Towards the unification of vowel spaces

Boë L.J.[1], Gabioud B.[2], Perrier P.[1], Schwartz J.L.[1] & Vallée N.[1]
[1]Institut de la Communication Parlée, URA CNRS 368, INPG & Université Stendhal
46, avenue Félix Viallet, 38031 GRENOBLE Cedex 01, France
[2]Laboratoire d'Analyse Informatique de la Parole - Université de Lausanne
Switzerland.
E-mail : boe@icp.grenet.fr

ABSTRACT

Representational spaces enable the study, comparison, and classification of vowel systems of different speakers in several languages. Two approaches compete at the articulatory level: the traditional description is based on the tongue body position (front/back and low/high) in the patatal region and on the lip geometry, whereas, in a more recent approach, classificatory parameters are the acoustically most relevant geometric features, namely the lip area and the position and size of the oral constriction. We intend in this paper to show that it is possible to preserve the traditional phonetic description and to link it to the acoustic one.

INTRODUCTION

In 1781, Hellwag [1] proposed to characterize the different vowels by articulatory and proprioceptive features. He thus defined a bi-dimensional, continuous space, in which vowels were distributed in three classes following three criteria: vocal tract aperture degree, front-back tongue position in the palatal part, and lip shape. These criteria were evaluated with respect to vowel [a].

In the same vein, the International Phonetic Association elaborated in 1888 the International Phonetic Alphabet (API) where vowels are depicted within a quadrilateral. The vowel quadrilateral is defined on the basis of three parameters, namely the lip shape, and the horizontal and vertical tongue positions, as described by the location of the highest point of the tongue in the palatal part of the vocal tract. It remains today the basis for different phonetic and phonological descriptions of languages.

However some limitations are emphasized by acoustical studies, which underline that such geometric descriptions are essentially limited to the visible palatal part of the vocal tract. And this part is acoustically, and then auditory, not relevant for some vowels such as [a].

From an auditory point of view, vowel production can be characterized with the first formant values. Thus Delattre [2] and Joos [3] proposed to base vowel representations on the first two formants.

In fact the link between the vowel quadrilateral and this latter acoustic space is fairly simple: it consists in permuting X- and Y-axes, and in inverting their orientations. F_1 is thus associated with the vertical tongue body position, and F_2 cumulates the effects of the anterior/posterior position of the tongue and of the protrusion/retraction of the lips. The three "star vowels" [i, a, u] (they are present in more than 90% of the languages, see UPSID, [4]) are located at the edges of the triangle. This description of the production space is more quantitative that the preceding one, and it will be systematically used for the description of linguistic systems and their prediction.

VOWEL REPRESENTATION BASED ON CONSTRICTION PROPERTIES

The acoustic theory of speech production ([5]; [6]) reduces vocal tract shape description to three relevant parameters: the constriction position X_c, its area A_c, the labial area A_l. Fant's nomograms showed that these parameters can be easy interpreted in reference to the spectral properties of the acoustic signal.

Moreover, from the study of the articulatory-to-acoustic relationships in an articulatory model of the vocal tract [7], Boë et al [8] proposed that these three geometrical parameters could play a major role in the control process of vowel production. Experimental works based on X-ray measurements by Wood [9] support this hypothesis, which is in the vein of Gay et al. suggestion [10] that vowels could be neurophysiologically coded in terms of oral constriction position and dimension.

Nevertheless, such a description presents strong limitations:

- The three main cardinal vowels [i, a, u] are not located at the edges of the space: in X_c/A_c and X_c/A_l planes, extreme vowels are respectively [i, a, œ] and [y, o, a, ɛ].
- The relative locations of [i], [a] and [u] do not respect the classical order, since [u] is between [i] and [a]; moreover [ɛ] is far from [æ] and [a], whereas data on phonetic changes support the idea that these vowels belong to similar classes.
- The notion of constriction is not suitable for vowels such as [œ, ɔ, æ, ɛ] which involve a open vocal tract.

TOWARDS A LINK BETWEEN CONSTRICTION AND TONGUE POSITIONS.

The analysis of the relations between the acoustic representation of the vowels, based on the formants, and both other geometrical representations (tongue position *versus* constriction) underlines a paradox: the traditional represention is quite isomorphic with the acoustic one, whereas it is not the case for the one based on the constriction, although theoretically linked to the acoustics. Hence it appears that it should be possible to preserve the traditional phonetic representation and to link it to the one based on constriction properties.

In fact, for the majority of the vowels the problem is really simple: the position of the highest point of the tongue in the palatal part corresponds to the constriction point. Only for vowels [æ, ɐ, ɒ, ạ, ạ, a, ɑ, ɒ] divergences are observed between both descriptions.

The tongue is not compressible. Hence, as long as there is no contact between tongue and hard vocal tract walls, it is likely that vertical movements in the palatal part could be correlated with tongue shape changes in the pharynx. For a quantitative evaluation of this correlation, we propose an analysis of natural speech X-ray data: coordinates (X_h, Y_h) of the highest point of the tongue are analyzed in relation with those of the most backward point in the pharynx (X_{ph}, Y_{ph}).

519 X-ray images, available at the Phonetic Institute in Strasbourg [11] and corresponding to 10 sentences produced by a female speaker, have been labelled, at the center of the utterances for vowels and fricative consonants, and at the center of the closure phase for stop consonants. Vocal tract outlines were then superimposed onto a grid [7], defining thus two 32 dimensional vectors for each sagittal view. For each frame, the values of X_h, Y_h, X_{ph} and Y_{ph}, were automatically extracted by a parabolic interpolation between the samples on the grid.

Stop consonants [t, d, n] were not considered for this study, because the vertical movement of the tongue is more due to tongue tip than to tongue body displacements. In the same way, stop consonants [k, g] were also not considered, since the tongue trajectory is limited by the contact with the palate. Correlations (R^2 values corresponding to the percentage of variance explanation) for the 490 remaining frames are shown in the table on figure 1 (all values are significant at .0001). The position of the most backward point of the tongue is obviously correlated with the position of the highest point: if the tongue body moves backward, X_h and Y_h vary simultaneously; if the tongue dorsum goes down, the pharyngeal area becomes smaller.

A principal component analysis on the same X_h, Y_h, X_{ph} and Y_{ph} values proposes two principal axes, that explain 92% of the variance. Figure 2 shows, after a Kaiser rotation, the distribution, in the palatal (X_h Y_h) and in the pharyngeal (X_{ph} Y_{ph}) parts, of the points of the tongue associated with same values for factors 1 and 2. It can thus be observed that each factor describes a specific movement in the palatal part, inducing respectively a horizontal and a vertical movement in the pharyngeal part. Figure 3 displays this phenomenon more precisely.

relationships	R^2
$X_{ph} = f(X_h)$	0.64
$X_{ph} = f(Y_h)$	0.54
$X_{ph} = f(X_h, Y_h)$	0.78
$Y_{ph} = f(X_h, Y_h)$	0.49

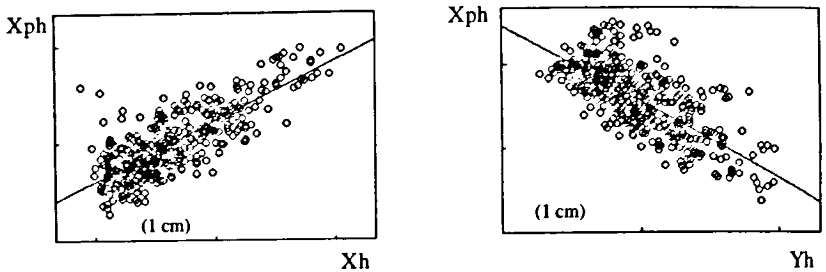

Figure 1 – (a) Correlations between the vertical position of the highest point of the tongue (X-axis) and the horizontal position of the most backward point in the pharynx (Y-axis)
(b) Correlations between the horizontal position of the highest point of the tongue (X-axis) and the horizontal position of the most backward point in the pharynx (Y-axis)
(measurements from X-ray data collected at Phonetic Institute in Strasbourg).

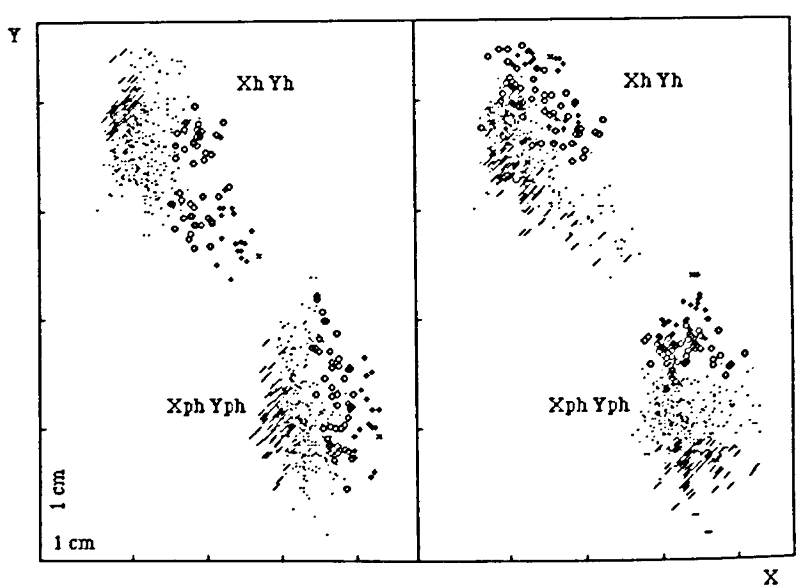

Figure 2 – Distribution of the highest point (X_h Y_h) and the most backward point (X_{ph} Y_{ph}) of the tongue associated with to a same value of factor 1 (left panel)and of factor 2 (right panel).

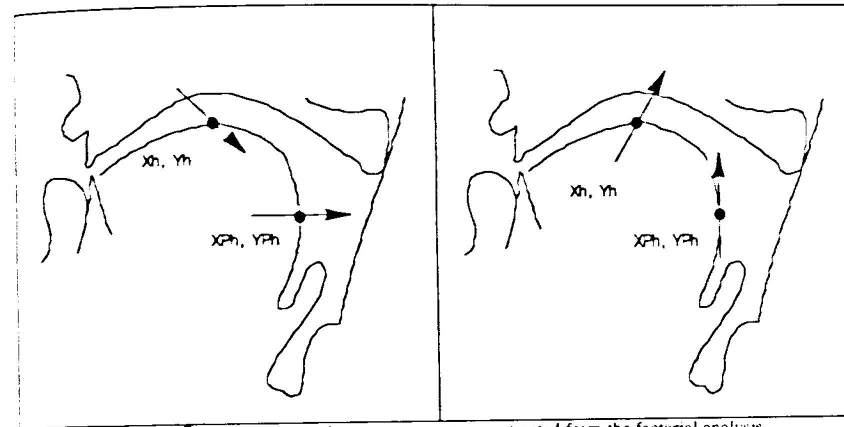

Figure 3 – The two articulatory movements as extracted from the factorial analysis of X-ray data for natural speech:
- A movement of the highest point of the tongue at 45 degrees with respect to the vertical line induces a horizontal displacement of the most backward point of the tongue.
- A movement of the highest point of the tongue at -30° with respect to the vertical line induces a vertical displacement of the most backward point of the tongue.

CONCLUSION

If the structural correlations between tongue dorsum and tongue root are taken into account, the traditional vowel representation *(high/low, front/back, rounded/unrounded)* is acoustically interpretable. It is thus possible to unify articulatory and acoustic representations. Tongue dorsum lowering entails a root movement towards the pharyngeal walls. The two remaining degrees of freedom *high/low* and *front/back* are clearly related to the characteristics of the constriction zone, either palatal, velar or pharyngeal.

The line of research started by Maeda and Honda [12], which associates muscular control and articulatory gestures, should enable thorough examination of vowel representation involving considerations on speech production control .

REFERENCES

[1]**Hellwag C.F.** (1781) De formatione loquelae. Dissert. Tübingen. Traduction française M.-P. Monin in *Bulletin de la Communication Parlée*, ICP, Grenoble, 1, 108 p.

[2] **Joos M.** (1948) *Acoustic Phonetics*. Waverly Press, Baltimore.

[3]**Delattre P.** (1948) Un triangle acoustique des voyelles orales du français. *French Review*, 21, 477-484.

[4]**Maddieson I.** (1986) *Patterns of Sounds*. Cambridge Univ. Press. Cambridge.

[5]**Dunn H.K.** (1950) The Calculation of Vowel Resonances and an Electrical Vocal Tract. *J. Acoust. Soc. Am.*, 22, 740-753.

[6]**Fant G.** (1960) Acoustic Theory of Speech Production. Mouton, The Hague.

[7]**Maeda S.** (1990) Compensatory Articulation during Speech : Evidence from the Analysis and Synthesis of Vocal-Tract Shapes using an Articulatory Model. In W.J. Hardcastle & A. Marchal. (Eds.) *Speech Production and Speech Modelling* (pp.131-150). Dordrecht, The Netherlands: Kluwer .

[8]**Boë L.-J., Perrier P. & Bailly G.** (1992) The Geometrical Vocal Tract Variables Controlled for Vowel Production : Proposals for Constraining Acoustic-to-Articulatory Inversion. *J. of Phonetics*, 20, 27-38.

[9]**Wood S.A.J.** (1979) A Radiographic Analysis of Constriction for Vowels. *J. Phonetics*, 7, 25-43.

[10]**Gay T., Lindblom B.E.F. & Lubker J.** (1981) Production of Bite-Block Vowels : Acoustic Equivalence by Selective Compensation. *J. Acoust. Soc. Am.* 69, 802-810.

[11]**Bothorel A., Simon P., Wioland F. & Zerling J.-P.** (1986) *Cinéradiographie des voyelles et des consonnes du français*. Publications de l'Institut de Phonétique, Strasbourg.

[12]**Maeda S. & Honda K.** (1994) From EMG to Formant Patterns of Vowels : the Implication of Vowel Systems Spaces. *Phonetica*, 51, 17-29.