# 20 Toward a Phonetic and Phonological Theory of Redundant Features

Kenneth N. Stevens, Samuel Jay Keyser, and Haruko Kawasaki
*Center for Cognitive Science and*
*Research Laboratory of Electronics*
*Massachusetts Institute of Technology*

## INTRODUCTION

This study investigates the role of redundant features in natural language. When a certain feature marks a distinction in a language, one or more "redundant" features may accompany that feature. Examples are discussed that suggest that certain types of redundant features strengthen the acoustic representation of distinctive features and contribute additional properties that help the listener to perceive the distinction. Some specific hypotheses are then set forth regarding the properties of redundant features and of distinctive features enhanced by redundancy. For example, it is hypothesized that only certain feature pairings play a role in this enhancement process, and that these pairings or linkings are determined by the properties of the speech-production and perception systems; that enhancement by redundant features can be realized in a continuous rather than quantal fashion; and that redundant features are more likely to come into play when the perceptual distinctions signalled by distinctive features are weak. It is further hypothesized that the linking of features for the purpose of enhancement plays an important role in determining the morphophonemic alternations in language.

## DISTINCTIVE FEATURES AND REDUNDANCY

There exists a small set of 20-odd phonetic features that are available for signalling distinctions between words in language (Chomsky & Halle, 1968a; Jakobson, Fant, & Halle, 1963; Jakobson and Waugh, 1979; Ladefoged, 1975). These features are normally organized into bundles that form segments, and





words are constituted from sequences of segments or from matrices of features. These features serve the function of organizing segments into classes. Our point of view is that, when a word is spoken, a particular acoustic property appears in the sound whenever a given feature is being used to identify the word, and thus to distinguish it from other words. This acoustic property is invariant in the sense that it is independent of the context of other features and segments in which the given feature occurs.

The acoustic properties that are used to define the inventory of features are determined by the characteristics of the human auditory system and of the speech-generating system. There are some acoustic properties to which the auditory system responds in a distinctive way, and there are certain acoustic properties that the speech-production mechanism is capable of generating without requiring excessive precision in the control of the articulatory structures (Stevens, 1972). For example, forming a complete closure in the vocal tract and then releasing that closure gives rise to an abrupt increase in amplitude of the sound over a range of frequencies, and the auditory system seems to respond in a distinctive way to this kind of abrupt amplitude change, in contrast to a more gradual change. Thus, a clear distinction can be made between continuant consonants such as [f, s, v, z] and stop consonants such as [p, t, b, d]. Or, when a vowel is formed with a fronted position for the tongue body, the second-formant frequency is high and close to the third formant, whereas for a backed tongue-body position, the second formant is low and close to the first. It has been postulated that there is a critical separation of two formants that divides vowel-like stimuli into two classes that give rise to two distinct kinds of auditory response (Chistovich & Lublinskaya, 1979). These are just two examples of the role played by these quantal principles in establishing an inventory of features.

While all of the 20-odd features are used distinctively in different languages, any given language does not employ all of the features to signal distinctions between words, and certainly does not use the features in all combinations. In principle, assuming 20 features, then $2^{20}$ combinations of these features are possible, but any one language uses only a small fraction of these combinations. A consequence is that there is a potential for redundancy (Jakobson, Fant, & Halle 1963; Jakobson & Waugh 1979). That is, it is possible that a minimal distinction between words can be carried not by not just one feature but by some combination of features. This redundancy could provide the listener with additional acoustic cues that could be used to reduce the possible confusion between words, particularly in situations in which there is noise or in which the speech is not clear for some other reason.

Among the approximately $2^{20}$ combinations of features that are theoretically possible, some cannot occur simply because of constraints on the sound-generating system. That is, some acoustic properties cannot be produced unless others are produced at the same time. For example, the generation of a nasal vowel or consonant (i.e., a segment with the feature [+nasal]) requires that negligible pressure be built up in the vocal tract (i.e., the feature [+sonorant]). In other words, the feature [+sonorant] is required if the sound is to contain the property that indicates the feature [+nasal]. As another example, the feature [strident] can only operate to signal a distinction in segments that are [−sonorant]. That is, it is necessary to build up pressure behind a constriction in order to provide an acoustic property appropriate for the feature [+strident]. These kinds of redundancy are inherent in the speech-production process, and a speaker has little freedom as to whether or not to implement them. They are also inherent in the speech-perception process since the detection of one of the features (e.g., [nasal]) requires that other features (e.g., [sonorant]) be present to act as a carrier for the feature. Ascribed to such physiological and perceptual constraints, this type of redundancy should be and indeed is observed cross-linguistically.

The redundancy just discussed is rooted in the intrinsic nature of the vocal tract and the auditory system. It is impossible for one feature to be implemented unless a second feature or feature-set is implemented. There is, however, another kind of redundancy which takes advantage of the fact, noted above, that languages select only a subset of the available combinations of distinctive features. In other words, some features are redundant in particular contexts by virtue of their not having been selected to do distinctive duty in some language. For example, the feature [back] is not distinctive for the system of obstruent consonants in English. Do such features simply remain unused or does the speech-perception process and the phonological system make use of them in some systematic way? It is suggested that features which have become available in this fashion can be used to enhance the perceptual cues for distinctive features in a given language. In the process of achieving this goal of strengthening the existing acoustic property, the redundant feature results in the introduction of a new acoustic property that co-occurs with the already present property, or may, in some contexts, exist in place of the property associated with the distinctive feature.

As will be seen below, enhancement by redundant features of the primary distinctions marked by distinctive features takes a variety of forms. There appear to be some classes of distinctive features that require the help of redundant features and others that do not. Such classes of features need to be identified and accounted for, and the mechanisms whereby enhancement is achieved through the introduction of redundant features needs to be clarified. Furthermore, exploring the ways in which redundant features interact with distinctive features may help to clarify the motivations for some synchronic patterns and historical changes which are not readily accounted for by current phonological theory.



## EXAMPLES OF REDUNDANCY AS AN ENHANCING MECHANISM

Below are examples of cases where redundant features are utilized for enhancing primary distinctions. We describe four of these examples in some detail, and then list a few additional examples with only a brief discussion of each.

### Enhancement of [back]

We define the feature [back] by beginning with a neutral configuration of the tongue body in the front-back dimension, illustrated in Figure 20.1a. The spectrum envelope for a vowel with this vocal-tract configuration is shown in Figure 20.1b. The second formant is midway between the first and the third. Fronting or backing of the tongue body while maintaining the tongue height about the same causes a raising or lowering of the frequency of the second formant, and the effect on the overall spectrum shape is shown in the figure. When the tongue body is displaced back in the mouth, the second formant moves downward in frequency, and becomes closer to the first than to the third formant, as the figure shows. The result is a valley in the spectrum between $F_1$ and $F_2$ that is not as deep as it was for the more neutral configuration, a weaker spectral peak for $F_3$, and a deeper valley in the spectrum between $F_2$ and $F_3$. On the other hand, a forward movement of the tongue body in the mouth leads to a deeper valley in the spectrum between $F_1$ and $F_2$, a shallower valley between $F_2$ and $F_3$, and a raising of the amplitude of the spectral peaks corresponding to $F_2$ and $F_3$ (and possibly $F_4$). The second formant $F_2$ has an affinity with $F_1$ for back vowels and with $F_3$ (and $F_4$) for front vowels.

For the nonlow vowels—i.e., for a nonlow position for the tongue body—backward displacement of the tongue body, without lip rounding, leads to a modest lowering of $F_2$. However, $F_2$ still remains some distance away from $F_1$—the difference is in the range 560–800 Hz. This separation of $F_1$ and $F_2$ can be observed for the unrounded back vowels [ɯ] and [ʌ] in Figure 20.2, which shows formant frequencies for several rounded and unrounded vowels in Korean (Han, 1963). A further lowering of $F_2$—i.e., a further decrease in the distance between $F_2$ and $F_1$—can be achieved by rounding the lips. Formant frequencies for these rounded back vowels are also shown in Figure 20.2, and we see that the spacing between $F_2$ and $F_1$ is reduced by 300 Hz or more by rounding. Thus the acoustic property corresponding to the feature [+back]—i.e., a closer proximity of $F_2$ to $F_1$ than to $F_3$—is enhanced by the feature [+round] for these nonlow vowels. The effect of rounding on the spectrum of a back vowel is illustrated in Figure 20.1c. It is suggested that introduction of the feature [+round] brings $F_2$ and $F_1$ close enough that

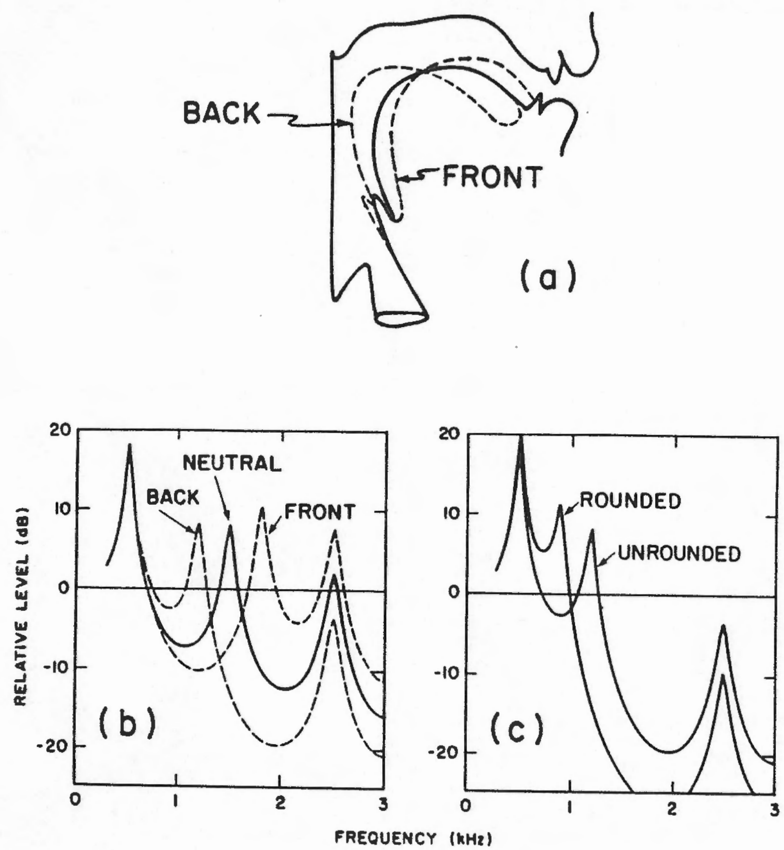


FIG. 20.1. (a) Schematization of midsagittal section of vocal tract for a neutral vowel (solid contour), and for a back and front tongue-body position. (b) Idealized spectrum envelopes corresponding to the three tongue-body configurations in (a). (c) Approximate effect of lip rounding on the spectrum envelope for a back vowel. (A small downward shift in $F_1$ that could also accompany rounding is not shown.)

the two-peaked spectral prominence caused by these formants is interpreted by the auditory system as a single peak whose frequency is at the center of gravity of the prominence, as proposed by Chistovich and Lublinskaya (1979). The feature [round], then, can effect an enhancement of the feature [back] for these nonlow vowels.

For vowels that are [+high], additional factors can come into play to enhance the front-back distinction. Vowels having the feature [+high] are characterized by a low first-formant frequency and by a high tongue-body position



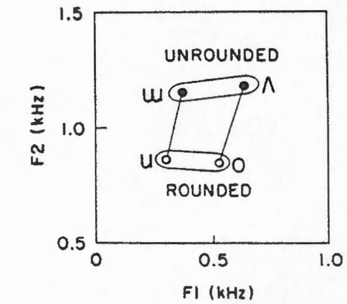


FIG. 20.2. First and second formant frequencies for two unrounded (closed circles) and two rounded (open circles) back vowels in Korean. (Data from Han, 1963.)

that forms a relatively narrow constriction in the oral portion of the vocal tract. For a vowel that is [+high, −back], $F_2$ is close to $F_3$, and both $F_2$ and $F_3$ can be increased still further by raising the tongue blade against the roof of the mouth. The front part of the tongue forms a narrow channel with the hard palate, and this configuration causes $F_3$ to come close to $F_4$. Thus the "center of gravity" of the broad spectral peak formed by $F_2$, $F_3$, and $F_4$ becomes quite high, between $F_3$ and $F_4$ (Carlson, Fant, and Granström 1975). The spectrum resulting from this relatively extreme tongue configuration is shown in the left panel in Figure 20.3. Raising of the tongue blade is the articulatory correlate of the feature [+coronal]. In effect, then, we are using the feature [+coronal] to enhance the feature [−back] for high vowels. The acoustic consequence of this enhancing feature is to produce a diffuse and upward sloping spectrum at high frequencies, similar to the acoustic correlate of [+coronal] for stop consonants (Blumstein & Stevens, 1979).

In the case of [+high, +back] vowels, the feature [+back] can be enhanced by rounding the lips as discussed above. Even further lowering of $F_2$ toward $F_1$ can be achieved by greater narrowing of the lip opening, i.e., by introducing the feature [+labial]. The spectrum for the vowel with these properties is shown in the right panel of Figure 20.3. The diffuse and downward sloping spectrum that results is similar to the acoustic correlate of [+labial] for stop consonants. Thus the feature [+labial] (as well as [+round]) is used to enhance the feature [+back] for high vowels.

## Enhancement of [distributed] and [anterior]

The feature [distributed] appears to be distinctive in some languages for consonants that are [+coronal], i.e., consonants that are produced by raising the tongue blade. From the point of view of articulation, a [-distributed]

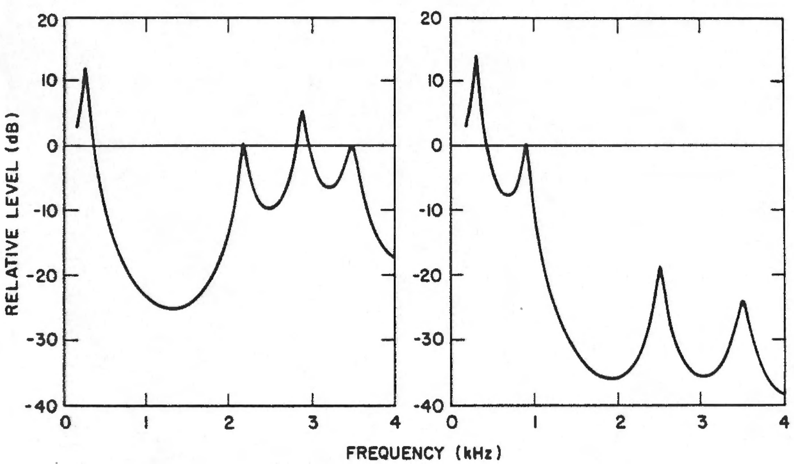


FIG. 20.3. Idealized spectrum envelopes for a high front vowel (left), and a high back rounded vowel (right).

consonant is formed by making a constriction with the apex or tip of the tongue, so that the constriction is quite short. On the other hand, a more extended area of the surface of the tongue blade is used to form the constriction for a [+distributed] consonant, and the length of the constricted region is somewhat greater. This distinction is sometimes called apical/laminal. A dental or palatal consonant is laminal or [+distributed], whereas an alveolar or retroflex consonant is apical or [-distributed]. Drawings that schematize these different configurations of the tongue blade are given in Figure 20.4.

Release of the short apical constriction, involving only movement of the tongue tip, is presumably accomplished much more rapidly than release of the longer laminal constriction. The apical release can be accomplished by a forward and downward movement of only the front part of the tongue blade, whereas the laminal release requires a downward movement of a substantial part of the tongue, including the tongue body. Acoustically, the apical release shows a much more abrupt onset than does the laminal release. This abruptness is a consequence of the very rapid increase in size and decrease in acoustic impedance (proportional to ratio of constriction length to constriction area), which causes a very rapid rise in the frequency of the first formant ($F_1$), probably of just a few msec duration. The laminal release is slower and the constriction is longer, and, as a consequence, the rise in $F_1$ would tend to be slower. This slow release and slow rise in $F_1$ leads to a relatively slow increase in amplitude following the release.

This difference in the release characteristics is illustrated in Figure 20.5, which shows waveforms and spectrograms at the release of the so-called den-



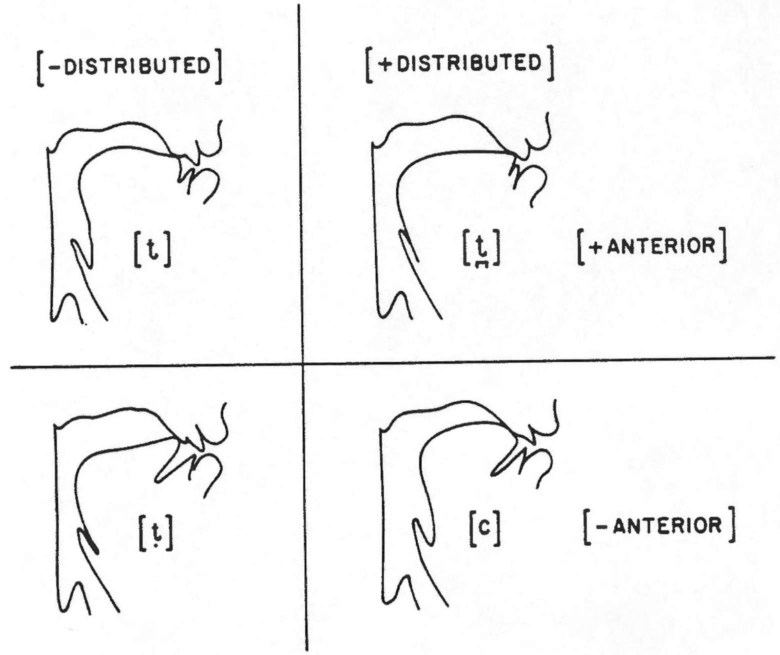


FIG. 20.4. Schematized versions of midsagittal sections showing positions of the tongue blade for several coronal consonants. The two sketches on the right represent laminal consonants (i.e., [+distributed]), and the two on the left are apical (i.e., [−distributed]). The anterior-nonanterior distinction is schematized on the two rows. The estimates of the configurations in the vicinity of the tongue blade and tongue tip are based on midsagittal sketches and other articulatory data reported by Perkell (1969) (for [t]), Ladefoged and Bhaskararao (1983) and Wierzchowska (1965) (for [t̪]), Perkell, Boyce, and Stevens (1979) (for the palatoalveolar configuration), and Dixon (1980) (who shows sketches of all four configurations). The tongue-body shape posterior to the tongue blade has been estimated, with the constraint that the midsagittal length of the dorsal tongue surface is about the same in all cases. This constraintresults in a more backed tongue body for the upper-right and lower-left configurations than for the other two. These estimated tongue-body shapes are also based on acoustic data of the type shown in Figure 20.6 and described in the text. There are differences in the positioning of the lateral surfaces of the tongue blade that are not shown in these midsagittal sections.

tal (laminal) and alveolar (apical) consonants [t̪] and [t], which contrast in Malayalam) (Ladefoged, 1971) and in a number of Australian languages (Dixon, 1980). The slightly faster rise in $F_1$ for the apical is evident from the spectrogram and the more abrupt rise in amplitude followed immediately by a decay can be seen on the waveform for the apical. Similar acoustic properties can be seen for the palatal stop consonant [c], which is [+distributed]

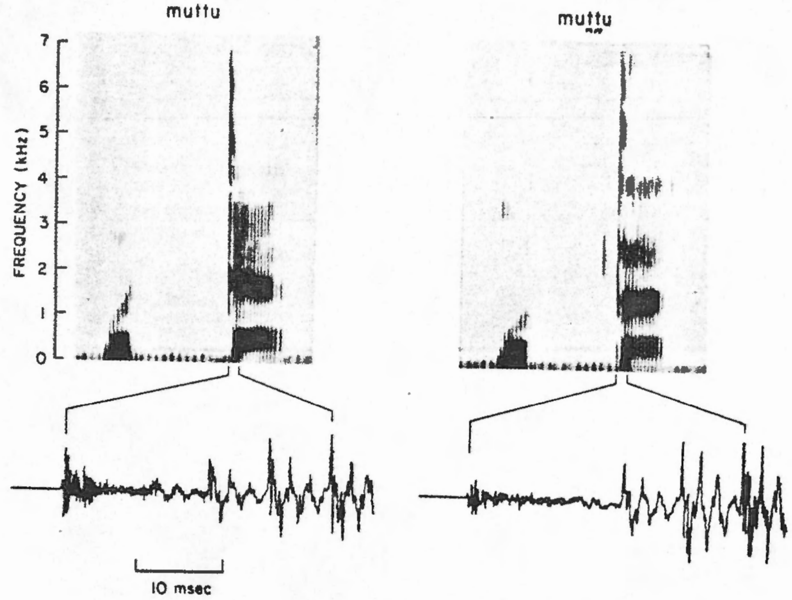


FIG. 20.5. These spectrograms and waveforms show the distinction between dental [t̪] and apical [ṭ] in Malayalam. The waveforms at the bottom illustrate the difference in abruptness of the release for the two consonants. The spectrograms above illustrate differences in $F_2$ immediately preceding and following the consonantal closure, indicating a difference in tongue body position for the two classes of consonants.

and the retroflex [ṭ], which is [−distributed], as well as for the corresponding pairs of nasal consonants.

Closer examination of the acoustic data for distributed and nondistributed consonants reveals that another property comes into play for the distributed/nondistributed contrast, and we interpret this additional property to be the result of the introduction of enhancing or redundant features. In the spectrograms in Figure 20.5, we observe marked differences in the frequency of $F_2$ immediately preceding the implosion and also immediately following the release. $F_2$ is considerably higher for the [−distributed] alveolar than for the [+distributed] dental, reflecting a more fronted tongue-body position for the [−distributed] member of the pair. Examples of data from measurements of $F_2$ at four points preceding and following the closure, for several different utterances, are shown in Figure 20.6 for three speakers. The kinds of differences shown in Figure 20.6, as well as the release characteristics illustrated in Figure 20.5, have been observed in many examples of intervocalic stops and nasals produced by Malayalam speakers.



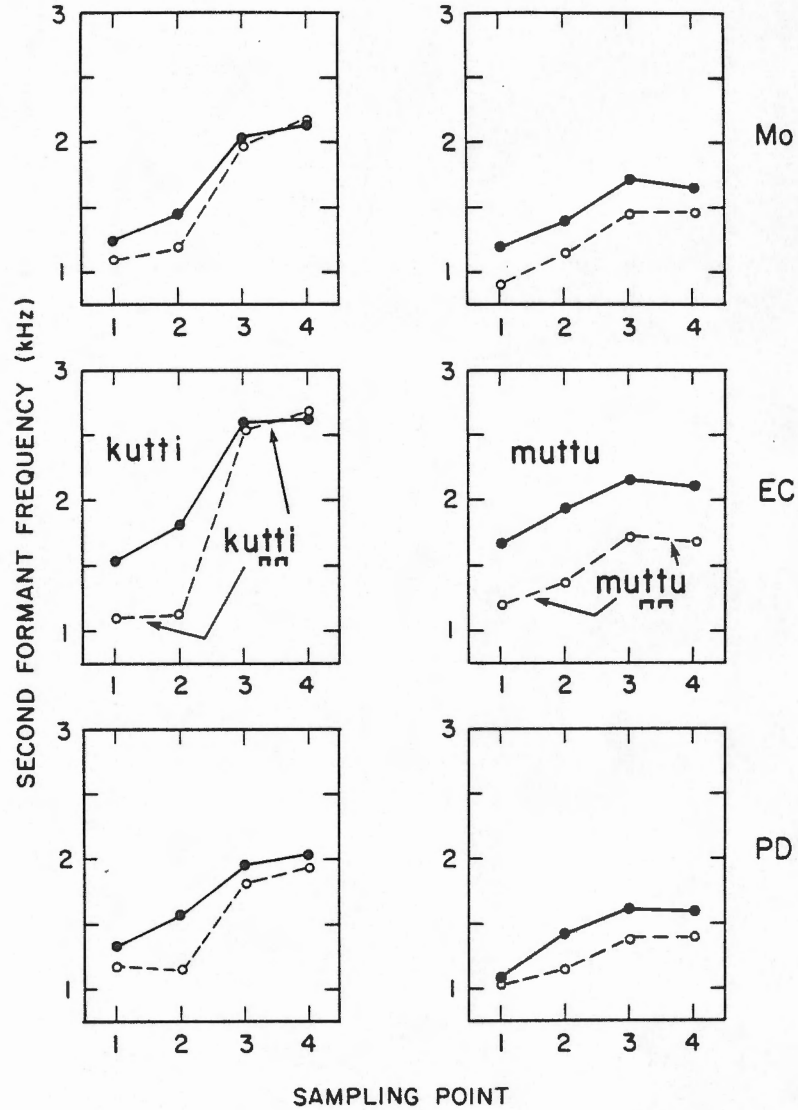


FIG. 20.6. Data on second formant frequency sampled at four points preceding and following the consonantal closure for the word pairs [kutti-kut̪t̪i] (left) and [mut̪t̪u-muttu] (right) in Malayalam, spoken by two male speakers and one female (EC). Each point represents average data from two utterances. The sampling points are: (1) 30 msec prior to consonant closure; (2) as consonant closure; (3) at onset of voicing; (4) 30 msec after onset of voicing.

We suggest, then, that the feature [back] is playing an enhancing role for these coronal consonants: [+back] enhances [+distributed] and [−back] enhances [−distributed]. A fronted tongue-body presumably provides a favorable posture from which the apico-alveolar constriction can be achieved, whereas a backed tongue-body position provides a posture that favors formation of the dental or interdental constriction. Formation of the dental or interdental constriction requires that lateral edges of the tongue blade make contact with the alveolar ridge over an extended region from the central incisors to the molars. This mode of contact in turn can be achieved more readily if the surface of the tongue blade is lowered. Since the tongue is a noncompressible mass, this tongue blade lowering may only take place if the tongue-body is displaced backward. The sketches in Figure 20.4 show how these configurations might appear in the midline. They indicate more backed tongue-body configurations for [t̪] than for [t].

We conclude from these observations that the feature) [back] is introduced to enhance the feature [distributed] for these coronal consonants. Not shown in Figure 20.6 are data indicating a small but consistent difference between $F_1$ at points 3 and 4 (following the release) for the apical and laminal consonants, with $F_1$ for the apical being lower. We interpret this difference as indicating a slightly higher tongue-body position for the apical. (The sketches in Figure 20.4 reflect this difference.)

In several Australian languages, there is a series of four coronal consonants, defined by the two features [distributed] and [anterior] (Dixon 1980). For a [−anterior] consonant, the constriction formed by the tongue blade is posterior to the alveolar ridge, and is produced in such a way as to create a space under the tongue blade, as shown in Figure 20.4 above, leading to a low-frequency resonance of the acoustic cavity in front of the constriction. This resonance usually appears to be associated with the third formant of the adjacent vowel. Thus, for example, the languages Kaititj and Alyawarra have two [+distributed] stop consonants: a [+anterior] consonant [t̪] and a [−anterior]) consonant [c]. Spectrograms illustrating these two consonants are shown in Figure 20.7. Acoustic analysis shows that there is a strong $F_3$ peak in the noise for the [c], in relation to the corresponding spectral peak for the vowel, but this peak (and in fact the entire burst) is weaker for the [t̪]. Two other acoustic properties are evident, however, in addition to the property corresponding to the feature [anterior]. First, the palatal stop is affricated, so that it has a [+strident] component. Second, the second formant at the onset of the vowel is much higher for the palatal than it is for the dental. This acoustic property can be taken as evidence that the palatal is [−back] whereas the dental is [+back]. These properties are evident in many exemplars of these consonants in different contexts. We have already provided an articulatory rationale for the backness of dental configurations. Fronting of the tongue-body for palatal configurations permits the tongue blade to bunch up into the palatal arch, as Figure 20.4 shows. We surmise, then, that the features [strident] and [back] are playing an enhancing role to provide additional perceptual cues for the feature [anterior] for the [+distributed] consonants. We shall observe later that the role of [back] in enhancing [anterior] has consequences in certain alternations that occur in Lardil, another Australian language.



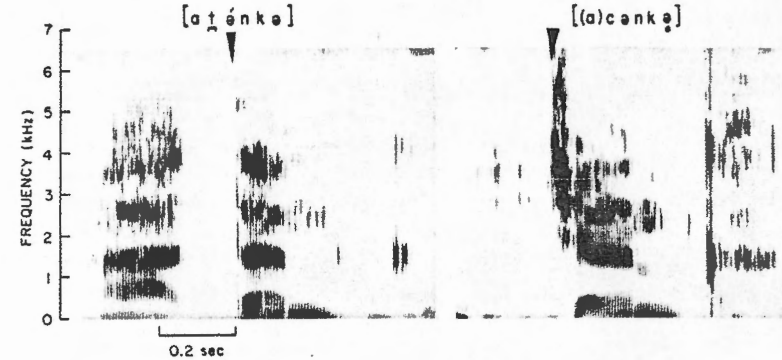


FIG. 20.7. Spectrograms of the words [at̪ ˈənkə] (top) and [(a)cənkə̣] produced by a speaker of Kaititj. These spectrograms illustrate the contrast [±anterior] for these two laminal consonants. The stop release for the [t̪] and [c] are indicated by arrows at the top of the spectrograms. We thank Peter Ladefoged for making available the recording from which this spectrogram was made.

## Enhancement of the Voiced-voiceless Distinction

The basic acoustic correlate of the distinction voiced/voiceless is usually considered to be the presence or absence of low-frequency periodicity in the sound. Vocal-fold vibration is maintained by appropriately adjusting the configuration and state of the vocal folds and by creating a flow of air through the glottis. For example, the low-frequency periodicity usually is observed during the labial closure interval for *rabid* but not for *rapid.* We restrict our discussion here to the voiced/voiceless distinction for consonants, where there appears to be an opportunity for several redundant features to play a role in enhancing this property that signals the presence or absence of low-frequency periodicity due to vocal-fold vibration.

One of the enhancing mechanisms is the feature of length. The feature [-voice] can be enhanced by increasing the duration over which there is lack of low-frequency periodicity. This increased duration of the closure interval for a voiceless consonant relative to that for a voiced consonant has been observed (Peterson & Lehiste, 1960), and is known to provide a cue for distinguishing word pairs like *rapid-rabid* (Lisker, 1957). On the other hand, the duration of the vowel preceding a voiceless consonant can be decreased relative to that for a voiced consonant, again enhancing the property corresponding to the feature [−voice] (Denes, 1955; House & Fairbanks, 1953). This effect is especially strong when the voiceless consonant is in utterance-final position. Both of these duration changes increase the proportion of the syllable duration within which there is lack of low-frequency periodicity. It is assumed that the perception of voicelessness is enhanced if there is an increase in the relative duration of the portion of the syllable without low-frequency periodicity. (See, for example, Port, 1981a and Port & Dalby, 1982.)

Another feature that is available to enhance [−voice] is [+spread glottis]. Adjusting the configuration of the vocal folds so that they are abducted along much of their length can contribute to the inhibition of vocal-fold vibration during the constricted interval. If the glottis is spread at the instant of release of the consonantal constriction, then aspiration noise will occur for a period of time (usually a few tens of msec) until the vocal folds are maneuvered to a more adducted configuration for which vocal-fold vibration is possible. Thus the time interval within which no vocal-fold vibration occurs extends a few tens of msec beyond the release of the constriction (Lisker & Abramson, 1964), and the property of lack of low-frequency periodicity is enhanced. This mode of enhancement is effective in syllable- or word-initial position for which there may not be an opportunity to use a lengthened voiceless interval prior to consonant release as a means to enhance [−voice]. In word-initial position in English, the presence or absence of low-frequency periodicity does not distinguish between voiced or voiceless stops (e.g., [pel] and [bel]). This is a case, then, where an enhancing feature (aspiration or [spread glottis] in this example) provides the primary acoustic evidence, and the acoustic property corresponding to the underlying distinctive feature is not present. (Phonological arguments, for example the plural rule in English, argue for the view that it is voicing which is distinctive in English, and not aspiration.)

Vocal-fold vibration during the closure interval can also be inhibited by increasing the stiffness of the vocal folds. With increased stiffness of the vocal folds, greater pressure across the glottis is needed to initiate vibration. This increased stiffness is usually achieved by increasing the length of the vocal folds, thereby stretching them. Evidence for this increased stiffness comes from examination of the fundamental frequency ($F_0$) immediately following the release of a voiceless consonant (cf. House & Fairbanks, 1953; Jeel, 1975; Lehiste & Peterson, 1961; Löfqvist, 1975; Slis & Cohen, 1969). For many languages, there is an increased $F_0$ during the first few glottal periods following the release of a voiceless consonant, but not for a voiced consonant. Thus the feature [+stiff vocal folds] (or a feature similar to this that causes a raised $F_0$ during a vowel) acts as an enhancing feature for [−voice] for obstruent consonants.



Enhancement of the feature [+voice] can be achieved by increasing the amplitude of the low-frequency sound during the constricted interval, thus increasing the audibility of the low-frequency periodicity. In the case of stop consonants, this increased amplitude can be obtained by opening the velopharyngeal port during the initial part of the closure interval. That is, the feature [+sonorant] is introduced to enhance [+voice], and the [sonorant] feature is implemented with [+nasal]. The resulting nasal murmur has a higher amplitude at low frequencies than does the normal prevoicing that results from sound transmission through the neck tissues. During the latter part of the closure interval, the velopharyngeal port is closed so that intraoral pressure can be built up to obtain the appropriate obstruent release. This prenasalization is used in the production of the intervocalic voiced stops [d] and [g] in the Tosa dialect of Japanese. In the Tokyo and other dialects, the opposition between [k] and [g] is sometimes realized intervocalically as that between [k] and [ŋ]. In this instance as well, the distinctive feature) [+voiced] is enhanced by the redundant feature [+nasal] (Kawasaki, 1981; Oishi & Uemura, 1975).

In the prenasalized example just given, the low-frequency amplitude of the sound is increased by preventing a buildup of intraoral pressure during a portion of the closure interval for the consonant. Significant pressure buildup can also be prevented either by expanding the volume of the vocal tract during the closure interval, or by not making a complete oral closure, thus allowing air to pass through the constriction. While this method enhances the audibility of the low-frequency periodicity, it also reduces or even eliminates the friction noise that occurs at the consonant release. The reduction of intraoral pressure during the closure interval can be viewed as another consequence of the introduction of the feature [+sonorant].

## Enhancement of [continuant] by [strident]

The acoustic correlate of the feature [−continuant] is a relatively abrupt onset of sound pressure, preceded by an interval in which the sound pressure is relatively low. Thus for a consonantal segment characterized by the feature [+continuant], there is some acoustic energy in the constricted interval for the consonant, and possibly an onset that is not so abrupt. For the feature [+strident], the acoustic correlate is a relatively high level of turbulence noise, so that in some part of the high-frequency range the noise exceeds the level of the vowel. The production of stridency is achieved by directing a rapid air stream against some obstacle, thus creating an increased level of turbulence. The acoustic correlate of stridency is illustrated in Figure 20.8, which shows spectrograms of the utterances *a sin* and *a thin* in English, as

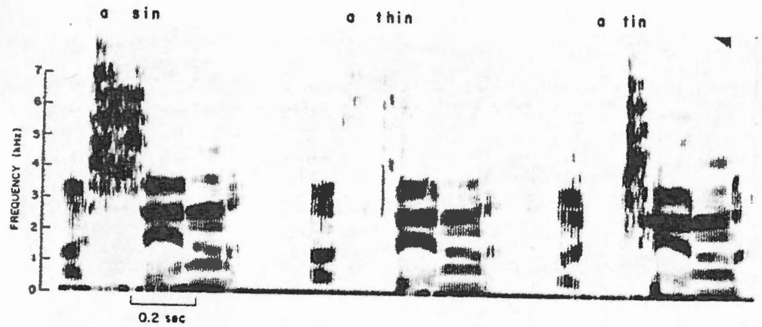


FIG. 20.8. Spectrograms of three utterances: *a sin* (left), *a thin* (middle), and *a tin* (right).

well as the utterance *a tin*, which contains a stop consonant which is inherently [−strident] (although in this example some stridency is introduced following the release, thereby enhancing the acoustic cue for the [+anterior] stop, as noted under "Other Examples" below). The weaker noise for the [θ] is quite evident in these utterances. This acoustic correlate of stridency is also supported by perceptual experiments (Heinz & Stevens, 1961; Stevens, 1981a). It is reasonable to assume, then, that [+strident] is a candidate for enhancing the feature [+continuant] for obstruent segments. The presence of the relatively high level of noise during the constricted interval serves to accentuate the contrast with a [−continuant] or stop consonant, for which the sound pressure during this interval must be minimal. A consequence of this observation would be that, for a language with just one labial fricative, that fricative would tend to be labiodental, or strident (with the upper lip acting as the "obstacle" against which the air stream is directed). Examination of the Stanford Archive (Crothers, Lorentz, Sherman, & Vihman, 1979) supports this prediction, since it reveals that, among 68 languages that have a single voiceless labial-fricative phoneme, 62 have the strident labiodental [f] and only six have the nonstrident bilabial [Φ]. Likewise, coronal fricatives would tend to be strident; i.e., a language would tend to have [s] rather than [θ], if there were just one coronal fricative. There may also be a tendency to avoid fricatives produced with a place of articulation (such as velar) that does not allow for stridency, since there is no readily available obstacle against which to direct the air stream.

## Other Examples

*[coronal] and [strident].* In English the stop consonant [t] is distinguished from [p] and [k] by the feature [coronal]. The acoustic correlate of this feature appears as a spectrum amplitude that increases with increasing frequency



(Blumstein & Stevens, 1979). Some speakers introduce affrication into the burst for [t] in some phonetic environments, as in the example in Figure 20.8 above. In effect, the feature [strident] is introduced to produce greater high-frequency energy immediately following the consonant release, and thus acts to enhance the [coronal] feature. In this context the feature [strident] is redundant.

*[high] and diphthongization.* In English and in Swedish, the long [+high] vowels [i] and [u] are diphthongized to [iy] and [uw], respectively (Fant, 1973; Jones, 1966). It can be argued that the diphthongization toward [y] or [w] accentuates the acoustic attribute associated with [+high]—i.e., a low first-formant frequency—since these offglides cause a movement of $F_1$ to an even lower frequency. (Presumably the entire vowel is not shifted toward [y] or [w] because a degree of syllabicity must be retained, and this requires that $F_1$ not be too low near the vowel onset.) This accentuation of the feature [high] is accomplished by the introduction of a nonsyllabic offglide toward the middle of the vowel. The acoustic manifestation of this offglide is a falling of the frequency of the first formant and a concomitant reduction in amplitude.

*[low tone] and [constricted glottis].* In several Southeast Asian languages, including Yao and Mandarin, phonemic low/falling/dipping tones are often accompanied by creaky voice (cf. Chao, 1968; Purnell, 1965). Presumably, creaky voice enables the low-frequency vocal-fold vibration needed for low tones to be even lower. This enhancement of the low tone is accompanied by a modified voice quality, associated with the feature [constricted glottis]. (In this respect, creaky voice and diphthongization are alike in that both are introduced midway in a vowel to enhance an acoustic property associated with a distinctive feature.)

*[voice] and [sonorant].* The distinguishing articulatory characteristic of the feature [+sonorant] is that there is essentially no pressure buildup behind a constriction in the vocal tract above the glottis. Thus, essentially no frication noise is generated along the vocal tract above the glottis. The generation of sound during a sonorant interval is achieved with an acoustic source at the glottis, and this source can either be voicing as a consequence of vocal-fold vibrations or aspiration as a consequence of turbulent airflow in the vicinity of a spread glottis. The airflow is much greater for the voiceless or aspirated case than for the voiced case, and this greater airflow during aspiration is more likely to give rise to turbulence noise at a narrow portion of the vocal tract above the glottis. Voicing is the preferred mode of sound generation during a sonorant interval, since it lessens the possibility of generating frication noise. In this sense, then, the feature [+voice] enhances [+sonorant].

## IMPLICATIONS OF REDUNDANCY AS ENHANCEMENT

The picture that emerges from the examples and the discussion given is the following. A given language makes use of a set of distinctive features to specify the inventory of segments that make up the lexical items of that language and in terms of which the phonological component of the language is structured. The distinctive features are drawn from a larger inventory of features that are universally available. Associated with each feature is an acoustic property and an articulatory realization. The features from this inventory that are not distinctive within the language are often used to assist in the implementation of the features that are. Each of these enhancing, or redundant, features serves to strengthen the acoustic property associated with a distinctive feature and, in addition, to contribute its own acoustic property. In what follows, we describe three implications that arise out of these examples.

### Nonindependence of Features

In all of the examples given above, we observe the pairing of features; namely, a redundant feature enhances a distinctive feature. An immediate question is: What is the basis for this linking of redundant features with distinctive features? It is our view that this pairing is not random, but rather is based upon characteristics of speech production and auditory perception. For a particular distinctive feature, there are only certain redundant features that can enhance or strengthen the acoustic property associated with the distinctive feature. Thus, rounding enhances backness by virtue of its reinforcing the low-frequency spectral prominence resulting from the proximity of $F_1$ and $F_2$. If the possible pairings of distinctive and redundant features are based upon the characteristics of the production and perception systems, then it is expected that this affinity between particular features is universal, i.e., it will be observed across languages. While this affinity may lurk, as it were, in the background, a given language may choose not to exploit the affinity. Further work is clearly needed in order to investigate in detail what principles give rise to those pairings.

An implication of this point of view is that redundant and distinctive features are all drawn from the same inventory. That is, among the list of 20-odd features discussed above, potentially any member of the list either can serve as a distinctive feature or can play the role of a redundant or enhancement feature. There is, however, another possibility. A given distinctive feature/redundant feature pair (or set of pairs) may actually represent different acoustic and articulatory manifestations of a single more abstract feature, a so-called "cover feature." Examples of possible candidates for such a feature are the feature [tense] for vowels, or a generalized feature of voicing or tenseness for consonants, which at first glance appear to have a variety of acoustic correlates. In the examples given above, it does not seem to be necessary to introduce this kind of feature. Nonetheless, we would like to keep open the possibility that distinctive feature theory may make use of "cover features" as well as distinctive features that have well-defined simple acoustic correlates such as nasality or stridency.



### Discrete and Continuous Manifestations of Features

A second implication which emerges from the examples reviewed above is that each case of enhancement results not only in the introduction of additional acoustic properties, but also in the strengthening of an acoustic property associated with a distinctive feature. An implication is that the acoustic correlate of a distinctive feature can be instantiated with varying degrees of strength. A theory of this kind would represent the merging of quantal theory (Stevens, 1972) with viewpoints which assign a more continuous character to phonetic events (Liljencrants & Lindblom, 1972).

### Enhancement and Saliency

The previous two sections concern the nature of enhancement. We would also like to look at the question of when enhancement occurs. The above examples suggest that enhancement may occur when the acoustic property associated with a particular distinctive feature cannot be reliably detected by listeners. That is, enhancing occurs when confusion is likely to exist between utterances that are distinguished only on the basis of that feature. The introduction of a redundant feature in these situations can serve to increase the "perceptual distance" between these utterances (cf. Klatt, 1982a, b), and thus to increase the reliability with which a listener can make the distinction.

In most of the examples given above, redundant features come into play in the presence of features indicating place of articulation or the state of the larynx . One might ask whether these classes of features are less salient perceptually, and are, therefore, most in need of enhancement. In order to answer this question, we will need to explain in a more quantitative way what is meant by "saliency" or "perceptual distance."

## PHONOLOGICAL IMPLICATIONS

An important outcome of the discussion of redundancy given above is that there is a linking, or dependency, between phonetic features. According to this view, there are other redundant features associated with certain distinctive features that serve to enhance the acoustic properties that signal the distinctive features. If features are interdependent, then one would expect to find evidence for this interdependence in phonological rules. In fact, rule systems have not been viewed from this perspective. In what follows, two examples are discussed which re-examine two phonological rules from the point of view of enhancement. The first rule comes from Lardil and attempts to show that a well-motivated phonological rule of that language which relates the features [distributed], [anterior], and [back] can be viewed in the light of the discussion above. The second rule comes from English and attempts to demonstrate that the role of stridency in the phenomena of Spirantization and Palatalization is, in fact, an enhancement role of the feature [continuant] by the feature [strident].

## [back] and [anterior] in Lardil

We turn first to the Lardil example. The relevant data are drawn from Klokeid (1976), who notes the following distribution in Lardil, an Australian Aboriginal language:

(1)

| *nominative* | *accusative* | *nonfuture* | *future* | |
|---|---|---|---|---|
| ngawid | ngawij-in | ngawith-ad | ngawith-ur | "stomach" |
| yarpud | yarpuj-in | yarputh-ad | yarputh-ur | "snake, bird" |
| ngampid | ngampij-in | ngampith-ad | ngampith-ur | "lumpy" |
| kaljid | kaljij-in | kaljith-ad | kaljith-ur | "urine" |

In Klokeid's (1976) orthography, the symbol j stands for a laminoalveolar coronal while the symbol th stands for a laminodental. Thus, we find in (1) alternations whereby a laminodental appears before back vowels and a laminoalveolar appears before front vowels. This is a pattern that appears to be widespread in Australian languages (Dixon, 1970; Dixon & Blake, 1979, 1981). Klokeid's description, as well as the detailed descriptions of others such as Dixon (1980), makes it clear that the laminodental consonants are [+anterior] while the laminoalveolar consonants are [−anterior]. Klokeid's account also establishes that both of these allophones stem from an underlying /t/ is optionally flapped in word-final position (Klokeid, 1976, pp. 39–40). To account for these alternations Klokeid postulates the following rule:



(2)

$$/t/ \longrightarrow \begin{bmatrix} +\text{distributed} \\ \alpha\ \text{anterior} \end{bmatrix} \ / \ \underline{\qquad} \begin{bmatrix} +\text{V} \\ \alpha\ \text{back} \end{bmatrix}$$

This rule states that the apical root-final /t/ becomes distributed and that, furthermore, there is a correlation between backness and anteriority which takes place among the resultant distributed coronals. In particular, back vowels induce [+anterior] coronals while front vowels induce [−anterior] coronals.

As it stands, there is no motivation for the correlation exhibited in (2). Such a motivation can be found, however, in the discussion given earlier and illustrated in Figure 20.4, in which we inferred from acoustic data that, for [+distributed] coronal consonants in some other Australian languages, [+anterior], or dental, consonants are produced with a more backed tongue-body position than are [−anterior], or palatal, consonants. We also argued from anatomical considerations that this front or back positioning of the tongue-body provides a configuration that enhances the ability of the speaker to achieve the acoustic characteristics appropriate for the palatal or dental consonant, respectively. It could also be argued that the reverse is true. That is, for the class of [+distributed] coronal consonants, achievement of a backed tongue-body configuration is enhanced by using a dental (i.e. [+anterior]) tongue-blade articulation, whereas achievement of a fronted tongue-body configuration is enhanced by using a palatal (i.e. [−anterior]) tongue blade articulation. This correlation between [back] and [anterior] can be formalized by hypothesizing the following universal linking convention (cf. Chomsky & Halle, 1968a, pp. 421ff):

(3)

$$[u\ \text{back}] \longrightarrow [\alpha\ \text{back}] \ / \ \begin{bmatrix} \underline{\qquad\qquad} \\ +\text{distributed} \\ \alpha\ \text{anterior} \end{bmatrix}$$

where [*u* back] refers to "unmarked for back." This convention states that universally for distributed consonants, the features [back] and [anterior] are correlated whenever they are free to do so by virtue of either one or the other being redundant in the phonology of a given language.

Given linking convention (3), we restate (2) as follows:

(4)

$$/t/ \longrightarrow \begin{bmatrix} +\text{distributed} \\ \alpha\ \text{back} \\ \alpha\ \text{anterior} \end{bmatrix} \ / \ \underline{\qquad} \begin{bmatrix} +\text{V} \\ \alpha\ \text{back} \end{bmatrix}$$

The value of the feature [back] on the V determines the value of the same feature on the preceding consonant. We assume that linking convention (3) motivates the correlation seen in (4) between [back] and [anterior]. Thus, (4) replicates the correlation that (3) establishes on grounds of enhancement.

If we are correct in supposing that linking conventions are based upon enhancement properties of feature pairs or pair-sets, then we may be able to shed new light on such conventions. In particular, we may be able to sustain the claim that in such conventions, the relevant features that occur in phonological rules do so precisely because they are in an enhancement relationship with one another. In the present case, we have argued that the correlation of [back] and [anterior] is explicable in terms of a prior theory of constraints between the tongue-body, the tongue blade and the mode of making contact with the hard palate.[1]

There is a further consequence of this point of view which is worth considering. In Chomsky and Halle (1968a), markedness and linking conventions were primarily employed in order to provide an evaluation metric for distinguishing between simple and complex lexical entries. This was accomplished by using values such as +, −, *m* and *u* in the underlying specification of feature bundles. Those representations which contained the fewest pluses and minuses and the most *u*'s were the simplest. This metric, it should be noted, is based upon the notion that underlying representations have to be fully specified for all features. Recent work (Kiparsky, 1982) has suggested that this may not be so and that underspecification of features in the lexicon is necessary. In this case, the same kind of metric would hold. Instead of giving the highest valuation to those entries containing the most *u*'s, one would give the highest valuation to those entries containing the most empty cells.

## Stridency in English: Spirantization and Palatalization

In English, there is a battery of rules (cf. Chomsky & Halle, 1968a, 223ff) whereby a stem final /t/ or /d/ (in words like *Egypt* and *invade*), when followed by a morpheme beginning with a palatal glide such as [yən], becomes corresponding) [−anterior,+continuant,+strident] segments. Thus, /t/ becomes [š] and /d/ becomes [ž] in *Egyptian* and in *invasion*, respectively. These forms are assumed to derive from the underlying morpheme sequences [Egypt + yən] and [invad + yən].[2] These alternations are in keeping with the tendency noted earlier for [+continuant, −sonorant] segments to be [+strident]; i.e., for stridency to enhance the property of continuancy. In English, stridency is distinctive for [+anterior, +coronal] continuants (e.g., [s] and [θ], but is redundant for [−anterior, +coronal] segments such as [š]. In

[1] The linking conventions discussed above offer one way of formalizing the enhancement relationship. However, one must examine critically whether linking conventions do, in fact, constitute a viable formalism for enhancement or whether some other formalism is required.

[2] The major rules involved in these alternations are Spirantization and Palatalization.



the view proposed here, then, stridency is available for enhancement in this latter context. Thus, we add the feature [+strident] to the list given above for a palatal fricative, resulting finally in [š] or [ž], as in words like *Egyptian* and *invasion.*

As with linking convention (3), we may propose a universal linking convention such as the following:

(5)

$$[u\ \text{strident}] \rightarrow [+\text{strident}]\ /\ \begin{bmatrix} \_\_\_\_ \\ +\text{continuant} \\ -\text{sonorant} \end{bmatrix}$$

This convention marks a nonsonorant continuant as [+strident] unless stridency is being used distinctively for that segment. In particular, in English a palatal fricative is redundantly strident. The convention will operate on the segments /t/ and /d/ after the rules of spirantization and palatalization have applied to render these segments [+strident] as well as [−anterior, +continuant], i.e., [š] and [ž]. Thus, the process of enhancement of continuancy by stridency is captured in our theory through the device of a linking convention which automatically introduces [+strident] when the phonology sets the stage for it.

## SUMMARY AND CONCLUDING REMARKS

This paper has been concerned with a relationship between distinctive and redundant features which has been noted in the literature but has not been studied in detail. We have observed that no language utilizes fully the entire set of 20-odd distinctive features which are hypothesized to describe the phonological and phonetic components of human language. A feature in a given language is often distinctive only in the environment of certain other features in that language.

Redundancy appears to be of two types. One type is obligatory, in that certain feature combinations always occur together. This type of redundancy is a direct consequence of the properties of the vocal tract and the auditory system. The other type, which is of interest in this paper, is optional in character, in that it may or may not be selected for use in particular language. We have proposed that languages make use of these redundant features in order to enhance acoustic properties associated with distinctive feature oppositions. Several examples have been given to illustrate how certain features might reasonably be viewed, within the framework of contemporary theories of speech production and perception, as enhancing or being enhanced by other features. Features chosen for detailed discussion included [distributed], motivates the correlation seen in (4) between [back] and [anterior]. Thus, (4) replicates the correlation that (3) establishes on grounds of enhancement.

If we are correct in supposing that linking conventions are based upon enhancement properties of feature pairs or pair-sets, then we may be able to shed new light on such conventions. In particular, we may be able to sustain the claim that in such conventions, the relevant features that occur in phonological rules do so precisely because they are in an enhancement relationship with one another. In the present case, we have argued that the correlation of [back] and [anterior] is explicable in terms of a prior theory of constraints between the tongue-body, the tongue blade and the mode of making contact with the hard palate.[1]

There is a further consequence of this point of view which is worth considering. In Chomsky and Halle (1968a), markedness and linking conventions were primarily employed in order to provide an evaluation metric for distinguishing between simple and complex lexical entries. This was accomplished by using values such as +, −, *m* and *u* in the underlying specification of feature bundles. Those representations which contained the fewest pluses and minuses and the most *u*'s were the simplest. This metric, it should be noted, is based upon the notion that underlying representations have to be fully specified for all features. Recent work (Kiparsky, 1982) has suggested that this may not be so and that underspecification of features in the lexicon is necessary. In this case, the same kind of metric would hold. Instead of giving the highest valuation to those entries containing the most *u*'s, one would give the highest valuation to those entries containing the most empty cells.

### Stridency in English: Spirantization and Palatalization

In English, there is a battery of rules (cf. Chomsky & Halle, 1968a, 223ff) whereby a stem final /t/ or /d/ (in words like *Egypt* and *invade*), when followed by a morpheme beginning with a palatal glide such as [yən], becomes corresponding) [−anterior,+continuant,+strident] segments. Thus, /t/ becomes [š] and /d/ becomes [ž] in *Egyptian* and in *invasion*, respectively. These forms are assumed to derive from the underlying morpheme sequences [Egypt + yən] and [invad + yən].[2] These alternations are in keeping with the tendency noted earlier for [+continuant, −sonorant] segments to be [+strident]; i.e., for stridency to enhance the property of continuancy. In English, stridency is distinctive for [+anterior, +coronal] continuants (e.g., [s] and [θ], but is redundant for [−anterior, +coronal] segments such as [š]. In

[1] The linking conventions discussed above offer one way of formalizing the enhancement relationship. However, one must examine critically whether linking conventions do, in fact, constitute a viable formalism for enhancement or whether some other formalism is required.

[2] The major rules involved in these alternations are Spirantization and Palatalization.



the view proposed here, then, stridency is available for enhancement in this latter context. Thus, we add the feature [+strident] to the list given above for a palatal fricative, resulting finally in [š] or [ž], as in words like *Egyptian* and *invasion*.

As with linking convention (3), we may propose a universal linking convention such as the following:

(5)

$$[u\ \text{strident}] \rightarrow [+\text{strident}] \ / \ \begin{bmatrix} \_\_\_\_ \\ +\text{continuant} \\ -\text{sonorant} \end{bmatrix}$$

This convention marks a nonsonorant continuant as [+strident] unless stridency is being used distinctively for that segment. In particular, in English a palatal fricative is redundantly strident. The convention will operate on the segments /t/ and /d/ after the rules of spirantization and palatalization have applied to render these segments [+strident] as well as [−anterior, +continuant], i.e., [š] and [ž]. Thus, the process of enhancement of continuancy by stridency is captured in our theory through the device of a linking convention which automatically introduces [+strident] when the phonology sets the stage for it.

## SUMMARY AND CONCLUDING REMARKS

This paper has been concerned with a relationship between distinctive and redundant features which has been noted in the literature but has not been studied in detail. We have observed that no language utilizes fully the entire set of 20-odd distinctive features which are hypothesized to describe the phonological and phonetic components of human language. A feature in a given language is often distinctive only in the environment of certain other features in that language.

Redundancy appears to be of two types. One type is obligatory, in that certain feature combinations always occur together. This type of redundancy is a direct consequence of the properties of the vocal tract and the auditory system. The other type, which is of interest in this paper, is optional in character, in that it may or may not be selected for use in particular language. We have proposed that languages make use of these redundant features in order to enhance acoustic properties associated with distinctive feature oppositions. Several examples have been given to illustrate how certain features might reasonably be viewed, within the framework of contemporary theories of speech production and perception, as enhancing or being enhanced by other features. Features chosen for detailed discussion included [distributed], [anterior], [back], [voice], [continuant], and [strident]. A number of other features and processes were touched upon in less detail. These examples indicate ways in which the enhancement process may operate. The manipulation of the articulatory mechanisms corresponding to one feature (the redundant or enhancing feature) interacts with the process of producing another feature (the distinctive feature). The redundant feature produces its own acoustic property but the interaction also results in conditions whereby the acoustic property for the distinctive feature is produced with greater strength, with greater reliability, or with less articulatory effort. These examples point toward a *linking* between two or more features: one of these functions distinctively and the other(s) redundantly to enhance the acoustic property of the distinctive feature.

We raise the question as to whether phonological rules associate features which are related in terms of this linking. To shed light on this question, we have discussed two rules in some detail, one drawn from Lardil and the other from English. A conclusion from these and other examples is that this pairing of redundant and distinctive features does seem to play a role in certain phonological rules, and that it might be useful to review phonological rules from the point of view of enhancement linkages. In fact, the theory may be able to account for a significant number of feature pairings that one finds in such rules, particularly those which reoccur from one language to another.

The concepts of redundancy and enhancement that have been introduced here could conceivably form a basis for gaining insight into several areas not considered in this paper:

1. When a redundant or enhancing feature exists alongside a distinctive feature, there is a possibility that the redundant feature can take over distinctive duty, and that the role of the distinctive feature is reduced or even eliminated. The contrast between lexical items is now carried by a different feature and sound change has occurred. Do enhancing or redundant features play a role in indicating what sound changes are possible?

2. A consequence of the fact that some features can enhance others is that certain combinations of features are more likely to occur in certain languages than others. (For example, back vowels tend to be rounded, continuant obstruents tend to be strident, sonorants tend to be voiced and so on.) A further possible consequence is that, in the process of acquiring the phonological aspects of language, a child is more likely to produce and to respond to some feature combinations than others. Is is possible to interpret acquisition data in terms of the feature linking and enhancement processes of the type discussed in this paper?

3. Redundant features may be utilized optionally by speakers in some contexts, since these features are not required to distinguish between words, particularly when the communication link is relatively free of interference. Thus when redundant features occur, there is an opportunity for variability in the way a particular word is actualized. Some of this variability is at the option of the speaker and some is determined by the context in which the features occur. The concepts of redundancy thus provide us with a framework within which to examine invariance and variability in the speech process - invariance in the relation between features and acoustic properties, and variability in the selection of features to utilize in a particular context and speaking situation.



This paper is offered as a first step in approaching these and other questions.

## ACKNOWLEDGEMENTS

Preparation of this manuscript was supported in part by the MIT Center for Cognitive Science under a grant from the A. P. Sloan Foundation's Particular Program in Cognitive Science, in part by Grant No. NS-04332 from the National Institute of Neurological and Communicative Disorders and Stroke, and in part by Grant No. NSF MCS-81-12899 from the National Science Foundation. We would like to thank K. Hale for discussion about Lardil.

## Bruno H. Repp: Comment

I find the theory sketched by Stevens, Keyser, and Kawasaki interesting and promising. It has important implications for the description of phonological structure and for the explanation of this structure in terms of articulatory and perceptual principles. Thus the theory must be welcomed as a significant attempt to bridge the often-lamented gap between phonology and phonetics.

Since I am not a linguist, I have little to say about the phonological side of the theory. I also presume that this aspect is less directly relevant to the concerns of this volume. As to the phonetic side, my aim will be to comment on some of the basic assumptions underlying the theory.

One fundamental assumption is that there is a small set of phonetic features that are common to all languages in the world and are in one-to-one correspondence with certain invariant properties in the speech signal. Although this assumption is stated as if it were an established truth, I presume Stevens et al. would agree that it can be decomposed into several semi-independent hypotheses, all of which are in need of further empirical support. These hypotheses are: (1) the number of phonetic features is small; (2) this small set is universal; (3) the acoustic signal contains certain invariant properties; (4) these properties are in a one-to-one relation with phonetic features. None of these component assumptions is universally accepted, and the last two hypotheses in particular cannot be considered well-supported at present and are the subject of ongoing research. Therefore, they are best considered working assumptions, even if they serve as axioms within the present theory.

The main focus of the authors' theory is on acoustic properties of the speech signal. Those critical properties that are taken to carry the distinctive information are assumed to be the joint result of articulatory and perceptual constraints in the evolution of language. On the articulatory side, a preference for certain regions of maximal stability is predicted; on the perceptual side, a maximally distinct sound output will be preferred. These are familiar assumptions, but they are nevertheless very much in need of further empirical support. While pressures toward articulatory and perceptual ease—to the extent that these concepts can be defined clearly—certainly must play a role in speech communication, there are many other factors at work as well. Skilled behavior, of which I take speech to be an instance, often requires the overcoming of preferences inherent in the perceptual and motor systems. In fact, it might be argued that, if this were not so, the behavior in question should not be called a skill. While there are clearly defined physical limits that cannot be exceeded, there is considerable freedom and flexibility within these limits. I agree, however, that in order to discover and describe this flexibility, a good hypothesis to pursue is that the perceptual and motor systems are tightly constrained.

The authors' main concepts are those of feature redundancy and enhancement. Feature redundancy is predicted on purely formal grounds: There are more possible combinations of features than there are phonetic distinctions in a language, even though many feature combinations do not actually occur because of articulatory restrictions, and also because some distinctive features devised by phonologists are by definition dependent on others. In addition to this *necessary* redundancy, however, Stevens et al. point out that "some features are redundant in particular contexts by virtue of their not having been selected to do distinctive duty in some language."

I am slightly confused by the argument here. If a feature is not used at all, then it does not vary and therefore cannot be redundant. In order to be redundant, a feature must exhibit some variation. Whether it performs distinctive duty in that case is apparently decided by Stevens et al. on the basis of phonological theory. However, we know from many perceptual experiments that virtually every acoustic property that covaries with a phonetic distinction can aid a listener in making that distinction; thus, from a perceptual viewpoint, it is not clear how a feature can vary redundantly and at the same time not do distinctive duty. In other words, it seems that the authors must first negate the distinctive perceptual role of certain features in order to be able to introduce purposeful enhancement as a phonological mechanism. This may be a form of theoretical bootstrapping.

It also seems that there are more acoustic properties in the speech signal than there are distinctive features in phonology, notwithstanding all attempts to describe acoustic feature correlates in a maximally integrated fashion. Although the authors clearly assume that there is a one-to-one correspondence between distinctive features and acoustic properties, some of their examples of feature enhancement suggest that they may be willing to consider the perceptual contribution of acoustic properties other than the primary feature correlates.



The feature enhancement hypothesis claims that redundant features may serve to enhance perceptual distinctions. Stevens et al. state that "in the process of achieving this goal of strengthening the existing acoustic property, the redundant feature results in the introduction of a new acoustic property that co-occurs with the already present property." In the specific cases considered, however, a redundant feature generally contributes to the *same* acoustic property as the feature that is assumed to do distinctive duty in the first place. If merely a new acoustic property were added, its effect would not be a strengthening of the existing property but a facilitation of the perceptual distinction according to the principle that two differences are easier to discriminate than one. If I understand the authors correctly, however, they are only concerned with perceptual facilitation due to enhancement of the primary acoustic property, not with facilitation due to several independent acoustic cues. But since they do consider the possibility that a new property due to a redundant feature may replace the original distinctive property, I conclude that redundant features are assumed both to contribute to existing acoustic properties and to introduce new acoustic properties. This seems to contradict the assumption that features are in a one-to-one correspondence with acoustic properties, for how can a single redundant feature contribute to two different acoustic properties? At the very least, it must be assumed that the primary acoustic properties associated with different distinctive features are not independent of each other but overlap in varying degrees. Another way of expressing this is to state that each acoustic property can be analyzed into a number of more detailed acoustic aspects or cues, some of which are shared by several distinctive features.

I will not attempt to discuss the specific examples provided by the authors, except for noting that they are not equally straightforward. In the first two examples, in particular, it is not fully clear to me whether they are cases of necessary feature redundancy due to articulatory linkages or of optional feature enhancement. This may simply reveal my ignorance, however. I would merely like to point out that, in order to claim that acoustic feature enhancement has occurred in a given language, it is necessary to show that: (1) the primary acoustic property can be found in its unenhanced state; and (2) the supposed enhancement indeed results in an acoustic change in the expected direction. It is necessary to keep in mind the possibilities that some optional feature linkages are due to articulatory facilitation that does not substantially alter the acoustic output, or that a predicted acoustic change is counteracted by some other simultaneous adjustments, so that there are no acoustic or perceptual consequences. These are precisely the questions that Stevens et al. are planning to investigate in more detail.

I would like to comment briefly on the third example. It is argued here that, in English stops, the feature [−voice], characterized by absence of low-frequency periodicity, may be enhanced by several other features, all of which extend the acoustic interval during which voicing is absent. These enhancing features include aspiration, shortening of a preceding vowel, and lengthening of the closure interval.

In addition, the authors mention the higher fundamental frequency at voicing onset. This is an acoustic property that, as far as I can see, is not associated with a particular distinctive feature (unless one invokes tones) and that does not directly enhance the primary acoustic property. Another acoustic feature not mentioned by the authors is the higher onset of $F_1$ at voicing onset for voiceless stops. Although a high $F_1$ is associated with the feature [low] for vowels, I presume the authors would consider it here as a byproduct of delayed voicing onset and not as an enhancing feature. All these acoustic properties, of course, can be shown to be perceptually salient. There may be additional examples of such satellites accompanying the primary acoustic properties that are the fixed stars of the present theory.

Stevens et al. also argue that, since the presence or absence of low-frequency periodicity is taken to be the primary acoustic correlate of the voicing feature, the enhancing feature of aspiration takes the place of the primary distinctive property in word-initial position in English. It is argued that, for phonological reasons, voicing is distinctive in English, not aspiration. (In support of this point, the authors cite the plural rule, although it applies, of course, only to word-final consonants.) For phonological theory, however, it should make no difference whether the feature that characterizes the voiced-voiceless distinction in all phonetic environments is called [voice], [aspiration], or [omega]. Clearly, so-called voiced stops all have something in common at the level of abstract description, as do so-called voiceless stops. At the acoustic level, however, there are important contextual differences, and Stevens et al. have perhaps gone a bit too far in reifying phonological nomenclature. If there must be a primary acoustic correlate, why not simply take aspiration to be that correlate of the voicing feature for stops in initial position? Apparently, such context-dependent definitions of acoustic correlates are not permitted in the present theory, which strives to maintain the strict one-to-one correspondence of distinctive features and acoustic properties across all positions and contexts. This highly restrictive and purely formal assumption is the lifeline of the present theory. A reformulation in terms of many-to-one correspondences would presumably be considered inelegant and unparsimonious by the authors, and it would reduce considerably the explanatory power of the present theory. The philosophical problem one encounters here is that whether or not something counts as an explanation depends very much on the assumptions one starts out with.

Toward the end of their paper, Stevens et al. mention several implications of their theory. One implication concerns the nonindependence of distinctive features, which is of interest primarily to phonologists. One additional possibility raised by the authors, that of devising "cover features" encompassing a variety of acoustic correlates, would be a significant step toward reconciling the theory of acoustic invariance with other theoretical views. (The next step might be to consider each acoustic property to be the cover term for an array of contextual variants.)

A second implication concerns the continuous, rather than discrete, manifestation of features in the acoustic signal. At first glance, this conclusion seems hardly novel; clearly, there are many sources of variability in the acoustic signal. However, I believe more is being implied. The theory points toward a particular source of systematic variability in the acoustic instantiation of phonetic distinctions across languages (assuming, of course, that distinctive features are universal). Some languages may make use of redundant features and thereby show stronger acoustic correlates of certain features than languages that make little use of the enhancement possibilities. This is an interesting and nontrivial implication. In other words, there may be systematic cross-linguistic variation in the idealized or prototypical acoustic representation of universal features. On the other hand, the variability around the norm that occurs as a function of context and speaking rate, is not of immediate concern to the present theory. The theory operates at a relatively abstract level even as far as the acoustic signal is concerned; hence the term "acoustic invariance" in the face of ubiquitous variability in natural speech. Stevens et al. are dealing with acoustic competence, as it were, not with acoustic performance. This is to be expected in a theory whose basic conceptual ingredients are supplied by phonological theory.



Clearly, the most interesting aspect of the theory of Stevens et al. is its potential explanatory power, even if that power is evident only within the particular framework adopted. On one hand, the theory aims to provide a systematic account of features and acoustic properties that are likely to enhance each other, based on a careful consideration of articulatory maneuvers and their acoustic consequences. On the other hand, the theory should be able to predict (or, rather, postdict) under what circumstances feature enhancement will actually occur in a given language or even for a given speaker, assuming that some specific feature redundancy is not obligatory and universal. Stevens et al. contribute the observation that features in need of enhancement are likely to be those that are difficult to perceive. This seems a reasonable proposal, although perceptual salience needs to be defined more clearly. Still, even a perceptual criterion cannot readily explain differences across languages and individual speakers. One of the most interesting tasks for the future will be the explanation of these variations, by examining the role enhancing features play in the total phonological system of a language, and—in the case of individual differences—how enhancement may be conditioned by individual variations in dialect, anatomy, and articulatory strategies.

In conclusion, it is clear that the theory of acoustic invariance in speech is alive and well. However, on closer inspection it seems that Stevens et al.'s paper presents an expanded theory that, behind a protective screen of orthodox assumptions, makes steps toward providing a systematic account of variability in the speech signal. One day, out of its chrysalis there might emerge a full-fledged theory of acoustic variability.

## AFTERTHOUGHT: INVARIANCE MAY BE A COGNITIVE ILLUSION

Here are a few very general thoughts on invariance: Our cognitive capabilities have evolved to help us deal constructively with the objects and events in our environment. Our cognitive concepts are categorical and enter our consciousness in serial order. We know a fair amount about how our thought processes work because, to the extent that they are open to introspection, they too form part of our environment. However, introspection reveals absolutely nothing about perceptual and (preparatory) motor processes, which are unconscious and are not part of the world that our minds have evolved to understand. The temptation is great to devise theories of speech perception and production whose ingredients are the conscious products of our cognitive analysis of speech and language. We are all born linguists because of our analytic abilities; however, nobody is born with the ability to understand perceptual or motor processes. These processes are likely to operate according to principles that are not immediately obvious to us and that can be understood—if that is possible at all, excluding tautology as a form of explanation—only by discovering and adopting a completely new conceptual framework. The study of complex organic systems may yield such a framework.

The process of perceptual categorization reduces physical variability to a discrete conscious percept. It is the discreteness and constancy of this percept that suggests to us that there might be invariance somewhere in the physical world. However, this may well be an illusion. Of course, to the extent that there is physical invariance, perceptual processes are simply not needed. Therefore, the search for invariance should by no means be abandoned. However, if the search is not successful, we should not be worried but rather turn our attention to models of perceptual categorization in the context of a large knowledge base.

The appropriate functional unit for that enterprise is most likely the meaningful word, and the contributions to this volume by Cohen and by Elman and McClelland (Chapters 24 and 17, respectively) point in the right direction. Underlying these theories is the assumption that our brains are powerful storage devices with the capability of retaining millions of words, including their precise acoustic specifications abstracted from the variable input, and of processing each new input in a highly efficient fashion. How such a vast lexicon is actually represented in the brain we cannot conceive at present, but note that professional musicians store potentially even larger amounts of musical information, complete with acoustic details of timbre, dynamics, and orchestration. It is from this interconnected array of partially similar lexical entries that our knowledge of linguistic units is derived in an analytic manner. *When we talk about the recognition of phonemes or features, we are using a rhetorical device for describing how much information about a word has accumulated and how that information constrains the possible choices from the lexicon.* This also applies to the nonsense material that is frequently used in our experiments and which is wordlike and perceived by analogy to the most similar words we know. We often feel the need to describe to ourselves the process of information accumulation in familiar analytic terms, and this is all right as long as we realize that the perceptual system need not describe anything to itself. Our perceptual mechanisms do not think or talk to themselves, as we do; they simply do their job. Similar arguments apply to speech production, substituting information dissemination for accumulation. Although speech errors are often cited as evidence for the existence of linguistic units in the production process, I suspect that here, too, we are merely using a familiar descriptive device for characterizing phenomena that ultimately can be predicted from a more intimate knowledge of motor planning.



To conclude, it seems to me that the problem of invariance, as juxtaposed with variability, is more apparent than real and stems in large part from our need to talk to each other about what we are investigating, and to investigate what we are talking about. I am not suggesting that we abandon all verbal communication, but that we take our intuitions with a grain of salt and begin to consider, as Elman and McClelland (Chapter 17) have begun to do, classes of models that capture some properties of our nervous system. The power of these models goes way beyond what can be expressed in words, but it can be demonstrated by mathematical arguments and computer simulation. Perhaps paradoxically, our understanding of perceptual and motor processes may be increased by resisting our urge to describe them in familiar terms.

## Gunnar Fant:
## Comment

Stevens et al.'s study is basic for the theme of our volume and represents a much needed continuation of the work of Jakobson, Fant, and Halle (1952). On the one hand, it is a stimulating contribution to phonological theory. On the other hand, it should sharpen our attention on the perceptual salience of various acoustic properties entering both redundant and distinctive features. In one sense, we are thus back to viewing phonemes and allophones and feature realizations as individuals shaped by their context specific combination of phonological genes. Does perception pick out the entire gestalt rather than each of the underlying genes? Perception could work both ways but it is hard to prove any specific model. As also pointed out by Repp, (Chapter 20a), the way we conceive of the process is already biased by our terminology and metalinguistic projections.

I have a specific comment on acoustic detail properties. The Malayalam alveolar (apical) stop contrasted with the dental (laminar) stop in Figure 20.5 does not only have a higher $F_2$ locus but also a positioning of $F_4$ close to $F_3$ typical of a small retraction of the point of articulation. This modification also occurs in Swedish when a dental is preceded by a phonological constituent [r] which is retained in the spelled form but is otherwise lost in pronunciation. In the example the alveolar stop has a component [−distributed] along the same dimension as the $F_3$ lowering due to a more retracted [−anterior] articulation. Is this an additional enhancement for the Malayalam [+distributed]? In any case, it would be interesting to have results from synthesis experiments evaluating the various components in the entire pattern contrast. How important are the dynamics of vowel onset and how important is the F-pattern versus time variation? Both are, of course, conditioned by one and the same underlying articulatory gesture.

## Patti J. Price: Comment

I'd like to make a few comments about redundancy. I will start with an example: Given two features that are not *both* needed to maintain a distinction, how is it determined which of the two is the redundant one? One could argue that if one is always there and the other is not, then the one that is always there is not redundant. I am not convinced that this situation is typical. Further, there may well be cases in which there are, for example, three features and the occurrence of any two out of the three would be sufficient to maintain the distinction. How would redundancy be described in this case?

The authors outline two types of redundancy, that arising from constraints of the human speech and auditory systems and that arising from language-specific characteristics. I maintain that there is another type of redundancy, namely that arising from our selection of sets of features used to characterize these aspects of language. Because the features are not orthogonal to each other, not all combinations occur, and redundancy is built in: Thus, the choice of the set of features determines some of the redundancies. This is a problem because we don't know or can't agree on how to choose features, i.e., on what counts as evidence for setting up a particular feature. Now, for a diverse set of reasons, people have come up with systems of features that are not orthogonal. One of the reasons that there is no good agreement on what the features are is that devising an orthogonal or elegant system for a given language is a quite different problem from devising a system of all possible features in all possible languages. For some tasks, we want features that can make distinctions as fine as the speech perception and production systems are capable of making. For other tasks we want features that are capable of uniting a great variety of different sets of sounds that are, for various reasons, linguistically "the same." What is elegant and economical for one task is not necessarily so for others. Thus, assuming that there is one set of features for these various tasks implies excess baggage—i.e., redundancy—and, as such, says more about how we are choosing to look at our data than it says about language.

A fundamental problem here is that there is no agreed upon set of criteria (if there is *any* set of criteria) for the existence of a given feature. In the old days, the minimal pair was critical in their definition. If, as the paper by Stevens et al. implies, there are no truly minimal pairs, we may be worse off than before. The distinction between distinctive and redundant features could offer some salvation, but this distinction is blurred since redundant features can become distinctive and vice versa, i.e., in language change.

The paper presents many interesting ideas, the most important of which is, I feel, its aim to capture linguistically significant generalizations about both the discrete and the continuous aspects of language. As with most interesting theories, it also leads to many new problems. This does not detract from its attractiveness.

I'd like to repeat something that was said by the structuralists: It is not the case that all things *are* structured, but we can always impose a structure on them. A variety of structures can be imposed for a variety of reasons. For example, in physics it is often useful to assume that force equals mass times acceleration ($f = ma$). This is descriptively (and otherwise) adequate for many purposes, but not all. It is inaccurate and inadequate for some purposes and the theory of Relativity was developed to deal with these problems. Inadequacy for some purposes does not preclude usefulness for other purposes. A "unified" theory is, of course, aesthetically appealing, but assuming that we have one may be procrustean. Further, the fact that many different structures are useful, for many different reasons, is an important fact about language, about language as a tool that we use as speakers, and also about language as a source of data for linguists and for speech scientists. Language is flexible enough to accommodate multiple and changing purposes.



## John J. Ohala: Comment

Stevens et al. claim that redundant features in speech are present in order to reinforce the primary distinctive features by contributing additional properties that help the listener to perceive the primary distinctions. I see at least two problems with that claim. First, in many of the cases which might be cited, it seems obvious that these redundant features are present due to purely mechanical factors, i.e., the physical constraints of the vocal tract (Ohala, 1981b, 1983b). For example, the shorter duration of voiced obstruents vis-a-vis voiceless obstruents doesn't simply result in more voicing (of the surrounding sonorants), it is also—perhaps principally—motivated by the necessity to keep the closure interval short enough so that voicing can be maintained by the most easily implemented enlargement of the oral cavity so that the accumulating air flow does not reduce the transglottal pressure drop so much that voicing would be extinguished (Ohala & Riordan, 1979). The $F_0$ perturbation on vowels after voiced and voiceless stops seems to be an automatic consequence of whatever laryngeal gestures are necessary to produce the voicing difference. The greater voice onset time (VOT) before high vowels than before low vowels seems to be due to the fact that the narrower oral constriction for close vowels creates greater resistance to air flow and thus delays the attainment of the transglottal pressure drop required for voicing. Glottalized consonants cause adjacent vowels to be tense or creaky because the constricted glottis condition is anticipated or perseverated. These redundant features *may* serve perception but their presence would seem to be most directly explained by the physical constraints of speech production.

Second, it is not clear to me how many of these redundant features can be seen as reinforcing the primary distinctive feature. How does the high $F_0$ on vowelbs following voiceless (and, as it turns out, voiced implosive) consonants reinforce the primary features of these segments? How does the greater VOT of stops before high vowels reinforce the primary distinctive feature, presumably [+high] (or [+diffuse])? Given the large number of redundant features in speech, it is not difficult to find some which seem to enhance the primary feature, e.g., nasalization of vowels before nasal consonants, but it is not obvious that Stevens et al.'s hypothesis is sufficiently general to handle all relevant cases.

## Stevens, Keyser, and Kawasaki: Response

### RESPONSE TO DR. OHALA

We agree that the linking of distinctive and redundant features often originates in the physical constraints of the vocal tract. However, the "physical" theory of redundancy alone does not explain all possible ways in which distinctive and redundant features interact.

The shortening of voiced obstruents does not seem to result from absolute physical necessity, for these consonants could alternatively be pronounced as partially voiced without any change in their closure duration. We hypothesize that the shortening of voiced obstruents is motivated by a perceptual factor, and therefore is employed for the purpose of enhancement. Voiceless stops typically have stronger bursts than voiced ones, and this difference in burst amplitude may serve as one of the cues to the voicing distinction. It can be argued that a strong burst interrupts the representation of periodicity in the auditory system, and thus itself serves as an enhancing property for [−voiced]. A weaker burst, on the other hand, may not interrupt this auditory representation, allowing the percept of periodicity, i.e., voicing. (It is necessary, then, to postulate a feature that has an acoustic correlate a strong burst or a strong onset—possibly the same feature as that discussed in the paper in connection with Figure 20.5.) Electro-physiological studies of the first-order peripheral auditory system suggest that the response of the auditory neurons to an abrupt rise in amplitude is greater in magnitude when such a rise is preceded by a longer silent interval (Delgutte 1982). (A discussion of the consequences of this response characteristic in modelling the auditory processing of speech is given by Goldhor (1983a).) Thus, the shorter is a stop, the less would be the auditory system's response to its burst. In the case of fricatives, a short interval may reduce the percept of noise, thereby enhancing the perceived periodicity. The shortening of voiced obstruents (as well as lengthening of voiceless ones) would therefore maximize the auditory difference between the voiced and voiceless cognates.

As Dr. Ohala stated, a certain physiological constraint seems responsible for the $F_0$ variations induced by consonantal voicing or voicelessness. And yet this cause-effect relationship alone does not tell us why, for example, the $F_0$ perturbations last as long as 100 ms after vowel onset in English but last for a considerably shorter interval in such tone languages as Yoruba and Thai and such a pitch-accent language as Japanese (Gandour 1974, Hombert 1978, Kawasaki 1983). Such a cross-linguistic observation suggests that the redundant systematic $F_0$ variation is exploited to a greater extent where the introduction of such redundancy will not disturb the properties of other phonemic distinctions. We speculated in our paper that one of the redundant features reinforcing the distinctive feature [−voice] is [+stiff vocal cords]. This feature [+stiff vocal cords] inhibits vocal-fold vibration during the consonantal interval, and has the acoustic correlate of high $F_0$, which is observed at voicing onset. In this sense, we claim that an $F_0$ perturbation is involved in enhancing the voicing feature, though the relationship between the two is an indirect one. Though Dr. Ohala considers it a mere aerodynamic consequence, the affinity between [−voice] and [+high] in vowels, witnessed in such phenomena as longer VOT's of voiceless stops before high vowels and greater tendency of high vowels toward devoicing, seems to be another kind of enhancement that languages have freedom to implement to varying degrees. The following observations support our claim. First, only a small subset of languages in the world have the processes of high vowel devoicing. Second, EMG and fiberscopic studies have shown that the laryngeal adjustments for devoiced vowels are very different from those for voiced vowels; the former not only lack and adductive movement of the vocal folds but also show active abductive gestures (Hirose 1971). Thus the devoicing in these cases appears to be an active process under control of the speaker. Dr. Ohala raises a valid point when he notes that there seems, at first glance, to be little basis for an argument that [−voice] causes an enhancement of the acoustic correlate of the feature [+high]. A possible argument might be something like the following. The feature [+high] requires that $F_1$ be low, and this feature is enhanced if the frequency of $F_1$ is made as low as possible. That is, $F_1$ is sufficiently low that it has no prominent character that is distinguishable from the source. Devoicing essentially eliminates $F_1$, creating a perceptual consequence similar to a lowered $F_1$.



We can cite numerous cases where an enhancing feature is not an automatic consequence of the production of a distinctive feature. Examples are: rounding of [š] in English, rounding of back vowels, diphthongization of high vowels, stridency in coronal consonants, and stridency in continuants. In these cases the redundant or enhancing feature is imposed strictly because it strengthens the acoustic property of the distinctive feature. There is little linkage between the two in their articulatory manifestations.

In short, our counterarguments for a purely physical account of redundancy are: (1) not all redundant features arise due to physical constraints, and (2) physically-induced redundancy can be optionally further exploited for the purpose of enhancement.

### RESPONSE TO DR. REPP AND DR. PRICE

We thank Dr. Repp for his insightful and helpful comments. As a consequence of remarks by Dr. Repp and others, we have recognized some inadequacies in our exposition and in some of our arguments, and we have made some revisions in our manuscript. Thus some of the points made by Dr. Repp may already be answered in the paper.

Near the beginning of his comments, Dr. Repp notes that several hypotheses concerning phonetic features, acoustic properties, and invariance form the starting points for the discussion in our paper. He points out that it may be difficult to provide justification for each of these hypotheses when they are examined individually. However, these hypotheses are not independent; any one of them would be difficult to support unless it were made in the context of the others.

In effect, we are choosing to *define* phonetic features as being characterized by invariant acoustic and articulatory attributes. We will not accept a phonetic feature in the inventory of universal features unless we can define an acoustic and an articulatory property for the feature. This requirement is different from a definition of a feature that allows multiple properties to occur when the feature is actualized. For example, according to some definitions, the feature [+voice] can have a variety of acoustic correlates. In our proposed formulation, we account for this apparent variability by postulating that selected additional features can come into play to enhance or even replace the original voicing feature. Each of these additional features has an invariant acoustic property, and in some cases this property can be present with various degrees of strength.

Thus we are proposing a theoretical framework in which there may be variability in the selection of features to be invoked (each with an invariant acoustic correlate) rather than postulating a one-to-many relation between features and acoustic properties. The features that can be utilized, whether in a distinctive or in a redundant role, are all drawn from the same inventory, and all have the potential of serving in a distinctive role in some language.

The question then arises (and Dr. Repp indeed poses this question) as to whether every auditory-acoustic property that is a potential cue for a distinction can be associated with a feature. That is, can this property be associated with an item on the list of features than can function distinctively in language? A complete answer to this question will require a careful and systematic review, across a variety of languages, of the inventory of features and of auditory-acoustic cues that have been proposed by various researchers. An initial review suggests that the list of acoustic cues arising from perceptual experiments can in almost all cases be interpreted as the acoustic correlates of particular features. This finding is not unexpected if we postulate that the inventory of features is intimately related to, and in fact defined in terms of, the capability of the auditory system to respond distinctively to particular acoustic properties and the capability of the articulatory system to be controlled in such a way as to produce these properties independent of the context.

The hypotheses that form the starting point for our discussion can be summarized, then, as follows. We define the features and their acoustic correlates in such a way that invariance in the implementation of a feature is guaranteed. We propose that variability arises in the features that are selected (largely on the basis of rules) to operate in particular utterances, and not in the relation between features and acoustic properties. With appropriate selection of auditorily-based acoustic properties, we believe that the number of features needed to perform these functions is not large—probably about twenty.

We turn next to comment on the question raised by both Dr. Price and Dr. Repp concerning the selection of which features are to be identified as distinctive and which are redundant in a particular situation. In order to focus the discussion of this question, we shall consider two examples.

One example comes from the the language Yawelmani (Archangeli, 1984), which has four vowels. Examination of the articulatory and perceptual characteristics of these vowels indicates that they can be described in terms of the four major vowel features as follows:



| | i | u | o | a |
|---|---|---|---|---|
| high | + | + | − | − |
| round | − | + | + | − |
| back | − | + | + | + |
| low | − | − | − | + |

This description is, however, an overspecification, i.e., there is redundancy in this classification of the vowels. In principle, it is possible to distinguish among four vowels using only two features. A similar overspecification in terms of features occurs for three coronal fricatives in English. The fricatives [s] and [š] are classified as [+strident] and [θ] as [−strident], while the feature [anterior] distinguishes [s] and [θ] from [š]. Acoustic evidence also suggests that [θ] is produced with a backed tongue body position, whereas [s] and [s] can be considered as [−back]. Thus we have the following feature representation for these fricatives

| | θ | s | š |
|---|---|---|---|
| strident | − | + | + |
| anterior | + | + | − |
| back | + | − | − |

Again we have redundancy in the representation of these segments in terms of features.

In situations like this, where it appears that a distinction is carried by more than one feature, is there a basis for arguing that a minimal subset of the features is primary or distinctive, while the remaining features are redundant? Examination of the properties of the acoustic signal does not provide a basis for selecting which of the features form the minimal or distinctive set. Each feature is represented by a property in the sound wave, and although some properties may be present with greater strength or greater reliability, there is often sufficient variability that no one of these properties could be selected as representing the primary or most reliable feature based simply on observation of the sound.

We propose, however, that in situations where more than one feature is involved in making a distinction, such as the Yawelmani or English examples just given, some of the features should be considered as distinctive features and the other features are derived from the distinctive features by a set of redundancy rules. We suggest that a native speaker of a language has access to a representation of words in the lexicon in terms of these underlying distinctive features. A full phonetic representation of the words that includes all of the features may also exist, but the additional features are derivable from the distinctive features through the redundancy rules.

The process of determining which features to label as the distinctive features follows several guidelines:

(1) The distinctive features that are selected should be a minimal set. That is, the underlying feature representation should be maximally simple and should not overspecify the segments in the language.

(2) The rules that are used to derive the redundant features should be universal in character. These rules should be maximally simple, and they should be motivated by relationships among the features that are based on universal physiological and perceptual principles. One of these principles is that a system of oppositions for a particular language evolves in such a way that there is maximum perceptual distinctiveness between contrasting utterances (similar to the ideas proposed by Liljencrants and Lindblom, 1972). Thus, for example, if a language has only five vowels, the vowels will tend to distribute themselves in such a way as to be maximally distinctive.

The universal rules in (2) have not yet been worked out in detail, and the notion of simplicity needs to be quantified more precisely. (See Archangeli, 1984, for a recent proposal in this regard.) Consequently we cannot at this time provide a complete account of how a set of distinctive features is selected.

The nature of the arguments can be illustrated, however, with reference to the Yawelmani vowel example just given. The procedure we shall follow is to *postulate* a set of distinctive features, and then show that, with these features as a starting point, a simple and well-motivated set of rules can be written for deriving the redundant features. It can be argued that if an alternative initial set of distinctive features were selected, the redundancy rules that would be necessary would be both less simple and less well motivated in terms of universal principles.

We begin with a description of the Yawelmani vowel system in terms of the distinctive features marked in the following table:

| | i | u | o | *a* |
|---|---|---|---|---|
| high | | | – | – |
| round | | + | + | |
| back | | | | |
| low | | | | |

The features that are not specified in this table are redundant and are specified by a set of rules.

One type of redundancy rule simply fills in the complements of the features that are marked in the underlying representation. In this example, two such rules are needed:

1. [ ] → [+high]

2. [ ] → [−round]

We now need some rules that will fill in the values of the features [back] and [low]. These rules, we suggest, will follow certain general principles related to auditory perception and to articulatory and acoustic phonetics. One of these principles is that the nonhigh vowel in the vowel triangle that is perceptually the most distinct from all other vowels is the vowel labeled as [+low, +back, −round], that is, the vowel [*a*]. This vowel has the highest possible first formant frequency, and this attribute separates it maximally from other vowels. If this [+low] vowel were [+round], or if it were [−back], the first formant frequency would tend to be lower,



and hence the perceptual distance of the vowel from other vowels would be less. Thus there are perceptual grounds for postulating that a vowel that is [−high, −round] is maximally distinct from other vowels if it also has the features [+low, +back]. At the same time, vowels that are *not* [−high, -round] should not be [+low], since the various [+low] vowels are perceptually rather close together. For example, the first and second formant frequencies for the three low vowels [*a*], [ɔ], and [æ] are relatively close together, and, in a sparse vowel system this clustering of low vowels is to be avoided. These notions of perceptual distinctiveness related to the low vowels can be captured by two ordered rules:

3. $\begin{bmatrix} -\text{high} \\ -\text{round} \end{bmatrix} \longrightarrow \begin{bmatrix} +\text{low} \\ +\text{back} \end{bmatrix}$

4. [ ] → [−low]

Finally it is necessary to fill in values for the feature [back] for the remaining vowels. Here we draw on a relation between the features [back] and [round] that is discussed in the paper we have prepared for this symposium. In the paper it is shown that the acoustic property that indicates backness for a nonlow vowel (proximity of $F_2$ and $F_1$) is enhanced if the vowel is rounded. In a similar fashion one might argue that the acoustic property that indicates rounding (tentatively identified as a greater degree of spectral prominence provided by a formant pair) can be most strongly represented in the sound if the vowel is [+back]. That is, rounding can be signaled in back vowels more effectively than it can in front vowels. This relation between rounding and backness is captured by a fifth redundancy rule, which fills in the unspecified values for the backness feature:

5. [round] → [back].

Application of all these rules in order leads to the completely specified matrix of features given in the table above.

Except for the first two rules, which fill in complement values for the underlying features, these rules can be motivated by relations that exist among the four features [high], [round], [back], and [low]. Given a set of underlying features that have been specified, the rules indicate how other features should be marked in order to enhance or to highlight the acoustic and perceptual representation of the underlying features. Attempts to find an alternative solution to the derivation of the Yawelmani vowel system, using a different underlying representation and/or a different set of redundancy rules, lead to a more complex specification, and the rules that are needed to derive the complete feature matrix are less well motivated in terms of natural relations among the features.

This example, then, illustrates the sense in which, in a given language, we would like to specify some features as distinctive and others as redundant.