Quantal Phonetics and Distinctive Features: a Review

George N Clements
Rachid Ridouane

1Laboratoire de phonologie et phonétique (UMR 7018, CNRS/Sorbonne Nouvelle, Paris)
2ENST/TSI/CNRS-LTCI UMR 5141, Paris.

Abstract
The abstract text should be less than 120 words, in Roman New Times (or Times), size 10pt, line spacing single, justified.

1. The quantal basis of distinctive features

Even though most linguists and phoneticians agree that features are defined in terms of concrete physical and auditory properties, there is little agreement on exactly how they are defined. One can distinguish two main lines of "classical" thought on this question. According to the tradition launched by Jakobson and his collaborators (for example, Jakobson, Fant and Halle 1952), features are defined mainly in the acoustic (and occasionally auditory) domain. This approach continues to have adherents, and some researchers insist that "speech perception is hearing sounds, not tongues" (Ohala 1996). According to a second tradition initiated by Chomsky and Halle (1968), features are defined primarily in articulatory terms. This hypothesis, though somewhat less appealing at first sight, is solidly grounded in the motor theory of speech perception as developed by A.M. Liberman and his colleagues at Haskins Laboratories (e.g. Liberman 1996) and underlies contemporary approaches such as the theory of articulatory phonology developed by Browman and Goldstein (e.g. 1989, 1992).

In recent years, a new initiative has emerged within the framework of what is termed the Quantal Theory of speech, inspired by Planck's theory of quanta, as developed by K.N. Stevens and his colleagues (e.g. Stevens 1972, 1989, 2005). Stevens assumes 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 consequences. The equal status accorded to the acoustic and articulatory dimensions of spoken language constitutes a main originality of Quantal Theory. In this view, features are defined with respect to certain articulatory dimensions within which small shifts in the position of the tongue, lips, or vocal cords do not have major consequences for perception.

A central claim in this framework is that there are regions in the phonetic space in which the relationship between an articulatory configuration and its corresponding acoustic outputs is not linear. Within such regions, small changes along the articulatory dimension have little effect on the acoustic output. According to Stevens, 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 to which the auditory system is sensitive.

The articulatory-acoustic coupling is reproduced schematically in Fig. 1, legends (a). It shows that for a series of equal movements along a given articulatory parameter, a discontinuity in the acoustic effect is observed between two stable zones (I and III), separated by a transitional unstable zone (II). An acoustic-auditory coupling is also shown in legends (b), where a discontinuity in auditory response to an acoustic continuum is observed (Stevens 1989). The articulatory and acoustic properties of such stable zones are recruited to define distinctive features across languages.
Figure 1. Schematic representation of the relation between (a) an articulatory and an acoustic parameter and (b) an acoustic and an auditory parameter as the first parameter is manipulated through a range of values.

A simple example of such an acoustic-articulatory coupling is found in the parameter of vocal tract constriction. Degrees of constriction can be ordered along an articulatory continuum extending from a large opening (corresponding to low vowels) to complete closure of the vocal tract (corresponding to ordinary 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 (Reynolds number, see Catford 1997), 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 2, where the three stable regions correspond to the three plateaux.

Figur 2. Continuous changes along the articulatory parameter "constriction degree" define three stable acoustic regions in voiced sounds.

In voiceless sounds, the falling slope 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 noise production, defines the class of approximants sounds, whose acoustic product is noiseless when they are voiced and 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.

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. It 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 crosslinguistic variation. Finally, it must identify the classes of sounds in which the definition holds. This will usually be the class within which the feature is potentially distinctive (Analogous remarks hold for features defined in terms of acoustic/auditory couplings, an example of which is discussed below.)

There are two general families of quantal feature definitions: a) *contextual* definitions, in which the acoustic or auditory cue to the feature only appears 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 [-continuant] is an example of a contextual definition, as the discontinuity only appears when the consonantal sound occurs in the context of a nonconsonantal sound (as in tack [tæk]). 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 (Stevens 1985, 2000, 2002).

An example of an intrinsic definition is the one recently proposed by Stevens for the feature [±back] which distinguishes front vowels from central and back vowels (Stevens 2004: 79-80). 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 those 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 implementations across languages. Finally, it identifies the class of sounds in which the definition holds (vowels). This definition is an intrinsic definition as to apply it we need only examine the steady state of the vowel. One advantage of using an intrinsic definition in this case is that it accounts for the fact that vowels can be easily identified out of context.
Another is that vowels typically occur next to consonants, where no jump in F2 provides a landmark. (Landmarks effects can be found in front-to-back vowel transitions, as in the diphthong [ai] (Honda and Takano 2006), but such sequences are too infrequent to provide a usable basis for feature definition.)

3. Expressing variability within quantal theory

A major challenge to most feature theories comes from the existence of variability in speech output. The realization of a given speech form is never quite the same across utterances, and varies considerably when we take into account differences in speech rate, style, and speaker. It is easily observed that a given distinctive feature may be incompletely realized, or not realized at all, in certain utterances; indeed, whole segments can be deleted in rapid speech, leaving no traces of their defining features. Furthermore, the realization of a given feature is not necessarily the same from one language to another. For example, as noted above, high front vowels tend to be prepalatal in Swedish but mediopalatal for many English speakers, showing that different languages may favor different subareas of the same stable region.

It should be kept in mind that variability is not in the first instance a problem for Quantal Theory, but for the notion of Invariance (e.g. Blumstein and Stevens 1979, Perkell 1986, Hawkins 1999). Quantal Theory attempts to provide a basis for explaining why some articulatory and acoustic dimensions are favored over others for feature definitions across languages, and is not logically committed to the claim that all features are realized with their defining properties in all contexts, situations, and languages. The two views are orthogonal to each other: either can be true or false, independently of the truth of the other. It is entirely possible, therefore, to maintain a strong form of Quantal Theory while attempting to develop independent explanations for variability in feature realization.
Enhancement theory (e.g. Stevens et al. 1986, Stevens and Keyser 1989, Keyser and Stevens 2006; Diehl 1991, Kingston and Diehl 1995) has precisely this mission. Starting out from the premise that much crosslinguistic variation is not random but systematic, it attempts to seek the reasons for which some languages systematically prefer one type of feature realization while others prefer another. According to this theory, when the acoustic difference between two sounds is insufficiently great, risking their confusion, a supplementary gesture may be introduced to increase the acoustic difference between them. This gesture corresponds to a redundant feature in some cases, as when the feature [+rounded] is introduced to enhance the difference between back vowels and front vowels (Stevens, Keyser and Kawasaki 1986). This feature has the effect of increasing the auditory difference between front and back vowels by increasing their difference in second formant (F2) frequency. In other cases, the enhancing gesture may be subfeatural, as in the case of the nondistinctive lip-rounding usually added to palato-alveolar sibilants like /S/ in English, increasing their auditory difference from alveolar sibilants like /s/. In this case, too, the enhancing gesture targets the inherent acoustic parameter distinguishing the two sounds and adds a gesture that increases the difference along this parameter. In the case of coronal sibilants, the universal correlate of a post-alveolar [-anterior] fricative appears to be turbulence noise in the region of F3. Adding lip-rounding to such a sibilant accents the spectral prominence in this frequency region and thus increases its perceptual distance from its [+anterior] counterpart /s/. Such enhancement would not, of course, be expected in languages that have no contrast of this sort. Many additional examples of this sort are provided by Stevens and Keyser in support of Enhancement Theory. For each case, an enhancing articulatory gesture is introduced to increase the perceptual difference between two similar sounds, or classes of sounds.

In addition to reinforcing nonsalient phonological distinctions, enhancement displays another rather unexpected property. While feature-defining gestures are often weakened or obliterated in causal speech due to reduction and gestural overlap, enhancement gestures tend to survive
intact, preserving underlying distinctions. Consider an example from English (Keyser and Stevens 2006). In this language, the distinctive feature [-strident] of /D/, which distinguishes it from [+strident] /z/, is enhanced by a dental or interdental positioning of the tongue blade which counteracts any tendency toward the production of strident noise. As a result, /ð/ is (inter)dental and /z/ is alveolar. In casual speech, the soft palate lowering for /n/ in an utterance like win those may completely overlap the articulation of /ð/, causing it to be realized as a nasal murmur (Manuel 1995). The potential risk is that nasalization suppresses the auditory distinction between win those and win nose. However, the interdental tongue blade position which enhances the underlying defining feature [-strident] of /ð/ remains, as is evidenced in the transition of the second formant into the following vowel at its release. Thus the only evidence for the underlying segment /ð/ resides in the acoustic consequences of the tongue blade gesture introduced to enhance the deleted feature [-strident]. (Note that [-strident] does not characterize the nasal murmur, as stridency is only distinctive in obstruents.)

In sum, enhancement theory offers a principled basis for understanding some types of regular cross-linguistic variation. Drawing upon the notion that languages tend to preserve useful contrasts, it suggests that supplementary features and gestures may be marshalled to reinforce existing contrasts by increasing the auditory distance between two sounds or sound classes along the primary dimension that distinguishes them.

4. Summary
Quantal/enhancement theory provides a promising basis for studying distinctive feature definitions, for explaining preferred types of phonological distinctions across languages, and for understanding certain patterns of system-dependent variability across languages. However, a number of important questions remain open, and must provide the subject matter of empirical studies in the future.
References


Stevens, K.N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. JASA 111, 1872-1891.


