



HAL
open science

ESPACE VOCALIQUE MAXIMAL

L. Boe, Pascal Perrier, B. Guerin, J. Schwartz

► **To cite this version:**

L. Boe, Pascal Perrier, B. Guerin, J. Schwartz. ESPACE VOCALIQUE MAXIMAL. Journal de Physique Colloques, 1990, 51 (C2), pp.C2-503-C2-506. 10.1051/jphyscol:19902118 . jpa-00230411

HAL Id: jpa-00230411

<https://hal.science/jpa-00230411>

Submitted on 4 Feb 2008

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.

1er Congrès Français d'Acoustique 1990

ESPACE VOCALIQUE MAXIMAL

L.J. BOE, P. PERRIER, B. GUERIN et J.L. SCHWARTZ

*Institut de la Communication parlée, UA CNRS n° 368, Université
Stendhal, BP. 25X, F-38040 Grenoble Cedex, France
INPG/ENSERG, 46 Avenue Félix Viallet, F-38030 Grenoble Cedex, France*

Résumé - Une voyelle peut être caractérisée par ses trois premiers maxima spectraux, les formants ; mais toutes les combinaisons $F_1 F_2 F_3$ ne sont pas possibles : l'ensemble des voyelles humainement productibles est contenu dans l'Espace Vocalique Maximal. Le but de ce travail est d'analyser quelles sont les causes acoustiques et les contraintes articulatoires qui sont à l'origine des caractéristiques de cet espace et d'en proposer une description bi- et tridimensionnelle.

Abstract - A vowel sound can be characterized acoustically in terms of its formant pattern, generally by the first three formants. However, all combinations of $F_1 F_2 F_3$ are not produced by human speakers. Vowel space is contained in what we call Maximal Vowel Space (MVS). The aim of this work is to determine acoustic and articulatory respective contributions to characteristics related to the MVS and to propose a bi- and tridimensional description of the MVS.

1 - INTRODUCTION

Parmi toutes les voyelles produites par les locuteurs de différentes langues (c.f. /1/ pour l'inventaire actuel le plus exhaustif) toutes les combinaisons $180 \text{ Hz} \leq F_1 \leq F_2 \leq F_3 \leq 3500 \text{ Hz}$ ne sont pas possibles. Par exemple la réalisation $F_1 = 800 \text{ Hz}$ $F_2 = 2500 \text{ Hz}$ $F_3 = 2600 \text{ Hz}$, qui peut être générée artificiellement par un modèle à n-tuyaux sans contrainte articulatoire ou par un synthétiseur à formants, n'est pas anthropophonique. L'ensemble des voyelles potentiellement réalisables par un être humain est entièrement contenu dans ce que nous appellerons l'Espace Vocalique Maximal (EVM). Peu décrit en 3D, sa projection dans le plan $F_1 F_2$ est bien connue : ce "sabot de cheval" a même été décrit analytiquement /2/ à partir d'un modèle articulatoire. La connaissance de l'EVM est très importante pour comparer les systèmes vocaliques de différents locuteurs d'une même communauté linguistique, pour mener des études contrastives entre différentes langues ou pour proposer des hypothèses de distribution des systèmes vocaliques /2, 3, 4/. Pour une langue donnée il est possible d'avoir une idée assez précise de la projection de l'EVM dans le plan $F_1 F_2$ à partir d'un corpus limité, prenant en compte les effets maximaux de coarticulation /5/, mais c'est évidemment en génération, c'est à dire par le recours à un modèle du conduit vocal, qu'il sera plus facile de définir précisément les contours de l'espace $F_1 F_2 F_3$. Le but de ce travail est d'analyser quelles sont les causes acoustiques et les contraintes articulatoires qui sont à l'origine des caractéristiques de l'EVM et d'en mieux dessiner le contour dans l'espace $F_1 F_2 F_3$.

2 - LIMITES ACOUSTIQUES INTRINSEQUES

Découpons la projection de l'EVM dans le plan $F_1 F_2$ en trois segments $L_1 L_2$ et L_3 (figure 1). La limite L_1 est associée à la fréquence la plus basse du conduit vocal avec vibrations des parois /6/ : elle correspond à la résonance du Helmholtz constitué, pour les voyelles hautes [i, y, ɨ, u] par l'ensemble "cavité arrière + constriction" , et / ou, pour la voyelle [u], de la "cavité avant + lèvres" /7, 8/. L_2 et L_3 dépendent respectivement de l'ensemble des valeurs maximales et minimales de F_2 pour une valeur donnée de F_1 . L_3 est évidemment entièrement située au dessus de la courbe $F_1 = F_2$ puisque par définition $F_2 \geq F_1$.

Afin de préciser les limites L2 et L3 utilisons d'abord un modèle à 4-tuyaux que nous prendrons ici de longueurs égales ℓ_i et dont les aires A_i sont choisies indépendamment les unes des autres. Nous nous sommes seulement fixés des contraintes réalistes pour les valeurs des A_i avec $A_{\min} = 0.30 \text{ cm}^2$, constriction en dessous de laquelle on ne peut plus considérer qu'il s'agit d'un son vocalique, et avec $A_{\max} = 15 \text{ cm}^2$ - maximum rarement atteint, mais observable. Nous avons retenu une longueur totale $\ell = 17.5 \text{ cm}$, une bonne moyenne entre [i] et [u] pour un conduit vocal masculin. Dans le cas d'une modélisation sans perte, il existe entre F1 et F2 les relations suivantes /9, 10/ :

$$(1) \tan^2(\tau F_1) + \tan^2(\tau F_2) = C_1 \quad \text{et} \quad (2) \tan^2(\tau F_1) \times \tan^2(\tau F_2) = C_2$$

avec : $\tau = 2\pi \ell / 4c$ (c : vitesse du son)

$$C_1 = (k_1 + k_2 + k_3 + k_1 k_2 + k_2 k_3 + k_1 k_2 k_3) / k_1 k_3 \quad C_2 = 1 / k_1 k_3 \quad \text{et} \quad k_i = A_i / A_{i+1}$$

L2 est déduit de l'ensemble des solutions des équations (1) et (2) et plus simplement grâce à l'utilisation d'un abaque /9, 10/ ; la position de L3 dépend de la valeur minimale de k_2 (que nous avons fixée à 0.02 pour différentes valeurs de $k_1=k_3$). La figure 2 présente les résultats ainsi obtenus avec mise en évidence de 4 configurations caractéristiques.

Pour vérifier ces résultats analytiques, nous avons généré l'espace maximal de ce modèle à 4 tuyaux avec les 1396 arrangements possibles pour $A_i = 0.30 \ 0.65 \ 1.40 \ 3.00 \ 7.50 \ 15.0 \text{ cm}^2$ et avec $\ell = 17.5 \text{ cm}$. Le programme de simulation harmonique intègre les pertes par friction, conduction de chaleur et vibrations des parois, ainsi que le rayonnement aux lèvres. On retrouve bien les valeurs théoriques prédites par les équations (1) et (2) : $F_{2\max} = 2000 \text{ Hz}$ et $F_{1\max} = 1048 \text{ Hz}$ pour $F_2 = 1480 \text{ Hz}$. Cette simulation permet de tracer L2 et L3, mais la forte pente de L2, qui donne à cet EVM une forme triangulaire plutôt que l'arrondi d'un "sabot", autorise des réalisations de $F_1 > 1000 \text{ Hz}$ et des configurations $F_1 = 800 \text{ Hz} \ F_2 = 1800 \text{ Hz}$ qu'on ne peut retenir comme "naturelles". En fait un modèle à n-tubes peut-il prétendre à l'anthropomorphisme tant qu'il n'intègre pas les contraintes inhérentes au processus articulatoire de la production ?

3 - CONTRAINTES ARTICULATOIRES

Pour affiner la forme de l'EVM nous avons retenu deux modèles qui, chacun à leur manière, prennent en compte des caractéristiques articulatoires fondamentales.

3.1 - Le modèle à 4 tuyaux de Fant /11/

Ce modèle de connaissance repose, pour l'essentiel, sur l'hypothèse fondamentale de la prépondérance du rôle joué, pour la production des voyelles, par la zone de constriction (position et dimension) et par les lèvres. Il permet de produire l'essentiel de la substance vocalique à partir de 4 paramètres fondamentaux : la longueur ℓ du conduit vocal, la constriction intra-buccale caractérisée par sa position X_c (par rapport à la glotte), son aire A_c et enfin l'aperture labiale A_ℓ . Il est possible d'explorer l'espace vocalique en déplaçant une constriction de longueur constante (tuyau n° 2 $\ell_2 = 5 \text{ cm}$) entre les deux extrémités (la glotte et les lèvres). Les aires des tuyaux 1 et 3 (pharynx et cavité d'avant) sont fixées à 8 cm^2 , la somme de leurs longueurs restant constante ($\ell_1 + \ell_3 = 10 \text{ cm}$) et la longueur des lèvres ℓ_4 est maintenue à 1 cm (tuyau 4). Nous avons généré 5152 réalisations vocaliques avec 28 valeurs de X_c correspondant à un déplacement de la constriction par pas de 0.5 cm entre 3 cm et 14 cm (des limites réalistes /12/), pour 23 valeurs de A_c comprises entre 0.2 et 4.0 cm^2 et 8 valeurs de A_ℓ réparties entre les mêmes limites. L'EVM ainsi généré présente, par rapport au précédent, de nettes améliorations. Sa forme globale est plus arrondie et toutes les configurations produites peuvent être retenues comme réalistes. Mais si $F_{2\max}$ est toujours voisin de 2000 Hz , le point le plus bas du contour (environ $F_{1\max} = 650 \text{ Hz}$ et $F_2 = 1000 \text{ Hz}$) ne peut prétendre bien représenter la classe des [a]. C'est là un point faible de ce modèle qui ne peut bien générer la forme caractéristique de cette voyelle avec des limites réalistes pour X_c .

3.2 - Le modèle de Maeda /13 /

Ce modèle offre l'avantage du réalisme physiologique en intégrant les contraintes articulatoires : l'analyse en composante linéaire, qui a servi à son élaboration, a été guidée par une connaissance a priori du rôle des articulateurs. La coupe sagittale est ainsi générée à partir de 5 paramètres associés à la mâchoire, au corps, au dos, à l'apex de la langue et aux lèvres. Nous avons décrit l'espace articulatoire à partir d'un maillage relativement serré : 13 valeurs pour chaque paramètre; au total plus de 200.000 configurations ont été retenues (avec les mêmes contraintes que précédemment sur Xc, Ac et Aℓ) pour constituer un dictionnaire vocalique /14/. On retrouve bien les aspects positifs de l'EVM généré avec le modèle de Fant, et cette fois-ci F1max est nettement plus élevé - environ 800 Hz - ce qui permet de bien décrire la voyelle cardinale extrême [a]. Nous considérons que cette projection est bien représentative et qu'elle peut servir de référence pour les limites L2 et L3 ajustées respectivement par les deux paraboles :

$$F_1 = -1.04 \cdot 10^{-3} F_2^2 + 2.59 F_2 - 797 \quad F_1 = -4.00 \cdot 10^{-4} F_2^2 + 2.59 F_2 - 797$$

avec l'intersection $F_1 = 790 \text{ Hz}$ $F_2 = 1100 \text{ Hz}$.

4 - UNE REPRESENTATION 3D de l'ESPACE VOCALIQUE MAXIMAL

Grâce à la simulation il nous est donc possible d'améliorer les représentations en 3D de l'EVM données jusqu'ici /3/. Le volume vocalique ainsi obtenu a la forme d'une "corne" (figure 3). A partir de cette simulation nous avons déjà proposé une paramétrisation des coupes F1 / F2 de la "corne vocalique" pour des valeurs de F3 réparties entre 2000 et 2900 Hz par des arcs de parabole Pi et des droites Di (figure 4) que nous avons déjà utilisée dans le cadre de la prédiction des systèmes vocaliques /4/.

5 - CONCLUSION

En utilisant une modélisation à 4 tuyaux nous avons pu avancer une explication acoustique des limites intrinsèques de l'EVM. Mais seule la prise en compte de contraintes articulatoires permet d'obtenir la forme caractéristique en "sabot" dans le plan F1 F2. Le modèle anthropomorphique de Maeda fournit ainsi la possibilité de déterminer une référence pour la caractérisation de l'EVM en 2 et 3D à partir d'une génération en extension : se dessine ainsi, pour la première fois, les contours de la "corne (d'abondance) vocalique" dans l'espace F1 F2 F3.

REFERENCES

- / 1 / MADDISON I. (1986) Patterns of sounds. Cambridge University Press, Cambridge.
- / 2 / LINDBLOM B. & LILJENCRANTS J. (1972) Language 48, 839-862.
- / 3 / LINDBLOM B. (1986). In : Experimental Phonology, 13-44. Academic. Press, New York.
- / 4 / SCHWARTZ J.L. BOE L.J. PERRIER P. GUERIN B. & ESCUDIER P. (1989) EuroSpeech 89 1, 63-66.
- / 5 / ABRY C. & BOE L.J. (1984) 13èmes JEP GCP - GALF 55-58.
- / 6 / FANT G. (1972) STL QPSR 2-3, 28-52.
- / 7 / BADIN P. & BOE L.J. (1987) 11th Int. Congr. of Phon. Sci. 2, 352-355.
- / 8 / BADIN P. PERRIER P. BOE L.J. & ABRY C. (1990) J. Acoust. Soc. Am. (à paraître).
- / 9 / BONDER L.J. (1982) Acustica 52, 216-226.
- / 10 / BONDER L.J. (1983) 10th Int. Congr. Phonetic Sci. vol IIB, 347-353.
- / 11 / FANT G. (1960) Acoustic Theory of Speech Production. Mouton & Co. The Hague.
- / 12 / LADEFOGED P. & BLADON A. (1982) Speech Communication 1, 185-198.
- / 13 / MAEDA S. (1979) J. Acoust. Soc. Am. 65, S22 (A).
- / 14 / PERRIER P. BOE L.J. MAJID SHIHAB R. & GUERIN B. (1985) 14èmes JEP GCP - GALF, 55-58.

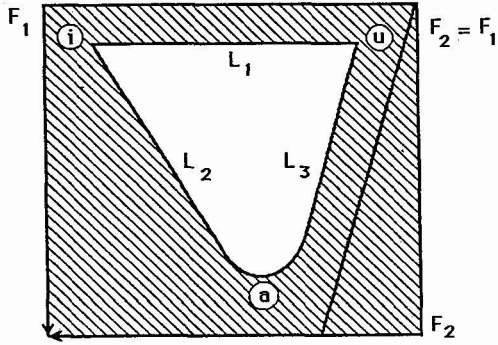


Fig. 1 - The limits L_1 L_2 L_3 of the Maximal Vowel Space

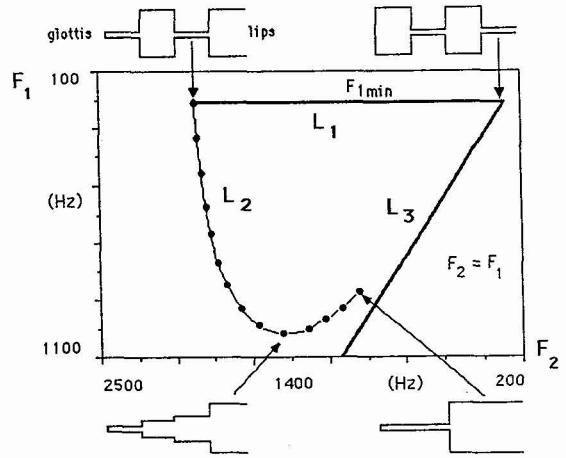


Fig. 2 - Limit L_2 calculated with a 4-tube model : $k_1 \times k_2 \times k_3 = 0.02$

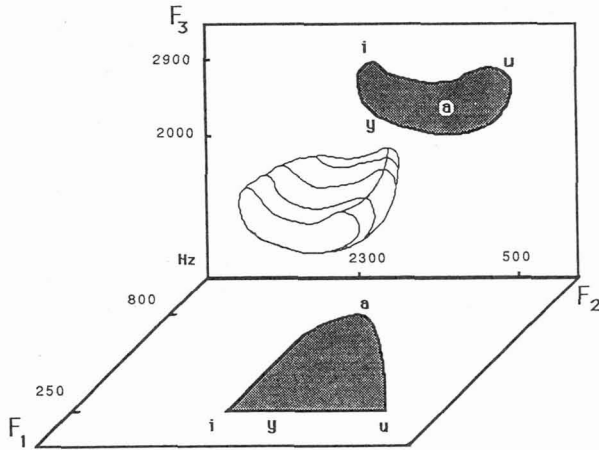


Fig. 3 - The 3-D "vocalic horn" obtained with Maeda's Model.

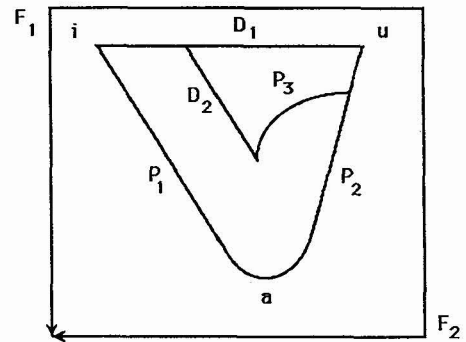


Fig. 4 - Parametrization of $F_1 - F_2$ views for F_3 values distributed between 2000 Hz and 2900 Hz.