



HAL
open science

Altérations des formants par l'utilisation de microphones “ non conventionnels ”

J. Richard, Véronique Zimpfer, Thomas Joubaud, Pascal Hamery, Sebastien
Roth

► **To cite this version:**

J. Richard, Véronique Zimpfer, Thomas Joubaud, Pascal Hamery, Sebastien Roth. Altérations des formants par l'utilisation de 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-03848422

HAL Id: hal-03848422

<https://hal.science/hal-03848422v1>

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^{ème} Congrès Français d'Acoustique
11-15 Avril 2022, Marseille

Altérations des formants par l'utilisation de microphones «non conventionnels»

J. Richard^{a,b}, V. Zimpfer^a, T. Joubaud^a, P. Hamery^a, et S. Roth^b

^a ISL, Institut de Recherche de St Louis, 5 rue du Général Cassagnou, BP 70034, 68301 St Louis, France,

^b Laboratoire Interdisciplinaire de Carnot-Bourgogne, département Conception, Optimisation et Modélisation Mécanique. F-90400 SEVENANS.



Dans les environnements bruyants la communication à distance est un véritable défi. L'utilisation de microphones classiques enregistrant l'onde aérienne émise par la bouche et les narines peut poser problème tant sur le plan de l'ergonomie que sur celui de la robustesse dans l'acquisition du signal de parole. Les microphones non conventionnels (laryngophones, microphones à conduction osseuse, microphones intra-auriculaires ...) qui mesurent le signal de parole se propageant au travers du corps, apportent une nouvelle solution. En effet, ils sont moins sensibles aux bruits environnants et sont compatibles avec des équipements de tête comme les masques à gaz. Cependant l'intelligibilité et la qualité du signal enregistré avec ce type de microphones est un frein à leur utilisation. Ainsi l'étude de ces derniers représente un grand intérêt. Pour comprendre leurs effets sur le signal de parole, des volontaires équipés de différents types de microphones ont prononcé une série de voyelles phonétiques. A l'aide des suivis des formants, une comparaison des microphones « non conventionnels » avec un microphone aérien de référence, a été réalisée. La présente étude donne des premiers résultats, premiers pas vers la compréhension de la perte d'intelligibilité engendré par ces microphones.

1 Introduction & Contexte

La communication dans les environnements bruyants est une problématique intervenant dans plusieurs domaines aussi bien militaires (le champ de bataille) que civil (l'industrie). Les microphones aériens que l'on qualifiera de « conventionnels » enregistrant l'onde aérienne émise par la bouche et les narines peuvent être utilisés pour la prise de parole dans ces milieux. Cependant, l'utilisation d'un seul microphone aérien ne permet pas de débruiter efficacement la parole. En effet, lorsqu'un seul capteur est utilisé, la phase du bruit ne pouvant être connue, des artefacts apparaissent lors du traitement du signal. Ainsi, pour pallier à cela des méthodes utilisant un réseau de microphones peuvent être mises en place, mais leurs performances dépendent du nombre de capteurs utilisés. Les microphones de ce type les plus employés aujourd'hui sont les microphones différentiels qui combinent deux microphones orientés dans des sens opposés. Ceci permet un bon compromis entre le nombre de microphones et l'encombrement. Ils sont par exemple utilisés par les gendarmes motocyclistes ainsi que par les pilotes et passagers d'hélicoptère. Cependant, l'utilisation de microphones « conventionnels » est incompatible avec le port d'équipement recouvrant la bouche tel que les masques à gaz et, en présence de bruit de très fort niveau continu ou transitoire, le signal de parole est souvent noyé dans le bruit. Ainsi pour pallier à ces inconvénients d'autres types de capteurs ont été développés.

Les microphones qualifiés ici de « non-conventionnels » (NC) n'enregistrent pas le signal de parole aérien émis par la bouche et les narines, mais un signal provenant des vibrations produites lors de la création de la parole. Il peut s'agir d'un capteur enregistrant la vibration en un point du corps ou d'un microphone aérien placé dans une cavité du corps. L'un des premiers microphones NC utilisé est le laryngophone [1] qui capte la vibration de la peau au niveau du cou et donc à proximité des cordes vocales. Ce type de microphone à contact est encore employé aujourd'hui, mais l'étude des microphones à conduction osseuse s'est poursuivie. Des études portant sur d'autres localisations ont été menées pour trouver le meilleur placement de ces microphones [2] [3]. D'autres types de microphones NC se plaçant dans le conduit auditif, propose la combinaison de deux fonctions principales à savoir la protection auditive et la communication [4] [5]. Dans ce dernier cas, deux solutions existent : la mesure des vibrations avec un capteur placé sur le bouchon acoustique de protection ou l'enregistrement de

l'onde aérienne reconstruite dans le volume résiduel entre le bouchon et le tympan.

Une étude réalisée par l'OTAN [6] a comparé des microphones NC (un laryngophone, un ensemble de microphones à conduction osseuse placé sur la tête et un microphone intra-auriculaire enregistrant les vibrations dans l'oreille) à un microphone différentiel. Ce rapport montre que dans le silence et dans des milieux où le bruit environnant est diffus, les microphones différentiels ont une meilleure intelligibilité tandis que le rapport signal sur bruit est meilleur pour les microphones NC. Ces derniers présentent donc une moins bonne intelligibilité dans ces milieux et cela est aujourd'hui un frein à leur utilisation.

Cette étude, se plaçant dans un projet d'amélioration de l'intelligibilité des microphones NC, a pour premier objectif d'essayer de comprendre l'origine de cette faible intelligibilité. En effet, l'expérience réalisée par Zimpfer et Buck en 2008 [7] montre que certaines voyelles peuvent être confondues lorsqu'elles sont enregistrées avec un microphone à conduction osseuse. Le groupe de voyelles où l'inter-confusion est la plus grande est composé des voyelles /u/, /y/, /i/. Ainsi, le choix a été fait de se pencher dans cette étude sur ce groupe de voyelles partageant le même premier formant. En se basant sur cette confusion et sur l'importance des deux premiers formants dans la reconnaissance des voyelles [8] [9], cet article propose de se focaliser sur les deux premiers formants de ces trois voyelles enregistrées par différents microphones NC.

2 Matériels et Méthodes

Cette partie rend compte des différents microphones NC utilisés, du protocole de mesure ainsi que du traitement des signaux enregistrés à l'aide des microphones NC et d'un microphone aérien de référence.

2.1 Microphones non conventionnels

Quatre types de microphones NC ont été évalués lors de cette expérience :

- Un laryngophone provenant du commerce et commercialisé sous la marque RETEVIS.
- Un microphone intra-auriculaire captant l'onde acoustique dans le conduit auditif ayant été réalisé par l'ISL (Institut franco-allemand de recherches de Saint-Louis). Il est composé d'un

microphone Knowles (série FG) et d'un bouchon d'oreille à trois ailettes.

- Un microphone intra-auriculaire enregistrant les vibrations du canal auditif, le M3h/X5, provenant du commerce. Ce microphone, réalisé par Invisio, est considéré comme une référence dans la prise de son dans l'oreille et est protégé par un brevet.
- Un microphone à conduction osseuse qui n'est autre qu'un accéléromètre placé très proche d'une surface osseuse. Ce microphone est aussi le fruit de travaux menés à l'ISL et est similaire à celui présent dans le BCH300 commercialisé par Elno [10].

Pour simplifier la lecture, une dénomination est donnée à chaque microphone. Les microphones nécessitant un contact seront nommés sous la forme CM# et les autres sous la forme AM#.

Tableau 1 - Dénomination des microphones.

Microphones	Dénominations
Microphone à conduction osseuse	CM1
Microphone intra-auriculaire enregistrant l'onde acoustique dans le conduit auditif	AM1
Microphone intra-auriculaire mesurant les vibrations du canal auditif	CM2
Laryngophone	CM3

Les microphones utilisés sont présentés sur la Figure 1.



Figure 1 – Photographies des microphones NC utilisés.

2.2 Enregistrement

L'expérience est réalisée dans une cabine d'audiométrie à double paroi où le bruit ambiant est inférieur à 30 dB(A).

Le sujet volontaire est équipé des quatre microphones NC présentés dans le paragraphe précédent. Le microphone CM1 est positionné sur le front du sujet puisqu'il s'agit de la localisation où l'intelligibilité est la meilleure parmi toutes les zones explorées par McBride [2] et Tran [3]. Le microphone CM2 est lui positionné à l'intérieur du canal auditif en suivant les recommandations du constructeur [11]. La bulle d'air (protubérance sur le microphone visible en Figure 1) doit être en contact avec le canal auditif. Le microphone CM3 est fixé sur la partie haute de la gorge et AM1 est placé dans le canal auditif tel un bouchon d'oreille. La Figure 2 illustre la position des différents microphones.

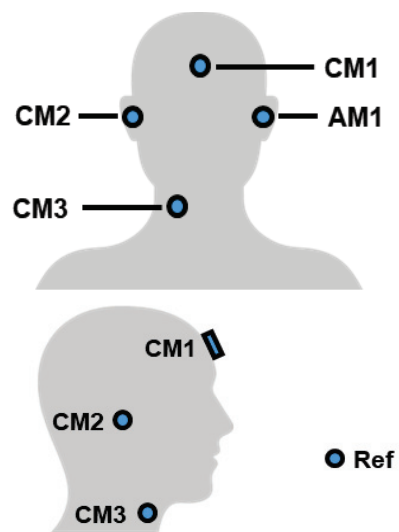


Figure 2 – Position des différents microphones NC.

Le sujet équipé de ces microphones prononce les 3 voyelles /i/ /y/ /u/ parmi une série de voyelles orales et nasales qui ne seront pas traitées dans cet article. Les trois voyelles étudiées sont :

- /i/ comme dans « **pie** »
- /y/ comme dans « **rue** »
- /u/ comme dans « **sous** »

Un microphone aérien est utilisé comme microphone de référence. Le capteur choisi est un BK ½ pouce Type 2669 placé à 50 cm de la bouche. L'ensemble des signaux provenant des différents capteurs sont enregistrés simultanément. Cette tâche est gérée par ordinateur via une carte son ZOOM F8n. Les signaux sont enregistrés par le logiciel Adobe Audition avec une fréquence d'échantillonnage de 48 kHz et une quantification de 16 bits. Aucun traitement n'est appliqué au signal du microphone AM1 contrairement aux autres microphones NC utilisés. C'est pourquoi il est filtré par un passe-haut de second ordre dont la fréquence de coupure est 1,9 kHz. Ce filtre a pour objectif de réduire l'effet d'occlusion observé dans l'oreille et est déterminé en se basant sur l'étude de Dircks et Swindeman [12]. L'expérience a été réalisée sur un groupe

de 12 sujets volontaires composés de 3 femmes et de 9 hommes.

2.3 Traitement

Pour les trois voyelles /i/ /y/ et /u/, une analyse de suivi de formants a été menée sur les enregistrements à l'aide du logiciel Praat [13]. Le logiciel calcule les coefficients cepstraux de prédiction linéaire (LPCC) selon l'algorithme de Burg [14] [15] après avoir ré-échantillonné le signal et l'avoir pré-accentué. Pour un locuteur donné et une voyelle prononcée, la moyenne des valeurs obtenues au cours de la prononciation a été calculée et utilisée comme fréquence du formant. En posant l'hypothèse que la perte d'intelligibilité peut être due à la modification de valeurs des formants, une évaluation de celle-ci est proposée. Pour ce faire, une métrique ΔF_n est définie. Celle-ci correspond à la valeur absolue de l'écart observé entre les fréquences du n^{ième} formant lorsque le signal est enregistré par un microphone NC et un microphone aérien de référence :

$$\Delta F_n = |F_{n_{reference}} - F_{n_{microphone\ NC}}|, \quad (1)$$

où F_n correspond à la fréquence du n^{ième} formant.

Pour chaque locuteur, microphone et voyelle, ΔF_1 et ΔF_2 sont déterminés.

3 Résultats

3.1 Effet Microphone – Locuteur

Pour représenter synthétiquement les résultats de l'étude, le choix a été fait de recourir au diagramme de Tukey. Les diagrammes de Tukey de ΔF_1 et ΔF_2 en fonction des microphones NC ont été réalisés pour chacune des trois voyelles étudiées et sont présentés en Figure 3, Figure 4 et Figure 5 pour les voyelles /i/, /y/ et /u/, respectivement.

Les Figure 3, Figure 4 et Figure 5 mettent en évidence que, pour un microphone NC donné, la valeur du formant et plus spécifiquement du second formant peut être très proches ou très éloigné de la valeur de référence en fonction du locuteur. La dispersion et l'écart interquartile renseignent sur l'influence du locuteur en explicitant la variabilité de la valeur du formant. Ainsi, pour le premier formant, les microphones CM1 et CM2 présentent le plus d'effets liés au locuteur. Pour le second formant, le microphone AM1 semble le moins sensible au changement de locuteurs. Pour les voyelles /i/ et /u/, l'effet du locuteur est bien plus significatif pour le second formant que pour le premier. Cet effet apparaît pour tous les microphones NC étudiés. En outre, le microphone AM1 présente les valeurs médianes de ΔF_2 les plus faibles. La médiane étant prise comme indicateur pour observer l'effet du microphone, le microphone AM1 est dans la majorité des cas meilleur sur ce formant que les autres microphones NC étudiés.

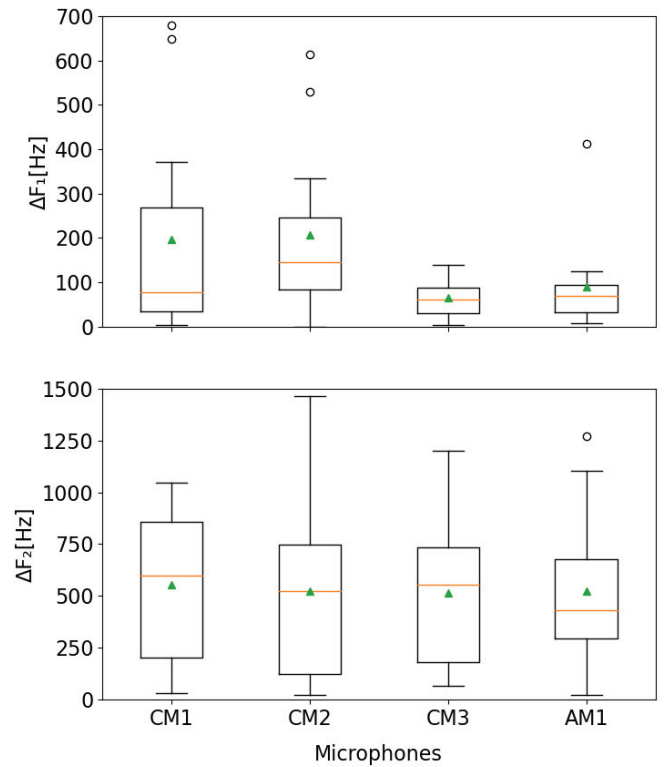


Figure 3 : Diagrammes de Tukey de ΔF_1 et ΔF_2 de la voyelle /i/ en fonction des microphones NC. Le graphique du haut correspond à ΔF_1 et le graphique du bas à ΔF_2 .

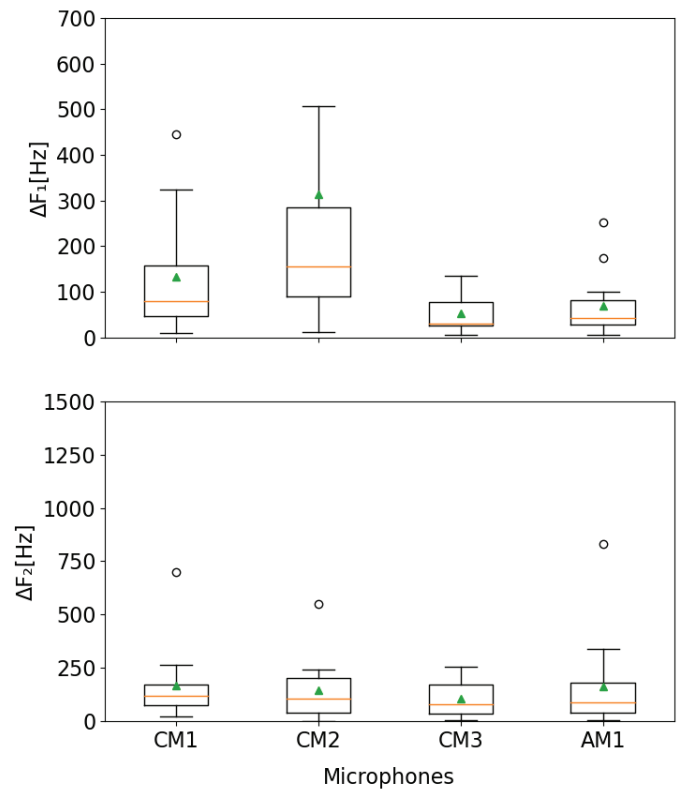


Figure 4 : Diagrammes de Tukey de ΔF_1 et ΔF_2 de la voyelle /y/ en fonction des microphones NC. Le graphique du haut correspond à ΔF_1 et le graphique du bas à ΔF_2 .

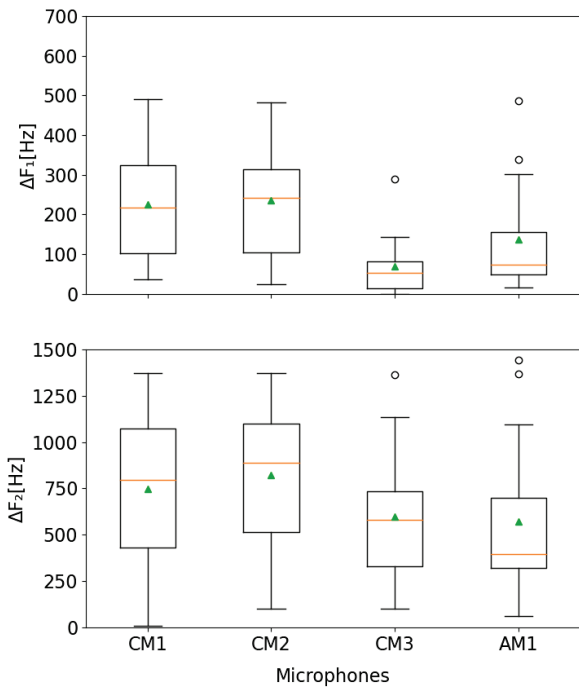


Figure 5 : Diagrammes de Tukey de ΔF_1 et ΔF_2 de la voyelle /u/ en fonction des microphones NC. Le graphique du haut correspond à ΔF_1 et le graphique du bas à ΔF_2 .

Les valeurs médianes des seconds formants sont présentées en Figure 6 pour chacune des voyelles prononcées en fonction du microphone. Le microphone de référence est placé en premier sur le graphique pour permettre de comparer plus aisément les autres microphones NC à ces valeurs. La fréquence du second formant des voyelles /u/ et /i/ est particulièrement affectée par la plupart des microphones NC. Elle semble plus proche, dans la majorité des cas, du second formant de référence de la voyelle /y/ que de celui de la voyelle prononcée. Ceci pourrait créer une confusion lors de la reconnaissance des voyelles se basant sur les deux premiers formants. De plus, un auditeur pourrait reconnaître la voyelle /y/ alors que la voyelle /i/ ou /u/ est prononcée. Cette potentielle ambiguïté a été montrée à l'aide d'un test subjectif pour des microphones à conduction osseuse par Zimpfer et Buck [7].

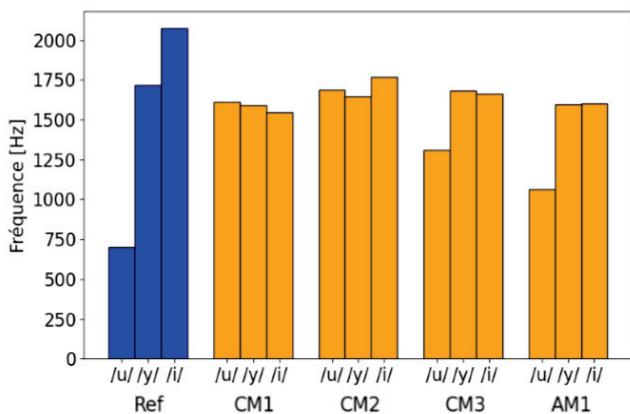


Figure 6 : Valeurs des seconds formants pour les voyelles /u/ /y/ et /i/ en fonction des microphones.

3.2 Effet du genre du locuteur

Sachant que les microphones NC enregistrent le signal de parole ayant été conduit à travers le corps, la morphologie de chaque personne semble être une cause importante de la variabilité des valeurs de formant observées entre les locuteurs. Une différence dépendant du genre du locuteur a été considérée. Les valeurs médianes de ΔF_1 et ΔF_2 ont été déterminées en séparant les locuteurs en deux groupes selon le sexe. La Figure 7 représente la médiane de ΔF_1 et ΔF_2 pour les voyelles /i/, /u/ et /y/ en fonction du genre et du microphone utilisé. Les locutrices semblent avoir une erreur plus faible que les hommes lors de la prononciation de la voyelle /i/. L'inverse se produit pour la voyelle /u/. Cependant les écarts entre les locutrices étant très grands pour le second formant, le genre n'explique pas la variabilité de ΔF_2 vue précédemment. De plus le microphone CM3 ne semble pas être meilleur pour un groupe tandis que la grande différence morphologique entre ces deux groupes se situe au niveau de la gorge. Ainsi, le genre du locuteur ne semble pas être un facteur déterminant permettant d'expliquer la variabilité de ΔF_1 et ΔF_2 observée précédemment pour les microphones. Une grande différence entre les locuteurs hommes et femmes est cependant observée pour le premier formant lorsque le signal est enregistré par le microphone CM2. Le graphique met en évidence que pour ce microphone, la valeur de ΔF_1 est bien plus faible chez les locuteurs féminins que chez les locuteurs masculins. Cet effet pourrait être la conséquence du traitement intégré au microphone agissant comme passe-haut pour réduire l'effet de la fondamentale, mais coupant le premier formant de ces voyelles chez les hommes (en général, plus bas que celui des femmes). Afin de confirmer ces résultats, il serait possible de réitérer l'expérience avec un plus grand nombre de locutrices.

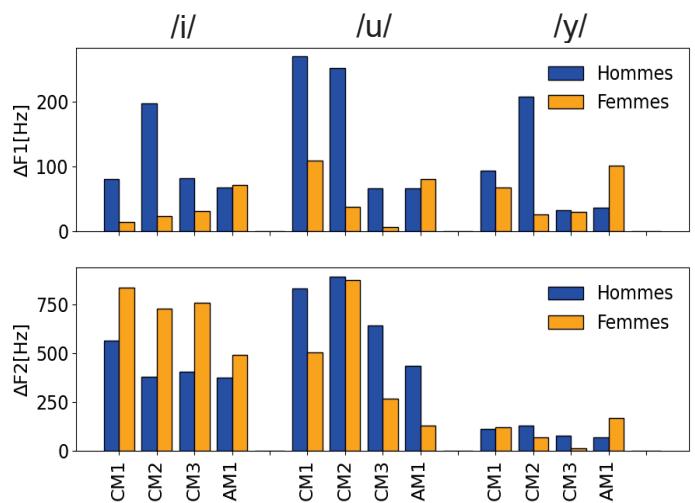


Figure 7 : Valeurs médianes de ΔF_1 et ΔF_2 en distinguant les locuteurs en deux groupes (hommes, femmes). La première série de résultat correspond à la voyelle /i/, la seconde /u/ et la troisième à la voyelle /y/. Le graphique du haut correspond à ΔF_1 et le graphique du bas à ΔF_2 .

4 Discussion & Perspectives

Le signal recueilli par des microphones qualifiés dans cet article de « non-conventionnels » peut présenter des valeurs de formants différentes du signal enregistré par un microphone aérien classique. Cette différence a une réelle importance puisque pour la majorité des locuteurs, la valeur du second formant pour les voyelles /i/ et /u/ enregistré par les microphones NC étudiés est plus proche du second formant de la voyelle /y/ que de leur second formant de référence. Ainsi, l'ambiguïté relevée par Zimpfer & Buck [7] entre ces trois voyelles s'étendrait à la reconnaissance objective par l'utilisation des premiers formants. La confusion existe ainsi pour les microphones à conduction osseuse, mais semble aussi s'étendre aux autres microphones étudiés ici à savoir les microphones intra-auriculaires et le laryngophone. L'expérience pourrait être réitérée avec plusieurs microphones NC de même nature permettant de confirmer les résultats obtenus et de se soustraire aux différences existantes entre les constructeurs.

Une étude subjective sur la reconnaissance des voyelles pourrait venir compléter ces résultats. Elle permettrait de déterminer si une métrique composée de ΔF_2 peut être cohérente avec les résultats d'un test subjectif et ainsi de rendre compte à elle seule de l'ambiguïté observée entre les voyelles enregistrées par les microphones NC.

Il est à noter, par ailleurs, que les microphones non-conventionnels présentent une efficacité dépendant du locuteur. Cette différence ne semble pas explicable par le genre, cela pourrait donc résulter de différences morphologiques ou de légères différences de positionnement du capteur. Ce positionnement pourrait traduire l'observation d'une plus faible influence du locuteur pour le microphone AM1 ; celui-ci n'ayant pas de surface de contact.

Cette étude s'inscrit dans une perspective d'amélioration de l'intelligibilité des microphones NC. Elle a permis d'évaluer l'effet de différents microphones NC sur différents locuteurs. L'étude a mis en exergue que l'effet d'un même microphone peut être différent en fonction du locuteur ou de la zone de contact. Ceci incite à s'intéresser à l'étude des chemins de propagation de la parole aériens et tissulaires vers ces microphones. La modélisation de ces derniers pourrait révéler la cause de la perte d'intelligibilité et poser les bases d'une future amélioration.

Remerciements

Les auteurs remercient l'ensemble des locuteurs de s'être portés volontaires et d'avoir permis la réalisation de cette étude.

Références

- [1] B. Stuart, «THROAT MICROPHONE». Brevet US2165124, 1939.
- [2] M. McBride, P. Tran, T. Letowski et R. Patrick, «The effect of bone conduction microphone locations on speech intelligibility and sound quality,» *Applied Ergonomics*, vol. 42, p. 495–502, March 2011.
- [3] 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.
- [4] A. L. Aakervik, «An earplug passes the word,» Gemini, 2000.
- [5] O. M. Strand, T. Holter, A. Egeberg et S. Stensby, «On the feasibility of ASR in extreme noise using the PARAT earplug communication terminal», 2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721).
- [6] B. Acker-Mills, A. Houtsma, W. Ahroon “Speech Intelligibility with Acoustic and Contact Microphones.” (2005).
- [7] V. Zimpfer et K. Buck, «Ambiguity in the recognition of phonetic vowels when using a bone conduction microphone,» *The Journal of the Acoustical Society of America*, vol. 123, p. 3067–3067, May 2008.
- [8] J. Clark et C. Yallop, *An Introduction to Phonetics and Phonology*, 1990.
- [9] A. Landercy, *Éléments de phonétique*, Bruxelles Mons: Didier Centre international de phonétique appliquée, 1982.
- [10] ELNO Bandeau communicant BCH300 [accessed 2021 Nov 3]. <https://www.elno.fr/products/bch300/>
- [11] INVISIO INVISIO X5 [accessed 2021 Nov 3]. <https://invisio.com/communication-systems/headsets/invisio-x5/>
- [12] D. Dirks & J.G. Swindeman, The variability of occluded and unoccluded bone-conduction thresholds, *J. of Speech and Hearing Res.*, 10, 232-249, (1967).
- [13] P. Boersma & D. Weenink(2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.09, retrieved 15 February 2022 from <http://www.praat.org/>
- [14] D. Childers, *Modern spectrum analysis*, New York: IEEE Press Sole worldwide distributor (exclusive of IEEE), Wiley, 1978.
- [15] W. Press, *Numerical recipes in C: the art of scientific computing*, Cambridge Cambridgeshire New York: Cambridge University Press, 1992.