



HAL
open science

Classification automatique de voyelles nasales pour une caractérisation de la qualité de voix des locuteurs par des réseaux de neurones convolutifs

Lila Kim, Cédric Gendrot

► **To cite this version:**

Lila Kim, Cédric Gendrot. Classification automatique de voyelles nasales pour une caractérisation de la qualité de voix des locuteurs par des réseaux de neurones convolutifs. Journées d'Etude de la Parole 2022 (JEP 2022), Nantes Université, Jun 2022, Noirmoutier, France. halshs-03980367

HAL Id: halshs-03980367

<https://shs.hal.science/halshs-03980367>

Submitted on 9 Feb 2023

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.

Classification automatique de voyelles nasales pour une caractérisation de la qualité de voix des locuteurs par des réseaux de neurones convolutifs

Lila KIM, Cédric GENDROT

Laboratoire de Phonétique et Phonologie

19 rue des Bernardins 75005 Paris

lila.kim@sorbonne-nouvelle.fr, cedric.gendrot@sorbonne-nouvelle.fr

RESUME

L'objectif de cette étude est d'évaluer, au moyen de réseaux de neurones convolutifs (CNN pour Convolutional Neural Networks), la capacité à discriminer acoustiquement une voyelle orale d'une voyelle nasale (voyelles /a/ et /ã/) sur un corpus de 45 locuteurs francophones. Pour une classification voyelle nasale vs. voyelle orale, les résultats obtenus peuvent atteindre 96% d'identification correcte. Ces résultats sont détaillés en fonction du locuteur, du contexte phonémique et de la durée. Dans la discussion, nous présentons une comparaison de ces résultats avec une mesure physiologique de nasalance. Une technique de visualisation (Grad-cam) est également mise en avant pour comprendre les zones du spectrogramme utilisées par le CNN.

ABSTRACT

Automatic classification of nasal vowels for speaker voice quality characterization by convolutional neural networks

The objective of this study is to evaluate, the ability to acoustically discriminate an oral vowel from a nasal vowel (vowels /a/ and /ã/) on a corpus of 45 French speakers with convolutional neural networks (CNN). We reached a 96% correct identification for a nasal vowel vs. oral vowel classification. These results are detailed according to speaker, phonemic context and duration. In the discussion, we present a comparison of these results with a physiological nasalance measure. A visualisation technique (Grad-cam) is also presented to understand the areas of the spectrum used by the CNN.

MOTS-CLES : classification automatique, traitement automatique, réseaux de neurones convolutifs, nasalité, qualité de voix, caractérisation du locuteur

KEYWORDS: automatic classification, automatic processing, convolutional neural networks, nasality, voice quality, speaker characterization

1 État de l'art

La qualité de la voix est considérée dans la littérature comme ayant de grandes implications pour la caractérisation des locuteurs (Gold & French, 2019). Elle peut être un élément permanent de la voix d'un locuteur en raison de particularités physiologiques (principalement laryngées et supralaryngées) ou d'habitudes du locuteur (facteurs sociolinguistiques), mais elle est également sujette à la variabilité intra-locuteur, notamment dans le style de discours ou l'émotion (Nolan, 2005). Quelques exemples connus de la qualité de la voix sont la nasalité, la voix tendue/relâchée, la voix sombre/claire ou encore le type de phonation. Si la qualité de voix ne permet pas d'identifier un locuteur stricto sensu, elle permet en revanche d'apporter une caractéristique fiable en complément d'autres caractéristiques comme la fréquence fondamentale ou l'articulation des voyelles et des consonnes.

La nasalité est utilisée pendant la production de parole, pour la distinction phonologique entre nasales et orales, que ce soit pour les voyelles (un /a/ et un /ã/ par exemple) ou bien les consonnes (un /b/ et un /m/ par exemple). Elle s'effectue par un abaissement du voile du palais qui permet au flux d'air de traverser la cavité nasale, en plus de la cavité orale. Cette double cavité dans laquelle l'air vient résonner a pour effet acoustique de créer des anti-résonances dans le spectre sonore. On parle de « zéros » ou « d'anti-formants » qui font baisser l'énergie dans le spectre au lieu de le renforcer comme le font les résonances de la cavité orale (Maeda, 1982). Ces anti-formants complexifient les analyses acoustiques de la nasalité. Celles-ci sont rendues d'autant plus difficiles que l'abaissement du voile du palais peut également avoir pour conséquence d'altérer la forme de la cavité orale et ainsi de modifier les formants (Carignan, 2017). À partir de la modélisation acoustique et de la compréhension de ces anti-formants, des analyses mesurant la différence d'amplitude entre les formants et les anti-formants ont été entreprises (Chen, 1997 ; Styler, 2017), mais ces analyses sont très dépendantes des caractéristiques articulatoires de chaque son, voire du locuteur, et ne peuvent être effectuées que dans des conditions très spécifiques où tous les contextes sont contrôlés. Dans l'optique d'une extension à l'analyse de corpus de parole plus écologiques, il est donc nécessaire d'essayer d'autres méthodes pour l'analyse acoustique de la nasalité.

Dans le domaine de la reconnaissance automatique du locuteur, il a fréquemment été observé (Kahn et al., 2011) que les nasales sont plus pertinentes pour caractériser le locuteur que d'autres phonèmes. L'explication apportée est que la cavité nasale est à la fois différente d'un locuteur à l'autre, et également peu malléable pendant la production de parole (Dang et al., 1994 ; Serrurier, 2006), ce qui en ferait une cavité de résonance stable et propre au locuteur. L'abaissement du voile du palais dure en moyenne 50 ms (Amelot, 2002), mais est sujet à d'importantes variations inter-locuteurs. Si ces variations sont souvent impliquées dans l'évolution diachronique d'une langue (Ohala, 1993), elles pourraient donc impliquer des différences entre locuteurs.

Notre travail a pour but de démontrer la capacité des réseaux de neurones convolutifs à reconnaître la nasalité dans les voyelles en nous penchant également sur l'analyse des erreurs de classification, particulièrement selon les contextes phonémiques, le locuteur et la durée des voyelles. La classification entre les voyelles orales et nasales nous permet d'obtenir une probabilité de nasalité avec laquelle nous nous proposons de comparer la nasalance sur un corpus de vérification. Nous analyserons également les zones utilisées par les CNN pour chercher la nasalité, et ce, au moyen d'une technique de visualisation appelée grad-cam (Gradient-weighted Class Activation Mapping) que l'on peut visualiser sous forme de carte de chaleur.

2 Protocole expérimental

2.1 Corpus

Les données du corpus utilisées ici ont été recueillies dans le corpus NCCFr (The Nijmegen Corpus of Casual French, Torreira et al., 2015) avec au total plus de 36 heures de conversation amicale de 46 locuteurs français. La sélection des échantillons s’est basée sur une paire de deux voyelles (/a/ et /ã/) ; elles ont le même degré d’aperture, et se distinguent par la nasalité (données 1). Pour objectif d’étudier le phénomène de coarticulation nasale, un autre jeu de données a été recueilli sur le même corpus (données 2) : les voyelles orales /a/ dans un contexte oral et les voyelles orales /a/ dans un contexte nasal. Elles ont été déclarées respectivement comme « non nasales » et « nasales ». Dans le jeu de données 3, un regroupement des données 1 et 2 a été opéré. Ces données sont résumées dans le tableau 1.

	Catégorie	Phonème	Entraînement et validation	Test
Données 1	nasal	/ã/	7 200	1 800
	non nasal	/a/ dans un contexte oral	7 200	1 800
Données 2	nasal	/a/ dans un contexte nasal	4 725	1 170
	non nasal	/a/ dans un contexte oral	4 725	1 170
Données 3	nasal	/ã/	4 725	1 170
	oral	/a/ dans un contexte oral	4 725	1 170
	oral dans un contexte nasal	/a/ dans un contexte nasal	4 725	1 170

Tableau 1: résumé des données utilisées

2.2 CNN

Nous avons utilisé des réseaux de neurones convolutifs (CNN) qui font la classification à partir d’images. Pour convertir les voyelles en image, nous avons eu recours aux spectrogrammes au moyen d’un script praat. Comme il existe peu de voyelles /ã/ comparé aux voyelles /a/, une restriction sur la durée des voyelles /a/ a été mise en place. Dans notre corpus, la durée des voyelles /a/ en contexte oral en moyenne est égale à 62 ms, les voyelles /a/ ayant une durée supérieure à cette durée en moyenne ont été considérées. La durée moyennée des voyelles /ã/ est 78 ms et celle des /a/ en contexte

nasal est 61 ms. Comme vu dans le tableau 1, 80% de spectrogrammes ont été utilisés pour entraîner et valider les modèles, et 20% de données pour le test. La taille d'image en entrée d'une couche de convolution est de 28x28. Trois modèles CNN ont été initialisés, ils contiennent deux blocs de couches de convolution et de pooling, et terminent par deux couches denses. Le premier bloc contient 32 filtres et le second en a 64. Le même modèle a été utilisé séparément pour chacun des 3 jeux de données.

3 Résultats

Le tableau 2 récapitule les résultats obtenus par chaque modèle : les proportions des voyelles correctement identifiées et celles par rapport à leur durée ainsi que la mesure de Kappa de Cohen. La performance de nos trois modèles sera mesurée par les valeurs de précisions (accuracy) ainsi que par la méthode du kappa de cohen. Les valeurs du kappa sont calculées comme suit : $\kappa = (p\theta - pe)/(1 - pe)$ où $p\theta$ signifie la valeur de précision (accuracy) du modèle et pe représente les probabilités de l'accord entre les classes réelles et les prédictions faites par le modèle au hasard. Plus la valeur du kappa est basse, plus le modèle prédit les classes au hasard. Nous nous intéressons également à la répartition des erreurs dans les voyelles selon leur durée, le locuteur et les contextes phonémiques. Trois catégories ont été établies pour la durée qui sont réparties équitablement en termes de quantité de données au sein de chaque catégorie de phonème : court, moyen et long. Une t-SNE est également effectuée pour observer la similarité entre les instances de chaque jeu de données, elle permet de visualiser les données dans un espace à deux dimensions en réduisant la dimensionnalité des caractéristiques de la dernière couche cachée de pooling.

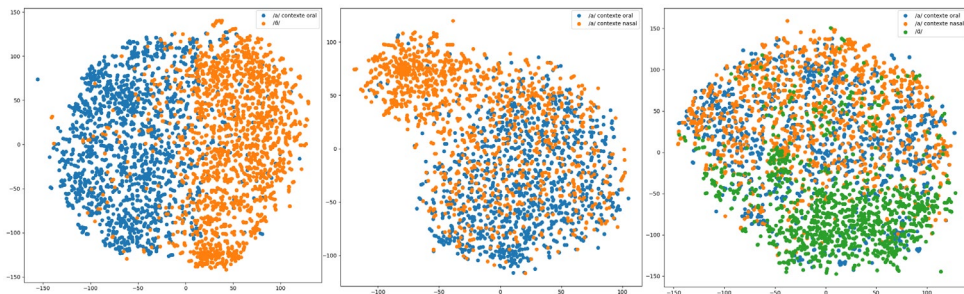


Figure 1: t-SNE obtenues par trois modèles 1, 2, 3

La figure 1 récapitule au moyen d'une t-SNE les classifications pour nos trois modèles. L'image à gauche de la figure 1 montre qu'il existe une distinction entre les instances nasales et non nasales. Les échantillons de deuxième image se rassemblent et sont séparés selon leur catégorie. Concernant l'excroissance en haut à gauche, 80% des instances dans cette zone ont une durée entre 30 et 40 ms, soit 32% sur la totalité des instances identifiées comme nasales. Ceci permet de visualiser que le modèle 2 est plus performant pour classer les voyelles nasales courtes comme nasales. Enfin, la troisième t-SNE montre une similitude entre les instances des /a/ en contexte oral et /a/ en contexte nasal car ils sont fréquemment confondus. La position des voyelles /ã/, quant à elle, est bien distincte des autres catégories.

Nous observons que, dans le modèle 1, 94% des voyelles nasales et 96% des voyelles orales sont correctement identifiées ($\kappa = 0.9$). Quelle que soit la durée, la classification n'est pas affectée. Pour les locuteurs, nous avons repéré d'une part les locuteurs avec le plus de voyelles orales identifiées comme nasales, d'autre part les locuteurs avec le plus de voyelles orales et nasales confondues

classées comme nasales. La classification baisse à 80% sur les voyelles orales prononcées par le locuteur M23 et à 80% sur les voyelles nasales produites par le locuteur M11. Nous avons également analysé dans quels contextes le modèle prédit de façon correcte. Les voyelles nasales sont gérées de la même manière sur 15 contextes phonémiques sauf le contexte dorsal_dorsal. 24 % des voyelles nasales dans ce contexte ont été correctement identifiées contre environ 94 % pour le reste. Pour les voyelles orales, les contextes phonémiques n'ont pas d'impact.

	Modèle 1 (/a/ vs. /ã/)		Modèle 2 (/a/ vs. /a/ dans un contexte nasal)		Modèle 3 (/a/ vs. /a/ dans un contexte nasal vs. /ã/)		
	nasal	non nasal	nasal	non nasal	nasal	contexte nasal	non nasal
court	96%	96%	98%	82%	75%	81%	48%
moyen	92%	96%	75%	88%	92%	79%	57%
long	95%	96%	35%	89%	93%	63%	62%
total	94%	96%	71%	87%	88%	75%	56%
Kappa	0,9		0,58		0,59		

Tableau 2: matrice de confusion pour trois modèles

Le modèle 2 se montre moins performant que le premier pour la tâche de classification nasal-non nasal : 71% pour les voyelles /a/ en contexte oral et 87% pour les voyelles /a/ en contexte nasal avec une chute de valeur de kappa dans le modèle 2 ($\kappa = 0.58$). Les voyelles non nasales qui ont une durée moyenne (88%) ou longue (89%) ont été plus correctement identifiées. La classification des voyelles nasales courtes est très élevée alors qu'elle est très basse pour les longues. Quant aux locuteurs, 65% des voyelles orales du locuteur F6 ont été détectées correctement et 73% pour le locuteur M6. 38% de bonnes classifications ont été effectuées sur les voyelles nasales des locuteurs M17 et F8. Le modèle atteint 84% d'identification correcte lorsque les voyelles nasales sont précédées et suivies d'un phonème nasal. Le taux de bonnes classifications tombe à environ 70 quand un seul phonème nasal est présente dans leurs contextes quelle que soit la position.

Le modèle 3 montre une performance faible pour classer les voyelles /a/ en contexte non nasal. 56% des voyelles /a/ en contexte oral ont été identifiées comme orales tandis que 88% des voyelles /ã/ et 75% des voyelles /a/ en contexte nasal ont été correctement détectées ($\kappa = 0.59$). En ce qui concerne la durée sur la classification, les voyelles /ã/ et les voyelles /a/ en contexte oral ont tendance à être identifiées correctement lorsqu'elles ont une durée longue (93% et 62% respectivement). Les voyelles /a/ en contexte nasal ont été identifiées correctement lorsque ces dernières sont courtes, soit 81% des voyelles /a/ en contexte nasal. Le modèle se montre performant pour classer les /ã/ dans la plupart des contextes. Mais une pause devant fait baisser le taux de bonnes classification, 57 % des /ã/ en contexte

pause_pause et 67 % en contexte pause_labial sont concernées. Dans les contextes pause_pause et pause_dorsal, la classification des voyelles /a/ en contexte oral baisse à 70% et à 29 % respectivement. Une bonne classification a eu lieu pour les voyelles /a/ en contexte nasal notamment quand elles sont suivies et précédées d'une consonne nasale. 96% des voyelles nasal_nasal ont été correctement identifiées alors qu'environ 73% des voyelles avec un phonème nasal dans leurs contextes.

4 Discussion

Dans cette section, nous présentons également deux aspects en cours : des mesures physiologiques de nasalité pour valider notre méthode et une technique de visualisation qui permettra de comprendre ce qu'utilise le CNN dans sa prise de décision. Seul le modèle 1 a été retenu pour la suite car il a montré une meilleure performance de classification pour la tâche nasal-non nasal.

4.1 Mesure physiologique de nasalance

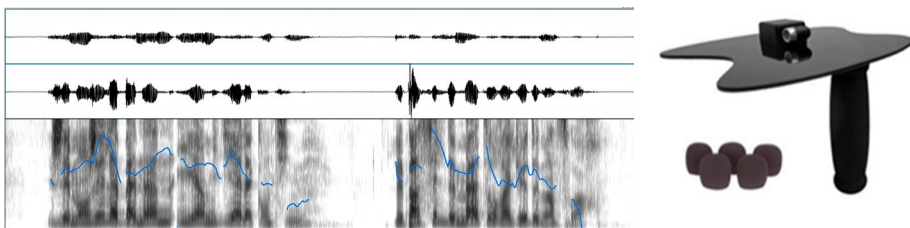


Figure 2: Exemple du signal enregistré avec le nanomètre et la photo du nanomètre

Locuteurs	F1	F2	M1	M2	M3
Nasalance /ã/	0.5733	0.5822	0.5622	0.5813	0.4872
Probabilités	0.99	1	0.99	0.99	0.97
Nasalance /a/	0.2096	0.1879	0.2984	0.2134	0.2218
Probabilités	0	0	0	0.2	0.16

Tableau 3: résultats de nasalance et probabilités obtenus pour le corpus de vérification

Les mesures physiologiques de nasalance sont utilisées pour comparer et valider les probabilités de nasalité obtenue avec notre premier modèle. Pour cela, nous avons eu recours à un nasomètre permettant d'obtenir des mesures directes et non altérées du degré de nasalité dit "nasalance". Deux phrases ("j'ai mangé la banane hier midi à la cantine", "le partiel de maths aura lieu demain soir") ont été prononcées par 5 locuteurs différents du corpus NCCFr¹. L'enregistrement au moyen d'un nasomètre donne en sortie un fichier wav en stéréo composé d'une sortie nasale et orale. La nasalance est obtenue avec la formule : $rms.ch1 / (rms.ch1 + rms.ch2)$ où *ch1* et *ch2* représentent respectivement

¹ Dans KIM (2021), une généralisation du modèle 1 a été effectuée sur des locuteurs non entraînés par le CNN, avec des scores entre 90% et 95%, soit seulement légèrement inférieurs à ceux présentés ici.

la sortie nasale et orale. Dans le but de comparer avec la probabilité de nasalité sur une séquence, la valeur de la nasalance est calculée sur une séquence qui correspond à la durée d'un phonème. Deux voyelles nasales sont présentes pour chaque phrase de vérification, la valeur moyenne de nasalance est considérée comme le seuil de détection de nasalité selon le locuteur pour comparer les valeurs de probabilités de nasalité moyennées selon les voyelles.

La nasalance confirme les probabilités du modèle 1. Nous observons, dans le tableau 3, que le seuil de détection est variable selon les locuteurs. La valeur de nasalance moyennée selon les voyelles / \tilde{a} / est la plus faible pour le locuteur M3 (soit 0.4872), et la plus élevée pour le locuteur F2 (soit 0.5822). Le locuteur ayant une valeur de probabilité de nasalité la plus faible est M3 (soit 0.97) et la plus élevée est F2 (soit 1). En outre, le seuil de nasalance pour les voyelles /a/ varie également en fonction des locuteurs. La valeur de nasalance la plus faible est égale à 0.1879 pour le locuteur F2. Une valeur de nasalance égale à 0.2984 est la plus élevée pour les voyelles /a/ prononcées par le locuteur M1. La valeur de probabilités de nasalité obtenues par le modèle 1 pour les voyelles /a/ est égale à 0 pour trois locuteurs F1, F2, M1. Une valeur de 0.2 pour le locuteur M2 est considérée comme la plus élevée parmi les probabilités de nasalité moyennées selon les locuteurs.

4.2 Techniques de visualisation

La technique appelée "Grad-cam" est abordée dans cette partie pour mieux comprendre comment les CNN ont procédé pour effectuer leur classification, notamment sur quelles zones du spectrogramme. Plus la couleur d'une zone est proche du jaune et loin du noir, plus la zone est considérée comme intéressante selon le modèle d'apprentissage pour catégoriser l'image.

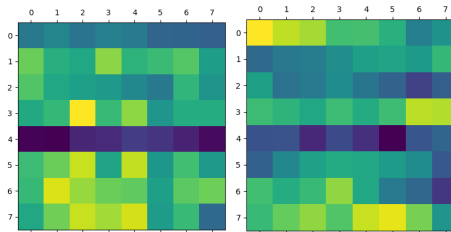


Figure 3: Grad-cam moyenné des /a/ (gauche) et / \tilde{a} / (droite) obtenu par le modèle 1

Lorsque les caractéristiques sont moyennées selon les voyelles, les indices que considère le modèle 1 comme importants se trouvent juste au-dessus et au-dessous de 4 000 Hz pour les voyelles non nasales /a/. Pour les voyelles nasales / \tilde{a} /, les caractéristiques intéressantes selon le modèle sont réparties en plusieurs zones, par exemple, à 8 000 Hz au début du phonème, à 1 000 Hz au milieu du phonème et à 5 000 Hz vers la fin du phonème.

4.3 Conclusion

Nous avons montré que les réseaux de neurones convolutifs ont permis une bonne détection acoustique de la nasalité. Il existe certains facteurs qui affectent la classification comme le contexte phonémique, la durée de la voyelle et le locuteur. Les /a/ suivies d'une pause ou d'une consonne dorsale ont été détectées comme nasales. Le mouvement d'abaissement du voile du palais est rapide chez certains locuteurs vers la fin d'une phrase par le relâchement (Amelot, 2004). Également, nous pouvons émettre l'hypothèse que pour les consonnes dorsales comme un /k/ ou / κ /, le velum est légèrement abaissé pour faciliter le contact avec la langue, ce qui donne une impression de nasalité

(KIM, 2021). La durée a un impact sur la classification de façon parfois surprenante. Ceci est probablement dû au fait que nos jeux de données ne contiennent pas les /a/ orales courtes car une restriction sur la durée a été mise en place sur elles dans le but d'équilibrer le nombre d'échantillons entre les catégories. Il est possible que les voyelles longues soient parfois des hésitations du locuteur avec un timbre qui change au cours de la voyelle. Suite à l'absence des voyelles orales entre 30 et 40 ms, le modèle 2 a considéré cette fourchette de durées comme caractéristique des voyelles nasales comme expliqué dans les résultats du t-SNE. Ceci permet d'expliquer pourquoi seuls 35% des voyelles nasales longues ont été identifiées correctement.

Nous avons par la suite vérifié les probabilités de nasalité obtenues par le modèle 1 sur un corpus de vérification et tenté de les valider par la comparaison avec la nasalance. L'intérêt de cette validation par nasalance sera de confirmer que les locuteurs avec plus de voyelles identifiées comme nasales ont une voix plus nasale que d'autres locuteurs. Nous avons constaté que les probabilités de nasalité pour les voyelles /a/ et /ã/ ont été validées par la nasalance. Le locuteur qui a des valeurs de nasalance les plus faibles obtient des probabilités de nasalité les plus basses (locuteur F2), et le locuteur qui a une nasalance la plus élevée obtient des probabilités de nasalité les plus élevées (locuteur M3). Mais l'écart entre les valeurs est faible pour la nasalance contrairement aux probabilités. La nasalance se fluctue entre 0.2 et 0.6 alors que les probabilités se situent entre 0 et 1. Nous supposons que les probabilités de nasalité ne représentent pas le degré de nasalité car ils semblent chercher à reconnaître une voyelle /ã/ ou /a/ par rapport à une autre plutôt que de chercher la nasalité dans les voyelles. C'est pourquoi les valeurs de probabilités pour toutes les voyelles non nasales sont très basses voire nulles tandis que la valeur de nasalance change selon les voyelles.

Nous avons également procédé à la technique de visualisation (Grad-cam). Pour les /a/, les zones où sont présents les formants des voyelles /a/ ont été considérées comme intéressantes par le modèle 1. Le résultat pour les voyelles nasales n'est pas encore interprétable, il doit être approfondi par des techniques de masquage et des tests de perception.

Bien que nos modèles montrent leur capacité à détecter la nasalité sur un ensemble de voyelles orales et nasales, cette tâche de classification aurait plutôt un but de discriminer les voyelles /a/ par rapport aux voyelles /ã/ et il ne serait pas question de détecter la nasalité dans un phonème. Dans la suite de travail, nous avons opté pour un modèle auto-supervisé "PASE+" permettant d'encoder et d'extraire des caractéristiques des fichiers audio avec des réseaux de neurones (Ravanelli M., et al., 2020) sur la totalité du signal, nous pouvons ensuite obtenir une courbe de probabilité de nasalité que nous pourrions corrélérer avec la mesure de nasalance. Nous tenterons également de normaliser la nasalance dans une optique d'effectuer des comparaisons entre les locuteurs. Des résultats encourageants pourront être présentés pendant la conférence.

Références

- AMELOT, A. (2004). Étude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français. *Université de la Sorbonne nouvelle-Paris III*.
- CARIGNAN, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics*, 63.
- CHEN, M. Y. (1997b). Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America*, 102(4), 2360-2370.
- DANG, J., HONDA, K., SUZUKI, H. (1994). Morphological and acoustical analysis of the nasal and the paranasal cavities. *The Journal of Acoustical Society of America*, 96(4), 2088-2100.
- GOLD, E., FRENCH, P. (2019). International practices in forensic speaker comparisons: second survey. *International Journal of Speech, Language & the Law*, 26(1).
- KAHN, J., AUDIBERT, N., BONASTRE, J.-F., ROSSATO, S. (2011). Inter and Intra-speaker Variability in French: An Analysis of Oral Vowels and Its Implication for Automatic Speaker Verification. *ICPhS*, 1002-1005.
- KIM L. (2021). Classification automatique de voyelles nasales pour une caractérisation de la qualité de voix des locuteurs : réseaux de neurones convolutifs, analyse acoustique et perception. *Université Sorbonne-Nouvelle - Paris III*.
- MAEDA, S. (1982). ACOUSTIC CUES OF VOWEL NASALIZATION: A SIMULATION STUDY. *Journal of Acoustical Society of America*, 7(S1), S102.
- NOLAN, F. (2005). Forensic Speaker Identification and the Phonetic. *A Figure of Speech: A Festschrift for John Laver*, 385.
- OHALA, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13(1-2), 155-161.
- RAVANELLI M., ZHONG J., PASCUAL S., SWIETOJANSKI P., MONTEIRO J., TRMAL J., BENGIO Y. (2020). Multi-task self-supervised learning for robust speech recognition. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- SERRURIER, A. (2006). Modélisation tridimensionnelle des organes de la parole à partir d'images IRM pour la production de nasales - Caractérisation articulatoire-acoustique des mouvements du voile du palais. PhD Thesis, INP Grenoble.
- STYLER, W. (2017). On the acoustical features of vowel nasality in English and French. *The Journal of Acoustical Society of America*, 142(4), 2469-2482.
- TORREIRA, F., ADDA-DECKER, M., ERNESTUS, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52 (3), 201.