BASES FOR PHONETIC UNIVERSALS IN THE PROPERTIES OF THE SPEECH PRODUCTION AND PERCEPTION SYSTEMS

Kenneth N. Stevens, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U.S.A.

This paper discusses how the properties of the human articulatory and perceptual systems play a role in determining certain phonetic universals. In particular, our concern is with the inventory of phonetic segments that are found in language, and the way in which these segments are organized into a set of natural classes. We shall review how the articulatory and the perceptual systems place certain constraints on the classes of sounds that are used universally in language. The classificatory features that play a role in the phonological rules in language are determined by these natural classes that are based on observation of the capabilities of the articulatory and perceptual mechanisms.

Articulatory evidence for natural classes of speech sounds.

The actualization of a given speech sound in context requires a complex sequence of articulatory activity. The articulatory structures must be maneuvered from positions or states appropriate to one sound to states corresponding to the next sound to be produced. We shall follow the traditional method, used by phoneticians for years, of specifying a phonetic segment in terms of a set of goals or target states that the articulators are to achieve or that are intended by the speaker rather than in terms of the movements between these targets. The hypothesis is that these target configurations or states, if appropriately specified for a given sound, are much less dependent on the phonetic context than are the articulatory movements or muscle contractions necessary to produce the sound in context. Thus the articulatory descriptions are static, in the sense that they describe stationary states or configurations. While the production of some sounds or sound sequences may involve movement, this movement is always from one target state to another.

How are these articulatory target states to be described and how does this description lead to a specification of natural classes of speech sounds? Examination of lateral radiographs gives us one view of the articulatory target states in terms of the positions of the various articulatory structures that are visible on the midline. This kind of evidence has traditionally been used in phonetics to describe articulatory targets in terms of place of articulation identified along the length of the vocal tract. Another way of describing articulatory configurations examines the pattern of contact that occurs between structures such as the tongue and palate. This pattern is presumably registered in the talker's speech control system through the responses of receptors located on the surfaces of the structures (Stevens and Perkell, 1977). Still another aspect of the target state is the physical properties of the surfaces of the structures, particularly the vocal folds and the tongue. These properties have an influence on the manner in which the airflow from the lungs is controlled and on the way in which the articulatory structures are forced against one another. Which of these ways (or combinations of ways) of describing articulatory states is most salient for grouping speech sounds into natural classes is a question about which we can only speculate at present.

We consider now several lists of phonetic segments. For all of the items on a given list, some aspect of the articulation is achieving the same state, as defined in at least one of the ways listed above. We suggest, then, that these items can be candidates for forming a natural class of phonetic segments.

[m n ŋ ɑ̃ ũ ...] These items are all produced by creating velopharyngeal opening, usually by placing the velum in a lowered position. From the point of view of the speaker, an indication that the velum is lowered comes from several possible sources: (1) the muscles used to lower the velum have been contracted; (2) the lowered state of the velum is sensed through receptors that signal the position of the velum or its contact with other structures; (3) there is airflow through the velopharyngeal opening and possibly acoustic energy in the nasal cavity that is sensed and registered in some way.

[k g ŋ i u ...] These sounds are all produced by placing the tongue body in a raised position within the oral cavity. More specifically, the common articulatory activity for the sounds can be described in one of two ways: (1) there is contraction of a common muscle or group of muscles to produce the raised tongue body, or (2) there is a common pattern of activity in particular groups of sensory receptors in the tongue musculature or on the dorsal surfaces of the tongue as these surfaces make contact with other structures, particularly the hard palate (Stevens and Perkell 1977).

[p t k č f θ s š ɑ́ í ú ...] For this group of sounds, it is hypothesized that the common articulatory attribute is a stiffening of the surfaces of the vocal folds (Halle and Stevens, 1971). The articulatory state that characterizes each member of this class can be described either as contraction of a particular laryngeal muscle or group of muscles or as the stiffened state of the vocal fold surfaces, independently of the muscle activity used to produce that state.

[p t k č b d g ǰ m n ŋ ...] The sounds in this group are all produced by forming a complete closure of the vocal tract at some point along its length. The articulatory description for this group of segments cannot be specified in terms of the contraction of particular muscles, since different muscles are clearly involved depending on where in the vocal tract the constriction is made. Rather, it is assumed that an instruction to form a complete closure is a basic component of articulatory control which, when coupled with a further instruction indicating which articulator is to be activated, effects the proper consonantal constriction. It is possible also that the sensory consequences of forming a complete closure are registered in some unique manner independently of the location of the closure in the vocal tract.

[p b f v m ...] The segments on this list have the common articulatory attribute that they are produced with a constriction at the lips. Thus a particular set of muscles - those making a lip closure - is involved in the generation of all of these sounds. The lower lip comes in contact with either the upper lip or the upper incisors, and this gesture leads to a unique pattern of excitation of sensory units in the lower lip.

[t d n θ ð s z š ž ℓ r ...] These phonetic segments are all actualized by raising the tongue blade to make contact with some part of the maxilla. The exact region of contact or the force of contact may vary from one sound to another in the set, but the common gesture is that of raising the tongue blade, presumably through contraction of certain intrinsic tongue muscles. There is a unique sensory consequence of this raised pattern of the tongue blade: the edges of the superior portion of the tongue come in contact with fixed surfaces of the hard palate or teeth, presumably leading to a special response of tactile receptors on these surfaces of the blade.

The six lists of segments given above are examples of a longer inventory of lists of segments that could be generated. Furthermore, there is no attempt to make each list exhaustive; additional items could be appended to the lists. These examples serve to indicate, however, that natural classes of speech sounds can be constructed through examination of the articulatory target configuration or states. In giving these examples, we have shown a certain amount of ambivalence as to how the common articulatory attributes for the items on a list should be specified. Until we know more about how motor systems operate, and, in particular, how the speech-production systems operate, the question of how best to characterize natural classes of speech sounds in terms of articulatory attributes must remain open.

Acoustic and psychoacoustic evidence for natural classes

Acoustic analysis of speech shows that there are groups of speech sounds that share common acoustic properties. If it is assumed that the auditory system responds in some unique way to sounds with a common acoustic property, then this unique response provides the listener with a means for organizing speech sounds into natural classes based on their acoustic properties. As examples, we shall consider several lists of speech sounds, and we shall show that for the items in any one of these lists there is a common distinctive acoustic property. The basis for these classifications is derived largely from the work of Fant (1960), Jakobson, Fant and Halle (1963), and others.

[m n ŋ] For the items on this list, there is a rather steady nasal murmur persisting for several tens of milliseconds, with an amplitude just a few dB below that of the adjacent vowel. The unique acoustic attribute of this nasal murmur is a strong spectral peak at low frequencies and a relatively uniform distribution of weaker spectral peaks at higher frequencies, with these peaks tending to be rather broad (Fujimura, 1962).

[t d n s z š ž č ǰ] For these consonants, the spectrum sampled at or near the consonantal release (in a consonant-vowel syllable) shows a diffuse spread of energy across the frequency range, but with greater spectral energy at high frequencies (Fant, 1960; Zue, 1976; Stevens and Blumstein, 1978).

[k g ŋ] The spectrum at the consonantal release for these sounds has a single prominent peak in the midfrequency range (Fant, 1960; Zue, 1976, Stevens and Blumstein, 1978).

[i ɪ u ʊ] The vowels in this list all have a relatively low first formant.

[ã ũ ĩ] These nasalized vowels have a spectrum in which the lowest peak, corresponding to the first formant region for a nonnasal vowel, is split or broadened to cover a wider frequency range than that for a nonnasal vowel.

[p t k č b d g ǰ m n ŋ] The items in this list all show an abrupt onset of spectral energy over much of the frequency range when the consonant is released into the following vowel. The rise in amplitude in any one frequency region occurs in a time interval of just a few milliseconds. A sound with an abrupt onset has been shown to produce a distinctive response in a listener (Cutting and Rosner, 1974).

[á ú í] These vowels all have a fundamental frequency (F_0) that is high in comparison with the average F_0 for the particular speaker and the particular position of the vowel within an utterance.

[p t k č f θ s š] The common acoustic characteristic of the sounds in this list is the absence of low-frequency periodicity in the sound in the vicinity of the consonantal closure interval.

As in the case of the lists based on articulatory attributes, the above lists are examples of a longer inventory of lists such that the items in each list have a common acoustic property to which the auditory system is assumed to respond in a unique way. Given our present rudimentary knowledge of the response of the auditory system to complex sounds, we have only been able to speculate on the kinds of acoustic properties that qualify for defining groups of speech sounds.

The classificatory features

Examination of the two sets of lists - these based on common articulatory attributes and those based on common acoustic attributes - reveals that there is much overlap in the two sets. This overlap is not surprising, since on the basis of acoustical theory it is not unexpected that sounds produced with common aspects of the articulatory configuration should also have similar acoustic characteristics.

Another way to organize speech sounds into natural classes is to examine the phonological rules of language, and to observe the various groups of segments that are operated on by these rules or that determine the environments in which the rules operate. The grouping of segments according to this criterion leads to a description of segments in terms of bundles of classificatory or distinctive features. These classificatory features also show a great deal of overlap with the groupings based on articulatory and acoustical considerations.

We would like to propose a rather simple condition on the definition of a classificatory feature: a set of speech sounds shares the same classificatory feature if the sounds share a common articulatory attribute and a common acoustic or perceptual attribute. That is, the sounds in a given class should give rise to response patterns that have a common property in the auditory system of the listener and the speaker, and, in addition, the production of the sounds should have common attributes in the speech-generating mechanism of the speaker, such as common patterns of orosensory response.

A consequence of this definition is that vowels and consonants will tend not to share the same features. Thus, for example, nasal vowels and nasal consonants would not have the same feature, although it might be desirable to mark in some manner the fact that they share an articulatory property. The strong definition of a classificatory feature would not capture in terms of feature specifications the fact, for example, that vowels preceding nasal consonants tend to be nasalized (or in fact that nasalization of the vowel often is accompanied by

elimination of the consonant), or the fact that the pitch of vowels following voiceless consonants tends to be raised. These kinds of modifications are, in a sense, simply mechanical consequences relating to the coarticulation that is a nature consequence of the juxtaposition of two segments.

The classificatory features defined in the way we have proposed would, however, specify major classes of segments that play a role in the phonological rules of language. These features would owe their existence, so to speak, both to the property-generating characteristics of the speech production system and to the property-detecting characteristics of the speech perception system.

References

Cutting, J. and B. Rosner (1974): "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.

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

Fujimura, O. (1962): "Analysis of nasal consonants", JASA 34, 1865-1875.

Halle, M. and K.N. Stevens (1971): "A note on laryngeal features", Research Laboratory of Electronics Quarterly Progress Report No. 101, M.I.T., Cambridge, Massachusetts, 198-213.

Jakobson, R., G. Fant, and M. Halle (1963): Preliminaries to speech analysis, Cambridge, Massachusetts: M.I.T. Press.

Stevens K.N. and S.E. Blumstein (1978): "Invariant cues for place of articulation in stop consonants", JASA 64, 1358-1368.

Stevens, K.N. and J.S. Perkell (1977): "Speech physiology and phonetic features", in Dynamic aspects of speech production, M. Sawashima and F.S. Cooper (eds.), 323-341, Tokyo: University of Tokyo Press.

Zue, V.W. (1976): Acoustic characteristics of stop consonants: A controlled study, Ph.D. thesis (unpublished), M.I.T., Cambridge, Massachusetts.