A METHOD FOR THE STATISTICAL TREATMENT OF VOCALIC FORMANT VALUES

L. Cerrato, F. Cutugno, P.Maturi

Centro Interdipartimentale di Ricerca
per l'Analisi e la Sintesi dei Segnali
Università degli Studi di Napoli "Federico II"

ABSTRACT - A method of analysis to represent temporal formant variations and vowel dynamic is proposed in this work.
This method is based on the following steps:
1) LPC analysis of the first two formants of the vocalic portion to be examined;
2) linear regression computation and extraction of linear equations;

3) polar representation of slopes in terms of $r=\sqrt{s_1^2+s_2^2}$ and $\phi=\mathrm{arctg}\left(\frac{s_1}{s_2}\right)$.

INTRODUCTION

In the description of vocalic sounds particular attention is often paid to the variations of formant values; in fact phoneticians usually try to:

a) follow the temporal development of formant variations; hence drawing formant temporal patterns on a time/frequency diagram (Fry,1991, Albano Leoni, Cutugno & Maturi 1992),

b) determine, instant by instant, which vowel can be related to a particular formant pattern; hence representing data as a curve on F1/F2 diagrams (Ladefoged 1982, Peterson & Barney 1952).

It is difficult to describe temporal development and vowel dynamic at the same time without making the representation too complex. Moreover it is difficult to compare diphthongs of the same type and/or different types of diphthongs, and quantify the extent of their diphthongization.

In this paper we wish to introduce a new method for the dynamic representation and the quantification of formant gliding in a temporal pattern. Some phonetic implications of the results will also be described. Among other things, this method enables us to define a parameter indicating the extent of gliding which can be correlated to various segmental and suprasegmental data.

PRELIMINARIES

In order to illustrate the proposed method, we have used speech materials selected from recorded spontaneous conversation of six male English speakers with RP pronunciation. Cutting up the stream of speech into segments by means of a Kay DSP Sonagraph 5500 (using a wide-band filter of 300 Hz to show the formants) we have gathered:

1) vocalic segments representing the five English "pure" vowels: [a:, 3:, Ɔ:, u:, i:] in stressed position;

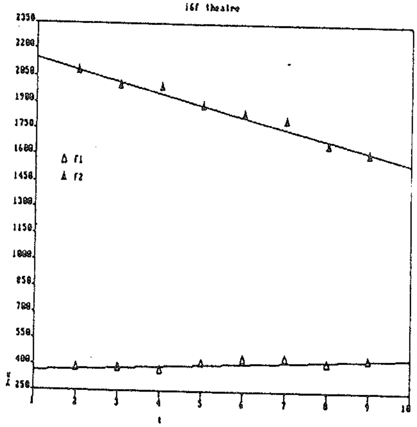
2) vocalic segments representing the centring diphtongs [ɪə;ɛə;ɔə;uə] in stressed position.

THE METHOD

It is based on the following three steps:

a) LPC analysis of the first two formants of each vowel-sound .
The vocalic segments representing long vowels and diphtongs have been analysed by means of LPC procedure (16 coefficients, frame length=1/10 of total segment length, frame advance without overlapping) this way each vowel-sound has been divided into ten bits of the same length . For each frame, only the first two formants values (F1, F2) have been taken into account.
In the elaboration of data the first and the tenth bit have not been taken into account as these are mostly affected by the coarticulation phenomena.

b) Linear regression computation and extraction of linear equation.
Lines best fitting the time patterns of formant values (obtained in the previous step) have been calculated . The standard error, resulting from the study of more than one hundred regression, is sistematically less than 20%, which proofs a satisfying degree of fitting.



-- Fig.1 Schematic representation of a realization of [iə] for speaker M.E. Triangles stand for the output of LPC formant extraction algorithm, while lines are the results of a best fitting procedure.

c) Polar representation of slopes.
The slope coefficients of the linear equation of both formants (calculated in the previous step) have been transferred on a polar diagram. In this kind of diagram each segment is represented as a vector whose length indicates the extent of the diphthongization expressed in terms of:

$$r = \sqrt{s_1^2 + s_2^2}$$

Where s_1 and s_2 are the slope coefficients.

The angle formed by the vector with the positive direction of the x-axis, is expressed in terms of:

$$\phi = \mathrm{arctg}\left(\frac{s_1}{s_2}\right)$$

and varies according to the type of diphthong. Dividing the circle in four quadrants we have that each quadrant classifies a type of diphthong (fig.2):

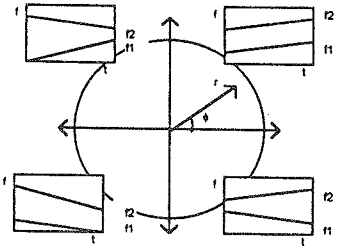


Fig 2. Polar representation of diphthongs
I quadrant 0-90 both increasing formants [u∂]
II quadrant 90-180 F1 decreasingi F2 increasing [Ↄ∂]
III quadrant 180-270 both decreasing formants [e∂]
IV quadrant 270-360 F1 increasing F2 decreasing [ɪ∂]

RESULTS

The extent of the diphthongization can be quantified at first sight as the radius expresses the degree of gliding:

-observing the graphs we can see that non-diphthongized vowels (fig.3) are all situated in the centre or very very close to it being their radii nul or almost equal to zero.

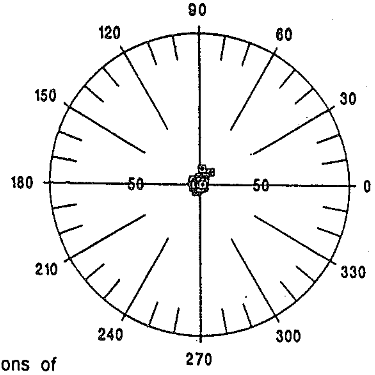


Fig. 3 Polar Representation of all the realizations of [a:, 3:, Ↄ:, u:, i:] uttered by the six analised speakers

-in the case of the diphthongs (fig.4,a,b,c,d) the length of the radius increases according to the increase of the extent of the diphthongization.

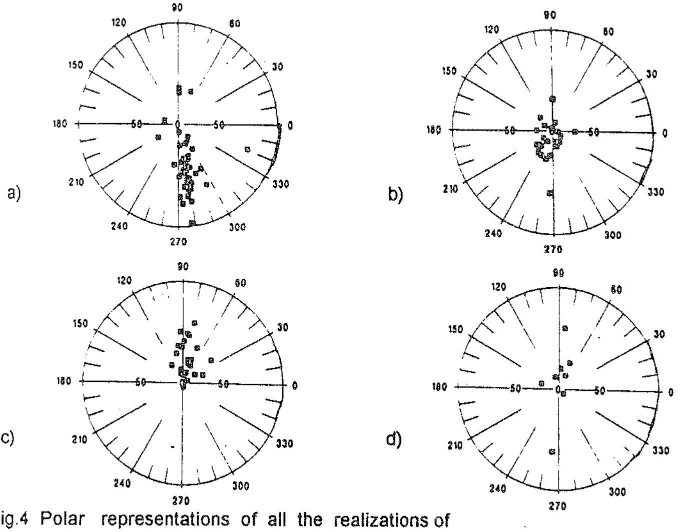


Fig.4 Polar representations of all the realizations of [Iə (a);ɛə(b);ɔə(c);uə(d)] uttered by the six analised speakers

This method offers the following advantages:

- it represents formant variations in time and vowel dynamic simultaneously, somewhat combining the F1/F2 representation, and the time frequency representation: the duration of the analysed phenomenon strongly influences the value of radius, in fact in macroscopical terms we have:

$$s_1 = \frac{\Delta F_1}{\Delta t} \quad \text{and} \quad s_2 = \frac{\Delta F_2}{\Delta t} \quad \text{hence we get} \quad r = \sqrt{s_1^2 + s_2^2} = \frac{\sqrt{\Delta F_1^2 + \Delta F_2^2}}{\Delta t}$$

in other words the radius is the sum of both absolute variation of F_1 and F_2 "weighted" by a factor that is just the duration of the observed phenomenon.

- it compacts data coming from different vocalic sounds in one representation which only shows the formant variations; in fact the differences among the first elements of the diphthongs are vanished as they are all led back to the same starting point: the centre of the circle.

To validate the model proposed, we show some correlations among the extent of the diphthongization and some segmental and suprasegmental parameters, such as duration of the segment being analysed (fig.5), speed of elocution,(fig.6), accented/unaccented (fig.7), given/new.(fig.8).

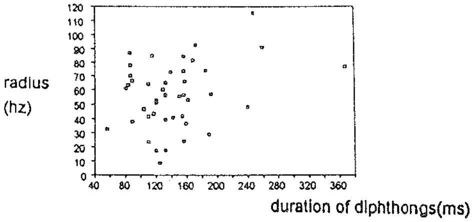


Fig. 5 Correlation between radius and duration of the segment

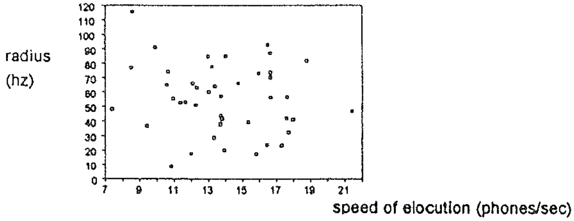


Fig. 6 Correlation between radius and speed of elocution

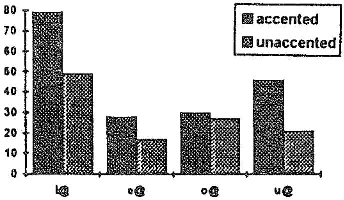


Fig. 7 Histogram showing relations between the radius and the parameter of accentedness

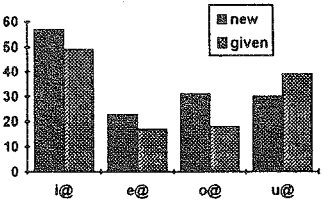


Fig. 8 Histogram showing relations between the radius and the given/new parameter
(please note that in fig. 7,8 [ə] is symbolized by @)

SOME SHORT CONSIDERATIONS

As we have seen, the method shows a clear distinction between long ,"steady" vowels and diphthongs and among different kind of diphthongs (figg.3,4), although some particular cases seem to contradict this trend.

Despite of the linguists' common opinion (Lindblom 1963, 1990) the direct proportionality expected between radius and duration and the inverse proportionality expected between radius and speed of elocution are not regular at all in our spontaneous speech materials (figg. 5,6). The large spreading of data is presumably due to intra- and interindividual variability. In figg. 7,8, on the contrary, the extent of the diphthongization is mainly greater when the item containing the diphthong is "new" (Brown 1983, Halliday 1967) and accented (Liberman & Prince 1977, Nooteboom & Efting1991), just as expected. It seems that the degree of actual diphthongization is more sensitive to functional parameters than to mechanical, articulatory ones.

REFERENCES

Albano Leoni F., Cutugno F., Maturi P. (1992), *Representation of Frequency Variations in Time in Speech Signals*, in M. Cooke & S. Beet (eds), *Visual Representations of Speech Signals* (John Wiley & sons: Chichester), in press;

Brown G. (1983), *Prosodic Structure and the Given/New distinction*, in Cutler A., Ladd D.R. (eds.), *Prosody: Models and Measurements* (Springer & Verlag: Berlin), pp. 67-77;

Fry D.B. (1991), *The Physics of Speech* (Cambridge Univ. Press: Cambridge);

Halliday M.A.K. (1967), *Notes on Transitivity and Theme in English, II"*, J. of Linguistics, 3, pp.149-163;

Ladefoged P.. (1982), *A Course in Phonetics*, p.165 (Harcourt Brace Jovanovic Publishers: New York);

Liberman M.Y., Prince A. (1977), *On Stress and Linguistic Rhythm*, Linguist. Inq., 8, pp. 249-336;

Lindblom B. (1963), *Spectrographic Study of Vowel Reduction*, J. of Acoust. Soc. Am., 35, pp. 1773-1781;

Lindblom B. (1990), *Explaining Phonetic Variations: a Sketch of the H&H Theory*, in Hardcastle W. J., Marchal A. (eds.), *Speech Production and Speech Modelling*, pp. 403-439 (Kluver Academic Press: Dodrecht);

Nooteboom S.G., Efting W. (1991), *The Effect of Accentedness and Information Value on Word Durations: a Production and Perception Study*, Actes du XXII Congrès des Sciences Phonétiques, Aix-en-Provence, 3, pp. 302-306;

Peterson G.E. & Barney H.L. (1952), *Control Method Used in a Study of Vowels*, J. Acoust. Soc. Am. 24, pp.175-184.