ENHANCEMENT AND OVERLAP IN THE SPEECH CHAIN

SAMUEL JAY KEYSER
Massachusetts Institute of Technology

KENNETH NOBLE STEVENS
Massachusetts Institute of Technology

A model of speech production is proposed in which the input is a planning stage at which lexical items are arrayed, accompanied by the full panoply of phonological representations from distinctive features to their attendant tree structures. A set of instructions for control of the vocal tract is calculated leading to a sound output. Two parallel processes are involved in the calculation of these instructions, both of which replace the planning-stage representation by the appropriate motoric instructions. One of these processes is universal and involves replacing each distinctive feature with an appropriate set of motoric instructions, either unmodified or modified by the process of overlap. We postulate a parallel language-specific process that is sensitive to those features in danger of losing their perceptual saliency as a consequence of the environment in which they appear. This process, referred to as ENHANCEMENT, adds additional motoric instructions to enhance the saliency of the jeopardized features. We provide a number of examples to illustrate how enhancement works. We conclude from these examples that whereas defining gestures related to distinctive features are, in many instances, weakened or even absented from the speech stream, enhancement gestures, once added to the set of motoric instructions, appear never to be subject to obliteration by overlap.*

1. INTRODUCTION: GENERAL CONSIDERATIONS. We assume as a starting point the correctness of distinctive feature theory. This statement may to many seem on a par with a theoretical physicist claiming to have made his/her peace with, say, quantum mechanics. That said, there are many working in the field of phonetics for whom distinctive feature theory is by no means so firmly entrenched. Such researchers are not swayed by results of phonology such as accounts that distinctive feature theory makes possible for omnipresent phenomena such as diachronic sound change, or the morphological variation one finds in such systems as the English plural. Rather they are swayed by the tremendous amount of variation in the instantiation of phonemes in a given language, for example, the multitude of realizations of English /t/.

This skepticism is discounted in phonological circles because of the tremendous gains in explanation offered by distinctive feature theory. But that in and of itself does not make the reasons for the skepticism disappear. There is, indeed, substantial variation in the realization of distinctive feature configurations, and it is an important question for those working at the interface of phonology and phonetics to try to come to grips with it. The purpose of this article is to address the question of variability head-on while maintaining the integrity of distinctive features as a cornerstone advance in our understanding of human language.

In what follows we develop the theory of enhancement presented in earlier manifestations (e.g. Stevens et al. 1986, Stevens & Keyser 1989, Keyser & Stevens 2001). Our strategy involves preserving the character of a given segment's distinctive feature representation as suggested by phonological considerations in the face of phonetic properties that seem completely unrelated. A paradigm example is the rounding that accompanies the production of /ʃ/ in English as opposed to the lack of rounding in English /s/. There is no reason to suppose that rounding is a distinctive part of the representation of the segment /ʃ/. But this is true only from a phonological point of

* The authors wish to acknowledge the late Ken Hale for his ready advice and constant support of this research. We would also like to thank the editors and referees of *Language* for their extensive and insightful criticisms of earlier versions of this manuscript. Finally, we would like to single out Keith Johnson, whose comments have been of great value to us as we came to understand the place of this research in ongoing work on the interface between phonetics and phonology.

view. There is no contrast between a rounded and an unrounded /ʃ/ or a rounded and an unrounded /s/. For the phonologist, then, rounding is extraneous. But to the phonetician it is important. In English no description of /ʃ/ that does not include rounding can be said to be complete.

In the face of this, some phoneticians might take the position, not unreasonably, that however one describes the segment, rounding must be a part of that description. From our perspective we are forced to ask the question: if rounding is not distinctive for /ʃ/ and /s/ in English, why is it there at all? This question makes sense only if one is willing to honor the intentions of both phonologists and phoneticians. In what follows this is what we try to do.

2. THE PLANNING STAGE. There is widespread agreement (e.g. Ladefoged & Maddieson 1996, Johnson 2003) that speech production begins in discrete representation and ends in continuous sound. This shift takes place in the vocal tract and its attendant musculature, the arena where representation becomes sound-producing gesture. Underlying this view is the assumption that representation and sound are fundamentally different in character. The former is digital in that it is composed of a variety of discrete entities: for example, features, syllables, feet. The latter, however, is primarily—though not exclusively—a continuous sound stream generated by a flow of gestures in the vocal apparatus from one intention to the next. This dichotomy between a discrete representation and an analogue output device is represented schematically in Figure 1 with the box labeled GESTURE CALCULATIONS mediating between the two.

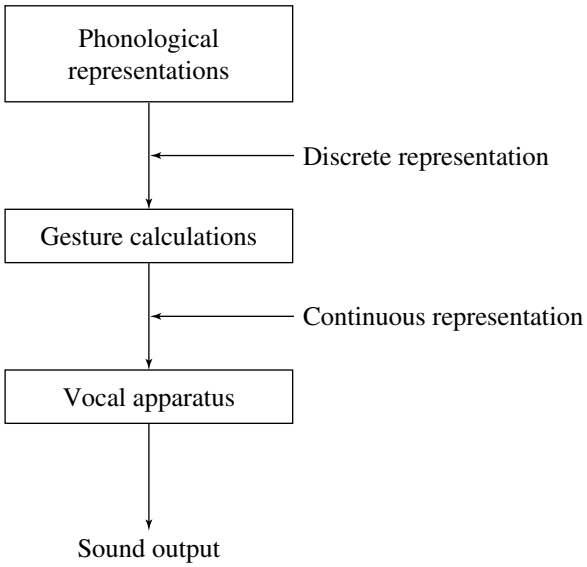

FIGURE 1. Basic components along the path from discrete phonological representations to sound output.

The existence of a gestural stage has led some researchers to conclude that lexical and gestural representation should be collapsed into a single stage. Were it true, this conclusion would greatly simplify the system. That is, instead of going from a distinctive feature representation and its attendant representations (foot structure, syllable structure, etc.) to gestures to sound, one would go directly from gesture instructions to sound (see Browman & Goldstein 1986).

Phonologists have, in general, rejected the view that gestures are lexical, the main reason being that phonological representation in the form of distinctive features allows

for the expression of a vast number of what appear to be true generalizations about language, for example, phenomena of vowel harmony, vowel lengthening and shortening in specific phonological environments, the Great Vowel Shift in English, stress placement, and so forth.

While there is some truth in the gestural notion of representation, it can be argued that gestural representations should not be localized in the lexicon, but must occur at a later stage in the speech process. One type of evidence for this view comes from a consideration of speech-error phenomena. These phenomena argue for the existence of something called a PLANNING STAGE in the production of speech utterances. We suggest that gestures must occur after this stage (see §4.2 below). Here we focus on arguments for the existence of the planning stage itself.

Consider two typical speech errors drawn from Fromkin (1973:245–46). Suppose a speaker wants to produce a form like *wind mill* but, in fact, produces the utterance [mɪnd wɪl] by transposing the onsets of each member of the compound. The erroneous [mɪnd wɪl] does not occur as a separate entry in the lexicon nor is it the result of phonological rules. Similarly, Fromkin reports the phrase *hash or grass* becoming the erroneous *hass or grash* as a result of final consonant transposition. Here, too, the resultant utterance contains nonsense words which could not possibly have been drawn from the lexicon. Speech-error theorists have postulated that these errors must occur outside the lexicon and the phonology but prior to articulation. They propose a level that we call the planning stage. Errors occur here (see Shattuck-Hufnagel 1986, Levelt et al. 1999).

We assume the planning stage is a level at which discrete items drawn from the lexicon and modified by the phonological component are arrayed in a serial fashion. In Figure 2 we expand on Fig. 1 by inserting the planning stage after phonological

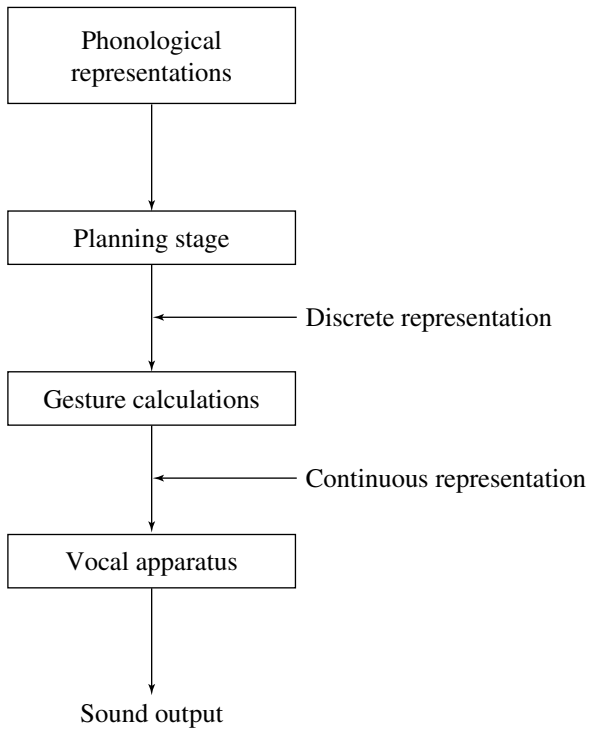

FIGURE 2. A planning stage is added to the basic components in Fig. 1.

representations. Representations in the planning stage are fully articulated. That is to say, at this stage we find the full panoply of phonological representation.

3. GESTURE CALCULATIONS.

3.1. THE BASIC FUNCTION. In the simplest case, representations in the planning stage are read off by a component whose function it is to replace representations by gestures. We suppose that the component that the planning-stage string feeds into is one that calculates how the various articulators in the vocal apparatus should be controlled, including the relative timing of the movements of these articulators.

To see how this component works, let us imagine that the planning stage contains the sequence *the tag* (see the spectrogram in Figure 3a). We focus on the gestures that are needed to produce the consonant /t/, which has the features [+ anterior, + consonantal, − continuant, − sonorant, + stiff vocal folds]. We note also that this segment is in syllable-initial position, which coincides with word-initial position, and that the entire word is dominated by a single syllable. The gesture-calculations component needs to turn this representation into a series of instructions to the musculature of the vocal apparatus which produces the relevant acoustic event. Thus, this component encounters the feature [+ anterior] and translates it into an instruction to position the tongue blade anterior to the alveolar ridge. The features [+ consonantal, − continuant] require that this articulator make a complete closure in the midsagittal region of the vocal tract, followed by a release. An acoustic consequence is a noise burst with spectrum energy at high frequencies, as the spectrogram shows. During the time that this closure is made, the pressure should build up behind the constriction, as specified by the feature [− sonorant]. The feature [+ stiff vocal folds] requires that the tension in the vocal folds be increased during the closure interval and be relaxed following the release of the consonant. As a consequence of these instructions, the vocal apparatus produces an acoustic pattern that reflects the set of features for this segment. This acoustic pattern extends from the closure through the release of the alveolar consonant. We see later that these instructions are modified by additional calculations that are not derived directly from the planning stage.

Note that while the input to the gesture-calculations component is a phonological representation, the output is not. Rather, the output is a series of instructions to the musculature. This entails that the phonological representation disappears at this point, being replaced by motor instructions. Hence, if the birthplace of lexical representation is in the lexicon, its demise is in the gesture-calculations component.

3.2. VARIABILITY IN GESTURE SEQUENCES. Consider now the difference in *the tag* and *top tag*, focusing on the gestures relevant to /t/ in *tag*. While the gestures for /t/ in both expressions are identical, the output is not. In particular, the tongue-blade closure for the second occurrence of /t/ in *top tag* usually occurs during the time when the lips are closed for the preceding /p/. Thus, unlike *the tag*, the time of closure for /t/ is not registered in the sound. The absence of evidence for the /t/ closure is illustrated in the spectogram in Figure 3b. Likewise, the release for /p/ occurs while the tongue blade is in the closed position, and consequently this release is either not registered or only weakly registered in the sound. However, acoustic evidence for the tongue-blade feature is still present in the release of the /t/, just as acoustic evidence for the labial is still present in the closure of the /p/. This masking of acoustic evidence is the result of overlap of gestures for two different articulators (Browman & Goldstein 1990). In the case of *top tag* the closure of the /p/ masks the closing gesture of the /t/ and the closure for the /t/ masks the release of the /p/. Of course, in careful speech there can be an

acoustic record of the /p/ release and the /t/ closure. Such an utterance is illustrated in the spectrogram in Figure 3c. Such overlap is a universal phenomenon and reflects a universal tendency to conserve time and, presumably, energy (see Lindblom 1990).

Overlap pervades the speech stream. Consider the /nt/ sequence in *teen tag*. For both segments the instruction from the planning stage is to make a closure of the tongue blade and then to release it. In this case, however, the closure is retained through both segments and there is, in effect, no separate release for /n/ and no separate closure for /t/. In the phrase *teen bag* (see the spectrogram in Figure 3d) the closure for the labial /b/ would normally occur during the alveolar closure for /n/. Soon after this labial closure the soft palate is raised and pressure is built up in the oral cavity. Thus we have an example not only of overlap of a labial and alveolar closure for the consonant sequence /nb/, but also overlap of the velopharyngeal opening into the labial-closure gesture. That is, the soft palate is open part of the time while the lips are closed. This is the sequence of events illustrated in Fig. 3d. In an extreme case the labial closure overlaps the alveolar closure completely, and the speaker may omit the tongue-blade gesture altogether (see Zsiga 1994). When that happens, there is no direct evidence for tongue-tip closure. Even so, speakers have no difficulty recognizing the occurrence of an /n/ preceding the /b/ in *teen bag*. How does this happen? The answer to this question involves a general strategy of speech production, of which *teen bag* is merely an example. We turn to the general strategy in §§4, 5, and 6.

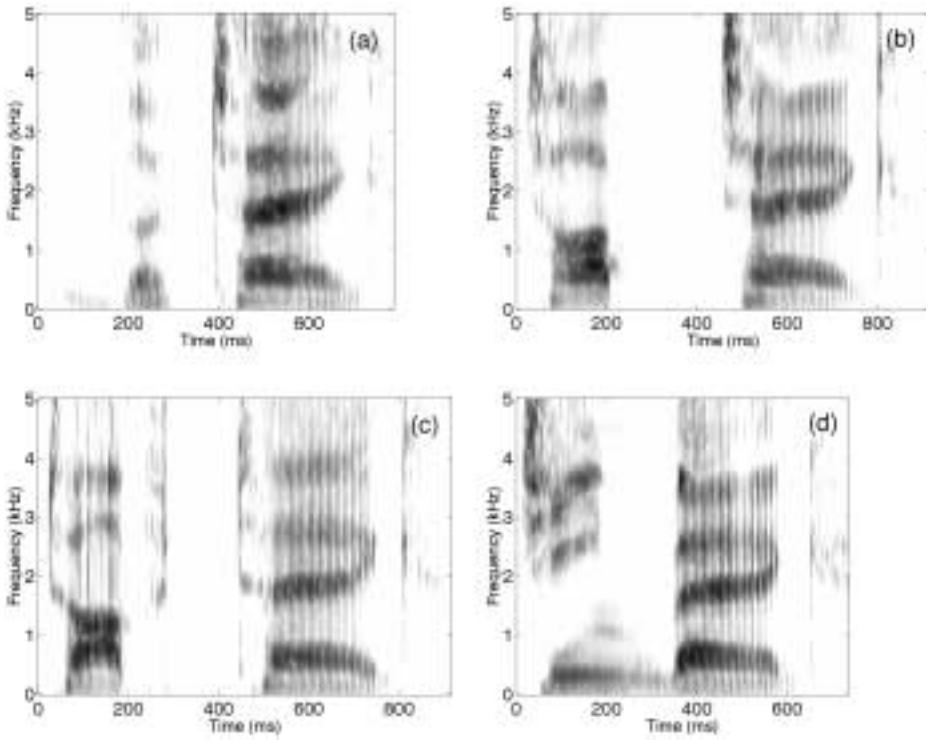

FIGURE 3. Spectrograms of utterances illustrating gestural overlap: (a) *the tag*, in which both closure and release for /t/ are represented in the sound; (b) *top tag*, in which there is no acoustic evidence for the /p/ release and /t/ closure; (c) *top tag*, in which there are acoustic traces of the /p/ release and /t/ closure; (d) *teen bag*, in which the labial closure is formed during the time when there is an alveolar closure for /n/.

4. ENHANCEMENT.

4.1. OVERVIEW. In any given language we assume that there is a set of distinctive features that is selected from a universal set of such features. Associated with each distinctive feature there is a basic gesture together with an acoustic pattern that defines that feature (see Stevens 1989). The basic, or defining, gesture for a feature can be grounded in a configurational aspect of an articulator (e.g. the feature [+back]) or in a somatosensory aspect (e.g. the feature [−sonorant]) or both. In either case the gesture gives rise to a distinctive auditory pattern. In a given language the sound output generated by these feature-defining gestures in accordance with Fig. 2 may lack saliency. In these cases the perceptual saliency may be enhanced by introducing gestures secondary to the feature-defining gestures (see Diehl 1991, Kingston 1992, Kingston & Diehl 1995). For example, in a language like English where the feature [anterior] defines a contrast between /s/ and /ʃ/, speakers learn to modify the universal gesture associated with /ʃ/ by adding lip rounding in order to enhance the distinction. The universal acoustic correlate of a [−anterior] fricative appears to be excitation of the third formant, which is basically a front-cavity resonance of the constricted vocal tract (Stevens 1989, 1998). The rounding gesture helps to 'tune' the front-cavity shape to accentuate the spectrum prominence in this frequency region.

Similarly, in a language like Spanish, the nonlow back vowels are always rounded. This rounding can be viewed as an enhancement of the nonlow back defining gestures (see Stevens et al. 1986). We take the addition of gestures like rounding to the universal gesture associated with the performance of [−anterior] in English stridents or [−low, +back] in Spanish vowels to be a common process, something that occurs as a normal part of speech production. To accommodate this enhancement we introduce new components into Fig. 2, as in Figure 4.

4.2. HOW ENHANCEMENT WORKS. Up to this point we have focused on the universal aspects of speech production represented in Fig. 2 whereby selected lexical entries make their way from the lexicon through the planning-stage and gesture-calculations components to a sound output from the vocal apparatus. We now propose that, parallel to this universal process, there is one based on language-specific properties whose sole purpose is to enhance the output of the universal process at points where that output is perceptually lacking in saliency.

For example, with respect to the rounding of /ʃ/ discussed above, we assume that enhancement adds rounding to its basic tongue-blade gesture independent of the environment in which it occurs (see §5.3 for a more detailed discussion). The model in Fig. 4 attempts to capture this modification in the following way. The box labeled ENHANCEMENT scans the planning stage, notes that the feature [anterior]—which is distinctive only for strident fricative consonants—has limited perceptual saliency, and flags this feature wherever it occurs in the planning stage. When the representation leaves the planning stage, it carries this flag into the component labeled gesture calculations. The gesture calculations for the nonflagged features proceed as before. In the case of the ones that are flagged by the enhancement box, additional gestures are added, for example, rounding in the case of /ʃ/. This is the function of the ENHANCEMENT GESTURE-CALCULATIONS component. In the following sections we provide a number of examples of enhancement phenomena from English and other languages. In these examples, a wide range of gestures is added to the universally defined set that interprets—according to universal principles—the feature representations of a particular language. (We return in §7 to a more detailed discussion of the relationship between enhancement gesture calculations, universally derived gesture calculations, and the vocal apparatus.)

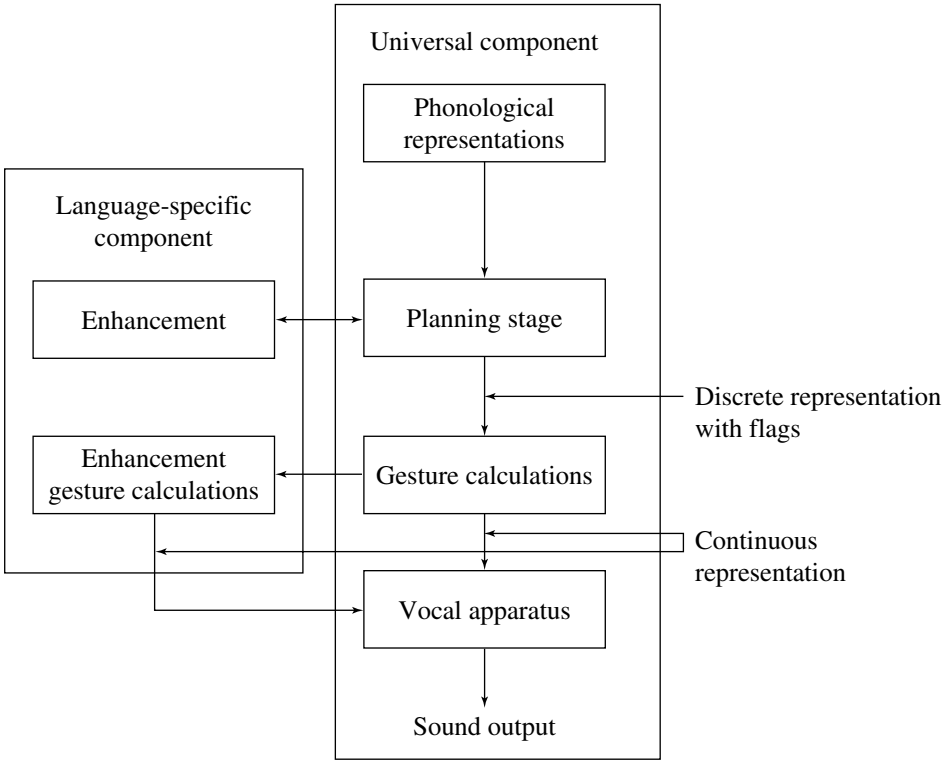

FIGURE 4. Enhancement and enhancement gesture calculations are added to Fig. 2. The enhancement component augments planning-stage representations by flagging relevant features. Enhancement gestures are calculated for these flagged features.

We mentioned above that strong evidence for the existence of a planning stage comes from speech errors where forms that do not actually exist in the lexicon can be erroneously generated, for example, [ni:kwi:d] for *weak-kneed*. To the best of our knowledge there are no examples of speech errors that involve solely enhancement gestures. Thus, in a word like *sunshine*, where the /ʃ/ in the second syllable is rounded, a planning-stage speech error might occur in which the word-initial /s/ and the syllable-initial /ʃ/ become interchanged. In our model, enhancement gestures would be applicable to this modified planning stage but only at the enhancement gesture-calculations stage (see Fig. 4), ultimately yielding the assignment of an enhancing gesture of rounding to the new initial /ʃ/. We know of no speech errors in which the rounding of the /ʃ/ appears on the initial /s/ while the unrounding of the initial /s/ appears on the /ʃ/ segment. This is a significant fact since it suggests that enhancement gestures are not featurally bound. That is to say, even though English /ʃ/ and, say, English /u/ are both rounded, the source of rounding of the former is not featural whereas that of the latter is.[1] While

[1] Pouplier (2003) has examined tongue-body movements for speech errors induced by rapid repetition of contrasting sequences of monosyllabic words. Her data show that some errors display tongue movements that are not identical to the movements one would expect if a true substitution error had been made. That is, the error is made in a way that is intermediate between the intended segment and a true substitution. Whether or not these articulatory representations of errors contain enhancement gestures for the target segment is indeterminate from Pouplier's data.

there are studies that seem to suggest that enhancement gestures may participate in speech errors (see Frisch & Wright 2002), we have taken the view that studies based on stressing the motoric system (e.g. eliciting data through tongue-twisters) do not offer a window on naturally occuring errors and are therefore not directly relevant. Additional studies of the articulation and acoustics of more naturally occurring speech errors are needed to resolve this question.

If a gestural theory makes no distinction between basic gestures and enhancement gestures, then one ought to find enhancement gestures playing a role in speech errors. One can imagine complicating a gestural theory of representation to separate out the two kinds of gestures, basic and enhancing. However, since we believe that at bottom a gestural theory cannot capture phonological generalizations, we do not pursue this alternative.

This brings us to a second point about enhancement, namely, that its implementation appears to be graded. For example, the degree of rounding of /ʃ/ is more variable than that in a featurally initiated rounding such as that in /u/. The phonological input is discrete and quantal in nature whereas the enhancement gestures tend to be nondiscrete and continuous. This is consistent with a different origin for each in the overall system, something that the diagram in Fig. 4 attempts to capture.

The investigation that follows shows that the acoustic properties that result from a feature-defining gesture can be enhanced or strengthened by recruiting another gesture. This second gesture is manipulated in combination with the feature-defining gesture to strengthen the relevant acoustic property or to add a new one. While there is some degree of orthogonality between movements of particular articulators and the resulting acoustic properties, there are also interactions between articulators in generating certain acoustic attributes. Thus the introduction of an enhancing gesture can not only strengthen the feature-defining acoustic correlate, but can also introduce new acoustic properties that can serve as cues for the defining feature. For example, a glottal-constricting gesture is often utilized in certain contexts to inhibit vocal-fold vibration, but it adds the acoustic attributes of glottalization as well. These attributes then serve to help identify the voiceless segment that engendered them.

5. SOME EXAMPLES OF ENHANCEMENT. Our purpose in this section is to provide a fuller account of the variety of enhancement gestures. Each example involves (a) a feature-defining articulatory gesture associated with the distinctive feature, (b) an acoustic output associated with that gesture, and (c) a mechanism for recruiting other articulatory gestures to enhance (b). We discuss five types of enhancement: (i) voicing for obstruent consonants, (ii) nasalization for vowels, (iii) place features for tongue-blade consonants, (iv) tongue-body features for vowels, and (v) stridency for tongue-blade consonants. In this survey not all examples are treated at the same level of detail.

5.1. VOICING FOR OBSTRUENT CONSONANTS. We consider first the generation of the sound source at the larynx, and we examine the conditions under which vibration of the vocal folds is facilitated or inhibited during obstruent consonants. When an obstruent consonant is produced, there is a reduced pressure drop across the glottis, and continued vibration of the vocal folds is in jeopardy. Maintenance of vocal-fold vibration during an obstruent consonant can be facilitated in several ways: (i) by slackening the vocal folds (Halle & Stevens 1971, Stevens 1977), (ii) by decreasing the stiffness of the walls of the supraglottal airway (Svirsky et al. 1997), (iii) by actively expanding the supraglottal volume during the consonantal interval, either through advancing the tongue root or by lowering the larynx, or both (Bell-Berti 1975, Westbury 1983), and

(iv) by opening the velopharyngeal port for some period of time during the consonant closure. Likewise, inhibition of vocal-fold vibration is facilitated in the opposite way: stiffening the vocal folds, stiffening the vocal-tract walls, actively preventing the supraglottal volume from expanding, and maintaining a closed velopharyngeal port. In addition, spreading or constricting the glottis relative to its normal position for phonation makes it more difficult for the vocal folds to vibrate.

It has been proposed that in English the feature-defining articulatory gesture that contributes to inhibition of glottal vibration is stiffening of the vocal folds. This feature is called [+ stiff vocal folds] (Halle & Stevens 1971).[2] For voiced obstruent consonants the feature is [− stiff vocal folds]. These features can be enhanced in various ways. For example, for [− stiff vocal folds], lowering the larynx can be considered an enhancing gesture. It has been pointed out by Honda, Hirai, and Kusakawa (1993) that there is a convex curvature of the spine, so that a change in the tilt of the cricoid cartilage in relation to the thyroid cartilage occurs as the larynx is lowered. As the cricoid cartilage moves downward, there is an increase in the angle between the cricoid and thyroid cartilages, leading to a shortening of the vocal folds and a decrease in their stiffness. Lowering the larynx is accompanied by a general slackening of the walls of the vocal tract. The downward movement of the larynx, then, creates two of the conditions that facilitate vocal-fold vibration for obstruents: decreasing the stiffness of the vocal-tract walls and increasing the supraglottal volume. This latter gesture can be achieved by advancing the tongue root during the consonantal interval. Enhancing the feature [+ stiff vocal folds] can be achieved by raising the larynx. This gesture stiffens the vocal folds, decreases the vocal-tract volume, increases the stiffness of the vocal-tract walls, and therefore inhibits maintenance of vocal-fold vibration. As noted above, this reduction in glottal vibration can be enhanced by spreading the glottis. Glottal vibration is also inhibited by constricting the glottis.

Raising or lowering the larynx, thereby increasing or decreasing the vocal-fold stiffness, has the effect of increasing or decreasing the fundamental frequency if these gestures are made during sonorant intervals when there is no increase in the supraglottal pressure. The effect of these movements on the fundamental frequency in vowels adjacent to voiced and voiceless consonants is well known (House & Fairbanks 1953, Erickson 1993).

The laryngeal and other adjustments for obstruent consonants to facilitate or to inhibit vocal-fold vibration apply equally to both [− continuant] and [+ continuant] consonants, that is, to both stops and fricatives. In the case of fricatives, however, there is a requirement of continued airflow through the supraglottal constriction, and consequently the default configuration for the glottis is somewhat spread.[3] In the case of a voiceless fricative, the spreading occurs over the length of the glottis. For a voiced fricative, the membranous part of the vocal folds is not spread, since continued glottal vibration is required. Sufficient airflow is presumably achieved by spreading the posterior cartilaginous portion of the glottis.

[2] The feature opposition tense/lax or fortis/lenis is sometimes used to describe the oppositions that we view in terms of the features [stiff vocal folds] and [slack vocal folds] (see e.g. Kohler 1984). This is relevant to our discussion at the end of §5.1.

[3] In a language like English the spreading of the glottis for fricatives is not regarded as a mechanism for enhancing the feature [+ continuant] but rather is part of the definition of how that feature is to be implemented. In some languages, however, glottal spreading can have the status of a feature, that is, [+ spread glottis] (Vaux 1998).

In Table 1 we list some examples of voiceless obstruents in a number of different environments. We assume these obstruents are specified by the phonological feature [+stiff vocal folds].

I	II	III	IV	V
rapid	still	p'an	nap	writer
rupee	instill	rep'eat	neck	ricing

TABLE 1. Word pairs illustrating voiceless consonants in five different phonetic environments.

Column 1 of Table 1 illustrates implementation of the feature [+stiff vocal folds] solely by the absence of glottal vibration. No enhancement is needed for consonants in this intervocalic environment. In column 2 enhancement is also unnecessary because there is no contrast between voiced and voiceless stops after an initial /s/ in English.

ASPIRATION AND VOICING. In certain environments, as in column 3 of Table 1, the voiceless consonants are produced with a spreading of the glottis following the consonant release. These are the so-called aspirated consonants. They occur when the consonant is in onset-initial position pretonically.[4]

All of the columns other than 3 contain examples of nonaspirated stop (or in one case fricative) consonants. These forms are in the complement set to those in column 3. That is, they are neither pretonic nor onset-initial. Typically, accounts of English have supposed that voiceless stops are aspirated and that a rule of deaspiration applies in the complement set just mentioned (Gussmann 2002). A different interpretation is, however, possible. Note first that the feature [spread glottis] is not distinctive for obstruents in English and, therefore, does not appear in the phonological representation specifying any given English obstruent. Nonetheless, it does accompany the gesture motivated by the feature [+stiff vocal folds] in certain environments, namely, those in column 3 above. We take this fact to be due to the use of glottal spreading as an enhancement gesture in onset-initial pretonic stop consonants.

What is there about pretonic, onset-initial position that requires enhancement of voicing? Recall that, in our view, enhancement may take place whenever a given distinction can be made more salient than it might otherwise be. In onset-initial position the distinction that is threatened pretonically is that between voiced and voiceless consonants, for example, *pan* versus *ban* and the verbs *repel* versus *rebel*. In English, prevoicing in /bæn/ is presumably weakened because of the vocal-fold stiffening in anticipation of the following stressed vowel. To counteract this, the distinction is enhanced by spreading the glottis for /p/, that is, by extending the voiceless interval into the beginning of the following vowel.

We are now in a position to observe that in the example of /t/ in *the tag*, discussed in §3.1, the feature-defining gesture for [+stiff vocal folds] is enhanced by a glottal spreading gesture since the /t/ is in pretonic position. The aspiration observed for /t/ in Fig. 3a, therefore, is the result of an enhancement gesture.

WORD-FINAL VOICING FOR STOP CONSONANTS. A common phonetic accompaniment of syllable-final alveolar stops is glottalization, something that is much less consistently

[4] Similar arguments apply in word-initial position, even though the consonant is not pretonic. Thus, the initial /p/ in *Potomac*, *Potemkin*, and *potato* are aspirated despite the following unstressed and reduced vowel. Despite the lack of minimal pairs in these instances, aspiration nonetheless occurs as a result, we suspect, of a more general phenomenon of English whereby word-initial segments are enhanced whenever possible (Gow et al. 1996).

true for velar and labial stops in this context. Why should there be this asymmetry? For any postvocalic voiceless stop consonant, it is important that glottal vibration be extinguished rapidly at the termination of the vowel. In the case of a labial stop consonant, the consonant closure appears to be accompanied by a stiffening of the tongue surface posterior to the constriction, thus inhibiting expansion of the vocal-tract volume and hence accelerating the buildup of intraoral pressure (Svirsky et al. 1997). A consequence is rapid inhibition of vocal-fold vibration. It is suggested that this same action applies to a velar stop consonant. In the case of the alveolar closure, however, there is need for flexibility in making the gesture for the forward part of the tongue. Consequently, after closure the pressure buildup causes expansion of the vocal tract. This, in turn, causes a pressure drop across the glottis to be maintained, which supports continued vocal-fold vibration. This continued vocal-fold vibration would then decrease the saliency of the voiced/voiceless contrast. In order to countervail the influence of the pressure drop across the glottis, speakers employ glottalization, which extinguishes glottal vibration even though a pressure drop remains. This gesture enhances the voicelessness of the postvocalic alveolar stop. Thus in the sequence *batboy*, the voiceless stop tends to be glottalized, whereas in *backbone* or *lapdog* the voiceless stop tends not to be glottalized.

In some dialects, words like *bottle* are produced with a glottal stop in place of the medial /t/: [baʔl̩]. However, the labial and velar consonants in the same environment are not glottalized, for example, *topple* and *nickel*. This asymmetry is accounted for if one assumes that only the medial /t/ is enhanced with a glottalizing gesture.

In utterance- and/or phrase-final position, however, one often observes glottalization in all three voiceless stop consonants. For example, in the case of *nap* (Table 1, column 4) the word-final voicing contrast with *nab* is compromised utterance-finally for two reasons. First, glottal vibration during the closure for /b/ may be weak because of the reduced subglottal pressure that occurs utterance-finally. Second, the difference in fundamental frequency that is normally observed in a vowel following a voiced or voiceless stop is unobservable due to the absence of a following vowel. An enhancing gesture is, therefore, appropriate in utterance- or phrase-final position. To inhibit glottal vibration for a voiceless stop, the glottis is often constricted immediately preceding the stop closure. A consequence of this glottal-constricting movement, when combined with the feature-defining stiffening gesture for the feature [+stiff vocal folds], is to cause a rapid decrease in the amplitude of glottal vibration.

PHRASE-FINAL VOICING. In phrase-final position, the voicing distinction for obstruent consonants (both stops and fricatives) is likely to be at risk because there is often a reduction in subglottal pressure in this environment. This reduced pressure tends to make continued glottal vibration more difficult, putting at risk the distinction between a voiced and voiceless consonant based only on glottal vibration within the obstruent interval. In English, and in other languages as well, this distinction is in some jeopardy. Shortening the preceding vowel before a voiceless consonant and lengthening the consonant increases the overall duration of voicelessness present in the signal, thereby enhancing that property. This increase in the duration of voicelessness parallels that of syllable-initial stop consonants in pretonic position (see 'Aspiration and voicing' above). An example is the oft-noted contrast between the word *misuse* as a verb and/or a noun when it occurs in phrase-final position. In many utterances of these words in phrase-final position, differences in vocal-fold vibration within the alveolar fricative are difficult to detect, but there are substantial differences in the duration of the preceding vowel.

TONGUE-ROOT ADVANCING AS AN ENHANCEMENT GESTURE FOR VOICING. The relevant distinction in column 5 of Table 1 is that of the voicing contrast that distinguishes *writer* and *rider* or *ricing* and *rising*. This distinction can be carried in part by the duration of the preceding vowel. However, closer examination of this vowel often shows differences in the formant frequencies in the vowel offglide, depending on whether the following consonant is voiced or voiceless (Kwong & Stevens 1999). Immediately preceding a voiceless consonant, as in *writer* or *ricing*, the first formant is lower and the second formant higher than for the contrasting voiced consonants in *rider* and *rising*, respectively. Furthermore, the first-formant frequency in the first part of the diphthong is often lower when the following consonant is voiceless than when it is voiced, though this vowel raising is not consistent in American English. From these findings it can be inferred that the tongue root has been advanced in the vowel that precedes the voiceless consonant. Advancing the tongue root results in a raising of the tongue body. Consequently, the pharynx is more fully extended, and little further expansion is possible during the consonantal portion of the gesture. This fully extended pharynx, therefore, prevents further airflow through the glottis and, consequently, inhibits consonantal voicing. This method of enhancing the voicing contrast seems to apply only when the preceding vocalic nucleus consists of a vowel followed by an offglide with a high tongue-body position. If the vowel is a lax vowel, as in *bitter/bidder*, *better/bedder*, or *latter/ladder*, and so on, the voicing distinction is apparently neutralized by flapping (see §6.2).

The case of Canadian raising is relevant to this discussion (see, for example, Joos 1942, Kenstowicz 1993). In the preceding paragraph we note that advancing the tongue root is responsible for the formant distribution in the offglide in words like *writer* in American English and, presumably, in Canadian English as well. A major difference between American and Canadian English has often been observed, namely, that the vowel of the vocalic nucleus in words like *writer* is considerably higher than its American counterpart. We have noted, however, that in some speakers of American English there is a marked tendency, not as prominent as in Canadian English, but discernible nonetheless, to raise the vowel nucleus as well as the glide before voiceless consonants. We suggest that the characteristic formant distribution in glides before voiceless consonants in both varieties of English is due to advancing the tongue root, and that regressive assimilation is responsible for the accompanying raising in a preceding vowel nucleus.

Advancing the tongue root for the glide, then, is intended to enhance the voicelessness of the following consonant. In common with other types of enhancement, this enhancement is not a rule-governed phenomenon.

NASALIZATION AND VOICING IN MIXTEC. The above examples illustrate the enhancing of the voicing feature [stiff vocal folds] by manipulating duration, by spreading or constricting the glottis, by adjusting the stiffness of the vocal-tract walls, or by manipulating the vocal-tract volume. Another way in which vocal-fold vibration during an obstruent consonant can be maintained is to partially open the velopharyngeal port during a portion of the consonant-closure interval. This gesture prevents buildup of pressure in the vocal tract and hence permits vocal-fold vibration to continue. However, if the consonant is to be an obstruent (in contrast to a sonorant nasal consonant), it is necessary to close the velopharyngeal port a few tens of milliseconds before the consonant is released. The result is a consonant with attributes similar to those of a prenasalized stop.

In Mixtec (Iverson & Salmons 1996) prenasalization occurs optionally before labials in word-initial position and obligatorily before alveolars.

(1) a. $^{(m)}$bàʔà ‘good’
b. $^{(m)}$báʔú ‘coyote’
c. ^{n}daʔa ‘hand’
d. ^{n}dákɪ ‘stiff, stale’

Iverson and Salmons argue that these examples reflect superficial prenasalization as a property of phonetic implementation rather than as a fundamental phonemic feature (p. 170), and elsewhere that prenasalization is an instantiation of a low-level phonetic phenomenon serving to help maintain a distinction that is otherwise difficult to produce (p. 172). In this case the relevant contrast is that between voiced and voiceless unaspirated stop consonants. We agree with their assessment of the role of prenasalization in Mixtec. In our framework prenasalization is an enhancement gesture.

ENHANCEMENT OF LENIS/FORTIS IN O’ODHAM. In a personal communication, the late Ken Hale offered the following example of enhancement of the lenis/fortis distinction in O’odham. Because of the intrinsic interest of the example, we quote his communication in full here.

> Many years ago, one of my O’odham (Papago) colleagues, Albert Alvarez, told me that the practically inaudible feature opposition between the two stop series in initial position could be heard, actually, if you whispered the words. A pair like /gai/ ‘roast:PERF’, and /kai/ ‘seed’, he told me, were actually different by virtue of tension in the articulation, but what you actually could HEAR was not THAT but rather an opposition which he named *s-hewbagim* (mellow, bland, like the yielding ground that you can safely fall in when thrown by a horse) versus *s-mu’ukam* (sharp, like the point of an awl, or a shrill scream). He claimed that this distinction was clearly audible when the words were whispered and you cupped your hands over your ears. I learned to do this, and he seemed to be right. But he was clear in his belief that this was a CLUE and that the true opposition had to do with articulatory tension of some sort.
>
> This is in initial position. In postvocalic position, the distinction is abundantly clear. The b,d,j,g-series is preglottalized, and the p,t,c,k-series is very audibly preaspirated. But again, these are CLUES to the distinction, which is still something like articulatory tension, not itself obviously audible.
>
> I think he was talking about enhancement—the features mellow, sharp, preglottalized and preaspirated are not real features of the segments, but enhancements. And, sure enough, there are no rules of O’odham phonology that refer to preaspiration, preglottalization, mellowness, or sharpness; but there are plenty that refer to a more abstract distinction like tense-lax or lenis-fortis.

5.2. NASALIZATION AND SPREAD GLOTTIS. Another example illustrating the common acoustic result of two different articulatory gestures is related to nasalization for vowels. The nasalization of a vowel is normally achieved by lowering the soft palate, resulting in acoustic coupling between the main vocal tract and the nasal cavity. The principal consequence of this acoustic coupling is a flattening of the spectrum in the vicinity of the first formant. This decreased low-frequency prominence is due to increased acoustic losses on the extensive surface area of the nasal cavity as well as the introduction of additional spectral peaks in the vicinity of the first formant. One of these peaks, thought to be due to a sinus resonance, is typically at low frequencies in the region of 300–400 Hz (Dang et al. 1994, Chen 1997). This resonance enhances the amplitude of the first or second harmonic in the spectrum of a nasalized vowel. The enhancement of the first harmonic in relation to higher harmonics has been reported as a concomitant of vowel nasalization in French (Hattori et al. 1958, Chen 1997).

An increased first-formant bandwidth and a general flattening of the spectrum at low frequencies also occurs when there is vocal-fold vibration with a spread glottis (Hanson 1997). The glottal opening contributes acoustic losses to the lowest vocal-tract resonance and also enhances the amplitude of the first harmonic in relation to higher harmon-

ics. Thus some of the acoustic consequences of spreading the glottis are similar to the acoustic correlate of nasalization.

Through experiments involving manipulation of the glottal source in synthetic speech, Klatt and Klatt (1990) have shown that enhancing the amplitude of the first harmonic of a vowel and increasing the first-formant bandwidth bias perception of the signal toward nasalization. We suspect that spreading the glottis enhances the first harmonic of nasal vowels in French in order to make them sound more nasal.

Ohala (1982) calls attention to an Indo-Aryan phenomenon, first studied by Grierson (1922), in which nasal vowels appear in certain words where the following consonant is voiceless. This process is of interest here since these words never had a nasal consonant in their prehistory. One can surmise that the final voiceless consonant induced some spreading of the glottis in the preceding vowel, leading to an increased first-formant bandwidth and an enhanced first harmonic.[5]

5.3. CONSONANTAL PLACE OF ARTICULATION: TONGUE-BLADE FEATURES. The tongue-blade place of articulation provides a range of examples in a variety of languages where enhancement frequently plays a role in identifying place of articulation. This section provides examples of such enhancements.

ALVEOLAR STOP CONSONANTS AND TONGUE-BODY POSITION IN ENGLISH. The production of obstruent consonants involves an increase in intraoral pressure and the generation of noise due to turbulent airflow in the vicinity of a constriction in the oral region of the vocal tract. The constriction can be a complete closure to form a stop, in which case the turbulent noise is generated immediately following the release of the closure, or it can be a narrow opening to form a fricative. The constriction can be implemented by the lips, the tongue blade, or the tongue body. When the constriction is at the lips, there is essentially no acoustic cavity in front of the constriction, and the sound output has the properties of the unfiltered turbulence-noise source, and has no major spectrum prominence at high frequencies. For a tongue-blade constriction, there is a cavity in front of the constriction, with a length of about 2 cm for an alveolar fricative or stop consonant in English. This short front cavity has a resonance in the range of 4–5 kHz, and excitation by a turbulence-noise source near the constriction results in a spectrum prominence in the sound output in this frequency range. For a velar consonant, the cavity in front of the constriction is generally in the range of 4–5 cm, and noise excitation of this cavity leads to a spectrum prominence in the range of F2 or F3, depending on whether the velar consonant is backed or fronted. These spectrum characteristics of the frication noise can be considered to be the defining acoustic attributes for the consonants produced by the three articulators—lips, tongue blade, and tongue body.

An important cue that distinguishes the three places of articulation for stop consonants in English, in addition to the noise spectrum, is the F2 transition in the adjacent vowel. This transition is largely the result of movement of the tongue body and lips as the articulators perform the gesture from the consonant closure to the following vowel. In the case of the velar consonant, the movement of the

[5] The discussion by Whalen and Beddor (1989) of intrusive vowel nasalization for long low vowels in Eastern Algonquian may provide an additional example that could be dealt with in terms of enhancement. However, we do not explore that possibility here.

tongue body is prescribed, because the features for this consonant require that the tongue body be in a raised position.

Spectrograms of a voiced labial and alveolar stop consonant before a back vowel and a front vowel are shown in Figure 5. For a labial consonant, the release of the lip opening and the lowering of the jaw cause F2 to rise, and the tongue-body position during the consonant anticipates the following vowel. Thus the starting frequency for F2 is lower when the labial is before a back vowel than before a front vowel. The F2 starting frequency for labials is always equal to or less than that for the adjacent vowel since F2 always rises when a constriction is released at the front of the oral cavity.

The alveolar consonant is produced with closure of the tongue blade, and there would appear to be some freedom in selecting the position of the tongue body. This is not the case, however. The tongue body for an alveolar before a back vowel is fronted so that the F2 starting frequency is higher than that for a labial before a back vowel, as

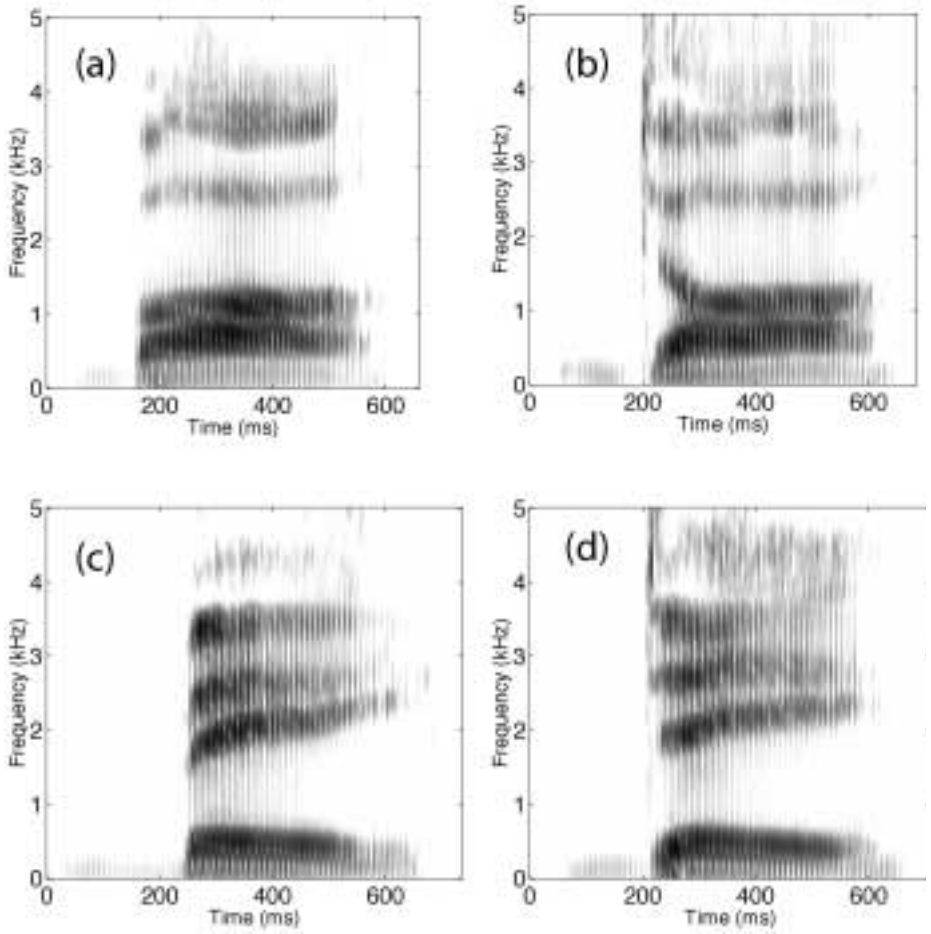

FIGURE 5. Spectrograms of (a) /ba/, (b) /da/, (c) / bɛ/, and (d) /dɛ/. These spectrograms illustrate the higher starting frequencies of the F2 and F3 transitions for the alveolar consonants relative to the labial consonants, indicating a more fronted tongue-body position for the alveolars. The F2 starting frequency is less dependent on the following vowel for /d/ than for /b/.

shown in Figs. 5a and 5b. Before a front vowel, the F2 starting frequency for an alveolar is only slightly higher than it is before a back vowel, indicating about the same fronted tongue-body position as that for the alveolar preceding a back vowel. This F2 starting frequency is greater than that for a labial stop before a front vowel, so that the F2 transition of the alveolar consonant is again distinct from that for the labial consonant, as Figs. 5c and 5d show. Tongue-body positioning for the alveolar consonants, then, appears to constitute an enhancing gesture.

Velar consonants play no role in this account because the convergence of F2 and F3 along with their unique release burst provide such a distinctive acoustic output that saliency is rarely an issue.

THE FEATURE [anterior] FOR FRICATIVES IN ENGLISH. In English the fricative consonants /ʃ/ and /s/ are both produced by raising the tongue blade against the hard palate. For the [− anterior] /ʃ/ the blade is shaped to form a narrow opening between the blade and the palate with a length of 3–4 cm. In addition, a narrow cavity is usually formed on the underside of the tongue blade (Perkell et al. 1979). A turbulence-noise source is generated when the air stream emerging from the tongue-blade constriction impinges on the lower incisors. The primary acoustic consequence of this source and of the surrounding acoustic cavities is that the lowest major spectral prominence in the sound is in the third-formant region—about 2800 Hz for female speakers and about 2500 Hz for male speakers. With proper positioning and shaping of the tongue blade, this lowest resonance can usually be adjusted to be in the proper frequency range. This affiliation of F3 with a front-cavity resonance, resulting in a spectrum prominence in the F3 range, can be regarded as a defining acoustic attribute for the feature [− anterior]. In the case of /s/, which is [+ anterior], the acoustic source is also at the lower incisors, but the tongue blade is positioned in a more anterior position so that the front-cavity resonance is F4 or F5 (or even F6), rather than F3.

As we have seen above, the length of the cavity in front of the tongue tip can be fine-tuned by adjusting the degree of lip rounding so as to lower the frequency of the lowest resonance until it becomes aligned with the F3 region for the speaker, thereby enhancing the saliency of the contrast between /s/ and /ʃ/. This effect of rounding for /ʃ/ (in English) is illustrated in the three spectrograms and spectra of Figure 6. The words *sip* and *ship* for the first two spectrograms were made normally. The utterance on the right is an attempt to produce *ship* without lip rounding. It is evident from both the spectrogram and the spectrum that the prominence corresponding to the third formant is at a higher frequency for the unrounded utterance and that the spectral peak corresponding to the third formant is not as prominent as that corresponding to the fourth formant. Thus the contrast between the unrounded /ʃ/ and /s/ is not as salient as it is when the lips are rounded. Measurements made from a database of utterances displaying movements of points on the lips (Westbury 1994) show that protrusion of both upper and lower lips is greater for fricative consonants in words like *ship* than in words like *sip*. Lip rounding, then, appears to be used to enhance the perceptual saliency that is characteristic of the feature [anterior]. Perceptual experiments on the identification of consonants in the presence of noise show that the distinction between /s/ and /ʃ/ is one of the more salient place-of-articulation distinctions (see Miller & Nicely 1955).

THE FEATURES [anterior] AND [distributed] IN MANDARIN CHINESE. In Mandarin Chinese there are three strident fricatives that are produced with the tongue blade. These fricatives are represented in featural terms in Table 2. For the [+ anterior]

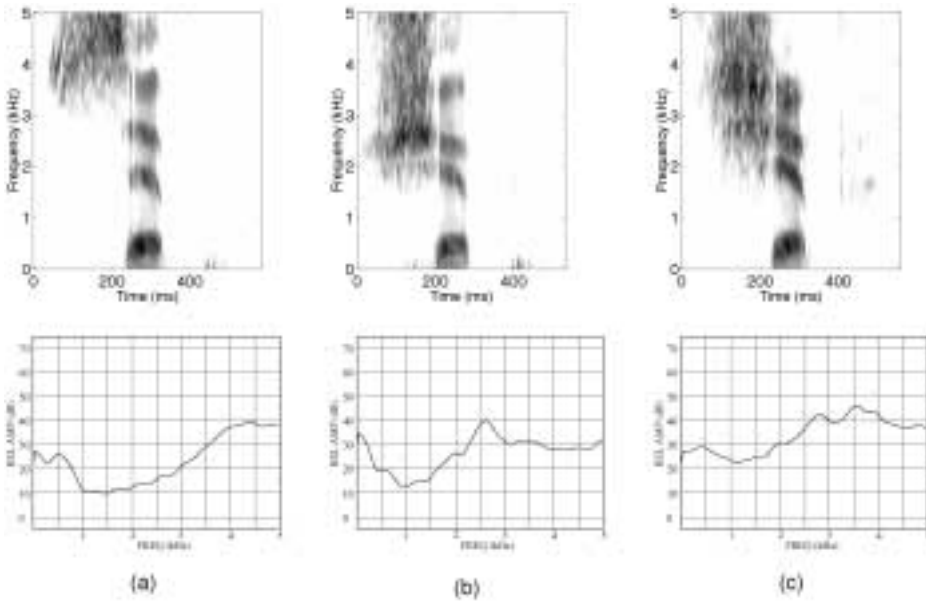

FIGURE 6. Spectrograms and spectra illustrating the acoustic properties of /s/ and ʃ: (a) the word *sip*, (b) normal production of the word *ship*, and (c) production of the word *ship* but with no rounding in the fricative. Spectrograms are at the top, and spectra, averaged over a 100-millisecond time interval within the fricative, are below the spectrograms. For the unrounded /ʃ/ the acoustic spectrum is intermediate between those of /s/ and rounded /ʃ/.

consonant /s/, the spectrum of the frication noise has a peak in the F5–6 range, as expected for an alveolar consonant. The [− anterior, + distributed] consonant /ɕ/, normally called a palatal, exhibits strong acoustic excitation of F4 in the frication noise, with weaker excitation of F3. The [− anterior, − distributed] fricative /ʃ/ has strong excitation of F3 and some excitation of the back-cavity resonance F2 as has been reported for this [− distributed] consonant (see Stevens et al. 2004, and, for a similar consonant in Polish, Halle & Stevens 1996). These articulatory and acoustic attributes can be regarded as defining the distinctive features [anterior] and [distributed] for these obstruent consonants produced with the tongue blade.

	s	ʃ	ɕ
[anterior]	+	−	−
[distributed]	−	−	+

TABLE 2. Distinctive features for Mandarin fricatives (and affricates).

Two kinds of enhancement appear to be introduced to strengthen the perceptual distinctions between these three fricatives (see Stevens et al. 2004). These include tongue-body fronting for the [+ distributed] (i.e. palatal) fricative, and a more backed tongue body for the [− distributed] consonants. The distinction between [+ anterior] and [− anterior] for these fricatives appears to be enhanced by creating a space under the tongue blade for the [− anterior] consonant, thereby enlarging the cavity in front of the constriction. Comparison with English shows some differences in these enhancement gestures, since English does not have [− anterior, + distributed] tongue-blade fricative, that is, /ɕ/.

SYLLABLE-FINAL NASAL CONSONANTS IN SOUTHERN CENTRAL AUSTRALIAN. The principal cues for place of articulation for syllable-final nasal consonants are the transitions of the formants from the vowel into the consonant closure. Normally, the latter part of the vowel in which these transitions occur is nasalized. This vowel nasalization is expected to make the formant movements less distinct, and hence to make the distinction between different places of articulation less salient perceptually (Repp & Svastikula 1988).

In Hercus 1972 it is observed that in certain Southern Central Australian languages, Arabana, Aranda, and WaNgaNuru, medial nasal consonants /m/, /n/, and /n̲/ in poststress position are produced with a brief obstruent interval at the time of consonantal closure.[6] This phenomenon is called PRESTOPPING. It occurs in a wide variety of Southern Central Australian languages as well as in Olgolo, a North Queensland language (see Dixon 1970). For these nasal consonants the soft palate is lowered immediately following the obstruent interval, and the consonant becomes a nasal. In this way, nasalization is postponed until the consonantal closure is complete, so that the formant transitions occur while the vowel is still nonnasal. The result is expected to be formant transitions that are more distinctive and hence the distinctions between places of articulation for the consonant are more salient. We suggest that this denasalization is invoked to enhance the place distinction for the poststress, medial consonants.[7]

Since there can be as many as six places of articulation for nasal consonants in these languages (four of which involve the tongue blade), it is not surprising that an enhancing gesture like prestopping is invoked. Prestopping produces formant transitions with greater clarity in a postvocalic, poststress position where the preceding vowel may be heavily nasalized and where the following unstressed vowel may not carry the necessary fine distinctions in the formant transitions.

As Hercus shows, there is considerable variation among the languages of South Central Australia with respect to this phenomenon. She notes, in fact, that /m/, /n/, and /n̲/ are by far the most commonly prestopped nasals with the distribution of the remaining nasals being variable.[8]

Whereas prestopped nasals are in complementary distribution with nonprestopped nasals in Arabana-WaNgaNuru, this is not the case in Aranda. In the latter language the loss of certain initial consonants appears to have produced a situation in which prestopped nasals occurred in vowel-initial words where they contrasted with nonprestopped nasals. Assuming this to be the case, we have an example in which what began as an enhancement phenomenon was elevated to phonological status; that is, prestopped nasals became distinctive in Aranda while remaining nondistinctive in Arabana-WaNgaNuru and related languages (see Hercus 1972). In Western Aranda, for example, prestopped nasals occur freely in initial position, a consequence of their becoming distinctive (Ken Hale, p.c.).

THE LIQUIDS. We saw earlier that lip rounding is used in English to enhance the distinction between [+/−anterior] consonantal fricatives. Lip rounding appears to be

[6] We are indebted to Andrew Butcher for bringing this reference to our attention.

[7] Prestopped laterals also occur in similar environments in these languages, although their occurrence is less robust than prestopped nasals. We assume that prestopping of laterals is, as with nasals, an enhancing gesture.

[8] There is one important restriction on the occurrence of prestopping. It does not occur in poststress position if the prestressed segment is either a vowel or a nasal. It appears that an obstruent in the relevant onset licenses prestopping in a following nasal. We have no explanation for why this should be so.

used in English in /r/ (cf. Ladefoged & Maddieson 1996) but not /l/. Our own measurements, based on data from an x-ray microbeam system (Westbury 1994), support this observation. We conclude that lip rounding functions to enhance the difference between the liquids /l/ and /r/ as in *light* and *right*. In the case of /r/, the frequency of F3 is low, and it arises from a resonance of the cavity in front of the constriction formed by the tongue blade. The lowering of F3 is enhanced if the front-cavity length is increased by rounding the lips. For /l/, by contrast, F3 is higher in frequency and the front-cavity length is shorter. Lip rounding is avoided in order to guarantee a higher F3.

It should be noted that in both /l/ and /r/ tongue backing plays a role. However, it is not the same role in each case. In the case of /r/, the lowering of F2 caused by tongue backing makes it possible for F3 to be low and hence to enhance the feature [− lateral]. In the case of /l/, the lowered F2 can increase the perceptual contrast with the glide /j/ which, like /l/, has a higher F3. Tongue backing functions to maximize the acoustic difference between /l/ and /j/ by lowering F2 in the case of the former. Note that there are some similarities between /l/ and /w/, but the lack of rounding for /l/ helps to keep these apart.

5.4. TONGUE-BODY FEATURES. In this section we discuss two examples where tongue-body features are enhanced, either by adjusting lip gestures or laryngeal gestures.

TONGUE-BODY FRONTING FOR VOWELS AND GLIDES. When the tongue body is raised and fronted, the first formant decreases and the second formant increases to become close to the third formant. The raised second formant is a well-known acoustic correlate of the feature [− back]. The frequencies of the second and third formants can be increased still further by raising the tongue blade so that a long, narrow channel is formed between the blade and the hard palate. Thus raising the tongue blade can enhance the acoustic correlate of the feature [− back], especially for a [+ high] vowel or glide. The frequencies of F2 and F3 can also be raised by spreading the lips, thereby shortening the front cavity and increasing the natural frequency of that cavity. We observe, then, that the acoustic correlates of vowels and glides that are [− back, + high] can be enhanced by raising the tongue blade to shape the front cavity and by spreading the lips. Likewise, for nonlow back vowels the feature [+ back] is often enhanced by rounding the lips. In many languages this rounding gesture is not distinctive and, in our view, is simply a graded gesture (e.g. Spanish as noted earlier). The primary acoustic correlate of the feature [+ back] is an F2 that is low and close to F1. Rounding positions F2 even closer to F1 and hence enhances the feature [+ back].

ADVANCED TONGUE ROOT FOR VOWELS. The articulatory correlate of the feature [+ advanced tongue root] ([+ ATR]) is a widening of the vocal tract in the pharyngeal region. This widening is accompanied by a reduction in the cross-sectional area of the vocal tract in the oral region. The acoustic consequences of this kind of perturbation in the vocal-tract shape is a lowering of the first-formant frequency F1. Depending on the vowel, there may also be some lowering or raising of F2. Likewise, implementation of [− ATR] causes a raising of F1. A decrease in F1 causes a decrease in the amplitudes of the spectral prominences for formants above F1. Measurements of vowels in languages that have a distinction in this feature show a shift in F1 of roughly 100–150 Hz between a [+ ATR] and a [− ATR] vowel (Lindau 1979). Calculations show that a decrease of 100 Hz in F1 causes a reduction in the amplitudes of higher formant peaks in the range of 2.5 to 5.0 dB, depending on whether the vowel is a low or a high vowel.

It has been reported that vowels with the feature [+ATR] are often produced with a breathy-voice quality (Stewart 1967), whereas [−ATR] vowels are produced with a tenser or creaky voice. A breathy glottal configuration leads to a glottal source with weaker high-frequency amplitude or a spectrum that slopes downward more steeply with increasing frequency, compared with a modal or pressed voice. A decreased spectrum amplitude at 3 kHz of 10 dB is not unusual for a breathy voice compared with a modal voice. Use of this modified glottal configuration for a [+ATR] vowel, then, would enhance the downward-sloping spectrum that occurs as a consequence of the low F1 accompanying such a vowel. The role of the breathy glottal configuration for enhancing the perception of [+ATR] vowels in English has been shown in Kingston et al. 1997 and Kluender et al. 1995.

5.5. STRIDENCY. The feature [strident] is distinctive in English only for [+continuant] consonants produced with the tongue blade. Fricatives that are [+strident] are implemented by shaping the tongue blade in such a way that the air stream impinges on the lower teeth, creating a robust excitation of the acoustic cavity anterior to the constriction. In the case of [−strident] fricatives (i.e. [ð] and [θ]) the tongue blade must be positioned and shaped in such a way that the air stream does not impinge on the lower teeth, and consequently only weak frication noise is generated. Thus the tongue blade assumes a laminal or dental configuration by backing the tongue body. This tongue-body backing can be considered as a gesture that enhances implementation of the feature [−strident] in English. A consequence of this tongue-body backing is a lowered F2 at the consonant release. This lowered F2 has been shown to distinguish /ð/ and /θ/ in English from labial fricatives like /f/ and /v/ (see Harris 1958). For alveolar consonants like /t/, /d/, and /n/, F2 is significantly higher, indicating a more fronted tongue-body position (see Manuel 1995).

6. GESTURE OVERLAP.

6.1. OVERLAP AND CONSERVATION. In any mode other than careful enunciation, speakers overlap adjacent gestures. For example, in the phrase *up to* (see discussion of *top tag* in §3.2 above) the sequence /pt/ requires a series of four gestures of the oral articulators: a closing of the lips, an opening of the lips, the making of a closure with the tongue blade, and the opening of that closure. In careful enunciation each of these closings and openings can have an acoustic consequence. In casual speech, however, the tongue-blade closing gesture occurs before the labial release, so that there is no acoustic record of these two gestures. That is, there is no source of sound at the time the tongue-blade closure is made, and there is no pressure behind the lips at the time the labial release occurs (see Fig. 3b). Acoustic evidence for the place of articulation of these two consonants, however, remains in the labial closure and in the coronal release.

Normally the production of a voiceless stop consonant requires that the glottis be opened and closed. Thus the sequence /pt/ in carefully enunciated speech requires two opening/closing gestures of the glottis and a concomitant expiration of respiratory energy. In the casual production of /pt/, by contrast, the requisite sequence of glottal gestures is reduced to a single opening and closing movement, thereby reducing the amount of respiratory energy associated with these gestures. This compression of gestures is a characteristic of noncareful speech modes and can be observed in a wide variety of environments, some of which we explore below. We refer to processes of this sort as conservation, or overlap.

We acknowledge this phenomenon by inserting an operation called OVERLAP in Figure 7 between gesture calculations and VOCAL APPARATUS, indicating that output from gesture calculations is subject to overlap. The interaction of overlap and enhancement gesture calculations is discussed in §7 below.

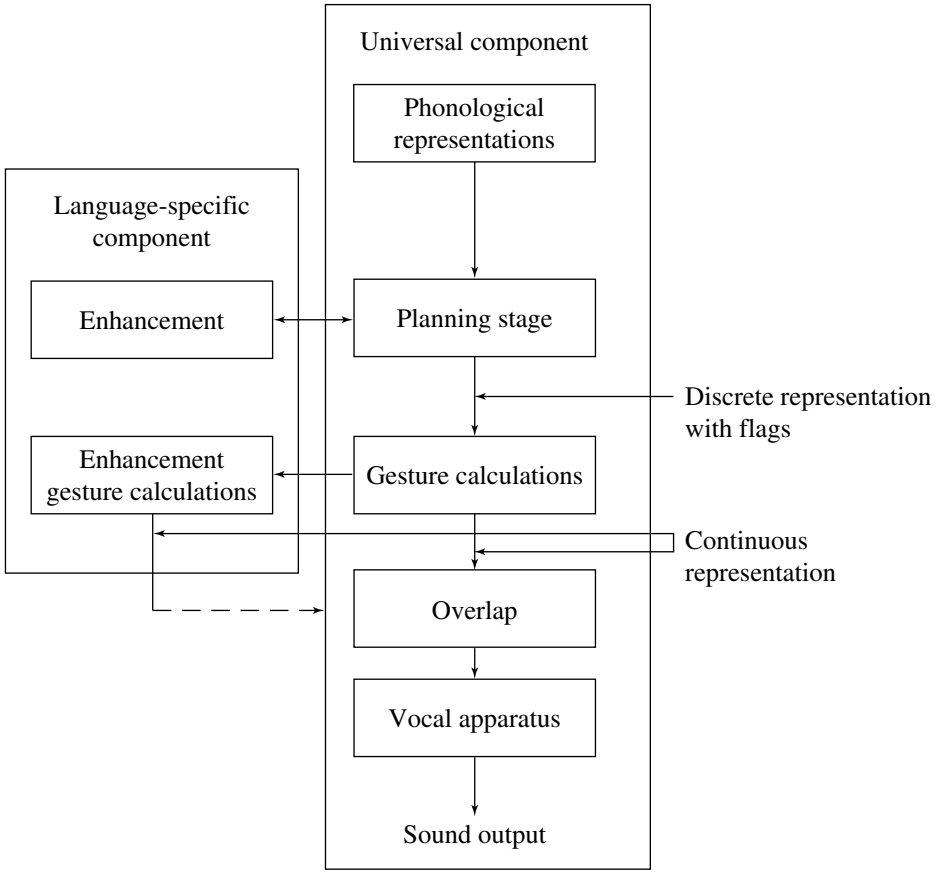

FIGURE 7. Fig. 4 is modified to include the capability for gestural overlap. Discussion of the destination of the dotted arrow is postponed to §7.

Figure 7 illustrates the view that speech production involves two important phenomena, enhancement and overlap, and that the separate as well as combined operations of these phenomena modify the representation derived directly from the defining acoustic and articulatory patterns for the features in the planning stage. In spite of these modifications, the underlying discrete representation remains recoverable by the listener. In the examples that follow we trace the route of this recoverability.

6.2. SOME EXAMPLES. There are several phenomena that have been treated as rule-governed. These include aspiration for stop consonants and vowel lengthening before voiced consonants in English. In Keyser & Stevens 2001 we tried to show that an alternative account of these and other phenomena is possible, one that treats the phenomena as resulting from enhancement. From this point of view, they do not involve featural manipulation. In what follows we pursue the same strategy, examining examples of gesture conservation or overlap in casual speech to show that here, too, rules are not

involved. We then examine certain more complicated cases in which both enhancement and conservation coexist.

VOWEL REDUCTION: *potato*, *bedazzled*, AND *monopoly*. It is well known that in non-careful speech the vowel in the initial syllable of *potato* can be voiceless. This absence of vocal-fold vibration during the vowel is a consequence of the influence of the voiceless consonants that precede and follow the vowel. Voicelessness for the initial consonant /p/ is enhanced with a spread glottis that extends a few tens of milliseconds beyond the consonant release. The following consonant /t/ also requires a spread-glottis gesture, again an enhancement gesture. In casual speech these two gestures can overlap to such an extent that the glottis never achieves a position that permits vocal-fold vibration to occur. A consequence of this gesture conservation is that the only evidence for the vowel resides in the aspiration for the voiceless stop /p/.

What role does enhancement play, if any, in the casual production of *potato*? Each of the spread-glottis enhancing gestures for /p/ and /t/ encroaches on the time available for the production of the reduced vowel /ə/. In practice the length of time that the /ə/ has available for glottal vibration can be reduced to zero: hence, its voicelessness. The enhancement gestures are preserved in spite of this overlap and are not truncated to permit voicing to occur in the vowel. In this case, it is important to note that enhancement gestures survive overlap while feature-initiated ones are severely weakened.

In *bedazzled*, by contrast, there is no enhancement of the syllable-initial voiced consonants. In particular, the glottis is not spread. Consequently, the vowel in the initial syllable never reduces to the voiceless state that overlap produces in *potato*.

In *monopoly*, the vowel of the initial syllable can be similarly squeezed by the gesture that opens the velum for the word initial /m/ and the following /n/. In this case, however, there is no concomitant enhancing gesture; that is, the lowering of the velum is associated with the distinctive feature [+nasal]. There is conservation, however, in the sense that the velopharyngeal gestures for the two nasal consonants are reduced to a single gesture that extends through the vowel.

/ən/ SEQUENCES: *lesson* AND *sudden*. In normal production of *lesson*, the consonant /s/ is released into the /ə/, the soft palate is lowered during the /ə/, and the tongue is then closed for the final /n/. Overlap is responsible for shifting the lowering of the soft palate back to the fricative /s/. A consequence is that the intraoral pressure for the /s/ suddenly drops to zero and the tongue blade makes a closure against the alveolar ridge. This action causes a nasal consonant to be produced immediately following the frication for /s/. The tongue blade is never released for /s/. In fact, were release to occur after this same shifting of the soft-palate lowering gesture, an extra segment would be introduced, producing the utterance [lɛsnən]. This consequence is avoided by failure to release the consonant /s/, so that it is followed simply by a gesture that has the properties of a vowel (with a low-frequency amplitude peak) and a nasal consonant. Thus a gesture is conserved, but nevertheless the presence of a reduced syllable is preserved as well as the nasal feature of the final /n/.

Similarly in the word *sudden* the soft-palate gesture is shifted backwards to a time immediately following the closure for /d/. This produces the bisyllabic [sʌdn̩] that terminates in a syllabic [n̩]. The onset of the syllabic [n̩] occurs when the soft palate is opened and the intraoral pressure for the [d] suddenly drops to zero, just as in the example of *lesson* above.

SYLLABLE-FINAL /t/ AND GLOTTALIZATION: *button* AND *batman*. The /t/ in the word *button* can be produced either with a tongue-blade closure at the end of the vowel [ʌ]

or with a glottal stop, with the tongue-blade closure being formed later. In either case the enhancing gesture for the alveolar, consisting of a fronting of the tongue body during the vowel /ʌ/, remains. This is evidenced by a rising second-formant transition during the vowel that distinguishes the alveolar place of articulation from a velar or a labial. If, in addition, the /t/ is glottalized, then the alveolar closure may not occur—for example, in a word like *batman*. Once again this is a case of conservation. In this case glottalization constitutes an additional enhancing gesture that distinguishes the alveolar from the labial and velar (see the discussion of word-final voicing in §5.1 above).

COARTICULATION OF VOICING: *his farm* AND *his vote*. In *his farm* the duration of glottal vibration in the /z/ of *his* is noticeably shorter than that for the /z/ in *his vote*. This is because the glottal spreading gesture in the /f/ of *farm* overlaps with the glottal gesture in /z/ of *his*, thereby cutting short the duration of the gesture for /z/. In *his vote*, by contrast, the glottal gestures remain in a position for voicing through the /z/ and the /v/. This overlap is, however, not obligatory. In more careful speech the duration of glottal vibration in /z/ preceding /f/ and /v/ is the same; that is, the /z/ is identical in *his vote* and *his farm*.

This is an example of a pure overlap strategy. Enhancement plays no role. This is in contrast to another well-known phenomenon of casual speech, namely, flapping. We turn to that now.

FLAPPING. By FLAPPING we mean the word-internal shift of voiced and voiceless alveolar stops to a flap intervocalically when the second vowel is unstressed.[9] In American English it is ubiquitous. In a word like *fraternity*, for example, only the second /t/ is subject to flapping. The first /t/ does not flap because it occurs in pretonic position. Other examples of /t/ that may undergo flapping include *atom*, but not *atomic*. In a word like *adiabatic* both alveolar consonants can undergo flapping since both are intervocalic and neither is pretonic.

We consider flapping to be an example of overlap, as represented in Fig. 7. In particular, while a nonflapped /t/ or /d/ requires a separate closing and opening gesture, flapping can be achieved with a single tongue-blade gesture. Furthermore, because the time of consonant closure in a flap is so short, there is no opportunity to represent the voicing distinction within the closure interval, and the glottal gesture required to signal this distinction can be omitted. A cue to the voicing distinction can remain in the form of different formant transitions in the preceding vowel. For example, in the word *writer* the transitions of the first and second formants preceding the flap are often more extreme than those in the word *rider*. Furthermore, the vowel preceding the flap is often shorter in *writer* than in *rider* (Fox & Terbeek 1977). Consequently, there need be no neutralization in those two words, contrary to many accounts. Compare this to *bidder* and *bitter* or to *hodder* and *hotter*, where no such cue to the voicing distinction remains when the consonants are flapped and where neutralization is quite common.

Apparently, the opportunity to expand the pharynx in advance of the flap in order to inhibit voicing exists for nonlow tense vowels only, all of which have an offglide toward an expanded pharynx, that is, toward /j/ or /w/. Thus, we can hypothesize that this type of enhancement of the voicing contrast by manipulating the formant structure at the end of the vowel is possible for some vowels and not for others. Creating such

[9] Flapping also occurs in other non-word-internal environments, for example, *that apple* and *bad apple*. Here flapping occurs even though the following vowel is stressed. Elsewhere we find flapping across major constituent boundaries as in *I had an apple*. We return to these cases in §7 below.

an offglide in words like *bitter* or *hotter* would compromise the quality of the preceding vowel. Therefore, enhancement by this method occurs only in those forms where the possibility of such compromise does not exist, that is, in words like *writer* and *hater*.

This example illustrates in a single utterance the operation of (i) enhancement via diphthongization and (ii) conservation through gesture overlap. Here each operates independently of the other. Other examples relevant to (i) and (ii) include those in 2.

(2) a. biter : bider
b. rater : raider
c. beater : beader
d. pouter : powder
e. coater : coder
f. looter : lewder

Among each of these words there are examples of the offglides /j/ and /w/ which often exhibit more extreme values of F1 and F2 when they precede a voiceless consonant than a voiced consonant.

SPREADING OF NASALIZATION: *win those*. In casual speech, phrases like *win those* exhibit radical compression (Manuel 1995). The frication that marks the distinctive nonstridency of the initial /ð/ in *those* is produced with an interdental positioning of the tongue blade. This tongue-blade position is itself achieved with a more backed tongue-body position than that for the preceding alveolar /n/. As described in §5.5, this tongue-body shape is represented in the sound by a lowered starting frequency for the second-formant transition. In the sequence *win those* the velopharyngeal opening for the nasal spreads to the right and the frication noise disappears. What remains is a nasal murmur in which the position of the tongue blade and tongue body changes from that for /n/ to that for /ð/. The presence of the underlying /ð/ is still represented in the formant transitions into the following diphthong, in this case /ow/. However, the implementation of the basic acoustic attribute for the feature [− strident] is masked by the spreading of the soft-palate gesture, thereby preventing pressure buildup.

In the framework of our model two processes are at work. The loss of frication occurs as the result of conservation through overlap. The gesture of lowering the soft palate for /n/ spreads into the adjacent consonant. Consequently, there is no increase in intraoral pressure and no forming of a narrow constriction for the fricative. The accompanying loss of nonstrident information in the speech signal is compensated for by the gestures of the tongue body and tongue blade that were originally made to prevent the airflow from impinging on an obstacle, namely, the lower teeth. These tongue-body and tongue-blade gestures are needed to mark the acoustic presence of [− strident] in the signal. They remain even though frication has been masked by the advancing nasal gesture as an indication of the presence, in the abstract representation, of the feature [− strident]. This example, like the previous one, exemplifies the simultaneous operation of enhancement and conservation through overlap.[10]

[10] The number of words that begin with /ð/ in English is very small, including *this*, *that*, *these*, *those*, *there*, *them*, *then*, and *the*. Such words often undergo severe truncation in casual speech; for example, *like that* loses initial frication of *that* to gemination or glottalization, that is, either [lajkkaet] or [lajkʔaet]. *Over there* can become [owvar] in some southwestern American dialects. Severe truncation of this sort probably goes beyond the scope of enhancement and overlap as conceived in this model. Rather it is probably the case that the so-called /ð/-initial closed-class items are able to be severely compressed precisely because their numbers are so small and so highly constrained syntactically and semantically.

7. ENHANCEMENT AND OVERLAP: SOME CONSEQUENCES. In the previous section some of the examples illustrated the introduction of enhancement gestures as well as gesture overlap. In *win those* the feature [− strident] is enhanced by shaping the tongue blade (and therefore adjusting the tongue body) to prevent the air stream from impinging on the lower teeth. The soft palate gesture for /n/ extends into the /ð/ to the point where the gesture representing the obstruency (and hence the stridency) of /ð/ disappears. That is, among the gestures for /ð/, there are none that directly represent the features [− strident, − sonorant]. The only evidence that remains for these features is in the enhancing gesture, that is, in the tongue-blade shaping. In the acoustic record this evidence is the transition of the second formant following the release into the diphthong /ow/.

Or again consider *ballgame*, *Elmer*, *help*, *fall-guy*, and so on. In casual speech the final /l/ of the initial syllable is often made without contact of the tongue tip with the alveolar ridge. Nonetheless the tongue backing for /l/, which enhances the feature [+ lateral], remains (see discussion of liquids in §5.3). Thus acoustic evidence for the syllable-final lateral can reside exclusively in the enhancing gesture itself.

Finally, consider the phrase-final voiced contrast for obstruent consonants. An example is in verb/noun pairs like $misuse_{\text{verb}}$ with final /z/ and $misuse_{\text{noun}}$ with final /s/. As discussed in §5.1, in phrase-final position the contrast residing in the presence or absence of glottal vibration is virtually lost. Nonetheless, the distinction is maintained through the use of duration of the preceding vowel, an enhancement gesture.

Another illustration of the simultaneous operation of overlap and enhancement is drawn from the word *batman*. The /t/ in the word *batman* or the /n/ in *teen bag* (see discussion in §3.2 above) may be produced without forming a complete closure of the tongue blade. Evidence for this consonant, however, appears in the second-formant transition in the preceding vowel, which is a consequence of tongue-body fronting—a gesture that is introduced to enhance the feature-defining tongue-blade gesture.[11]

Consider how the model in Fig. 7 generates the /t/ in *batman* in casual speech. In the planning stage, we begin with the distinctive feature representation for /t/. In the word *batman*, this /t/ is in syllable-final position. In the model, two features of syllable-final /t/ in such a position are in need of enhancement: place of articulation and voicing. This is noted at the planning stage by flagging these features.

At the gesture-calculations stage, the presence of a flag is communicated to enhancement-gesture calculations, where it is determined how these features are to be enhanced. In particular with respect to place of articulation, tongue-body fronting enhances alveolar placement of the tongue blade. With respect to voicing, glottalization is the appropriate enhancing gesture. Instructions for the implementation of these enhancement gestures are added at this point.

The feature-defining tongue-blade gesture in *batman* is typically overlapped in casual speech with the following labial gesture. In the extreme case the labial closing gesture of the following /m/ can even terminate the initial syllable. Nevertheless the tongue-body gesture signals the presence of the alveolar consonant (Gow 2002).

The above discussion has underscored an unexpected property; namely, while feature-defining gestures are, in certain contexts, subject to severe weakening up to

[11] Even when *teen bag* is produced without a tongue-blade closure, the word-final /n/ is still heard in *teen bag*. However, if the word *bag* is excised, listeners are likely to hear *team* rather than *teen*; that is, the presence of the labial closure influences the listener's perception. Thus, the interpretation of the word-final /n/ in *teen* is influenced not only by the immediate acoustic properties, but also by the following context (see Manuel 1995 for a similar conclusion with respect to the word-final /n/ in *win* in the sequence *win those*; see also Gow 2002).

and including obliteration, enhancement gestures are far more robust and are apparently never obliterated. We hypothesize that overlap is responsible for the deviations in careful speech illustrated by the preceding examples. We also suppose that, unlike feature-defining gestures, enhancement gestures are never subject to overlap severe enough to mask their acoustic consequences. This is reflected in Figure 8 where the output of the enhancement gesture-calculations box goes directly to vocal apparatus rather than indirectly via overlap. In terms of this model we suppose that whatever instructions enter the vocal-apparatus component must be realized. All enhancements will surface acoustically but only those gestures from the basic gesture-calculations component that have survived overlap will do the same.

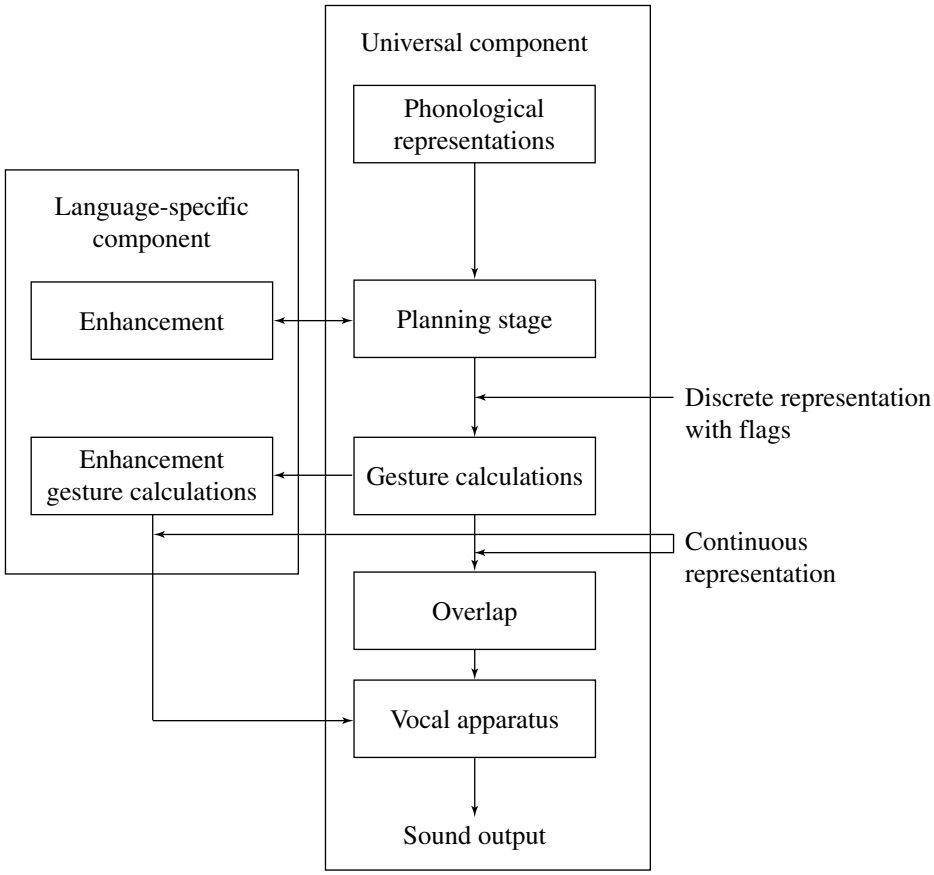

FIGURE 8. Fig. 7 is modified to indicate that enhancement gestures are not subject to overlap.

8. CONCLUDING OBSERVATIONS. In this section we touch upon a number of issues that relate to the model we have presented, including its relationship to speech production and perception, followed by a summary of the framework presented here.

8.1. OTHER SPEECH-PRODUCTION MODELS. Other models aim to account for the sequence of processes starting from a representation of an utterance in a planning stage; articulatory movements and acoustic patterns are derived from this input. One set of models regards lexical representations as being composed of gestural primitives and the articulators as a set of dynamic units (see Browman & Goldstein 1986, 1990, 1992,

Saltzman & Munhall 1989). The primary concern of these models is to account for the coordination and overlap of articulatory movements. The model in Perkell et al. 2000 views the planning stage as sequences of segments. The motor-control system is organized to produce speech-sound sequences based on auditory goals. This model utilizes an internal representation of the relation between vocal-tract configurations and their acoustic consequences so that the gestures required to achieve a given auditory goal can be calculated.

There are two components to the model proposed in this article: a set of gestures that is part of the definition of the universal distinctive features from which a given language selects a subset, and enhancing gestures that are language- and context-dependent. The defining gestures have acoustic and perceptual consequences that are universal, whereas the enhancing gestures are introduced to support and strengthen the perceptual saliency of these defining gestures. Our model differs from the others in that we propose a planning stage that includes a sequence of bundles of distinctive features. In common with other models we include modules that calculate articulatory actions, although the Saltzman and Munhall (1989) model includes a more quantitative procedure for calculating these actions. Our model implies that all articulatory gestures are not equal; some are universal and are defined by distinctive features and others are introduced in a language-dependent way to enhance the acoustic consequences of those features. The present model is similar to that of Perkell et al. 2000 in that attention is paid to the perceptual consequences of some gestures.

8.2. Relevance to perceptual models. There is a wide variety of models of speech perception. In terms of the framework developed here, each model has as its goal the reconstruction of the planning stage of the speaker in the mind of the listener, assuming, of course, that the semantic content of words is represented at this stage as well. Many models proceed by searching the acoustic input for characteristic bits of information that can be used to ferret out the word that it is intended to convey. A model such as the TRACE model of speech perception (McClelland & Elman 1986, Johnson 2001) does this by an interactive activation framework involving a series of nodes that represent features, phonemes, or words. This model assumes that features are continuous and evolve over time. In the course of this evolution, known principles of contextual variability in the speech stream are exploited in identifying segments and ultimately words. It is at this point that such models make contact with the framework developed here.

The interactive aspect of the TRACE model includes the possibility of postulating a word sequence, and, through feedback to a node at an acoustic level, verifying or rejecting this sequence. A speech-production strategy of the type we are proposing could perform this function in an analysis-by-synthesis model like TRACE. Through exploitation of the systematic variations that occur in enhancement and overlap, enhancement theory could provide an account of the so-called lawful variability that TRACE proposes to take advantage of (see Johnson 2001).

8.3. Parallel processes. As Fig. 8 illustrates, we assume two parallel processes at work in speech production. The first is the one normally assumed, that is, a planning stage where phonological representations exist in some quasi-linear array. These representations form the input to a component that replaces them with motoric instructions which are then implemented in the vocal tract, either unmodified or, more usually, modified by the process of overlap. Parallel to this process is the one we have referred to as enhancement, which takes note of features that are in jeopardy of losing their saliency when they are translated into feature-defining gestures. These gestures, we

have hypothesized, are shored up by enhancement gestures prior to implementation by the vocal apparatus. We have suggested that these gestures, unlike feature-defining gestures, are never subject to obliteration at the hands of overlap.

Within our framework the question arises as to where the enhancement-modified gestures reside. Are they in the lexicon as precompiled instructions? Or are they computed online, that is, never precompiled? Our theory commits us to the view that enhancement gestures are always computed online as opposed to defining gestures which are, of course, universal.

The reason for this is that enhancement gestures depend almost entirely on context. They are in the majority of instances introduced precisely because context puts their defining attributes in jeopardy. If these gestures were precompiled, then the precompilation would have to take into account the context in which the enhancement portion of the gesture arose. This strikes us as inordinately complicated. It would, in effect, require that the derivational history of each enhancement component be incorporated into the gesture itself. If, by contrast, the enhancement component is computed online, as it were, no such complication is required.

There are, of course, some enhancement gestures that are context-free, for example, the rounding of /ʃ/ in English. The rounding component would be a prime candidate for precompilation. Though nothing precludes having both precompiled and online-compiled enhancement gestures, for simplicity's sake we propose that all enhancement gestures are computed online.

8.4. SUMMARY. The model presented here rests on three fundamental assumptions. The first is that most defining gestures require enhancement. The second is that enhancement gestures are separate from defining gestures. The latter are universal and apply to every language in the world. Thus, the gesture associated with [anterior] in English is identical to the gesture associated with that same feature in Mandarin Chinese. Enhancing gestures, by contrast, vary from language to language depending upon the particular set of contrasts in that language. Thus, we find /ʃ/ rounded in English but not in Mandarin Chinese.

Our third assumption is that a planning stage exists that mediates between the lexicon and the apparatus needed to transform a given lexical item into sound. This immediately raises the question of where articulatory gestures associated with a given segment are assigned. It is a crucial assumption of our model that defining gestures and enhancement gestures are assigned at different stages in the model.

Our third assumption is not generally accepted. Some researchers take the position that segments are best represented as gestures. These researchers would presumably combine both defining and enhancing gestures into a single representation (see Browman & Goldstein 1992 and, more recently, the work of Ellis and Hardcastle (2002)). We have not taken this view because it offers no explanation for why certain segments are enhanced in one language but not in another. In addition, we have not taken this point of view because we see enhancement as a perceptual phenomenon, one designed to provide increased perceptual saliency. It follows that enhancement needs to be examined through perception and acoustics as well as through articulatory studies.

Further studies are certainly needed in order to test the validity of the assumptions made here. Future investigations into other languages will hopefully uncover other forms of enhancement than those discussed above. One consequence of such investigations is the possibility that what investigators have identified as separate phonemes are, in fact, not phonemes at all but rather enhanced or overlapped forms of already existing

phonemes (see Ken Hale's remarks at the end of §5.1 above). A corollary to this is that enhancement gestures can become phonologized. That is, they may undergo the same kind of change over time that defining gestures have shown.

REFERENCES

BELL-BERTI, FREDERICKA. 1975. Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America* 57.456–61.

BROWMAN, CATHERINE, and LOUIS GOLDSTEIN. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3.219–52.

BROWMAN, CATHERINE, and LOUIS GOLDSTEIN. 1990. Tiers in articulatory phonology with some implications for casual speech. *Papers in laboratory phonology 1*, ed. by John Kingston and Mary Beckman, 341–76. Cambridge: Cambridge University Press.

BROWMAN, CATHERINE, and LOUIS GOLDSTEIN. 1992. Articulatory phonology: An overview. *Phonetica* 49.155–80.

CHEN, MARILYN Y. 1997. Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America* 10.2360–70.

DANG, JIANWU; KIYOSHI HONDA; and HISAYOSHI SUZUKI. 1994. Morphological and acoustical analysis of the nasal and the paranasal cavities. *Journal of the Acoustical Society of America* 96.2088–100.

DIEHL, RANDY L. 1991. The role of phonetics within the study of language. *Phonetica* 48.120–34.

DIXON, ROBERT M. W. 1970. Olgolo syllable structure and what they are doing about it. *Linguistic Inquiry* 1.2.273–76.

ELLIS, LUCY, and WILLIAM J. HARDCASTLE. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30.373–96.

ERICKSON, DONNA. 1993. Laryngeal muscle activity in connection with Thai tones. *Research Institute of Logopedics and Phoniatrics Annual Bulletin* 27.135–49.

FOX, ROBERT A., and DALE TERBEEK. 1977. Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics* 5.27–34.

FRISCH, STEPHAN, and RICHARD WRIGHT. 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30.139–62.

FROMKIN, VICTORIA A. (ed.) 1973. *Speech errors as linguistic evidence*. The Hague: Mouton & Co.

GOW, DAVID W., Jr. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance* 28.163–79.

GOW, DAVID W., Jr.; JANICE MELVOLD; and SHARON MANUEL. 1996. How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing. *International Conference on Speech and Language Processing* 1.66–69.

GRIERSON, GEORGE A. 1922. Spontaneous nasalization in the Indo-Aryan languages. *Journal of the Royal Asiatic Society* July.381–88.

GUSSMANN, EDMUND. 2002. *Phonology: Analysis and theory*. Cambridge: Cambridge University Press.

HALLE, MORRIS, and KENNETH N. STEVENS. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Progress Report* 101.198–213.

HALLE, MORRIS, and KENNETH N. STEVENS. 1996. The post alveolar fricatives of Polish. *Speech production and language: In honor of Osamu Fujimura*, ed. by Shigeru Kiritani, Hajime Hirose, and Hiroya Fujisaki, 177–93. Berlin: Mouton de Gruyter.

HANSON, HELEN M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101.455–81.

HARRIS, KATHERINE S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1.1–7.

HATTORI, SHIRO; KENGO YAMAMOTO; and OSAMU FUJIMURA. 1958. Nasalization of vowels in relation to nasals. *Journal of the Acoustical Society of America* 30.267–74.

HERCUS, LUISE. 1972. The pre-stopped nasal and lateral consonants of Arabana-WaNga-Nguru. *Anthropological Linguistics* 14.293–305.

HONDA, KYOSHI; HIROYUKI HIRAI; and NAOKI KUSAKAWA. 1993. Modeling vocal tract organs

based on MRI and EMG observations and its implication on brain function. *Research Institute of Logopedics and Phoniatrics Annual Bulletin* 27.37–49.

HOUSE, ARTHUR S., and GRANT FAIRBANKS. 1953. The influence of consonantal environments upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25.105–13.

IVERSON, GREGORY K., and JOSEPH C. SALMONS. 1996. Mixtec prenasalization as hypervoicing. *International Journal of Anthropological Linguistics* 62.2.165–75.

JOHNSON, KEITH. 2001. Spoken language variability: Implications for modeling speech perception. *Proceedings of the Workshop on Speech Recognition as Pattern Classification (SPRAAC)*. Online: http://corpus.linguistics.berkeley.edu/~kjohnson/JohnsonSpraac.pdf.

JOHNSON, KEITH. 2003. *Acoustic & auditory phonetics*. 2nd edn. Oxford: Blackwell.

JOOS, MARTIN. 1942. A phonological dilemma in Canadian English. *Language* 18.141–44.

KENSTOWICZ, MICHAEL. 1993. *Phonology in generative grammar*. Oxford: Blackwell.

KEYSER, SAMUEL J., and KENNETH N. STEVENS. 2001. Enhancement revisited. *Ken Hale: A life in language*, ed. by Michael Kenstowicz, 271–91. Cambridge, MA: MIT Press.

KINGSTON, JOHN. 1992. The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech* 35.99–113.

KINGSTON, JOHN, and RANDY DIEHL. 1995. Intermediate properties in the perception of distinctive feature values. *Papers in laboratory phonology 4*, ed. by Bruce Connell and Amalia Arvaniti, 7–27. Cambridge: Cambridge University Press.

KINGSTON, JOHN; NEIL A. MACMILLAN; LAURA W. DICKEY; RACHEL THORBURN; and CHRISTINE BARTELS. 1997. Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America* 101.1696–709.

KLATT, DENNIS H., and LAURA KLATT. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87.820–57.

KLUENDER, KEITH R.; ANDREW J. LOTTO; and LORI L. HOLT. 1995. Effect of voice quality on the tense/lax distinction for English vowels. *Proceedings of the International Congress of Phonetic Sciences* 4.164–67.

KOHLER, KLAUS J. 1984. Phonetic explanation in phonology: The feature fortis/lenis. *Phonetica* 41.150–74.

KWONG, KATHERINE, and KENNETH N. STEVENS. 1999. On the voiced-voiceless distinction for writer/rider. *Speech Communication Group Working Papers—MIT Research Laboratory of Electronics* 11.1–20.

LADEFOGED, PETER, and IAN MADDIESON. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

LEVELT, WIM; R. D. ROELOFS; and ANTJE S. MYER. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–75.

LINDAU, MONA. 1979. The feature expanded. *Journal of Phonetics* 7.163–76.

LINDBLOM, BJORN. 1990. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modeling*, ed. by William J. Hardcastle and Alain Marchal, 403–39. Dordrecht: Kluwer.

MANUEL, SHARON Y. 1995. Speakers nasalize /ð/ after /n/ but listeners still hear /ð/. *Journal of Phonetics* 43.453–76.

MCCLELLAND, JAMES L., and JEFFREY L. ELMAN. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18.1–86.

MILLER, GEORGE A., and PATRICIA NICELY. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27.338–53.

OHALA, JOHN J. 1982. The phonological end justifies any means. *International Congress of Linguistics* 13.232–43.

PERKELL, JOSEPH; SUZANNE E. BOYCE; and KENNETH N. STEVENS. 1979. Articulatory and acoustic correlates of the [s-sh] distinction. *Speech communication papers: 97th meeting of the Acoustical Society of America*, ed. by J. J. Wolf and Dennis H. Klatt, 109–13. Melville, NY: Acoustical Society of America.

PERKELL, JOSEPH S.; FRANK H. GUENTHER; HARLAN LANE; MELANIE L. MATTHIES; PASCAL PERRIER; JENNELL VICK; REINER WILHELMS-TRICARICO; and MAJID ZANDIPOUR. 2000. A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics* 28.233–72.

POUPLIER, MARIANNE. 2003. *Units of phonological encoding: Empirical evidence*. New Haven: Yale University dissertation.
REPP, BRUNO H., and KATYANEE SVASTIKULA. 1988. Perception of the /m/-/n/ distinction in VC syllables. *Journal of the Acoustical Society of America* 83.237–47.
SALTZMAN, ELLIOT, and KEVIN G. MUNHALL. 1989. A dynamic approach to gestural patterning in speech production. *Ecological Psychology* 1.333–82.
SHATTUCK-HUFNAGEL, STEFANIE. 1986. The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook* 3.117–49.
STEVENS, KENNETH N. 1977. Physics of laryngeal behavior and larynx modes. *Phonetica* 34.264–79.
STEVENS, KENNETH N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17.3–46.
STEVENS, KENNETH N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
STEVENS, KENNETH N.; SAMUEL J. KEYSER; and HARUKO KAWASAKI. 1986. Toward a phonetic and phonological theory of redundant features. *Invariance and variability in speech processes*, ed. by Joseph S. Perkell and Dennis H. Klatt, 426–47. Hillsdale, NJ: Lawrence Erlbaum.
STEVENS, KENNETH N., and SAMUEL J. KEYSER. 1989. Primary features and their enhancement in consonants. *Language* 65.81–106.
STEVENS, KENNETH N.; ZHIQIANG LI; CHAO-YANG LEE; and SAMUEL J. KEYSER. 2004. A note on Mandarin fricatives and enhancement. *From traditional phonology to modern speech processing*, ed. by Gunnar Fant, Hiroya Fujisaki, Jianfen Cao, and Yu Xi, 393–403. Beijing: Foreign Language Teaching and Research Press.
STEWART, JOHN M. 1967. Tongue root position in Akan vowel harmony. *Phonetica* 16.185–204.
SVIRSKY, MARIO A.; KENNETH N. STEVENS; MELANIE L. MATTHIES; JOYCE MANZELLA; JOSEPH S. PERKELL; and REINER WILHELMS-TRICARICO. 1997. Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America* 102.562–71.
VAUX, BERT. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29.3.497–511.
WESTBURY, JOHN R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73.1322–36.
WESTBURY, JOHN R. 1994. *X-ray microbeam speech production database user's handbook*. Madison, WI: University of Wisconsin.
WHALEN, DOUGLAS H., and PATRICE S. BEDDOR. 1989. Connections between nasality and vowel duration and height: Elucidation of the Eastern Algonquian intrusive nasal. *Language* 65.457–86.
ZSIGA, ELIZABETH C. 1994. Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics* 22.121–40.

Keyser
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
32-D770
77 Massachusetts Ave.
Cambridge, MA 02139
[keyser@mit.edu]

Stevens
Research Laboratory of Electronics
Department of Electrical Engineering and Computer Science,
and Division of Health Sciences and Technology
Massachusetts Institute of Technology
36-517
77 Massachusetts Ave.
Cambridge, MA 02139
[stevens@speech.mit.edu]

[Received 20 March 2003;
revision invited 14 November 2003;
revision received 16 September 2004;
accepted 29 May 2005]