969

In: Bell-Berti, F., & Raphael, L. J. (1995). Producing Speech: Contemporary Issues. For Katherine Safford Harris. AIP Press: New York.

2 Gestural Syllable Position Effects in American English

Catherine P. Browman
Haskins Laboratories

Louis Goldstein
Haskins Laboratories
Yale University

INTRODUCTION

In a 1984 paper, Harris and Bell-Berti observed that "it remains in the future, then, for us to develop a theory of syllabification and coarticulation using evidence gathered from the articulatory domain with a net of mesh that has a smaller gauge than that which has produced our present views" (p. 94). The framework of Articulatory Phonology, which we have been developing over the last several years (Browman & Goldstein, 1986, 1989, 1992), provides one approach to meeting the challenge that they laid out.

In Articulatory Phonology, the basic phonological units are gestures, and lexical units are composed of dynamically defined articulatory gestures and their organization. Each gesture corresponds to the formation of a constriction within one of the relatively independent articulatory subsystems of the vocal tract, and can be modeled as a task-dynamic system or control regime (Saltzman & Munhall, 1989). For oral gestures, constriction tasks are defined along pairs of coordinates called tract variables, each of which specifies one of the two spatial dimensions of a constriction involving the lips, the tongue tip, or the tongue body (for English). In order to characterize the constrictions, movements are analyzed into discrete underlying gestures, and the values of the coefficients for a dynamic equation for each tract variable, or dimension of the constriction, are determined. Velic and glottal opening-and-closing gestures are considered to be unidimensional, and therefore consist of a single tract variable each. Gestures are associated with each other in terms of their phasing; the resultant gestural group is referred to as a gestural constellation.

One phonological property for which gestural constellations provide a particularly interesting account is syllable structure. Often, in nonarticulatory phonologies (e.g., Kahn, 1976) syllable structure is expressed as a hierarchical structure imposed on a linear (or multilinear) sequence of phonological units. One of the reasons that such structures are posited is to account for variation in phonological units as a function of their position within this structure. For example, Kahn attempted to account for allophonic differences in consonants (particularly /t/) as a function of being in syllable initial vs. noninitial position. However, in such approaches there is no intrinsic connection between the configurational structure (the syllable hierarchy) and the kinds of variation that it is described as conditioning.

In the gestural approach, there are configurational properties intrinsic to gestural constellations that identify syllable-initial vs. syllable-final consonants without necessarily invoking any hierarchical structure. Thus, affiliated consonants both preceding and following a vowel ('syllable-initial' and 'syllable-final' consonants) are phased with respect to the vowel, but in different ways. Moreover, because the configurational properties (the phase relations) also determine the patterns of overlap among gestures, differences in configuration are automatically associated with differences in overlap, and in turn, with the articulatory and acoustic consequences of such overlap. Thus, there is an intrinsic, lawful relation between configurational patterns that correspond to different syllable positions, and (at least some of) the kinds of articulatory and acoustic variation that can be observed for a given phonological unit. Put another way, within this view syllable structure is a characteristic pattern of coordination among gestures; certain types of variation are automatic consequences of this pattern of coordination, and are, therefore, necessary correlates of syllable structure. Moreover, this view entails that, although the exact nature of the observed correlates will vary as a function of the nature of the gestures being organized, the patterns of coordination among disparate gestures may be identical. We will see an example of such identity of coordination in the next section.

TIMING OF GESTURES FOR NASALS AND LATERALS

In American English, both nasals and laterals consist of two gestures. Nasals, of course, involve a velic opening gesture in addition to an oral closure gesture.

Apparently, laterals also consist of two gestures. Sproat and Fujimura (1993) showed that in American English both light and dark [l] (i.e., both syllable-initial and syllable-final [l]) consist of two movements, one involving the tongue tip and one involving the tongue dorsum. In the language of Articulatory Phonology, this would mean that [l] consists of two gestures. This double aspect of [l] can be seen in Figure 1, which shows a mid-sagittal view for a sustained [l] for six different speakers of American English, acquired using Magnetic Resonance Imaging. From this perspective, any lateral side-channel could be purely a secondary consequence of the stretching of the tongue caused by the tongue tip raising toward the upper teeth and tongue dorsum retraction. That these two gestures, without specification of lateral channels, can indeed produce a percept of [l] was demonstrated using a modified version of the Haskins vocal tract model to produce the shape shown in Figure 2, to which there are no side channels added (see Rubin, Baer, & Mermelstein, 1981, for a description of the original model). This shape produced an acoustic signal that sounded like an [l] in informal perceptual testing.

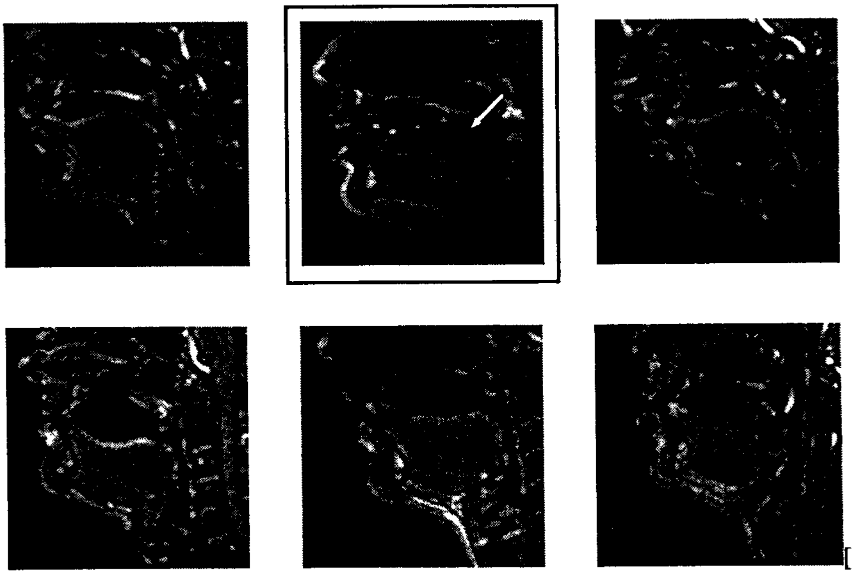


FIGURE 1. MRI images of [l] (held for several seconds) for 6 speakers of American English. The speakers lived in different regions between the ages of 0 and 17: left to right, top row, Connecticut; Chicago (0-13)/Cincinnati (13-17); Washington state; left to right, bottom row: Texas/Oklahoma; San Francisco area (0-7)/primarily Australia (7-12)/Minnesota (12-17); Philadelphia. The image in the middle of the top row that is outlined with a box is from the speaker for whom X-ray information on [l] is reported in Figures 3 through 5. The white ellipses on the surface of the tongue approximate the placement of the gold pellets in the microbeam experiments. Movement data for the frontmost pellet and for a pellet in the position indicated by the arrow are shown in Figure 3.

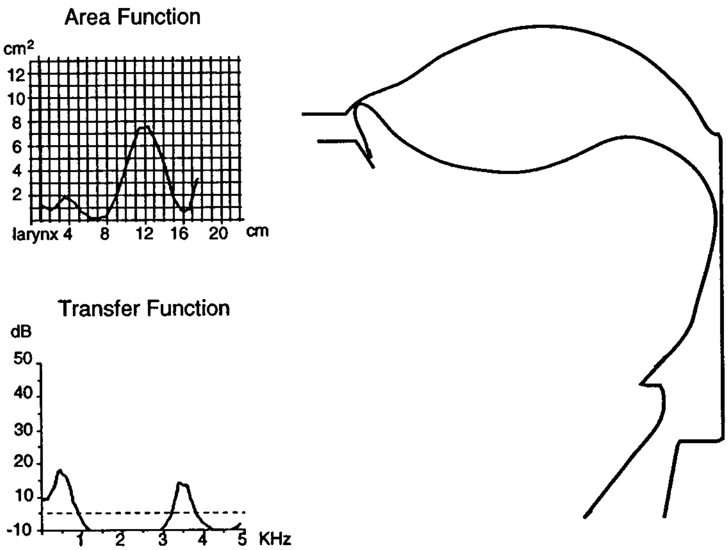


FIGURE 2. Bi-gestural [l] in vocal tract model (with no side channels). Midsagittal view on right, area function and transfer function on left.

Both nasals and laterals in American English show differential timing between their component gestures syllable-initially and -finally. Moreover, as discussed in Browman and Goldstein (1992), the temporal patterns appear to be the same for the nasals and laterals. For the nasals, Krakow (1989) has shown that in initial nasals (e.g., "see more") the end of velum lowering roughly coincides with the END of the lip closing movement, whereas in final nasals (e.g., "seem ore") it coincides with the BEGINNING of the lip closing movement. That is, the velic opening gesture occurs much earlier with respect to its associated oral closure when syllable-final than when syllable-initial. Sproat and Fujimura (1993) showed analogous behavior for laterals, with the tongue dorsum gesture occurring much earlier with respect to its associated tongue tip closure when final than when initial (they attribute the differential timing to the affinity for the syllable margin of gestures having extreme constrictions).

This timing pattern involving the tongue tip and tongue dorsum (tongue rear) for American English laterals is illustrated in the movement of the relevant tongue pellets for "leap" and "peel." These utterances are shown in Figures 3(a) and 3(b), respectively, which display representative tokens of movement data collected using the X-ray microbeam system in Madison, Wisconsin. (Four productions each of "leap" and "peel" in the carrier phase "Give __ buttons," where the words "leap" and "peel" were accented, were collected from a female speaker of midwestern American English, approximately 40 years old. The speaker read the "leap" phrase and the "peel" phrase in alternation.)

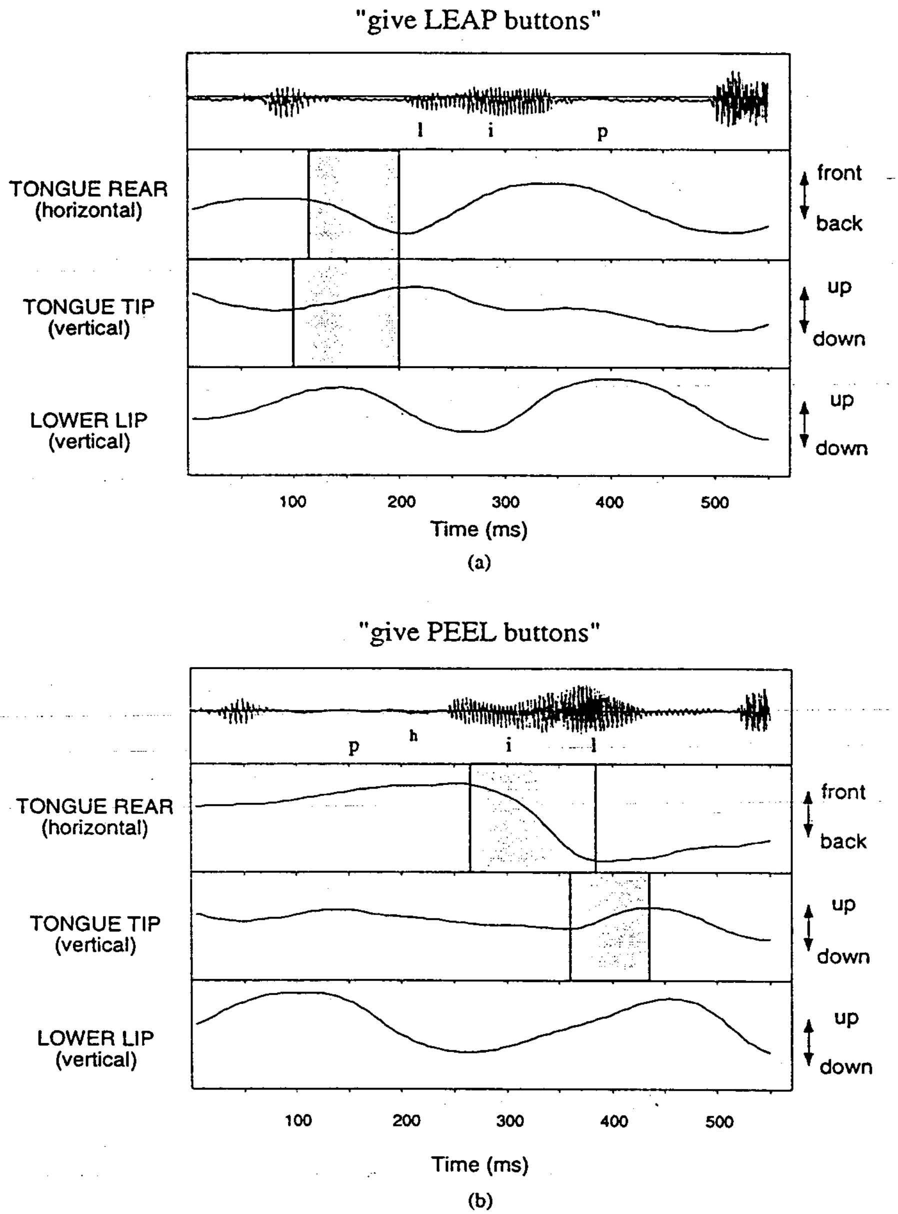


FIGURE 3. X-ray microbeam pellet data showing difference in relative timing of tongue tip raising and tongue dorsum (tongue rear) retraction for initial and final [l]. Tongue tip pellet is the frontmost pellet in the outlined image in Figure 1; tongue rear pellet is in the position indicated by the arrow in that image. (a) "leap" (b) "peel."

The shaded regions in Figure 3 highlight the movements presumably being controlled for the laterals. We would expect these regions to correspond roughly to the intervals of activation of the two gestures, as would be represented in the gestural score employed in Articulatory Phonology. Note there are two gestures present both syllable-initially and syllable-finally, that is, in both "leap" and "peel." The two gestures are roughly synchronous in "leap," but the dorsum gesture leads substantially in "peel." In fact, as can be seen in Figure 4 for all the tokens, in "leap" the end of the tongue dorsum retraction is nearly synchronous with the END of tongue tip movement (Figure 4(a)), while in "peel" the end of the tongue dorsum retraction is nearly synchronous with the BEGINNING of the tongue tip movement (Figure 4(b)). These patterns are analogous to the timing patterns discussed above for the nasals.

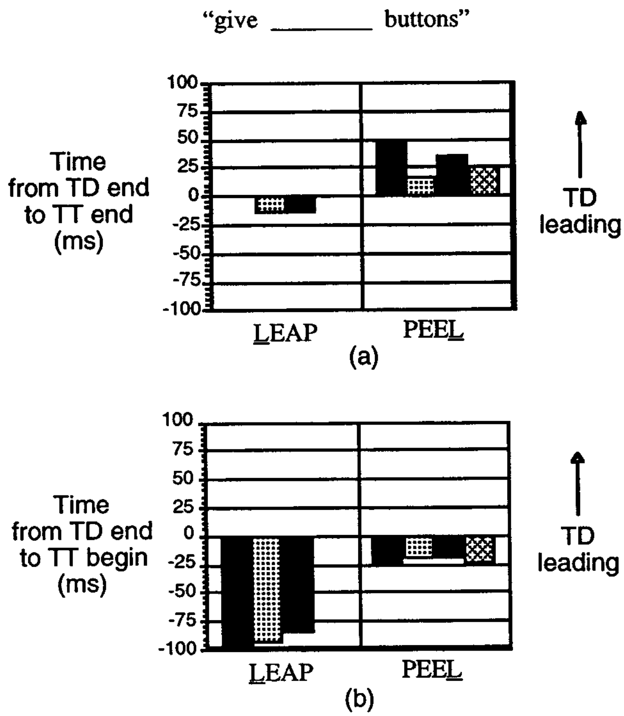


FIGURE 4. (a) Time from the end of the tongue-dorsum retraction (TD) movement to the end of the tongue-tip raising (TT) movement for "leap" and "peel." Each bar represents one token; there are three tokens of "leap" (the first of which has a time value of 0 ms) and four of "peel." (b) Time from the end of the tongue-dorsum retraction (TD) movement to the beginning of the tongue-tip raising (TT) movement for "leap" and "peel." Each bar represents one token; there are three tokens of "leap" and four of "peel." Movement beginning and end times for both (a) and (b) were automatically detected when the movement velocity for a given pellet/dimension increased above (beginning) or decreased below (end) a threshold of 30 mm/s.

Thus, for both the laterals and the nasals, it is the wider constriction that precedes the narrower constriction syllable-finally, as pointed out also by Sproat and Fujimura (1993). That is, for the nasal [m], the velum-lowering gesture has a wider constriction than the labial closing gesture, and this wider gesture precedes the narrower when syllable-final. Similarly, for the laterals, the wider tongue-dorsum constriction precedes the narrower tongue-tip closure syllable-finally. Gesturally there appears to be a single generalization here, applicable at least to both nasals and laterals (and perhaps to other gestures), namely that syllable-final linked gestures are phased so that the gesture with the wider constriction degree comes earlier, whereas syllable-initially these gestures are phased roughly synchronously. This single configurational principle results in different lawful acoustic consequences in the two examples—consequences which have been described as a nasalized vs. non-nasalized vowel, or a light vs. dark /1/.

The gestural generalization, of course, contrasts with the feature- and segment-based characterizations of syllable-position effects for nasals and laterals. American English syllable-initial and -final nasals are typically considered to differ in a featural approach only in that, in final position, the feature [+nasal] is extended to co-occur with the preceding vowel. That is, the same feature value is used for initial and final nasals (cf. Cohn, 1993). Laterals, however, are often considered to have different feature values initially and finally, so that, for example, American English initial and final laterals are often presumed to be [–back] and [+back], respectively (see discussion in Sproat & Fujimura, 1993, who also argue against this view). The syllable position similarity between nasals and laterals is totally missed in this kind of feature-based approach.

It is interesting to note that the tongue tip closure in American English final laterals can occur even in the silence following an utterance (Recasens & Farnetani, 1992). That is, the tongue tip articulation can occur even though vocalization has ceased. Such a situation might conceivably serve as an explanation for the errors in language acquisition in which a word ending with [1] is apparently mispronounced by the child as having a final [w] instead of a final [1] (e.g., in Dutch, /kas'tel/ 'castle' pronounced by a child of 2 years, 1 month as [tas'te:u], reported in Fikkert, 1992). In such a situation, presumably the inaudibility of the tongue tip closure for [1] would contribute causally to the interpretation of the final consonant as an approximant containing a dorsal gesture, i.e., [w]. One can speculate that similar inaudible articulations might be causally involved in those historical sound changes that involve the loss of a final nasal consonant, with concomitant nasalization of the preceding vowel. Presumably in such a case the velic opening and oral closure gestures would be timed as per the above discussion, with the velic opening gesture occurring during the vowel when syllable-final. That would mean, of course, that no additional process would need to be invoked to nasalize the vowel. If, as for [1], the closure gesture for the nasal could occur so late as to occur in utterance-final silence, then presumably the deletion of the closure gesture for the nasal would be a listener-based sound change (Ohala, 1981) caused by the inaudibility of the closure gesture when utterance-final.

FINAL REDUCTION

The perceptual deletion of the closure gesture could also be the result of its articulatory reduction. That is, regardless of what other gestures are involved, the position of the tongue tip is reduced in final position for [l] as well as for [t] and [n] (this appears to be a general positional effect, as we shall see later, although it is more prominent for articulations involving the tongue tip than for those involving other articulators). The cause of this reduction is not known, although it might be caused by a general reduction of speaking effort over the time course of the unit. As Browman and Goldstein (1992) have argued, this reduction can be stated as a single generalization about gestures, while in traditional phonological descriptions, a number of different types of rules (some selecting allophones, some specifying quantitative values) involving different features would be invoked.

In the microbeam data for "leap" and "peel" discussed above, reduction of the tongue-tip closure in syllable-final [l] can be observed. (Reduction for final-position [l] was found earlier by Giles & Moll, 1975.) Figure 5(a) shows the reduction of the tongue-tip component of syllable-final [l], by comparing the vertical positions of a gold pellet placed about 7 mm behind the tongue tip in each of the four repetitions of "leap" and "peel," measured at the maximum tongue-tip height for the [l]. In all cases, the tongue tip for the syllable-final [l] is lower than that for syllable-initial [l]. It is not the case that the tongue-tip gesture for syllable-final [l] is absent, since in most cases there is tongue-tip movement present, as is shown by the fact that tongue-tip height is generally above the baseline (which is the value of tongue-tip height that occurs when no tongue-tip gesture is present). That is, in general, these cases appear to be instances of syllable-final reduction of the tongue-tip movement. Figure 5(b) shows the same information when a low vowel, instead of a high vowel, occurs.

The reduction of the tongue-tip movement for syllable-final [l] is consistent with traditional transcriptions of variability in syllable-final [l], for example, those indicating the possibility of its "vocalization" or deletion in certain contexts, for some speakers (Bailey, 1985). However, this reduction is not specific to [l]: the tongue tip shows a reduction in the maximum height achieved finally in other gestural contexts as well—for example, in words like "pot" and "pawn," for which the traditional descriptions differ both from [l] and from each other (Bailey, 1985).

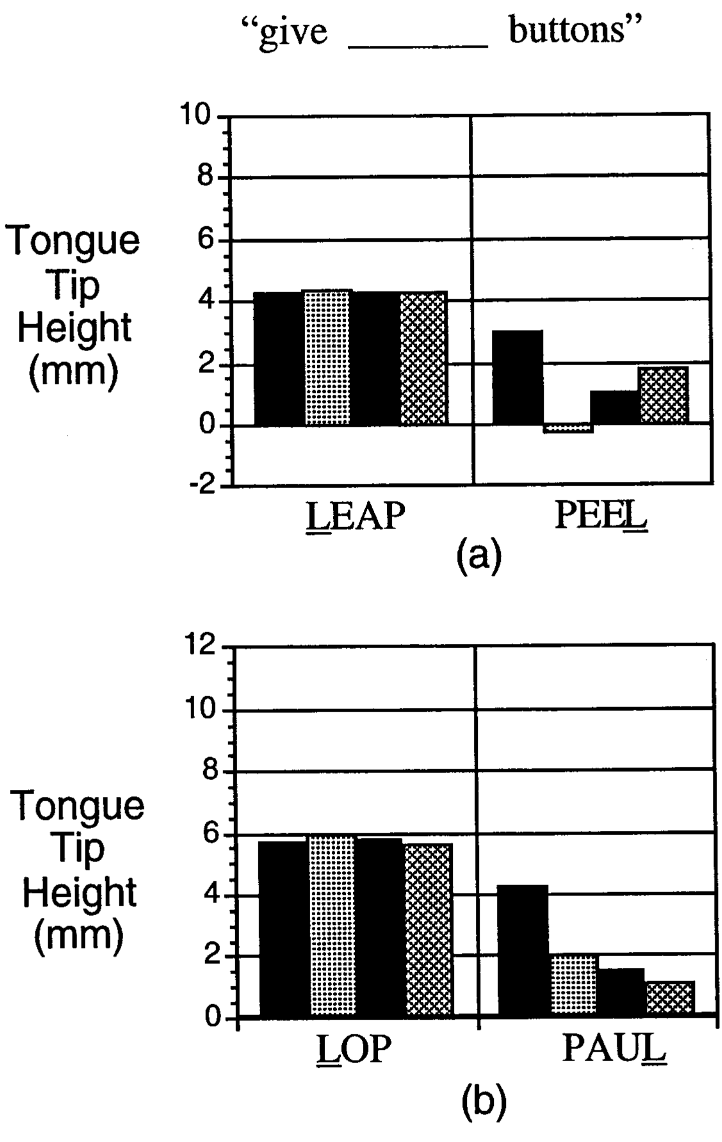


FIGURE 5. Reduction in tongue-tip height achieved for final [l]. Tongue-tip height was measured at the temporal center of a 1 mm noise band at the peak of the movement. Each bar represents one token. (a) "leap" vs. "peel." The tongue-tip height displayed for each token is relative to the baseline tongue-tip height (0 mm), which in this subfigure is the average maximum vertical tongue-tip position occurring during the final labial gesture in tokens of "leap" for this speaker. (b) "lop" vs. "Paul." Here the baseline (0 mm) is the average maximum vertical tongue-tip position occurring during the final labial gesture in tokens of "lop" rather than in tokens of "leap."

Again using the X-ray microbeam system in Madison, Wisconsin, we collected C1VC2 words in carrier phrases varying in accent and in the gesture following the target word. The speaker was a college-aged male speaker of California English. Figure 6 shows the results for five consecutive repetitions of "tot," "top," "pot," "Don," and "pawn," with the targeted word always accented and occurring in each of two contexts "(My BEAR doesn't huddle, but) my___ huddles" and "(My BEAR doesn't puddle, but) my ___ puddles."

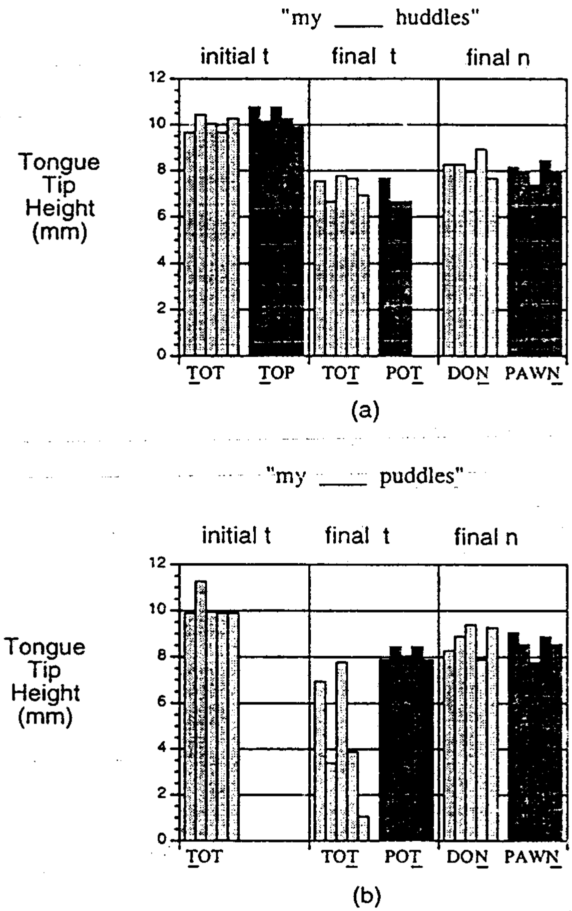
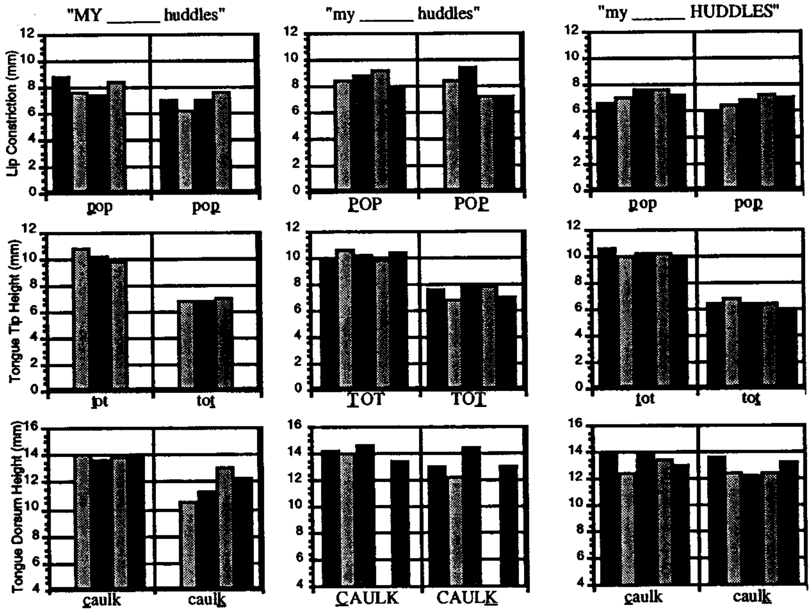


FIGURE 6. Reduction in tongue-tip height achieved for final [t] and [n]. Time points for tongue-tip height measurement were chosen as described for Figure 5. The baseline (0 mm) is the mean maximum vertical tongue-tip position occurring during the final labial gesture in tokens of "top" for this speaker in this accent condition. Each bar represents one token; all the utterances have five tokens except "pot huddles," which has three. (a) context "huddles," (b) context "puddles."

(The parenthetical portions of the sentences were shown to the speaker to elicit the desired accent placement, but were not actually produced by the speaker. Also, the speaker produced the same vowel in all the utterances, as judged by informal listening, which is not surprising since there is no [O] in the dialect of English spoken in California, as in many other dialects; i.e., the vowel in "pawn" is pronounced like the vowel in the second syllable of "upon.") The figure shows the final reduction in tongue-tip height of both [t] and [n], by comparing the maximum heights of a gold pellet (placed 10 mm behind the tongue tip) for the final tokens to the maximum heights of the same tongue tip pellet in initial [t].

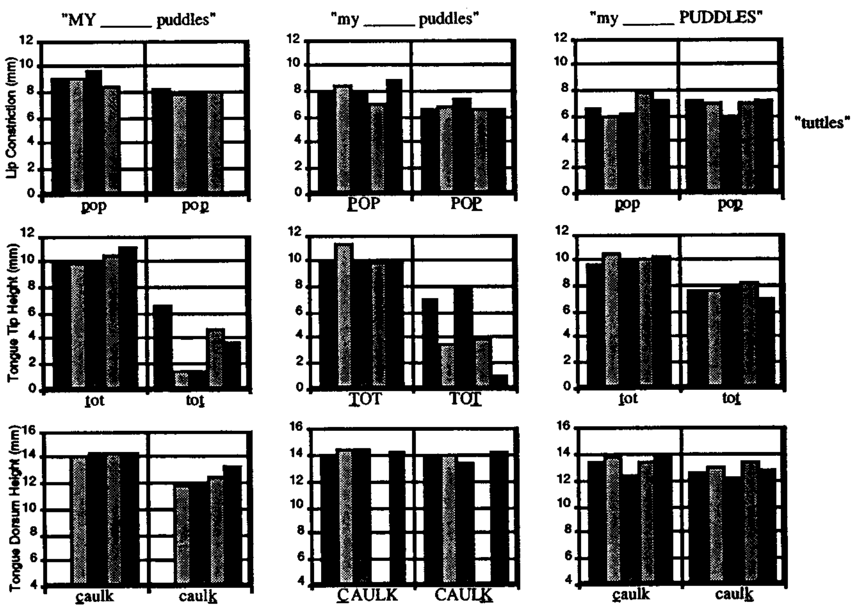
Looking first at Figure 6(a), in which the word immediately following does NOT begin with an oral closure gesture ("huddles"), it is clear that all the final tokens of [t] and [n] have reduced tongue-tip positions when compared to the initial [t] on the left. That the tongue-tip gestures are reduced but not deleted is indicated by the fact that all the positions are far above the baseline value of 0 mm, which represents the position of the tongue tip during a final labial stop. Similarly, in Figure 6(b), in which the word immediately following DOES begin with an oral closure gesture ("puddles"), the final tokens are reduced. Several of the tokens in "tot puddles" do behave somewhat differently, however, with two of the tokens being reduced even further, and a third token being effectively deleted. While there is something additional going on in the case of "tot puddles" (additional reduction and possibly also deletion), this does not change the fact that, in general, final tongue-tip gestures are reduced. Similar results were obtained with the other accent patterns that were also collected—where either the word before or the word after the targeted word was accented, e.g., "(YOUR __ doesn't huddle, but) MY __ huddles," and "(It may not DANCE, but) my __ HUDDLES." A subset of these results (for "tot" but not for "pot") is shown in the middle rows of Figures 7(a) and 7(b).

While tongue-tip movements are reduced the most, the reduction effect also occurs in gestures that involve other oral articulators. Figure 7 shows the results for some of the other target words employed in this experiment—"pop" and "caulk" (pronounced [kAk]), as well as for "tot." As can be seen in the top row of Figures 7(a) and 7(b) for labials, and the bottom row of Figures 7(a) and 7(b) for dorsals, gestures involving the lips and tongue dorsum are also reduced finally. The position effect is significant at the 0.01 level for each of [p], [t], and [k], using analysis of variance (with 3 factors: accent placement, following word starting with oral closure gesture or not, and position within the word/syllable, which was treated as a repeated measures factor). For coronals ($p < 0.01$) and marginally for dorsals ($p < 0.021$), the magnitude of final reduction also varies according to the accent placement, presumably because there is more final reduction when the accent is placed on the word preceding the word being measured. In general, then, oral closure gestures decrease in magnitude when in final position.



(a)

FIGURE 7. Reduction in position achieved of relevant articulators for final voiceless stops, with three different places of articulation and three different accent patterns. Top row shows reduction in lip constriction for the labials in [pAp]; the middle row reduction in the height of the tongue tip for the coronals in [tAt]; and the bottom row reduction in the height of the tongue dorsum for the dorsals in [kAk]. The three columns show reduction for three different patterns of accent placement with respect to the measured word: left, accent preceding; middle, accent on the measured word; right, accent following. Time points at which the measurements were made were all determined using the method described in the caption for Figure 5. The values displayed in the top 2 rows were computed as follows: Lip constriction was measured by finding the vertical distance



(b)

between the pellets on the upper and lower lips (Lip Aperture) and then, to invert the curve for visual compatibility with other rows, subtracting that distance from 30 mm, which is a rough estimate of the Lip Aperture during initial tongue-tip gestures for this speaker. Thus, the greater the value of lip constriction, the smaller the distance between the upper and lower lips. For tongue-tip height (middle row), the baseline (0 mm) is the mean maximum vertical tongue-tip position occurring during the final labial gesture in tokens of "top" for the appropriate accent condition. For tongue dorsum height, no special baseline is employed. (a) context "huddles," (b) context "puddles" (or "tuttles", following [pap]).

CONCLUSION

In this chapter we have attempted to demonstrate how, within a gestural perspective, the notion of syllable position corresponds to different modes of configuration of physical events (the gestures). Three implications of this view are worth summarizing. First, it is often possible to capture the phonological facts that make reference to syllable position without reifying any hierarchical syllable structure, because syllable position inheres in the patterns of gestural coordination. Second, the gestural approach allows a single generalization to be stated for some superficially unrelated cases involving nasals and laterals. These phenomena, which have been described as involving variation of phonological units as a function of syllable position (e.g., allophonic variation), can be better understood, we argue, as the lawful physical consequences of the different patterns of gestural overlap that in effect define the different syllable positions. Third, as also posited by Sproat and Fujimura (1993), there may be a single gestural generalization for English, namely that, syllable-finally, gestures involving wider constrictions precede those with narrower constrictions. While this chapter considered only relatively simple gestural constellations (involving two gestures), the regularity uncovered suggests that it might be rewarding to investigate the more complex codas of American English phonology (constellations involving multiple gestures, e.g., in "howled") to see what additional generalizations emerge, and to see what light can be shed on the distributional patterns of such structures when viewed as gestural constellations rather than being parsed into a string of segment-sized units.

It seems also to be the case that there is a very general reduction in the size of gestures in final position. Such a result is perhaps not surprising, given the general weakness of the syllable-final position in the languages of the world (many languages have been described as having no syllable-final consonants at all), but it is striking that such reduction occurs in English, a language with a sizable inventory of final consonants. Further investigation of this reduction effect should determine its precise domain (is it really syllable-final, or word-final, phrase-final?) and how the reduction is distributed over the component articulators that constitute a gesture's coordinative structure (do all the articulators cooperate to produce a less extreme gesture, or is the reduction caused by the activity of a single articulator, possibly the jaw?). Answers to such questions could provide insight both into the causes of the reduction effect, and into the ways it might be expected to show up in the phonology of languages.

ACKNOWLEDGMENTS

This work was supported by NSF grant DBS-9112198 and NIH grants HD-01994 and DC-00121 to Haskins Laboratories. Our thanks to the Yale School of Medicine, General Electric Company, and the Esther A. and Joseph Klingenstein

Fund for providing the capacity to obtain the Magnetic Resonance Images reported on herein. Thanks also to Alice Faber for comments on an earlier version.

REFERENCES

Bailey, C.-J. N. (1985). *English phonetic transcription.* Arlington, TX: Summer Institute of Linguistics.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3,* 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology 6,* 201-251.

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

Cohn, A. C. (1993). Nasalization in English: Phonology or phonetics. *Phonology, 10,* 43-81.

Fikkert, P. (1992). A prosodic account of truncation in child language. Presentation at 7th International Phonology Meeting, Krems, Austria, July 1992.

Giles, S. B., & Moll, K. L. (1975). Cinefluorographic study of selected allophones of English /1/. *Phonetica, 31,* 206-227.

Harris, K. S., & Bell-Berti, F. (1984). On consonants and syllable boundaries. In L. J. Raphael, C. B. Raphael, & M. R. Valdovinos (Eds.), *On consonants, and syllable boundaries* (pp. 89-95). Plenum Publishing Corporation.

Kahn, D. (1976). *Syllable-based generalizations in English phonology.* Bloomington, IN: Indiana University Linguistics Club.

Krakow, R. A. (1989). *The articulatory organization of syllables: a kinematic analysis of labial and velar gestures.* Unpublished doctoral dissertation. Yale University.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *The listener as a source of sound change* (pp. 178-203). Chicago: Chicago Linguistic Society.

Recasens, D., & Farnetani, E. (1992). Spatiotemporal properties of different allophones of /1/. Phonological implications. Presentation at 7th International Phonology Meeting, Krems, Austria, July, 1992 .

Rubin, P. E., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America, 70,* 321-328.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333-382.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /1/ and its implications for phonetic implementation. *Journal of Phonetics, 21,* 291-311.