The nature of distinctive features

Fant, G.

journal: STL-QPSR
volume: 7
number: 4
year: 1966
pages: 001-014

http://www.speech.kth.se/qpsr
I. THEORY OF DISTINCTIVE FEATURES

A. THE NATURE OF DISTINCTIVE FEATURES *
G. Fant

Introduction

The following essay is intended as a review of my own thinking on distinctive features. Of all that has been written on this topic I have followed up only a part. I am more practically oriented than linguists and less bound to orthodox acceptance of working principles but I find it fascinating to sit down and work out alternative solutions, e.g., to order the Swedish vowels which was one of my early interests.

A substantial revision of our old "Preliminaries" has long been overdue. The major principles are still valid but I feel we need much more factual data before a substantial revision can be undertaken. Till then the following material may serve as an expression of my views on the subject.

Specialists in language and speech have displayed rather diverging reactions to the theory of distinctive features and misunderstandings have been frequent. Does it provide a condensed presentation of the most useful facts about speech or is it just an intellectual game, a purpose of itself for the structural linguist? High initial expectations of finding new and simple solutions to central problems of speech analysis, such as automatic speech recognition, have been followed by distrust. Now, the theory of distinctive features is not intended as a working recipe for technical application but it can provide some organizational principles and suggestions. The rather specific terminology of the "Preliminaries" does not stand for either radically new or very special features. The articulatory, acoustic, or perceptive correlates of distinctive features should comprise condensed transforms of the most relevant information from any of these stages within the complete

* This is a preprint version of a paper to be published by Mouton & Co. in Festschrift for Roman Jakobson on the occasion of his 70th birthday, October 11, 1966.
speech communication systems. A continued study of alternative solutions and descriptive forms is needed in the development of the distinctive feature theory as in speech analysis in general.

What is then really the concept of distinctive features? How are they defined, from the speech wave, from articulation, or from perception? Are they simply a part of the linguistic code for decomposition of phonemes in a bundle of smaller units? Reading the "Preliminaries" one finds that the distinctive features operate on all four levels but from which one do you start the analysis? Here as in phonemic analysis the set of distinctive features constitutes abstract units of the message code. The distinctive feature is a choice between one of two alternatives. In "Preliminaries" distinctive features are referred to by terms as "discriminations", "choice", and "selection" stressing the linguistic level. A distinctive feature generally recurs as a choice situation in several minimal distinction pairs within a language and it is of course required that the physical or physiological manifestations be consistent, i.e., one and the same feature shall have qualitatively the same articulatory, acoustic, and perceptive correlates independent of context of other features within the bundle. The modification "qualitatively" here implies that the relation between the two opposites is the same in all contexts. Absolute values of descriptive parameters, however, generally vary with context. Failure to recognize the role of contextual bias is a frequent source of misunderstanding of the nature of distinctive features. A distinctive feature is by definition the same in all contexts. The associated physical phenomena, on the other hand, referred to as "correlates", "cues", or "parameters" need exhibit only relational invariance.

Distinctive features are really distinctive categories or classes within a linguistic system but just like in accepted phonemic analysis it is required that they are consistent with the phonetic facts and these phonetic facts on various levels have lent their name to the features. It is not within my competence to discuss the generality of distinctive features but it is apparent that they comprise, as they should, the essentials of the framework of classical phonetics and in addition some categorizations of a more novel appearance.

I shall now proceed to some specific points concerning distinctive features some of which are often brought up in discussions.
How important is the binary principle?

The binary principle obviously has its basis in the presence versus absence of an articulatory or phonatory event, e.g., presence versus absence of voicing, nasality, occlusion, etc. or in a selection of one of two polar alternatives along a continuous parameter scale, e.g., more open as opposed to less open. In the analysis of vowels it can be motivated to recognize more than two significant levels of one and the same parameter, c.f., the discussion on compactness in "Preliminaries", paragraph 2.414. As indicated in the analysis of the Swedish vowel system later in this paper one could conceive of instances, where up to 4 distinct levels of one and the same feature might be considered. In these instances a decomposition in terms of two binary categories is generally undertaken in order to allow a consistent use of the binary principle within the whole system.

The distinctive feature represents the linguist's condensed view of the minimal units for composing speech messages. If properly applied, categorization according to binary principles need not come in conflict with the physical reality. It is a matter of coding convenience only.

Economy, but at what price?

If alternative solutions of distinctive features are possible it is established policy to adopt the one providing the best economy in terms of the smallest number of features or rather the least redundancy minimizing the number of alternatives that can be generated by the specific set of categories. However, this requirement can come in conflict with the principle of consistency, i.e., there is the risk that one or more minimal pairs in which a feature is supposed to operate do not conform sufficiently well with the rest of the system. Seemingly elegant solutions may thus have to be rejected by failures to apply in specific contexts. It is also apparent that the economy gained in treating consonants and vowels with the same features at times leads to somewhat remote analogies.

Are distinctive features always orthogonal?

Distinctive features are handled as independent units on the linguistic level but their phonetic manifestations often lack orthogonality. The phonetic quality of a vowel may to a first approximation be specified by $F_1$ and $F_2$. In the $F_1$, $F_2$ plane, however, gravity, compactness,
flatness, and tenseness all occupy specific vectors and interdependency is thus unavoidable. Even when such parameters as $F_3$, $F_0$, overall intensity and duration are added in order to provide a better approximation it is not phonetically realistic to choose a consistently orthogonal set of features.

How are distinctive features distributed in time?

People who lack training in experimental phonetics are generally rather surprised when they learn that the acoustic speech wave does not stand up very well to the ideal concept imposed by our intuitive phonemic view of speech as a sequence of discrete units with distinct boundaries. One major shortcoming of "Preliminaries" is the lack of a realistic discussion of the time-varying aspects of speech patterns and the temporal distribution of the acoustic, articulatory, and perceptual characteristics underlying the distinctive features which might have saved phonetically inexperienced people from developing an oversimplified, often naive, view of the segment structure of speech.

It is said in paragraph 2.14 of "Preliminaries" that "For practical purposes each phoneme can be represented by a quasi-stationary spectrum in which the transfer function is invariable with respect to time, except in the manner stated for transient effects". Phoneme boundaries are said to be related to rapid changes either in the source function or in the vocal transfer function. It is also said that inherent features in contrast to prosodic features are definable without a reference to the sequence. These rules are oversimplified and need to be reformulated and expanded.

The speech spectrogram displays a mixture of continuous and discontinuous elements. A successivity of "segments" is to be seen but what from the spectrograms appears to be a natural unit may constitute only a fraction of a phoneme, e.g., the aspiration segment or the occlusion segment of a noninitial unvoiced stop. In other instances a piece of speech that stands out from the rest of the sequence as a separate unit, e.g., in virtue of a continuity of voicing, may be associated with several successive phonemes.

No unique and simple rule exists for segmentation of speech on the basis of nonphonemic criteria. I would claim, however, that a nonphonemic segmentation (1)(2) could be of value as a rationale for articulatory or spectrographic systematizations per se and as an introductory
form of transcription before imposing the linguistic message concepts on the signal data. The outcome of such confrontations of preconceived linguistic structure with the observed acoustic-phonetic structure is that the number of physical sound segments comes out to be larger than the number of phonemes. Because of coarticulation effects one sound segment generally carries information on two or more successive phonemes. Conversely, a single phoneme exerts an influence on several successive sound segments of the signal structure.

When it comes to discussing the distribution of distinctive features in time one must make clear if it is the abstract message structure or the physical manifestation of speech signals that is intended. In the former case distinctive features are bounded as the phonemes although there can be continuity of a feature from one phoneme to the next. It is thus said in paragraph 1.1 of "Preliminaries" that "The difference between the distinctive features of continuous bundles permits the division of a sequence into phonemes. This difference may be either complete, as between the last two phonemes /i/ and /n/ in the word wing (which have no distinctive features in common) or partial as between the last two phonemes of the word apt".

This statement is correct on the message level. Indeed, the phoneme /i/ is categorized as nonconsonantal, vocalic, nongrave, acute, and noncompact whereas the phoneme /n/ is labelled consonantal, nonvocalic, nasal, compact. One would accordingly expect a maximum of acoustic contrast between the /i/ and the /n/ of wing.

In a multidimensional articulatory or acoustic space, however, the contrast between the two corresponding segments is minimal only. The place of tongue articulation is identical or almost identical in American English, and both segments are produced with a lowered velum, the anticipatory nasalization being a normal feature generally affecting the entire segment assigned to the /i/. The raising of the tongue against the palate closing off the mouth cavity does not affect the sound much since a substantial part is directed through the nose already in the /i/ segment. In the case of a more reduced articulation the tongue never reaches the stages of full contact with the palate and the phoneme /n/ is signalled merely as the nasalization of the sound segment. The perceptual importance of nasalization of a vowel as a cue for identification of an adjacent nasal phoneme is considerable (3).
The theory of segmentation of speech on various levels is a well worth object for further research and descriptive studies. The essential point to consider is that we can measure the duration of physical events such as sound segments in a spectrogram but there exists no unique method or convention of measuring the duration of a phoneme or of a distinctive feature.

Specific features. Vocalic and consonantal

One of the weaker parts of the distinctive feature theory is that of defining consonants and vowels. It is in my opinion quite motivated to categorize liquids as being both vocalic and consonantal but the classification of the consonant h (and glides) as being nonvocalic and non-consonantal is a more arbitrary construction although arguments can be raised in favor of such a classification. The physical criteria for the vocalic and consonantal features have not been very rigid. A small damping of vowel formants has been one of the requirements in all versions of the system. The first edition of "Preliminaries" stressed the voiced source of vocalic sounds but the Addenda and Corrigenda chapters of the later edition turned the emphasis on the formant pattern and formulated the consonantal feature as almost the negative of the vocalic feature in terms of formant reduction.

In "Fundamentals of Language" Jakobson and Halle (4) limit the consonantal feature to a low intensity alone. In a study of the classification of Swedish phonemes (5) I introduced a new formulation retaining the concept of formant reduction in defining the consonant feature but with intensity associated with the vocalic feature. One gain of this formulation is that the phoneme /h/ accordingly contrasts with vowels as being less intense. Also /h/ differs from other consonants in the lack of pattern contrast with adjacent vowels, motivating the minus consonantal feature.

The study of Fant (1960), see ref. (5), also includes comments on the theory of the syllable. The syllable nucleus must possess the vocalic feature and it displays a temporal contrast with respect to adjacent sounds in terms of either higher intensity or a more vowellike structure. Syllabiciry should not be ascribed to intensity alone.
The subject of tense and lax vowels and consonants has been given a thorough treatment by Jakobson and Halle (6). Their view on the subject does not depart substantially from that expressed in our earlier joint work. The tense versus lax opposition is intended to operate in vowels as well as in consonants. Tenseness is phonetically described by an articulation with greater overpressure behind the place of the active source; in the case of vowels a higher subglottal pressure and in the case of stops and constrictives a higher pressure behind the place of articulation. Furthermore tenseness is associated with a more extreme articulation and with a greater time spent in an extreme articulatory position.

The two last mentioned characteristics were mentioned in our earlier work but I am somewhat sceptic about the higher overpressure. This factor when present indeed adds emphasis to tense consonants but in my opinion it has not been sufficiently well documented in experimental work. Recent studies of Malécot suggest that the combined effect of pressure and duration expressed as a pulse integral could have a role in proprioceptive feedback. In my experience Swedish voiced and unvoiced stops are produced with the same subarticulatory pressure at the instance before release. This does not prove anything for English but it seems probable that the pressure factor, if present, as a constituent in the opposition between American English unvoiced and voiced stops, is of relatively small significance and a secondary effect of glottal articulation. Several recent studies support the view of Lisker and Abramson (8) that it is the glottal articulation that is the basic factor.

The longer duration and higher intensity of the noise interval following at the release of an unvoiced stop compared with voiced stops are physiologically due to a later closing of the vocal cords after the release. The pulmonary pressure appears to be the same. Studies of subglottal pressure in Swedish speech do not reveal any difference in the pulmonary activity comparing voiced and unvoiced consonants or short or long vowels. I really doubt that subglottal pressure would have anything to do with the tense-lax opposition among English and French vowels.
On these grounds I hesitate to accept the use of the tense-lax opposition among American English consonants as well as vowels suggested by Halle (9). It should be used in the vowel system but without reference to subglottal pressure. Within the consonants it would be just as motivated to use the voiced-voiceless distinction as the tense-lax distinction. The economy gained by one and the same feature operating in vowels as well as in consonants is of course desirable but I find it more important to retain a close correspondence between phonetic facts and feature criteria. We need more experimental data to illuminate this very interesting problem.

**Distinctive features and perception**

I do not hold the view that the decoding of speech in the brain up to the level of phonemic identification has to follow a functional scheme strictly conforming with a distinctive feature system of language analysis. This does not imply that I consider distinctive features unimportant in speech perception. Distinctive features as phonetic classes are a psychological reality as judged from confusion tests under varying types of distortion or mental disturbance. Even in rapid mimicking the decoding proceeds along phonetic classes so that, e.g., place of articulation may be confused whilst the category of stop sound is correctly recognized.

When constructing models of speech perception we should not limit our choice to a representation in terms of either allophones or features. On the contrary it seems reasonable that a decoding in terms of phonetic classes (distinctive features) is paralleled by a direct attempt of allphone decoding. I am thus more in favor of a parallel analysis of features than of a serial analysis with a succession of decisions.

The question whether a translation to equivalent motor instructions precedes phonemic identification is not important. Of greater importance is to study what aspects of known features are of primary importance to perception (11). It is observed that in some instances it is the temporal contrast of two successive sound segments (e.g., stops, laterals, nasals) rather than the inherent quality of each of the segments that evokes the particular auditory sensation associated with the particular class (feature). The pertinent problem is to find what transforms we should apply to the speech wave data in order to extract the information bearing elements that operate in speech perception. Attempts to avoid the search for auditorily relevant sound characteristics by an
uncritical acceptance of the view that perception is merely a reconstruction of the production does not appear very fruitful.

Given a distinction, let us say between the phonemes /k/ and /t/ or between the phonemes /g/ and /d/, the procedure would first of all be to study a number of minimally contrasting pairs in different vowel contexts. As a precaution for securing that we have not lost any significant information in the acoustic specification of a feature, it is advisable to consider the underlying production events and apply known transformation rules. Second formant loci are known to be effective cues for discriminating place of articulation but they are not sufficient. Indeed, a vowel context can be found such that the $F_2$ transition is almost the same for /k/ and /t/ or for /g/ and /d/. This is the vowel $[a]$. This ambiguity in $F_2$, observed by Liberman 1957 (12) (see Fig. 3 in his paper) led him to conclude that "articulation goes one way and the speech wave another way" and since we can discriminate /ga/ from /da/ perception follows articulation, not the speech wave. The ambiguity is resolved if we add the information of the third formant and the burst portion.

In a recent experiment (see Section III, A) by Fant, Lindblom and de Serpa-Leitão we low-pass filtered Swedish and English stops at 2000 c/s. The result was that our bilingual speaker, now serving as a listening subject, classified all his filtered Swedish /t/ as /k/ and his /d/ and /g/ and likewise almost all his British English /t/ as /k/. Since the filtering removed the upper part of the burst spectrum of the dentals there remained very little difference between the unfiltered /k/ and the filtered /t/.

This experiment supports the view that the effective acoustic feature is the combination of the second and the third formant and the burst concentrating the energy in the palatals and velars in a frequency region at or above the effective upper formant of the vowel, see Fant (13), pp. 217-218. My formulation above of the perceptually relevant characteristics of compactness is not claimed to represent the final answer. Also I do not claim that a formulation found effective for contrasting /k/ and /t/ must be exactly the same as that contrasting /ŋ/ and /n/. The human brain surely has a sufficiently large capacity to identify /ŋ/ as different from /n/ in a space separate from that discriminating /k/ and /t/. It seems quite likely that feature analysis is an important aspect of speech recognition, adding to the economy of the system.
On the other hand, it seems likely that perception works according to more redundant principles than those guiding linguistic analysis. Here I believe Roman Jakobson agrees with me that speech perception is not merely the neural decoding of his twelve features.

Recent studies of Chistovich et al in Stockholm have revealed interesting results concerning vowel perception. These studies support the view of a categorical perception of isolated vowels and the existence of an $F_1F_2$ quantization in accordance with our working principles of acoustic phonetic analysis.

**Swedish vowels**

The Swedish vowel system is generally presented in terms of 9 long and 9 short phonemes. Special pre-r variants of the open unrounded and rounded front vowels in Swedish orthography å and ö stand out as well recognized allophones of an especially "open" quality.

The following phonemic symbols referred to as the STA alphabet will be used for the 9 vowels. Approximate IPA symbols are included as examples of phonetic values in contexts other than before [r]. Index 1 stands for long vowel, 2 for short vowel, and no index implies either long or short, i.e., length distinction omitted.

<table>
<thead>
<tr>
<th>Group</th>
<th>STA</th>
<th>IPA</th>
<th>STA</th>
<th>IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Back vowels</td>
<td>$o_1$</td>
<td>$u$</td>
<td>$o_2$</td>
<td>$u$</td>
</tr>
<tr>
<td></td>
<td>$\ddot{o}_1$</td>
<td>$\ddot{o}$</td>
<td>$\ddot{o}_2$</td>
<td>$\ddot{o}$</td>
</tr>
<tr>
<td></td>
<td>$a_1$</td>
<td>$a$</td>
<td>$a_2$</td>
<td>$a - a$</td>
</tr>
<tr>
<td>II Unrounded front vowels</td>
<td>$i_1$</td>
<td>$i$</td>
<td>$i_2$</td>
<td>$i$</td>
</tr>
<tr>
<td></td>
<td>$e_1$</td>
<td>$e$</td>
<td>$e_2$</td>
<td>$e - e$</td>
</tr>
<tr>
<td></td>
<td>$\ddot{e}_1$</td>
<td>$\ddot{e}$</td>
<td>$\ddot{e}_2$</td>
<td>$\ddot{e}$</td>
</tr>
<tr>
<td>III Rounded front vowels</td>
<td>$y_1$</td>
<td>$y$</td>
<td>$y_2$</td>
<td>$Y$</td>
</tr>
<tr>
<td></td>
<td>$u_1$</td>
<td>$u$</td>
<td>$u_2$</td>
<td>$o$</td>
</tr>
<tr>
<td></td>
<td>$\ddot{u}_1$</td>
<td>$\ddot{u}$</td>
<td>$\ddot{u}_2$</td>
<td>$\ddot{o} - e$</td>
</tr>
</tbody>
</table>

In my first attempt of ordering I chose to oppose group I to group II and III in terms of the grave/acute distinction and group III was naturally opposed to group II in terms of the flat/plain feature. So far my views have not changed. However, a division within the three major groups in terms of articulatory opening, i.e., the compactness feature classifying not only /$\ddot{a}_1$/ and /$e_1$/ but also by an
intermediate + degree of compactness did not conform well with phonetic facts, the $F_1$ of /u₁/ being the same as the $F_1$ of /γ₁/. Also within the short vowels the $F_1$ of /u₂/ was generally not significantly different from the $F_1$ of /o₂/. I therefore chose in later works (5)(16)(17) to use three degrees of flatness assigning an intermediate value of flatness to the phonemes /y/ and /œ/ and to the phoneme /u/ the maximal degree of flatness, as suggested by Malmberg (18). This conforms well with the flatness criteria of low $F_1 + F_2 + F_3$ and the extreme degree of lip-rounding in the /u₁/ and /o₁/ phonemes. In group I either flatness or compactness in three levels can be used for the further division.

<table>
<thead>
<tr>
<th>Feature</th>
<th>o</th>
<th>å</th>
<th>a</th>
<th>i</th>
<th>e</th>
<th>ä</th>
<th>y</th>
<th>u</th>
<th>ö</th>
</tr>
</thead>
<tbody>
<tr>
<td>grave</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>flat</td>
<td>+</td>
<td>†</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>†</td>
<td>+</td>
<td>†</td>
<td>+</td>
</tr>
<tr>
<td>compact</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>†</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

An alternative solution which perhaps comes closer to the views of Roman Jakobson would be the following. In group I flatness is used to separate phonemes /o/ (IPA û) and /å/ (IPA o) from /a/, and /å/ is opposed to /o/ by the greater compactness. Now in group II and III the opposition sharp/plain is introduced to differentiate /i/ and /e/ from /å/ and /y/ and /œ/ from /u/. * The criterion of sharpness is higher $F_2$ and $F_3$ everything else being equal. In this sense sharpness is given a function similar to that of diffuseness in Halle’s vowel analyses (9). By this arrangement one avoids introducing three degrees of flatness or compactness and four binary features specify the entire system, as shown on next page.

* Suggested by S. Öhman, personal communication.
In this solution /u/ differs from /ö/ by two features. Within the system of short vowels /ö₂/ is opposed to /u₂/ primarily by the sharpening (palatalization) whereas within the long vowels /ö₁/ differs from /u₁/ primarily in terms of greater compactness.

A small variation of the system allowing a closer relational correspondence of /e/ to /u/ is the following:

<table>
<thead>
<tr>
<th>Feature</th>
<th>o</th>
<th>å</th>
<th>a</th>
<th>i</th>
<th>e</th>
<th>ä</th>
<th>y</th>
<th>u</th>
<th>ö</th>
</tr>
</thead>
<tbody>
<tr>
<td>grave</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>flat</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>compact</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>sharp</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Concluding remarks

The concept of distinctive features requires an integrated view of speech patterns at each stage of the speech communication chain and should not be confused with the concept of cue which is limited to a detail of the pattern. Several cues make up a feature. The primary aim of the phonologically oriented analysis such as the one underlying my work together with Jakobson and Halle was to make an integrated formulation of those cues which are in common in all contexts where the feature operates. There is a potential danger that this demand for generality dilutes the discriminatory power of the formulation in any specific context. In spite of frequent reference to contextual bias in our work misunderstandings have occurred, especially among those entering the field of phonetics.
The generalized formulations of our features primarily serve a function in summary views of a phonological system. If we are interested in speech communication mechanisms we must take into consideration all perceptually important information in each specific context where the feature operates. Contextual bias represents "redundant features" from a restricted structural point of view.

In my view this terminology is conceptually misleading when discussing the perception of speech. What is redundant from a narrow metalinguistic point of view may comprise essentials of the actual code. This amounts to saying that the major allophones of a language would have a perceptual reality as individuals whereas we need not anticipate that the brain possesses a catalogue of all possible contextual variations of a feature or of a phoneme. The concept of the distinctive feature appears to be a reality in perception.

Most of the manner of production features are clear cut and the same in all contexts. It is outside my competence to discuss the universality of features except in a very general sense. I consider the distinctive features to be a very powerful concept based on the natural constraints of our speaking mechanism. The set of features we have been operating with is not unique. Alternative solutions, affecting some of the features, are possible within some languages as well as when dealing with similar phenomena in two different languages.

The universal acceptance of the basic concept of feature analysis is much due to the inspiring work of Roman Jakobson.

References


