Knowledge of Language and the Sounds of Speech

Morris Halle and Kenneth N. Stevens 1991

1. The Representation of Utterances in Memory: Phonological Evidence

1.1. Speech is Composed of Words

Speakers have the clear intuition that when speaking they say words and when spoken to they hear words. It comes as a considerable surprise to naive speakers to discover that the utterances that they produce and hear are in fact not divided by short pauses into words. And this impression of hearing and speaking words is not lost even by experienced speech researchers who are well aware of the fact that utterances are not acoustically segmented into words.

The proposition that we hear words is nicely supported by the results of the following *Gedankenexperiment*. Speakers of English can readily divide the utterance

The Lord is my shepherd I shall not want

into its nine component words. But when presented with the utterance

[jaw'ɛrɔʕ'ilɔʔɛħs'ar],

which is the original Hebrew phrase of which the English is a translation, English speakers are no longer able to divide the utterance into its component words.

There is of course no mystery about this result. The subjects of our experiment do not know Hebrew, and having no knowledge of Hebrew words, they are unable to segment a Hebrew utterance into its words. The inference to be drawn from this experiment—as well as from countless other facts—is that the knowledge that speakers have of their language plays a central role in all phenomena that are the subject matter of phonetics. On this view, which we have graphically represented in

Reprinted with permission from J. Sundberg, L. Nord, and R. Carlson, eds., *Music, Language, Speech, and Brain*, MacMillan Press, London, pp. 1–19, 1991.

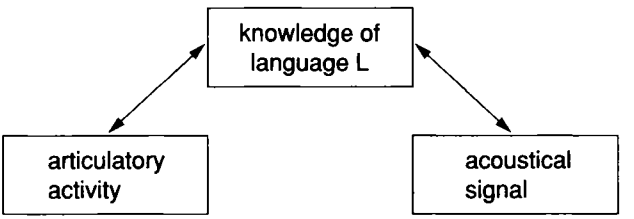


Fig. 1. Knowledge of language and its physical manifestations.

Fig. 1, the subject matter of phonetics cannot be limited to the acoustical speech signal and the articulatory behavior that produces the signal, but must always take explicit account of the role that is played by speakers' knowledge of their language.

In what follows we provide some information about the nature of this knowledge and discuss some fundamental problems of phonetics in light of this knowledge.

1.2. Words are Sequences of Sounds

We are not born knowing the words of our mother tongue; this knowledge is something that we acquire during the first few years of our life. We therefore begin by asking in what form this knowledge is laid down in speakers' memories.

When speakers learn a word, they learn its sound or phonetic shape, its meaning and also its grammatical features.[1] Since we are interested here in phonetic issues, we shall disregard the latter information, and limit ourselves to the form in which a word's phonetic shape is stored in memory.

We assume that the traditional view that words are stored in memory as sequences of discrete sounds is correct in its essentials. Part of the evidence in support of this view derives from the fact that in most, if not all languages there are systematic relationships between words of the kind illustrated in (1), i.e., we find sets of words related by affixation.

(1)	a.	power—em-power	courage—en-courage	force—en-force
	b.	list—en-list	rage—en-rage	mesh—en-mesh
	c.	possible—im-possible	conspicuous—in-conspicuous	fallible—in-fallible
	d.	legal—il-legal	regular—irregular	moral—im-moral

In (1) we have illustrated the behavior of two word-forming prefixes of English. The two prefixes are pronounced alike in many dialects, yet there are important differences between them. The prefix spelled *en* forms verbs from nouns and adjectives; it therefore changes not only the meaning of the word, but also its lexical class. By contrast, the prefix spelled *in* changes the meaning of the adjective from positive to negative, but it keeps its lexical class intact. This morphological difference between the two prefixes is paralleled also by differences in their phonetic behavior. Before stems beginning with an obstruent both prefixes behave alike, as illustrated in the examples in (1a,c).[2] As shown by the examples (1b,d), however, before sonorant cousonants the prefixes behave differently. In this context, *en* remains unchanged—e.g., *enmesh*, *ennoble*, *enrage*, *enlist*—whereas in *in* the [n] either undergoes total assimilation or is deleted (cf. Zwicky 1970); e.g., *immoral*, *innumerable*, *illegal*, *irregular*.

The behavior of the prefixes *in* and *en* shows thus that affixation does not always leave everything else intact; it frequently is accompanied by a phonetic modification of the component elements. What is especially important from our point of view is that the overwhelming majority of such modifications affect individual speech sounds: in (1a,c) it is the initial consonant of the stem whose place of articulation is assimilated by the last consonant of the prefix, whereas in (1d) it is the last sound of the prefix *in* that is deleted or undergoes assimilation when followed by a sonorant consonant. This way of characterizing the facts presupposes crucially that affixes and stems are sequences of discrete sounds, and since affixes and stems make up words, this evidence also supports the proposition that words are sequences of discrete sounds.

As is well known when we investigate the acoustic speech signal and the actions of the articulators that give rise to the signal, we often find it impossible to determine where one sound ends and the next begins. This fact has raised questions about the correctness of the proposition that words are composed of discrete sounds. The absence of clearly marked sound boundaries is analogous to the absence of pauses between words. As noted in the first paragraph of this paper, although words in ordinary utterances are not separated by short pauses, there is little question that speakers intend to produce word sequences and they are so understood by their interlocutors. Similarly, there are innumerable facts such as those in (1) that can only be accounted for by presupposing the existence of discrete sounds. These facts warrant the reality of discrete sounds as units of language regardless of the extent to which this discreteness is masked in the physical speech event.

1.3. Segments are Feature Complexes

If words are indeed composed of discrete sounds, our next task is to say something about the nature of the sounds that compose the words. It is obvious that we cannot define the sounds of language simply as acoustic signals produced by the larynx, the lips, tongue, etc., for this would not allow us to distinguish the vowels and consonants of any natural language from sighs, moans, groans, burps, coughs, etc. How these two types of acoustic output of the human vocal tract are to be distinguished one from another was discussed by E. Sapir in a famous paper *Sound Patterns in Language* (1925). Sapir remarked that from an articulatory and acoustic point of view the sound made when blowing out a candle is indistinguishable from the sound that, in many English dialects, is found at the beginning of words such as *when*, *whale*, *white*. Sapir asks: "Does this identity amount to psychological identity of the two processes?" and answers "Obviously not." He suggests that the salient difference between speech and nonspeech sounds is that every speech sound has a specific place in a system of sounds, whereas there is no such systematic relation between the various nonspeech sounds. In Sapir's words: "A sound that is not unconsciously felt as 'placed' with reference to other sounds is no more a true element of speech than a lifting of the foot is a dance step unless it can be 'placed' with reference to other movements that help to define the dance. Needless to say, the candle-blowing sound forms no part of any such system of sounds." Sapir's point thus is that a given speech sound is a token of a type that stands in a cognitive relation to other sound types by virtue of speakers' knowledge of language, but this is of course not true of non-speech sounds.

A major contribution to our understanding of the nature of language to be credited to the Russian linguists R. Jakobson and N. Trubetzkoy (see Jakobson, 1928) was the discovery that speech sounds are not the ultimate, further unanalyzable building blocks of language, but that speech sounds are complexes of features such as nasality, rounding, continuancy, etc. We quoted just above Sapir's remark that for a sound to be an element of language it must be "placed" with reference to other sounds. The proposition that speech sounds are complexes made up of distinctive features makes explicit the manner in which this "placement" is to be understood. Each sound belongs to several different subsets of sounds, where the subset is composed of sounds sharing one or more features. No such "placement" is recognizable for nonspeech sounds such as sighs, groans, moans and burps. Since these are indeed further

unanalyzable entities rather than complexes of features, they stand in no cognitive relation to each other.[3]

The features not only reflect phonetic attributes of the different sounds, but as we tried to illustrate in our discussion of the examples in (1), they also play a fundamental role in many of the rules that speakers of a language must know. It is a striking fact about these rules that they involve only certain sets of sounds, but not others. In particular, the sets of sounds encountered in rules of very different languages have very simple characterizations in feature terms; the sets typically share one or two features, which distinguish them from all other sounds of the language. Sets that require more complex characterizations are never encountered. We have illustrated this with the paired examples in (2). In the first member of each pair we have cited the sets of sounds encountered in the rules underlying the treatment of the prefixes *in* and *en* in (1). In the second member of the pair we have cited sets that contain the same number of sounds as the former, but which are never encountered in any phonological rule and which lack the simple characterization available for the first set. In (2a) we have characterized the two labial stops that trigger assimilation of the point of articulation in careful speech. In (2b) we have given the class of obstruents, which trigger assimilation in fast speech. In (2c) we have given the set of sonorant consonants before which the /n/ of the prefix *in* is deleted.

(2) a. i. [p b] = Labial, [−continuant]
ii. [p e]

b. i. [p b f v t d s z θ ð č ǰ š ž k g] = [+consonantal, −sonorant]
ii. [a b c d e f g h i j k l m n o p]

c. i. [r l m n] = [+consonantal, +sonorant]
ii. [r k m o]

The difference between the paired sets illustrates the fact that rules are feature-sensitive and that rules admit only groups of sounds that are readily characterized in terms of features.

1.4. Feature Hierarchies

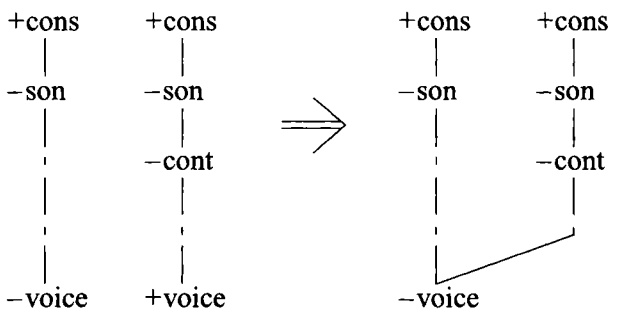
Among the phonological processes that accompany affixation, one of the most common is *assimilation*, where one speech sound absorbs part—

or all—of the features of its neighbor. A somewhat complex example of assimilation is the behavior of the prefixes *en*, *in* discussed above. A simpler example is the treatment of the English regular past tense suffix illustrated in (3), where we get /t/ after voiceless consonants and /d/ elsewhere.

(3) [t]: sipp-ed, cough-ed, plac-ed, blush-ed, work-ed
[d]: grabb-ed, love-d, prize-d, garage-d, hugg-ed
cramm-ed, crane-d, fill-ed, spear-ed,
play-ed, crie-d, tango-ed, conga-ed

This type of assimilation is commonly characterized by postulating that the regular past marker is /d/, which assimilates voicelessness when affixed to a verb ending in a voiceless consonant. We might picture this formally as a process of spreading the voicelessness of the stem-final consonant to the past tense suffix as shown in (4).

(4)

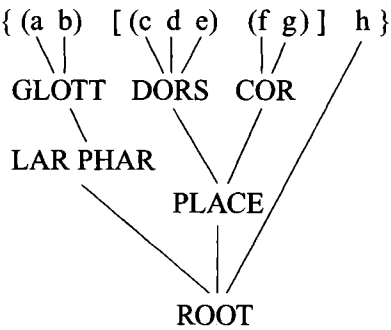


Processes of assimilation that involve more than one feature are subject to severe restrictions. In fact, only a small number of such multiple assimilations are attested, and the overwhelming majority of logically possible assimilations have never been encountered. For example, we never find instances where nasality and lip rounding are assimilated together, or voicing and tongue height. To account for this fact, it was proposed by Clements (1985) that the features of a speech sound are not just a simple list without structure, of the kind illustrated in (4), but are rather organized into a hierarchical tree structure of the kind illustrated in (5).

As indicated at the right side of (5) the terminal nodes labelled with lower case letters are features. The nonterminal nodes labelled with

capital letters represent different feature *groupings*. It is readily seen that if assimilation is expressed formally by spreading a node in a tree such as (5) to a similar tree representing the features of an adjacent sound, then only certain feature sets can spread simultaneously, namely those exhaustively dominated by a node in the tree. Thus, given (5) it is possible to assimilate the feature pair [a,b] or the quintuplet [c,d,e,f,g], but it is impossible to assimilate the pair [e,f] or the quadruplet [a,b,c,d].

(5)



where a = [stiff vocal folds]
b = [slack vocal folds]
c = [high]
d = [low]
e = [back]
f = [anterior]
g = [distributed]
h = [continuant]

The purely formal proposal that features are grouped into sets of constituents and these into higher-order sets as in (5) accounts for the observed restrictions on assimilation and on other phonological processes. It has, moreover, a side-result of more than routine interest. As shown by Sagey (1986), many of the groupings of the features in (5) have straightforward phonetic interpretations. Thus, the features [c,d,e], i.e., [high], [low], and [back], are grouped together, and so are [f,g], i.e., [anterior] and [distributed]. In each case we have grouped together features executed by a given articulator: [low, high, back], by the dorsal articulator or tongue body; [anterior, distributed], by the coronal articulator or tongue blade.

1.5. Articulator-Free and Articulator-Bound Features

That features executed by a given articulator belong together is surely a truism for phoneticians. Nonetheless, the central role of the articulator is not taken into account in the design of the alphabet of the International Phonetic Association, and this is also true of O. Jespersen's (1889) analphabetic notation, K. Pike's (1943) phonetic framework as well as the feature system of Jakobson, Fant and Halle (1952) and its later

modifications. As a matter of fact, phonetic texts do not standardly list the articulators that execute the gymnastics which produces the speech signal. We list them in (6) with the caveat that at this time the list is not yet definitely settled. For some motivation of the organization in (5), see Halle (1992).

(6) lips = Labial
tongue blade = Coronal
tongue body = Dorsal

soft palate

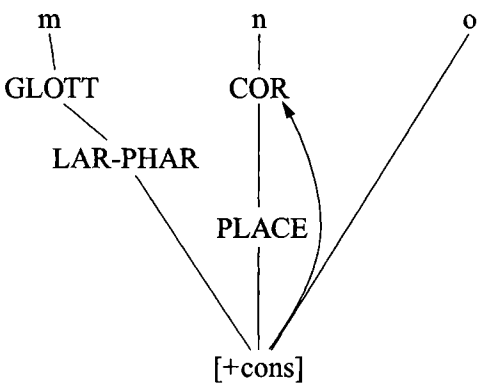
tongue root
glottis

We note that in the feature tree (5) the feature [continuant] (=h) at the right differs from the rest in that no articulator node dominates it. This difference reflects the important phonetic fact that whereas each of the other features in (5) is executed by a specific articulator exclusively, the feature [continuant] has no dedicated articulator, but is implemented either by the lips in [pbm], or by the tongue blade in [tdn], or by the tongue body in [kgŋ]. The same is true also of the features [consonantal], [sonorant], [lateral] and [strident]; these features too lack a dedicated articulator. Thus, there are Labial, Coronal and Dorsal stops, and there are Glottal, Dorsal, and Labial glides. We will designate this class of features as *articulator-free* and contrast them with the rest which are *articulator-bound.* It is obvious that when a sound includes an articulator-free feature it also must include a stipulation of the articulator that is to execute this feature.

Following McCarthy (1988) we place the feature [consonantal] at the root of the feature tree, and extending somewhat a suggestion of Sagey (1986) we call the stipulated articulator the *major* articulator of the sound. We indicate formally the major articulator of a sound by a pointer connecting the Root node of the feature tree with this articulator, as shown in (7).

It has been proposed in Halle (1992) that for every speech sound the feature [consonantal] *must* be specified. As a rough approximation this means that every sound must be either vowel or consonant. One or more additional articulator-free features must be specified in the case of a [+consonantal] sound. For [−consonantal] sounds no additional articulator-free features are available.

(7)



where m = [+stiff vocal folds]
n = [+anterior]
o = [+continuant]

Since [consonantal] is an articulator-free feature, it follows from Halle's proposal that every speech sound must have its own major articulator, i.e., the one that executes the feature [consonantal]. It is an interesting further fact that if for a given sound additional articulator-free features must be specified, the major articulator of the sound executes these features as well. Thus, in the Coronal continuant [s] represented in (7) both the features [+consonantal] and [−continuant] are executed by the tongue blade. A sound where [+consonantal] is executed by the tongue blade, whereas [−continuant] is executed by the dorsum has never been observed and is, in fact, anatomically impossible. The pointer mechanism is our way of implementing this constraint formally.

2. The Phonetic Module: On the Links between Articulation and Acoustics

Up to this point we have assumed without discussion that feature representations can be translated into articulator movements with specific acoustic effects and that the acoustic patterns can be decoded into arrays of features, but we did not go into this matter. The phonetic implementation of the features must now be discussed.

We follow Liberman and Mattingly (1985, 1989) and other workers in assuming that part of the genetic endowment of humans that makes them capable of acquiring command of one or more languages is a special *phonetic* module that selects specific actions of the articulators and links them to selected aspects of their acoustic consequences. For example, the forward and backward placement of the tongue body is cor-

related with specific differences in the frequency of the second formant, or the different placements of the tongue blade—before or behind the alveolar ridge—are correlated with the differences in the acoustic spectrum between "hissing" and "hushing" sounds. And similar correlations between articulatory activity and acoustic signal are genetically provided for each of the nineteen or so features that make up the universal set of phonetic features (cf. Halle, 1992).

We share with Liberman and Mattingly and other students of speech the supposition that this link between articulatory and acoustic aspects of speech is *not* "a result of the fact that what people hear is what they do when they speak. Rather the link is innately specified, requiring only epigenetic development to bring it into play." (Liberman and Mattingly, 1985, p. 3.) For example, humans have available as part of their genetic endowment the information that a sound with a second formant (F2) that is high and close to F3 is produced by moving the tongue body forward. Similarly, information that a sound with a compact spectral peak in the mid-frequency range is produced by raising the tongue body against the roof of the mouth is part of a child's genetic endowment.

Some experimental evidence in support of this proposition comes from the studies of Kuhl and Meltzoff (1982, 1984), who showed that infants as young four months "appear to know that /a/-sounds go with faces displaying wide open mouths, /i/-sounds [go] with faces displaying retracted lips ... [and] u-sounds go with pursed lips." (Kuhl, 1988, p. 39). These authors (Kuhl and Meltzoff, 1988) also observe that "infants who heard the vowel /a/ produced vowel-like sounds whose formant frequencies were closer to adults' /a/'s than to adults' /i/'s," and similarly for /i/. (For other experimental evidence, see McGurk and MacDonald, 1976.)

Since the link between acoustic signal and articulatory activity is genetically established, infants exposed to language need not discover the existence of this link. What they need to discover for themselves is the particular features that play a functional role in the language of their community as well as details about this role.[4]

2.1. The Categorial Nature of Articulatory-Acoustic Links: Phonetic Bases for Features and Segments

It has been shown by Stevens (1972, 1989) that the relation between articulatory displacement and the perceptually relevant acoustic effect

tends to be quantized. In the case of a great many features there appear to be in the acoustic-articulatory relation two extreme regions where moderate changes in positioning an articulator have essentially negligible acoustic consequences, while in an intermediate region located between these two extremes small articulatory movements have significant acoustic effects. Thus, for example, as the position of the consonantal closure formed by the tongue blade is retracted from the dental to the alveopalatal region in the production of fricative consonants, the spectrum of the resulting sound undergoes abrupt changes as the length of the cavity anterior to the constriction passes through the unstable region corresponding to the boundary that separates [+anterior] from [−anterior] sounds.

One important consequence of these quantal relations between articulation and sound is that the positioning of the articulators does not need to be precise in order to achieve the desired acoustic result. Considerable variation can occur without altering in a significant way the acoustic properties relevant to the feature being implemented. This fact has an obvious advantage in speech production: speakers will implement the feature correctly as long as they locate the articulator anywhere in the quasi-stable region at the proper side of the boundary with an adjacent region.

Since each speaker's anatomy is somewhat different (shape of palatal vault, thickness of vocal folds, shape of alveolar ridge, dimensions of nasal cavity, etc.), each individual must discover the stable regions for each of the features utilized in his/her language. Like other kinds of skilled muscular activity this learning process requires some time. Moreover, it is subject to readjustment when the shape of the articulator and hence the location of the stable regions is changed (for example, by the insertion of special dental protheses).[5]

2.2. Discontinuities in the Acoustic Signal

A second consequence of the existence of these quantal relations between articulation and sound is that the parameters of the sounds often exhibit discontinuities as the articulatory structures traverse certain critical regions. An acoustic discontinuity occurs for example as the movement of an articulator creates a sufficiently narrow constriction in the vocal tract or causes the release of such a constriction. Thus, as the articulators form openings and constrictions during speech production,

the resulting discontinuities mark a succession of discrete events or landmarks that are readily discerned in the acoustic record.[6] These discontinuities in the signal are correlated to a degree—though not totally—with the discrete sounds by means of which words are represented in memory.

2.3. Features with Multiple Acoustic Correlates

The discontinuities in the acoustic signal can be of several types, and are usually a consequence of implementation of the articulator-free features. On one side of a discontinuity the vocal tract is relatively constricted, whereas on the other side the constriction is less severe as the major articulator moves towards or away from the constriction. The acoustic manifestations of the articulator-bound features are different on the two sides of this discontinuity.

For example, the articulatory action of stiffening the vocal folds and the musculature of the lower pharynx has quite different acoustic consequences depending on whether or not significant pressure is built up in the supralaryngeal cavity. When there is such pressure, stiffening results in suppression of all vocal-fold vibration. In order for the vocal folds to vibrate under this condition, both the folds and the lower pharynx must be slack. In the absence of intra-oral pressure, on the other hand, stiffening increases the frequency of vibration of the folds, and slackening decreases it.

This significant pressure buildup in the supralaryngeal cavity happens especially during the production of obstruents, when the air flow from the lungs to the ambient air is greatly impeded or stopped altogether. By contrast, in the production of a sonorant there is an open passage from the lungs to the ambient air and no intraoral pressure is built up. As a result, we have two distinct consequences of vocal-fold stiffening: in sonorants vocal-fold stiffness-slackness is correlated with lower vs. higher frequency of vocal-fold vibration, whereas in obstruents vocal-fold slackness-stiffness is correlated with the presence-absence of vocal-fold vibrations. This fact accounts for the well-documented phenomenon that in a sequence of an obstruent consonant followed by a vowel, the frequency of vocal-fold vibration in the initial portion of the vowel is higher when the obstruent is voiceless than when it is voiced (cf. House and Fairbanks, 1953). Here vocal-fold stiffness in the consonant is signalled as a voicing difference on one side of the consonantal divide, and as a pitch difference on the other.

This is one of a number of examples in which radically disparate acoustic properties on two sides of a discontinuity provide information about a given feature.[7] The fact that listeners effortlessly integrate sequences of such disparate acoustic cues into a unitary phonological feature is a prime example of the operation of the special phonetic module dedicated to the processing of speech signals that has "its own modes of signal analysis and its own primitives" (Liberman and Mattingly, 1989, p. 489). Both in its primitives and its analytic procedures this linguistic module differs from the perceptual module triggered by non-speech sounds.

2.4. The Feature [Stiff Vocal Folds]

The question that arises at this point is what reason there is for pitch-register distinctions to be viewed as a special manifestation of the same feature as consonant voicing. The uniform articulatory action suggests that we are dealing with a single feature (*Halle and Stevens, 1971*), whereas the disparate acoustic consequences of this uniform articulatory behavior suggest that voicing and register distinction are best treated as distinct. The articulatory and acoustic facts do not exhaust the relevant data at our disposal. We must also bring to bear on our question linguistic data, i.e., facts deriving from the knowledge possessed by speakers of different languages.

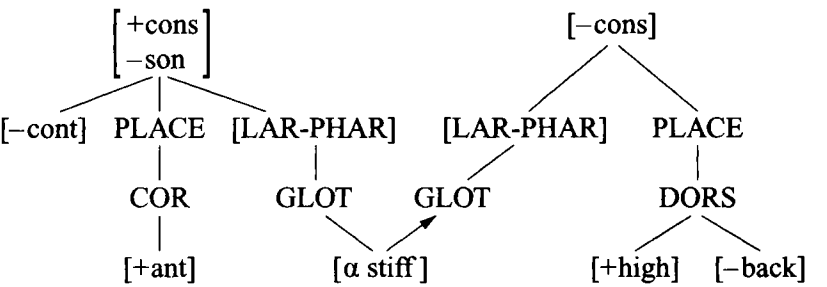
Before reviewing these data, we note that recent work by Bao (1990) and Duanmu (1990) has provided strong evidence for the proposal originally made by Yip (1980) that not all differences in pitch are implemented by this type of stiffening of the vocal folds. The pitch differences implemented by these differences in vocal-fold stiffness are identical with upper vs. lower register differences referred to by Yip (1980). It is pitch register rather than the other pitch distinctions that is correlated with voicing in obstruents and that is under discussion here.

Students of the languages of East Asia have long drawn attention to the close correlation in these languages between voicing in consonants and register differences in vowels.[8] A typical example cited by Bao (1990, p. 64) is the tonal inventory of the Songjiang dialect of Chinese illustrated in (8).

(8)	ti	53	"low"	di	31	"lift"
	ti	44	"bottom"	di	22	"brother"
	ti	36	"emperor"	di	13	"field"

In (8) the numerals to the right of each word indicate the pitch contour of the vowel in the Chao (1930) notation, which distinguishes five tone heights with 1 denoting the lowest tone and 5, the highest. The words on the left in (8) have tones in the upper range or register—i.e., between 3 and 5—whereas the words on the right have tones in the lower register—i.e., between 1 and 3. Moreover, the words on the left have as onset the voiceless obstruent [t], whereas those on the right have as onset the voiced obstruent [d]. This distribution can readily be accounted for by assuming that the feature [stiff] spreads from the consonant to the following vowel as illustrated in (9).

(9)



This account is impossible in principle if the voicing feature in the consonant is treated as distinct and separate from the register feature of the vowel. Since the phenomenon illustrated in (9) is common both among the languages of China as well as elsewhere (Verner's Law of Indo-European is a special case), it must be taken as linguistic evidence in favor of treating vocal-fold stiffness as a single feature in spite of its apparent acoustical diversity.

2.5. Implications for Speech Perception and Production

We have argued that words are represented in memory as sequences of discrete segments each of which is characterized by a complex of distinctive features. The phonetic substratum for each feature establishes a link between a specific articulatory action and an acoustic and perceptual consequence of this action.

During speech perception, the phonetic module interprets the acoustic signal as an array of features and utilizes this information to identify words that are stored in memory as arrays of features. We propose that the events happening at discontinuities in the signal play an especially salient role in the process of speech perception. Discontinuities of

different kinds provide information about different articulator-free features. These events identify regions in the signal where other acoustic properties are detected.[9] Identification of the articulator-bound features is based on these properties. In particular, as sketched above, the changes in the spectral patterns of certain features that occur at these discontinuities are utilized by the phonetic module for direct identification of these features.

In the *production* of utterances the same process runs in the reverse, as it were. Speakers create the landmarks or events in the sequence of sounds that are specified in the Vocabulary representation of the words which compose the utterance. Speakers must further insure that at the times these landmarks are created the articulators assume the states or positions specified by the other features in the representation.[10] This process involves a complex coordination among movements of different articulators, taking into account their different rates of response so as to insure that the correct acoustic attributes surface in the vicinity of the different landmarks. In order for this to happen the movements of the various articulators contributing to the output must be initiated at times prior to the occurrences of the different landmarks. For example, when a Vowel-Nasal sequence is to be produced the lowering of the velum is started early in the vowel so that at the instant of oral closure the cross-sectional area of the velopharyngeal opening is in the range that results in a noticeable acoustic discontinuity.[11]

The timing of actions of different articulators is coordinated so that the appropriate set of acoustic properties appear together in time. This complex coordination is one of the most striking things that one sees on x-ray motion pictures of speech. Young children acquire this coordination much too rapidly, with too little trial and error to allow one to entertain realistically the hypothesis that learning is involved in speaking in the same sense in which learning is involved in such other activities of young children as using spoons, forks and other eating utensils, or tying their shoes, or catching and throwing balls. Unlike these activities, but like bipedal gait, which also involves complex coordination of actions of several anatomical structures, the ability to produce speech must therefore be assumed to be largely innate; it is a genetically prewired function of the speech module, and therefore need not be painstakingly learned by speakers when they acquire their mother tongue.

3. Concluding Remarks

We have tried to illustrate here the role that knowledge of language plays in phonetic phenomena of all kinds. We argued that such a fundamental concept of phonetics as the speech sound is best viewed as a unit in terms of which words are encoded in speakers' memories. We presented evidence showing that in speakers' memories the features of a given sound are not just random collections, but are organized into a specific hierarchy of the kind illustrated in (5). This hierarchy distinguishes between features that are *articulator-bound* in that they are executed by a single dedicated articulator, and features that are *articulator-free* and hence not so restricted; and we found that features executed by a given articulator are grouped together in the feature hierarchy. Of particular importance to our argument was the observation that this essentially anatomical grouping of the features is also required for a proper description of how the features function in the different phonological rules. Thus, considerations of fundamentally different kinds—i.e., anatomical, on the one hand, and grammatical, on the other—converge on a single result.

In the second part of the paper we focused on the phonetic correlates of the features and noted two important characteristics: their quantal character and the coupling of a single articulator action with acoustically and perceptually disparate effects in differing phonetic contexts. In the examples we have studied, a particularly prominent contextual role is played by the feature [consonantal], but other articulator-free features may also exhibit similar behavior.

We noted that some aspects of the process of speech production and perception must be learned; e.g., the location of quasi-stable regions for each of the features. The majority of the properties that we have examined, however, cannot plausibly be attributed to learning. The idea that utterances are composed of words, that words are stored in memory as sequences of discrete sounds, that sounds are made up of hierarchically organized features—none of these can plausibly be attributed to learning. One can readily see this when one tries to describe a scenario whereby three-year-olds could acquire these insights into the nature of language as a by-product to their ordinary interaction with a home environment that may vary from an igloo to a Bedouin tent, and from a tree house in New Guinea to an apartment in Stockholm. Since learning must thus be excluded, the only remaining alternative is to assume that much of our knowledge is part of our genetic endowment—part of the

language module that all humans possess at birth and that makes us human.

Acknowledgements

We are grateful to Sylvain Bromberger and Sharon Manuel for help and advice as well as for some of the ideas in this paper. Preparation of the paper was supported by grants CD00075 from the National Institutes of Health and IR1-8910561 from the National Science Foundation.

Notes

1. Thus, when some years ago the word *glasnost* was introduced into English and other languages, we acquired a new phonetic item and also learned that it represents an abstract noun, and that it has roughly the same meaning as the English *candor, openness.*
2. The prefix-final nasal assimilates the place of articulation of a following labial stop, i.e., n → m before p, b as in *embed, empower, impossible,* but /n/ is preserved before other obstruents at least in careful speech, i.e., [n] in *incorrect, encourage.* (In less careful speech, the assimilation occurs before obstruents of all kinds.)
3. The proposition that speech sounds are complexes of features must be understood in the same light as the proposition that words are composed of discrete sounds. Both propositions refer to the form in which these entities are represented in our memories. The implementation of these entities in actual utterance tokens will, of course, differ in various ways from their ideal form, much as the scrawled notes that we make at a lecture differ from the sequence of discrete letters that they are intended to represent.
4. The phonetic module is part of the human language competence and is activated in the ordinary hearing speaker whenever utterances are being processed. The phonetic module is not activated by non-linguistic articulatory activity or by non-linguistic acoustic stimuli. This is a further difference between blowing out a candle and the acoustically and articulatorily indistinguishable phenomenon of pronouncing the English labialised glide [ʍ] discussed by Sapir.
5. According to Fowler *et al.* (1980, p. 406) "Amerman and Daniloff (1971) relying on listener judgments, found normalization of vowel production within 5 minutes of the insertion of a prothesis in a subject's mouth. In a similar procedure Hamlet and Stone (1976) fitted subjects with three different types of protheses. The effects on the production of vowels were striking. Compensation for vowel changes was variable among the subject pool and was not always accomplished, even after a week of adaptation. In addition, a period of readjustment was required by subjects subsequent to the removal of these protheses." These findings are in contradiction with the well-known bite-block experiments, where compensation appears to be immediate. Thought should be given as to how these contradictory observations are to be reconciled.

6. Attention was drawn to the existence of these acoustic discontinuities and their significance many years ago by Gunnar Fant (1961).
7. For additional examples exhibiting this kind of diversity, see Stevens (1985).
8. See, for example, Haudricourt (1954) and Matisoff (1973).
9. A model of speech recognition based on this concept has been proposed by Stevens (1986, 1988).
10. A similar view has been advanced by M. Huffman (manuscript).
11. Or, in a Vowel-Consonant-Vowel utterance such as *a pie* in which the Consonant is a voiceless aspirated stop, the beginning of the glottis-spreading maneuver will usually occur prior to the end of the first Vowel, so as to create a more abrupt offset of voicing as well as to insure that there is sufficient glottal spreading at the time of the Consonant release.

References

Amerman, J. D. and R. G. Daniloff
1971 Articulation patterns resulting from modification of oral cavity size. *ASHA*, **13**, 559.

Bao, Z.
1990 On the nature of tone. Ph.D. dissertation, Massachusetts Institute of Technology.

Chao, Y.-R.
1930 A system of tone letters. *Le Maitre Phonétique*, **34**, 24–47.

Clements, G. N.
1985 The geometry of phonological features. *Phonology Yearbook*, **2**, 223–250.

Duanmu, S.
1990 A formal study of syllable, tone, stress and domain in Chinese languages. Ph.D. dissertation, Massachusetts Institute of Technology.

Fant, G.
1961 The acoustics of speech. In L. Cremer (ed.), *Proceedings of the Third International Congress on Acoustics, Stuttgart, 1959*. Amsterdam: Elsevier Publishing Company. Reprinted in G. Fant, *Speech Sounds and Features*, pp. 3–16. Cambridge MA: MIT Press.

Fowler, C. A., P. Rubin, R. E. Remez, and M. T. Turvey
1980 Implications for speech production of a general theory of action. In B. Butterworth (ed.), *Language Production, Vol. 1; Speech and Talk*. London: Academic Press, pp. 373–420.

Halle, M.
1992 Phonological Features. In W. Bright (ed.), *Oxford International Encyclopedia of Linguistics*. New York: Oxford University Press, vol. 3, pp. 207–212.

Halle, M. and K. N. Stevens
1971 A note on laryngeal features. *Report No. 101*, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge MA, pp. 198–213. (Reprinted in this volume.)

Hamlet, S. L. and M. Stone
1976 Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *J. Phonetics*, **4**, 199–218.
Haudricourt, A.-G.
1954 De l'origine des tons en vietnamien. *Journal Asiatique*, **242**, 68–82.
House, A. S. and G. Fairbanks
1953 The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.*, **25**, 105–113.
Huffman, M.
manuscript Articulatory landmarks: Constraining timing in phonetic implementation.
Jakobson, R.
1928 Quelles sont les méthodes les mieux appropriées à un exposé complet et pratique de la grammaire d'une langue quelconque? *Actes du Premier Congrès International des Linguistes*, Leiden, 1930, pp. 33–36. Reprinted in *Selected Writings*, I, The Hague: Mouton, 1962, pp. 3–6.
Jakobson, R., C. G. M. Fant, and M. Halle
1952 Preliminaries to speech analysis: The distinctive features and their correlates. *Acoustics Laboratory Technical Report*, 13, Massachusetts Institute of Technology, Cambridge MA. Reprinted by MIT Press, Cambridge MA, 1967.
Jesperson, O.
1889 *The articulations of speech sounds represented by analphabetic symbols.* Marburg.
Kuhl, P. K.
1988 Auditory perception and the evolution of speech. *Human Evolution*, **3**, 19–43.
Kuhl, P. K. and A. N. Meltzoff
1982 The bimodal perception of speech in infancy. *Science*, **218**, 1138–1141.
Kuhl, P. K. and A. N. Meltzoff
1984 The intermodal representation of speech in infants. *Infant Behavior and Development*, **7**, 361–381.
Kuhl, P. K. and A. N. Meltzoff
1988 Speech as an intermodal object of perception. In A. Yonas (ed.), *Perceptual development in infancy: Minnesota symposia on child psychology*, Vol. 2, pp. 235–266. Hillsdale NJ: Erlbaum.
Liberman, A. M. and I. G. Mattingly
1985 The motor theory of speech perception revised. *Cognition*, **21**, 1–36.
Liberman, A. M. and I. G. Mattingly
1989 A specialization for speech perception. *Science*, **243**, 489–494.
Matisoff, J. A.
1973 Tonogenesis in Southeast Asia. In L. M. Hyman (ed.), *Consonant types and tone, Southern California Occasional Papers in Linguistics No. 1*, University of Southern California, Los Angeles, California, pp. 73–95.
McCarthy, J. J.
1988 Feature geometry and dependency: A review. *Phonetica*, **45**, 84–108.
McGurk, H. and J. McDonald
1976 Hearing lips and seeing voices. *Nature*, **264**, 746–748.

Pike, K. L.
1943 *Phonetics*. Ann Arbor: University of Michigan Press.
Sagey, E.
1986 The representation of features and relations in nonlinear phonology. Ph.D. dissertation, Massachusetts Institute of Technology.
Sapir, E.
1925 Sound patterns in language. *Language*, **1**, 37–51.
Stevens, K. N.
1972 The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr. and P. D. Denes (eds.), *Human communication: A unified view*, pp. 51–66. New York: McGraw-Hill.
Stevens, K. N.
1985 Evidence for the role of acoustic boundaries in the perception of speech sounds. In V. Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, pp. 243–255. New York: Academic Press.
Stevens, K. N.
1986 Models of phonetic recognition II: An approach to feature-based recognition. In P. Mermelstein (ed.), *Symposium on Units and their Representation in Speech Recognition*, 12th International Congress on Acoustics, Montreal.
Stevens, K. N.
1988 Phonetic features and lexical access. *The Second Symposium on Advanced Man-Machine Interface Through Spoken Language, Hawaii*, **10**, 1–23.
Stevens, K. N.
1989 On the quantal nature of speech. *J. Phonetics*, **17**, 3–45.
Yip, M.
1980 The tonal phonology of Chinese. Ph.D. dissertation, Massachusetts Institute of Technology.
Zwicky, A. M.
1970 The free-ride principle and two rules of complete assimilation in English. *Papers from the Sixth Regional Meeting*, Chicago Linguistic Society, pp. 579–588.