Vers un nouveau modèle acoustique des nasales basé sur l'enregistrement bouche - nez séparé

Gang Feng et Cyril Kotenkoff

Institut de la Communication Parlée
Université Stendhal – INPG – CNRS, ICP Campus, BP 25, 38040 Grenoble Cedex 9, France
Tél.: +33 (0)4 76 82 43 38 - Fax: +33 (0)4 76 82 43 35
Mél: feng@icp.inpg.fr - http://www.icp.fr

ABSTRACT

An experimental setting is described in this paper which allows to separately record the speech signals emitted from the mouth and the nostrils. The speaker is asked to pronounce a series of transitions from an oral configuration to the pharyngonasal configuration, in order to test acoustic models of vowel nasalization. Analysis of the recorded signals seems to show that the main effect of the velum lowering in the oral tract would be due to the modification of its shape, while the connexion of the nasal tract could be neglected when considering the mouth output. A new model is proposed based on this hypothesis. The simulation results seem to match better with the observed spectrum than the classical model.

1. INTRODUCTION

Tout le monde sait à quel point il est difficile d'analyser un son nasal, que ce soit une voyelle nasale, une voyelle nasalisée ou une consonne nasale. Si le geste articulatoire principal pendant la production de ces sons est relativement simple (l'abaissement du vélum permettant la connexion du conduit nasal au conduit oral), les effets acoustiques de ce geste restent néanmoins très complexes. Outre la grande complexité anatomique du conduit nasal, le fait d'avoir deux conduits (oral et nasal) en couplage complique beaucoup le spectre d'un son nasal : ajout des pôles supplémentaires et apparition des zéros. Ces pôles et zéros dépendent de la configuration orale et surtout de la position du vélum, mais également de la condition de prise de son car un son nasal est émis par les narines et la bouche. A cette complexité acoustique s'ajoute la difficulté rencontrée lors de la caractérisation de ces sons. En effet, bien que le spectre d'une voyelle orale ne soit jamais aussi simple que la fonction de transfert simulée par un modèle, le fait de savoir que ce sont les formants (en particulier les 2 ou 3 premiers) qui caractérisent efficacement une voyelle orale nous permet de comparer le spectre réel avec la modélisation sans trop de difficultés. Malheureusement, un tel filtre de caractérisation n'existe plus pour les nasales pour lesquelles il est extrêmement difficile de confronter un spectre réel avec sa modélisation.

Les études concernant la modélisation acoustique et la caractérisation des nasales sont très abondantes dans la littérature. Citons ici quelques références majeures : Fant [2], Fujimura & Lindqvist [4], Maeda [6], Hawkins & Stevens [5], Dang et *al*. [1]. Dans une étude précédente (Feng & Castelli [3]), la nasalisation d'une voyelle est considérée comme étant une transition qui part d'une configuration orale et qui évolue vers une configuration nasopharyngale proche de la consonne [ŋ]. Le conduit nasopharyngal peut être modélisé dans un premier temps par un conduit unique et possède deux premiers pics spectraux qui se situent respectivement à 250-300 Hz et aux alentours de 1000 Hz. Ainsi, nous considérons que c'est la présence de ces deux pics spectraux dans le spectre d'une nasale qui constitue le corrélat acoustique le plus important de la nasalisation, en accord avec la proposition de Maeda [6].

Nous pouvons ainsi simuler toute sorte de transitions du type "configuration orale vers configuration naso-pharyngale" et considérer une voyelle nasale comme un état intermédiaire correspondant à un abaissement plus ou moins grand du vélum. Cependant, il reste toujours très délicat de comparer le spectre réel d'une voyelle nasale avec la simulation, en particulier en raison du manque d'information concernant la position du vélum. Pour s'affranchir de ce problème, nous pensons qu'il est plus judicieux de comparer la simulation non pas avec une voyelle statique mais avec une transition complète prononcée.

Nous disposons d'un locuteur qui, après des entraînements est capable de prononcer des transitions en faisant varier presque uniquement la position du vélum, ce qui offre une très bonne condition pour tester les différents modèles. Il lui est toutefois difficile de prononcer une transition complète "orale – naso-pharyngale". En revanche, il peut partir d'une des voyelles nasales françaises et évoluer ensuite soit vers la configuration nasopharyngale correspondante en abaissant le vélum, soit vers la configuration orale correspondante en levant le vélum. Lors d'une étude articulographique des mêmes articulations [7], nous avons vérifié que dans ces transitions la hauteur de la mâchoire ne varie pas de plus de 0,03 cm, et que la variation maximale de la position de la langue ne dépasse pas 0,2 cm. Les premiers signaux enregistrés du locuteur ont montré que dans les basses fréquences (< 1500 Hz), l'évolution des différents formants semble concorder avec

celle du modèle. En revanche, pour les fréquences plus élevées, les spectres diffèrent sensiblement des simulations. Prenons comme exemple la transition [ɑ̃] – [ɔ]. Notons qu'ici [ɔ] représente la configuration orale correspondante de [ɑ̃] avec le vélum levé et non la vraie voyelle [ɔ]. On peut constater en particulier un rapprochement très net des deux formants se situant aux alentours de 3000 Hz lors de l'évolution [ɔ] vers [ɑ̃], alors que la simulation ne reproduit pas vraiment cette tendance.

Cette expérience montre qu'il est difficile d'évaluer les modèles et de les améliorer tant que nous n'avons pas accès, de manière séparée, aux signaux issus des narines et de la bouche. Nous avons donc décidé de réaliser un dispositif permettant d'enregistrer séparément ces deux signaux. Dans la section 2, nous décrivons la réalisation de ce dispositif et ses performances, et nous analysons les enregistrements obtenus avec le dispositif dans la section 3 et proposons un nouveau modèle permettant de rendre mieux compte des observations.

2. DISPOSITIF D'ENREGISTREMENT

La grande difficulté lors d'un enregistrement séparé réside dans le fait qu'il n'est pas facile d'isoler les deux signaux issus des narines et de la bouche. Une solution consiste à fabriquer une planche isolante de taille suffisamment grande permettant de couper une chambre sourde en deux parties (cf. Schnell & Lacroix [8] par ex.). On peut aussi construire une boite de grande taille avec une planche séparatrice à l'intérieur (Suzuki et *al*. [9]). Nous avons adopté cette deuxième solution.

Pour des raisons pratiques, les dimensions de notre boite sont 60x60x80cm. Une planche horizontale sépare la boite en deux enceintes de 40 cm de hauteur. Avec une telle taille, les différentes résonances des enceintes se trouvent pleinement dans les fréquences où il y a les formants de la parole. Il est donc indispensable de les atténuer. Nous avons choisi de recouvrir toutes les parois par de la mousse alvéolée spécialement conçue pour l'isolation phonique. Pour résoudre le problème de fuites par l'extérieur de la boite, nous avons taillé un trou sur la face avant de la boite permettant au locuteur d'insérer une partie de son visage (du haut du nez jusqu'au menton) dans la boite. Un joint en caoutchouc placé autour du trou permet au locuteur de coller son visage sur la boite, assurant une bonne étanchéité. Deux microphones identiques sont placés, avec des suspensions nécessaires, à l'intérieur de chaque enceintes, à environ 10 cm des narines et de la bouche du locuteur.

Nous avons effectué une série de mesures afin de nous assurer que la boite convient à un enregistrement séparé du locuteur. On exige en particulier que la fonction de transfert soit suffisamment plate sans résonances marquées et que l'atténuation entre les deux enceintes soit suffisamment grande pour que les signaux enregistrés soient exploitables.

Les mesures ont été réalisées à l'aide d'un petit haut-parleur collé contre la face avant de la boite, avec un trou spécialement taillé à cet effet. Ce trou concerne l'une des deux enceintes, l'autre étant entièrement fermée. Le haut-parleur est excité par un bruit rose et les fonctions de transfert sont estimées à partir des signaux enregistrés en utilisant une technique d'analyse spectrale (le périodogramme moyenné).

Les premières mesures effectuées avec la boite sans mousse ont montré la présence de nombreux pics correspondant aux différentes résonances d'une enceinte. Cette dernière ayant une forme géométrique très simple, nous avons pu calculer sa réponse théorique et vérifier que celle-ci concorde parfaitement avec les mesures. La figure 1 montre la caractéristique fréquentielle d'une enceinte avec de la mousse d'isolation phonique. On peut constater que, bien que la courbe ne soit pas tout à fait plate, les résonances de l'enceinte sont totalement atténuées et ne constituent plus une gêne pour notre enregistrement.

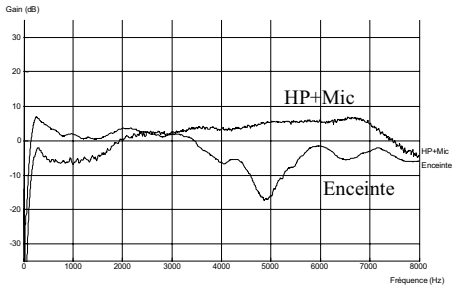

Figure 1 : Réponse fréquentielle d'une enceinte, ainsi que la caractéristique du couple "HP - microphone".

Rappelons par ailleurs que le petit haut-parleur utilisé n'a pas un spectre très plat. Nous avons donc mesuré séparément ce spectre en n'utilisant que le couple haut-parleur – microphone (voir la courbe de la figure 1 "HP+Mic"). En retranchant ensuite ce spectre au spectre du signal issu de l'enceinte, on obtient la fonction de transfert de celle-ci (un filtre passe-bas très doux globalement) qui est par la suite utilisée pour corriger les spectres des signaux enregistrés.

En ce qui concerne le spectre du signal capté par le microphone se situant dans l'enceinte où il n'y a pas d'excitation, les mesures ont montré que, pour toutes les fréquences, ce signal est au moins 20 dB en dessous du signal issu de l'enceinte où il y a le haut-parleur. Nous considérons que cette atténuation est suffisante pour notre étude.

3. ANALYSE DES RÉSULTATS

Une fois notre dispositif validé, nous avons procédé à l'enregistrement du locuteur. Dans un premier temps, le corpus contient des transitions qui partent des voyelles nasales françaises pour évoluer vers des configurations orale et nasopharyngale correspondantes. Nous nous intéressons ici uniquement aux transitions [ɑ̃] – [ɔ] et [ɑ̃]

– [ŋ]. Notons que [ɔ] désigne toujours la partie orale de la voyelle nasale [ɑ̃] avec le vélum levé, et [ŋ] la configuration nasopharyngale correspondante. Le locuteur prononce une dizaine de fois chaque transition, ainsi qu'un certain nombre de fois [ɔ] et [ŋ] sans transition, afin de vérifier si ces deux positions extrêmes sont bien réalisées dans les transitions.

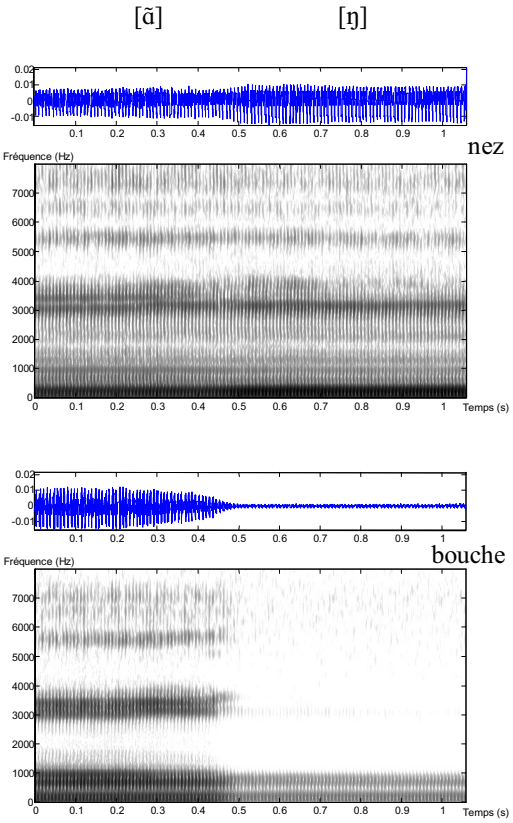

Figure 2 : Sonagrammes de la transition [ɑ̃] – [ŋ]. En haut : sortie nez ; en bas : sortie bouche.

Nous montrons en figures 2 et 3 les sonagrammes des signaux enregistrés pour les deux transitions étudiées. Nous pouvons d'abord formuler quelques remarques générales concernant ces résultats. Les grands contrastes entre les sorties nez et bouche nous laissent penser que l'isolation de la boite est bonne. Si théoriquement les deux sorties possèdent les mêmes formants, leurs amplitudes relatives varient considérablement. Ainsi certains formants sont visibles dans une sortie et non dans l'autre. Dans la partie [ŋ] de la transition [ɑ̃] – [ŋ] (figure 2), on peut constater que même si le vélum est abaissé, la bouche rayonne encore le premier formant du conduit nasopharyngal (environ 250 Hz), ainsi qu'une résonance propre à la cavité buccale (environ 800 Hz). Pour la partie [ɔ] de [ɑ̃] – [ɔ] (figure 3), l'amplitude du signal issu du nez reste grande. Il semble que cela est dû au fait que le locuteur a du mal à remonter complètement le vélum dans cette transition.

Nous avons comparé ces résultats avec ceux obtenus en simulation avec les sorties nez / bouche séparées (figure 4). Le rapprochement très net des deux formants aux alentours de 3000 Hz lors de l'évolution [ɔ] vers [ɑ̃] reste difficile à trouver dans les fonctions de transfert simulées. Par ailleurs, l'évolution vers les basses fréquences du F3 dans la simulation est totalement absente dans les spectres réels.

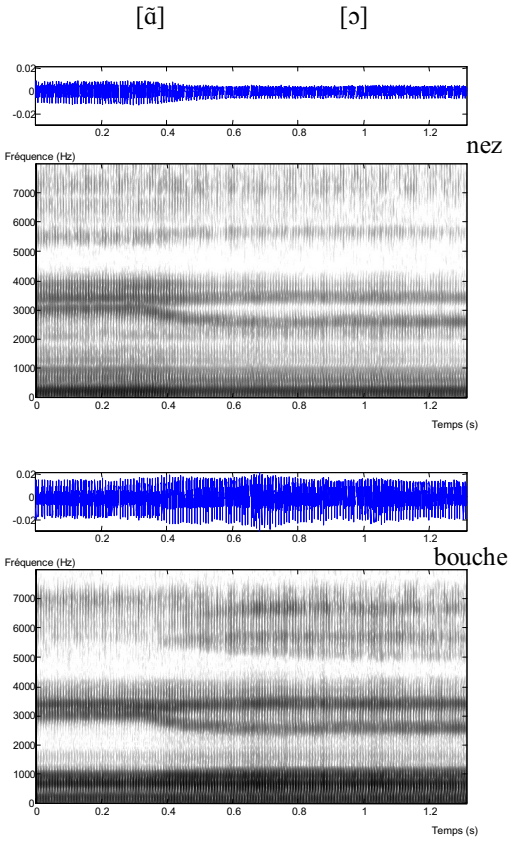

Figure 3 : Sonagrammes de la transition [ɑ̃] – [ɔ]. En haut : sortie nez ; en bas : sortie bouche.

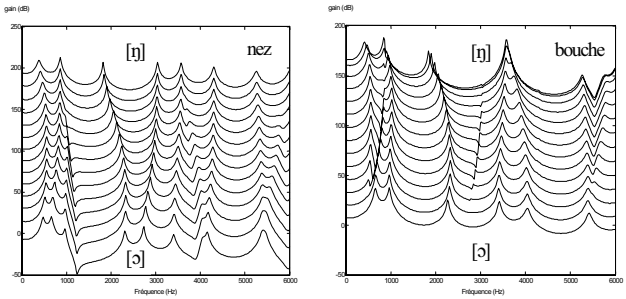

Figure 4 : Simulation de la nasalisation de [ɔ] : modèle de couplage. A gauche : sortie nez ; à droite : sortie bouche.

4. NOUVEAU MODÈLE PROPOSÉ

Nous admettons bien que la grande simplicité du modèle utilisé ne permette pas de rendre compte de tous les phénomènes complexes lors de la production d'une

nasale. C'est pourquoi nous avons choisi d'évaluer les modèles en utilisant les transitions complètes et avec les deux sorties séparées. Un modèle valable doit, dans ces conditions, au moins rendre compte des grandes tendances observés dans les spectres réels.

Il nous a semblé que le rapprochement très net des deux formants observé (figure 3) n'est pas forcément le fruit d'un couplage de deux conduits, mais simplement (ou principalement) l'effet de l'abaissement du vélum dans le conduit oral seul. Il est en effet aisé de prédire la variation de F3 et F4 dans une configuration proche de [ɔ] lorsque le conduit oral subit un "pincement" au niveau du vélum.

Si cette hypothèse est vraie, cela revient à dire que l'abaissement du vélum a deux effets distincts : 1) modifier la fonction de transfert du conduit oral mais cette modification serait essentiellement due au rétrécissement du conduit vocal au niveau du vélum et non (ou peu) à la connexion du conduit nasal ; 2) permettre la connexion du conduit nasal et par conséquent le rayonnement au niveau des narines, ce qui a pour effet d'émettre une "signature" du nez, responsable de la perception d'un son nasal.

Notons que cette vision revient à supposer que l'impédance d'entrée du conduit nasal est beaucoup plus grande que celle du conduit orale. En attendant les futures mesures de l'impédance d'entrée du conduit nasal, nous avons effectué une simulation pour tester cette hypothèse. On traite donc le problème en deux temps : d'abord le conduit vocal seul avec le vélum abaissé et ensuite la connexion du conduit nasal en prenant comme excitation le débit acoustique au niveau du vélum. La figure 5 montre le résultat de cette simulation. Notons que les deux courbes du haut correspondent à une position du vélum bien abaissé mais le conduit oral n'est pas fermé (l'aire minimale = 0,2 cm^2). Nous pouvons constater que l'évolution des formants F3 et F4 (sortie bouche), qui résulte du rétrécissement du conduit vocal, rend mieux compte du spectre observé. La sortie nez est plus complexe mais elle semble plus proche de l'observation par rapport au modèle de couplage. En effet, la concentration d'énergie dans les basses fréquences, ainsi qu'aux alentours de 3000 Hz est cohérent avec les sonagrammes (figure 2 et 3), ce qui n'est pas le cas du modèle de couplage.

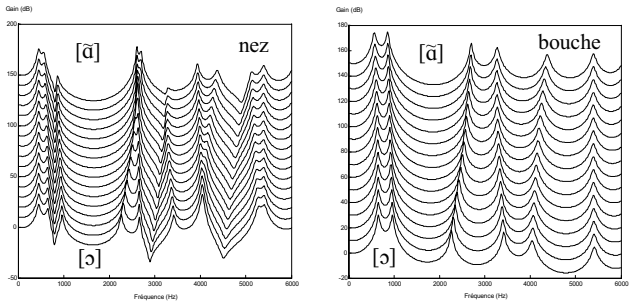

Figure 5 : Simulation de la nasalisation de [ɔ] : modèle proposé. A gauche : sortie nez ; à droite : sortie bouche.

5. DISCUSSION - CONCLUSION

Notre expérience montre que l'enregistrement séparé constitue un moyen très important même incontournable pour élaborer et améliorer les modèles acoustiques des nasales. Nos premiers enregistrements semblent montrer que, au moins pour les fréquences supérieures à 1000 Hz, l'effet principal de l'abaissement du vélum pour le conduit oral serait plutôt un rétrécissement de celui-ci qu'une connexion du conduit nasal qui implique un couplage acoustique complexe. Notre simulation basée sur cette hypothèse semble rendre mieux compte des observations par rapport au modèle classique de couplage. Et surtout nous proposons une nouvelle façon de considérer les effets de l'abaissement du vélum, qui éviterait le problème complexe du couplage.

Dans l'avenir, seules les mesures précises des impédances des deux conduits permettraient d'améliorer les modèles. Il est aussi probable qu'une combinaison du modèle proposé avec le modèle de couplage, peut-être en fonction des zones de fréquences considérées, permettrait de se rapprocher de la réalité.

Remerciements : Nous remercions Pierre Badin, notre locuteur préféré ; Alain Arnal pour la fabrication de la boite et Xavier Pelorson pour les conseils en acoustique.

BIBLIOGRAPHIE

[1] J. Dang, K. Honda and H. Suzuki. Morphological and acoustical analysis of the nasal and the paranasal cavities. *J. Acoust. Soc. Am*. 96, 2088-2100, 1994.

[2] G. Fant. *Acoustic theory of speech production*. Mouton, The Hague, 1960.

[3] G. Feng and E. Castelli. Some acoustic features of nasal and nasalizes vowels : A target for vowel nasalization. *J. Acoust. Soc. Am*. 99, 3694-3706, 1996.

[4] O. Fujimura and J. Lindqvist. Sweep-tone measurements of vocal tract characteristics. *J. Acoust. Soc. Am*. 19, 541-558, 1971.

[5] S. Hawkins and K.N. Stevens. Acoustic and perceptual correlates of the non-nasal – nasal distinction for vowels. *J. Acoust. Soc. Am*. 77, 1560-1575, 1985.

[6] S. Maeda. Une paire de pics spectraux comme corrélat acoustique de la nasalisation des voyelles. *13ème J.E.P.*, 223-224, 1984.

[7] S. Rossato. Du son au geste, inversion de la parole : le cas des voyelles nasales. Thèse de doctorat, INPG, 2000.

[8] K. Schnell and A. Lacroix. Generation of nasalized speech sounds based on branched tube models obtained from separate mouth and nose outputs. *Proc. ICASSP*, 2003.

[9] H. Suzuki, T. Nakai, J. Dang and C.X. Lu. Speech production model involving subglottal structure and oral-nasal coupling through closed velum. *Proc. ICSLP*, 1, 437-440, 1990.