

Contribution à la quantification du degré d'organisation des systèmes vocaliques.

K. Huet, B. Harmegnies

Université de Mons-Hainaut

Département de Communication Parlée – Place du Parc, 18 – MONS, Belgique

Tél.: +32 (0)65 37 31 44 - Fax: +32 (0)65 37 31 42

Mél : Kathy.Huet@umh.ac.be, Bernard.Harmegnies@umh.ac.be

<http://www.umh.ac.be/~compa/>

ABSTRACT

This paper presents an exploratory study on the vocalic differences which can be observed under different speaking styles. A single speaker was recorded on 4 situations with gradual involvement of the speaker in the communication process. A new index is introduced for the assessment of the system's degree of organization in each speaking style. The results are compared with those drawn from a procedure based on discriminant functions.

1. INTRODUCTION

Durant les deux dernières décennies, divers travaux se sont intéressés aux variations des caractéristiques acoustiques des sons de parole qu'induisent des changements dans les conditions de production. Les premières recherches en la matière se sont principalement centrées sur l'étude de la variabilité des timbres vocaliques de diverses langues d'origine germanique [Koo80] et [Nor86]. D'autres travaux, notamment conduits par l'un d'entre nous, se sont attachés aux voyelles de diverses langues d'origine romane, telles l'espagnol [Har92], le catalan [Ble93], l'italien [Har95] et le portugais [Del97]. Peu de travaux se sont par contre centrés sur le français.

La plupart de ces recherches avaient pour objectif la confirmation de l'existence d'un effet systématique de la situation de production sur le signal de parole. Elles ont, dès lors, recouru, le plus souvent à deux modalités fortement différenciées : d'une part, la conversation *spontanée*, entretenue le plus naturellement possible avec l'informant, afin d'obtenir une qualité de production proche de celle qui caractérise ordinairement les échanges verbaux banals; d'autre part, une situation classique de production de sons de parole en *laboratoire*, faisant appel tantôt à la réalisation de logatomes, tantôt à celle de mots signifiants isolés.

Il ressort de ces études que -dans le cas des productions en langues romanes, à tout le moins- la différenciation acoustique entre les segments associés à des catégories phonologiques différentes est plus tranchée en parole *de laboratoire* qu'en parole *spontanée*. Projetés dans le plan

F_1/F_2 , les nuages vocaliques apparaissent ainsi plus mêlés dans le cas de la parole spontanée que dans celui de la parole de laboratoire.

Diverses modalités de quantification ont été envisagées pour rendre compte de ces observations. L'idée de *centralisation* du système vocalique en parole spontanée (fortement débattue par certains auteurs: cf. [Lin63] et [Lin90]) a ainsi fait l'objet de l'élaboration d'un indice de centralisation [Har92]. Celui-ci indique, dans la plupart des langues étudiées, une tendance dominante -quoique non omniprésente- à un rapprochement acoustique du timbre de schwa, voyelle théorique définie comme le résultat de l'excitation d'un tube de section uniforme de 17.5 cm de long et habituellement considérée comme le modèle de la voyelle neutre. Par ailleurs, il a été noté que la dispersion au sein des nuages de points dans le plan formantique est, en règle générale, plus importante en parole spontanée qu'en parole de laboratoire. La conjonction de ces deux tendances aboutit à l'idée d'un système vocalique moins différencié (présentant des nuages plus rapprochés les uns des autres, mais chacun plus dispersé) en parole spontanée qu'en parole de laboratoire.

Néanmoins, bien plus que la centralisation, c'est en fait la variation du degré d'*organisation-désorganisation* du système qui apparaît comme le phénomène le plus universel au travers de l'ensemble des expériences menées à ce jour. Il semble ainsi raisonnable de considérer que tout état d'un locuteur résultant de l'ensemble des variables de la situation d'émission peut être associé à un état de plus ou moins grande organisation/désorganisation du système vocalique.

Deux conditions au moins sont requises à la mise à l'épreuve de cette conception. D'une part, il importe de diversifier les situations de production de signal de parole, afin d'étayer l'étude sur une variabilité suffisante des styles de parole. D'autre part, il convient de se doter d'une mesure apte à rendre compte avec finesse des variations de l'état du système vocalique sous l'effet des changements de style de parole. Or, la technique utilisée jusqu'à présent (simulation d'une tâche de reconnaissance automatique sur base de fonctions discriminantes dérivées des valeurs formantiques) ne donne finalement qu'une

information quantitative indirecte et assez grossière sur la variable à l'étude.

La présente communication constitue une première contribution au comblement de ces lacunes. Elle procède ainsi à l'exposition d'un sujet francophone à quatre situations contrastées de production du signal de parole et présente une alternative (l'indice ϕ) à la procédure basée sur l'analyse discriminante pour l'évaluation du degré d'organisation/désorganisation du système vocalique dans les quatre situations.

2. PROCÉDURE EXPÉRIMENTALE

2.1 Corpus et recueil des productions

Le corpus est constitué de productions vocales d'un locuteur francophone de Belgique issu de la région de Mons (zone d'influence picarde) et placé dans 4 situations de communication différentes.

Dans un premier temps, le locuteur participe à une conversation ordinaire avec l'expérimentateur qui lui demande de parler de différents thèmes tels que sa situation familiale, le dernier livre qu'il a lu, ce qu'il pense de tel ou tel sujet d'actualité, etc..

Cette conversation, d'une heure, environ, est entièrement enregistrée puis transcrite orthographiquement. Un échantillon aléatoire d'au moins 45 réalisations de chaque voyelle étudiée est sélectionné. Pour la présente étude, seules les voyelles délimitant la périphérie du triangle vocalique (/i/, /e/, /ɛ/, /a/, /ɔ/, /o/ et /u/) sont prises en considération. Même si l'équipartition des effectifs entre catégories vocaliques est souhaitée, il arrive dans certains cas - la probabilité d'apparition de certains phonèmes étant faible - que le nombre d'occurrences apparaissant dans la conversation n'atteigne pas 45. C'est alors la totalité des réalisations disponibles qui sont retenues.

L'échantillon de 268 réalisations enregistrées à la faveur de la première situation de communication (que nous appellerons 'parole spontanée') constitue l'ensemble des mots cibles que nous utiliserons dans les 3 autres situations de communication. Afin de neutraliser les effets contextuels, le locuteur devra reproduire, dans les autres conditions de communication, les mots ainsi extraits du corpus recueilli lors de l'entretien spontané. Chacun des mots cibles contenant les réalisations sélectionnées sera donc disponible pour analyse dans chacune des situations de communication.

Lors de la seconde mise en situation de communication, il est simplement demandé au locuteur de lire les mots contenant les phonèmes préalablement sélectionnés. Afin d'éviter les effets de liste, chaque mot est présenté en isolation sur une fiche ; les fiches sont présentées en ordre aléatoire. On obtient ainsi le corpus de 'parole de laboratoire'.

Dans un troisième temps ('production de parole sous condition de bruit'), le locuteur est équipé d'un casque diffusant dans les deux oreilles un bruit uniformément masquant. La tâche consiste également en la production des mots cibles.

Pour la dernière situation de communication ('production de parole sur requête d'un interlocuteur'), un interlocuteur éloigné, dont notre locuteur peut percevoir la voix dans son casque, lui fait répéter les mots cibles sous prétexte d'une mauvaise compréhension. Dans ce cas précis, le stock de mots cibles est noyé dans un ensemble de distracteurs.

Ces 4 sessions d'enregistrement ont été effectuées dans la chambre sourde du Laboratoire de Phonétique de l'Université de Mons-Hainaut. Les productions du sujet ont été enregistrées grâce à un microphone Neumann U87UP48 relié via un NAGRA IV S à la chaîne d'acquisition constituée d'un digitaliseur PCM Sony et d'un magnétoscope Panasonic NH75 HQ.

2.2 Analyse acoustique

Les fréquences des premier ($F_1(j)$) et second ($F_2(j)$) formants estimées au centre de chaque voyelle (j) sélectionnée pour chaque situation de communication sont évaluées au moyen de l'analyseur CSL 4300B KAY et du logiciel d'analyse KAY Multi-Speech 3700 version 2.01, à partir de spectrogrammes à large bande.

Nous disposons donc de 1072 mesures (4 situations fois 268 réalisations, ventilées comme suit : 45 [i], 37 [e], 42 [ɛ], 45 [a], 45 [ɔ], 22 [o] et 32 [u]) et nous nous intéressons ici aux nuages de points formés dans le plan biformalique par ces mesures.

Chaque mesure f en Hertz a été convertie en mels, d'après la formule

$$m = 2595 \log(1 + f/700),$$

proposée par Van Bergem [Van93].

3. ANALYSE DES RÉSULTATS

3.1 Analyse discriminante

De même que dans les travaux précédemment cités, nous avons appliqué à nos mesures la procédure basée sur l'analyse discriminante. Pour chaque situation de communication, les phonèmes étudiés étaient considérés comme les catégories à priori, tandis que les fréquences formantiques F_1 et F_2 de leurs réalisations, pour une situation donnée, étaient considérées comme variables discriminantes. Une fois les fonctions de discrimination obtenues, nous les avons utilisées pour simuler une tâche de reconnaissance dans chacune des situations étudiées. De cette manière, chaque point du plan F_1/F_2 représentant une réalisation d'une voyelle était affecté à posteriori à l'une des 7 catégories phonémiques à l'étude.

Les matrices de confusion, que nous ne pouvons reproduire ici, font apparaître des taux de reconnaissance correcte respectivement égaux à 52.6% en production de parole spontanée, 73.5% en production de parole de laboratoire, 72.4% en production de parole sous condition de bruit et 75% en production de parole sur requête d'un interlocuteur.

Au vu de ces résultats, il apparaît donc que la parole spontanée est associée avec un état de désorganisation du système nettement plus prononcé que dans les 3 autres styles. Ces derniers font, quant à eux, apparaître une suite ordonnée (en ordre croissant du pourcentage de reconnaissance correcte) : production de parole sous condition de bruit - production de parole de laboratoire - production de parole sur requête d'un interlocuteur.

Les pourcentages correspondants se différencient cependant très peu (différence maximale de 2.6%).

3.2 Indice ϕ

Ainsi qu'on l'a souligné plus haut, les informations quantitatives recueillies au moyen de l'analyse discriminante sont assez grossières. D'une part, ce n'est qu'à la faveur d'un raisonnement indirect que l'on considère que, si la reconnaissance est mauvaise, c'est que le système est désorganisé. D'autre part, le raisonnement est de l'ordre de l'inclusion ou de l'exclusion d'un élément par rapport à un groupe, et on peut donc suspecter une sensibilité particulière de la procédure aux valeurs exceptionnelles.

En vue de pallier ces défauts, nous avons mis au point une nouvelle procédure, inspirée de l'analyse de variance. Celle-ci se base sur l'établissement d'une analogie entre d'une part, l'écart (en analyse de variance) d'une valeur -donnée ou moyenne- à la moyenne de référence et, d'autre part, la distance euclidienne (pour nos voyelles) entre un point dans le plan F_1/F_2 -voyelle ou centre de gravité d'un nuage vocalique- et le centre de gravité de référence. Le but est de quantifier, au moyen d'un indice que nous dénommerons ϕ , et qui s'inspire de la statistique F de Fisher-Snedecor, le rapport de la variabilité inter-classe (i.e., inter-nuage) à la variabilité intra-classe (i.e., intra-nuage).

Nous disposons, à cet effet, de 7 groupes (constitués des 7 catégories de voyelles) et pour chacun de ces groupes, nous pouvons poser :

N_k (nombre de productions disponibles pour la catégorie k , avec $0 < k < 8$)

$N = \sum_{k=1}^7 N_k$ (nombre total de productions, toutes catégories confondues)

d_{intra} (distance entre le point j du nuage k et le centre de gravité de celui-ci).

$$d_{intra}(j) = \left[\left(F_1(j) - \langle F_1(k) \rangle \right)^2 + \left(F_2(j) - \langle F_2(k) \rangle \right)^2 \right]^{1/2}$$

avec $\langle F_1(k) \rangle$ la moyenne des F_1 de tous les points formant le nuage k et $\langle F_2(k) \rangle$ la moyenne des F_2 de tous les points formant le nuage k .

d_{inter} (distance entre le centre de gravité du nuage k et celui de l'ensemble des nuages de points).

$$d_{inter}(k) = \left[\left(\langle F_1(k) \rangle - \langle F_1 \rangle \right)^2 + \left(\langle F_2(k) \rangle - \langle F_2 \rangle \right)^2 \right]^{1/2}$$

avec $\langle F_1 \rangle$ la moyenne des F_1 , tous nuages confondus et $\langle F_2 \rangle$ la moyenne des F_2 , tous nuages confondus.

Ces distances étant considérées comme des écarts, nous pouvons calculer des 'sommes de carrés d'écarts' (SCE) :

$$SCE_{intra} = \sum_{k=1}^7 \sum_{j=1}^{N_k} d_{intra}^2(j)$$

$$SCE_{inter} = \sum_{k=1}^7 \left(N_k * d_{inter}^2(k) \right)$$

avec leurs nombres de degrés de liberté respectifs :

$$L_{intra} = N - k \text{ et } L_{inter} = k - 1$$

Enfin, nous pouvons calculer l'indice ϕ pour chacune de nos situations de communication :

$$\phi = \frac{CM_{inter}}{CM_{intra}}$$

où les carrés moyens (CM) sont obtenus en divisant les SCE par le nombre de degrés de liberté correspondant :

$$CM_{intra} = \frac{SCE_{intra}}{L_{intra}} \text{ et } CM_{inter} = \frac{SCE_{inter}}{L_{inter}}$$

Appliquée à nos données, cette procédure fournit les valeurs suivantes : 3463 en production de parole spontanée, 8970 en production de parole de laboratoire 7572 en production de parole sous condition de bruit et 10792 en production de parole sur requête d'un interlocuteur.

Au vu de ces résultats, il apparaît que peut être opéré le même classement des situations de production que celui qui avait émané du traitement à base d'analyse discriminante. Ainsi que le montre la figure 1, la répartition des valeurs obtenues sur le domaine de définition de la variable est cependant plus régulière.

4. DISCUSSION

Les observations effectuées à la faveur de l'expérience relatée ici confirment celles découlant des expériences antérieures, singulièrement pour les condition de *parole spontanée* et *parole de laboratoire*, dont le contraste

s'apparente à ceux déjà relevés. Les pourcentages de reconnaissance correcte dérivés de la tâche de simulation de reconnaissance apparaissent en effet dans un rapport comparable à ceux déjà identifiés. Les indices ϕ rendent compte également du même phénomène.

En outre, les deux situations nouvelles (*production de parole sous condition de bruit* et *production de parole sur requête d'un interlocuteur*) font apparaître, pour la première, un état de désorganisation du système *intermédiaire* à ceux qui caractérisaient la parole spontanée et la parole de laboratoire et, pour la seconde, un état d'organisation *supérieur* à celui correspondant à la parole de laboratoire. Le classement des situations de production de parole qui peut ainsi en être dérivé (1. production de parole sur requête d'un interlocuteur; 2. production de parole de laboratoire; 3. production de parole sous condition de bruit; 4. production de parole spontanée) est identique, que l'on s'appuie sur les pourcentages de reconnaissance correcte dérivés de l'analyse discriminante ou sur les indices ϕ . Néanmoins, on note que la répartition des valeurs de ϕ sur le domaine de définition de la variable est plus régulière que celle des pourcentages de reconnaissance. En effet, si ces derniers montrent une nette différence entre la parole spontanée et les trois autres styles de parole, ils se singularisent par une concentration importante de ces trois valeurs autour de celle de la parole de laboratoire.

Bien que ces observations, établies à partir d'un nombre de données trop peu important ne permettent pas de tirer de conclusion définitive, elles paraissent cependant justifier la poursuite des investigations relatives à une quantification du degré d'organisation/désorganisation du système vocalique à base de techniques numériques dérivées de l'analyse de variance.

Par ailleurs, si jusqu'à présent, le signal de parole produit artificiellement sur injonction d'un expérimentateur était apparu comme la plus proche manifestation du système canonique du locuteur, on observe ici que les signaux de parole émis sur requête de clarification d'un interlocuteur éloigné s'associent à un état encore plus différencié du système.

5. FIGURES

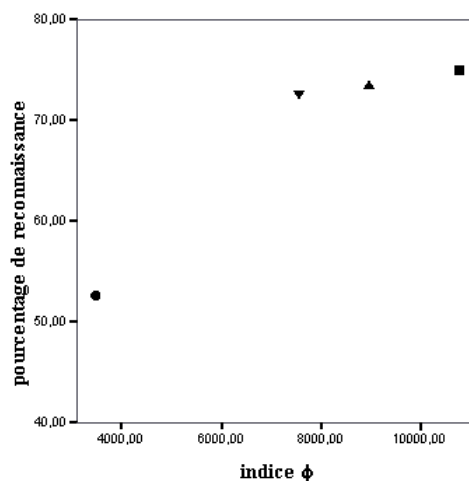


Figure 1: Taux de reconnaissance correcte (%) obtenus par l'analyse discriminante en fonction des indices ϕ calculés pour chacune des situations de communication envisagée : ● production de parole spontanée, ▲ production de parole de laboratoire, ▼ production de parole sous condition de bruit et ■ production de parole sur requête d'un interlocuteur.

BIBLIOGRAPHIE

- [Koo80] KOOPMANS-VAN BEINUM F. (1980) "Vowel Contrast Reduction. An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions", Amsterdam Academische Presse.
- [Nor86] NORD L. (1986), "Acoustic studies of vowel reduction in Swedish", STL-QPSR 4, pp. 19-36.
- [Har92] HARMEGNIES B., POCH-OLIVE D. (1992) "A study of style-induced vowel variability: laboratory versus spontaneous speech in Spanish", Speech Communication, 11, pp. 429-437.
- [Ble93] BLECUA-FALGUERAS B., POCH-OLIVE D., HARMEGNIES B. (1993), "Variaciones en la organización de la vocales del español y del catalán en function del estilo de habla", Proceedings of the International Conference of Applied Linguistics, Granada, pp. 97-107.
- [Har95] HARMEGNIES B., POCH-OLIVE D. (1995), "A dynamic Approach of Vowels Systems in Italian", Proceedings of the XIIIth International Congress of Phonetics Sciences, Stockholm, 1, pp. 408-412.
- [Del97] DELPLANCQ V., HARMEGNIES B. (1997), "Une modélisation à base angulaire pour l'étude de la réduction vocalique", Actes du 4ème

congrès français d'acoustique, Toulouse, Teknea, pp. 389-393.

- [Lin63] LINDBLOM B. (1963), "Spectrographic Study of Vowel Reduction", *J. Acoust. Soc. Amer.* 35, pp. 1771-1781.
- [Lin90] LINDBLOM B. (1990), "A sketch of the H and H theory". In A. Marchal et W. Hardcastle (eds.). *Speech Production and Modelling* pp. 403-440, Dordrecht, Kluwer, Academic Publishers.
- [Van93] VAN BERGEM D.R. (1993), "Acoustic vowel reduction as a function of sentence accent, word stress, and word class", *Speech Communication* 12, pp. 1-23.