

CONTRIBUTION A LA CLASSIFICATION ARTICULATORI-ACOUSTIQUE DES VOYELLES :  
"ETUDE DES "MACRO-SENSIBILITES A L'AIDE D'UN MODELE ARTICULATOIRE.

R. MAJID<sup>1</sup>, C. ABRY<sup>1</sup>, L.J. BOE<sup>1</sup> & P. PERRIER<sup>2</sup>

Institut de la Communication Parlée. UA CNRS 368, FRANCE.

<sup>1</sup> Institut de Phonétique, Univ. Langues & Lettres, BP 25X 38040 GRENOBLE CEDEX.

<sup>2</sup> Laboratoire de la Communication Parlée, ENSERG/INPG, 46 Av. Félix Viallet, 38031 GRENOBLE CEDEX.

RESUME

Le domaine de validité des fonctions de sensibilité de FANT & PAULI (1974) est trop limité pour qu'elles puissent être exploitées dans le cadre de la synthèse articulatoire. Avec le modèle articulatoire de MAEDA (1979), sont proposées les "macro-sensibilités" articulatoire-acoustiques de treize voyelles (dont les onze du français) qui décrivent la totalité du triangle vocalique. Les résultats sont analysés à la lumière de la théorie quantique de STEVENS et nous proposons deux paramètres articulatoires qui permettent une orthogonalisation du triangle vocalique inter-prétable en termes articulatoires.

INTRODUCTION

L'étude des conséquences de petites perturbations de la fonction d'aire du conduit vocal sur le signal acoustique correspondant, a fait l'objet de nombreux travaux. Citons parmi les plus importants ceux de FANT & PAULI [1] qui, à la suite de SCHROEDER [2] qui avait évalué les énergies cinétique et potentielle mises en jeu, ont proposé les "fonctions de sensibilité" associant les perturbations de la fonction d'aire et les variations des fréquences formantiques correspondantes.

Pour intéressantes qu'elles soient, ces fonctions présentent, à notre avis, trois limitations importantes :

- \* la fonction d'aire n'est pas une donnée directement interprétable en termes de commandes articulatoires ;
- \* les perturbations introduites dans la fonction d'aire ne tiennent pas compte des contraintes intra- et inter-articulatoires ;
- \* le domaine de validité des perturbations (10% max.) est faible, en regard des variations de forme du conduit vocal pouvant intervenir lors de la production réelle de parole.

Pour cet ensemble de raisons, nous avons étudié les "macro-sensibilités" articulatoire-acoustiques, c'est-à-dire les conséquences sur les fréquences formantiques de variations, éventuellement importantes, de paramètres articulatoires. Les résultats que nous avons ainsi obtenus, peuvent être interprétés en termes de stabilité articulatoire-acoustique [3], et font apparaître deux paramètres permettant une "orthogonalisation"

non arbitraire de l'espace vocalique  $F_1/F_2$  [4].

LA METHODE UTILISEE

Pour ce travail, nous avons adopté une modélisation globale du processus de production vocalique, susceptible d'intégrer au mieux les connaissances physiologiques et acoustiques.

La clé de voûte de cette étude est constituée par le modèle articulatoire proposé par MAEDA [5]. Notre choix trouve sa justification dans la conception même du modèle : il résulte d'une analyse statistique effectuée, dans un premier temps, sur des radiographies pour la langue, et, dans un deuxième temps, sur des labiofilms pour les lèvres, et il intègre a priori des connaissances sur les articulatoires, ce qui autorise à la fois une bonne adéquation avec la réalité physiologique et une grande facilité d'interprétation en termes de phonétique articulatoire. Les premiers tests globaux [6] ont très largement confirmé la validité de ce modèle. Il est commandé par cinq paramètres : les "lèvres", la "mâchoire", le "corps", le "dos" et la "pointe" de la langue. Il fournit une coupe sagittale, à partir de laquelle est évaluée la fonction d'aire [7] qui n'est ici qu'une donnée intermédiaire.

La réponse harmonique est ensuite calculée à l'aide d'un analogue électrique du conduit vocal [8] implanté par CHARPENTIER (1982) [9], qui inclut les pertes par vibrations des parois, par viscosité, par chaleur et par rayonnement.

LA SELECTION DE STANDARD VOCALIQUES.

A partir d'un "dictionnaire articulatoire-acoustique" contenant environ 200.000 formes vocaliques potentielles générées selon la procédure présentée ci-dessus, et décrivant la totalité du triangle acoustique  $F_1/F_2$  [6], nous avons déterminé les configurations susceptibles de correspondre aux voyelles orales du français, décrites par des ellipses de dispersion formantiques [10]. De cet ensemble, par approximations successives, nous avons déduit des valeurs standard en nous astreignant à respecter une quadruple cohérence structurelle au niveau : des paramètres articulatoires, des coupes sagittales, des fonctions d'aire et des formants.

Pour compléter ce triangle  $F_1/F_2$ , nous avons générés deux voyelles supplémentaires qui remplissent l'espace laissé libre entre le [y] et le [u] :

\* [w], voyelle obtenue à partir du [u] par un avancement de la langue et une plus grande ouverture aux lèvres ;

\* [x], une voyelle "inconnue", située dans l'étroite zone comprise entre [w] et [u] ; MAEDA [11], FENG et al. [12], ont avancé l'hypothèse que cette zone correspond à la cible des voyelles nasales ; [x] est obtenue à partir du [u] par un abaissement du dos de la langue et de la mâchoire.

LES MACRO-SENSIBILITES PARAMETRIQUES.

Dans le modèle de MAEDA, les paramètres de commande évoluent autour d'une valeur moyenne, dans les limites de - 3 à + 3 fois l'écart-type. Nous avons décidé de faire varier, un à un, chacun des paramètres de commande, avec une dynamique totale de 2 fois l'écart-type, répartie, quand c'était possible, symétriquement de part et d'autre de la valeur cible de chaque paramètre, et ceci pour chaque voyelle.

Examinons maintenant un à un l'effet de chacun des paramètres articulatoires.

Les voyelles françaises.

Le paramètre "lèvres". (Figure 1.a.)

On constate une tendance générale de toutes les voyelles à converger vers la zone du [u]. Les macro-sensibilités du [o], du [ɔ] et du [u] à ce paramètre sont particulièrement importantes. Pour la voyelle [u], en particulier, la rétraction des lèvres, liée à l'augmentation de l'ouverture labiale, induit un important déplacement vers le centre du triangle acoustique, via le [o].

C'est en revanche pour [i] et pour [e], et pour les voyelles d'avant en général, que le paramètre "lèvres" a le moins d'influence. Notons cependant qu'il est loin d'être négligeable, puisque pour le [i] la fermeture-protrusion des lèvres permet d'atteindre la zone du [y]. Enfin, pour les voyelles [a] et [o] et [ɔ], on constate essentiellement une variation du premier formant  $F_1$ , dont on sait qu'il est sensible à l'ouverture labiale.

De manière générale, on obtient un résultat déjà souligné par LINDBLOM & SUNDBERG [13] et PERKELL & NELSON [14] : les fréquences des formants diminuent avec la fermeture des lèvres.

Le paramètre "mâchoire". (Figure 1.b.)

Dans tous les cas, l'abaissement de la mâchoire provoque une augmentation de  $F_1$  et une diminution plus ou moins sensible de  $F_2$  (sauf pour le [u]). Mais c'est avec [y], [i] et [œ] qu'apparaissent les variations maximales. Ces résultats confirment pleinement ceux de LINDBLOM & SUNDBERG [13] et les précisent. Remarquons que, dans le cas du [a], l'effet de l'abaissement de la mâchoire n'est pas une fonction monotone. On peut expliquer ce phénomène par le fait que la position du lieu de constriction, situé dans la région pharyngale, est affectée lors de l'ouverture de la mâchoire : on observe ainsi à partir d'un certain point, les effets conjoints de l'ouverture de la

mâchoire et du déplacement du lieu d'articulation. Notons enfin la grande insensibilité du groupe [u]. [o], [ɔ] aux variations du paramètre "mâchoire".

Le paramètre "corps de la langue"

(Figure 1.c)

Le déplacement de la langue vers l'avant provoque, de manière générale, une tendance à la convergence vers la zone du [i]. Ceci est très marqué pour les voyelles d'avant qui sont de loin les plus sensibles.

La précision nécessaire sur [i] est très importante puisque dès qu'on recule la langue, on passe à un [e], puis à un [œ], en se dirigeant rapidement vers le centre du triangle vocalique.

Le paramètre "dos de la langue".

(Figure 1.d)

Ce sont [i], [y] et [e] qui sont les voyelles les plus sensibles aux variations de ce paramètre. En revanche, les voyelles d'arrière sont relativement peu affectées. On observe que l'abaissement du dos de la langue fait tendre le [i] vers le [e], alors que [u] évolue légèrement dans la direction du [y]. Les déplacements dans le plan  $F_1/F_2$  se font en gros selon deux directions quasi orthogonales, de façon beaucoup plus nette que les simples tendances mises en évidence par GOLDSTEIN [15]. L'abaissement du dos de la langue peut avoir quatre conséquences différentes :

- \* pour [i, e, ε]  $F_1$  augmente et  $F_2$  diminue ;
- \* pour [y, œ, œ]  $F_1$  augmente et  $F_2$  augmente ;
- \* pour [u, o, ɔ]  $F_1$  est stable et  $F_2$  augmente ;
- \* enfin, pour [a],  $F_1$  reste stable et  $F_2$  diminue.

Le [a] se différencie des autres voyelles d'arrière par le sens de la variation de  $F_2$ . Là encore, l'explication réside dans le recul du lieu de constriction associé à l'aplatissement de la langue.

Le paramètre "pointe de la langue".

(Figure 1.e)

L'abaissement de la pointe de la langue affecte, comme on peut facilement le prévoir, essentiellement les voyelles d'avant fermées [i, e, y]. Il s'accompagne généralement d'une diminution de  $F_2$  et d'une augmentation de  $F_1$ .

Conclusions.

De manière générale, il ressort de l'observation des figures 1.a-1.e que [i, e, y] sont les voyelles les plus sensibles, avec en particulier de grosses variations de  $F_1$ , et qu'inversement, les voyelles d'arrière présentent une assez bonne résistance à la variation des paramètres.

Les voyelles [w] et [x].

L'observation de la figure 1 montre que le comportement de la voyelle [w] est en grande partie comparable à celui des voyelles d'avant, sauf pour le "dos de la langue" analogue à celui

du [y]. Cette voyelle se situe donc un peu à mi-chemin entre les voyelles d'avant et les voyelles d'arrière du français.

Mis à part pour les "lèvres", la voyelle [x], ne présente pas de grandes sensibilités aux paramètres articulatoires : elle se rapproche en cela du [u]. Le fait que, pour l'ensemble des langues, peu de voyelles fréquentent cette zone du plan  $F_1/F_2$  [11], ne semble donc pas pouvoir être expliqué par une grande instabilité articulatoire-acoustique. S'agit-il d'instabilité articulatoire-perceptive, de "mauvaise forme perceptive" [16], ou faut-il introduire des contraintes articulatoires encore plus sévères sur le modèle de MAEDA ? Le débat reste ouvert...

#### ORTHOGONALISATION DU TRIANGLE VOCALIQUE.

#### Deux paramètres articulatoire-acoustiques non ambigus : lèvres et corps de la langue.

En 1983, FANT [4] a proposé deux paramètres spectraux, "spread" et "flatness", correspondant dans l'espace acoustique à  $F_2 - F_1$  et  $F_2 + F_1$ , permettant une optimisation de la représentation du système vocalique du suédois. Ainsi la distinction entre les voyelles d'arrière et les autres, d'une part, et les voyelles arrondies et les autres d'autre part, est bien mise en évidence. Cependant la rotation ainsi effectuée dans le plan  $F_1/F_2$  présente un aspect arbitraire. Nous pensons que l'utilisation des résultats de l'étude des macrosensibilités peut être à cet égard très intéressante.

Observons, en effet, les figures 1.a et 1.c : la fermeture-protrusion des lèvres tend à déplacer toutes les voyelles dans la direction du [u]; tandis que toutes ont tendance à converger vers le [i] lors d'un mouvement du corps de la langue vers l'avant. Ce phénomène n'apparaît que dans ces deux cas : les paramètres "lèvres" et "corps de la langue" sont donc les seuls paramètres non-ambigus dans la relation articulatoire-acoustique. Utilisons cette propriété en traçant les droites de régression, dans le plan  $F_1/F_2$ , de l'ensemble des points des trajectoires obtenues pour les "lèvres" d'une part (droite  $D_1$ ), et pour le "corps de la langue" d'autre part (droite  $D_2$ ). Considérons alors le plan orthogonal défini par les droites  $D_1$  et  $D_2$ . Les positions cibles se placent alors sur un triangle orthogonal défini par les droites [i-a] et [a-u] (Figure 2). Cette première représentation souligne l'importance du "corps de la langue" et des "lèvres" pour le classement articulatoire des voyelles. Voyons maintenant quelle peut être l'interprétation géométrique de ces deux axes.

#### Interprétation géométrique.

Nous proposons de relier : l'axe [i-a] à la notion de postériorité (Backness [17]), que nous définissons comme la distance entre l'incisive inférieure et le centre d'un cercle ajusté à la plus grande courbure de la langue ; l'axe [a-u] à la notion de profondeur (depth [18]), correspondant à la distance entre

l'extrémité des lèvres et le point d'élévation maximal de la langue. La figure 3 montre bien la possibilité d'une adéquation satisfaisante entre la représentation des positions cibles des voyelles dans cet espace et leurs positions dans l'espace décrit plus haut ( $D_1/D_2$ ).

#### Relation avec l'acoustique.

Dans le plan acoustique, on peut faire la correspondance entre ces axes et les axes  $F_1$  et  $F_2$ , où  $F_2 = 0.6 * F_1$ , ainsi que le confirme la figure 4. Notons qu'articulatoirement, la postériorité est une grandeur absolue (référence fixe) et la profondeur une grandeur relative (référence mobile), alors que, curieusement, c'est l'inverse dans le plan acoustique.

Nous avons ensuite examiné les relations entre les paramètres articulatoires, les paramètres géométriques et les paramètres acoustiques ainsi proposés. Les résultats sont donnés figure 5. On remarque la bonne corrélation entre les données géométriques et acoustiques, et les paramètres articulatoires et acoustiques. La corrélation entre "lèvres" et "profondeur" est moins bonne, mais cela est sans doute dû aux caractéristiques du modèle articulatoire, dans lequel, rappelons le, les "lèvres" ne sont pas liées statistiquement à la mâchoire et à la langue.

Nous pouvons donc proposer, dans le plan  $F_2/F_1$ , une rotation optimisant la représentation du système vocalique parfaitement interprétable en termes articulatoires et reliée au rôle joué par les lèvres et surtout par le corps de la langue.

#### CONCLUSIONS.

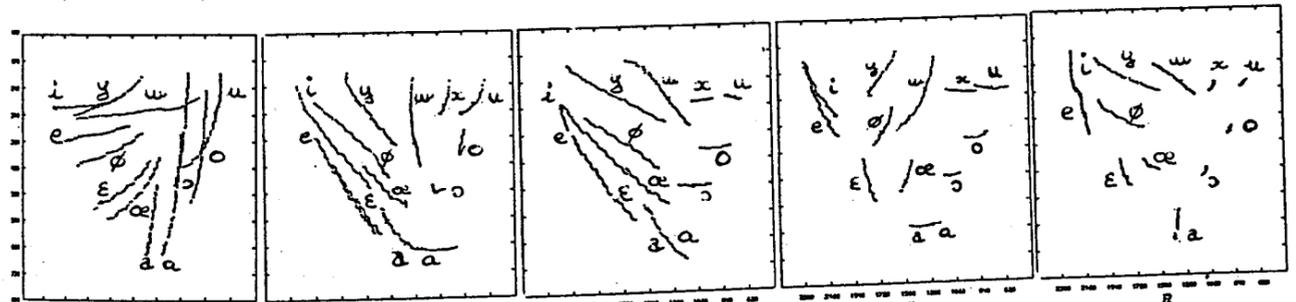
Nous avons défini pour chaque voyelle des positions cibles et pour chacune d'elles nous avons présenté leurs macro-sensibilités aux variations des paramètres articulatoires. Les résultats que nous avons obtenus confirment les grandes tendances déjà mises en évidence par LINDBLOM & SUNDBERG [13] et GOLDSTEIN [15], et les affinent. Ils vont dans le sens général de la théorie quantique de STEVENS [3] : il existe, en effet, très clairement des régions acoustiques qui sont peu sensibles aux variations de certains paramètres articulatoires. Enfin l'exploitation de ces résultats nous a permis d'orthogonaliser l'espace vocalique, selon une rotation dans l'espace acoustique, parfaitement interprétable en termes articulatoires. Il nous reste cependant à affiner la définition de la notion de profondeur.

#### REFERENCES.

[1] G. FANT & PAULI S., "Spatial Characteristics of Vocal Tract Resonance Modes.", Speech. Comm. Seminar 2, 121-132, 1974.  
 [2] SCHROEDER M.R., "Determination of the Geometry of the Human Vocal Tract by Acoustic Measurement.", J. Acoust. Soc. Am. 41, 1002-1010, 1967.

[3] K.N. STEVENS, "The Quantal Nature of the Speech : Evidence from Articulatory-Acoustic Data.", in Human Communication, Mc Graw Hill, New York, 1972.  
 [4] G. FANT, "Speech Production : Feature Analysis of Swedish Vowel. A revisit.", STL QPSR 2-3, 1-15, 1983.  
 [5] S. MAEDA, "Un modèle articulatoire de la langue avec des composantes linéaires.", 10èmes JEP, GALF-GCP, 152-162, 1979.  
 [6] P. PERRIER, L.J. BOE, R. MAJID & B. GUERIN, "Modélisation articulatoire du conduit vocal : exploration et exploitation.", 14èmes JEP, GALF-GCP, 55-58, 1985.  
 [7] H. SANCHEZ & L.J. BOE, "De la coupe sagittale à la fonction d'aire du conduit vocal.", 13èmes JEP, GALF-GCP, 23-25, 1984.  
 [8] J.L. FLANAGAN, K. ISHIZAKA & K.L. RIPLEY, "Synthesis of Speech from a Dynamic Model of the Vocal Cords and Vocal Tract.", B.S.T.J. 54, 485-506, 1975.  
 [9] F. CHARPENTIER, "Un logiciel de simulation électrique du conduit vocal.", C.N.E.T. Lannion, Comm. Perso., 1982.  
 [10] C. ABRY, BOE L.J. & R. DESCOUT, "[i,a,u] ? Pas si fou ? Ou les lèvres des consonnes maximisent elles l'espace des voyelles ?", 13èmes JEP, GALF-GCP, 205-207, 1984.

[11] S. MAEDA, "Une paire de pics spectraux comme corrélat acoustique de la nasalisation des voyelles.", 13èmes JEP, GALF-GCP, 223-224, 1984.  
 [12] G. FENG, C. ABRY & GUERIN B., "The Nasopharyngeal Tract : A Target for Nasality. Acoustic Simulation vs. Sweep Tone Measurements.", Proc. 12th. I.C.A., A 3.3, 1986.  
 [13] B.E.F. LINDBLOM & J.E.F. SUNDBERG, "Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement.", J. Acoust. Soc. Am. 50, 1166-1179, 1971.  
 [14] J.S. PERKELL & W.L. NELSON, "Variability in Production of the Vowels /i/ and /a/.", J. Acoust. Soc. Am. 77, 1889-1895, 1985.  
 [15] L. GOLDSTEIN, "Vowel Shifts and Articulatory-Acoustic Relations.", 10th Int. Congr. Phonetic Sci. IIA, 267-273, 1983.  
 [16] C. ABRY & J.L. SCHWARTZ, "Quelques éléments pour une théorie des objets phonétiques du langage... Autour d'une voyelle inconnue.", Revue de l'I.C.P., Bulletin du L.C.P., 1, 1987.  
 [17] N. CHOMSKY & M. HALLE, "The Sound Pattern of English.", Harper & Row, New York, 1968.  
 [18] H. TRAUNMULLER, "Some Aspects of the Sounds of Speech Sounds.", Workshop of the Psychophysics of Speech Perception, Utrecht, 1986.



1.a : Les lèvres 1.b : La mâchoire 1.c : Le corps 1.d : Le dos 1.e : La pointe  
 Figure 1 : LES MACRO-SENSIBILITES

