

# Quantal Phonetics and Distinctive Features: a Review

George N Clements<sup>1</sup>, Rachid Ridouane<sup>1,2</sup>

<sup>1</sup>Laboratoire de phonologie et phonétique  
(UMR 7018, CNRS/Sorbonne Nouvelle, Paris)

<sup>2</sup>ENST/TSI/CNRS-LTCI (UMR 5141, Paris)

## Abstract

This paper reviews some of the basic premises of Quantal-Enhancement Theory as developed by K.N. Stevens and his colleagues. Quantal theory seeks to explain why some articulatory and acoustic dimensions are favored over others in distinctive feature contrasts across languages. In this paper, after a review of basic concepts, a protocol for quantal feature definitions is proposed and problems in the interpretation of vowel features are discussed.

## 1. The quantal basis of distinctive feature

Though most linguists and phoneticians agree that the distinctive features of spoken languages are realized in terms of concrete physical and auditory properties, there is little agreement on exactly how they are defined. According to a tradition launched by Jakobson and his collaborators (for example, Jakobson, Fant and Halle 1952), features are defined mainly in the acoustic (or perhaps auditory) domain. In a second tradition initiated by Chomsky and Halle (1968), features are defined primarily in articulatory terms. After several decades of research, these conflicting approaches have not yet led to any widely-accepted synthesis.

In recent years, a new initiative has emerged within the framework of the Quantal Theory of speech, developed by K.N. Stevens and his colleagues (e.g. Stevens 1989, 2002, 2005). This theory maintains that the universal set of features is not arbitrary, but can be deduced from the interactions between the articulatory parameters of speech and their acoustic effects. The central claim is that there are phonetic regions in which the relationship between an articulatory configuration and its corresponding acoustic output is not linear. Within such regions, small changes along the articulatory dimension have little effect on the acoustic output. It is such regions of acoustic *stability* that define the articulatory inventories used in natural languages. In other words, these regions form the basis for a universal set of distinctive features, each of which corresponds to an articulatory-acoustic coupling within which the auditory system is insensitive to small articulatory movements.

A simple example of an acoustic-articulatory coupling can be found in the parameter of vocal tract constriction. Degrees of constriction can be ordered along an articulatory continuum extending from a large opening (as in low vowels) to complete closure (as in simple oral stops). In most voiced non-nasal sounds, the passage along a scale of successively greater degrees of constriction gives rise to three relatively stable acoustic regions with separate and well-defined properties. Sounds generated with an unobstructed vocal tract constriction, such as vowels, semivowels, and liquids, are classified as sonorants. A sudden change in the acoustic output occurs when the constriction degree passes the critical threshold for noise production (the Reynolds number, see Catford 1977), giving rise to continuant obstruent sounds (fricatives). A further discontinuity occurs when the vocal tract reaches complete closure, corresponding to the configuration for noncontinuant obstruents (oral stops). These relations are shown for voiced sounds in Figure 1, where the three stable regions correspond to the three plateaux.

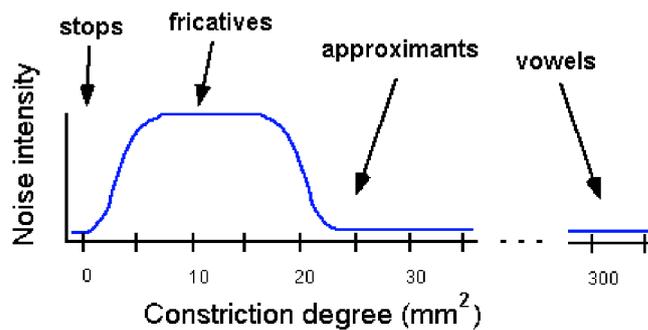


Figure 1. Continuous changes along the articulatory parameter "constriction degree" define three stable acoustic regions in voiced sounds.

In *voiceless* sounds, the falling slope in this figure shifts some distance to the right (to around 90  $\text{mm}^2$ ), and the region between the shifted and unshifted slopes (about 20 to 90  $\text{mm}^2$ ), corresponding to voiceless noise production, defines the class of approximant sounds (liquids, high semivowels, etc.), whose acoustic realization is noiseless when they are voiced but noisy when they are voiceless (Catford 1977).

Languages prefer to exploit articulations that correspond to each of the four stable regions defined in this way. These regions give rise to the features which define the major classes of speech sounds, as shown in Table 1. (The feature [+vocalic], used here to define vowels and semivowels, is equivalent to the classical feature [-consonantal]).

Table 1. The four major classes of speech sounds

	stops	fricatives	approximants	vocoids
[continuant] :	no	<b>yes</b>	<b>yes</b>	<b>yes</b>
[sonorant] :	no	no	<b>yes</b>	<b>yes</b>
[vocalic] :	no	no	<b>yes/no</b>	<b>yes</b>

These features are commonly used across languages. All known languages have stops and vowels, and most have fricatives and approximants as well.

## 2. A protocol for quantal feature definitions

A feature definition, if it is quantal, must identify an articulatory continuum associated with one or more acoustic discontinuities, and must specify the range within this continuum that corresponds to relatively stable regions in the related acoustic output. The range is the articulatory definition of the feature, and the associated output is the acoustic definition. A feature definition must also identify the stable region in terms specific enough to distinguish it from other regions, yet general enough to apply to all articulations within this region, allowing for observed crosslinguistic variation. It must effectively distinguish segments bearing this feature (e.g. /t<sup>h</sup>/) from otherwise similar segments that do not (e.g. /t/). Finally, it must identify the classes of sounds in which the definition holds. This will usually be the class in which the feature is at least potentially distinctive.

As an example, consider a proposed definition of the feature [+consonantal], which distinguishes true consonants from vocoids (vowels, semivowels) and laryngeals: "The defining acoustic attribute for this feature is an abrupt discontinuity in the acoustic signal, usually across a range of frequencies. The defining articulatory attribute is the formation of a constriction in the oral cavity that is sufficiently narrow to create such an acoustic discontinuity. This description applies to both [-sonorant] and [+sonorant] consonants." (Stevens 2004, B79). This definition conforms to the protocol suggested above. It identifies an articulatory continuum (constriction degree) and identifies the range within this continuum ("narrow constriction") associated with a discontinuity -- specifically, a rapid drop in F1 frequency and amplitude, as further explained and illustrated in the extended discussion of this feature in Stevens (1998), 244-6. It will be noted that this definition is specific enough to distinguish [+consonantal] sounds from other sounds, yet general enough to apply to a variety of realizations, for example by the lips, tongue blade, or tongue body. Finally, the definition is general enough to hold across all consonants, including both obstruents and sonorants.

There are two general families of quantal feature definitions: a) *contextual* definitions, in which the acoustic or auditory cue to the feature can only be detected when the sound bearing the feature occurs in an appropriate context, and b) *intrinsic* definitions, in which the cue can be found within the segment itself. The feature [-consonantal] just discussed is an example of a contextual definition, as the discontinuity in question occurs when the consonantal sound occurs in the context of a nonconsonantal sound (as in *may* or *aim*). A strong advantage of contextual cues is that they are linked to "landmarks" in the signal often associated with phoneme boundaries. Such "landmarks" are perceptually salient and tend to be rich in feature cues. It is suggested that they may facilitate speech segmentation and lexical access (e.g. Huffman 1990, Stevens 2000, 2002).

An example of an intrinsic definition is the following, as proposed for the feature [ $\pm$ back] which distinguishes front vowels from central and back vowels. "[During the] forward displacement of the tongue body, the second natural frequency F2 of the vocal tract passes through the second natural frequency of the airway below the glottis, which we will call F2T, for the second tracheal resonance. For adult speakers, F2T has been observed to be in the range 1400 to 1600 Hz, and it is relatively constant for a given speaker ... As F2 passes through F2T, the spectrum prominence corresponding to F2 often does not move smoothly, but exhibits a discontinuity or abrupt jump in frequency. Thus there tends to be a range of values of F2 within 100 Hz or so where the frequency of the spectrum prominence is unstable. It appears that languages avoid vowels with F2 in or close to this region...and put the F2 their vowels on one side or the other of this region; corresponding to [+back] vowels for lower F2 and [-back] vowels for higher F2. Thus there appears to be a dividing line between two regions with a low F2 for a backed tongue body position and a high F2 for a fronted tongue body position." (Stevens 2004, B79-80)

This definition again follows the protocol. The articulatory continuum is tongue fronting (assuming a central position at rest), and the two stable regions correspond to positions in which the associated F2 is either above or below F2T. The definition is specific enough to distinguish this feature from others, but general enough to apply to various types of front, central and back vowels as well as to the same vowel in different contexts. Finally, it identifies the class of sounds in which the definition holds (vowels). This definition is an *intrinsic* definition, since to apply it we need only examine the internal properties of the vowel. An advantage of using an intrinsic definition in this case is that it accounts for the fact that vowels can be usually be identified as front or back in isolation. Another is that vowels typically occur next to consonants, in which F2 is less prominent or absent. (Landmark effects can be found in front-to-back vowel transitions, as in the transition

from [a] to [i] (Honda & Takano 2006), but vowels in hiatus are too infrequent in most languages to provide a primary basis for feature definition.)

### 3. Quantal acoustic-auditory relations

Further types of discontinuity can be found among certain acoustic-auditory relations (Stevens 1989). We consider an example involving vowels.

Vowels are often considered problematic for quantal analysis and it has been suggested that they may organize themselves instead according to an inherently gradient principle of maximal dispersion in perceptual space (e.g. Lindblom 1986). However, the fact that vowels pattern in terms of natural classes just as consonants do suggests that they are also organized in terms of features (see much phonological literature, as well as Schwartz et al. 1997: 281), raising the question of what these features are, and whether they are also quantal. A proposed quantal definition for the feature [ $\pm$ back] has been cited above, based on a region of F2 instability located in the mid-frequency range. Here we will examine evidence for the same feature from natural acoustic/auditory discontinuities.

Vowel-matching experiments have shown that vowel formant patterns are perceived not just on the basis of individual formant frequencies, but also according to the *distance* between formants. In such experiments, synthetic vowels with several formants are matched against synthetic one- or two-formant vowels. Subjects are asked to adjust the frequency of the only (or the higher) formant of the latter vowel so that it matches the former as closely as possible in quality. Results show that when two formants in the normal range for F1 and F2 are well separated, they tend to be heard as two separate spectral peaks, but when two formants approach each other across a certain threshold value, their mutual amplitude is strongly enhanced and they are perceptually integrated into a single peak whose value is intermediate between the two acoustic formants. The crucial threshold for this integration is usually estimated at a value around 3.5 bark (Chistovich & Lublinskaja 1979). The implication of these experiments is that some aspect of the response of the auditory system undergoes a qualitative change -- a discontinuity -- when the distance between two spectral prominences falls under a critical value.

Experiments with data involving Swedish vowels have confirmed this effect for higher formants as well (Carlson et al. 1970). In these experiments, synthetic vowels with five formants were matched against two-formant synthetic vowels. The first-formant frequency was the same for both vowels. Subjects were asked to adjust the second frequency F2' of the two-formant vowel to give the best match in quality to the corresponding five-formant vowel.

The results of the experiment are shown in Figure 2. Here, the frequencies of the first four formants in Hz are shown as lines and the F2' frequencies of the matching vowel are shown as rectangles. It is observed that when the spacing between F2 and F3 is less than about 3.0 bark, as it was for the front vowels (the first six in the figure), subjects place F2' at a frequency between F2 and F3 for all vowels except /i/. (In /i/, in which F3 is closer to F4 than to F3, they place F2' between F3 and F4.) In back vowels, in which higher formants have very low amplitude, F2' is placed directly on F2.

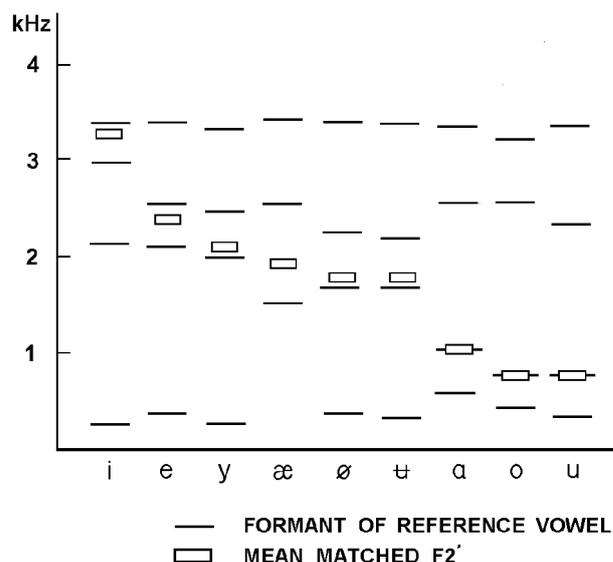


Figure 2. Results of a matching experiment in which subjects adjusted the frequency F2' of a two-formant vowel to give the best match in quality to each of nine Swedish five-formant vowels; only the four lowest formants are shown here. (After Carlson et al. 1970.)

These results indicate that there is a critical spacing of higher formants (F2, F3 and F4) leading to the interpretation of closely-grouped two-peak spectral prominences as single broad perceptual prominences. They give independent support for the view that the feature [ $\pm$ back] has a natural basis, in this case in terms of audition. We see that for [-back], but not [+back] vowels, the distance in Hz between F1 and the effective F2' is always greater than the distance between F1 and the acoustic F2. In other words, perception *magnifies* the front/back vowel distinction present in the acoustic structure.

While the difference between [-back] and [+back] vowels seems well-founded in quantal terms, it is much less clear that other features, such as those of vowel height and lip rounding, can be defined in these terms. For

example, there is no obvious discontinuity in the comparison of Swedish [+high] /u/ and [-high] /o/ in Figure 2. For reasons such as these, phoneticians usually tend to speak of quantal *vowels* rather than of quantal features. Quantal vowels are those in which two formants approach each other maximally, an effect known as focalisation (Schwartz et al. 1997). It is sometimes thought that /i/, /u/, /a/ and perhaps /y/ or /æ/ may constitute quantal vowels in this sense, though experimentally-based, multispeaker data bearing on this question is still rather scarce.

We do not propose, however, to abandon the search for nongradient definitions for vowel features. We tentatively suggest that features of vowel height -- setting aside the problematic feature [ $\pm$ ATR] -- may be defined in terms of the *absolute boundary values* set by the upper and lower range of each speaker. In this point of view, a vowel bearing the feature [+high] would be one whose perceived lowest prominence - let us call it P<sub>1</sub> -- falls within an auditorily indistinguishable subrange of values at the bottom of a given speaker's total range of values for this prominence, while a [+low] vowel would be one whose perceived lowest prominence falls within the corresponding subrange at the top. A mid vowel, bearing the values [-high, -low], would be defined as falling within neither of these subranges. In other words, the speaker's total range of values for a given prominence P<sub>n</sub> establishes the frame of reference with respect to which a given production is evaluated. While this account is not strictly quantal (as there appears to be no natural discontinuity as we pass up and down the vowel height scale), it has the advantage of tying the feature definition to a set of fixed reference points, defined in a way that is applicable to any speaker, regardless of the size and shape of their vocal tract. If it is true that vowel identification is more reliable as a vowel's values approach the periphery of the vowel triangle (see Polka & Bohn 2003), we can explain why distinctions among mid vowels (such as /e/ vs. /ɛ/) are much less stable across languages, in both historical and synchronic terms, than distinctions involving high vs. mid or mid vs. low vowels. These suggestions are quite tentative, of course, and we believe that future research should continue to seek possible quantal correlates of vowel height.

#### 4. Summary

Our aim in this short tutorial has been to present a brief overview of a number of basic concepts of Quantal Theory, proposing a protocol according to which quantal feature definitions may be given. Quantal Theory offers a promising basis for redefining features in both articulatory and acoustic terms, overcoming the tradition competition between these two apparently incompatible approaches.

## References

- Carson, R., Granström, B., and Fant, G. 1970. Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report* 2-3, 19-35. Royal Institute of Technology, Stockholm.
- Catford, J. C. 1977. *Fundamental Problems in Phonetics*. Bloomington, Indiana University Press.
- Chistovich, L.A. and V.V. Lublinskaja. 1979. The "center of gravity" effect in vowel spectra and critical distance between the formants : psychoacoustical study of the perception of vowel-like stimuli, *Hearing Research* 1, 185-195.
- Chomsky, N. and Halle, M. 1968. *Sound Pattern of English*. New York, Harper and Row.
- Honda, K. and Takano, S. 2006. Physiological and acoustic factors involved in /a/ to /i/ transitions. Invited talk, Colloquium on the Phonetic Bases of Distinctive Features, Paris, July 3.
- Huffman, M.K. 1990. Implementation of nasal: timing and articulatory landmarks. *UCLA Working Papers in Phonetics* 75, 1-149.
- Jakobson, R., Fant, C.M., and Halle, M. 1952. *Preliminaries to Speech Analysis*. Cambridge MA, MIT Press.
- Lindblom, B. 1986. Phonetic Universals in Vowel Systems. In J. J. Ohala and J. J. Jaeger (eds.), *Experimental Phonology*, 13-44. Orlando: Academic Press, Inc.
- Polka, L. and O.-S.Bohn. 2003. Asymmetries in vowel perception, *Speech Communication* 41, 221-231.
- Schwartz, J-L., Boë, L-J. Vallée, N. and Abry, C. 1997. The Dispersion-Focalisation Theory of Vowel Systems, *Journal of Phonetics* 25, 255-286.
- Stevens, K.N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Denes, P.B. and David Jr., E.E. (eds.), *Human Communication, A Unified View*, 51-66. New York, McGraw-Hill.
- Stevens, K.N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17, 3-46.
- Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Stevens, K. N. 2000. Diverse Acoustic Cues at Consonantal Landmarks. *Phonetica* 57, 139-51.
- Stevens, K.N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustic Society of America* 111, 1872-1891.
- Stevens, K. N. 2004. Invariance and variability in speech: interpreting acoustic evidence. *Proceedings of From Sound to Sense*, June 11 – June 13, 2004, B77-B85. Cambridge, MA, Speech Communication Laboratory, MIT.
- Stevens, K.N. 2005. Features in Speech Perception and Lexical Access. In Pisoni, D.E. and Remez, R.E. (eds.), *Handbook of Speech Perception*, 125-155. Cambridge, MA, Blackwell.