

THE RECENT CRISIS IN SPEECH ACOUSTICS AND ITS POSSIBLE IMPLICATIONS FOR LINGUISTIC THEORY

Rostislav Pazukhin

The Pedagogical College, Institute of Neophilology, PL 42-201 Czenstokhova, Poland

Abstract: The century-long acoustic investigation of speech has unmistakably confirmed the lack of any invariant and discrete phonic units in speech flow. This, in turn, implies that our traditional theory of the language, seen as a set of segmental elements combinable into linear sequences, is inadequate and needs to be replaced by a novel language doctrine. The sought invariance of speech signals may be found in the trajectories of articulatory movements, which are the only beareres of linguistic information. The sound modulations merely make such speech gestures audible.

Keywords: speech, non-discretness, articulation, invariance,sound echoes

1. What is now going on in speech acoustics is to be seen rather as a severe practical testing of some basic principles of modern linguistic theory. Acoustics has offered to linguists — for the first time in history — the opportunity of an *absolutely objective* checking of our pronouncements on language. In this very way the popular postulate that every correlated spoken and written utterances must be structured similarly was overthrown:

> Since the earliest spectographic, cineradiographic and electromyographic studies we have known that neither articulatory nor the acoustic flow of speech can be divided into a sequence of segments corresponding to the invariant segments of linguistic description. Whether the segments are words, syllables, phones or features, the case is the same. (Studdert-Kennedy 1987:68).

ISBN: 0 08 043 438X

See also (Prince 1992:384; Pazukhin 1993:9fol., and others).

2. Regrettably, acousticians are still irrationally trying to split the non-denumerable continuous speech wave into „sound-chains" corresponding to the discrete letter-sequences (Joos 1950:702-703; Cohen 1981; Fowler 1986; Kent and Read 1992:48fol.; Pazukhin 1993:12-18). The continuous unsuccess of such attempts and the resulting crisis of acoustical methods in speech research (cf. Fant 1990:174) also stimulate a general crisis in modern linguistics. Our traditional view of language as of an atomistic-combinarorial system upon „discrete speech units" has been derived, indeed, from our *reading* practice. Having been accustomed to alphabetic systems of writing, researchers intuitionally seek strict "acoustic correspondents" of letters allegedly hidden in speech flow (Pazukhin 1996a:8fol.). If then the assumption of an analogous composition of written and oral speech is, in fact, untenable, most theoretical constructions of contemporary linguistics must of necessity be groundless (Pazukhin 1996a:13-14).

3. A natural way out of this crisis would be a concentration on the direct acoustic analysis of living speech (free from any of the disturbing effects of graphics). This new approach, however, stumbles upon a still greater barrier: the observable lack of *invariance* in speech signal (Perkell and Klatt 1986:i-ii).

Generally, material signals belonging to a particular signal-type must possess some common (invariant) properties admitting signals' successful discrimination by users from signals of other types. For instance, concrete traffic lights may have various colour shades; at the same time, however, they conserve certain basic colour properties characteristic of all red lights (distinguishing them from every yellow, or green, signal). Curiously, no such invariant acoustic traits can be identified on spectrograms:

> Physically speaking, speech production can be (and usually is) described as a process in which a sequence of spectral shapes is generated by a sequence of vocal tract shapes *via* the laws of physical acoustics. The speech invariance problem is observed when this analysis of gestures or sounds fails to reveal a clear relation between phonetic messages and either the sequence of vocal tract shapes or the sequence of spectral 'slices' which result. No single shape or spectrum, nor any particular sequence of shapes or spectra, can be found to map invariantly onto phonetic segments or their sequence. (Porter 1987:87).

See also (Mack 1991:48).

4. This obstacle may be circumvented by adopting a rather unusual premiss: the invariance in speech flow is present not in the phonic but in the *articulatory* component of speech acts! (cf. Porter 1987:88)

The Echo-theory of speech (Pazukhin 1993:29fol.; Pazukhin 1996c:21-23) posits that the true bearers of linguistic information in speech wave are not the postulated discrete „phones" but the complex and continuous manoeuvres of speech organs. These are being made audible through the fairly variable modulations of the voice. In the experiments

ISBN: 0 08 043 438X

(Pazukhin 1997:521-523), two artificial „utterances", (1) *iieaou* and (2) *aouiie*, spoken by men, women, and children, have been recorded (see the diagrams).

5. The bare „vocalic" composition of the utterances (1) and (2) helps us to realize that the floating formants can legibly reflect the trajectories of speech organs. Note that the formants are not the signals, by themselves, but merely the *echoes* of the articulatory movements (which are the genuine signals). Proceeding from these acoustic „allusions", hearers, in their imagination, restore the invisible spatial changes in the vocal tract (Pazukhin 1993:34-38). The invariance in speech is, thus, due not to the configurational stability of acoustic images but to their capacity to point to one and the same dynamic articulatory type (despite of their having not identical shapes).

The articulatory manoeuvres themselves may in turn be highly variable, since no „precise parameters" have any practical importance in gesticulation. It is the general dynamic impression (*Gestalt*), loose and irregular as it may be, that determines the identity of a gesture (cf, e.g., hand gestures, used as a common greeting sign, being, at the same time, so much diversified kinematically in different persons, situations, and emotional states).

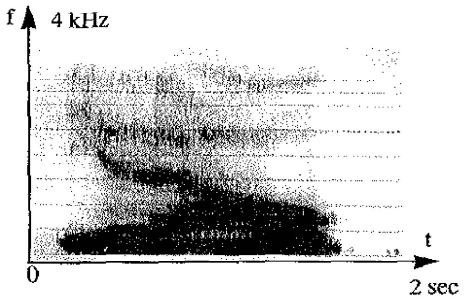
6. At the same time, the „floating" formants schemes can remain acoustically invariant in the speech of persons differing in their age, sex, and the individual (incidental) characteristics of the vocal tract. It thus is hard to distinguish the pronunciation of a man (MO), a woman (LR), and a child (MK) on our diagrams, while these are not all that identical. The differences among the displays can be more properly defined as „topological deformations" (cf. Atiyah 1992:936) of certain „basic patterns", here: (1) and (2).

We can thus see that the Eco-theory is one of the alternatives, more promising than the well worn-out "segmental theory of speech". It is capable, at least, to suggest a concrete solution of the invariance problem, which will for ever remain an insuperable barrier to the "atomistic-combinatorial" methodology. The kinematic-configurational study of living speech may also explain the codal unity of different modalities of oral speech, such as the whisper, whistled, oesophageal, and "inner" speech, lip-reading, TADOMA communication, singing, etc., which are constantly based on standard articulations proper to a particular language (English, Polish, etc.). In all of these cases, the same "vernacular" movements of the speech organs are being reproduced by different kinds of energy (not only by voice effusion!) through different material manifestations (cf. Pazukhin 1993:22-26, 32-34; Pazukhin 1996c:20-21). Most of these oral speech modalities are fully disregarded by phono-acousticians, who concentrate exclusively on so-called "voiced" speech.

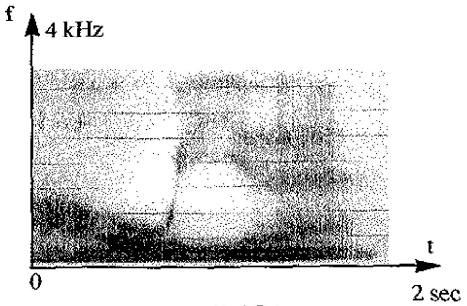
Such approach may thus open a realistic way out of the lasting overproduction of sterile phono-acoustic experiments, discussions, declarations, and publications, which revolve around discovering the legendary "speech particles".

There are also other indications that a new theoretical linguistics of the oncoming century will not, certainly, be „atomistic-combinatorial" by its nature.

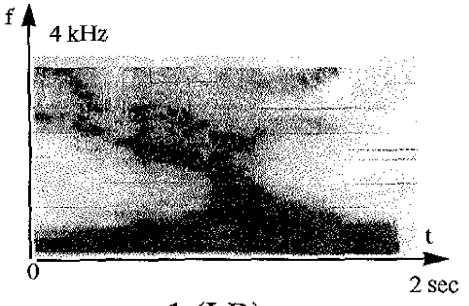
ISBN: 0 08 043 438X



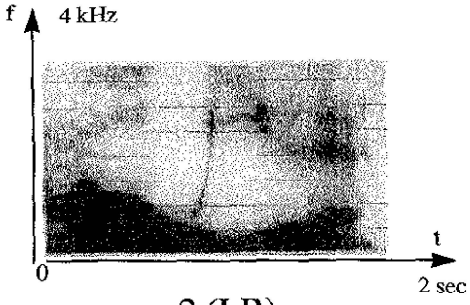
1 (MO)



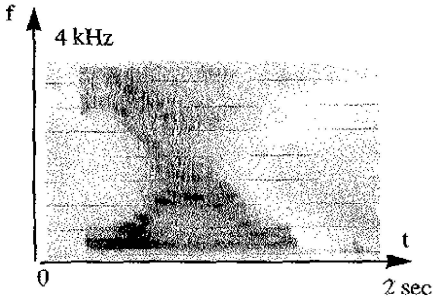
2 (MO)



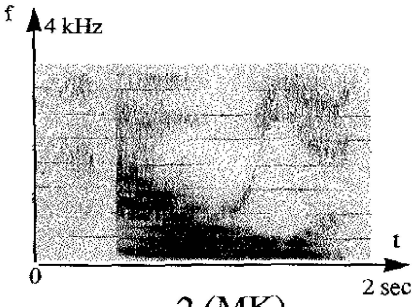
1 (LR)



2 (LR)



1 (MK)



2 (MK)

(executed by Mariusz Owsianny)

ISBN: 0 08 043 438X

REFERENCES

Allport, A., D. Mackkay, W. Prinz, E. Scheerer (Eds). (1987). *Language Perception and Production*. Academic Press, London.

Atiyah, M. F. (1992). General topology. In: *The New Encyclopaedia Britannica*, vol. XIX, 936-950. Enc. Britannica Inc., Chicago.

Cohen, J.R. (1981). Segmenting speech using dynamic programming, *Journal of the Acoustical Society of America*. 69(5), 1430-1438.

Fant, G. (1990). Speech research in perspective, *Speech Communication* 9(3), 171-176.

Fowler, Carol A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*. 14, 3-28.

Joos, M. (1950). Description of language design, *Journal of the Acoustical Society of America*. 22(6), 701-708.

Kent, R. and Ch.Read. (1992). *The Acoustic Analysis of Speech*. Whurr, San Diego CA.

Mack, Molly. (1991) Auditory phonetics. In: *The Linguistic Encyclopedia*, K. Malmkjaer (Ed), 42-48. Routledge, London.

Pazukhin, R. (1993). *Oral Speech: Controversies and Perspectives*. Author's Edition, Lodz.

Pazukhin, R. (1996a). *Topics in Semiotics and Linguistics*. WSP, Czenstokhova.

Pazukhin, R. (1996b). Speech perception and "paper" linguistics. In: Pazukhin (1996a), 7-17.

Pazukhin, R. (1996c). A linguistic-acoustic-computational analysis of spoken communication: Illusion and reality. In: Pazukhin (1996a), 18-24.

Pazukhin, R. (1997). Floating formants and invariance in speech signal [in Polish]. In: *XLIV Otwarte Seminarium z Akustyki*, vol. II, 519-524. PTA, Gdańsk-Jastrzębia Góra.

Perkell, J. and D. H. Klatt (Eds). (1986). *Invariance and Variability in Speech Processes*. Erlbaum, Hillsdale NJ.

Porter R.J. (1987). What is the relation between speech production and speech perception. In: Allport *et al.* (1987), 85-106.

Prince, A. (1992). Segments. In: *International Encyclopedia of Linguistics*, W. Bright (Ed), vol.III, 348-386. Oxford University Press, New York.

Studdert-Kennedy, M. (1987). The phoneme as a perceptual structure. In: Allport *et al.* (1987), 67-84.

ISBN: 0 08 043 438X