SOUND, FEATURES, AND PERCEPTION*

GUNNAR FANT**

THE SPEECH COMMUNICATION CHAIN

Speech communication may be considered as the transmission of information through a succession of stages within a speaker, a connecting medium, and a listener. Flow diagrams of this process can be elaborated in various forms depending on the detail of the analysis attempted and the aspects of the communication process on which the investigator focuses his descriptive efforts. The following tabulation of stages will be considered here.

A. Production	(1) Intended meaning of message
	(2) Message sentence form
	(3) Neural production program
	(4) Myodynamic activity
	(5) Aerodynamic and acoustic processes
B. Technical medium	(1) The acoustic speech wave emitted by the speaker
	(2) Speech signal representation in various parts of a technical communication system
	(3) The acoustic speech wave affecting the listener
C. Perception	(1) Cochlear response
	(2) Primary neural analysis
	(3) Identification of phonetic elements
	(4) Identification of sentence structure
	(5) The message received

The terminal stages remain rather hypothetical in view of our limited insight in the organization of brain functions. Therefore, the formulation of the stages $A(1)$, $A(2)$ and $C(4)$, $C(5)$ above reflects our general concepts of successive levels of language structure rather than established neurological functions.

---

* Condensed version of the oral presentation at the 6th Int. Congr. of Phonetic Sciences. The detailed material including illustrations is to be found in the author's contribution to the forthcoming *Manual of Phonetics*, edited by B. Malmberg, North-Holland Publishing Co. [see Fant (1968)].

** Dept. of Speech Communication, Speech Transmission Laboratory, Royal Institute of Technology (KTH), 10044 Stockholm 70, Sweden.

Each stage is to be characterized by an inventory of specific signals specified by parameters which possess certain time and space characteristics that combine into patterns according to general rules and constraints. A major ambition is to derive rules for translating a representation on one stage to a corresponding representation on any other stage of the complete system. Stage $A(4)$ which comprises the dynamics of the speech organs may accordingly be described by a set of time varying articulatory parameters. One of the primary aims of general phonetics and speech research is to derive the rules for translating from this articulatory stage to that of the speech wave $B(1)$.

At the stage $A(5)$ comprising the acoustic production processes the signal structure can be divided into source and filter categories and each of these may be considered at two substages. Thus the filter-function is initially represented by the vocal tract "area-function", i.e. its resonator dimensions, from which their sound shaping properties may be derived by acoustic theory. Similarly, the source has a primary aspect of mean pressures and flows characterizing the aerodynamics of the exhaled air whilst the superimposed periodic or random disturbances constitute the raw material of voiced and unvoiced sounds.

This model of a successivity of encoding stages that the speech message has to pass from the transmitter to the receiver through the entire speech communication chain cannot be quantitatively studied with the same rigor as for instance a telegraph communication system. The main purpose of the model is to serve as a frame for formulating research objectives and discussing descriptive theory whilst the application of a quantitative signal and information analysis generally is beyond our capacity.

One sometimes encounters statements proposing that the information rate is very low at higher brain centers and increases towards the periphery with a maximum at the speech wave. This reasoning suffers from a confusion of the message and signal aspects of the communication. Ideally, the message is the same at all stages and the rate of information flow thus the same everywhere. It is more valid to speak of an ncreasing redundancy in the sense that the signal structure gets more complex and utilizes a larger number of parallel pathways whilst the information remains the same. Even this statement is rather loose in view of our limited insight in the neurological levels.

At present it is not possible to accomplish anything like a complete description of signal structure at any stage with the exception of the acoustic speech wave where all details of the waveform may be sampled and studied. However, even if we cared to carry out a maximally detailed sampling it would not be worth the labor. Also, there exists an infinite variety of transformations for expressing one and the same fact by different parameters, i.e. by different descriptive systems. Thus, in spite of apparent visual differences a narrow-band spectrum contains essentially the same information as a broad-band spectrum.[1]

[1] The signal data contained in a spectrogram are mathematically equivalent to that of an oscillogram providing the phase information is retained in the spectral representation. Relative phases within the spectrum would mathematically account for one half of the information concerning the signal structure but they are of rather minor communicative importance. Spectrographs are not designed to preserve phase information which in effect reduces the "redundancy" of spectral specifications by a factor of two compared to oscillographic specifications.

A formant representation is more condensed than a harmonic representation since a small number of formants can have the same descriptive power as a large number of harmonics. This economy is generally gained at some reduction of the accuracy in signal analysis. However, the harmonic representation is more detailed only when the voice fundamental frequency is low. The information gained in a low F_0 harmonic spectrum concerns irregularities of the voice source rather than the more important properties of the vocal tract transfer function.

We have to accept the limitation of any quantification being approximate only but we require that it shall preserve a maximum of message information with as simple a signal description as possible. The extent to which such "minimum redundancy" or "compact" descriptive systems can be worked out is first of all a matter of how well the investigator is acquainted with the stage and its constraints and how complex abstractions he is capable of introducing.

THE NATURE OF DISTINCTIVE FEATURES

Complete formant specifications of a piece of speech is of practical use for synthetic reproduction only and is too detailed for comparative phonetic studies. What we need is a phonetically oriented data sampling system that allows us to sample the speech wave less densely than at intervals of the inverse of the bandwidth. The segmentation theory outlined by Fant and Lindblom (1961) and Fant (1962A and B) is a starting point for developing such a system. Segmental boundaries are mainly derived from changes in the "manner of production" whilst the "place of production" determines a more continuously varying element of segment patterns, in the first place the F-pattern (F_1, F_2, F_3, F_4) reflecting the continuous movements of the speech articulators.

This system operates with a terminology of speech production categories that is in part identical to that of the distinctive feature system of Jakobson, Fant and Halle (1952). The main difference is that the distinctive feature system serves a phonemic minimal redundancy classification purpose whereas the segment classification of Fant and Lindblom accounts for any production category irrespective of its communicative significance and is thus more phonetically detailed.

It should be appreciated that distinctive features in the sense utilized by Jakobson, Fant and Halle (1952) primarily constitute a system for subdividing phonemes and other components of the message ensemble. A distinctive feature has certain correlates on each stage of the speech communication chain and these correlates are described in terms of various parameters and cues, e.g. formant locations. A distinctive feature is thus a unit of the message ensemble rather than a property of the signal ensemble. The term "distinction" or "minimal category" would have been more appropriate and might have led to less confusion concerning their nature and use.

The distinctive features are not intended as absolute descriptors of spectrographic qualities. The production or speech wave correlate of any feature will differ somewhat with the particular context of simultaneous and subsequent features. The invariance is generally relative rather than absolute. For instance, an invariable cue of compactness is the higher F_1 of the compact phoneme compared to the non-compact phoneme in the same context irrespective of which minimal pairs are inspected.

The relation between phonemes or features on the message level to speech segments and parameters on the signal level is generally complex. One segment may contain information about several successive phonemes and a single phoneme is generally related to several successive segments of the speech signal. As a rule the number of segments determined according to the principle of Fant and Lindblom comes out to be larger than the number of phonemes in the utterance. However, this is not always the case since in less careful articulation one or several phonemes of the intended message turn out to be produced in an extremely reduced fashion or omitted altogether without affecting the intelligibility. In practice we do not measure the duration of phonemes in the speech spectrogram but we measure the duration of sound segments and other characteristics of the speech signal.

A feature classification system can thus retain more or less redundancy and it can be more or less representative of actual encoding dimensions of the speech signal. The system of Jakobson, Fant, and Halle is too condensed for practical purposes such as comparative phonetic studies and development of automatic speech recognition schemes. The strength and novelty of the system is that it attempts to break the barrier between phonology and phonetics, linking the theory of message signs with the theory of their physical realization through the concept of the speech signal as a multi-dimensional event.

However, the specific choice of units still remains a disputable compromise between the two aspects. The extreme minimum redundancy objectives inherent in phonemic analysis have been the guiding principle for the selection of features. Accordingly, these constitute a very condensed and handy set for transcription of speech messages. Most of the features represent conventional phonetic categories which undoubtedly have a physiological and psychological significance. In a more phonetically oriented solution, on the other hand, one should increase the number of features so as to avoid or at least reduce the number of features operating in both vowels and consonants. In search for independent units on the signal level as opposed to a linguistic message level one might have to include major allophones of a language. The underlying principle would be to search for an inventory of speech production categories at our disposal for programming the phonatory and articulatory events. EMG, cineradiography, and direct recordings of the dynamical patterning of speech articulation will be helpful tools for such studies.

The search for generative rules of speech production may be exemplified by the studies of some of my colleagues, Lindblom (1963), Öhman (1966, 1967), and Öhman and Lindqvist (1965), who have tackled the problems of formulating rules for predicting vowel reduction, coarticulation, and intonation contours. Given a phonemic or allophonic unit of the assumed production inventory the corresponding speech wave realization may be thought of as the output of "black box" labelled production mechanism the input of which is the selected unit plus a set of other discrete units representing the immediate context of other simultaneous, preceding and following units, prosody included. By a consistent analysis in terms of such models it should be possible to reach a more profound insight in the actual inventory of independent signal categories.

The model of Öhman operates with separate sets of control signals for vowels and consonants and this principle is also followed by Borovičková and Maláč (1966). The frequent use of one and the same feature in vowels as well as in consonants of the Jakobson feature system cannot be supposed to reflect an actual sameness of neurological encoding. Thus it would not be hypothesized that one and the same neural motor command labelled compactness is triggered off in the production of a consonant (*k*) and a vowel (*a*).

As a consequence of the high degree of economy aimed at in the Jakobson system and the unavoidable pay off for this economy in terms of a reduced phonetic similarity of a feature in widely different contexts it is not advisable to scale the phonetic distance between two speech sounds in terms of the number of distinctive features by which the corresponding phonemes differ. An extreme example that I have elaborated on earlier, Fant (1966A), is that the last two phonemes of the word "wing" the (*i*) and the (*ng*) do not have any distinctive features in common as pointed out by Jakobson whereas the temporal contrast between the sound segments related to [*i*] and [ŋ] is minimal only. The place of articulation being the same and the consonant anticipated already by the nasalization of the [*i*] the transition from [*i*] to [ŋ] merely involves a closing gesture of the tongue towards the palate.

One weakness of the phonological feature system leading to this paradox is that the palatal articulation goes with compactness in the consonantal system and with noncompact acute sounds in the vowel system. However, from an abstract acoustic feature point of view the (*uia*) interrelation show some similarities with the (*ptk*) relations. The relation [*p*/*t*] is a good parallel to [*u*/*i*] acoustically and the analogous role of (*k*) and (*a*) can also be supported in spite of the articulatory sameness of (*i*) and (*k*). From a perception point of view this similarity is superficial. In my view vowels and consonants are perceived through separate "feature channels", if any.

SPEECH PERCEPTION

From the accumulated experience on speech perception and especially experiments with speech-like synthetic stimuli it is apparent that speech is perceived categorically, Liberman et al (1967). We respond phonemically and tend to identify phonemes and allophones in the first place even when we are asked to discriminate small variations in quality, Liberman et al (1957). According to Liberman et al (1963) this effect is pronounced with consonants, whereas vowels are not perceived categorically. Stevens (1966) reports on categorical effects in vowel perception providing the vowel is embedded in a syllabic frame. This effect is interpreted by Stevens as an instance of a principle that all factors that contribute to make the stimulus or the general conditions of the experiment representative of actual speech condition the listener to perform in a "speech mode" characterized by his making message identifications rather than quality gradations, Stevens (1966). This effect is a result of the listener's language experience rather than a unique property of the acoustic signal, Liberman et al (1967), Stevens and House (1966).

The significance of the concept of distinctive features is quite apparent from perception experiments. However, some investigators have interpreted the term distinctive feature at its face value only and accordingly identified it with the concept of a single important parameter or a cue. This has caused some confusions and distrust in the principle of distinctive features. As already stressed a feature is a recurrent phonemic distinction within a language and a major purpose of perception research is to evaluate the physical parameters and cues which signal the distinctions and phonemes of a language. The term cue is the same as an important physical parameter but can also be more complex in the sense that certain parameters combine to a characteristic pattern.

The Haskins Laboratories' systematic studies of the perception of simple stylized formant patterns have contributed greatly to our knowledge of the perceptual significance of formant data. However, the potential risk when working with simplified synthetic stimuli is that they may become insufficient carriers of phonemic cues and that the conclusions drawn from such experiments will be valid for the particular synthesizer only and not for human speech. This was the cause of the somewhat pessimistic conclusions Liberman et al (1957) made concerning the ambiguity of acoustic data as opposed to articulatory data in a study of F_2-locus as a cue for identifying (*d*) and (*g*). It can be shown that the syllables (*da*) and (*ga*) have approximately the same F_2-transition but this ambiguity is resolved by combining F_3, F_2 and the release burst into a single cue, Fant, Lindblom, and de Serpa-Leitão (1966), Fant (1968).

There remains much to be studied concerning the speech wave characteristics of phonemes and distinctions. A practical strategy is to start out with a detailed list of observable spectrographic pattern cues. In order to make a specification of contextual variants feasible it is advisable to present the data on each phoneme or feature in a few reference contents only and add contextual rules derived from studies of coarticulation, reduction, etc. After this preliminary analysis there follows an evaluation by synthesis. One should not start directly with synthesis experiments and an incomplete knowledge of the speech wave characteristics. It is helpful to construct alternative hypotheses concerning effective cues already in the analysis stage of the work.

A method of parameter evaluation which has been extensively used in perception research is to make systematic variations of the sound stimulus and determine the

boundary where the response shifts from one phoneme to another. When this technique is applied to several minimally contrasting pairs of phonemes the data can be interpreted on a distinctive feature basis. The absolute values of the boundaries will vary with the particular context of simultaneous, preceding, and following features of the sound matter as well as with prosodic elements. This is the so-called "contextual bias". In an integrated view based on all parameters of importance for a distinction the distinctive feature or rather its speech wave correlate can be conceived of as a vector perpendicular to the hypersurface constituting the multidimensional boundary. A similar formulation was given by Chistovich (1967) in her paper at the Congress. The main direction of this vector is the sole remaining attribute of the feature if a common denominator of all possible contexts is to be expressed as was the ambition of Jakobson, Fant, and Halle (1952).

However, a knowledge of this mean direction of the feature vector is not a sufficient end result in speech research. For general descriptive phonetics as well as for automatic speech recognition we need the detailed information of how these boundaries shift with context in the general distributional sense adopted here. The search for formulas enabling us to calculate the contextual bias from the discrete inventory of conditioning factors has already been mentioned in the previous discussion on speech production.

The greater accessibility to the problem from a generative speech production point of view than from a perception point of view has had a certain inhibiting effect on the work at the perception end. We would all agree that the categorization inherent at the production end is quite similar to that at the perception end of the speech communication chain but only defenders of a motor theory of speech perception would argue that perception is nothing but the association of the incoming acoustic stimulus with production categories at the listener's disposal when acting as a speaker. If production and perception categories were identical there no longer remains any difference between a sensory theory and a motor theory but merely a concept of economy in our storage of phonetic categories in a place of the brain common for production and perception.

However, by introspection we can certainly study our own stored sound images of distinctive features and phonemes some of which we might not be able to produce correctly if they belong to a language we are not so well acquainted with. When we mimick speech rapidly the motor activity must be the automatic consequence of a phonetic identification in a previous stage. An identification through what is going on in the efferent motor pathways appears to be an unnecessary complication.

Speech perception is a process of successive and simultaneous identifications in a chain of successively higher levels of language structure. We cannot expect to find a specific brain center for each linguistic category: feature, phoneme, syllable, morpheme, word sentence, but at least a lower level $C(3)$ and a higher level $C(4)$ as proposed in the introductory section. To the inventory belong short term and long term memory functions as well as feedback mechanism which allow storage comparison and correction. Also it allows a generative prediction of what the speaker is going to say at least at the levels of syntax and semantics, $C(4)$ and $C(5)$. At a level corresponding to a complexity of the order of the syllable $C(2)$ I conceive of an analysis through a window of the width of a few phonemes through which the speech signal passes. I do not hypothesize a strict principle of all phonemes being first identified by their features. Some phonemes are probably identified directly and independent of context, e.g. the (*s*). Also the identification is probably not strictly sequential but of arbitrary order within the time span of the window.

This principle overcomes the difficulty of some of the features being specifically sensitive to context. Each identification is a decision based on the probabilities existing at the particular instance and each completed decision influences the distribution of probabilities for the previous and following elements within the window. The general sequential constraints imposed by the language structure and of the speech production mechanism effectively limits the number of alternatives in any decision. This model is also the best principle we can follow in attempts of automatic speech recognition.

The principle indicated above is close to the model of perception outlined by Chistovich in her paper and has an interesting parallel in her experiments on psychological scaling of perceived distances between each of two alternative phonemes and a synthetic sound, the composition of which is varied to produce variations around a perceptual boundarry. According to the experiments of Chistovich (1967), Chistovich, Fant, and de Serpa-Leitão (1966) there is some evidence of a gross quantization and scaling at a stage preceding the phonetic identification but this effect is not well established yet.

Vowels, glides, nasals, and laterals appear to offer greater descriptive problems than stops, affricates, and fricatives. One reason is the greater dynamic variability and affinity to coarticulation. The other lies in the variation of scale factors with different speakers. The first and second formant frequencies, F_1 and F_2, are known to be more important than other parameters, but they are not sufficiently descriptive. The third and higher formants are also of considerable importance in front vowels and serve to differentiate [*i*] [*y*] [*ʉ*] [*e*] and also [*ɛ*] and [*ö*]. F2 and higher formants of front vowels appear to constitute a single perceptual cue which plays a role similar to that of an F2 alone in mid and back vowels. This cue is probably not sufficiently specified by a center of gravity only, Fujimura (1967). Spectral width and relative intensity may also be of some importance.

NORMALIZATION OF ACOUSTIC DATA

The average female voice shows 20 per cent higher formant locations than an average male voice and the same average difference is also found between the spectrum patterns of the voices of children (age 8) and female voices, Peterson and Barney (1952). However, the scale factors vary not only with the speaker but also with the specific vowel and the formant under observation. Thus F_1 of [*o*] varies but little with the sex of the speaker whereas the scale factor for F_1 of open vowels such as [*a*] and [*æ*] are appreciably greater than the average. The origin of these nonuniform variations lies in the non-uniform scaling of the female vocal tract with respect to the male vocal tract, Fant (1966B).

Even if we include all formants in a specification we might find ambiguities such that a female [*ö*] might have almost the same formant frequencies F_1, F_2, und F_3 as a male [*e*]. Such ambiguities have not been studied in detail but they might be resolved in part by reference to the center of gravity, intensity, and width of the upper group of formants, to a small part by relative levels of peaks and valleys, and in part by a reference to the voice fundamental frequency. It is not known to what extent the normalization with respect to F_0 is a psychological effect, i.e. F_0 acting as a label for the specific female vowel category or whether F_0 enters already in a weighting of the effective timbre. In connected speech we can also expect a normalization with respect to both the immediate and remote context. A related phenomenon is that time variable formants affect the identification more than constant frequency formants. However, it should be appreciated that because of the general relations between formant frequencies on one hand and formant levels and spectral shape factors on the other hand, Fant (1960), a formant F4 is not audible in a vowel [*u*] but has a sensation level equal to that of F1 and F3 in the vowel [*i*].

The boundary shift techniques has been successfully adopted by Fujisaki and Kawashima (1967) for an evaluation of the trading relations between the various vowel parameters. We have used this technique for studying the effect on the source level in the region of F2 and F3 as a factor influencing the F_2—F_3-boundary between [*ʉ*] and [*y*]. A 20 dB reduction of F2 and F3 intensity level shifted the $(F_2, F_3)^{1/2}$ threshold by no more than 50 Hz, i.e. rather little. However, the probability of [*ʉ*] identifications rose significantly within the main [*y*] region.

The extent to which we can approximate a vowel specification by F_0, F_1, and a few measures related to an effective upper formant region is not yet determined but is one of the hypotheses that we can test with synthetic speech. For this purpose the upper formant should be generated in parallel with F1 and shaped with a filter of greater width and selectivity than a simple formant circuit.

CONCLUSIONS

The concepts of distinctive features and cues should be kept apart as belonging to the message inventory and the speech signal inventory, respectively. In search of the physical and psychological reality behind the categorical effects in speech production and perception we might find a system of features constituting a natural ensemble of minimal message units. Such an ensemble can only in part be expected to conform with the system of Jakobson, Fant, and Halle (1952) and I expect it to be more redundant. There remain many questions to be studied concerning the relations between speech parameters and members of a feature or phoneme inventory.

*ACKNOWLEDGMENTS**

I am indebted to Björn Lindblom and Sven Öhman for many fruitful discussions on form and contents of this paper.

REFERENCES

(1966) Borovičková, B. and Maláč, V.: "Towards the Basic Units of Speech from the Perception Point of View", *Proc. of the Seminar on Speech Production and Perception, Leningrad 1966*, 83—88 (Z.f. Phonetik, Sprachwissenschaft und Kommunikationsforschung, 21, Heft 1/2, 1968).

(1967) Chistovich, L.: "Method of Studying the Decision Rules Applied in Speech Perception", to be publ. in *Proc. of the 6th International Congress of Phonetic Sciences, Prague 1967*

(1966) Chistovich, L., Fant, G. ,and de Serpa-Leitão A.: "Mimicking and Perception of Synthetic Vowels. Part II", *STL—QPSR*, No. 3, 1—3 (Stockholm 1966).

(1960) Fant, G.: *Acoustic Theory of Speech Production* (The Hague 1960).

(1962A) Fant, G.: "Descriptive Analysis of the Acoustic Aspects of Speech", *Logos*, 5, 3—17 (1962).

(1962B) Fant, G.: "Sound Spectrography", *Proc. IVth Int. Congr. of Phonetic Sciences, Helsinki*, 14—33 (The Hague 1962).

(1966A) Fant, G.: "The Nature of Distinctive Features", *STL-QPSR*, No. 4, 1—14 (Stockholm 1966).

(1966B) Fant, G.: "A Note on Vocal Tract Size Factors and Non Uniform F-Pattern Scalings", *STL-QPSR*, No. 4, 22—30 (Stockholm 1966).

(1968) Fant, G.: "Analysis and Synthesis of Speech Processes", a chapter in *Manual of Phonetics*, ed. B. Malmberg, (North-Holland Publ. Co., Amsterdam 1968).

(1961) Fant, G. and Lindblom, B.: "Studies of Minimal Speech Sound Units", *STL-QPSR*, No. 2, 1—11 (Stockholm 1961).

(1966) Fant, G., Lindblom, B., and de Serpa-Leitão, A.: "Consonant Confusions in English and Swedish—A Pilot Study", *STL-QPSR*, No. 4, 31—34 (Stockholm 1966).

(1967) Fujimura, O.: "The Spectral Shape in the F2-F3 Region", *Models for the Perception of Speech and Visual Form*, ed. by W. Wathen-Dunn, 251-256 (Cambridge, Mass. 1967).

(1967) Fujisaki, H. and Kawashima, T.: "Roles of Pitch and Higher Formants in Perception of Vowels", *Digest of the 7th International Conference on Medical and Biological Engineering* Aug. 14—19, 1967, Stockholm, Session 24-2.

(1952) Jakobson, R., Fant, G., and Halle, M.: "Preliminaries to Speech Analysis. The Distinctive Features and Their Correlates", *Acoust. Lab., M.I.T. Techn. Rep.*, No. 13 (Cambridge, Mass. 1952); 4th printing publ. by The M.I.T. Press (Cambridge, Mass. 1963).

(1957) Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C.: "The Discrimination of Speech Sounds Within and Across Phoneme Boundaries", *J. of Exp. Psychol.*, 54, 358—368 (1957).

(1963) Liberman, A. M., Cooper, F. S., Harris, K. S., and MacNeilage, P. F.: "A Motor Theory of Speech Perception", Paper D3 in *Proc. of the Speech Communication Seminar, Stockholm 1962*, Vol. II (Stockholm 1963).

(1967) Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F., and Studdert-Kennedy. M.: "Some Observations on a Model for Speech Perception", *Models for the Perception of Speech and Visual Form*, ed. by W. Wathen-Dunn, 68—87 (Cambridge, Mass. 1967).

(1963) Lindblom, B.: "Spectrographic Study of Vowel Reduction", *J. Acoust. Soc. Am.*, 35, 1773—1781 (1963).

(1966) Öhman, S. E. G.: "Coarticulation in VCV Utterances: Spectrographic Measurements", *J. Acoust. Soc. Am.*, 39, 151—168 (1966).

(1967) Öhman S. E. G.: "Numerical Model of Coarticulation", *J. Acoust. Soc. Am.*, 41, 310—320 (1967).

(1965) Öhman, S. E. G. and Lindqvist, J.: "Analysis-by-Synthesis of Prosodic Pitch Contours", *STL-QPSR*, No. 4, 1—6 (1965).

(1952) Peterson, G. E. and Barney, H. L.: "Control Methods Used in a Study of the Vowels", *J. Acoust. Soc. Am.*, 24, 175—184 (1952).

(1966) Stevens, K. N.: "On the Relations Between Speech Movements and Speech Perception", *Proc. of the Seminar on Speech Production and Perception, Leningrad 1966*, 102—106 (Z. f. Phonetik, Sprachwissenschaft und Kommunikationsforschung, 21, Heft 1/2, 1968).

(1966) Stevens, K. N. and House, A. S.: "Speech Perception", a chapter prepared for *Foundations of Modern Auditory Theory*.

DISCUSSION

Akhmanova:

The division into "engineers" and "linguisticians" which Prof. Fant appears to have laid considerable stress on is the one point where I could not quite follow him—otherwise I could not agree more with Prof. Delattre's comment: this report of Prof. Fant is an excellent piece of work, an invaluable contribution to the progress of phonetic science. I am convinced that the successful development of phonetics depends on an even more minute analysis of what is *actually* going on in linguistic communication. We must emancipate ourselves as much as possible from all kind of phonological (or graphological?) preconception of globality.

Carnochan:

The linguist might profitably move closer to the engineer in accepting the sort of physical features of speech to be evaluated, and make use of different notations for different purposes. A phonemic transcription is always necessary for reading, but for analytical statement, a prosodic approach in phonology may bring out relevant systemic contrasts of the syllable, word and of the longer piece, relating the spectrograms and the phonology in a rather more illuminating way.

Lehiste:

In your lecture you presented ample evidence for the variability of acoustic features and for their interdependence when they function as perceptual cues. Two of the basic premises of the theory of distinctive features were orthogonality and invariance: the distinctive features were to be independent of each other, and the physical manifestations of a given feature were supposed to contain some invariant elements. After your lecture, neither of the premises appears to hold any more.

Pike:

Papyrologists find it necessary sometimes to read a "Whole document" before being able to "read" separate words. Cursive writing, even in English, may smear at the end of words into a scrawl with indistinguishable final letters. Isn't there an analogy here to the acoustic problem?

*) This study has been supported in part by Swedish and US governmental funds, more recently by NIH Research Grants NB 04003—05 and HD 02111-02 and by US Air Force Grant AF EOAR 67-34.

It seems to me that the experiments here are studying the contrasts at points of maximum differentiation. In normal speech something more like the smearing of cursive writing may also be found.

Zimnyaya:

Perception can be regarded as a working process of at least two channels, which are parallel. In real life we perceive speech on the basis of probabilistic guessing (prognosticating) the whole structure of a word. This guessing may be based on some features, so to say, critical points which evidently are located in the first segment but they may be placed elsewhere too. So the problem of a unit of perception may be regarded as a problem of a part and a whole.

Fant:

ad Akhmanova: I agree with Prof. Akhmanova that we should not build up separate descriptive theories on speech for engineers and linguists. We have a common interest in being able to introduce more or less redundancy in existing specificational systems and to develop new systems that hopefully would conform with an overall generative system.

ad Lehiste: Distinctive features are by definition independent and display combinatory contrasts only. The production-speechwave and perception correlates are in general not orthogonal. Thus in vowels F_1 and F_2 we have to define not only compactness and gravity but also flatness —$(F_1 + F_2)$. The lack of orthoquality is not important for the theory.