Prediction of Vowel and Consonant Place of Articulation

René Carré and Maria Mody
Département Signal, Unité Associée au CNRS,
ENST
carre@sig.enst.fr, mody@sig.enst.fr

Abstract

A deductive approach is used to predict vowel and consonant places of articulation. Based on two main criteria, viz. simple and efficient use of an acoustic tube, along with maximum acoustic dispersion, the Distinctive Regions Model (DRM) of speech production derives regions that closely correspond to established vowel and consonant places of articulation.

1 Introduction

Attempts to explain speech production data often suffer from circularity. That is, the classification of data tends to be purely descriptive without shedding light on the processes that give rise to them. In criticizing the data-based approach, Lindblom (1990) favors, instead, the use of substance-based approaches. Here, deductions are derived from specific criteria or from characteristics of the speech production and/or perceptual systems in order to explain the data. In fact, researcher have successfully exploited the substance-based approach in explaining vowel systems, using a perceptually-driven maximum acoustic dispersion criterion (Liljencrants and Lindblom, 1972; Lindblom, 1986). But this approach, too, may be thought of as being circular, to some extent, in that the criterion used may be viewed as, merely, a more encompassing description of the vowels themselves. The crucial task then is to explain the characteristics of the speech perception/production system itself. For example, what is it about the human speech apparatus that yields three and not four or five places of articulation? Is the vocal tract designed primarily for feeding or does it reflect adaptations specialized for speech? To answer some of these questions, we undertook to investigate what deformations (i.e. vocal tract configurations/shapes) of a simple acoustic tube, open at one end and closed at the other, and modeled after a male vocal tract, will yield (i) a good acoustic communication device, and (ii) an acoustic repertoire that closely matches the phonetic richness of human speech.

2 Criteria for building an acoustic device for communication

Our approach will be based purely on physical laws and communication theory principles, and not on any specific characteristics of the human production and/or perception systems. The acoustic device is shown in Fig.1. It is a closed-open tube corresponding to the vocal folds at one end and to the lip opening at the other, respectively. Commands deform the area function of the acoustic tube consequently producing changes in the acoustic signal. These commands constitute our phonological repertoire and may be viewed as analogous to the 'gestures' proposed by Browman and Goldstein (1989).

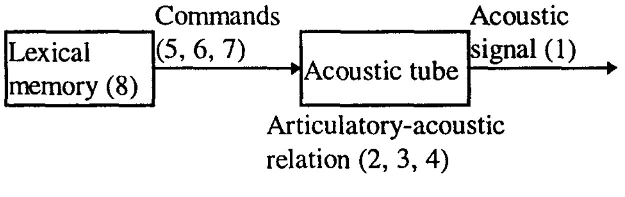


Figure 1. Acoustic device deformed by commands.

The criteria used to build a 'good' communication device are the following:

- acoustic contrast: maximization of the acoustic contrast of the sounds produced (1) in order to have a good signal-to-noise ratio. The spectral peaks (formants) having maximum energy are retained as acoustic parameters;

- least effort: this refers to the efficiency of the articulatory-acoustic relations (2) with monotonicity (3) and orthogonality (4) of the relations. The use of an efficiency criterion corresponds to increasing the role of the dynamics;
- simplicity: with respect to the commands of the deformation of the acoustic tube, the commands must be simple (5) (straight deformations), few in number (6), and with a reduced number of degrees of constriction (7). This is based on the assumption that fewer commands make for smaller demands on memory resources (8) that may be engaged in the learning and eventual mastery of these commands in the phonological acquisition process.

An algorithm to automatically and efficiently deform the area function of an acoustic tube in order to increase or decrease the frequency of a formant or combination of several formants has been proposed elsewhere Carré et al. (1994; 1995). As an example, figure 2 shows the automatic evolution of the area function of the tube from a closed-open neutral configuration, for increasing and decreasing F_2.

Four main regions naturally emerge which are not of equal length. The /a/ vowel is an automatic consequence of a back constriction associated with a front cavity, and, the /i/ vowel of a front constriction associated with a back cavity (anti-symmetrical behavior), a pharynx cavity, thus, being automatically obtained. If, however, the initial configuration is that of a tube closed at both ends (i.e. closed-closed) the /u/ vowel is automatically obtained with a central constriction (symmetrical behavior). A summary of the main conclusions that may be drawn from the above manipulations are as following:

- configurations using the maximum acoustic contrast criterion correspond to those for the three vowels /a, i, u/ of the vowel triangle;
- the deformation of the tube is minimum, because of the use of the sensitivity function (Fant and Pauli, 1974) and thus efficient;
- the deformation commands (or gestures) are simple (recti-linear), limited in number (only one in the case of figure 2, for the making of a back constriction is automatically associated with a front cavity and vice-versa, as in human), and applied at specific places called distinctive regions.

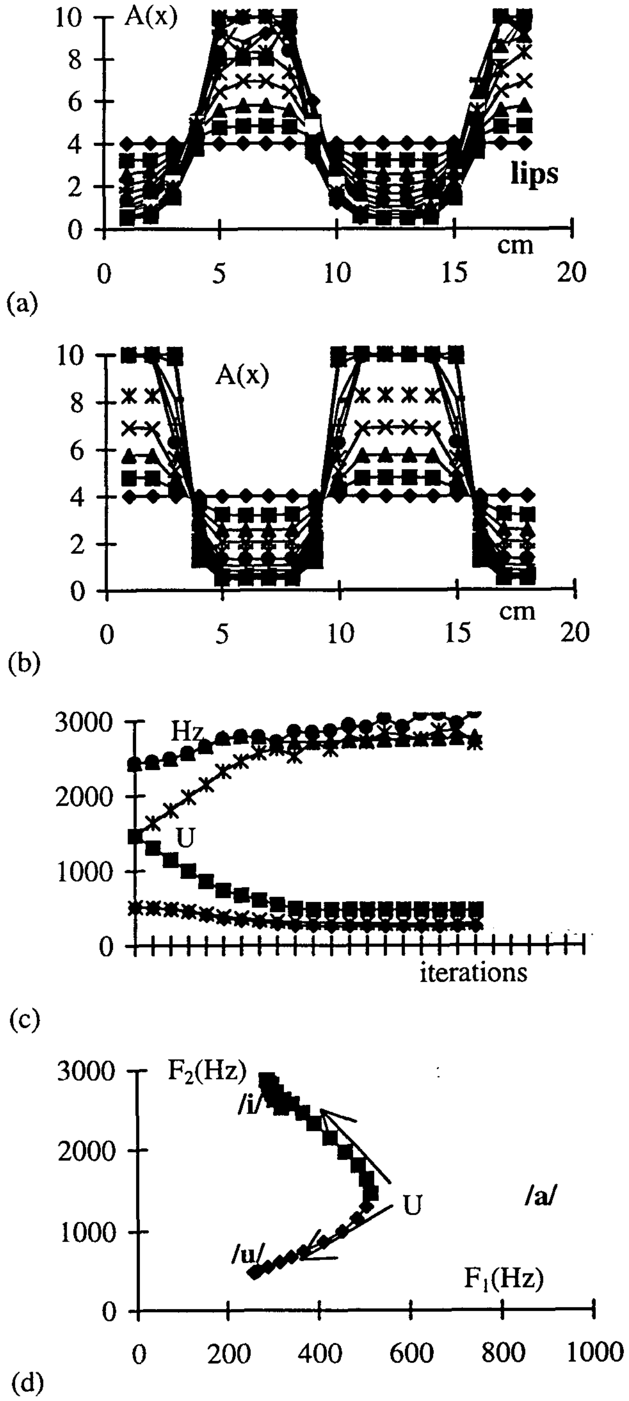
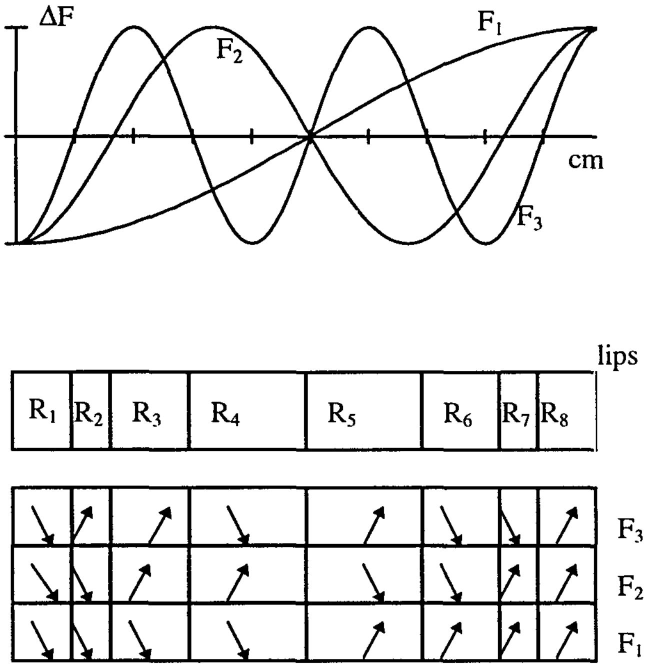


Figure 2. (a)-(b) Iterative deformation of the area function A(x) of a uniform acoustic tube (U) for increasing and decreasing (upper part) F_2.
(c)-(d) Corresponding three formant variations and formant trajectories in the F_1-F_2 plane (lower part).

In summary, characteristics important and integral to a 'good' acoustic communication device were obtained when an acoustic tube underwent specific deformations. It is this type of tube, so-structured in regions, that forms the basis of the Distinctive Region Model (DRM) (Carré and Mrayati, 1992; Mrayati, et al., 1988). Deformation gestures and efficient places of deformation (the regions) are thus deduced from acoustic theory.

3 The Distinctive Region Model (DRM)

The DRM model proposed in 1988 (Mrayati, et al., 1988) is structured in regions, the limits of which correspond to the zero-crossings of the sensitivity function computed on a uniform closed-open tube (figure 3).



	R1	R2	R3	R4	R5	R6	R7	R8
F3	↘	↗	↗	↘	↗	↘	↘	↗
F2	↘	↘	↗	↗	↘	↘	↗	↗
F1	↘	↘	↘	↘	↗	↗	↗	↗

Figure 3. Sensitivity functions (formant variations ΔF for local positive cross-sectional area perturbations) for the first three formants in the case of a uniform closed-open tube (upper part). The limits of the 8 regions of the DRM model correspond to the zero-crossings of the sensitivity functions (central part). Corresponding positive or negative formant variations for increases in the cross-sectional area of the regions (lower part).

The DRM is a 2-region model when only the first formant is controlled. It is a 4-region model when the first two formants (F_1 and F_2) are controlled (note that this 4-region structure is automatically obtained with a maximum acoustic contrast algorithm - see figure 2). It is an 8-region model when the first three formants (F_1, F_2, F_3) are controlled. Thus, increased control of the number of formants entails an increase in complexity and topological accuracy. Varying the cross-sectional area of any one region maximizes the formant frequency variations, with each binary combination of a positive or negative formant variation corresponding to a specific region. Thus, it is hypothesized that the regions of the model could represent the best places of articulation for vowels and consonants.

Critics of this model, may claim that defining regions from the sensitivity functions for a uniform tube would limit the functioning domain for small perturbations. But synergetic command of the regions allows one to maintain monotonic formant variations as well as the region structure, as shown in figure 2.

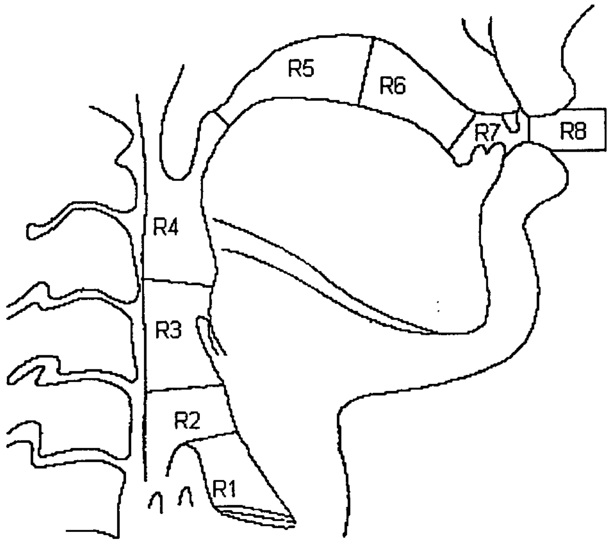


Figure 4. Vocal tract (from Perkell (1969)) and the eight regions of the model.

Note that the region boundaries correspond to a percentage of the total length of the tube. Since the physical length of the acoustic tube is around 19.5 cm for the vowel /u/ and 17.5 cm for the vowel /i/, the place of articulation would thus move according to the pronounced vowel. However, it is the 'effective' length that one must take into account ; for, in addition to the physical length, Lp, effective length incorporates the length, Lr, which corresponds to the radiation effect. Now then, Lr is more or less proportional to the lip opening. For the vowel /i/, Lr=2 cm, and for the vowel /u/, Lr=0 cm. Thus the effective length remains almost constant (Mrayati, et al., 1990) and the places of articulation, fixed, with region 7 always corresponding to the teeth (figure 4). Region 8 (with the radiation effect) corresponds to the lips, regions 3, 4, 5, 6 to the tongue, the region 1 to the larynx. Regions 2 and 7, of small length, are acoustically unimportant.

As such, a closed-open model may not be employed for an /u/ production because the vocal tract is closed at the source and more or less closed at the lips. A closed-closed model would be more convenient. The behavior of such a model is symmetrical, allowing for a central constriction to be automatically derived (Carré and Mrayati, 1992).

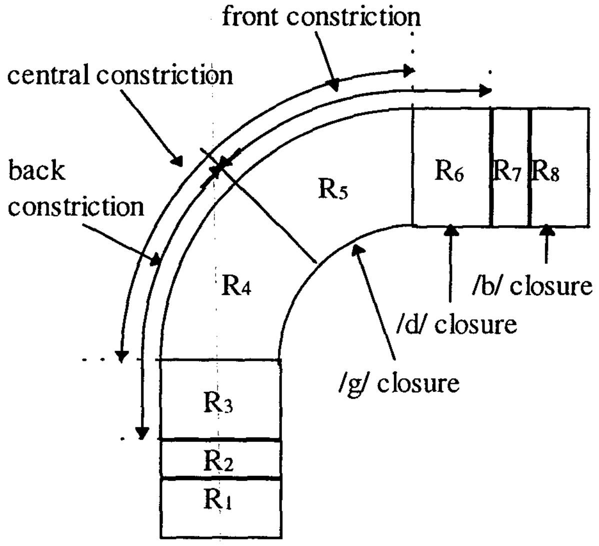


Figure 5. Places of articulation predicted by the closed-open DRM model. The central constriction is predicted by the closed-closed model.

The places of articulation predicted by the model as the best places for maximum formant variations are represented in figure 5. It can be observed that the three main places of articulation for vowel production may be predicted with the 4-region model (taking into account the first two formants). The closed-open model predicts the front constriction and the back constriction; the closed-closed model predicts the central constriction. For consonant production, the 8-region model is needed (taking into account the first three formants). Thus, prediction of consonant production requires a more complex model than that for the vowel production (insofar as the number of degrees of constriction is not taken into account).

4 Vowel places of articulation

The DRM model with its places of articulation was used to produce vowels. For example, figure 6 shows the execution of a command (using a closed-open DRM model) to pass from a front constriction to a back constriction along with a labial command. The acoustic results of these commands are also shown in the F_1-F_2 plane. The positions of the vowels obtained with such a closed-open model are given. Thus, to produce vowels the first two formants appear to be sufficient, the third one either being deduced from the first two (Ladefoged and Harshman, 1979) or is a speaker specific characteristic (Boulogne, et al., 1973).

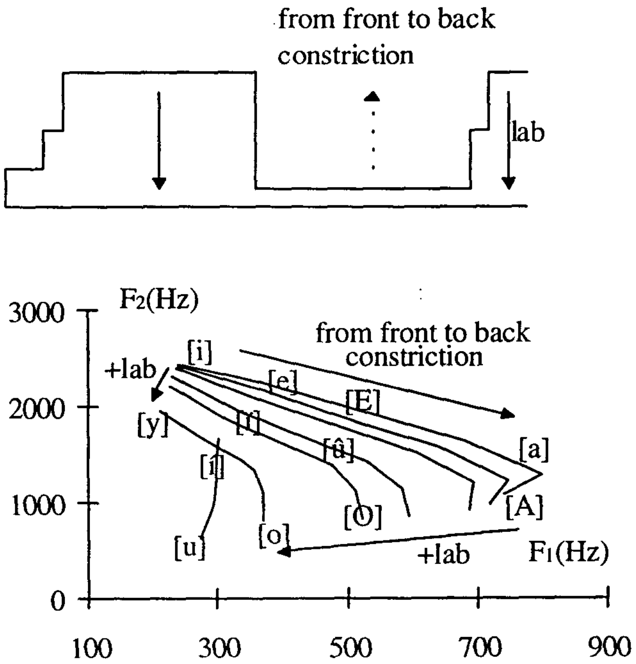


Figure 6. Commands of the model and corresponding formant trajectories in the F_1-F_2 plane. The trajectory /uɨ/ was obtained with a central constriction (closed-closed model).

The consequences of a change from a closed-open to a closed-closed configuration of the DRM model have been studied elsewhere (for more details, see Carré and Mrayati (1995)). Intermediate places of articulation between central and back and central and front are obtained. This approach allows for a good prediction of vocalic systems (Carré, 1996).

5 Consonant places of articulation

Figure 7 shows the formant variations associated with closing and opening of the eight regions of the DRM model with its uniform configuration. Using the places of articulation defined by this model (taking into account the first three formants), we were able to obtain the formant transitions measured by Öhman (1966) (Carré and Chennoukh, 1995). The use of the region 7 gives rise to the production of consonants perceived as labial. Thus, the classification proposed by Mrayati et al. (1988, figure 18) has been revised. Considering the vowel /E/ as close to the neutral partly contributed to the misclassification. The patterns of Delattre (1955) corresponding to /EcE/ where c is a plosive consonant are obtained by the model when only regions 5, 6 and 8 are used. The regions 3 and 4 of the back part of the model correspond to the places of articulation of pharyngeal plosives (Al Dakkak, et al., 1994).

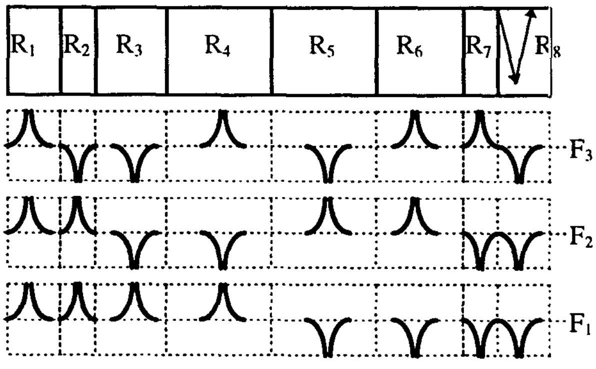


Figure 7. Variations of the first three formants corresponding to a closing-opening gesture in a specific region.

Since the regions of the DRM model can be obtained from the formant variations of a uniform tube and since these regions indeed correspond to the places of articulation, we may hypothesize the following:

- the bursts (Blumstein and Stevens, 1979; Stevens and Blumstein, 1981) associated with the plosives are consequences of closure-opening actions. Indeed, it has been shown that if the bursts are important, they are not determinant for plosive identification in comparison with the formant transitions (Kewley-Port and Pisoni, 1983; Walley and Carrell, 1983).
- the fricative consonants which are generally characterized by their noise sources and not by their formant transitions have to be studied according to another criterion.
- the third formant may be essential for differentiating certain consonants like /d/ and /g/ (the slopes of the first and second formant would not be sufficient). Indeed, it is noted in Harris et al. (1958) that the consonant /d/ cannot be obtained in the vocalic context /i/ without the third formant. The authors concluded that « The effects of third formant cues are independent of the two-formant patterns to which they are added. When a third formant cue enhances the perception of a particular phoneme, it typically does not do so equally at the expense of the other response alternatives ». Similarly, in categorical perception experiments, Godfrey et al. (1981) studied the boundary between /d/ and /g/ by changing the rate and direction of the third formant only, while the first two formants were changed for a /b/-/d/ contrast. Additional experiments have to be undertaken to further explore this point.
- the use of a uniform tube allowed for all combinations of formant variations (positive and negative), that form the basis of the DRM structure, to be revealed. Does this imply that perceptual identification is carried out with reference to a uniform state? This point is worth studying further.

6 Conclusions

The main places of articulation for vowels and plosive consonants can be predicted from a criterion of efficiency of articulatory-acoustic relations. For vowels, a 4-region model (taking into account the first two formants) is needed, and for the plosive consonants, an 8-region model (taking into account the first three formants). This has important implication for the role of complexity in the production of vowels versus consonants.

If the regions 5, 6, 8 are the best places for maximum formant variations (maximum dynamic acoustic contrast) in the case of non-nasal closure-opening, how does one explain the production of nasal consonant having the same places of articulation? For nasal consonants, it is not evident that the maximum dynamic contrasts are preserved. Perhaps, simplicity of production (same places of articulation) may be the more important criterion if the dynamic acoustic contrasts are sufficient? Additional experiments have to be undertaken to know if this apparently optimal acoustic behavior of the human speech apparatus represents adaptations specialized for communication, i.e., is it a consequence of communication theory (ecological point of view) or is it just a coincidence? Is it that speech is just an overlaid function on a system designed primarily for feeding? The answer to this question will determine the importance and validity of our deductions.

In summary, the DRM model uses as its basic elements, control commands, that allow one to deduce the places of articulation for vowels and consonants thereby uncovering the physical bases of phonological distinctions. Furthermore, speech synthesis using this model have yielded high quality results (Hill, et al., 1995).

References

O. Al Dakkak, M. Mrayati and R. Carré. 1994. Transitions formantiques correspondant à des constrictions réalisées dans la partie arrière du conduit vocal. *Linguistica Communicatio,* VI: 59-63.

S.E. Blumstein and K.N. Stevens. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.,* 66: 1001-1017.

M. Boulogne, R. Carré and J.P. Charras. 1973. La fréquence fondamentale, les formants, éléments d'identification des locuteurs. *Revue d'Acoustique,* 6: 343-350.

C.P. Browman and L. Goldstein. 1989. Articulatory gestures as phonological units. *Phonology,* 6: 201-252.

R. Carré. 1996. Prediction of vowel systems using a deductive approach. In *Proceedings of the ICSLP 96*, pages 434-437, Philadelphia.

R. Carré and S. Chennoukh. 1995. Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gesture. *J of Phonetics,* 23: 231-241.

R. Carré, B. Lindblom and P. MacNeilage. 1994. Acoustic contrast and the origin of the human vowel space. *J. Acoust. Soc. Am.,* 95: S2924.

R. Carré, B. Lindblom and P. MacNeilage. 1995. Rôle de l'acoustique dans l'évolution du conduit vocal humain. *Comptes Rendus de l'Académie des Sciences, Paris,* t. 30, série IIb: 471-476.

R. Carré and M. Mrayati. 1992. Distinctive regions in acoustic tubes. Speech production modeling. *J. d'Acoustique,* 5: 141-159.

R. Carré and M. Mrayati. 1995. Vowel transitions, vowel systems, and the Distinctive Region Model. In C. Sorin, J. Mariani, H. Méloni and J. Schoetgen, Editors, *Levels in Speech Communication: Relations and Interactions*. Elsevier: Amsterdam.

P.C. Delattre, A.M. Liberman and F.S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.,* 27: 769-773.

G. Fant and S. Pauli. 1974. Spatial characteristics of vocal tract resonance modes. In *Proc. of the Speech Communication Seminar.* Almqvist & Wiksell: Stockholm.

J.J. Godfrey, A.K. Syrdal-Lasky, K.K. Millay and C.M. Knox. 1981. Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology,* 32: 401-424.

K.S. Harris, H.F. Hoffman, A.M. Liberman, P.C. Delattre and F.S. Cooper. 1958. Effect of third-formant transitions on the perception of the voiced stop consonants. *J. Acoust. Soc. Am.,* 30: 122-126.

D. Hill, L. Manzara and C.R. Taube-Schock. 1995. Real-time articulatory speech-synthesis-by-rules. In *Proc. of AVIOS'95*, San Jose.

D. Kewley-Port and D.B. Pisoni. 1983. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *J. Acoust. Soc. Am.*, 73: 1779-1793.

P. Ladefoged and R. Harshman. 1979. Formant frequencies and movements of the tongue. *UCLA Working Papers in Phonetics*, 45: 39-52.

J. Liljencrants and B. Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language,* 48: 839-862.

B. Lindblom. 1986. Phonetic Universal in Vowel Systems. In J. J. Ohala and J. J. Jaeger, Editors, *Experimental Phonology*. Academic Press: Orlando.

B. Lindblom. 1990. On the Notion of "Possible Speech Sound". *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm (PERILUS),* XI: 41-63.

M. Mrayati, R. Carré and B. Guérin. 1988. Distinctive Region and Modes: a new theory of Speech Production. *Speech Communication,* 7: 257-286.

M. Mrayati, R. Carré and B. Guérin. 1990. Distinctive regions and modes: articulatory-acoustic-phonetic aspects. A reply to Boë and Perrier comments. *Speech Communication,* 9: 231-238.

S. Öhman. 1966. Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am.*, 39: 151-168.

J. Perkell. 1969. *Physiology of speech production. Results and implications of a quantitative cineradiographic study.* The MIT Press: Cambridge.

K.N. Stevens and S.E. Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In P. D. Eimas and J. L. Miller, Editors, *Perspectives on the study of speech.* Erlbaum: Hillsdale.

A.C. Walley and T.D. Carrell. 1983. Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *J. Acoust. Soc. Am.*, 73: 1011-1022.