Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 38 (2010) 10–19
Journal of Phonetics
www.elsevier.com/locate/phonetics

Quantal theory, enhancement and overlap

Kenneth Noble Stevens*, Samuel Jay Keyser*

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 10 April 2008; received in revised form 1 October 2008; accepted 7 October 2008

Abstract

This paper explores three aspects of a theory of speech production and perception: quantal theory, enhancement, and overlap. The section on quantal theory makes the claim that every phonological feature or contrast is associated with its own quantal footprint. This footprint for a given feature is a discontinuous (or quantal) relation between the displacement of an articulatory parameter and the acoustical attribute that results from this articulatory movement. The second and third sections address the question of how a listener might extract the underlying distinctive features in running speech. The second section shows that for a given quantally defined feature, the featural specification during speech production may be embellished with other gestures that enhance the quantally defined base. These enhancing gestures, together with the defining gestures, provide a set of acoustic cues that are potentially available to a listener who must use these cues to aid the identification of features, segments, and words. The third section shows that even though rapid speech phenomena can obliterate defining quantal information from the speech stream, nonetheless that information is recoverable from the enhancement history of the segment. We provide examples and discussion in each of these sections of the paper.

1. Introduction

A number of years ago (Stevens, 1972) it was observed that the relations between the acoustic and articulatory attributes of several distinctive features appeared to have quantal characteristics. That is, when a particular articulatory dimension is manipulated through a range of values, there is a nonlinear relation between this dimension and its acoustic consequence. The acoustic parameter is relatively insensitive to the change in the articulatory parameter over one portion of its range and shows a relatively rapid change with articulation over another part of its range. It was proposed that regions of insensitivity of acoustic attributes to changes in articulation could provide a quantitative basis for defining distinctive features. Over the years this initial quantal proposition has been observed to apply to a number of distinctive features across a range of languages (Keyser & Stevens 2006; Stevens, 1989).

In Section 2 we provide a quantal account of a number of distinctive features, and suggest a way of classifying the features based on the quantal articulatory and acoustic parameters that define them. One source of quantal properties arises from aerodynamic and compliant properties of vocal-tract surfaces. The other is related to the acoustic filtering of sound resulting from vocal-tract manipulation. The notion of a defining acoustic and articulatory basis of a distinctive feature is introduced, and its relation to a conventional view of distinctive features is discussed.

In Section 3 we motivate the need to postulate additional acoustic and articulatory attributes that are superimposed on the attributes defined by the quantal relations in order to enhance the perceptual saliency of the underlying features. These enhancing gestures and the resulting acoustic cues are shown to take several forms. Section 4 describes how the overlap of articulatory gestures in running speech can weaken some of the cues available to the listener. Examples are given to show that, in spite of this overlap, enough cues usually remain to permit the listener to uncover the distinctive features intended by the speaker. A summary of the principal points in the paper is given in Section 5.

2. Quantal theory and distinctive features

Fig. 1 gives an example of a relation between an acoustic parameter in the sound radiated from the vocal tract when

*Corresponding authors.
E-mail addresses: Stevens@speech.mit.edu (K.N. Stevens), Keyser@mit.edu (S.J. Keyser).

0095-4470/$ - see front matter © 2008 Published by Elsevier Ltd.
doi:10.1016/j.wocn.2008.10.004

an articulatory parameter is varied continuously through a range of values. For this idealized articulatory/acoustic relation, there is a range of values of the articulatory parameter, designated I, over which there is only a small variation in the acoustic parameter in the sound. Over an adjacent range II, there is a relatively abrupt change in the parameter describing the acoustic result. Over this range the acoustic parameter is quite sensitive to variations in articulation. In the adjacent region III the acoustic parameter, once again, becomes relatively insensitive to articulatory changes. It is hypothesized that an articulatory/acoustic relation of this type defines a distinctive feature. In region I the articulatory and relatively stable acoustic attributes are associated with the minus value for the feature, i.e., [−F], and region III defines [+F].

It is important to note that the feature-defining articulatory/acoustic relation in Fig. 1 is the result of a hypothetical experiment in which just one articulatory parameter is manipulated, with all other articulatory parameters remaining constant. For example, if the articulatory parameter represents the degree of vocal-tract constriction formed by the tongue blade, it is assumed that all other parameters, such as glottal opening, vocal-fold stiffness, stiffness of vocal-tract walls, sub-glottal pressure, etc. remain constant. An articulatory/acoustic relation may be difficult to measure experimentally under such constraints. Consequently, the exploring of defining articulatory/acoustic relations is usually done by modeling the acoustic consequences of various articulatory parameters.

2.1. Some phonological remarks

A word needs to be said about our conception of "defining" features. Phonologists have long made use of the notion of "distinctive feature" in describing processes of natural language—processes like the pronunciation of the plural in English, or vowel deletion in Russian or, diachronically, phenomena as central to Indo-European and Germanic phonology as Grimm's and Verner's Laws. These features have been embedded in rather gross descriptions of the vocal tract and the position of its articulators. Examples include observations that the velum is lowered to produce nasal sounds, that the tongue tip is used to produce coronal sounds, that the tongue body is used to produce dorsal sounds and the like.

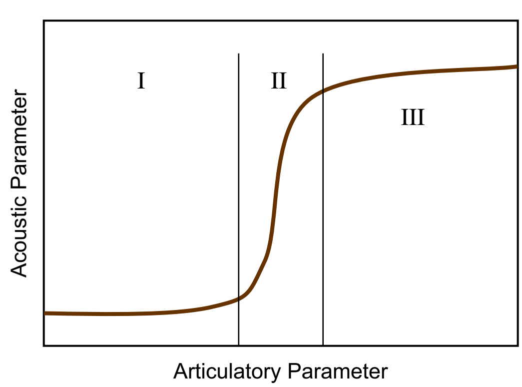

Fig. 1. Hypothetical articulatory/acoustic relation showing two relatively stable regions (I and III) and a region where there is a rapid change in an acoustic parameter for a relatively small change in the articulatory parameter.

Quantal theory can be seen as an independent corroboration of the features that phonologists have made use of over the years. That is to say, these features were not chosen because of the footprint that Fig. 1 suggests they have. Rather, they were chosen because of their efficacy in describing phonological processes like the ones adumbrated in the previous paragraph. That these same features all seem to share a common footprint we take not to be an accident. Quantal theory provides acoustic and articulatory evidence for the phonological features, and, in some cases, could suggest adjustments to some of those features.

2.2. Two sources of quantal relations

As already noted, a quantal relation emerges when a particular articulatory parameter is manipulated along a continuum, and the resulting acoustic parameter exhibits a stable property along one or two ranges of articulation, and more rapid changes along ranges in between these stable regions. The quantal relations that define the distinctive features appear to arise from two physical principles.[1]

The first physical principle involves acoustic coupling between resonating vocal-tract airways. In general, the quantal aspect is a consequence of the movement of zeros in the vocal-tract transfer functions. In English, most of these movements involve features that apply to vowels, place contrasts for obstruent consonants and nasal consonants, and some sonorant consonants such as laterals and rhotics. The place features also include [round], [atr] (also called [tense]), and [nasal]. The defining attributes of these features are frequently observed between abrupt landmarks. This class of features has been called articulator-bound features (Halle, 1992).

The second physical principle involves an aerodynamically based articulator parameter that creates airflows and acoustic sources, and these interact with compliant vocal-tract surfaces. This interaction creates different types of acoustic sources such as (i) quasi-periodic pulses, (ii) turbulence noise or (iii) transients. The acoustic sources can be at the glottis or at narrow constrictions within the vocal tract. The features that are defined by these principles include [stiff vocal folds], [slack vocal folds], [continuant], [sonorant] and [strident]. This class of features has been called articulator-free (Halle, 1992).

[1]In Stevens (2003) three principles were stated as involved in quantal generation. Here we have simplified the matter by merging two of the principles into one. We will assume this merger without justification here since nothing about the present argument hangs on it.

2.3. Examples of quantal articulatory-acoustic relations

2.3.1. Articulator-bound features

One articulatory/acoustic relation with quantal attributes of the type in Fig. 1 can arise in the production of certain vowels due to the coupling between the acoustics of the supraglottal vocal tract and the acoustics of the sub-laryngeal system (Chi & Sonderegger, 2004, 2007; Hanson & Stevens, 1995; Lulich, 2006, 2009). The sub-laryngeal airways are known to have three natural frequencies (for adult speakers) in the principle frequency range for vowels (about 500–2500 Hz) (Cranen & Boves, 1987). These sub-laryngeal resonances vary somewhat from one speaker to another, but are relatively fixed for a given speaker. During vowel production there is always some acoustic coupling between the sub- and supraglottal system. This acoustic interaction is especially strong when the vocal tract is in a configuration for which the second formant, F2, is close to the second sub-glottal resonance (F2sub), which is usually in the range of 1350–1600 Hz.

The effects of this coupling can be seen when the vocal-tract shape is manipulated to produce a tongue-body movement from a backed position to a fronted position within the tract, as illustrated in Fig. 2 for the word *hide*. This articulatory movement causes a gradual increase in the frequency of the second formant. If this articulatory displacement were made with a closed glottis, i.e., with no acoustic coupling between supra- and sub-glottal airways, the spectrum prominence corresponding to F2 would follow a gradually increasing frequency as illustrated by the dashed line of Fig. 2, labeled "model, without coupling." The fixed F2sub as measured for a particular adult male is shown as a constant frequency at 1370 Hz. If there were some acoustic coupling, acoustic analysis shows that the frequency of the F2 prominence in the output exhibits a relatively abrupt jump in the vicinity of the uncoupled value of F2, as indicated by the solid line in the figure (Chi & Sonderegger, 2007). That is, the frequency of this formant prominence is somewhat unstable as the tongue-body position passes through a region in which the original F2 is close to F2sub. Acoustic analysis also shows that the spectrum amplitude of the F2 prominence exhibits a rather abrupt decrease in this region. This behavior of the frequency and amplitude of the F2 prominence can be observed consistently in the production of natural speech when F2 passes through this critical region. The points in Fig. 2 show actual measurements of the frequency of the spectrum prominence for the male speaker. The jump in the frequency of the F2 prominence in this example is about 100 Hz. The acoustic analysis shows that the transfer function from the glottal volume velocity to the vocal-tract output deviates from the all-pole transfer function usually assumed for non-nasal vowels by the introduction of zeros. This deviation is significantly different from the all-pole approximation only at frequencies in the vicinity of F2, where the added zero is well separated from the original pole. The effect of sub-glottal coupling is small at frequencies remote from F2 where there is essentially pole-zero cancellation.

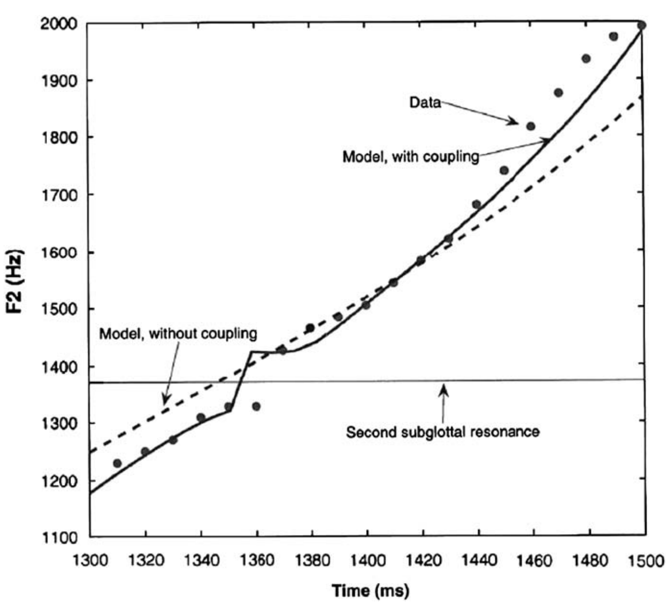

Fig. 2. Observed and modeled F2 tracts (with and without coupling to the second sub-glottal resonance F2sub) for diphthong /ai/ in the word *hide*, spoken by a male speaker. The modeled tract with coupling shows a jump in F2 near F2sub, while the modeled tract without coupling shows no discontinuities. Data were taken once per pitch period, with a window length of one glottal cycle (from Chi & Sonderegger, 2007).

This region of F2 instability corresponds to a boundary that separates vowels that are [+back] from those that are [−back]. Examination of vowel systems for several speakers of English, together with measurements of the sub-glottal resonances for those speakers, shows that F2 for the monophthongal vowels tends to occur outside of this region (Sonderegger, 2004). In other words F2sub defines a region that speakers avoid.

Another type of quantal articulatory/acoustic relation is one that can provide a basis for distinctive features that define the place of articulation for obstruent consonants. For such consonants, the principal acoustic source is turbulence noise generated in the vicinity of the vocal-tract constriction that defines the place of articulation. This acoustic source can be approximated by a sound-pressure source p_s that may be distributed in a region downstream from the constriction (Fant, 1960; Stevens, 1998). The transfer function from this source to the acoustic volume velocity U_m at the mouth opening contains zeros as well as poles, the poles being the natural frequencies of the entire vocal tract. Upstream from the constriction (if the constriction is short), the poles are almost canceled by nearby zeros; these poles result only in weak spectrum prominences at the output. The poles associated with the downstream portion appear as relatively strong spectrum prominences.

The frequencies of these spectrum prominences depend on the length l_{fa} of the front cavity anterior to the constriction, as schematized in Fig. 3a. The lowest of these frequencies F_f is approximately $c/(4l_{fa})$, where $c =$ velocity of sound. For example, if $l_{fa} = 2.5$ cm, as it might be for an alveolar stop or fricative, then F_f would be about 3500 Hz. This is usually a strong spectrum peak in the sound output, and constitutes an important acoustic cue for perception of place of articulation. The next lowest natural frequency in Fig. 3a, F_b, is associated with the cavity behind the constriction, which has a length l_{ba}. This frequency is excited only weakly by the turbulence noise source downstream from the constriction; this is equivalent to saying that in the transfer function from source to output there is a zero that is very close in frequency to this back cavity resonance. If the length l_{fa} of the front cavity is now increased, as in Fig. 3b, the frequency of the front cavity resonance decreases. At some point during this increase in l_{fa}, a pole F_b from the upstream portion of the vocal tract appears now as a downstream spectrum prominence, as shown by the arrow between Figs. 3a and b. Another way of stating this change is that the zero that was formerly close to the pole at F_b in the transfer function is now shifted away from the pole. The original spectrum prominence is shifted downwards in frequency. This abrupt shift in the acoustics as the length of the front cavity increases constitutes a ‘‘quantal’’ change in the output spectrum, and defines a change in a distinctive feature for place of articulation for an obstruent consonant (Stevens, 2003). If in the example given above the front-cavity length increases by 1 cm, and if what was F_b now becomes a front cavity resonance, its frequency is $c/(4 \times 3.5) = 2500$ Hz. This would put this configuration in the range of a velar stop or fricative instead of an alveolar consonant.

A similar kind of argument can be used in describing the place features for a nasal consonant. The nasal consonant is produced with a velo-pharyngeal opening, and with the oral cavity adjusted to a position specified by the place of articulation for the consonant. For a consonant preceding a vowel, there is an initial closure with no airflow beyond this point, but with continuing airflow through the nasal cavity. Immediately following the release, the front cavity is opened and is excited by the sound source generated by

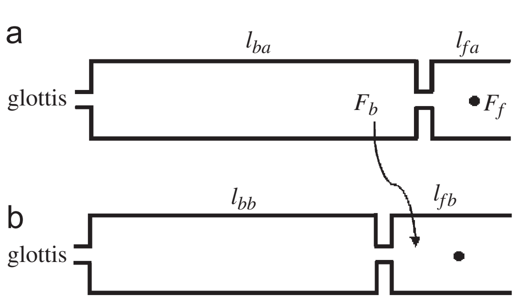

Fig. 3. Schematic representation of a vocal tract with a narrow constriction that simulates a shift in place of articulation for an obstruent consonant. Two configurations are displayed, representing a relatively short front cavity l_{fa} to a longer front cavity l_{fb}, where a back-cavity formant affiliation in (a) changes to a front-cavity affiliation in (b), as shown by the arrow. The dot in the front cavity is assumed to be in the region of the noise source.

glottal flow. This onset of excitation has a spectrum peak at a frequency similar to that of an obstruent consonant with the same place of articulation. Thus, the two consonant types have the same articulator and acoustic signatures with respect to the place of articulation feature.

There are several other quantal relations for articulator-bound features that can be uncovered by introducing modifications to the all-pole vocal-tract transfer function, with the source at the glottis and with the output at the mouth opening or at the nose. In English and in many other languages these appear to define the following features: [+nasal] for vowels, and the ‘‘liquid’’ consonants with the features [+lateral] and [+rhotic]. For all of these features, the articulatory action can be defined by starting with a normal configuration having an all-pole transfer function and no side branches produced by the articulators (ignoring the influence of the sub-glottal system). In order to produce each of these sonorant features a particular perturbation is made in the transfer function for a vowel, introducing a side branch in the acoustic path from the glottis to the lips. This side branch is the path from the velo-pharyngeal port through the nasal passage. For the lateral, a closure is formed along the vocal tract terminated a few cm posterior to the lip opening, creating a zero around 2000 Hz. The rhotic (in English) also has a side branch, usually under the tongue blade, resulting in a pole around 1500 Hz and a zero around 2000–2500 Hz. These pole-zero pairs create stable regions in the articulatory/acoustic relations for many of the articulator-bound features that have been studied.

Also in this class of articulator-bound sonorants are the features [+round] and [+atr]. These can often be considered as modifications of vowel features. The articulatory definition of the feature [+round] is a narrowing of the lip opening together with a lengthening of this opening consistent with the basic articulation of the vowel host. The result is to lower the natural frequencies of the vocal tract. This is usually most evident for F1 and F2, depending on the cavity affiliations of these formants. For the feature [−atr] (sometimes called [−tense]), the modification of the articulation involves a narrowing of the pharyngeal portion of the vocal tract, and some widening of the oral cavity. Consistent with acoustic theory, the acoustic consequence of this perturbation is to produce a vowel with F1 and F2 frequencies that lie inside the vowel triangle relative to those for tense vowels, i.e. with a higher F1 for front vowels and a lower F1 for back vowels. These two features do not have defining quantal attributes that are based on pole-zero combinations. Further studies are needed to determine whether there is a physically based quantal condition that underlies these two features. While we think there is such a basis, we postpone discussion of this question.[2]

[2]Another area altogether is that of tonal distinctions: what if anything is quantal about them? Candidates for quantal characterizations of tone do not leap immediately to mind. It may well be that such systems are outside

2.3.2. Articulator-free features

The articulator-free feature [−sonorant] is implemented by creating a narrow opening within a region of the vocal tract between the lips and the lower pharynx. This opening causes an increase in pressure in the region posterior to the opening. The acoustic consequence of this pressure increase is a reduction in the amplitude of the glottal vibration source relative to that for a vowel. Turbulence noise is also produced in the vicinity of the constriction. The presence of the narrow constriction, sometimes with no glottal vibration, and the consequent pressure behind this opening together constitute the defining articulatory attribute for this [-sonorant] feature. The defining acoustic attribute is the presence of a turbulent noise source.

The feature [+sonorant] is defined only for segments that are [+consonantal] and include nasal consonants, rhotics and laterals. A segment that includes this feature is produced with a glottal source with a low-frequency amplitude that is reduced relative to that for a vowel, but rising exclusively from the glottal source.

Several other distinctive features can be present in a segment that is [±sonorant], including place features as described in Section 2.3.1, and several articulator-free features such as [±stiff vocal folds] indicating the presence or absence of voicing, [±continuant] representing a stop or fricative consonant, and [±strident], representing the presence or absence of an affricate consonant (Clements, 1999). Each one of the features attached to [−sonorant] within a segment has its own articulatory and acoustic definitions, and each exhibits quantal attributes.

Some articulatory/acoustic relations for articulator-free features exhibit a more abrupt change or threshold in the acoustic parameter in region II of Fig. 1, so that a very minor change in articulation in effect creates a different acoustic state, as shown in Fig. 4a. An example of such an abrupt change is the threshold of vocal-fold vibration as the stiffness of the vocal folds increases from a slack configuration (i.e., [−stiff vocal folds]) to a stiffer configuration ([+stiff vocal folds]) (Halle & Stevens, 1971).

In another form, displayed in Fig. 4b, the acoustically stable region may be relatively flat, usually representing a maximum in an acoustic parameter such as the turbulence noise generated at a constriction (Stevens, 1998). To the left of this maximum, as the constriction area becomes small, there is a gradual decrease in the amplitude of the noise, until, for zero area, there is no airflow and no turbulence noise. The same comment applies to the nasal opening; in this case, however, the nasal resonance disappears when the area of the velo-pharyngeal port decreases to zero. This illustrates cases in which setting a constriction area to zero, including flattening of an articulator that creates this constriction in order to guarantee a complete closure, leads to a stable "region" with zero area. In Fig. 4b this creation of a closure with pressure, as for a stop consonant, is represented by zero constriction size. The articulatory/acoustic relations that define the features [±continuant] and [±nasal] are of this type. One end of a continuum represents the closure or end-point of an articulator, and the other end defines a maximum or minimum in an acoustic attribute, but the intervening region may not be as abrupt as that schematized in Figs. 1 and 4a.

(*footnote continued*)
the quantal system altogether, an interesting consequence in its own right. The same applies, of course, to stress systems such as occur in many languages of the world, English included. We simply have nothing to say at the moment about the quantal character of such systems. This does not mean, however, that we are willing to relegate those systems to a non-quantal characterization.

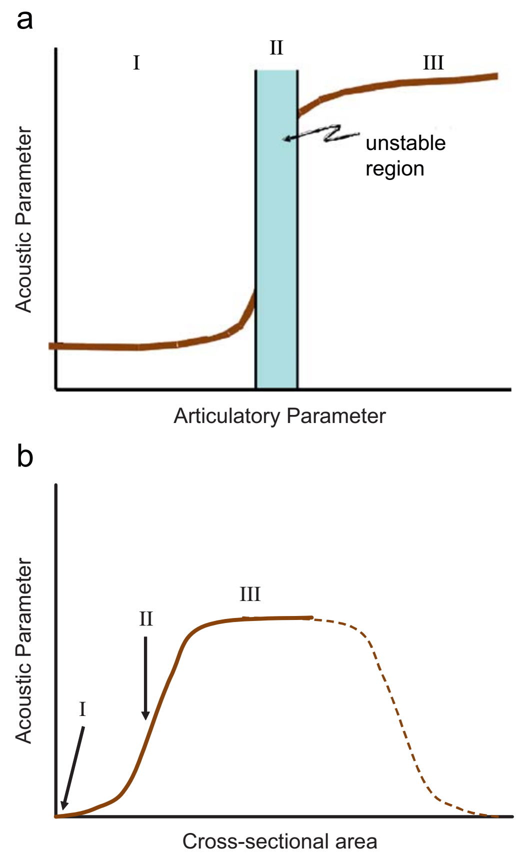

Fig. 4. Two examples of hypothetical articulatory/acoustic relations: (a) a case in which the shift from region I to region III is relatively abrupt, with an unstable region II; (b) at one end of the articulatory/acoustic relation (I) an articulator makes a complete closure leading to a fixed acoustic property, and at the other end (III) a broad maximum in the acoustic parameter is observed, with a relatively gentle change in the intermediate region II. See text.

2.4. Distinctive features and lexical representations

The distinctive features in a lexical item are typically represented in terms of a branching tree diagram. In the discussion to follow we dispense with the tree representation for ease of exposition. We focus instead on the aggregation of features that make up the terminal symbols of any phonological tree, using a feature matrix for our purposes. Consider the matrix representation of the lexical item *seem*, given in Table 1.

Table 1
Distinctive feature bundles that are required for the lexical representation of the word *seem.*.

/s/	/i/	/m/
+continuant	**+syllabic**	**+sonorant**
+stiff	−back	+nasal
+anterior	+high	+labial
+strident	+atr	

The bold-faced features in the second row are articulator-free features, i.e., they specify actions of a class of articulators but do not refer to individual articulators. Every phonological tree contains at least one such feature. Depending on the articulator-free feature for a segment, there are constraints on the possible articulator-bound features. For example a fricative consonant has one articulator-free feature ([+continuant]) but is not specified for the feature [nasal]. And a stop consonant may have only certain contrasts in features relating to place and laryngeal activity (voicing). Most segments, then, may have relatively sparse inventories of features that are distinctive or contrastive.

For the most part, the defining articulatory attribute for a given articulator-bound distinctive feature remains constant regardless of whatever articulator-free feature might co-occur with it in the same feature bundle. For example, for the coronal nasal consonant /n/, which is [+sonorant], the defining acoustic attribute for its place of articulation is the same as that for a coronal stop consonant. For the *stop* consonant, the defining acoustic attribute is the frequency of the spectrum peak in the burst at the consonant release, which reflects the resonance of the cavity anterior to the consonant constriction. For the coronal *nasal* consonant, the defining attribute that reflects the front-cavity resonance is the abrupt appearance of a spectrum peak in the F4 or F5 range at the time of the consonant release, a peak that, like the burst for the stop, also reflects the front cavity resonance and hence the length of the front cavity. This is not unexpected since the defining *articulatory* gesture for the place feature is the same for the nasal and the stop.

Another example is the feature [stiff vocal folds], which defines a contrast for obstruent consonants (i.e. specifies the presence or absence of glottal vibration) and a different kind of acoustic contrast for vowels (i.e. specifies a high tone in contrast to a lower tone). Again, however, the defining articulatory gesture is essentially the same. This distinction does not apply to articulator-free features like [+continuant].

There have been a number of feature systems proposed over the years. The feature system we assume in this paper is essentially that in Chomsky and Halle (1968), although with certain changes; for example, [stiff vocal folds] and [slack vocal folds]. We are unclear about the status of other features; for example, [tense] versus [atr]. We also suspect that other features not mentioned in Chomsky and Halle (1968) may be needed; for example, [dental] may be necessary to account for certain lamino-dental stop consonants in Australian aboriginal languages and elsewhere (Butcher, 2006). Our purpose here, however, is not so much to provide a definitive feature system as it is to demonstrate, we hope convincingly, that whatever features one introduces must show a template of the kind illustrated in Figs. 1 and 4.

2.5. General comments

Sometimes a quantal acoustic property defines an acoustic value or an acoustic range of values by virtue of its stability over this range. Other times a quantal acoustic property results from the avoidance of a numerical region of the vocal tract, for example, F2sub (see discussion in Section 2.3.1). Thus a defining quantal property does not specify an acoustic change or contour, such as the movement of a formant frequency over a particular range or with a certain trajectory. (The role of time-varying parameters as enhancing cues to phonological contrasts will be discussed in the following section on enhancement.)

We hypothesize that a quantal acoustic/articulatory relation underlies each distinctive feature, and consequently each feature can be said to be based on a defining articulatory range and a defining acoustic attribute. This acoustic attribute may depend upon the associated articulator-free feature of the segment. These defining attributes are properties of the human speech production system and are expected to be universal in language. It is hypothesized that the human speech production system is structured in such a way that the sounds that it can generate and the articulatory attributes that produce these sounds define a set of quantal states. As will be noted later, additional acoustic and articulatory attributes may be added in certain contexts to enhance the perceptual saliency of the defining acoustic attribute.

3. Enhancement

Quantal theory seeks to explain why the inventory of distinctive features that make up the phonologies of the languages of the world are what they are. It specifies defining articulatory and acoustic attributes for those distinctive features. It is not intended to be the principal basis of a model that describes how human speakers generate running speech or how listeners extract words from continuous speech. The surface representation of words and word sequences includes not only the feature-defining *acoustic* and *articulatory* attributes but also an array of articulatory gestures (and their acoustic consequences), including prosody, that enhance the *perceptual* saliency of the defining attributes.

There are two general ways in which enhancement gestures may be added to a defining gesture for a particular feature in a particular language.

(1) An articulatory gesture is superimposed on the defining gesture, and thereby enhances the perceptual saliency of the feature. In effect the acoustic attribute resulting from the enhancing gesture increases the perceptual distance between the feature and its neighbors. The enhancing gesture is not the defining gesture for a distinctive feature in that language, and thus by itself does not represent a contrast in the language. This type of enhancing gesture can be graded. It makes an adjustment to the defining acoustic attribute, and is implemented in all contexts in which the feature occurs. When a vowel is adjusted in this way, it has been said that the vowels are adjusted to give a uniform dispersion in the vowel space—the so-called dispersion theory (Diehl, 1991; Liljencrants & Lindblom, 1972; Lindblom, 1986).

An example of this type of enhancement for consonants is the rounding of the lips in the production of /ʃ/. This rounding tends to lower the natural frequency of the anterior portion of the vocal tract, so that the frequency of the lowest major spectrum prominence in the fricative spectrum is in the F3 range, well below the F4 or F5 range for the lowest spectrum prominence for the contrasting fricative consonant /s/. Another enhancing adjustment that achieves a similar effect is the shaping of the tongue blade to assume a domed configuration, thereby creating a longer narrow section in the oral cavity. Both of these gestures create a configuration that strengthens the spectrum prominence in the F3 range.

Other examples can be observed in vowels. In a five-vowel system, the nonlow back vowels are often produced with lip rounding, presumably to enhance the contrast with vowels having the feature [−back] (Keyser & Stevens, 2006; Stevens & Keyser, 1989). Similarly, the nonlow front vowels are often produced with lip spreading, thereby strengthening the acoustic attribute that defines [−back]. For these types of enhancement, the enhancing gesture itself does not create the contrast. These enhancements are usually implemented for all contexts in which the feature occurs.

(2) A second possible type of enhancement for a feature introduces a new acoustic attribute that is separate from the defining acoustic attribute for the feature. The new acoustic attribute created by this type of enhancement introduces additional perceptual cues to the feature. The form this enhancement takes can depend on the context in which the feature occurs. These types of enhancement are introduced in regions of the speech signal that are adjacent to the times when the defining acoustic attributes appear. The enhancements can be time-varying attributes, as opposed to the defining attributes which consist of target acoustic measures.

A typical example of this second type of enhancement is a measure of the movement of the formant frequencies at the release of an obstruent consonant. Such a measure could be the frequency of F2 immediately after the release of a stop consonant. This frequency is related to the length of the vocal-tract cavity behind the constriction. Or the time course of the formant movements may play a role. Another example is the formant movements that may be introduced toward the end of a vowel, as in the nonlow vowels in English. In this case, the [+atr] vowels often are produced with F2 movements toward more extreme values, and the [−atr] vowels show movements toward more central values. These offglides can be regarded as enhancements of the vowel feature [atr]. A number of examples of this second type of enhancement are given in Keyser and Stevens (2006).

There are some articulatory gestures for consonants that produce a particular defining acoustic property during the consonant and a different acoustic property in the vowel interval adjacent to the consonant. We consider the acoustic property in the vowel region to be an enhancement. This property provides a cue that is used by a listener to help to identify the consonantal segment. Examples of this kind include the feature [+stiff vocal folds] (as noted above), which causes an inhibition of glottal vibration in the consonant obstruent region and an increased fundamental frequency of glottal vibration in the following vowel adjacent to the consonant (House & Fairbanks, 1953). The pattern of this fundamental frequency contour may be under the control of the speaker, and may be language dependent. Another example is the feature [+nasal] for a consonant which has a particular defining acoustic attribute in the nasal murmur and a somewhat different enhancing attribute in the adjacent vowel region. There is evidence, however, that in cases like these, the timing and extent of the attribute in the adjacent vowel is not completely automatic, but rather, depending upon the language, is under speaker control (Butcher, 1999). In some cases, it may be influenced by prosodic factors (Hanson, 2004).

The acoustic manifestations of alveolar consonants in English (particularly the voiceless /t/) exhibit a much wider range of variability than that observed for other places of articulation for consonants. In some contexts these alveolar consonants have acoustic properties (and corresponding articulatory attributes) that appear to be related only indirectly to the defining attributes for the place and voicing features for the tongue-blade consonants. For example, the flap that is often used intervocalically in a post-stressed context (like *writer* or *rider*) signals the place feature, but often without a burst, although there is still evidence for a coronal place of articulation. Or, in word-final position like /t/ in *that boy*, there may be no alveolar closure for /t/, but often a glottal stop is produced in that position. Or again, in a word like *Alvin*, the alveolar lateral consonant /l/ is often produced with no tongue blade closure.

We assume that the alternation of /t/ and /ʔ/ is not due to a phonological rule but rather to enhancement. If it were a phonological rule, then the phonology would presumably need a rule that takes /æl/ to /æᵘ/ in *Alvin*. These two phonological rules would be unrelated. They would, in fact, not capture what we take to be a real commonality between them; namely, they both involve the same phenomenon, the failure of a coronal to close completely in normal speech. This failure results from overlap (see next section). From this perspective the introduction of a glottal stop in *batboy* becomes an automatic consequence of needing to signal [−continuant] for /t/ because the usual method, i.e. coronal closure, has been obviated by closure failure just as it has in *Alvin*. The difference in the two cases is that only the former is [−sonorant]. Because of these variations, a suggestion (with which we tend to agree) has often been advanced that the alveolar places of articulation (at least for stops) be given special status (cf. Butcher, 2006; Lahiri & Reetz, 2002; Paradis & Prunet, 1991).

4. Overlap

We have observed that multiple cues may be available to a listener to help identify the distinctive features that underlie the segments in an utterance. Some of these cues are directly related to the definition of the feature based on quantal articulatory/acoustic relations. Other cues can be regarded as having an enhancing role that contributes to the perceptual saliency of the distinctive feature. These various cues for a given feature may be distributed over time. For example, in running speech, there is often overlap of the articulatory gestures that produce these acoustic cues in adjacent segments. A consequence of this overlap is a weakening of some cues and sometimes a masking or obliteration of cues. The discussion that follows illustrates these options.

A simple example of articulatory overlap occurs in an utterance containing a sequence of two stop consonants, as in the casually produced utterance *top tag*. A spectrogram of this utterance is shown in Fig. 5. Each of the stop consonants like /p/ and /t/ is normally defined by a particular type of noise burst—a relatively flat spectrum for /p/ and a spectrum with greater amplitude in the high-frequency range for /t/. If a consonant like /p/ were in intervocalic position, some enhancing attributes would be generated as the articulators move from the region associated with the preceding vowel to the region of the defining gesture. Other enhancing gestures occur during the transition to the following segment. In this example, the transition toward the labial closure for /p/ generates enhancing cues for the labial place of articulation. However, the noise burst that would normally signal the labial place of articulation is obliterated because the tongue blade closure for /t/ occurs before the lip closure for /p/ is released, i.e., the two closures overlap. Any cue for the labial place of articulation immediately prior to the /t/ release is probably also obscured. In the case of /t/, there is little direct evidence of the presence of the alveolar place during the time preceding the /t/ release. The alveolar

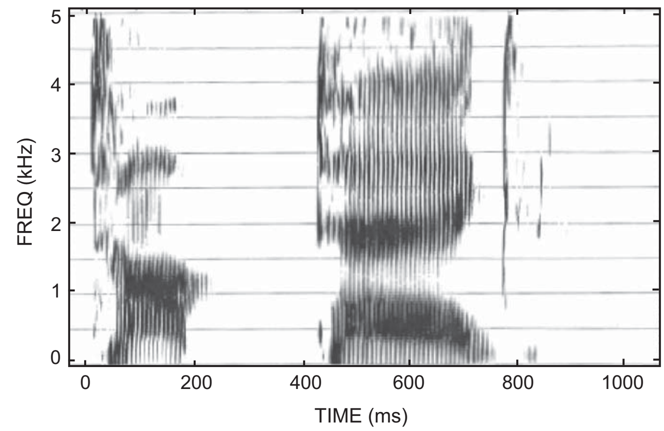

Fig. 5. Spectrogram of casually spoken version of *top tag* produced by a male speaker. The /p/ closure (≈200 ms) and the /t/ release (435 ms) are evident, but the /p/ release and the /t/ closure are not evident in the acoustic pattern. See text.

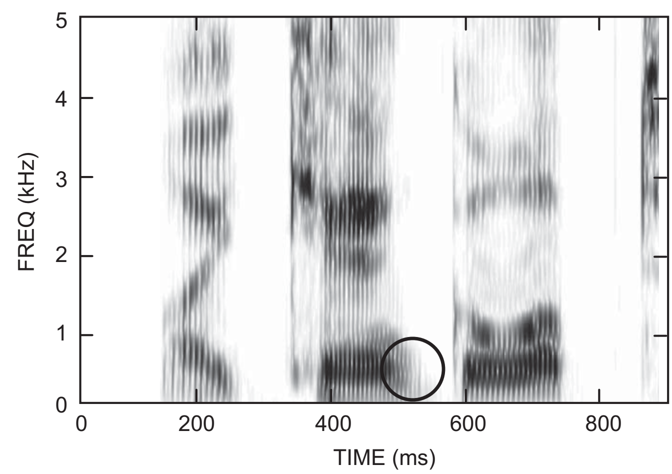

Fig. 6. Spectrogram of casually spoken version of *I can't go up* produced by a male speaker. See text for description of acoustic modifications due to articulatory overlap. The circled area of the figure encloses a section of the cluster /nt/.

burst, however, provides strong evidence for alveolar place, as does the transition from this burst into the following vowel /æ/. Thus some cues exist for /t/, but only weaker cues for /p/. The "defining" cue for /p/ is actually obliterated.

Perhaps a more extreme example of gestural overlap occurs with a casual production of the sequence *I can't go up* (see Fig. 6). Such a sequence can sometimes be produced with no alveolar closure to provide evidence for the cluster /nt/ as in the circled area of the figure. However, the vowel /æ/ is nasalized over much of its length, and the vowel ends with glottalization. In spite of these apparent modifications or deletions of significant cues for the features [nasal], [tongue blade] (for nasal consonant), [−continuant] (for alveolar consonant), and [tongue blade] (for alveolar consonant), there are still sufficient cues for a listener to decode the utterance. The nasalization of the vowel /æ/ can be interpreted as an enhancing attribute indicating the presence of a nasal consonant; the glottalization is an enhancing attribute for a syllable-final /t/; and phonotactics require that the preceding nasal consonant be /n/. Thus for this sequence of four syllables and nine segments, the defining attributes are obliterated for features in two of the segments, but the enhancing attributes for these features contain sufficient cues to preserve intelligibility of the phrase.

Another common example is the overlap of a sequence of a reduced vowel /ə/ and a following nasal consonant /n/, to produce a syllabic /n/ as in the word *lesson.* This syllabic nasal contains acoustic cues for the presence of a vowel (a maximum in low-frequency amplitude) and for a nasal consonant (a nasal murmur). However, the low-frequency spectrum prominence for the syllabic nasal is below the frequency normally required for a vowel, and there is no abrupt spectrum discontinuity that is a defining attribute for a consonant. Nevertheless there are sufficient enhancing cues that the syllabic nasal can be identified as a sequence of reduced vowel plus nasal consonant.

Examples of a different kind involve the various acoustic manifestations of the segment /ð/ in English, usually in function words. These include the apparent stop-like version in a sequence like *back the team* (Zhao, 2007), or the nasal version in the sequence *win those games,* or the apparent lateral manifestation in *will they come* (Manuel, 1995; Manuel & Wyrick, 1999). We assume the underlying distinctive feature composition of /ð/ to be [+continuant, −strident]. In the cases we are considering the feature [+continuant] disappears. Instead we find either a stop-like, or nasal-like, or lateral-like [−continuant] consonant that has taken on the continuancy of the preceding segment. Thus an enhancing gesture is needed to recover the underlying [+continuant] feature. This gesture appears in the second formant prominence at the release of the consonant into the following vowel. This frequency is lower than what would be observed if the consonant were produced as an alveolar consonant such as /t/ or /n/ and is a link to the underlying [+continuant, −strident] feature composition of the /ð/ segment.

5. Conclusion

The theoretical framework presented above rests upon the following:

1. The anatomy and physiology of the human sound-generating system assumes a set of discrete "states" based on "quantal" relations between certain articulatory parameters and their resulting acoustic properties. Each of these states defines the basic articulatory/acoustic attributes for the distinctive features that make up the universal inventory of phonemic contrasts available for use in language.
2. These defining acoustic and articulatory attributes may be augmented by the introduction of additional gestures that increase the perceptual saliency of the feature. The enhancing gestures may be language dependent and may depend on the context in which the feature appears.
3. In running speech the acoustic manifestations of a given distinctive feature for an underlying segment in an utterance may be modified by gestural overlap. Enhancing acoustic cues usually preserve evidence for the distinctive feature, even though the defining acoustic cue is weakened or even obliterated.
4. The model toward which we are moving, then, may be characterized as one in which:
   a. Underlying representations are entirely feature-based and contain only distinctive features.
   b. Differences between underlying and surface representations are mainly due to strategies of enhancement and overlap, which introduce, delete or extend gestures, but do not operate on features.

Acknowledgements

This research was supported in part by funds from NIH Grant no. DC00075. We are indebted to Keith Johnson and an anonymous reviewer for perceptive criticism and helpful suggestions and to G.N. Clements over the years for insightful discussion and criticism that has illuminated the direction of this work. We would also like to thank the members of his research group at the University of Paris for their invaluable contributions to our deeper understanding of these issues.

References

Butcher, A. R. (1999). What speakers of Australian aboriginal languages do with their velums and why: The phonetics of the nasal/oral contrast. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences, ICPhS* (pp. 479–482). Berkeley, CA.

Butcher, A. R. (2006). Australian aboriginal languages: Consonant-salient phonologies and the 'Place-of-Articulation Imperative'. In J. M. Harrington, & M. Tabain (Eds.), *Macquarie monographs in cognitive science. Speech production: Models, phonetic processes and techniques*. London: Taylor & Francis.

Chi, X., & Sonderegger, M. (2004). Subglottal coupling and vowel space. *Journal of the Acoustical Society of America*, *115*(5), 2540.

Chi, X., & Sonderegger, M. (2007). Subglottal coupling and its influence on vowel formants. *Journal of the Acoustical Society of America*, *115*(5), 1735–1745.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. (1999). Affricates as non-contoured stops. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Proceedings of LP '98: Item order in language and speech* (pp. 271–299). Prague: Karolinum Press.

Cranen, B., & Boves, L. (1987). On subglottal formant analysis. *Journal of the Acoustical Society of America*, *84*, 734–746.

Diehl, R. (1991). The role of phonetics within the study of language. *Phonetica*, *48*, 120–134.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Halle, M. (1992). Features. In W. Bright (Ed.), *Oxford international encyclopedia of linguistics*. New York: Oxford University Press.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report*, *101*, 198–213.

Hanson, H. M. (2004). The feature [stiff] interacts with intonation to affect vocal-fold vibration characteristics. *Journal of the Acoustical Society of America*, *116*(4), 2546.

Hanson, H. M., & Stevens, K. N. (1995). Sub-glottal resonances in female speakers and their effect on vowel spectra. In *Proceedings of ICPhS-95* (Vol. 3, pp. 182–185). Stockholm.

House, A. S., & Fairbanks, G. (1953). The influence of consonantal environments upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *25*, 105–113.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language*, *82*, 33–62.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 637–676). Mouton, Berlin.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839–862.

Lindblom, B. (1986). Phonetic universal in vowel systems. In J. J. Ohala, & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Orlando, FL: Academic Press.

Lulich, S. (2006). The role of lower airway resonances in defining feature contrasts. Ph.D. dissertation, Massachusetts Institute of Technology.

Lulich, S. M. (2009). Subglottal resonances and distinctive features. *Journal of Phonetics*, this issue, doi:10.1016/j.wocn.2008.10.006.

Manuel, S. Y. (1995). Speakers nasalize /ð/ after /n/ but listeners still hear /ð/. *Journal of Phonetics*, *43*, 453–476.

Manuel, S. Y., & Wyrick, G. C. (1999). Casual speech: A rich source of intriguing puzzles. In *Proceedings of ICPhS '99* (Vol. 1, pp. 679–682). San Francisco.

Paradis, C., & Prunet, J. -F. (Eds.). (1991). *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Sonderegger, M. (2004). Subglottal coupling and vowel space: An investigation in quantal theory. Bachelor's thesis in Physics, Massachusetts Institute of Technology.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In P. B. Denes, & E. E. David Jr. (Eds.), *Human communication: A unified view* (pp. 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. (2003). Acoustic and perceptual evidence for universal phonological features. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 33–38). Barcelona.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language*, *65*, 81–106.

Zhao, S. (2007). The stop-like modification of /ð/: A case study in the analysis and handling of speech variation. Ph.D. dissertation, Harvard-MIT Division of Health Sciences and Technology.