pour préserver les Indices de Localisation (IL) dans les audioprothèses afin de bénéficier de l'effet *cocktail-party* [7]. Cependant, l'interaction entre les deux traitements et la manière de les combiner a été peu étudiée alors que la CD est connu pour détériorer aussi les IL.

Le but de ce travail est, tout d'abord, de tester la proposition suivante : la façon usuelle d'associer FV et CD n'est pas pertinente lorsque la FV préserve les IL du bruit. Ensuite, nous proposons des architectures alternatives pour combiner le FV et le CD. Elles sont enfin comparées grâce à des métriques objectives à l'architecture de référence.

1.2 Modèle des données

A chaque microphone, le son est modélisé comme la somme des sources convoluées par la réponse du canal acoustique correspondant. Il est commode de le penser dans le domaine de la Transformée de Fourier à Court Terme (TFCT) où la convolution devient un produit simple. Alors, pour chaque point du

plan temps-fréquence, le modèle de données peut être écrit sous la forme : $\mathbf{x} = \mathbf{h}_0 s + \mathbf{v}$ où $\mathbf{x} \in \mathbb{C}^M$ désigne les signaux des M microphones composés de la cible s filtrée par les fonctions de transfert \mathbf{h}_0 associées à sa direction d'arrivée. $\mathbf{v} \in \mathbb{C}^M$ correspond au mélange des autres sources, bruit diffus et bruit électronique.

1.3 La formation de voie préservant les indices de localisation (MWF-N)

On choisit comme FV le filtre de Wiener multicanal avec estimation partielle du bruit (MWF-N) car son efficacité est bien documentée [8] et son architecture permet de le coupler de plusieurs manières avec un CD. Dans le domaine TFCT, à une fréquence et une trame donnée, la sortie du MWF-N y peut s'exprimer comme :

$$y = \mathbf{w}^H \mathbf{x}. \quad (1)$$

Il a été montré dans [1] que la solution du MWF-N, $\mathbf{w}_{\text{MWF-N,L}}$, peut s'écrire :

$$\mathbf{w}_{\text{MWF-N,L}} = (1 - \eta) \mathbf{w}_{\text{MWF,L}} + \eta \mathbf{e}_L \quad (2)$$

où \mathbf{e}_L est un vecteur nul de taille M sauf l'élément correspondant au microphone de référence de l'oreille gauche qui est à

un. $w_{\text{MWF,L}}$ est la solution du filtre de Wiener multicanal classique (MWF) pour la sortie gauche et η est le paramètre de compromis entre la sortie du MWF et le microphone de référence gauche. Les variables indicées par L *e.g.* e_L , sont définies pour la sortie gauche, leur homonymes indicés par R sont définis similairement mais pour l'oreille droite. Sans perte de généralité, seules les versions de l'oreille gauche des expressions seront écrites par la suite.

1.4 Le compresseur de dynamique (CD)

La CD atténue les sons d'un niveau supérieur au seuil L_T . Le gain de réduction G alors appliqué est défini en dB comme suit :

$$G = (S - L_p) \left(\frac{1}{R} - 1 \right) - K \quad (3)$$

où L_p est le seuil de douleur de l'auditeur, K le gain de compensation de la perte auditive et le ratio R dépend de K et L_T . En dessous de L_T , l'amplification est linéaire. Dans cette étude, L_T , exprimé en dB_{SPL} , est défini comme $L_T = \alpha(L_p - K)$ où α est la proportion de la plage dynamique pour laquelle la CD est linéaire. En pratique, G est filtré passe-bas en fonction des constantes de temps d'attaque et de relâchement. Comme la plage entre le seuil d'audition et le seuil de douleur dépend de la fréquence, un CD est appliqué à chaque bande de fréquence. Nous renvoyons à [9] pour une revue détaillée des techniques de CD.

2 Methodes

2.1 Combinaison du MWF-N et de la CD

L'architecture de référence La condition de référence consiste à appliquer la CD après le MWF-N. La sortie pour l'oreille gauche y_L est alors égale à

$$y_L = k\sigma((1 - \eta)\check{s}_L + \eta x_L) \quad (4)$$

où k est le gain de compensation de la perte auditive, $\sigma(\cdot)$ représente la CD, \check{s}_L est l'estimation de la cible à l'oreille gauche *i.e.* la sortie du MWF et x_L le signal du microphone gauche *i.e.* le plus près de l'entrée du conduit auditif gauche.

L'architecture à branche latérale pre- η limitée Pour dissocier la compression sur la cible et le bruit, une approche naïve consiste à appliquer la CD juste après la sortie du MWF et à laisser la branche latérale sans compression mais avec le gain k afin de ne pas affecter les IL du bruit qu'elle apporte. La sortie pour l'oreille gauche, y_L , devient alors :

$$y_L = (1 - \eta)k\sigma(\check{s}_L) + \eta k x_L \quad (5)$$

$$= k((1 - \eta)\sigma(\check{s}_L) + \eta s_L + \eta v_L). \quad (6)$$

On peut noter que les termes $(1 - \eta)\sigma(\check{s}_L) + \eta s_L$ correspondent à un *compresseur parallèle*, une architecture bien connue en production musicale. Un inconvénient critique de cette conception

est l'amplification linéaire de la branche latérale où le niveau peut dépasser L_p si $k\eta > 1$. Par conséquent, un limiteur $\lambda(\cdot)$ est ajouté à la branche latérale afin d'amplifier linéairement sur une plage la plus grande possible avant la compression. Un limiteur est un cas particulier du CD où $R \rightarrow \infty$ permettant de ne pas dépasser L_p . La sortie gauche peut être écrite comme suit : $y_L = (1 - \eta)k\sigma(\check{s}_L) + \eta\lambda(kx_L)$.

L'architecture à branche parallèle post- η limitée Comme $\eta < 1$ [8], il pourrait être plus pertinent de placer le limiteur après l'atténuation par η . En effet, si le limiteur est appliqué sur une version atténuée du signal, il sera moins souvent activé. La sortie gauche s'écrit alors : $y_L = (1 - \eta)k\sigma(\check{s}_L) + \lambda(\eta k x_L)$.

L'architecture à branche parallèle post- η limitée + nullformer Comme la cible est présente à la fois dans les deux branches, le problème de comodulation [10] dû au limiteur dans la branche latérale ne peut être résolu. Nous proposons d'ajouter un estimateur de bruit \tilde{v}_L (ou nullformer) avant le limiteur afin de réduire cet effet. Il est modélisé comme suit : $\tilde{v}_L = v_L + \epsilon s_L$ où v_L est le bruit de référence dans x_L et ϵ le gain résiduel complexe de la cible. Ainsi, la sortie de l'oreille gauche s'écrit : $y_L = (1 - \eta)k\sigma(\check{s}_L) + \lambda(\eta k \tilde{v}_L)$. Dans le cas où $\eta k \tilde{v}_L < L_p$ *i.e.* le limiteur n'est pas activé, y_L devient :

$$y_L = k((1 - \eta)\sigma(\check{s}_L) + \eta \epsilon s_L + \eta v_L). \quad (7)$$

Alors, l'expression de y_L entre Eq. 6 et Eq. 7 diffère seulement du facteur ϵ . Quand $|\epsilon| \rightarrow 0$ *i.e.* le nullformer est efficace, le compresseur parallèle devient un simple compresseur.

2.2 Évaluation objective

Trois critères sont utilisés pour évaluer les architectures présentées précédemment : \hat{L}_s , le niveau maximum de la cible pour lequel le niveau de sortie ne dépasse pas L_p ; si $\hat{L}_s > L_p$, cela devient dangereux pour l'auditeur; $L_{T\lambda}$, le niveau de la cible dans la branche latérale nécessaire pour activer le limiteur, permet d'évaluer la préservation des IL du bruit. En effet, c'est le comportement non linéaire qui distord les IL. Enfin, le Gain du Réseau (GR), le rapport entre le rapport signal à bruit (RSB) en sortie et le RSB à l'entrée.

Dans cet exemple, le seuil d'audition normal est de 0 dB_{SPL} et le seuil d'audition de l'auditeur est de 30 dB_{SPL} , donc un gain k de la même valeur est appliqué. Le L_p est fixé à 120 dB_{SPL} et le paramètre du CD $\alpha = \frac{2}{3}$. Par souci de simplicité, les constantes de temps sont négligées afin d'afficher des courbes d'entrée/sortie simples. Le paramètre $\eta = 0.2$ [8] permet de localiser une source hors axe. Le gain résiduel de la source cible $\epsilon = \frac{1}{k}$ afin d'assurer un $\hat{L}_s = L_p$ autant que possible. Une variable en minuscule suit une échelle linéaire et la même lettre en majuscule renvoie à la variable exprimée en dB *e.g.* $\eta = 0.2$ et $H \approx -14$ dB.

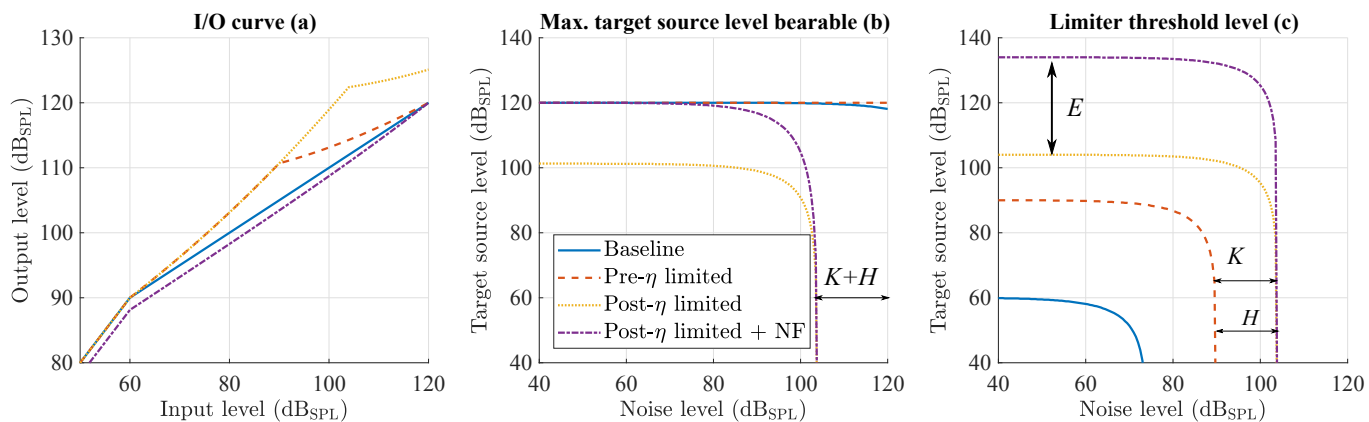


FIGURE 1 – Courbes d’entrée/sortie (a) pour les quatre architectures, \hat{L}_s est le niveau d’entrée maximal supportable (b) et le niveau d’activation du limiteur (c) en fonction du niveau de bruit.

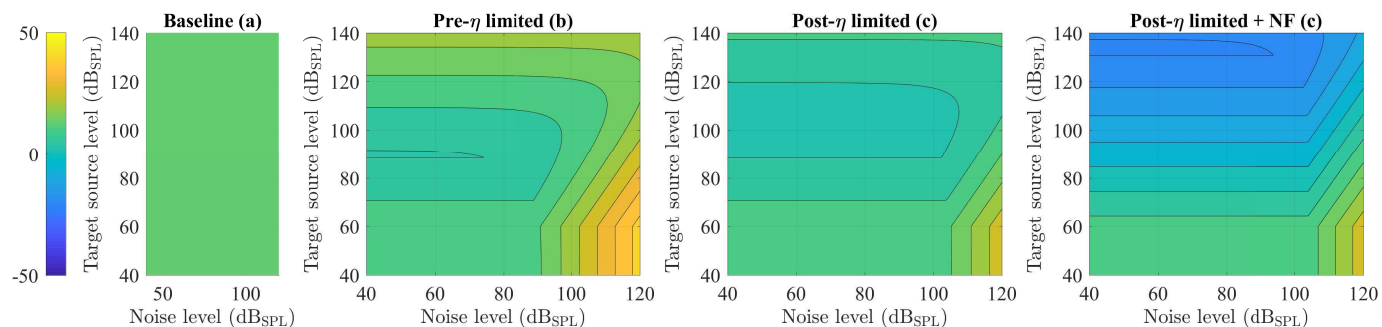


FIGURE 2 – Gain du réseau (dB) pour les architectures suivantes : référence (a), pre- η limitée (b), post- η limitée (c) et post- η limitée + nullformer (d). Les courbes de niveau sont espacées de 5 dB.

3 Resultats

Sur la Fig. 1(a) sont représentées les courbes d’entrée/sortie du système complet pour un RSB infini *i.e.* l’abscisse correspond au niveau de la cible seule. Sur la Fig. 1(b, c) sont représentés \hat{L}_s et $L_{T\lambda}$ en fonction du niveau de bruit. La surface sous les courbes peut être vue comme l’ensemble des combinaisons valides de niveaux de la cible et de bruit au sens de la réalisation des objectifs : maintenir le niveau de sortie au-dessous de L_p et préserver les IL du bruit dans la branche latérale. Enfin, est représenté en Fig. 2 le GR en dB pour chaque combinaison de niveaux de la cible et du bruit.

3.1 L’architecture de référence

Dans cette condition, le \hat{L}_s correspond au L_p sur la plage de niveau considéré pour la cible et le bruit par construction. On peut noter le début d’une chute de \hat{L}_s pour un niveau sonore supérieur à 110 dB_{SPL}. En effet, il y a une baisse similaire à celle observée pour les architectures *post- η limitée* et *post- η limitée + nullformer* à un niveau de $L_p - H$ (134 dB_{SPL} ici), qui correspond au niveau de bruit nécessaire pour passer le L_p en sortie sans la présence de la cible. Le principal inconvénient attendu de la condition de référence est l’activation du com-

presseur à faible niveau. En effet, il s’active pour un niveau de la cible supérieur à $\alpha(L_p - K)$ (60 dB_{SPL} ici), pour un RSB élevé. De plus, le bruit seul est capable de l’activer lorsque son niveau est supérieur à $L_T - H$ (≈ 74 dB_{SPL} ici). Il faut faire attention avec le paramètre α car plus il est proche de 1, plus l’aire sous la courbe est grande mais aussi plus le ratio R est grand, ce qui conduit à des distorsions plus importantes pour les sons forts. Enfin, le GR de la condition de *référence* est constant et égal à η pour un niveau du bruit à la sortie dominé par la contribution de la branche latérale.

3.2 L’architecture à branche latérale pre- η limitée

Le limiteur placé avant l’atténuation η permet de toujours maintenir le niveau de sortie inférieur à L_p tel que le \hat{L}_s est optimal et égal à L_p (voir Fig. 1(b)). C’est l’architecture la plus sûre possible pour le système auditif de l’auditeur. Cependant, un inconvénient majeur est l’atténuation η dans la branche latérale correspondant à une baisse apparente du niveau d’activation du limiteur de 14 dB ($\eta = 0.2$). Elle conduit à un préservation des IL du bruit sur une plage de niveau réduite. Néanmoins, elle est beaucoup plus grande que celle obtenue avec la condition de *référence*. Enfin, le GR est plus élevé pour un

faible RSB d'entrée et un fort niveau de bruit *i.e.* lorsque le limiteur est activé et que le CD ne l'est pas. Cet augmentation du GR pourrait compenser la distorsion des IL du bruit en terme de compréhension de la parole (cible). C'est un moyen de contourner l'inconvénient présenté dans [11, 10].

3.3 L'architecture à branche latérale post- η limitée

Le limiteur placé après l'atténuation de η , son seuil d'activation est décalé d'autant par rapport à la condition précédente. Ainsi, cette architecture permet de contourner en partie le problème susmentionné. Dans cet exemple, le bruit est amplifié linéairement pour un niveau de la cible et du bruit allant jusqu'à 105 dB_{SPL}. Cette plage permet de couvrir la majorité des scenarii. Cependant, cette architecture a un inconvénient majeur qui est l'abaissement du \hat{L}_s (environ 20 dB ici), lorsque le bruit est inférieur à celle de la cible. On peut également l'observer sur la courbe E/S sur la Fig. 1(a). De plus, si $k\eta > 1$ le bruit seul peut être suffisamment fort pour atteindre le L_p prématurément et conduire à la chute observée sur la Fig. 1(b). Enfin, le GR bénéficie ici du même type d'effet que pour le précédent dans une moindre mesure.

3.4 L'architecture à branche latérale post- η limitée + nullformer

D'abord, on peut noter sur la Fig. 1(a) un décalage entre la courbe de cette condition et les autres due à la inadéquation des termes $1 - \eta$ et $\eta\epsilon$ dans Eq. 7 conduisant à une atténuation de la cible (≈ 2 dB ici). Cependant, il est possible de montrer que $\hat{L}_s = L_p$ sous les hypothèses suivantes : $\epsilon = \frac{1}{k}$ et $v_L \ll s_L$. Par conséquent, \hat{L}_s correspond à la performance optimale atteinte par la conception *pre- η limitée* pour un niveau de bruit allant jusqu'à 85 dB_{SPL} dans l'exemple. Pour un RSB plus faible, \hat{L}_s chute et rejoint la courbe de l'architecture *post- η limitée*. L'amélioration majeure de cette architecture est l'augmentation de $L_{T\lambda}$ pour un niveau inférieur à $L_p - K + H$ (105 dB_{SPL} ici). En effet, pour un grand RSB, l'effet de l'atténuation ϵ sur $L_{T\lambda}$ devient sensible. Il faut noter que l'écart entre la courbe de cette condition et la courbe de celle *post- η limitée* est égal à ϵ . Dans notre exemple, il permet de préserver les IL du bruit sur une plage de niveaux extrêmement large. Au-delà de ce niveau, $L_{T\lambda}$ chute et atteint la même performance que le design *post- η limitée* *i.e.* le bruit seul active le limiteur.

De plus, un compresseur fréquentiel (CF) [12] peut venir compléter le CD ce qui mènerait à de fortes distorsions de la cible lors du mélange entre la branche principale et latérale. En atténuant la cible dans celle-ci, le nullformer pourrait les éviter. Néanmoins, l'étude de cet aspect est au-delà du cadre de ce travail.

4 Conclusion

Nous montrons que la combinaison naïve d'un CD et d'un MWF-N dans le cadre des audioprothèses pourrait conduire à des distorsions des IL du bruit, perdant ainsi le bénéfice apporté par ce dernier. Pour contourner ce problème, nous avons introduit une architecture compressant indépendamment la cible et le bruit en gardant ce dernier non compressé sur la plus grande plage possible. Ceci évite ainsi le problème de comodulation [10] tout en conservant les IL du bruit sur une grande dynamique. Cette étude se concentre sur l'intérêt théorique d'une telle architecture et ses conclusions doivent être confirmées par des tests perceptifs. Un travail futur doit être consacré à trouver une estimation du bruit de référence modélisé ici. Aussi, ce travail pourrait être étendu à la combinaison de la FV et du CF.

Références

- [1] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical Analysis of Binaural Multimicrophone Noise Reduction Techniques," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 342–355, Feb. 2010.
- [2] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed Binaural LCMV Beamforming," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, pp. 137–152, Jan. 2017.
- [3] I. M. Wiggins and B. U. Seeber, "Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise," *The Journal of the Acoustical Society of America*, vol. 133, no. 2, pp. 1004–1016, 2013.
- [4] A. H. Schwartz and B. G. Shinn-Cunningham, "Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2329–2339, 2013.
- [5] H. G. Hassager, A. Wiinberg, and T. Dau, "Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment," *The Journal of the Acoustical Society of America*, vol. 141, pp. 2556–2568, Apr. 2017.
- [6] S. M. A. Ernst, S. Kortlang, G. Grimm, T. Bisitz, B. Kollmeier, and S. D. Ewert, "Binaural model-based dynamic-range compression," *International Journal of Audiology*, pp. 1–12, Jan. 2018.
- [7] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, vol. 25, pp. 975–979, Sept. 1953.
- [8] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," *The Journal of the Acoustical Society of America*, vol. 124, pp. 484–497, July 2008.
- [9] J. M. Kates, "Principles of Digital Dynamic-Range Compression," *Trends in Amplification*, vol. 9, pp. 45–76, Mar. 2005.
- [10] M. A. Stone and B. C. J. Moore, "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," *The Journal of the Acoustical Society of America*, vol. 116, pp. 2311–2323, Oct. 2004.
- [11] P. E. Souza, L. M. Jenstad, and K. T. Boike, "Measuring the acoustic effects of compression amplification on speech in noise," *The Journal of the Acoustical Society of America*, vol. 119, pp. 41–44, Jan. 2006.
- [12] D. Glista, S. Scollie, M. Bagatto, R. Seewald, V. Parsa, and A. Johnson, "Evaluation of nonlinear frequency compression : Clinical outcomes," *International Journal of Audiology*, vol. 48, pp. 632–644, Jan. 2009.