QUANTAL THEORY, ENHANCEMENT AND OVERLAP

Kenneth N. Stevens and Samuel Jay Keyser

December 5, 2006

1. INTRODUCTION

A number of years ago (Stevens1972) 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 (Stevens 1989, Keyser & Stevens 2006). Based on these observations, we have attempted in Section 2 of this paper to define more clearly some of the general properties of the quantal relations and to indicate how the "quantal theory" is one aspect of a more general theory that links the discrete abstract phonological representation to human production and perception of the speech signal.

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

Figure 1 gives an example of a relation between an acoustic parameter in the sound radiated from the vocal tract when 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 Figure 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, subglottal pressure, etc. remain constant. An articulatory/acoustic relation is 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 movements. With proper interpretation, however, the nature of certain articulatory/acoustic relations can be inferred experimentally.

The canonical articulatory/acoustic relation in Figure 1 can take various forms. In one form, already displayed in Figure 1, the relation shows two regions (I and III) within which the acoustic attribute is relatively stable over a range of articulatory displacements, and one region (II) in which small changes in articulation lead to relatively large changes in the acoustic parameter. An example of a quantal relation that takes this form is the influence of acoustic coupling between the second subglottal resonance and the second formant F2 of the vocal tract proper, which leads to avoidance of a particular range of F2 frequencies --- a range that separates vowels that are [-back] from vowels that are [+back] (Chi and Sonderegger 2004). Another example is the relatively abrupt relation between

formant number (e.g., F4 and F3) as the length of the front cavity for obstruent consonants changes to form a [+anterior] to a [-anterior] feature (Stevens 2003).

Some articulatory/acoustic relations exhibit a more abrupt change or threshold in the acoustic parameter in region II, so that a very minor change in articulation in effect creates a different acoustic state, as shown in Figure 2a. 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 more stiff configuration ([+stiff vocal folds]) (Halle and Stevens 1971).

In another form, displayed in Figure 2b, the acoustically stable region is 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 velopharyngeal 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 Figure 2b 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 Figures 1 and 2a.

By defining a quantal acoustic property in the manner illustrated in the above examples, it is implied that this defining attribute has a particular numerical value or has a value in a particular range, possibly with appropriate normalization to the speaker. 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 acoustic attribute and a defining articulatory range. 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.

The distinctive features that emerge from the quantal relations are of two types (Halle 1990). For one type, called articulator-free features, the "articulator" in the

acoustic/articulatory relation specifies a class of articulatory actions independent of the articulator that performs the actions. These include the features [vowel], [consonant] and [glide], and, within the consonant class, the features [sonorant] and [continuant]. The defining acoustic attribute for segments with the consonantal feature is an abrupt change in the spectrum. The defining acoustic attribute for the articulator-free feature for vowels is a maximum in low-frequency amplitude in the region of the first formant. These attributes of abrupt spectrum change or maximum amplitude are called *landmarks*.

A second type of feature, called articulator-bound, specifies actions of articulators. Some of these articulators are peripheral to the vocal-tract proper. These include the features [stiff vocal folds], [spread glottis], [constricted glottis], and [nasal]. The defining acoustic attributes for these features are manifested in the speech signal in the vicinity of landmarks, usually in the low-frequency region. Other articulator-bound features specify actions of articulators that are part of the vocal-tract proper. They are frequently called place features. These specify the place of articulation of a constriction and the shaping of the articulator that forms the constriction. The defining acoustic attributes for these features describe the spectrum shape in the vicinity of the consonant landmark in the middle- and high-frequency region of 1 kHz to 8 kHz.

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.

/s/	/i/	/m/
+continuant	**+syllabic**	**+sonorant**
+stiff	-back	+nasal
+anterior	+high	+labial
-dental	+tense	

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

The bold-faced features in the second row are the articulator-free features. (Every phonological tree contains 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 [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 labial nasal consonant

/m/, which is [+sonorant], the defining acoustic attribute for its place of articulation is different from that for a labial stop consonant. The defining *articulatory* gesture for the place feature, however, 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.

There are other approaches to lexical representation. For example, Browman and Goldstein (1992) take the view that articulatory gestures are, in fact, primary and that lexical representations exist in terms of gestures rather than features. While we disagree with the status they accord articulatory gestures, we agree that articulatory gestures are an important part of a full description of lexical representation.

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) 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 (Liljencrants & Lindblom 1972, Lindblom 1986, Diehl 1991).

An example of this type of enhancement for consonants is the rounding of the lips in the production of /_/. 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] (Stevens & Keyser 1989, Keyser & Stevens 2006). 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 attributes 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 the movement of the formant frequencies at the release of an obstruent consonant. These formant transitions may indicate the starting frequency of a formant at the time of consonant release and the frequency of the formant in the following vowel. 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 tense vowels often are produced with F2 movements toward more extreme values, and the lax vowels show movements toward more central values. These offglides can be regarded as enhancements of the vowel features [high] and [tense]. 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] 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 and Fairbanks 1953). 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 in unstressed 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.

It has often been suggested that the alveolar places of articulation (at least for stops) be given special status (cf. Paradis & Prunet 1991, Butcher 2006). One proposal is that this stop place be considered as a sort of default place of articulation. In this view cues for labial and velar stops follow the normal pattern of defining gestures supplemented by enhancements, whereas the alveolar consonants are segments that do not have the attributes expected of labials and velars, but nonetheless are stops (cf. Lahiri and Reetz 2002).

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 feature. These various cues for a given feature may be distributed over time. Some can be "internal" to the segment and not strongly influenced by neighboring segments, and others are more associated with the edges of the segments, and may be more context-dependent. In running speech, there is often overlap of the articulatory gestures that produce these various acoustic cues in adjacent segments. A consequence of this overlap is a weakening of some cues and sometimes a masking or obliteration of cues.

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 Figure 3. 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. 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 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* (see Figure 4). Such a sequence can sometimes be produced with no alveolar closure to provide evidence for the cluster /nt/. 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 three syllables and six 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*, or the nasal version in the sequence *win those games,* or the apparent lateral manifestation in *will they come* (Manuel 1995). In these cases the [+continuant] feature for the obstruent consonant is not realized as such and appears as stop-like or nasal-like or lateral-like. However, we think that the place feature for /ð/ should be classified as [+dental], which distinguishes /ð/ from the [-dental] /s/. An enhancing gesture for this feature 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/.

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 preserves evidence for the distinctive feature, even though the defining acoustic cue is weakened or even obliterated.

6. ACKNOWLEDGEMENTS

This research was supported in part by funds from NIH Grant No. DC00075.

REFERENCES

Browman, C. & L. Goldstein (1992) Articulatory phonology: an overview. *Phonetica* 49:155-180.

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 & AC Bailey (eds) Proceedings of the XIVth International Congress of Phonetic Sciences Berkeley: ICPhS, 479-482.

Butcher, A.R. (2006) Australian Aboriginal languages: consonant-salient phonologies and the 'Place-of-Articulation Imperative.' In J.M. Harrington and M. Tabain. eds. *Macquarie Monographs in Cognitive Science. Speech Production: Models, Phonetic Processes and Techniques.* Taylor & Francis. London.

Chi, X. & M. Sonderegger (2004) Subglottal coupling and vowel space. Paper presented at the 147th Meeting of the Acoustical Society of America, New York, NY May 24-28, 2004.

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

Halle, M. (1990) "Features." In W. Bright. *Oxford International Encyclopedia of Linguistics*. New York. Oxford University Press.

Hanson, H.M. (2004) The feature [stiff] interacts with intonation to affect vocal-fold vibration characteristics. Presented at the 148th Meeting of the Acoustical Society of America, San Diego, California, November 15-19, 2004.

House, A. S. & G. Fairbanks (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. & K.N. Stevens (2006) Enhancement and overlap in the speech chain. *Language* 82.1.33-62.

Lahiri, A. & H. Reetz (2002) Underspecified recognition. In C. Gussenhoven, Natasha Werner, and Toni Rietveld. eds. *Labphon 7*.637-676, Mouton. Berlin.

Liljencrants, J. & B. Lindblom (1972) Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language.* 48.839-862.

Lindblom, B. (1990) Explaining phonetic variation: A sketch of the H & H theory. In W.J. Hardcastle and A. Marchal. eds. *Speech Production and Speech Modeling*. Kluwer. Dordrecht. 403-439.

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

Paradis, C. & J.-F Prunet. (1991) Eds. *The Special Status of Coronals: Internal and External Evidence.* Academic Press. San Diego.

Stevens, K. N. (1989) On the quantal nature of speech. Journal of Phonetics 17. 3-46.

Stevens, K.N. (2003) Acoustic and perceptual evidence for universal phonological features. Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, pp. 33-38, August 3-9, 2003.

Stevens, K. N. (1998) *Acoustic Phonetics*. MIT Press, Cambridge, MA.

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

FIGURE LEGENDS

Figure 1. Hypothetical acoustic/articulatory 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.

Figure 2. Two examples of hypothetical acoustic/articulatory 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 acoustic/articulatory 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.

Figure 3. Spectrograms of casually spoken version of *top tag* produced by a male speaker. The /p/ release and /t/ closure are not evident in the acoustic pattern. See text.

Figure 4. Spectrogram of casually spoken version of *I can't go* produced by a female speaker. See text for description of acoustic modifications due to articulatory overlap.

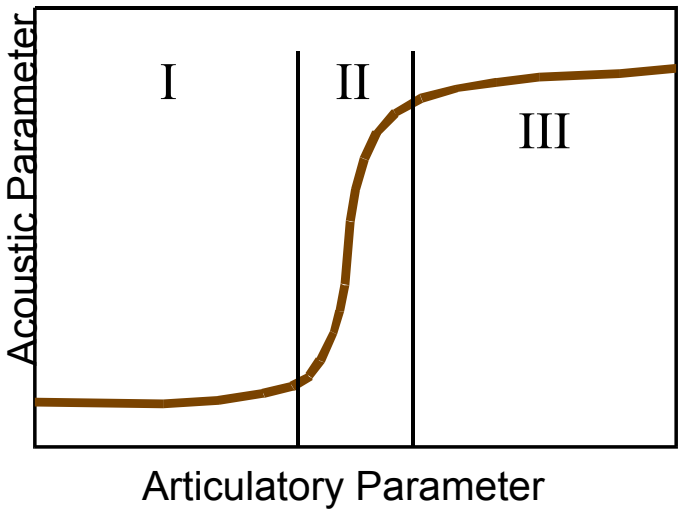

Figure 1

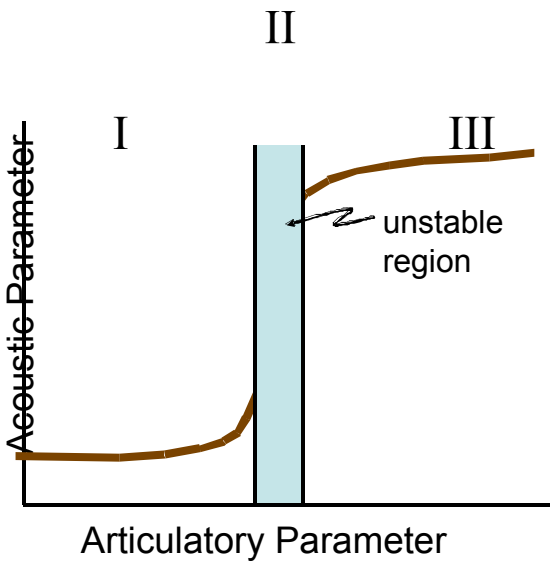

Figure 2a

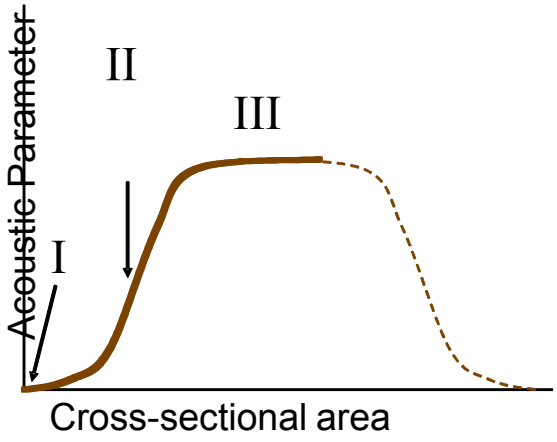

Figure 2b

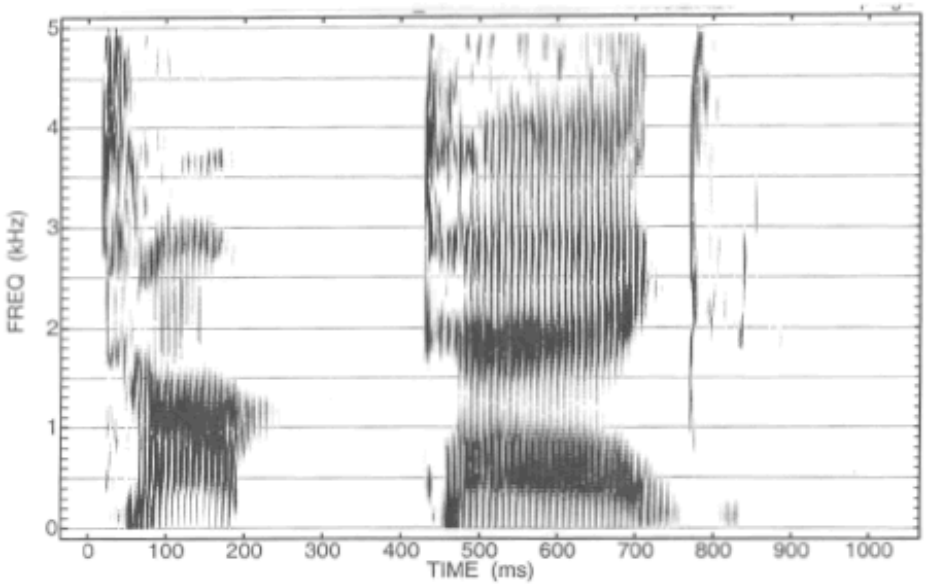

Figure 3

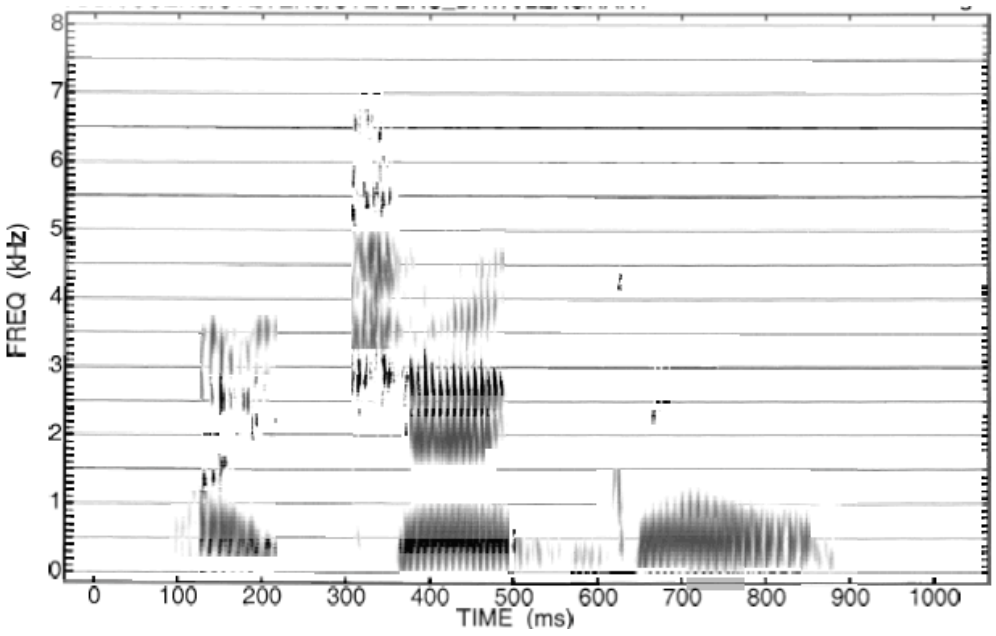

Figure 4