Détection de la nasalité du locuteur à partir de réseaux de neurones convolutifs et validation par des données aérodynamiques

Lila Kim, Cedric Gendrot, Amélie Elmerich, Angelique Amelot, Shinji Maeda

Actes de CORIA-TALN 2023
Paris, France, 5 au 9 juin 2023

Détection de la nasalité du locuteur à partir de réseaux de neurones convolutifs et validation par des données aérodynamiques

Lila Kim[1] Cédric Gendrot[1] Amélie Elmerich[1] Angélique Amelot[1]
Shinji Maeda[1]
(1) Laboratoire de Phonétique et Phonologie, 4 rue des irlandais, 75005 Paris, France
lila.kim@sorbonne-nouvelle.fr, cedric.gendrot@sorbonne-nouvelle.fr,
amelie.elmerich@gmail.com, angelique.amelot@gmail.com, smaeda75@gmail.com

RÉSUMÉ

Ce travail se positionne dans le domaine de la recherche d'informations sur le locuteur, reconnue comme une des tâches inhérentes au traitement automatique de la parole. A partir d'un nouveau masque pneumotachographe acoustiquement transparent, nous avons enregistré simultanément des données aérodynamiques (débit d'air oral et nasal) et acoustiques pour 6 locuteurs masculins français, impliquant des consonnes et voyelles orales et nasales sur des logatomes. Un CNN entraîné sur d'autres corpus acoustiques en français a été testé sur les données recueillies à partir du masque pour la distinction de nasalité phonémique, avec une classification correcte de 88% en moyenne. Nous avons comparé ces résultats CNN avec les débit d'air nasal et oral captés par le masque afin de quantifier la nasalité présente par locuteur. Les résultats montrent une corrélation significative entre les erreurs produites par le CNN et des distinctions moins nettes de débit d'air entre nasales et orales.

ABSTRACT

Detection of speaker nasality from convolutional neural networks and validation with aerodynamic data

This work is positioned in the field of speaker information retrieval, which is recognized as one of the inherent tasks of automatic speech processing. Using a new acoustically transparent pneumotachograph mask, we simultaneously recorded aerodynamic (oral and nasal airflow) and acoustic data for 6 French male speakers, involving oral and nasal consonants and vowels on logatomes. A CNN trained on other French acoustic corpora was tested on the data collected from the mask for nasality distinction, with a correct classification of 88% on average. We compared these CNN results with the nasal and oral airflow captured by the mask to quantify the nasality present per speaker. The results show a significant correlation between the errors produced by the CNN and less clear distinctions in airflow between nasal and oral.

MOTS-CLÉS : CNN, RI, nasalité, caractérisation des locuteurs, aérodynamique.

KEYWORDS: CNN, nasality, speaker characterisation, nasality, aerodynamic.

Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 4 : articles déjà soumis ou acceptés en conférence internationale, pages 101–108.

1 INTRODUCTION

La nasalité est un trait distinctif dans environ un tiers des langues du monde (Basbøll, 1985). Les connaissances de base impliquent que le palais mou doit être suffisamment abaissé pour que l'orifice vélopharyngé soit ouvert et permette à l'air de passer par le nez. L'abaissement du voile du palais et le passage de l'air par le nez ont une incidence sur la composante acoustique du signal de parole, ce qui gêne généralement l'analyse acoustique pour les phonéticiens(Styler, 2017).

Il existe des variations temporelles et spatiales dans la réalisation de la caractéristique [nasale]. Elle varie en fonction du sexe et de l'anatomie du locuteur (Clarke, 1975; Amelot, 2004), de la stratégie du locuteur (Croft *et al.*, 1981; Skolnick *et al.*, 1973; Vaissière, 1988), de la langue (Clumeck, 1976), du style d'élocution (Basset *et al.*, 2001), du débit de parole (Bell-Berti & Krakow, 1991), du type de son de parole, du contexte phonétique et prosodique (Krakow, 1993), etc.

Plus précisément, l'ouverture de l'orifice vélopharyngé diffère d'un locuteur à l'autre et la morphologie des fosses nasales est très variable d'un individu à l'autre (Clarke, 1975; Amelot, 2004). Les voyelles et consonnes nasales sont importantes pour l'identification du locuteur car elles contiennent plus d'informations acoustiques relatives aux locuteurs que les autres sons (Ajili *et al.*, 2016; Kahn *et al.*, 2011).

Les réseaux neuronaux profonds ont récemment connu un développement important dans le domaine de la parole. Des études ont été menées dans le domaine clinique avec des réseaux neuronaux artificiels pour diagnostiquer des pathologies du langage, notamment l'hyper- ou l'hypo-nasalisation (Wang *et al.*, 2019; Mohammed *et al.*, 2020; Abderrazek *et al.*, 2022b). En effet, il a été démontré que les réseaux de neurones artificiels ont la capacité de se spécialiser sur des caractéristiques phonétiques telles que le lieu d'articulation ou le mode articulatoire (Abderrazek *et al.*, 2022b,a; Pellegrini & Mouysset, 2016).

L'objectif de la présente étude est d'évaluer la détection de la nasalité à partir de données acoustiques avec un CNN en la comparant avec les données aérodynamiques collectées lors de l'enregistrement acoustique. Nous prenons l'étiquette phonémique de la voyelle "nasale" ou "orale" comme référence et vérifions si la classification CNN est correcte à partir des données acoustiques. Dans un deuxième temps, le débit d'air nasal fourni validera la classification CNN ou nous aidera à comprendre les erreurs de classification. Enfin, nous étudierons si le niveau de nasalité de chaque locuteur peut être approché avec le CNN.

Ce travail se situe dans le domaine de la recherche d'information en TAL. En effet, la nasalité est une caractéristique inhérente du locuteur dûe à la morphologie peu malléable des cavités nasales ainsi qu'aux habitudes de production idiolectales. Dans le cadre de la vérification du locuteur, les informations propres à la voix du locuteur sont cruciales pour l'explicabilité du résultat.

2 MATÉRIAUX ET MÉTHODES

2.1 Corpus et acquisition de données

Pour cette étude, 6 locuteurs natifs masculins français (âge moyen : 36 ans) ont été enregistrés dans une pièce insonorisée. Les échantillons de parole sont constitués de séquences VCV, où

C=[p,b,t,d,v,s,z,m,n] et V=[i,a,y,u,e,ɑ̃,ɛ̃,ɔ̃]. Les séquences ont été insérées dans la phrase cadre, par exemple : " Non, tu n'as pas dit apa quatre fois, tu as dit aba et ada quatre fois ". Finalement, nous avons un total de 270 séquences avec C= 270 et V= 540. Les données aérodynamiques et acoustiques ont été enregistrées simultanément à l'aide d'un masque pneumotachographe.

Les avantages de ce masque sont les suivants : i) le débit d'air oral et nasal peut être enregistré séparément, ii) il est possible d'adapter la taille et la position de la plaque pour séparer le débit d'air nasal (NAF) du débit d'air oral (OAF) pour chaque locuteur, iii) il n'y a pas de distorsion acoustique.

Il peut y avoir de légères différences dans les mesures de débit en fonction du masque, de la position du capteur et de la taille et de la position de la plaque séparant le débit d'air nasal et le débit d'air oral. Par conséquent, un étalonnage doit être effectué séparément pour chaque masque (masque individuel pour chaque locuteur) : un étalonnage pour le compartiment buccal et un autre pour le compartiment nasal. L'étalonnage des 2 modules de capteurs de pression permet de convertir les valeurs de débit d'air dans l'unité physique (litres/s, voir Figure 1). Le masque offre une faible résistance, nécessaire pour mesurer le débit d'air sans affecter la propagation du son.

Les données acoustiques ont été capturées à l'aide d'un microphone (AKG C520 L). Tous les capteurs aérodynamiques et acoustiques sont reliés à une carte d'acquisition (DT9003). Les données acoustiques et aérodynamiques ont été enregistrées à une fréquence d'échantillonnage de 20kHz. Les données ont été segmentées manuellement dans Praat. Un script Python a été utilisé pour extraire automatiquement la moyenne de l'OAF et du NAF pour chaque voyelle (l/sec).

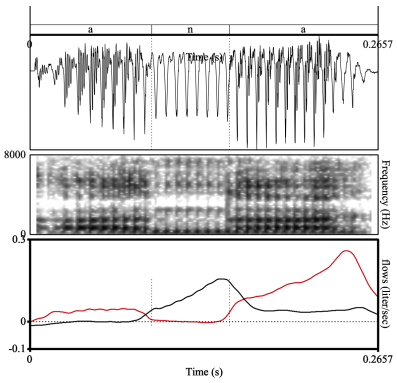

FIGURE 1 – Exemple d'enregistrements acoustiques et de débit d'air de [ana]. De haut en bas, (1) signal audio capturé avec un microphone, (2) spectrogramme, (3) débit d'air nasal (NAF en noir) et débit d'air oral (OAF en rouge)

2.2 Réseau de neurones convolutif

Pour la tâche de classification automatique acoustique nasal-non nasal, nous avons choisi de travailler avec des réseaux neuronaux convolutifs (CNN). Pour des raisons d'espace, nous détaillons uniquement les résultats sur 6 voyelles /a, e, o, ɑ̃, ɛ̃, ɔ̃/, donc 3 qualités de voyelles /a, e, o/ et leurs homologues nasales. Nous sommes conscients qu'il n'y a pas de correspondance articulatoire exacte entre nos 3 voyelles orales et nasales (Zerling, 1984) et nous avons décidé d'inclure les 6 voyelles dans le système de classification (au lieu de comparer par paires) afin de contourner cette asymétrie.

Le choix d'un CNN par rapport à d'autres réseaux neuronaux est lié à notre intention d'utiliser des spectrogrammes de voyelles, l'objectif final étant de localiser, à l'aide d'un algorithme de type

gradCam, l'endroit où se trouve l'information sur la nasalité.

L'ensemble de données d'entraînement est composé des productions de ces voyelles extraites de 3 corpus français avec différents types de discours : NCCFr (Torreira *et al.*, 2010), ESTER (Gravier *et al.*, 2004) et PTSVOX (Chanclu *et al.*, 2020). Dans tous ces corpus, des segmentations automatiques ont été fournies au niveau des phonèmes. Les voyelles ont été extraites aléatoirement à leurs frontières sous la forme d'un spectrogramme, sans aucune sélection du contexte prosodique, lexical ou phonémique. Pour les 2 premiers corpus, le nombre de voyelles de chaque type a été vérifié. 10 887 productions de chaque type ont été extraites de NCCFr et 9 186 d'ESTER. Pour PTSVOX, nous avons pris toutes les voyelles possibles en respectant la fréquence naturelle des phonèmes, ce qui a donné de meilleurs résultats.

	Entraînement & validation			Test
Source	NCCFr	ESTER	PTSVOX	Données enregistrées avec le masque
nasal	32,661	27,558	65,669	198
non nasal	32,661	27,558	135,119	198

TABLE 1 – Nombre de voyelles dans les jeux d'entraînement et de test selon les corpus

Dans la phase d'évaluation du modèle, nous avons sélectionné au hasard des productions de voyelles dans les données acoustiques présentées à la section 2.1. et extrait leurs spectrogrammes. Cet ensemble de test contient 66 productions de chaque type de voyelle (6 locuteurs * 11 occurrences), soit 198 voyelles pour chaque catégorie. Toutes ces images de spectrogrammes avec une bande de fréquence de 0 à 8000 Hz ont été réduites en 48x48 pixels et présentées comme entrée à notre réseau.

Pour la partie extraction des caractéristiques, le modèle est composé de deux blocs de couches de convolution et de pooling. Les couches de convolution ont été réalisées avec un noyau de taille 5x5 et ont donc produit respectivement 32 et 64 filtres. Après chaque couche de convolution, une couche de batch normalisation a été insérée avant d'appliquer une couche d'activation afin de permettre au modèle de se généraliser (Ioffe & Szegedy, 2015) sur différents types de corpus et de données. Les couches de max-pooling ont ensuite été utilisées pour réduire la taille des images avec une taille de pool de 2x2. Avec les caractéristiques extraites, 3 couches denses ont effectué la tâche de classification avec 1024 neurones. La fonction d'activation ReLU a été appliquée après chaque couche de batch normalisation et chaque couche dense. Enfin, une fonction d'activation softmax a été utilisée dans la dernière couche dense pour la classification nasale-orale. Au cours de l'apprentissage du modèle, nous avons tenté de minimiser les erreurs du modèle en appliquant Adam comme technique d'optimisation et categorical crossentropy comme fonction de perte pour mesurer la performance du modèle.

3 Résultats

3.1 Résultats du CNN

Pour une voyelle donnée, le classifieur renvoie une valeur entre 0 et 1, que nous appelons la probabilité de nasalité. Lorsqu'une voyelle est identifiée comme nasale par le modèle, la probabilité de nasalité attendue est supérieure à 0,5 et, inversement, une voyelle classée comme non nasale a une valeur proche de 0 (ou au moins inférieure à 0,5). Notre classifieur a pu identifier avec précision 95% des voyelles non nasales et 82% des voyelles nasales en atteignant une exactitude globale de 88% et

un score F1 de 88% (k = 0,77). Nous avons également testé un autre modèle incluant -en plus des voyelles- les consonnes /m, n, l, b, d, v/, et obtenu des résultats comparables : 89% d'exactitude globale en testant des voyelles et consonnes nasales et orales confondues.

La variabilité inter-locuteurs peut être observée sur la figure 2 (uniquement pour les voyelles pour une question de place). Certains locuteurs génèrent beaucoup d'erreurs de classification alors que pour d'autres, le modèle fait considérablement moins d'erreurs. Par exemple, le modèle fait le plus d'erreurs de classification pour les locuteurs MT01 et MT04. Sur le total des erreurs de classification, 37% des voyelles mal classées proviennent du locuteur MT04 (soit 17 erreurs sur 46). Le locuteur MT01 recueille 13 occurrences incorrectes (soit 28% du total des erreurs de classification) tandis que le locuteur MT03 n'a qu'une seule erreur. En outre, les erreurs sur les voyelles non nasales ne se produisent que pour les locuteurs MT01 et MT05. Pour les autres locuteurs, le modèle fonctionne correctement sur les voyelles non nasales, les erreurs ne se produisant que pour les nasales.

Nous observons que les erreurs de classification apparaissent principalement entre /a/ et /ɑ̃/. Plusieurs contextes phonétiques peuvent être considérés comme des facteurs de confusion. D'une part, lorsqu'il y a une pause dans le contexte gauche, les voyelles /a/ ont tendance à être classées comme nasales par notre modèle (5 erreurs sur 10 de /a/, soit 50%). En revanche, la présence d'une consonne labiale ou coronale devant les voyelles /ɑ̃/ peut influencer la décision de sa classe (respectivement 5 et 6 sur 14 erreurs de /ɑ̃/). Nous remarquons la même influence lorsqu'une pause est située après ces voyelles (5 des 14 erreurs de classification de /ɑ̃/). Ces 3 contextes pour les voyelles /ɑ̃/ apparaissent également dans les erreurs pour les autres voyelles nasales. Sur 36 classifications incorrectes de voyelles nasales, 12 erreurs sont causées par le contexte des consonnes labiales et coronales précédant les voyelles nasales, et 9 erreurs se produisent avec une pause en contexte gauche.

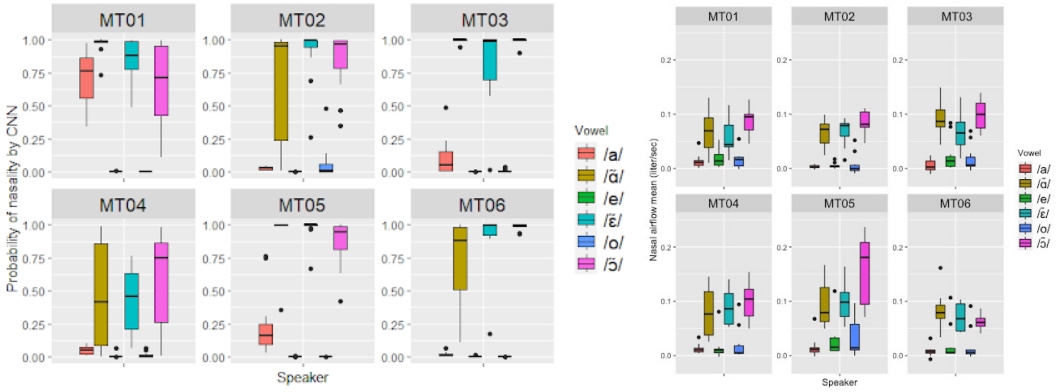

FIGURE 2 – Probabilités de nasalité obtenues par le modèle CNN pour chaque locuteur et chaque type de voyelle (à gauche), Moyenne du débit d'air nasal pour chaque locuteur et chaque type de voyelle (à droite)

3.2 Résultats aérodynamiques

Dans la figure 2 à droite, nous observons une plus grande quantité de débit d'air nasal pour les 3 voyelles nasales. Les résultats de l'ANOVA étaient statistiquement significatifs avec p<0,001 pour tous les locuteurs pour la distinction de la NAF entre les voyelles nasales et orales.

Nous constatons que chaque locuteur a un niveau minimum de débit d'air nasal qui diffère entre les voyelles. De plus, pour chaque paire de voyelles (/a/ vs. /ɑ̃/, /e/ vs. /ɛ̃/, et /o/ vs. /ɔ̃/), le débit d'air

nasal maximum pour les voyelles orales peut être plus important que le débit d'air nasal minimum pour les voyelles nasales. Ceci explique évidemment certaines erreurs de classification faites par le CNN, en particulier pour les locuteurs MT01, MT04 et MT05. Pour le locuteur MT02, le débit d'air nasal moyen pour /ɑ̃/ est de 0,065 l/s, et toutes les occurrences de /ɑ̃/, sauf une, ont été mal classées en dessous de ce seuil. En ce qui concerne le locuteur MT04, toutes les occurrences de /ɛ̃/, sauf deux, ont été mal classées en dessous d'un seuil de 0,059 l/s. De nombreux exemples de mauvaises classifications ont été trouvés selon ces critères. Globalement, un coefficient de corrélation de Pearson a révélé que la prédiction de la nasalité et de la non-nasalité est corrélée avec la mesure moyenne de la NAF (avec r = 0,66).

4 Discussion et conclusion

Le principal résultat de ce travail est la classification automatique correcte de la nasalité pour les voyelles jusqu'à 88% à partir d'un nouveau corpus acoustique, et 89% lorsqu'on inclue des consonnes dans l'entraînement et dans le test. Nous avons montré qu'il existe un lien significatif entre les probabilités CNN et les données aérodynamiques. Les erreurs de classification du CNN pour les locuteurs MT01 et MT04 ou pour la distinction entre /a/ et /ɑ̃/ sont corrélées à des différences plus faibles dans le débit d'air nasal. La même tendance des erreurs de classification de CNN entraîné avec des voyelles a également été observée pour le modèle incluant des voyelles et consonnes : les erreurs sont globalement les plus fréquentes chez les locuteurs MT01, MT04 et MT05.

Notre objectif était également d'établir une corrélation entre les probabilités CNN et le niveau moyen de débit d'air nasal par locuteur afin d'évaluer la nasalité globale par locuteur, et ce point doit encore être approfondi. Une première analyse a montré que les valeurs de probabilité données par le CNN n'étaient pas liées à la quantité de débit d'air nasal, mais que les erreurs de classification parviennent à donner de bonnes indications à ce sujet.

Reste à déterminer pourquoi les erreurs de classification portent plus fréquemment sur les voyelles orales pour le locuteur MT01 et sur les voyelles nasales pour le locuteur MT04. Pour le premier, les valeurs de débit d'air nasal sur les voyelles orales sont moyennes comparativement aux autres locuteurs alors que le débit d'air nasal est bas sur les voyelles nasales. Pour le deuxième, les valeurs de débit d'air nasal sur les voyelles nasales sont moyennes comparativement aux autres locuteurs alors que le débit d'air nasal est haut sur les voyelles orales. Ce résultat surprenant pourrait s'expliquer par d'autres paramètres mis en place par les locuteurs telles que la durée phonétique ou l'intensité, et qui montreraient une stratégie différente de la norme dans la distinction orale vs. nasale. Une étude perceptive de plusieurs items pour ces locuteurs permettrait de répondre à cette question.

Dans un futur proche, en incluant tous les phonèmes de la parole, nous travaillerons sur les fonctions d'activation afin de mieux relier les valeurs de probabilité avec le niveau de débit d'air nasal. L'analyse des zones spectrales utilisées par un CNN pourrait être un élément important de notre travail car la modélisation de la relation entre l'acoustique et l'articulation est encore problématique pour les nasales. Par exemple, des études articulatoires ont montré que le vélum du /a/ est plus bas que celui des autres voyelles, ce qui devrait avoir un impact sur le taux de classification (Durand, 1953). Dans l'ensemble, les implications de ces résultats devraient aider les phonéticiens dans leur analyse des voyelles nasales, et il est prévu de partager ce modèle et ce masque aérodynamique avec la communauté. Pour conclure, le projet de comparaison entre le test de perception et le CNN en anglais ainsi que l'implémentation du système Wav2vec sont en cours de développement.

Références

ABDERRAZEK S., FREDOUILLE C., GHIO A., LALAIN M., MEUNIER C. & WOISARD V. (2022a). Towards interpreting deep learning models to understand loss of speech intelligibility in speech disorders step 2 : contribution of the emergence of phonetic traits. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 7387–7391 : IEEE.

ABDERRAZEK S., FREDOUILLE C., GHIO A., LALAIN M., MEUNIER C. & WOISARD V. (2022b). Validation of the neuro-concept detector framework for the characterization of speech disorders : A comparative study including dysarthria and dysphonia. In *Interspeech 2022*.

AJILI M., BONASTRE J.-F., ROSSETTO S. & KAHN J. (2016). Inter-speaker variability in forensic voice comparison : a preliminary evaluation. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 2114–2118 : IEEE.

AMELOT A. (2004). *Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français*. Thèse de doctorat, Université de la Sorbonne nouvelle-Paris III.

BASBØLL H. (1985). Ian maddieson (1984). patterns of sounds. with a chapter contributed by sandra ferrari disner.(cambridge studies in speech science and communication) cambridge : Cambridge university press. pp. ix+ 422. *Phonology*, **2**(1), 343–353.

BASSET P., AMELOT A., VAISSIÈRE J. & ROUBEAU B. (2001). Nasal airflow in french spontaneous speech. *Journal of the international phonetic association*, **31**(1), 87–99.

BELL-BERTI F. & KRAKOW R. A. (1991). Anticipatory velar lowering : A coproduction account. *The Journal of the Acoustical Society of America*, **90**(1), 112–123.

CHANCLU A., GEORGETON L., FREDOUILLE C. & BONASTRE J.-F. (2020). Ptsvox : une base de données pour la comparaison de voix dans le cadre judiciaire. In *6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1 : Journées d'Études sur la Parole*, p. 73–81 : ATALA ; AFCP.

CLARKE W. M. (1975). The measurement of the oral and nasal sound pressure levels of speech. *Journal of Phonetics*, **3**(4), 257–262.

CLUMECK H. (1976). Patterns of soft palate movements in six languages. *Journal of phonetics*, **4**(4), 337–351.

CROFT C. B., SHPRINTZEN R. J. & RAKOFF S. J. (1981). Patterns of velopharyngeal valving in normal and cleft palate subjects : A multi-view videofluoroscopic and nasendoscopic study. *The Laryngoscope*, **91**(2), 265–271.

DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

DURAND M. (1953). De la formation des voyelles nasales. *Studia Linguistica*, **7**(1-2), 33–53.

ELMERICH A., AMELOT A., MAEDA S., LAPRIE Y., PAPON J. F. & CREVIER-BUCHMAN L. (2020). F1 and f2 measurements for french oral vowel with a new pneumotachograph mask. In *ISSP 2020-12th International Seminar on Speech Production*.

GRAVIER G., BONASTRE J.-F., GEOFFROIS E., GALLIANO S., MCTAIT K. & CHOUKRI K. (2004). The ester evaluation campaign for the rich transcription of french broadcast news. In *LREC*.

HONDA K. & MAEDA S. (2008). Glottal-opening and airflow pattern during production of voiceless fricatives : a new non-invasive instrumentation. *The Journal of the Acoustical Society of America*, **123**(5), 3738–3738.

IOFFE S. & SZEGEDY C. (2015). Batch normalization : Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, p. 448–456 : pmlr.

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. In *ICPhS*, p. 1002–1005.

KRAKOW R. A. (1993). Nonsegmental influences on velum movement patterns : Syllables, sentences, stress, and speaking rate. In *Nasals, nasalization, and the velum*, p. 87–116. Elsevier.

MOHAMMED M. A., ABDULKAREEM K. H., MOSTAFA S. A., KHANAPI ABD GHANI M., MAASHI M. S., GARCIA-ZAPIRAIN B., OLEAGORDIA I., ALHAKAMI H. & AL-DHIEF F. T. (2020). Voice pathology detection and classification using convolutional neural network model. *Applied Sciences*, **10**(11), 3723.

PELLEGRINI T. & MOUYSSET S. (2016). Inferring phonemic classes from cnn activation maps using clustering techniques. In *Annual conference Interspeech (INTERSPEECH 2016)*, p. pp–1290.

SKOLNICK M. L., MCCALL G. N. & BARNES M. (1973). The sphincteric mechanism of velopharyngeal closure. *The Cleft Palate Journal*, **10**(3), 286–305.

STYLER W. (2017). On the acoustical features of vowel nasality in english and french. *The Journal of the Acoustical Society of America*, **142**(4), 2469–2482.

TEAM R. D. C. (2009). A language and environment for statistical computing. *http ://www. R-project. org*.

TORREIRA F., ADDA-DECKER M. & ERNESTUS M. (2010). The nijmegen corpus of casual french. *Speech Communication*, **52**(3), 201–212.

VAISSIÈRE J. (1988). Prediction of velum movement from phonological specifications. *Phonetica*, **45**(2-4), 122–139.

WANG X., TANG M., YANG S., YIN H., HUANG H. & HE L. (2019). Automatic hypernasality detection in cleft palate speech using cnn. *Circuits, Systems, and Signal Processing*, **38**, 3521–3547.

ZERLING J.-P. (1984). Phénomènes de nasalité et de nasalisation vocalique : Étude cinéradiographique pour deux locuteurs. *Travaux de l'Institut de Phonétique de Strasbourg Strasbourg*, (16), 241–266.