PHONOLOGICAL/PHONETIC OPPOSITIONS: BINARY OR GRADUAL? SOME EXPERIMENTAL CONTRIBUTIONS TO THE CURRENT ISSUE BASED ON THE ANALYSIS OF ITALIAN DATA FROM THE POINT OF VIEW OF SPEECH RECOGNITION.

P.Bonaventura(*),L.Prina Ricotti(*),J.Trumper(**)

(*) Fondazione "Ugo Bordoni"
Viale Trastevere, 108 - 00153 Roma, Italy

(**) Dept. of Linguistics
Universita' delle Calabrie, Cosenza, Italy

Abstract- In the present paper some evidence' is given for the existence of a gradual phonetic change in Italian stop consonants from the point of view of their defining distinctive features.

The four features of Mode 1 (Voicing), Mode 2 (Continuity), Place and Timing are assumed to be perceptually effective and are examined from the point of view of their significant correlation.

The experiment used synthetic stimuli for CV groups of phonemes obtained from an acoustic model which allows one to vary continuously the acoustic characteristics associated with the distinctive features that are being examined.

1.Foreword

This paper is based on the assumption that the acoustic or articulatory categories detected on the physical continuum are not homogeneous with the corresponding perceptual ones; in order to define such categories, it is essential to describe the relative variation of the significant parameters (see also [8]) on their own perceptual scales.

We have prepared a table (see table 1) of the relative values of the four most significant perceptual features for Italian consonants: in this system Timing is considered as a feature in itself which varies in combination with the others, but along a typical continuum.

It is thus obtained the "intrinsic" time of each perceptual entity (or "res percepta").

The table shows all the possible ratios between the relative values fixed for the experiment but it could be expanded (within given houdaries) to give a more suitable frame for the complex reality of a natural language.

2.Theoretical approach

Multivalue features scales commonly used (see [3], improvement on [7], etc.) are based on extrapolation from experimental observation on articulatory processes.

Such processes are effectively gradual: their graduality is implicit to the performance time of muscular commands.

The basic assumption of such approaches is that the sum of values (no longer conceived in binary terms) of a fixed number of features that are selected on the bases of their economicalness or "naturalness" defines each phoneme.

Scales are defined by giving fixed values to the beginning and end of possible continua which are segmented into different range groupings according to the language in question: e. g. possible cuts along a place of articulation continuum are:

	0	1	2	
α)	/p/	/t/	/k/	say English, Italian, Finnish.

	0	1	2	3	
β)	/p/	/t/	/c/	/k/	say Albanian.

	0	1	2	3	4	
γ)	/p/	/t/	/k/	/q/	/?/	say Classical Arabic.

We evince from (α) - (γ) that the possible continuum that defines the Front - Back, Frontedness or Place features is composed of the following positions that define phoneme ranges:

0	1	2	3	4	5
p	t	c	k	q	?

However, rather than claim that such levels or possible ranges of a continuum (cut into a graded plane or Gradatum) - and we can allow that levels may be only two (usual for Voicing) at the phonological level, as in English or Italian, or even only one as in Finnish, though n - valued at the phonetic level - are bound to the production level of our model (articulation), we would claim that when in fact we say that the Frontedness or Place features has three levels in Italian we are really referring to the perceptual interpretation of a position of the articulatory-acoustic space, to a process in which speech production categories are mapped on to corresponding perceptual categories.

The scheme we propose to represent the three phases of the whole process involved is as in Diagram I.

We can admit that the gradualness of features is effectively codified at the 2nd LEVEL, where perception takes places on the basis of a set of perceptual features that refer to the constant relations between physical values individuated along a perceptual scale formed of acoustic parameters given in the LEVEL ONE INPUT (i.e. capacity to select given parameters of the human ear). This has a filter function with respect to the acoustic signal and allows for the trasmission of only certain components of the complex signal.

It is at this level that binary choices (see [2]) operate and are observed effectively to operate, though uniquely on the acoustic form of a given feature.

The scope of the present paper is - by means of straightforward perceptual experiment of identification - to give a demonstration of the non-categoricalness (or gradualness) of perception, that, as we shall see, operates on the basis of precise rules connected in the neural topology and functioning.

This renders discrete the continuum of acoustically selected parameters and combines segments obtained as a parameter of time (this parameter at the 2nd LEVEL corresponds to the intrinsic timing factor in [1].

The full set of these perceptual relation (particularly complex - we shall skip over details here, but the question is being studied) furnishes a definition of three phonemes belonging to the class of STOPS based on the reciprocal values of three essential perceptual features that we have so far identified.

Phonemes are organized on the basis of the values evinced from the perceptual scales for each features that describes a phoneme as a res percepta; this organization is schematized in table 1.

Numbers are not numbers in a set a natural numbers, but exist uniquely in a relational plane.

3. Method

We have varied a first parameter (F2) along a continuum composed of constant intervals of 100 Hz each obtaining 13 variable stimuli ranging across three levels identified respectively as the articulatory categories: LABIAL, ALVEO-DENTAL and VELAR.

The V.O.T. of the components of this sequence was also varied in 10 ms steps for negative V.O.T. and in 5 ms steps for positive V.O.T., affecting the first part of the transitions.

On such a sequence a simple labelling test was performed with a set of 18 unexperienced native italian listeners; a straightforward identification test was chosen as in our experience on sequences involving the variation of a single feature ranging within the classes of Stops, Fricatives and Affricates, any kind of discrimination tests, both ABX and 4IAX, gave poor results.

This is due to the procedure of the test in itself that affects perception in presenting too long sequences with respect to the temporal capacity of STM [4], in fact the initial part of the signal are lost and can't be processed in relation to the last parts.

Several other experimental procedures ([8], [10]) seem to give some evidence for a non categorial perception at the acoustic (our 1st) level, while categorical perception at the phonetic (our 2nd) level.

We are actually interested in verifying if the relative degrees attributed to the perceptual features in table 1 play an effective role in the perception process.

From our standpoint it is thus enough to check the labelling thresholds through an identification task.

4. Results and Discussion

The mean results of the identification tests are shown in Diagram 2. We assume that the correct proportion among the amounts of the cues for each perceptual feature involved must be maintained constant to obtain a definite categorization, otherwise the identification scores will not be significant, resulting in misunderstandings or even no labelling at all of the stimuli.

Timing is assumed to be fixed for all the consonantal parts of the syllables. It pertains to each phoneme and is calculated from the relative times of the parameters (see below).

Table 2a shows the values of the actual cues of the three features along the acoustic continuum; Table 2b shows the combinations of the actual values of the perceptual features as obtained from Table 1 with the corresponding labels attached to every stimulus according to their labelling rate in the test.

As shown in Table 2b the phoneme /g/ is characterized by the triple 0 1 0. While Mode 2 is unaltered, the cues of Mode 1 and Mode 3 are changed: Voicing moves, through regular intervals, from value 1 to value 0 and Place from 0 to 3.

The two extremes are characterized in Diagram 2 by the categorization as /ga/ for stimuli n. 1 to n. 5 and as /pa/ for n. 11 to n. 13. In the intermediate range (stimuli n.6 to n.10), the bias to perceive a /ka/ at the 5th and 6th steps can be explained considering that the variation of the cue of Voicing to the value 0 is not yet strong enough to interact with that of Place and to polarize identification in the Alveolar area. In the range of stimuli with a set of values similar to the initial one, only the variation of Voicing is relevant in perception (/ka/: 0 0 0).

Stimuli n.7 to n.10 are either perceived as /da/ or nonidentified at all as a speech sound. Perception of /da/ (triple 2 1 0) is explained as an interpretation of an ambigous Place value 1 (only values 0 to 2 are distinctive for Stops) as a level 2, coupled with a previous level 1 of Voicing

The possible evaluation of a level 1 of Place corresponds to the occurrence of nonexistent triples, either 1 0 0 or 1 1 0 , that determines the exit from the speech mode, given that this situation is not predicted in the subsystem of Italian we already described. The interval where there is a restructuring of unexpected values with a fictitious combination or quite none is considered a black hole in perception.

5. Conclusions

The results are consistent with the assumption that each single amount of any acoustic cue is not relevant in itself to select a feature, as the auditory system filters the signal transmitted by the receptors using a special code, activated when the listener is in the speech mode (see [9]). This code is a structure formed by the relation between the temporal amounts of every significant parameter in input and the whole temporal frame of the actual segment as it is realized. This means that a special rule binds the relative times of every single parameter among them, and with the global time that they share to form a " res percepta "; the correct proportions can be predicted from the relations of a complete model of perception and production (P.Bonaventura,"Preliminary studies for an MCC model of perception", to be published).

This could be a tentative explanation also for the insertion, erasure and reajustement of phonemes along the speech chain.

REFERENCES

[1] Fowler,C.A.: "Coarticulation and Theories of Extrinsic Timing". Report Speech Research SR-57

[2] JaKobson,R. Fant,G. Halle,M.: "Preliminaries to Speech Analysis". Massachussets Institute of Technology Press, Cambridge, Mass., 1963

[3] Ladefoged,P.: "Preliminaries to Linguistic Phonetics". Univ. of Chicago Press, 1971

[4] Massaro,D.W.: "Understanding Language". Academic Press, New York, 1975

[5] Massaro,D.W. Cohen,M.M.: "Categorical or Continuous Speech Perception: a New Test". Speech Commun.,2,1983

[6] Massaro,D.W. Cohen,M.M.: "The Contribution of Fundamental Frequency and Voice Onset Time to the /zi/-/si/ Distinction". JASA,60,1976

[7] Miller,G.A. Nicely,P.: "An Analysis of Perceptual Confusions among some English Consonants". JASA,27,1955

[8] Oden,G.: "Integration of Place and Voicing Information in the Identification of Synthetic Stop Consonants". Journal of Phonetics, 6, 1978

[9] Pisoni,D.B. Lazarus,J.H.: "Categorical and Noncategorical modes of Speech Perception along the Voicing Continuum". JASA,55,n.2,1974

[10] Sawusch,J.R. Pisoni,D.B.: "On the Identification of Place and Voicing Features in Synthetic Stop Consonants". Journal of Phonetics, 2, 1974

Diagram 1

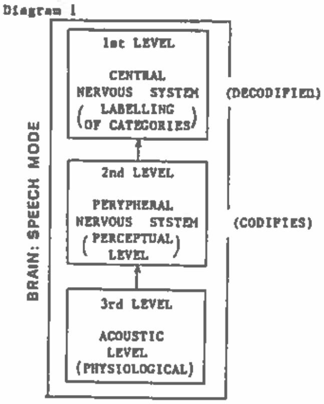

Table 1

Phoneme Symbols	m	n	ɲ	r	l	ʎ	f	s	ʃ	ts	tʃ	p	t	k	v	z	dz	dʒ	b	d	g
MODE 1 (Voicing)	2	2	2	2	2	2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
MODE 2 (Continuity)	0	0	0	3	1	1	3	3	3	1	1	0	0	0	3	3	1	1	0	0	0
MODE 3 (Place)	3	2	1	2	2	1	3	2	1	2	1	3	2	0	3	2	2	1	3	2	0
TIMING	t^{m}	t^{n}	$t^{ɲ}$	t^{r}	t^{l}	$t^{ʎ}$	t^{f}	t^{s}	$t^{ʃ}$	t^{ts}	$t^{tʃ}$	t^{p}	t^{t}	t^{k}	t^{v}	t^{z}	t^{dz}	$t^{dʒ}$	t^{b}	t^{d}	t^{g}

Diagram 2

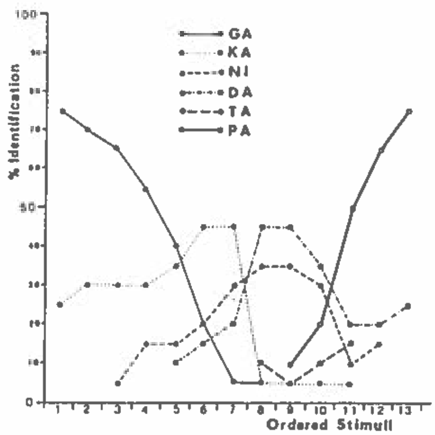

Table 2a

Stimulus Number	Parameter Values		
	F2	V.O.T.	Burst Duration
1	2200	+90	5
2	2100	+80	5
3	2000	+70	5
4	1900	+60	5
5	1800	+50	5
6	1700	+40	5
7	1600	+30	5
8	1500	+20	5
9	1400	+10	5
10	1300	0	5
11	1200	-5	5
12	1100	-10	5
13	1000	-15	5

Table 2b

Feature Values			Labels
Place	Voicing	Continuity	
0	1	0	ga
0	1	0	ga
0	1	0	ga
0	1	0	ga
0	1	0	ga
0	0	0	ka
0	0	0	ka
2/1	1/0	0	da/ni
2/1	1/0	0	da/ni
2/1	1/0	0	da/ni
3	0	0	pa
3	0	0	pa
3	0	0	pa