

PREDICTION OF VOWEL SYSTEMS USING A DEDUCTIVE APPROACH

René Carré

Département Signal, Unité Associée au CNRS,
ENST, 46 rue Barrault, 75634 Paris cedex 13.
e-mail: carre@sig.enst.fr

ABSTRACT

A deductive approach is developed to predict vocalic systems. First, vowels are proposed from an efficient and simple use of an acoustic tube. Then, a maximum acoustic dispersion criterion is applied to classify the obtained vowels.

1. INTRODUCTION

Vowel systems have been widely studied. Data on languages of the world have been accumulated [1], [2], and the sounds have been phonologically labeled and classified. It has been observed statistically that vowel systems have regularities. These regularities include the following:

- the vowels of systems with 3, 4, 5, ... vowels statistically are either /a, i, u/, /a, i, u, ε/, /a, i, u, ε, ɔ/,... or /a, i, u, /a, i, u, i/, /a, i, u, i, ε/,... Indeed, the vowel systems can be divided into two main classes: one with the central vowel /i/, and the other without, and they are obtained by the successive addition of a new vowel [1];
- the vowels can be classified into: labials or non-labials, nasals or non-nasals, ATR, retroflexes, diphthongs, etc.;
- the vowels can be classified according to their complexity [3], [4], [5]. The first vowels of any systems are initially differentiated by their places and degrees of tongue constriction, then by the use of a distinctive labial or nasal gesture, etc.;
- the vowel systems can be well-balanced or not [2];
- the vowel systems tend to make maximum use of the features describing them [6].

All these observations, like any data, cannot be fundamentally explained without circularity. Existing literature attempts to explain the observations from criteria or systems outside the data. Such explanatory approaches are called ‘substance-based’ [7] or deductive. Particularly, attempts were made to explain the observed characteristics of vowel systems from intrinsic properties of either the human speech production system or the perception system or both:

- in production, the stability zones of the articulatory-acoustic relation (quantal theory [8]) have been proposed by Stevens to explain the vowel production characteristics.

This criterion corresponds generally to spectral focalization, i.e. closeness of two consecutive formants [9]. As for ten Bosch [10], he used a least effort criterion;

- in perception, Lindblom [7], [11] hypothesized that vowel systems are built according to maximal perceptual contrast applied in a vowel space (F_1 - F_2 plane) obtained from an articulatory model;
- more recently, Vallée [9] tried to explain the vowel systems from characteristics of the vocal tract (quantal stabilities and focalization) and of the perception system (perceptual contrast).

However, all these interesting approaches cannot give fully satisfactory explanations. The quantal stability is not an intrinsic characteristic of the vowels. If the vowels /a, i, u/ are more or less stable, it is because they represent the extreme acoustic capabilities of the vocal tract production. Recall that the 4th most common vowel, /ε/, is not stable. Furthermore, the focalization criterion seems to be useful to explain the production of certain /y/ but not for other labial vowels like /ø, œ/. On the other hand, the maximum perceptual contrast criterion drives the vowels to change locations in the F_1 - F_2 plane with the addition of a new vowel consequently always yielding a well-balanced system; moreover, with such a criterion, more vowels appear on the axis /iu/ than observed in data, at the loss of central vowels.

In conclusion, we think that the preceding approaches does not take into account the need for **simplicity and efficiency** in the production of a **succession** of sounds. Indeed, a close examination of vowel systems shows that passages from one vowel to another, in the case of vowel systems having less than 7 or 8 vowels, need one distinctive gesture only [5]. Thus, it is possible to classify the vowel systems according to the number of distinctive gestures controlling the vocal tract as the simplicity criterion. On the other hand, it has to be noted that a characteristic of stability in the articulatory-acoustic relation is the exact opposite of the characteristic of least effort, and thus of efficiency. These last remarks lead us to adopt a more radical approach to the study of the problem: viz. to try and predict vowel systems by building a ‘good’ acoustic device which, through simple and efficient manipulations, maximize the number of possible acoustic variations to include the most significant ones. The approach based purely on physical laws and communication theory does not take into account characteristics of the production or/and perception human systems. The acoustic device is shown in Fig.1. Commands deform the area function of the acoustic tube (17.5cm length) which produces an acoustic signal. The criteria to build a ‘good’ device are the following:

- with regard to the acoustic signal, a maximization of the acoustic contrast of the sounds produced (1) in order to have a good signal-to-noise ratio;
- least effort: as far as the articulatory-acoustic relation is concerned, this corresponds to efficiency of the articulatory-acoustic relation (2) with monotonicity (3) and orthogonality (4);
- simplicity: with respect to the commands of the deformation of the acoustic tube, the commands must be simple (5) (recti-linear?), few in number (6), and with a reduced number of constriction degrees (7).

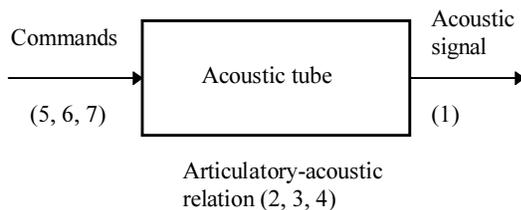


Figure 1: Acoustic device deformed by commands.

The stability criterion of the articulatory-acoustic relation (quantal situations) is not retained here as opposed to the least effort (efficiency) one.

A deductive approach is now applied on the acoustic tube using a criterion of maximum acoustic contrast.

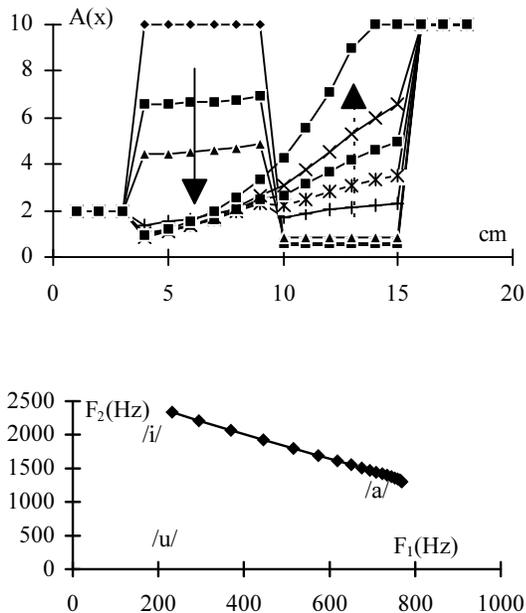


Figure 2: Evolution of the area function of the acoustic tube for an increasing of F_1 from a crude /i/ configuration and the corresponding formant trajectory in the F_1 - F_2 plane.

2. MAXIMUM ACOUSTIC CONTRAST: TOWARDS THE DRM MODEL

An algorithm to automatically deform the area function of an acoustic tube in order to increase or decrease the frequency of any formant or combination of several formant has been proposed [12]. This operation yields the maximum range of formants possible with a tube of a given length (17.5cm). Indeed, the formant variations obtained correspond to a large vowel triangle and the shapes of the tube for the angles of the triangle are crude configurations of the vowels /a, i, u/. Figure 2 shows the automatic evolution of the area function of the tube from a crude /i/ vowel configuration with an increasing F_1 . A larynx cavity is imposed.

The conclusions from this investigation are the following:

- configurations with maximum acoustic contrast are obtained (according to the initial criterion) which indeed correspond to 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) thus efficient, with monotonicity and pseudo-orthogonality of the articulatory-acoustic relation;
- the deformation commands (or gestures) are simple (recti-linear) and limited in number (only one in the case of figure 2, doing a back constriction automatically associated with a front cavity, as in human).

In summary, the important characteristics of a ‘good’ acoustic communication device are obtained when an acoustic tube is specifically deformed. It is this type of tube, so-structured in regions, that forms the basis of the DRM model [13], [14].

3. VOWEL PRODUCTION AND REPRESENTATION

With the preceding device, structured according to the DRM model, a topological approach is first developed in order to produce sounds, in a simple and efficient manner (represented by simple configurations, the passage from one to another being simple to realize and then simple to represent). The DRM model, exploiting the acoustically sensitive characteristics of the tube (between quantal stabilities!), divides the tube into corresponding regions and then uses a simple classification of the produced sounds according to these regions (for example back, front and central constriction). For simplicity reason, the number of degrees of constriction is always reduced as much as possible. Furthermore, a new degree of constriction has to be fully used (maximum use of a new possibility of control) so that the commands of the DRM model are kept topologically well-balanced. The passage between the two main configurations of the DRM model (between closed-open and closed-closed configurations) has been studied elsewhere (for more details, see [15]). For example, figure 3 shows the command of a closed-open DRM model to pass from a front constriction to a back constriction and the 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. The vowels /a, i, u/ are not located at the extreme possibilities of the tube characteristics where the stability situations are encountered.

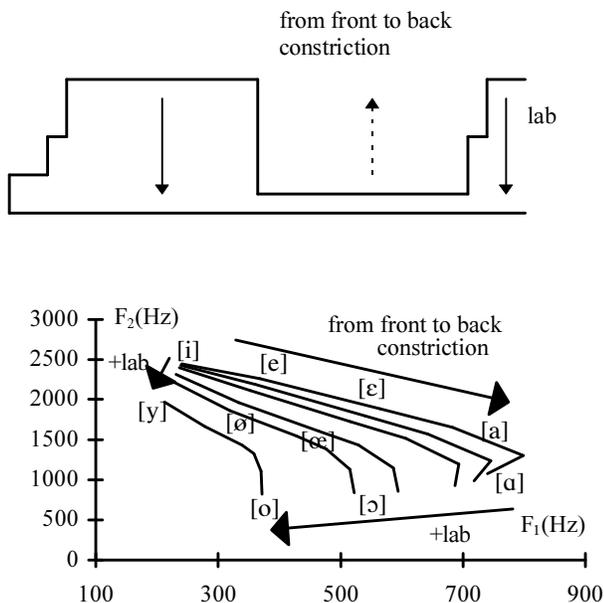


Figure 3: Commands of the model and corresponding formant trajectories in the F_1 - F_2 plane.

Finally, the vowels, produced with a maximum of two commands: displacement gesture of the constriction (with two degrees of constriction of 0.5 and 1.3cm² plus the neutral situation of 4cm²) and labial gesture (one degree of constriction), are represented in figure 4. It is noted that the position of the constriction is topologically located along the tube: at the extremities, at the center, etc. The division of the back part into thirds, results from a

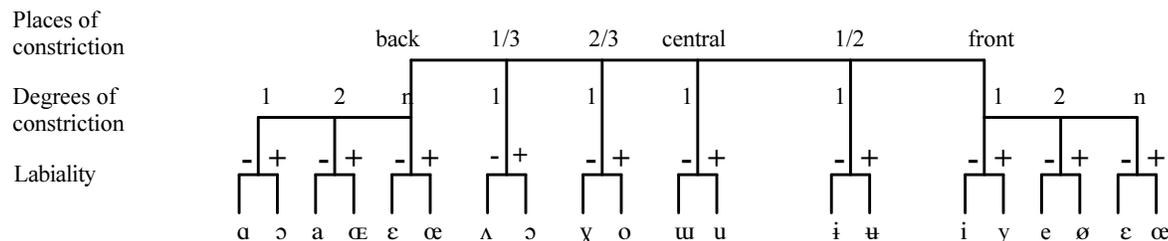


Figure 4: Vowels obtained with the DRM model (closed-closed and closed-open configurations plus passages from one to another). Constrictions 1, 2, n, are respectively: 0.5, 1.3, 4 cm².

5. CONCLUSIONS

Our approach, which consists first in exploiting the capabilities of an acoustic device to produce sounds simply and efficiently, leads to vowel systems having characteristics very closed to the observed data. Particularly, the vowels have their own specific locations in the F_1 - F_2 plane because they are simple to produce and the V-V

closed examination of the acoustic characteristics obtained, the vowel /ɔ/ being more or less acoustically equidistant of /a/ and /u/. The vowel /ɛ/ is here classed either with a back constriction or with a front constriction.

4. PREDICTION OF VOWEL SYSTEMS

To build vowel systems with 3, 4, 5,... vowels, we choose successively from among the vowels previously obtained, ones corresponding to the criterion of maximum acoustic contrast. The first three are /a (or /a/), i, u/, corresponding to maximum acoustic and topological (back, front then central constriction) contrast. Note, whereas the production of /u/ involves 2 commands (i.e. tongue and lips), the latter does not constitute a distinctive gesture due to the absence of a non-labial counterpart (distinctive gesture, as used here, is analogous to a distinctive feature). Two choices then are available (see [1]) depending one's retention or not of the central vowel /i/ (in the middle of the formant trajectory /iu/ and constriction in the middle of the back-central distance). This vowel is very unstable because the labial and constriction gestures do not produce orthogonal trajectories as is the case for /ɛ/ (in the middle of the formant trajectory /ai/ and neutral configuration). This approach yields a progressive development of systems (1, 2, 3,... see figure 5) to include vowels which do not need a distinctive labial gesture. As such, in the system class without /i/, the system with 5 vowels is well-balanced; it also happens to be the most common. With the central /i/, the system is well-balanced with 6 vowels and interestingly this too is the most common in the languages of the world. The inclusion of a new distinctive gesture: the labial one, automatically gives rise to new vowels then: which are the labial vowels /œ/, then /y/, then /ø/, already well located because of our production constraints, and according to the maximum acoustic contrast. In the system class with /i/, it will be necessary to study more the system with 7 vowels (apparition of /e/, /o/, /ə/).

transitions are efficient. The labial vowels have also their own locations according to a criterion of increasing complexity. Our approach explains also why some systems are more common depending on whether they are well-balanced or not. These good agreements strengthen the explanatory power of our approach and also the fact that the human communication system is more or less optimal in terms of efficiency. Furthermore, we can provide an answer about the acoustic locations of the vowels [16] (they have

specific locations because of criteria applied to their production) and about their limited or unlimited number (a new degree in complexity can result in a multiplication by two the possible number of vowels, if this new gesture is efficiently used. This characteristic is an obstacle for small new needs in communication). The pseudo-orthogonality of the articulatory-acoustic relation observed in an acoustic tube divided in distinctive regions (DRM model) explains the good correspondence and the complementary nature, easy to exploit, of the two systemic approaches: topological on the one hand, and acoustic, on the other. Hence, it should be noted that characteristics of acoustic tasks and corresponding deformation gestures of the area function can be deduced from our approach. These characteristics have to be tested in other situations.

Acknowledgments. The author thanks Maria Mody for her helpful and stimulating comments. This research was supported by the Science Program of the European Union.

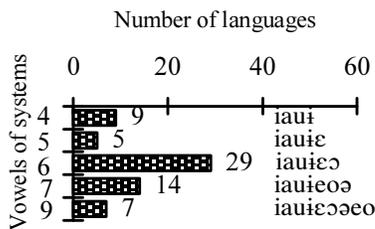
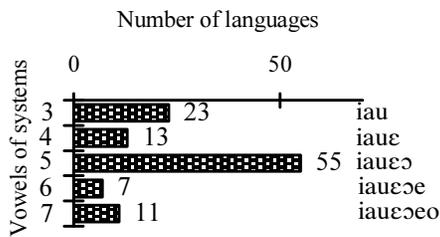
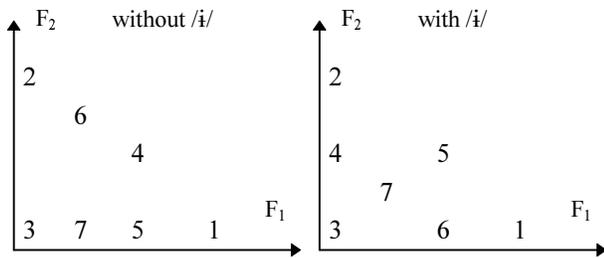


Figure 5: Choice of the successive vowels with /i/ or without (above) and statistics on vowel systems with /i/ or without (below) (after [1]).

6. REFERENCES

1. Crothers, J. "Typology and universals of vowel systems," in *Universals of human language. Vol. 2: Phonology*, ed. J. H.

- Greenberg, C. A. Ferguson and E. A. Moravcsik, 93-152, 1978.
2. Maddieson, I. *Patterns of sounds*, Cambridge University Press, Cambridge, 1984.
3. Lindblom, B. "Phonetic contents in phonology," *Perilus XI*: 101-118, 1990.
4. Carré, R. "'Speaker" and "Speech" Characteristics: a deductive Approach," *Phonetica* 51: 7-16, 1994.
5. Carré, R., Bourdeau, M. and Tubach, J.P. "Vowel-vowel production: the distinctive region model (DRM) and vocalic harmony," *Phonetica* 52: 205-214, 1995.
6. Ohala, J. *Moderator introduction to symposium on phonetic universals in phonological systems and their explanations*, Proc. of the 9th ICPhS, Copenhagen, 1980.
7. Liljencrants, J. and Lindblom, B. "Numerical simulation of vowel quality systems: The role of perceptual contrast," *Language* 48: 839-862, 1972.
8. Stevens, K.N. "The quantal nature of speech: evidence from articulatory-acoustic data," in *Human Communication: a unified view*, ed. E. E. David and P. B. Denes, 51-66, 1972.
9. Vallée, N. *Systèmes vocaliques : de la typologie aux prédictions*, Thèse de Doctorat en Sciences du Langage, Grenoble: Université Stendhal, 1994.
10. ten Bosch, L. *On the structure of vowel systems; aspects of an extended vowel model using effort and contrast*, Dissertation Thesis, Amsterdam: University of Amsterdam, 1991.
11. Lindblom, B. "Phonetic Universal in Vowel Systems," in *Experimental Phonology*, ed. J. J. Ohala and J. J. Jaeger, 13-43, 1986.
12. Carré, R., Lindblom, B. and MacNeilage, P. "Rôle de l'acoustique dans l'évolution du conduit vocal humain," *CR Acad. Sci. Paris t. 30, série IIb*: 471-476, 1995.
13. Mrayati, M., Carré, R. and Guérin, B. "Distinctive Region and Modes: a new theory of Speech Production," *Speech Communication* 7: 257-286, 1988.
14. Carré, R. and Mrayati, M. "Distinctive regions in acoustic tubes. Speech production modeling," *J. d'Acoustique* 5: 141-159, 1992.
15. Carré, R. and Mrayati, M. "Vowel transitions, vowel systems, and the Distinctive Region Model," in *Levels in Speech Communication: Relations and Interactions*, ed. C. Sorin, J. Mariani, H. Méloni and J. Schoetgen, 73-89, 1995.
16. Ladefoged, P. and Maddieson, I. "Vowels of the world's languages," *Journal of Phonetics* 18: 93-122, 1990.