



**HAL**  
open science

## Comparaison entre PESQ et MUSHRA pour l'estimation de la qualité audio d'un signal de parole enregistré par des microphones “ non-conventionnels ”

Véronique Zimpfer, Thomas Joubaud, Julien Richard, Pascal Hamery

### ► To cite this version:

Véronique Zimpfer, Thomas Joubaud, Julien Richard, Pascal Hamery. Comparaison entre PESQ et MUSHRA pour l'estimation de la qualité audio d'un signal de parole enregistré par des microphones “ non-conventionnels ”. 16ème Congrès Français d'Acoustique, (CFA) 2022, Société Française d'Acoustique; Laboratoire de Mécanique et d'Acoustique, Apr 2022, Marseille, France. hal-03848426

**HAL Id: hal-03848426**

**<https://hal.science/hal-03848426v1>**

Submitted on 10 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



16<sup>ème</sup> Congrès Français d'Acoustique  
11-15 Avril 2022, Marseille

## Comparaison entre PESQ et MUSHRA pour l'estimation de la qualité audio d'un signal de parole enregistré par des microphones «non-conventionnels »

V. Zimpfer <sup>a</sup>, T. Joubaud <sup>a</sup>, J. Richard <sup>a,b</sup>, et P. Hamery <sup>a</sup>

<sup>a</sup> ISL, Institut de Recherche de St Louis, 5 rue du Général Cassagnou, BP 70034, 68301 St Louis,  
France,

<sup>b</sup> Laboratoire Interdisciplinaire de Carnot-Bourgogne, département Conception, Optimisation et  
Modélisation Mécanique. F-90400 SEVENANS.



Lorsque l'on parle, on génère non seulement du signal aérien, mais aussi des vibrations qui se propagent dans les différents éléments constituant notre tête. Ces vibrations contiennent l'information du signal de parole et peuvent être enregistrées. On appelle microphones « non conventionnels » tous les capteurs permettant d'enregistrer ces signaux se propageant dans notre corps. Les microphones non conventionnels sont souvent moins sensibles au bruit ambiant que les microphones acoustiques classiques, ils peuvent être utilisés dans des environnements difficiles et ils sont compatibles avec les équipements de tête tels que les masques de protection NRBC. Le plus ancien microphone non conventionnel est le laryngophone mais il en existe d'autres, comme la prise de son aérienne dans le conduit auditif ou les microphones à conduction osseuse. Des études antérieures ont pu montrer que l'intelligibilité résultant de ces types de prise de son était dégradée. Le but de cette étude est alors de vérifier si une méthode objective telle que PESQ peut être utilisée pour estimer la qualité audio de ces microphones. Une comparaison entre les résultats obtenus par la méthode objective PESQ et celle subjective du test MUSHRA a été réalisée pour estimer la qualité audio de signaux de paroles provenant de quatre microphones « non conventionnels » de types différents.

## 1 Introduction & Contexte

Depuis une vingtaine d'années, on voit apparaître sur le marché civil et militaire, de nouveaux systèmes de communication par voie « osseuse ». Pour la réception, le principal avantage est d'offrir à l'utilisateur la possibilité d'entendre une conversation tout en ayant les oreilles dégagées afin de rester à l'affût de signaux d'alerte potentiels. Il est tentant d'utiliser les mêmes chemins pour recueillir la parole émise par l'utilisateur. En effet, lorsque l'on parle, on génère non seulement du signal aérien sortant de notre bouche, mais aussi des vibrations qui se propagent dans les différents éléments constituant notre tête. Les vibrations produites par l'ensemble « larynx – cavités pharyngo-buccale et nasale » se propagent dans le corps et peuvent être recueillies (très atténuées et déformées) en différents points du corps. L'utilisation faite des laryngophones pour le recueil de la parole en environnement très bruyant (véhicules blindés, avions...) est bien connue. Le Laryngophone [1] est sûrement l'ancêtre des microphones dit à « conduction osseuse » mais ce n'est pas le seul. On appelle microphones « non conventionnels » (NC) tous les capteurs permettant d'enregistrer les signaux provenant d'une propagation dans le corps. Un autre avantage de ces microphones NC est la compatibilité avec des équipements de têtes comme les masques à gaz. Des systèmes couplant protection auditive et fonction de communication, en utilisant les vibrations du signal de parole qui arrivent au canal auditif, sont en cours de développement avec la société Cotral [2].

Une étude du signal de parole provenant d'un laryngophone révèle que certaines des fréquences les plus élevées sont manquantes. Par conséquent, on peut observer une absence de fricatives dont la majorité de l'énergie est concentrée en hautes fréquences [3], d'où une diminution de l'intelligibilité et une dégradation de la qualité sonore.

Une étude faite à l'ISL [4,5] avait montré une ambiguïté sur la reconnaissance des voyelles orales lorsqu'elles étaient enregistrées par un microphone à conduction osseuse. Dans cette étude, le microphone à conduction osseuse était placé en haut de la mâchoire sous l'os temporal du côté droit. Des études [6,7] ont montré que la qualité sonore et l'intelligibilité dépend de la position du capteur sur la tête. Cette dernière montre également qu'avec ce type de

microphone, une dégradation de la qualité du signal de parole est observée.

Notre objectif principal est de trouver des solutions pour améliorer l'intelligibilité et la qualité du signal audio capté par les microphones NC. Dans un premier temps, on s'intéresse à la mesure et à la compréhension des dégradations que subit le signal de parole avant d'atteindre le microphone NC.

Il existe différents types de microphones NC qui enregistrent indirectement le signal de parole après une propagation au travers de la tête. Est-ce que tous ces microphones dégradent de la même manière le signal de parole, ou en existe-t-il un type plus performant que les autres ? Le travail présenté dans ce papier est une étude préliminaire sur l'obtention d'une métrique pour mesurer la qualité audio des signaux de paroles captés par un microphone NC afin de pouvoir répondre aux questions posées précédemment. De plus, on souhaite vérifier si on peut classer en termes de qualité audio les différents types de microphone NC en utilisant une méthode perceptive comme le MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) et une méthode objective tel que PESQ (Perceptual Evaluation of Speech Quality).

## 2 Matériels et Méthodes

Dans cette partie sont décrits les quatre types de microphone NC utilisés pour cette étude préliminaire. Elle présente également les deux méthodes permettant l'évaluation de la qualité sonore.

### 2.1 Microphones NC

Quatre types différents de microphones NC ont été testés dont les photos sont représentées en Figure 1 :

Le premier type récupère, comme les microphones à conduction osseuse, les vibrations sur la tête à l'aide d'un accéléromètre [4,5]. Le modèle utilisé pour l'étude est un prototype fabriqué à l'ISL et constitué d'un accéléromètre et d'une électronique dédiée. Ce capteur sera placé sur le front à l'aide d'un bandeau. En effet, Tran et McBride ont montré

que c'est à cet emplacement que la qualité audio et l'intelligibilité sont les meilleures pour ce type de capteur [6,7]. Dans la suite de l'étude ce microphone est noté CM1.

Le deuxième type est un système couplé à une protection auditive dans lequel un capteur va enregistrer les vibrations sur la paroi du canal auditif. Pour cette étude, nous avons utilisé le M3h/X5, provenant de la technologie de Invisio [8]. Ce microphone réalisé par Invisio est considéré comme une référence dans la prise de son dans l'oreille. Dans la suite de l'étude ce microphone est noté CM2.

Le troisième type est donc la récupération des vibrations du signal de parole au niveau du larynx qui est la source primaire du signal de parole. Pour ce type de microphone NC, un laryngophone provenant du commerce a été utilisé. Celui-ci est commercialisé sous la marque RETEVIS. Dans la suite de l'étude ce microphone est noté CM3.

Le dernier type permet l'enregistrement de l'onde aérienne reconstruite dans le canal auditif quand celui-ci est obstrué à l'aide d'un mini-microphone. Cette technologie est utilisée dans la protection auditive BANG (Bouchon Auriculaire Nouvelle Génération) [2]. Pour cette étude, une version simplifiée a été utilisée. Celle-ci est constituée d'un microphone à électret Knowles (série FG) inséré dans une pièce permettant d'intégrer un bouchon à trois ailettes. Dans la suite de l'étude, ce microphone est noté AM1.



Figure 1 – Les différents microphones NC

## 2.2 Enregistrement des signaux audio

12 locuteurs volontaires (9 hommes et 3 femmes) ont participé à ces enregistrements. Chaque locuteur a été équipé des 4 microphones NC testés comme le montre la Figure 2 : à savoir le CM1 sur le front, le CM2 dans l'oreille droite, CM3 sur le larynx et AM1 dans l'oreille gauche.

Le locuteur était assis dans une cabine d'audiométrie à double paroi avec un bruit de fond inférieur à 35 dB(A). Un microphone de référence (noté Ref sur la Figure 2) de type B&K ½ pouce était placé à 50 cm environ de la bouche du

locuteur. Ce signal de référence est nécessaire pour les différents tests envisagés.

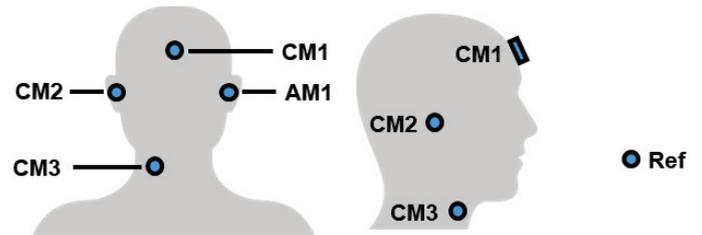


Figure 2 – Position des différents microphones NC

Chaque locuteur a prononcé deux phrases de la liste 6 de Combescure [9]:

1. « Ce chasseur projette encore de partir d'ici, ce matin »
2. « Là-haut, monte la voix du pâtre qui ramène ses moutons »

Ces deux phrases ont été enregistrées simultanément sur les 4 microphones NC et le microphone de référence à l'aide de la carte son ZOOM F8n avec une fréquence d'échantillonnage de 48kHz.

Pour le microphone AM1, n'ayant aucun traitement par rapport aux 3 autres microphones NC, le signal a été filtré par un passe-haut du second ordre, de fréquence de coupure 1,9 kHz. Ce filtre permet d'atténuer l'effet d'occlusion qui amplifie fortement les basses fréquences. Le gabarit du filtre a été déterminé à l'aide des valeurs de l'effet d'occlusion lié au conduit auditif par voie osseuse, selon Dirks et Swindeman [10].

## 2.3 PESQ

Pour la méthode objective, nous avons utilisé la méthode PESQ ; abréviation de « Perceptual Evaluation of Speech Quality » (évaluation de la qualité vocale perçue). C'est un outil permettant l'évaluation de la qualité vocale perçue transmise par un système de télécommunication. Il a été recommandé par l'Union internationale des télécommunications (UIT) sous le nom P.862 [11]. La méthode PESQ est donc un algorithme analysant le signal vocal, échantillon par échantillon, après un alignement temporel des extraits correspondants du signal de référence et du signal de test. Les résultats de PESQ peuvent modéliser principalement des notes moyennes d'opinions (MOS, Mean Opinion Score) qui couvrent une échelle de 1 (mauvais) à 5 (excellent). Une fonction de mise en correspondance avec MOS-LQO est décrite dans la Recommandation UIT-T P.862.1 [12] et était utilisée dans notre étude.

Pour estimer les scores de PESQ, nous avons utilisé une routine Matlab utilisant les codes sources en C développés par l'UIT. Avant l'utilisation de cette routine, un ré-échantillonnage des signaux à 16 kHz (fréquence d'échantillonnage demandé pour l'algorithme de la PESQ) a

été réalisé. La méthode PESQ fournit un score compris entre 0,5 et 4,5 où 0,5 correspond à une mauvaise qualité et 4,5 à une qualité excellente. Le score de PESQ a été déterminé indépendamment pour les deux phrases et la moyenne des deux notes a été retenue comme valeur de qualité audio pour un locuteur donné et un microphone donné.

Pour cette évaluation, nous n'avons pas choisi d'utiliser la version de PESQ large-bande puisque les signaux provenant des microphones NC ont tendance à supprimer les hautes fréquences, ce qui résulte en une réduction systématique du score de PESQ.

## 2.4 MUSHRA

Comme méthode perceptive, le test MUSHRA, abbréviation de "MULTiple Stimuli with Hidden Reference and Anchor" (multi-stimuli avec référence et ancrages cachés) a été utilisé. Il s'agit d'une méthodologie pour réaliser un test d'écoute afin d'évaluer la qualité perçue de la sortie des systèmes de communication audio avec perte. Elle est définie par la recommandation BS.1534-3 de l'UIT-R [13] de l'union internationale des télécommunications (UIT).

Un des avantages de cette méthodologie est que toutes les configurations testées sont présentées en même temps. Pour notre cas, l'auditeur a dû comparer en même temps, par rapport à un signal de référence, les signaux des 4 microphones NC ainsi que la référence cachée et le signal ancre. Le signal de référence est le signal enregistré par le microphone B&K placé devant la bouche. Le signal ancre doit correspondre à la qualité audio perçue la plus mauvaise. Le signal ancre a donc été défini comme le signal de référence filtré par un passe bande avec l'ajout d'un bruit blanc.

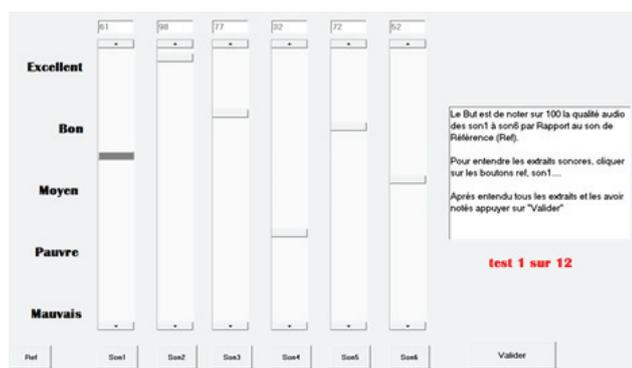


Figure 3 – interface créé pour la réalisation de l'évaluation perceptive de la qualité audio avec la méthodologie MUSHRA

Pour ce test, une interface homme/machine a été développée (voir Figure 3). Cette interface permet de noter de 0 à 100 chacun des 6 sons à l'aide d'un curseur vertical où 0 correspond à mauvais et 100 correspond à excellent. L'auditeur peut écouter autant de fois qu'il le souhaite le son de référence et les 6 sons à noter en appuyant sur les boutons. L'ordre des sons à noter est aléatoire et change à chaque itération. Après avoir noté les 6 sons, il appuie sur le bouton

« valider » pour enregistrer les notes. Chaque auditeur va devoir évaluer une phrase de chaque locuteur (12 évaluations) pris en alternance entre les deux phrases enregistrées. L'ordre des locuteurs est aussi aléatoire. La durée du test pour évaluer les 12 locuteurs est d'environ dix minutes.

Chaque auditeur a donné 72 notes sur 100 (12\*6). Pour homogénéiser les notes d'un auditeur à un autre, une homothétie a été réalisée afin que la dynamique des notes soit la même pour chaque auditeur. Les notes obtenues sont après ramenée entre 0,5 et 5 pour mieux les comparer avec le score obtenu avec la PESQ.

25 auditeurs volontaires ont participé au test d'écoute. Un auditeur a été retiré car il n'a pas reconnu la référence cachée. Pour tous les autres auditeurs, la référence cachée (meilleure note) ainsi que l'ancre (plus mauvaise note) ont été reconnues.

## 3 Résultats

### 3.1 Pour la méthode PESQ

En plus des quatre signaux provenant des quatre microphones NC, le score de PESQ a aussi été estimé pour le signal ancre utilisé pour le test MUSHRA. Donc, pour chaque microphone NC et pour le signal ancre, on a obtenu 12 scores avec PESQ, un par locuteur. La Figure 4 représente la boîte à moustaches pour chaque microphone et le signal ancre. Cette figure permet de vérifier que la qualité du signal ancre est la plus mauvaise. Elle montre, par ailleurs, que les scores moyens de PESQ pour les différents microphones NC sont très proches : le meilleur score est de 2,17 pour AM1 et le moins bon est de 2,04 pour le CM2. Des tests de Student deux à deux ont montré que les différences entre les microphones NC sont non significatives. En revanche, les différences entre chaque microphone NC et le signal ancre sont significatives.

On constate sur la Figure 4 que le signal ancre n'est pas influencé par le locuteur (faible dispersion). Ceci semble normal puisque tous les signaux ont subi les mêmes dégradations. Cependant, on constate qu'avec CM2 l'écart interquartile est le plus grand. Ainsi avec ce microphone, l'effet du locuteur est plus important. Ceci peut être expliqué par le fait que certains locuteurs ont eu des difficultés à le positionner correctement dans l'oreille. Pour ce microphone, il est impératif que le capteur touche la paroi du conduit auditif.

La Figure 5 représente le score moyen et l'écart-type en fonction du genre des locuteurs. On constate pour le signal ancre, qu'il n'y a pas d'effet du genre. En effet les signaux ont subi la même dégradation par rapport au signal de référence. En revanche, on constate que la qualité d'après la méthode PESQ est plus faible chez les femmes pour les trois microphones CM. On observe, cependant, l'opposé pour le microphone AM1.

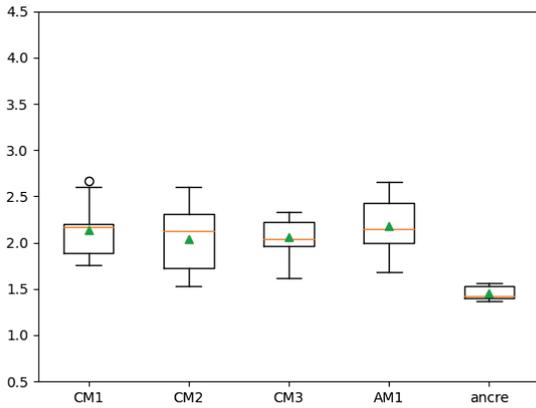


Figure 4 : Score PESQ obtenue pour chaque microphone NC et le signal ancre

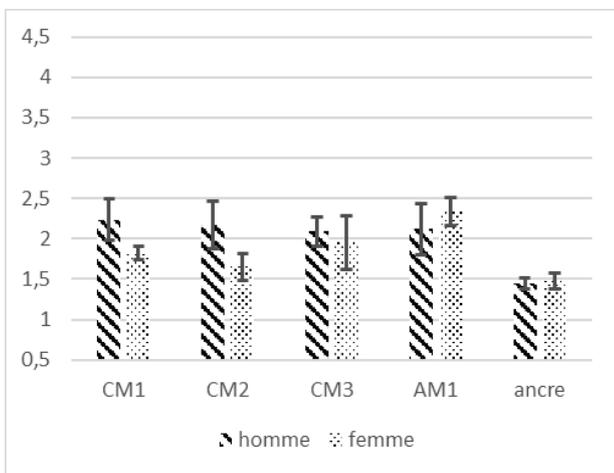


Figure 5 : Score PESQ moyen en fonction du genre du locuteur pour chaque configuration.

### 3.2 Pour le test MUSHRA

Pour exploiter les scores obtenus avec le test MUSHRA, deux possibilités ont été utilisées :

- La première permet de voir l'effet de l'auditeur. Pour se faire, une note moyenne (moyenne des 12 notes) pour chaque microphone a été estimée pour chaque auditeur.
- La deuxième permet de voir l'effet du locuteur (se rapprochant des résultats obtenus avec PESQ). Pour se faire, une note moyenne (moyenne sur les 24 auditeurs) pour chaque microphone a été estimée pour chaque locuteur.

La Figure 6 représente la dispersion des notes moyennes obtenues par microphone en fonction des 24 auditeurs. Cette figure montre une forte dispersion des notes données par les auditeurs. Cette forte dispersion peut s'expliquer par le fait que les auditeurs n'ont pas la même préférence en termes de

qualité perçue. On constate tout de même qu'un des microphones NC sort du lot, à savoir le microphone AM1.

La Figure 7 représente la dispersion des notes moyennes obtenues par microphone en fonction des 12 locuteurs. Dans ce cas les dispersions sont moins grandes, surtout pour le signal ancre et le microphone CM1 qui ont les écarts interquartiles les plus faibles. Ainsi CM2 et CM3 sont donc les deux microphones les plus sensibles aux locuteurs.

Ces deux figures ont permis de montrer que le signal ancre a toujours été reconnu comme étant le signal audio avec la plus faible qualité. On constate aussi que le microphone AM1 obtient une note moyenne meilleure que pour les trois microphones. En revanche, pour les trois autres, les notes moyennes sont proches avec pour CM2 la note moyenne la plus faible. Des tests de Student deux à deux ont montré que les différences entre le microphone AM1 et les 3 autres microphones NC sont significatives. Par contre, les différences entre les 3 microphones CM sont non significatives. Bien évidemment, les différences entre chaque microphone NC et le signal ancre sont significatives.

La Figure 8 représente la note moyenne et l'écart-type en fonction du genre des locuteurs. On constate pour le signal ancre et le microphone CM1 qu'il n'y a pas d'effet du genre. Par contre, on remarque que la qualité d'après les auditeurs est plus faible chez les femmes pour les 2 autres microphones CM et est plus élevée chez les femmes pour le microphone AM1.

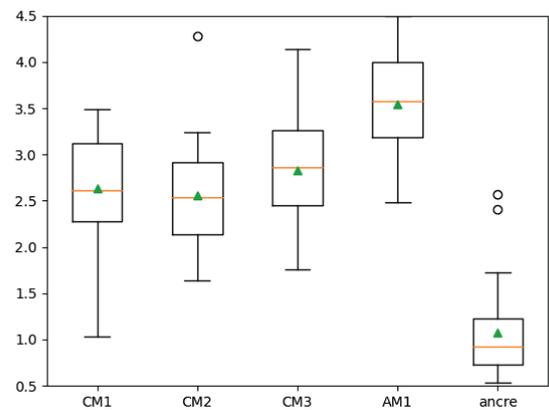


Figure 6 : Dispersion du score du test MUSHRA en fonction des auditeurs pour chaque microphone NC et les signaux d'ancre

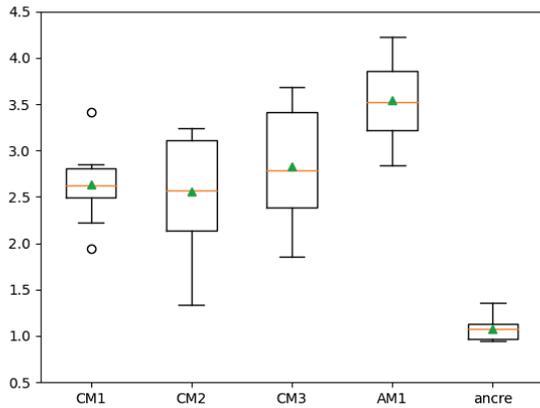


Figure 7 : Dispersion du score du test MUSHRA en fonction des locuteurs pour chaque microphone NC et les signaux d'ancre

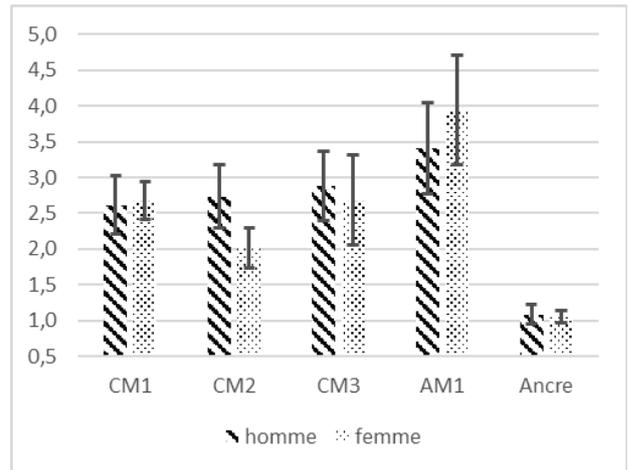


Figure 8 : Résultat du test MUSHRA moyen en fonction du genre du locuteur pour chaque microphone

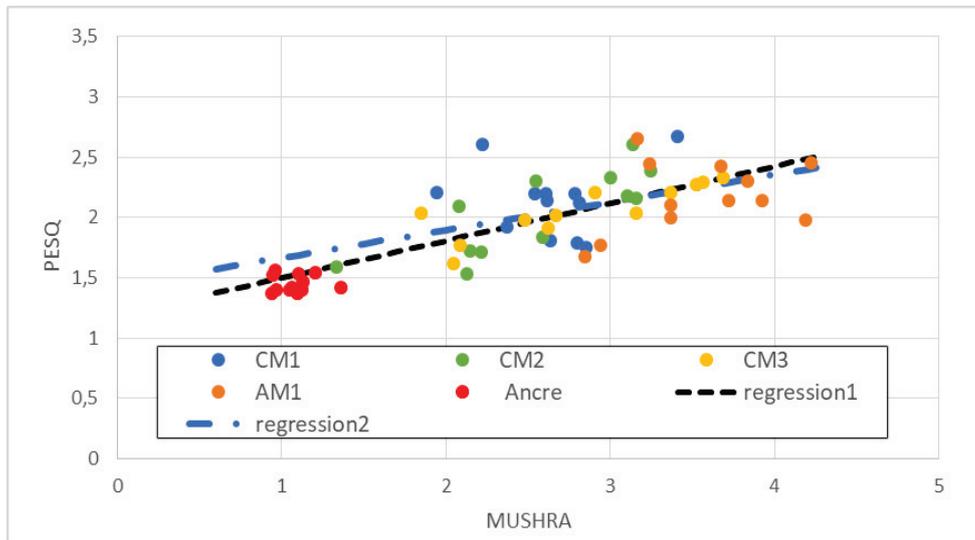


Figure 9 – Score PESQ par rapport à la note MUSHRA pour chaque locuteur et pour chaque configuration (4 microphones NC et signal ancre).

### 3.3 Discussion

L'objectif de cette étude est de voir s'il y a une corrélation entre la PESQ et le test MUSHRA. On a pu obtenir pour chaque locuteur et pour chaque microphone NC un score pour la méthode PESQ et une note moyenne de qualité perçue à l'aide du test de MUSHRA. La Figure 9 représente pour chaque configuration le score de PESQ en fonction de la note moyenne de la qualité perçue. Sur cette figure deux droites de régression ont été estimées :

- Droite de régression 1 en tenant compte de tous les points.
- Droite de régression 2 en excluant les points obtenus avec le signal ancre.

Pour chacune des deux droites de régression, le coefficient de corrélation  $R^2$  a été calculé. Ce coefficient vaut 0,6 pour la droite de régression 1 et chute à 0,26 pour la droite de régression 2. Ce faible taux de corrélation nous incite à dire qu'il n'y a pas ou peu de corrélation entre le score de la PESQ et la note moyenne de qualité perçue.

Même si la corrélation entre les scores de PESQ et la note moyenne de la qualité audio perçue est faible, on relève quelques similitudes entre ces deux méthodes :

- L'ordre des microphones est presque le même : en effet avec PESQ l'ordre est AM1, CM1, CM3 et CM2 et pour le test MUSHRA l'ordre est AM1, CM3, CM1, CM2.
- Que ce soit avec PESQ ou le test de MUSHRA, la comparaison entre les locuteurs féminins et masculins va dans le même sens sauf pour le microphone CM1

pour lequel on n'observe pas de différences avec le test de MUSHRA.

## 4 Conclusion & Perspectives

La méthode PESQ n'est pas parvenue à discriminer les microphones NC entre eux. Ce test montre que la qualité audio n'est pas bonne puisque le score moyen est de l'ordre de 2,1 sur 4,5. L'utilisation de cette méthode pour comparer les microphones non conventionnels entre eux n'est pas adéquate. Le score de PESQ est probablement dû principalement au manque de hautes fréquences que les signaux provenant de ces microphones ont.

Cependant, avec le test perceptif MUSHRA, un microphone NC a été jugé meilleur que les autres en termes de qualité audio. Mais le problème de cette méthode est la durée. En effet, faire passer les tests sur un nombre suffisant d'auditeurs demande beaucoup de temps.

Cette première étude a montré que la qualité audio du signal de parole provenant des microphones non conventionnels n'est pas excellente et mérite qu'on étudie plus profondément les dégradations causées par la « conduction osseuse ». Est-ce que la mauvaise qualité est seulement due au manque des hautes fréquences ou est ce qu'il y a d'autres phénomènes comme, par exemple, une résonance supplémentaire ?

L'objectif des études à venir est d'améliorer la qualité audio de ces microphones non conventionnels en essayant de comprendre les mécanismes de propagation du signal de parole entre la cavité buccale et les capteurs. Ce travail s'accompagnera d'une modélisation de cette propagation. Pour celle-ci, dans un premier temps, il sera nécessaire de répondre à la question du nombre de sources sonores à modéliser. En effet, l'hypothèse de considérer qu'une seule source sonore localisée dans la cavité buccale où sont construits les différents formants, permettrait de simplifier la modélisation, mais est-ce suffisant. Dans le cas contraire, il serait nécessaire de modéliser aussi la source primaire, à savoir les vibrations des cordes vocales.

## Remerciements

Nous remercions les 12 locuteurs volontaires qui ont accepté de s'équiper des différents microphones NC et autorisé l'utilisation de leur voix. Nous remercions aussi les 25 auditeurs ayant participé au test perceptif MUSHRA.

## Références

- [1] E. H. Greibach, & L. G. Pacent, Inertia throat microphones. Transactions of the American Institute of Electrical Engineers, 65(4), 187-191 (1946).
- [2] P. Hamery, V. Zimpfer, S. Hengy, G. Nexer, assessment feedback of intelligent earplug, 2019, in proceeding ICSV26, Montréal.
- [3] A. Shahina, B. Yegnanarayana., & M. R. Kesheorey, Throat microphone signal for speaker recognition, 2004, October In Proceedings of ICSLP.
- [4] V. Zimpfer, & K. Buck, Ambiguity in the recognition of phonetic vowels when using a bone conduction microphone, 2008, Journal of the Acoustical Society of America, 123(5), 3067.
- [5] K. Buck, & V. Zimpfer, Ambiguity in phoneme recognition when using a bone conduction microphone, 2011, the Journal of the Acoustical Society of America, 130(4), 2469-2469.
- [6] P. Tran, T. Letowski, & M. McBride, Bone conduction microphone: Head sensitivity mapping for speech intelligibility and sound quality. In 2008 International Conference on Audio, Language and Image Processing (pp. 107-111), IEEE.
- [7] M. McBride, P. Tran, T. Letowski, & R. Patrick, The effect of bone conduction microphone locations on speech intelligibility and sound quality, March 2011, Applied Ergonomics, 42, 495-502. doi:10.1016/j.apergo.2010.09.004
- [8] [https://www.talkingheadsets.co.uk/productpdfs/pdfs/INVISIO-Brochure\\_2905141020.pdf](https://www.talkingheadsets.co.uk/productpdfs/pdfs/INVISIO-Brochure_2905141020.pdf)
- [9] P. Combescure, 20 listes de dix phrases phonétiquement équilibrées, Revue d'acoustique, 34-38, 1981
- [10] D. Dirks & J.G. Swindeman, The variability of occluded and unoccluded bone-conduction thresholds, J. of Speech and Hearing Res., 10, 232-249, (1967).
- [11] P.862 : Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs". www.itu.int. Retrieved 2021-04-20.
- [12] P.862.1 : Mapping function for transforming P.862 raw result scores to MOS-LQO". www.itu.int. Retrieved 2021-04-11.
- [13] TU-R recommendation BS.1534-3