Distinctive features and phonetic dimensions

Fant, G.

journal: STL-QPSR
volume: 10
number: 2-3
year: 1969
pages: 001-018

http://www.speech.kth.se/qpsr
I. SPEECH ANALYSIS

A. DISTINCTIVE FEATURES AND PHONOETIC DIMENSIONS*
G. Fant

The purpose of this paper is to express some comments on the recent developments of distinctive feature theory with specific reference to the work of Chomsky and Halle (1968). On the whole I consider their feature system to be an improvement over that of Jakobson, Fant, and Halle (1952), one of the main advantages being the introduction of a set of tongue body features in common for vowels and consonants but separate from the consonantal "place of articulation" features. The basic philosophy of treating phonetics as an integral part of general linguistics demands that features in addition to their classificatory function shall have a definite phonetic function reflecting independently controllable aspects of the speech event or independent elements of perceptual representation. However, there is a danger that the impact of the theoretical frame with its apparent merits of operational efficiency will give some readers the impression that the set of features is once for all established and that their phonetic basis has been thoroughly investigated. This is not so. Many of their propositions are interesting and stimulating starting points for further research whereas others I find in need of a revision.

As pointed out by Chomsky and Halle there are still serious shortcomings in our general knowledge of the speech event. Their feature system is almost entirely based on speech production categorizations. The exclusion of acoustical and perceptual correlates was a practical limitation in the scope of their work but also appears to note the importance layed on the production stage. It is far easier to construct hypothetical feature systems than to test them on any level of the speech communication chain. This is really our present dilemma. Until we have reached a more solid basis in general phonetics any feature theory will remain "preliminary".

Here follows my reaction to some of the basic issues in chapter seven of Sound Pattern of English. My earlier comments on distinctive feature theory may be found in the list of references, Fant (1960a, b, 1966, 1967, 1968).

1. Will we ever have a language universal, finite, and unique set of distinctive features?

The universality aspects are attractive but I am somewhat pessimistic about the outlooks. Features are as universal as the sound producing constraints of the human speech producing mechanism and a finite number should suffice for the classificatory function. However, I am rather sceptical concerning the uniqueness and thereby a definite number of features since one and the same facts often can be described in alternative forms and the criteria for selecting an optimum system are not very rigid. Even if we had all the knowledge we needed the choice of features would be dependent on the particular weight given to phonetic and general linguistic considerations and the preferences of the investigator would in the last instance determine some of the selections. The problem is the following.

2. Are the demands on a feature system different on the classificatory level and the phonetic level?

There are two ways of arriving at features: (1) by selecting an inventory of classes suitable for encoding of language structures and then determine their phonetic correlates or (2), to start with an exhaustive analysis of the modes and constraints of the speech producing mechanisms and perception and determine their distinctive function in language. Feature theory has to develop along both lines and investigators differ only in the relative importance layed on one or the other. The main approach of Jakobson et al (1952) was to start out with an ordering of phonemic oppositions and to identify minimal distinctions as the same if motivated by phonetic similarities. The demand for a smallest possible number of features and the fargoing identification of features within the vowel and consonant systems, e.g. that of identifying the relation between dentals and labials with that of front and back vowels, resulted in an unavoidable pay-off between encoding efficiency and phonetic reality and specifiability. Chomsky and Halle (1968) avoided some of these difficulties by introducing a greater number of features.

One of their basic issues is that a feature system in addition to the classificatory efficiency should conform with a natural phonetic systematization. How have they managed in this respect? In many instances such as dealing with the classes of fricatives, stops, nasals, laterals, etc., the solution is straight forward. On the other hand, I find the encoding of
the class of labial consonants as [+ anterior] and [- coronal] to constitute a clear case of departure from the unifying principles. One single phonetic dimension, "labiality", which has a distinctive function has here lost its identity on the phonological level. It appears to be a rather far-fetched hypothesis that the actual neural encoding of labial consonants at some stage should include a selection of a maximal anterior point of articulation in the vocal tract and a lack of tongue tip evaluation in order for a lower level to find out that this command has to be executed by the lips and not the tongue.

The major class features "vocalic" and "consonantal" introduced already in the work of Jakobson et al and the features "sonorant" and syllabic display a complicated system of interdependencies as will be described in later sections.

The starting point for the major class features appears to have been the need to encode certain pre-established phonetic classes whereas the voiced-voiceless feature is a typical example of the opposite approach, i.e. to start out with a natural phonetic dimension and study its distinctive role in language. A natural linguistic class, i.e. all [r]-phonemes, may have rather complicated sets of phonetic correlates and a natural phonetic dimension as voicing may have to be studied together with several other dimensions as tensening, durations, and coarticulation when it comes to the discussion of its distinctive role.

Before we can accomplish the happy marriage between phonology and phonetics we have to work out the rules for predicting the speech event given the output of the phonological component of grammar. To me this is the central, though much neglected, problem of phonetics and it is of the same magnitude as that of generative grammar in general and will require a similar set of transformational rules. The starting point is the feature matrix of a message as successive phonological segments, i.e. columns each with its specific bundle of features, i.e. rows. The particular choice of classificatory features at this stage is not very important providing the conventions relating phonemes to alternative features systems are known.

The derivation of the rules of this "phonetic component" of language aims at describing the speech production, speech wave, or perception correlates of each feature given the "context" in a very general sense of
co-occurring features within the phonological segment as well as those of following and preceding segments. One set of sequential constraints are expressible as coarticulation rules which may be both universal and language specific.

In addition to these more or less inertia dependent laws of connecting vocal gestures there may exist rules of neural reorganization of control signals for modifying the physical manifestation of a feature in accordance with a principle of least effort articulation, or the contrary, a compensation for maintaining or sharpening of a phonetic distinction dependent on what features occur or follow in the time domain. In addition there enter rules for modifications dependent on stress patterns, intonation, tempo, speaker, sex, type, and dialect, attitude etc. Rules for speech segment durations and sound shapes have to be expressed in terms of larger phonological segments, generally several syllables defining a natural rhythmical unit in terms of stress and intonation. Very little is known about these rules. There is some evidence that the phase of maximal intensity increase within a syllable is a reference point for ordering rules concerning segment durations (B. Lindblom, personal communication).

This "phonetic component" of the speech event receives very little attention in the work of Chomsky and Halle who merely refer to the phonetic correlates of a feature as a scale with many steps instead of the binary scaling on the classificatory level. A knowledge of linguistic structuring is of great importance in practical communication engineering undertakings such as the administration of synthesis by rule or automatic identifications. However, without access to the rules of the "phonetic component" the phonetic aspect of features becomes as imaginary and empty as the "Emperor's New Clothes" in the story of H.C. Andersen. Observing the speech wave we are not faced with phonemes or features but sound segments and more or less continuous sound shapes with a reciprocal many-to-one relation between phonological and physical units. The same is true of speech production studied in relation to the phonological transcript. In both cases there is the need to define inventories of physical units, Fant (1968), which are not identical to the distinctive features but are used to define their phonetic correlates. It may be quite practical to refer to a specific sequence of segments as a stop followed by a fricative at the phonetic level while we may want to refer to the whole unit as an affricate on the phonological level.
Those who want to increase their perspectives on phonology in relation to phonetics should read Ladefoged's monograph "Linguistic phonetics" (1967a). A pure phonetic system was outlined by G.E. Peterson (1968).

3. **What is the psychological reality of features?**

As demonstrated in the previous section features must, at least under prototype conditions, have physical correlates as observed by an external observer of the speech communication act and they should hopefully reflect categorical phenomena in the encoding and decoding mechanism. This is not the same as ascribing each feature to a specific brain allocation. We can be aware of a feature by introspection but otherwise it may lack immediate neurophysiological correlates. The important thing is that the actual processes are phenomena that have some abstract relation to our feature matrices.

4. **Is the binary principle important?**

No, not necessarily, but it is convenient. Language regularities and language developments may in some instances be more easily described by scales of three or more levels, cf. Ladefoged (1967a). It is also questionable whether formulations in terms of feature matrices always reveals more fundamental rules than formulations in terms of phonemes.

5. **Are features independent and orthogonal?**

This question can pertain both to the classificatory, "phonological level", and to the phonetic level discussing the production speech wave and perceptual correlates. Besides the apparent constraints on possible sequences of phonological segments there exist universal constraints on feature combinations within one and the same segment. As discussed by Chomsky and Halle [+ high] would contradict [+ low]. Also, some features or combinations of features imply specific signs of other feature in the same bundle, as exemplified by [+ vocalic] implying [+ sonorant]. A closer analysis of interdependencies within the major class features reveals that the class of [+ sonorants] by definition also incorporates all [+ syllabics] and all [- consonantal] segments. Such constraints will be discussed in greater detail in the section of major class features. The phonological dependencies within this set of features are paralleled by phonetic similarities. Thus the class of [- consonantal] incorporating
vowels and glides must have much in common with the class of \([+\text{vocalic}]\) incorporating vowels and liquids. In other words "vocalic" is almost the negative of the "consonantal" feature.

The phonetic interdependencies are apparent even when they are not paralleled by classificatory constraints. The situation had been ideal in the vowel system if the perceptually relevant number of dimensions had been the same as the number of classificatory features. We would have had a perfect orthogonal system if limited to the \([+\text{low}]\) or \([-\text{high}]\) and the \([-\text{back}]\) dimensions corresponding to the \(+F_1\) and \(+F_2\) dimensions, respectively. The feature "rounding" is correlated with \(-(F_1+F_2+F_3)\) and thus only partially independent of other features. The same is true of the feature "tense" which is related to the formant pattern (direction towards an extreme target) and duration. Additional features and/or scale values are needed for the Swedish vowel system as will be discussed later.

We accordingly have to resort to the minimal claim of Chomsky and Halle that features should be at least partially independent. At the same time we have to be aware of considerable interdependencies. This applies to their classificatory function as well as to their phonetic correlates.

6. Are differences in feature contents of matrices a reliable measure of phonetic distance?

No, not always. On an average basis it might be permissible to express differences between languages or dialects by summing binary units in the classificatory domain and expect such differences to represent their phonetic differences, Ladefoged (1969). However, one cannot expect the phonetic difference between any two phonemes to be proportional to the number of features by which they differ. The situation was especially severe in the Jakobson, Fant, and Halle system. It was stated that the \([n]\) and the \([i]\) of the word "wing" do not have any features in common, the \([i]\) being \([+\text{voc}][-\text{cons}][-\text{compact}][-\text{grave}]\), the \([n]\) being \([-\text{voc}][+\text{cons}][+\text{nasal}][+\text{compact}]\). On the phonetic level, on the other hand, the difference between the \([i]\) and the \([n]\) is minimal since the entire \([i]\) is nasalized and the transition from \([i]\) to \([n]\) merely involves a gesture of tongue closure which in dialectal variants is omitted. Within the Chomsky-Halle framework the situation is indeed improved since the tongue body features \([-\text{back}][-\text{low}][+\text{high}]\) are in common for the two segments.
The [+nasal] feature is not phonemically distinct for the vowel and the tongue gesture toward closure is paralleled by the [+consonantal] feature on the classificatory level. The [-vocalic] feature of the [n] becomes an automatic consequence of the [+nasal][+consonantal] combination in both systems.

The moral is, eliminate inherent redundancies before counting the number of plus and minus features by which two segments differ.

7. Review of the major class features

The three major class features "sonorant", "vocalic" and "consonantal" relate to various aspects of the open-close dimension. Judging from a footnote on p. 302 of Sound Pattern of English, Chomsky and Halle recommend the use of a new feature "syllabic" instead of "vocalic". According-ly my analysis will be based on this revision whereas almost all construc-tions in Sound Pattern are based on the "vocalic" feature.

The nature of these features is best reviewed with reference to phonetic classes:

[-consonantal] = vowels and glides+h
[+vocalic] = vowels and liquids
[+syllabic] = vowels, syllabic liquids and syllabic nasals
[+sonorants] = vowels, liquids, nasals, and glides+h
[+consonantal] = liquids, nasals, and obstruents (fricatives, affricates, and stops)

[-vocalic] = glides+h, nasals and obstruents
[-syllabic] = glides+h, nonsyllabic liquids and nasals, obstruents
[-sonorant] = obstruents

It is thus seen that all [-consonantal], all [+vocalic] and all [+syllabic] segments must by definition also be [+sonorant] and, conversely, if a segment is classified as [-sonorant] it must by definition also be [+consonantal], [-vocalic] and [-syllabic]. The crucial difference between the "vocalic" and the "syllabic" features is that nonsyllabic liquids are plus in terms of the vocalic feature but minus in terms of the syllabic feature. This has important consequences in segment classifications.

Before further discussing the inherent redundancies of combinations of these features I shall briefly review their phonetic definitions and add some comments on their acoustic correlates as I see them.
Consonantal sounds are produced with a radical constriction in the midsagittal region of the vocal tract. This constriction limits the flow of air in the obstruents and in the closed phase of r-sounds whereas it is "shunted", i.e. by-passed in laterals and nasals. Because of the variety of sounds to be included by the feature a formulation of the acoustical correlates becomes rather complex, the common denominator being a deviation from the ideal "vocalic" pattern by a reduction of the second and/or higher formants.

Vocalic sounds are produced with an oral opening that shall not exceed that of the high vowels [i] and [u] and which by definition shall be greater than that of glides. In addition the vocal cords shall be positioned to allow for spontaneous voicing. This requirement rules out unvoiced vowels as being nonvocalic. Oral opening here includes lateral opening and in case of sonorant [r]-sounds the more open intervals. The acoustic correlate is a higher $F_1$ and higher overall intensity than in nonvocalic sounds.

Syllabic sounds form a syllabic peak in the sequence of sound events. Obstruents are by definition excluded from the possibility of forming syllabic peaks, whereas syllabic nasals and liquids between obstruents are basically characterized by the same criterion as that of vowels between obstruents or glides. A weighted sum of second and first formant intensity relative to that of adjacent phonetic segment would be the simplest acoustic correlate.

Sonorant sounds. The relative degree of sonority can be based on exactly the same criteria as for syllabicity except that the relative degree of sonority is related to alternative compositions of one and the same segment whereas syllabicity implies comparisons in the time domain. The production correlate of sonority is the sum of vocal tract openings including oral, nasal, and lateral passages which is larger than that found in obstruents. Thus [-sonorant] = obstruent. An interesting claim not yet verified is that nonsonorant sounds would not allow "spontaneous voicing" and that a compensation of glottal adjustment to counteract the impaired flow would be necessary.

The interdependencies between basic class features are as apparent on the phonetic level as on the classificatory level. The situation is even more complicated by the fact that the continuant-noncontinuant (stop) feature is the same as the consonantal feature, except that the degree of
primary stricture is total in stops and in the closed interval of affricates
but not total in the [+consonantal][+continuant] fricatives.

I fully agree with Chomsky and Halle on the need for replacing the
"vocalic" feature by the "syllabic" feature. The syllabicity seems to be
more easily testable than the vocalicity which employs a disputable thresh-
hold between liquids and glides which does not focus on the important dif-
f erences. Furthermore, I suggest a further reduction of the number of
features dealing with vocal tract opening by replacing the features "con-
sonantal" and "continuant" by one single feature (medially) "closed" which
is identical to the "consonantal" feature but for an extension to separate
stops and affricates from fricatives. Before applying this feature we shall
study how some of the main phonetic classes are encoded.

<table>
<thead>
<tr>
<th>vowels</th>
<th>nasals</th>
<th>laterals</th>
<th>r-sounds</th>
<th>glides+h</th>
<th>stops</th>
</tr>
</thead>
<tbody>
<tr>
<td>syllabic</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>consonantal</td>
<td>-</td>
<td>(+)</td>
<td>(+)</td>
<td>(+)</td>
<td></td>
</tr>
<tr>
<td>sonorant</td>
<td>(+)</td>
<td>+ (+)</td>
<td>(+)</td>
<td>+ (+)</td>
<td>(+)</td>
</tr>
<tr>
<td>nasal</td>
<td>+ +</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>lateral</td>
<td>(-)</td>
<td>(-)</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>continuant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inst. release</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Features that by definition are implied by other features of the same
phonological segment are marked with parantheses. Blank spaces repre-
sent other instances of "unmarkedness", i.e. (a) not applicable because
of physiological constraints, (b) irrelevant for the classificatory function,
or (c) occurrence in rare cases only. In detailed feature-analysis it
would be valuable to have separate notations for these four different as-
pcts of unmarkedness and also for the fifth aspect, that related to sequen-
tial constraints as implied by all higher levels of analysis. According to
Chomsky and Halle the [+nasal] feature when added to stops could stand
for prenasalization, i.e. instance (c) above, whereas +nasal, when added
to vowels or liquids, is a contextual variant due to adjacent nasal conso-
nants and can thus be omitted from the matrix (case (b) above).
It is interesting to note that if the feature matrix is to be used for description of actual phonetic states, it would not be possible to distinguish between proper nasal consonants and nasalized [r]-sounds. This is a consequence of liquids being opposed to nasal consonants in terms of [-nasal] feature alone instead of by a specific complex as the [+vocalic] [+consonantal] in the earlier conventions.

A similar case of defining a phonetic category by the negative of an other not directly related category is the encoding of [r]-sounds as [-lateral]. It is questionable whether an inhibition of the lateral command in the production of an [l] automatically results in an [r]-sound. Additional adjustment may be necessary. These examples are analogous to the [-coronal, +anterior] encoding of labial consonants which I consider more objectional. All these instances of classification in terms of combinations and selections from a finite set are acceptable provided we give up the demand that each feature shall represent an independent and specific production category.

A coding tree related to Table I-A-1 is shown in Fig. I-A-1. The syllabic feature presides in the top but this is not crucial. The same number of yes-no branching points would have been needed if we put the sonority feature on top. Now, coding trees are deceptive in a way since all sorts of variations and hierarchies are possible because of inherent redundancies. However, the manipulation of coding trees has the pedagogical merit of bringing out these redundancies.

Examples of coding trees for the reduced set of features I have proposed are shown in Figs. I-A-2 and I-A-3. In one the syllabic feature is placed on the top, in the other it is given the lowest place and sonorant the top place. The economy in terms of branching points is the same in all the three figures. Figs. I-A-2 and I-A-3 merely have the merit of a smaller number of features. It was actually during the construction of such trees that I observed the complementary distribution of [-continuant] and [+consonantal]. I prefer the tree of Fig. I-A-2 which starts out with the sonorant feature related to vocal tract opening irrespective of where it occurs. Then, logically follows the feature of closure in the vocal tract midsagittal plane, then the manner of release of this closure which applies to [-sonorants] only. The medially closed sonorants are then separated into nasals, laterals, and r-sounds as previously discussed.
Fig. 1-A-1. Coding tree with the basic Chomsky-Halle features, "syllabic" replacing "vocalic".
Fig. 1-A-2. Coding tree with the features conontantal and continuant replaced by a single feature "mid-closure". The feature "sonorant" is given the top level.
Fig. 1-A-3. Alternative coding tree with the same features as in Fig. 1-A-2 arranged in a different order, the feature "syllabic" in the top. Note the relation to Fig. 1-A-1.
and glides are opposed to vowels as nonsyllabic. The main acoustic correlate of voiced sonorants is their higher F1 intensity, whereas the acoustic correlates of "closure" is a reduction of formants higher than F1. The specification of the nasal and the lateral correlates are not so simple. They will not be discussed here.

Some detailed comments

The class of h-sounds has always been a problem in feature analysis. I accept the classification of glides (semivowels) and h-sounds given by Chomsky and Halle as [+sonorant], [-consonantal]* and [-syllabic] but I object to their contrasting of h-sounds to other glides as [+low]. This solution is an apparent mistake since h-sounds display perfect coarticulation with vowels whether [+low] or [-low]. The h-sounds, voiced or unvoiced, are produced with an active glottal readjustment.

The presence of the unvoiced h-sound in the class of sonorants weakens the simple acoustic correlate of intensity if this class since velar fricatives display similar acoustic patterns but with more noise in the region above F2. The degree to which the intensity is associated with the vocalic formant patterns is accordingly a necessary aspect to take into account. This fact also correlates with the affinity of sonorants to be found next to the syllabic nucleus.

Directly related to the classification of h-sounds is the treatment of aspiration. The statement of Chomsky and Halle that a feature of heightened subglottal pressures is a necessary requirement for aspiration is not tenable, see Fant, Acoustic Theory of Speech Production, pp. 277-279. Instead we need a new feature of "glottal relaxation" yet to be defined that covers aspiration in general as well as the class of h-sounds.

On the whole, there is a need for further studies of the phonatory mechanism in various situations before we can single out the various phonetic components involved in the various manner of articulations of stop sounds. The difference between English or Swedish [p, t, k] and [b, d, g] involves both aspiration, tenseness and voicing as phonetic parameters. In initial stressed position the aspiration, i.e. glottal relaxation is the obvious cause of the delay of voicing in [p, t, and k]. A higher intraoral stop pressure, when present, appears to reflect a larger glottal opening

* or [-"midclosure] instead of [-consonantal].
rather than a higher subglottal pressure. At the same time there appears to be a prolongation of the state of articulatory narrowing in [p, t, and k] which accounts for a high frequency "fricative" noise superimposed on the first part of the aspiration.

There are also coarticulation differences. The range of $F_2$-locus at the instant of release is greater for the voiced than for the unvoiced stops, especially so with [b] compared with [p]. This can be seen in the data of Lehiste and Peterson (1961) and I have measured similar distributions for Swedish (forthcoming article). At the instant of release of [b] before a back vowel the tongue takes a position close to that of the following vowel whilst the instant of release of the [p] before the same vowel displays a much higher locus, typical of neutral tongue articulation. After about 40 msec from the release of the [p] the formant pattern follows essentially that observed immediately after the release of the [b]. These temporal relations should be studied closer.

It could be, as stated by Chomsky and Halle, that the amount of vocal wall tensening could affect the possibility to maintain a prevoicing (before the release) but I consider the glottal adjustment to be primary and that it also is the primary cause of the small difference found in the time lag of voicing after release comparing the intervocalic [k, p, t] and [g, b, d] and associated with this time lag a difference in the $F_1$ contour ($F_1$ cut back).

The mere fact that there are certain "tense-lax" elements associated with the distinction between the English or Swedish [k, p, t] versus [b, d, g] in addition to the obvious glottal adjustments is not a sufficient basis for selecting the feature "tense" rather than the feature "voiced". According to Chomsky and Halle the criterion for classifying [p, t, k] as [+tense] rather than [-voiced] would be that vocal vibrations are stopped because of articulatory interaction rather than by glottal relaxation. With this criterion I would lay a greater importance in the voicing component than in the tense-ness component. Further studies are needed.

The feature "distributed" which on the articulatory level is defined as a long versus short constriction in the direction of the air flow has not been analyzed very closely as to its acoustic correlates, and these are far from obvious. Differences in source location, size of front cavity, and the degree of coupling to the back cavities may be affected. A high frequency extension of the noise could be an acoustic correlate but I cannot really say
anything definite before I have studied actual samples of spectrograms and cineradiograms. It appears to me that the main difference between labials and labiodentals is that of a less effective versus a more effective source and I am rather hesitant to equate it with differences in tongue articulations.

In Swedish there are both dental and apical alveolar stops, the latter being lexically induced by a previous /r/. The phonological component would have to work with classifications that differentiate these articulations. It is indeed questionable whether the phonetic difference is that of distributed-nondistributed.

Swedish vowels

The feature "covered" pertaining to narrowed, tensed pharynx wall and an elevated larynx is suggested to have some relevance for the difference between the Swedish vowels [y] and [u]. There is no evidence to support this suggestion as far as I can see.

The Swedish vowel system is of considerable interest in view of the large number of sounds contained. I shall attempt here to construct a phonetic feature matrix of Swedish long vowel phonemes, [u:], [o:], [a:], [e:], [i:], [y:], [a:], [ø:], and the pre-r allophones, [æ:] and [ɔ:] of [ɛ:] and [ø:], respectively. I shall first attempt to use the Chomsky-Halle tongue-body features back, low, high, and the rounding feature. In addition, I have defined two new features, which in the consistent articulatory terminology are named "palatal" and "labial". These function as extreme degrees of tongue-height and lip-rounding, respectively. It has been long recognized that all Swedish long vowels of extreme low first formant frequency, [i:], [y:], [u:], and [u:] are pronounced as diphthongs towards a homorganic glide or fricative. However, what is not so obvious and often overlooked is that the vowel [y:] is made with a palatal closing gesture just as in [i:] but with added lip-rounding and that the front vowel [u:] is produced with a labial gesture towards closure just as in the back vowel [u:], Fant (1968). The historical origin of [u:] is a tongue fronting of [u:] which was replaced by an [ɔ:] in a vowel shift. In the Swedish spoken in Finland [u:] and [u:] are not differentiated and are realized with a single sound shape. The tongue fronting of the "long" [u:] has now progressed to an articulation close to that of [i:], [y:], [e:], and [ø:], generally a little more open than [y:], and a little more close than [ø:]. As far as I can judge the
element of velarization has been completely lost*. The position of the
tongue in the palatal-velar direction is not more "velar" than
that of the other front vowels, and the apex is often slightly raised thus
tending to shift the location of the tongue-palate constriction somewhat an-
terior of [i:]. However, in the class of "short", i.e. lax Swedish vowels,
the tongue of [u]** is lower than that of [o] but more velarized.

When sampling formant data on vowels the distinction between Swedish
[o:] and [u:] and between [o:] and [u:] may be obscured if [u:] and [u:]
are sampled at their onset and not at their target values where F₁ and F₂
are lower. Similarly, the contrast between [y:] and [u:] is increased if
the sampling is performed at the later part of the vowel where F₂ of [u:]
has been progressively lowered and F₃ of [y:] has been progressively in-
creased. At the place of the vowel target the main constriction is at the
lips for [u:] and [u:] but at the tongue-palate region for [y:] and [i:].
The progressively decreasing tongue-height in the series [u:], [o:], [a:] and in [i:], [e:], [ε:], [æ:] and in [u:], [o:], [æ:] is paralleled by an in-
creasing jaw opening, Lindblom (1967). It has been demonstrated by
Lindblom and Sundberg (1969a and b) that with a minimum jaw opening but
otherwise normal tongue movements the F₁ range is considerably reduced.
The jaw opening thus adds not only to the tongue-palate distance but also
to the effective lip-opening, everything else being equal. The six vowel
features classify the Swedish long vowels as follows.

** TABLE I-A-2 **

Swedish long vowels

\[
\begin{array}{cccccccccc}
\text{u:} & \text{o:} & \text{a:} & \text{æ:} & \text{ε:} & \text{e:} & \text{i:} & \text{y:} & \text{u:} & \text{ö:} & \text{æ:} \\
\text{back} & + & + & + & - & - & - & - & - & - & - \\
\text{low} & - & - & + & + & - & - & - & - & - & - \\
\text{high} & + & - & - & - & + & + & + & + & + & - \\
\text{palatal} & - & - & - & - & - & + & + & - & - & - \\
\text{round} & + & + & + & - & - & + & + & + & + & - \\
\text{labial} & + & + & + & + & + & + & + & + & + & - \\
\end{array}
\]

* Lindblom and Sundberg (1969a) classified [u] as "velar" but
expressed doubts as to the phonetic validity.

** The quality of short /u/ is generally transcribed as [e].
In the consonant system the feature "labial" should be used instead of [+anterior][-coronal] to define the class of labial consonants. Labialized vowels are analogous to "retroflex", i.e. [+coronal] vowels. Long (tense) Swedish vowels are accordingly diphthongized if they possess the features "palatal" or "labial". These are the maximally "close" vowels, compare Lindblom and Sundberg (1969 a).

An alternative matrix may be set up with "jaw closure" instead of the "palatal" feature. The maximum degree of jaw closure is found in [i:], [y:], [u:], and [u:] which would be labeled [+closed]. With this solution one gains the distinction in actual tongue-palate opening comparing [u:] and [ɛ:] whilst the distinction between [u:] and [y:] is reduced to one of labialization only. One then has to add the rule that labialization always determines the diphthongal element when present in the close vowels. Note the minimal distinction of [-back] separating [u:] from [u:] in either system. A third and rather different alternative system was suggested by Lindblom and Sundberg (1969 a).

The variety of solutions possible in a system of interrelated physiological dimensions scaled according to binary principles is indeed a problem. One source of variability is that the number of possible combinations generated from a given ensemble is larger than the number of sounds to be encoded. Therefore there may result an ambiguity in feature selection. Two or more physiological parameters may contribute to one and the same acoustical and perceptual effect which may constitute a more natural candidate for the role of feature, at least in the sense of phonetic feature. Let us see what happens if we try to simplify the inventory of articulatory parameters by grouping together the features "low", "high", and "palatal" to one single dimension assigning the value 0 for the most "open" degrees [a:], [æ:], and [œ:] and the value 3 for the maximally palatal [i:]. Similarly the feature labial is added to that of rounding accounting for

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>back</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>high</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>round</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

A matrix of this sort is easier to comprehend than a multidimensional binary system. There are apparently three major classes within the system, the back vowels [u:], [o:], [ə:] in which an increase in tongue
height goes with increasing lip rounding (partially jaw dependent). The unrounded front vowels are differentiated by tongue (and jaw) height and the rounded front vowels are also differentiated by height and by extra rounding as a special feature of [uː], c.f. Malmberg (1956) and Fant (1966).

At this stage we might ask for the acoustic and perceptual correlates of these articulatory categories. The phonetic color is mainly dependent of $F_1$, $F_2$, and $F_3$ but it should be possible to find an optimal projection of this three-dimensional space on a plane. Pilot experiments now in progress at the Dept. of Speech Communication, KTH (Fant, Carlson and Granström) indicate that an $F_1$ versus $F_2$ plot would serve this purpose. $F_2$ is the frequency of the second formant in a two-formant approximation to the vowel. In mid- and back vowels $F_2$ is identical to $F_2$ and in high front vowels close to $F_3$.

A tentative $F_1$ versus $F_2$ plot of Swedish long vowels and some short vowels of specific identity have been plotted on a mel scale, Fig. I-A-4. In this diagram we find evidence of a fairly even spread in the perceptual domain. The average distance between any of the sounds and its closest neighbor is 180 mels. The articulatory correlate of increasing $F_1$ is increasing jaw opening and a shift of tongue place towards a pharyngeal position. The articulatory correlate of the ordinate $F_2$ is a shift of the tongue away from the velum and towards the palate.

It can be seen that back vowels may be separated from front vowels by a line of the slope +45 degrees and rounded vowels from unrounded vowels with a line of -45 degrees slope. Therefore a rotation of coordinates as in Fig. I-A-5 brings out the direct correlates to the main vowel classes. Back vowels are characterized by a distance between the first and the second formant lower than 400 mels. All unrounded front vowels lie close to a line of $M_1 + M_2 = 2200$ mel and the rounded front vowels have an abscissa of $M_1 + M_2$ less than 2100 mel. The quantal steps in the ordinate comparing [iː, eː, ɛː, and æː] are of the order of 250-300 mels whereas the quantal steps in the abscissa are of the order of 200-250 mels. Since we now have condensed the vowel space to a plane we have only two orthogonal parameters.

The abscissa ($M_1 + M_2$) is twice the center of gravity of the spectrum, giving equal weight to $M_1$ and $M_2$, and will be identified with the negative of the old feature "flat". Labialization, velarization, jaw closing, larynx...
Fig. I-A-4. Swedish, long vowels and some short vowels in a $M'_2$ versus $M'_1$ plot (mel scale). $M'_2$ is determined from an analysis-by-(two-formant) synthesis procedure.
lowering will all lower the center of gravity whilst the ordinate, here referred to as the spectral feature "spread" is a measure of dispersan. Note that it is related to but not identical to any of the old features such as [-compactness], [+diffuseness], or [-gravity]. The spectral spread is increased with moving the tongue from a pharyngeal to a palatal place of articulation. Five levels are indicated by the points [a:], [e:], [e:], [i:]. Note that increasing jaw opening increases in the first hand $M_1$ and thus makes the spectrum less flat and less spread. Fig. I-A-5 would motivate a quantization of the long vowels in scales of "flat" and "spread" as follows.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;spread&quot;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>&quot;flat&quot;</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

These scales are absolute but can of course be reduced according to the principle of complimentary distributions. The progressing "flatness" from [a:] over [o:] to [u:] is the effect of rounding + velarization whereas the flatness of [u:] is primarily a matter of small lip opening. As previously discussed no velarization appears to be involved in [u:] but possible an "anteriorization". The possibility of compensatory forms of articulations in the flatness domain are apparent. In the class of "short" i.e. lax Swedish vowels, the /u/, phonetically [e] is more "velar" than the short [ø], see Fig. I-A-6. These facts support a perceptual rather than an articulatory feature basis.

It has often been suggested that articulatory descriptions of vowels actually rely on underlying perceptual classifications, Ladefoged (1967b). Our data indicate that the Swedish vowels are not arbitrarily spaced individuals in the space of physically producible sounds but show a clear organization in terms of linear sequences and a tendency of equalistant spacings in an orthogonal perceptual space. This ordering appears to be a subset of a language universal system of maximal contrast. This idea was also expressed by Lindblom and Sundberg (1969 a). Further work along these lines is continuing. Earlier work on mel scale mapping of Swedish vowels was published by Fant (1959).

references on next page
Fig. I-A-5. Swedish vowels in a "spread" versus "flat" mel scale plot: bringing out some orthogonal vowel categories (back and front vowels) and a tendency of equidistant mel spacings.
Fig. I-A-6. X-ray tracings of Swedish vowels. (From Fant, G.: "The acoustics of speech", in Proc. of the Third International Congress on Acoustics, Stuttgart 1959, pp. 188-201, Fig. 9, Amsterdam 1961.)
References:


