Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Auditory patterns of speech

Fant, G.

journal: STL-QPSR
volume: 5
number: 3
year: 1964
pages: 016-020

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

B. AUDITORY PATTERNS OF SPEECH *

Any attempt to propose a model for the perception of speech is deemed to become highly speculative in character and the present contribution is no exception. Neurophysiologists[10] have detailed information on the topology of the brain but not so much to contribute concerning the mapping of the information processing in successive stages of the auditory system. The purpose of the present paper is merely to stimulate discussions on these processes.

The current trend in speech research[2] is to analyze speech spectra and other acoustic data in terms of speech production parameters and in specific in terms of a set of hypothetical motor commands controlling the articulatory and phonatory activity in speech. This has proved to be a very fruitful approach providing valuable organizational principles in dealing with the speech substance[4][9][13]. Some investigators have gone one step further and hypothesize that the brain utilizes the same principles for the decoding of speech. Typical examples of such "motor theory" of speech perception are those of Liberman et al[2][7][8] and Chistovich et al[1]. Their common element is that a primary stage of auditory signal analysis is followed by a stage of articulatory recoding before identification of words takes place. In one variant of Liberman's theory[8] the role of the speech motor functions is limited to the establishment of identification criteria during the process of language learning.

My own opinion[5] would be that the speaking ability is not a necessary requirement for the perception of speech but that it enters as a conditioning factor. I would guess, though, that the importance of the speaking capacity for the development of speech perception is less than the importance of the hearing capacity for the development of normal speech. Children learn to understand speech before they talk and people born with complete lack of hearing have great difficulties in learning to speak.

The block diagram of Fig. II-B-1 allows a discussion of various models, my own view included. A common principle would be that motor and sensory centers become more and more

---

*Paper to be presented at the Symposium on Models for the Perception of Speech and Visual Form in Boston/Mass., November 11-14, 1964.

HYPOTHETICAL MODEL OF BRAIN FUNCTIONS IN SPEECH PERCEPTION AND PRODUCTION

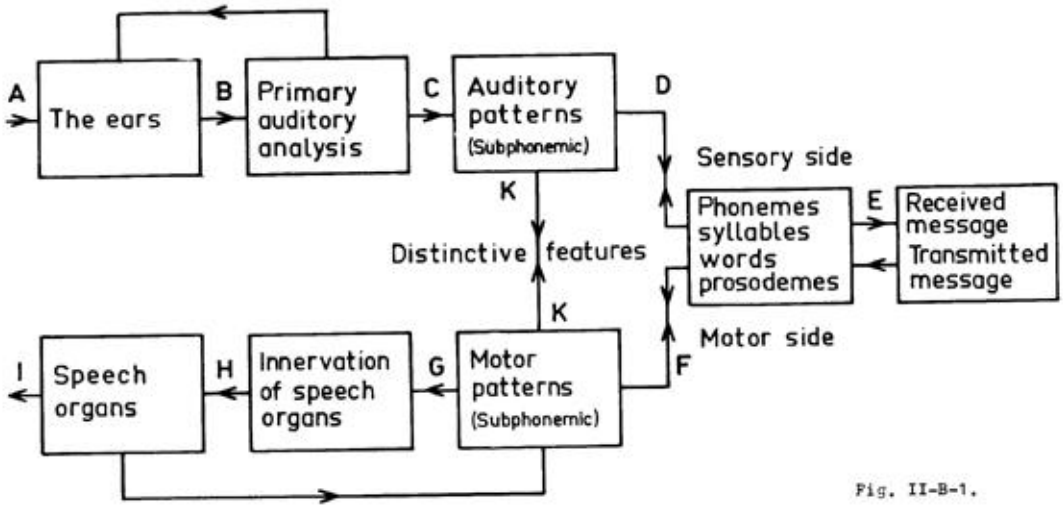

Fig. II-B-1.

involved, as we proceed from the peripheral stages to the central stages. According to Penfield and Roberts[10] there exist separate although closely connected language centers on the word level for motor and sensor functions. Local sensory feedback from the articulatory organs to the motor command center is included and the motor connections between the auditory cortex and the cochlea are also a mere detail of less significance. It is the extent and functional role of possible interconnection between the sensor and motor parts which is the main point of discussion.

According to Lieberman the criteria for selection are established by reference to what set of articulatory commands the subject would associate with perceived sound patterns. Once the correct association is set up, the decoding could proceed along the branch KFE or directly along CDE, if box CD contains an image of the neural motor patterns stored in box GF.

From my point of view, the capacity of perceiving distinctive auditory patterns on the subphonemic level would be delveloped in the early learning process, prior to and not critically dependent of corresponding motor patterns. However, because of the cause to effect relation between articulation, speech wave, and perception the inventory of auditory patterns would be structured rather similar to the patterns of motor commands. Each gross category of articulatory or phonatory events on the distinctive feature level in box GF would be paralleled by a corresponding auditory pattern in box CD. Liberman stresses the rare cases when apparent differences in articulation are associated with rather small differences in the acoustic structure. To me it is the collected auditory experience and the necessity of making linguistic discriminations that account for the distinctiveness of auditory patterns even if they are seemingly small.

Here I am back to the main theme of the "distinctive features", as proposed by Jakobson, Fant, Halle[6], although I wish to stress the general type of approach rather than a specific solution.

Distinctive features are of two kinds, manner features related to the particular sound source and resonator system involved, and place features related to the place of tongue articulation and the degree of lip-rounding. In this connection and since so much attention has been devoted to place features, I would like to emphasize the dynamic aspects of manner features. It is the temporal contrast effects of rapid changes in spectrum composition, intensity, and the particular onset or offset characteristics of successive sound segments that account for several distinct qualities. I have in mind stops versus fricatives and affricates, and vowels versus nasals and laterals, and syllable nucleus versus syllable boundaries. For example, the sound segment corresponding to a proper lateral articulation sounds like a vowel when gated out and presented out of context. It is the contrast with adjacent sound segments which provides the peculiar auditory effect that we associate with the /l/ phoneme. This can be conceived of as analogous to the "ringing" or "overhang" in a filter system with long time constants when subjected to a step change in the input signal.

In my opinion, there has been an overemphasis[7] on second formant transitions and more generally on F-pattern (F_1 F_2 F_3) specifications as acoustic correlates to place features of consonants. We could, for a change, view the spectrum through the constraints of a cochlear model rather than by means of a conventional spectrum analyzer, of course, with the understanding that such a model represents an extreme peripheral aspect of the auditory frequency-place analysis only. Experiments with synthetic speech can provide more direct evidence, see the contribution of O. Fujimura "The spectral shape in the F_2-F_3 region". Secondly, as I have suggested in earlier publications[3], we should attempt to set up hypotheses on how the auditory system would integrate various simultaneous and successive cues into one single distinct auditory sensation. A typical example would be to integrate "inherent" and "transitional" place cues sampled from a consonantal sound segment and an adjacent vocalic segment, e. g. in terms of a combined locus.

Motor theories of speech perception often incorporate an "analysis-by-synthesis" procedure for generating articulatory patterns[11]. In a pure sensory theory of speech perception on the other hand, analysis-by-synthesis implies a generation of auditory patterns according to rules for allophonic variations.

Independent of the particular emphasis laid on motor versus sensory patterns, it is apparent that speech perception at least on a higher level is combined with a running prediction which limits the necessary range of search before identification. The extent to which such prediction employs activity in the motor centers, e. g. along a loop EFKDE or EFDE and how far down towards the peripheral end of the system motor activity mediates sensory discriminations is a pertinent problem.

The presence of motor activity at a peripheral level during listening need not imply an active engagement of motor centers in auditory discriminations but merely the back scatter of neural activity along the path KGHI or FGHI or both. The path ABCKGHI is directly applicable to the shadowing experiments of Chistovich et al[1], who made synchronous oscillographic registrations of the articulation of a subject listening to nonsense VCV words and the articulation of the leading speaker. The timelag was found to be of the order of 100-150 msec and the conclusion was that the subject could start to reproduce some of the distinct articulatory features of the consonant before it was completed by the leading speaker. From the point of view of distribution in time of distinctive sound features it is of interest to find the same trend in the carefully controlled gating experiments of Öhman[12].

Inner speech would proceed along a closed loop EFDE, and when introspectively concentrating on isolated features along EFKDE. Even here a secondary flow FGHI is simultaneously present although with a reduced information, e. g. concerning phonation.

G. Fant

References:

(1) Chistovich, L. A., Klass, J. A. and Kuzmin, J. I.: "Running identification of speech sounds", Psychological Problems, No. 6 (1962) (USSR), pp. 26-39.

(2) Cooper, F. S.: "Instrumental methods for research in phonetics", paper presented at the Fifth International Congress of Phonetic Sciences, Münster, August 16-23, 1964.

(3) Fant, G.: Acoustic Theory of Speech Production ('s-Gravenhage 1960).

(4) Fant, G.: "Descriptive analysis of the acoustic aspects of speech", Logos 5 (1962), pp. 3-17.

(5) Fant, G.: "Comments to the SCS paper D3 by Liberman, A. M. et al 'A motor theory of speech perception' ", Proceedings of the Speech Communication Seminar (Stockholm 1963), Vol. III.

(6) Jakobson, R., Fant, G. and Halle, M.: "Preliminaries to speech analysis", Acoustics Laboratory, MIT, Technical Report No. 13, 1952 (1st to 3rd issue out of print); 4th printing published by The M.I.T. Press. Cambridge, Mass., 1963.

(7) Liberman, A. M.: "Some results of research on speech perception", J.Acoust.Soc.Am. 29 (1957), pp. 117-123.

(8) Liberman, A. M., Cooper, F. S., Harris, K. S. and MacNeilage, P. F.: "A motor theory of speech perception", Paper D3, Proceedings of the Speech Communication Seminar (Stockholm 1963), Vol. II.

(9) Lindblom, B.: "Spectrographic study of vowel reduction", J.Acoust.Soc.Am. 35 (1963), pp. 1773-1781.

(10) Penfield, W. and Roberts, L.: Speech and Brain Mechanisms (Princeton 1959)

(11) Stevens, K. N.: "Toward a model for speech recognition", J.Acoust.Soc.Am. 32 (1960), pp. 47-55.

(12) Öhman, S. E. G.: "On the perception of Swedish consonants in intervocalic position", Thesis work for fil.lic. degree at the University of Uppsala; also published as Report No. 25, March 1962, Royal Institute of Technology (KTH), Speech Transmission Laboratory.

(13) Öhman, S. E. G.: "Numerical model for coarticulation, using a computer-simulated vocal tract", J.Acoust. Soc.Am. 36 (1964), p. 1038 (A).