

PRODUCTION DES VOYELLES ET MODELE A REGIONS DISTINCTIVES

René Carré^o and Mohamed Mrayati^{oo}

^o)Département Signal, Unité de Recherche Associée au CNRS
ENST, 46 rue Barrault, 75634 Paris cedex 13

^{oo})Scientific Studies and Research Center, POB 4470, Damascus.

RESUME

On rappelle les propriétés acoustiques optimales d'un tube acoustique divisé en régions distinctives. Pour maîtriser la commande d'un, puis de deux puis de trois formants, on définit deux puis quatre puis huit régions distinctives. On exploite un tel modèle optimal pour produire les sons acoustiquement les plus différents. La production des voyelles est ensuite étudiée à la lumière de ce modèle théorique.

1. LE MODELE EN REGIONS DISTINCTIVES

1.1. Tube fermé à l'une des extrémités, ouvert à l'autre.

On a montré [1], [2], [6], [7] que le tube acoustique fermé à l'une de ses extrémités et ouvert à l'autre pouvait être divisé en régions distinctives permettant des modulations optimales des fréquences de résonance de ce tube. La longueur de ces régions est fixe et correspond à un pourcentage de la longueur totale effective du tube. Les commandes des sections des régions de cette modélisation sont donc transversales. Par modulations optimales, on entend : plages maximales possibles de variations de ces fréquences de résonance, et variations maximales des fréquences pour un déplacement minimal des parois latérales des régions autour de la position neutre. Par ailleurs, on peut souhaiter commander soit le premier formant (le modèle doit être constitué de seulement deux régions R1 et R2), soit les deux premiers formants (on a alors quatre régions, R1, ..., R4), soit des trois premiers formants (on a alors huit régions) (fig. 1), etc.

Le comportement de ce modèle est pseudo-orthogonal c'est-à-dire que, autour de la position neutre, les commandes des sections des différentes régions correspondent à toutes les

combinaisons de sens de variations des formants pris en considération (fig. 1). Ce modèle permet donc la maîtrise des sens de variations des formants.

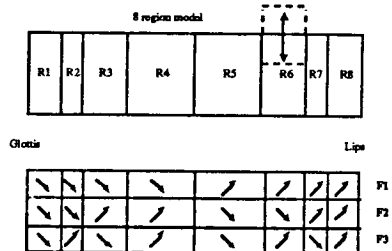


Figure 1. Modèle constitué de huit régions. On note la commande de type transversale et les sens de variations des trois premiers formants pour un accroissement de la section de l'une des régions autour de sa position neutre.

On peut aussi noter que la position de la commande demande de plus en plus de précision selon le nombre de formants considéré : pour le premier formant, les longueurs des deux régions sont de 1/2 si l est la longueur effective du tube ; elles sont de 1/6, 1/3, 1/3 et 1/6 pour prendre en compte les deux premiers formants, etc.

Par ailleurs, le comportement antisymétrique mis en évidence figure 1 peut être exploité pour multiplier les effets de modulation : il suffit de mettre en oeuvre une commande synergétique, c'est-à-dire, par exemple, qu'une commande de réduction de section d'une région de la partie avant du tube va être associée à une augmentation de la section de la région correspondante de la partie arrière. En mettant en oeuvre cette propriété, le nombre de commandes des régions du modèle est divisé par deux. Dans le cas

d'un modèle à deux régions permettant le contrôle du premier formant, un seul paramètre de commande suffit. La figure 2 montre l'évolution du premier formant (et des 2ème et 3ème formants). Dans ce cas, le produit de la section (SR1) de la région arrière (R1) par la section (SR2) de la région avant (R2) est constant. La longueur effective du tube est de 19cm et les différents types de pertes sont pris en compte dans la simulation. On note trois zones principales : deux zones de stabilité et une zone de variation rapide (autour de la position neutre qui est ici de 4cm²). Les zones de stabilité peuvent, en particulier, être utilisées pour coder l'information acoustique. Le passage rapide d'un état à un autre permet un débit maximal d'informations par seconde. La prise en compte du 2ème formant, réalisée dans un modèle à 4 régions, multiplie le nombre de niveaux de codage. Dans ce type de codage à modulation des fréquences de résonance du tube acoustique, on ne prend pas en compte les types de sources d'excitation qui multiplieraient encore les possibilités de produire des informations acoustiquement différentes.

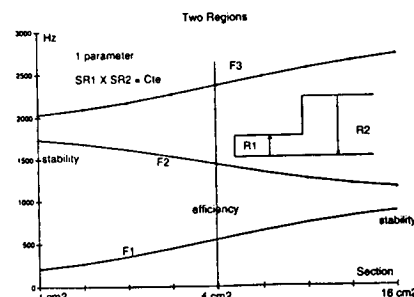


Figure 2. Evolution des trois premiers formants avec un modèle constitué de deux régions. Les sections varient de 0.5 à 32cm² par pas logarithmiques. La position neutre est de 4cm².

1.2. Tube fermé aux deux extrémités.

Selon les principes énoncés dans [7], on peut aussi définir des régions pour un tube acoustique fermé aux deux extrémités. Le premier formant corres-

pond alors à la fréquence de vibration des parois de ce tube, et le deuxième formant peut être contrôlé par un modèle à trois régions. Le comportement du modèle est symétrique et la constriction s'applique au milieu du tube. Ce modèle permet d'atteindre les valeurs les plus basses du couple F1, F2. Mais l'aspect pseudo-orthogonal mis en évidence dans le cas précédent disparaît ici.

Avec ces deux modèles de référence, on peut produire efficacement les sons les plus différents en ce qui concerne les fréquences de formants, tout en faisant appel à un minimum de paramètres de commande. On peut aussi jouer sur le nombre de références retenues. Il faut rappeler que cette modélisation implique intrinsèquement des commandes transversales à des endroits quantifiés du tube acoustique (les endroits les plus efficaces pour la modulation). La question qui se pose maintenant est de savoir si, pour parler, l'homme exploite les propriétés importantes du tube acoustique mises en évidence dans le modèle à régions.

2. LE MODELE A REGIONS DISTINCTIVES ET L'APPAREIL VOCAL.

Tout d'abord, notons que l'appareil vocal est bien adapté pour exploiter les caractéristiques d'un modèle à régions distinctives. C'est un conduit fermé à l'une des extrémités (coté glotte) et ouvert de l'autre (coté lèvres). Dans le cas d'une référence à huit régions, R8 serait contrôlé par les lèvres, R7 par la pointe de la langue, R3, R4, R5, R6 par le corps de la langue laquelle réalise par ailleurs la mise en oeuvre de la synergie grâce à son volume constant. En revanche, R1 et R2 sont relativement constants (larynx). Le modèle fermé-fermé peut être réalisé par une forte labialisation et par une constriction centrale obtenue par le corps de la langue.

En dynamique, le problème de la commande de l'appareil vocal chez l'homme se pose : est-elle transversale et réalise-t-elle des contractions à des endroits spécifiques lors de la production de la parole ? La bonne reproduction de la transition /ai/ par le modèle [2] par exemple est un élément de réponse positive. Un déplacement longitudinal de la constriction donnerait des résultats tout à fait différents.

3. LE MODELE A REGIONS DISTINCTIVES ET LA PRODUCTION DES VOYELLES.

Les hypothèses formulées dans ce paragraphe demandent vérifications et études approfondies. En procédant du plus simple au plus compliqué, en privilégiant une voyelle non labialisée (plus intense) à une voyelle labialisée, en privilégiant la distinction par le premier formant (la position de la commande demande la précision la plus faible) et en tenant compte du fait que l'avant de la langue est plus souple que l'arrière, on peut expliquer le contenu de systèmes vocaliques. Les voyelles extrêmes /a/ et /ɪ/ sont les plus simples à produire dans le cas de la référence fermée-ouverte. Ensuite, la mise en oeuvre de la référence fermée-fermée (avec constriction centrale) permet la production de la voyelle /u/ (F1

et F2 les plus bas) Puis, on retient la position intermédiaire de la transition /ai/, c'est à dire la voyelle /e/. A partir de la configuration de la voyelle /a/, la labialisation entraîne la production de la voyelle /ɔ/. On retrouve ici le système vocalique de 5 voyelles le plus répandu [5]

On a reproduit figure 3, les trajectoires formantiques obtenues par un modèle à quatre régions. R1 est fixé à 1.4cm², R2 et R3 sont commandées synergétiquement. Les sections varient entre 0.7 et 11cm². On a placé différentes voyelles dans le plan F1-F2. A titre indicatif, on a aussi représenté la trajectoire /u/ obtenue par ouverture aux lèvres d'un modèle fermé-fermé avec constriction centrale. Les voyelles /o/ et /ɔ/ et leurs correspondants non labialisés pourraient être obtenue de la même manière.

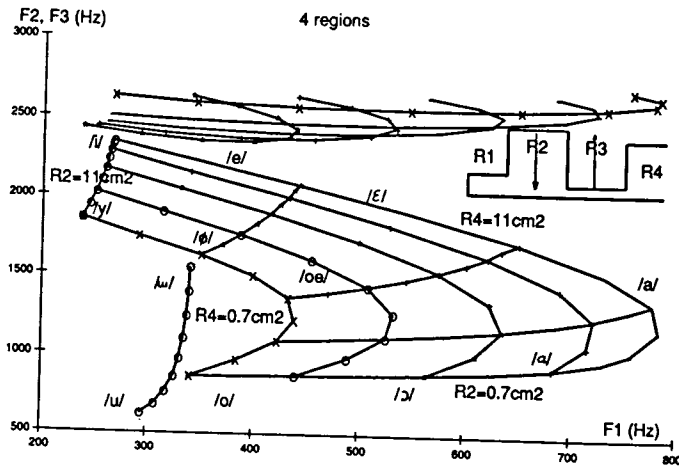


Figure 3. Position de voyelles dans le plan F2, F3(F1) et trajectoires formantiques obtenues avec un modèle constitué de quatre régions distinctives. On a aussi représenté la trajectoire /u/ obtenue avec un modèle fermé-fermé à trois régions.

La prise en considération du troisième formant (pour distinguer les voyelles /i/ et /y/ qui, dans certains cas, ont même deuxième formant par exemple) peut être obtenue avec un modèle en huit régions et commande synergétique de régions symétriques R3 et R6 par exemple (dans ce cas, la plage du deuxième formant est réduite et permet de réaliser les voyelles centrales) ou bien par synergie avant ou bien arrière (R3 et R4 par exemple) ou

bien par une légère asymétrisation de la commande d'un modèle à quatre régions. Un déplacement de 2 cm (figure 4) vers l'avant de l'axe de symétrie permet de stabiliser le deuxième formant lors de la transition /iy/ alors que la plage de variation du troisième formant augmente.

5. DISCUSSIONS

On a montré que l'on pouvait simplement produire les voyelles avec un mini-

mum de paramètres de commande. Une stratégie de commande transversale paraît réaliste : elle permet de réaliser simplement des trajectoires que l'on retrouve en

parole naturelle. On peut alors émettre l'hypothèse que l'homme exploite les caractéristiques acoustiques optimales d'un conduit vocal qui serait divisé en régions.

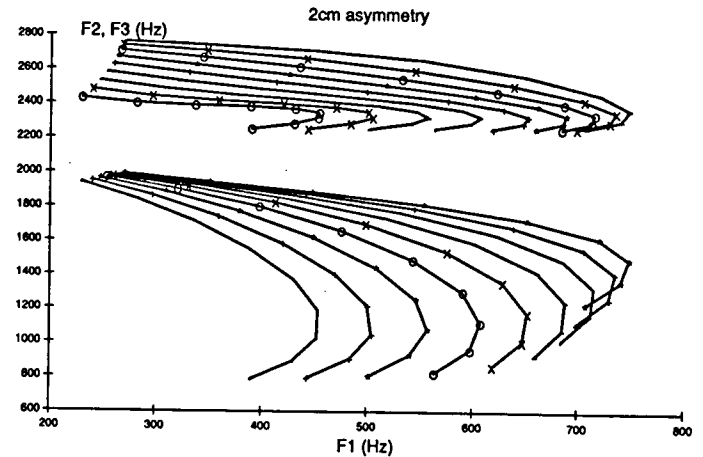


Figure 4. Trajectoires formantiques obtenues avec un modèle à quatre régions pour un déplacement de 2cm de l'axe de symétrie.

Cette interprétation diffère de celle émise par Lindblom [4] qui propose le système de perception comme référence. Dans la logique de l'hypothèse que nous formulons, le lexique serait constitué d'intentions de type réalisation d'une constriction à tel ou tel endroit quantifié du conduit vocal, associées à une intention de type labialisation ou non labialisation. La réalisation des intentions serait obtenue, au niveau périphérique, par une structuration des paramètres articulatoires pour réaliser les gestes phonétiques souhaités. L'objectif du système de perception serait alors de reconnaître les intentions initiales [3]. Le pilotage des systèmes de production et de perception de la parole chez l'homme par des lois physiques n'est pas irréaliste. Si cette hypothèse se vérifie, les conséquences seraient nombreuses ne serait-ce que pour proposer des schémas de l'évolution de l'appareil vocal humain au cours des millénaires.

REFERENCES

[1] Carré, R and Mrayati, M. (1990), "Articulatory-acoustic-phonetic relations and modeling, regions and modes", in

Speech Production and Speech Modelling (W.J Hardcastle and A. Marchal, editors), NATO ASI Series, Kluwer Academic Publishers.

[2] Carré, R. and Mrayati, M. (1991), "Vowel-vowel trajectories and region modeling", Second seminar on speech production: models and data (Leeds, 1990), to appear in The Journal of Phonetics.

[3] Fowler, C.A. and Rosenblum, L.L. (1989), "The perception of phonetic gestures", Haskins Labs., SR-99/100, 102-117.

[4] Lindblom, B. (1986), "Phonetic Universals in Vowel Systems", in Experimental Phonology, J.J. Ohala and J.J. Jaeger ed., Academic Press.

[5] Maddison, I. (1984), "Patterns of Sounds", Cambridge University Press, Cambridge.

[6] Mrayati, M., Carré, R. and Guérin, B. (1988), "Distinctive regions and modes: a new theory of speech production", Speech Communication, 7, 257-286.

[7] Mrayati, M., Carré, R. and Guérin, B. (1990), "Distinctive regions and modes: articulatory-acoustic-phonetic aspects", Speech Communication, 9, 231-238.