

# Converging sources of evidence for dissecting articulatory movements into core gestures

**Suzanne E. Boyce**

*M.I.T., Cambridge, MA U.S.A.*

**Rena A. Krakow**

*Temple University, Philadelphia, PA and Haskins Laboratories, New Haven, CT, U.S.A.*

**Fredericka Bell-Berti**

*St John's University, Queens, NY and Haskins Laboratories, New Haven, CT, U.S.A.*

**Carole E. Gelfer**

*William Patterson College, Wayne, NJ, U.S.A.*

*Received 20th November 1989, and in revised form 2nd March 1990*

---

Explaining and modeling the process of coarticulation is one of the central tasks of speech research. At the same time, the literature on the temporal extent of anticipatory coarticulation has been characterized by conflicting findings. In this paper, we bring together data from a number of different articulatory studies to argue that many of these inconsistencies can be resolved by careful comparison between minimally contrastive contexts. In particular, we examine reports of “one-” and “two-stage” coarticulatory patterns, and suggest that the occurrence of these patterns can be accounted for by factoring in the effects of neighboring segments and suprasegmental variables such as speaking rate. Our conclusions support the coproduction view of coarticulation as the consequence of overlap between spatio-temporally stable, context-independent gestures.

---

Over the last few decades, the temporal extent of coarticulation—in particular, anticipatory coarticulation—has been an ongoing subject of research and debate. Resolution of the issue has remained elusive, however, largely because the results reported by various investigators have been inconsistent with each other and/or internally inconclusive. In previous work (Gelfer, Bell-Berti & Harris, 1989; Boyce, Krakow & Bell-Berti, 1990; Bell-Berti & Krakow, 1990), we have argued that experimental controls in the form of minimally contrastive contexts are underutilized in studies of coarticulation. In this paper, we will argue that many of the apparent inconsistencies in the literature can be resolved by careful examination of such controls and that the results lend strong support to the “coproduction” model of articulation.

Conceptually, coarticulation is defined as the influence of segmental context on the articulatory/acoustic realization of a target segment. It is assumed that,

because of perceptual or articulatory constraints on target and surrounding segments, there are limits on the temporal extent of coarticulation (see, for example, Lindblom, 1983; Manuel & Krakow, 1984). Thus, the problem of modeling coarticulation is often cast in terms of the question "what are the constraints on the temporal spread of coarticulation"?

The debate on this issue has focused on two particular speech production frameworks, the "look-ahead" or "feature-migration" models exemplified by Henke (1966), Daniloff & Moll (1968), Benguerel & Cowan (1974) and, in a revised version, Keating (1988), and the "coproduction" models of Fowler (1980), Bell-Berti & Harris (1981), Browman & Goldstein (1986), and Saltzman & Munhall (1989). In the "look-ahead" models, an articulatory planning component determines which movements will be required for upcoming segments and initiates them as soon as possible, barring any competing articulatory requirements. Although they differ in the level at which requirements are specified, with Henke (1966) defining specification in purely articulatory terms and Keating (1988) advocating specification at the level of phonological contrast, these models have certain basic assumptions in common. First, segments are assumed to be concatenated in non-overlapping time slots. Thus, if movement for a target segment is observed in a context segment, that movement is assumed to have "spread" into the domain of the context segment. Second, because every different context poses a different set of conditions, the complex of neuromuscular commands (the motor "plan") associated with the target segment, and consequently the time at which coarticulation begins, will change in a predictable manner according to context.

The coproduction models, on the other hand, assume that each segment has an associated articulatory control structure for instantiation of a particular feature. These core structures, or gestures, are posited to be (a) consistently present, and (b) relatively stable with regard to segmental context. It is assumed that the ordinary time course of an articulatory gesture extends both before and after the time that its effects dominate the acoustic signal and that much, if not most, coarticulation can be explained as resulting from local interactions between overlapping gestures. Thus, the underlying motor control structure for a particular segment remains essentially the same regardless of the phonetic identity of surrounding phones. In contrast to the look-ahead view, changes in observed patterns of movement in different contexts stem from local interactions between context and target gestures rather than from any change in the motor plan for the target segment. It should be noted that early versions of the look-ahead model made a distinction between anticipatory coarticulation, as requiring advance computation, and carryover coarticulation, as proceeding strictly from physical factors such as inertia. In the coproduction approach, this distinction is blurred, since both intentional activity and physical forces are subsumed under the notion of a characteristic gesture.

Experimental efforts to resolve the debate between these two classes of theory (look-ahead and coproduction) have shown inconsistent results. Based on studies of anticipatory lip-rounding, Benguerel & Cowan (1974), Daniloff & Moll (1968), Sussman & Westbury (1981), and Lubker (1981) concluded that the onset of rounding may vary depending on the identity of the context segment(s); in contrast, Bell-Berti & Harris (1974, 1979, 1982), Engstrand (1981), McAllister (1978), and Boyce (1988) concluded that the time course of rounding is stable over different phonetic contexts. Studies of anticipatory nasalization have led to similarly



inconsistent findings (cf. Moll & Daniloff, 1971; Ohala, 1971; Kent, Carney & Severeid, 1974; Bell-Berti & Krakow, 1990).

One explanation for the variable outcome of coarticulation studies was put forward by Bladon & Al-Bamerni (1982), who proposed that observed coarticulatory patterns might be a combination of anticipatory feature spread plus stable gestures and that the inconsistent results of previous studies might be due to differences in measurement technique that weighted the effects of one or the other more heavily. In addition, Bladon & Al-Bamerni proposed that speakers may vary randomly in how they implement anticipatory coarticulation. Observation of token-to-token variability in lip-rounding patterns led Perkell & Chiang (1986) (see also Perkell, 1986, and Chiang, 1987) to a similar conclusion. For these investigators, then, coarticulation for a target segment, while showing general similarities from token to token, is not predictable in detail from the segmental context.

A different explanation was put forward by Gelfer *et al.* (1989), who showed that much of what investigators have taken to be anticipatory movement for an upcoming target segment may be due to movement associated with surrounding phones. In this paper, we will expand the original argument of these authors by bringing together data from different studies, from the literature or our own research, that examine coarticulation by different measurement techniques. To allow direct comparison with the existing studies cited above, we have used measures of lip movement for rounding and associated electromyographic (EMG) data, as well as measures of velic height for nasality. Taken together, these data demonstrate that, when utterances containing minimal contrasts for the features of interest are examined, it is often possible to separate the effects of target and context segments. Moreover, many of the apparent inconsistencies in the literature can be resolved by this method. We will argue further that the notion of stable but overlapping gestures can account for much of the variability observed in coarticulatory behavior, given certain assumptions about the nature of overlap and the effect of suprasegmental factors such as speaking rate and stress.

One of the crucial assumptions made by investigators studying coarticulation is that segments may be articulatorily neutral (or "unspecified") for a particular feature. This assumption seems to be drawn from current phonological theories which maintain that segments are distinguished from one another at the underlying level by the specification of a few contrastive features (usually translatable as articulatory configurations in the vocal tract), rather than by a maximally detailed description. In the application of this principle to theories of speech production, it is generally assumed that what distinguishes segments from one another phonologically also dictates the parameters by which they are allowed to vary articulatorily. Normally, investigators have taken this to mean that as long as the conditions represented by the specified features are met, the rest of the vocal tract, for which features are not specified, will be free to vary. Thus, if a segment is unspecified at the point in speech production when abstract phonological units are translated into a motor plan, it will contribute nothing of its own to a trajectory between segments that are specified for that feature (Keating, 1988).

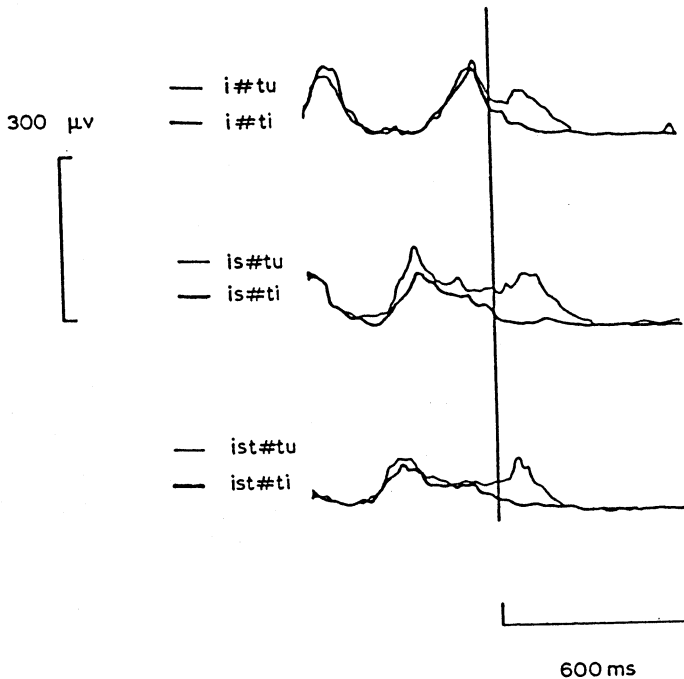
This assumption has been particularly important for studies testing the look-ahead model. Given a segment with a particular feature preceded by one or more segments regarded as neutral for that feature, researchers have tended to interpret the first sign of movement or muscle activity typical of that feature as an instantiation of that

segment. To take an example, in a sequence such as /is#tu/, where the feature of interest is lip-rounding, the first evidence of lip protrusion, or EMG activity in the orbicularis oris (OO) muscle, is assumed to be associated with the rounded vowel /u/. If lip protrusion or OO activity is observed during /i/, /s/, or /t/, then coarticulation for lip rounding is taken to have begun, and the time at which it occurs is taken to indicate the extent to which coarticulation can spread to earlier segments. This argument presumes, of course, that the consonants /s/ and /t/ are indeed neutral with respect to rounding. Similarly, in a sequence such as /sei#an/, where the feature of interest is nasality and the vowels /ei/ and /a/ are considered neutral for nasality, the first sign of velopharyngeal port opening, or velic lowering towards an open nasal port, is taken to indicate the temporal extent of anticipatory nasal coarticulation.

Practically speaking, however, when a single neutral segment precedes a segment which is specified for the feature of interest, it is hard to disentangle the look-ahead and coproduction models. This is because while the look-ahead models predict that articulatory onset will occur at the beginning of the neutral segment, the stable trajectory predicted by the coproduction model may begin at that point as well. To solve this problem, investigators have varied the number of supposedly neutral segments preceding a target segment that is positively specified for the feature of interest (e.g. Bell-Berti & Harris, 1982; Sussman & Westbury, 1981); for instance, in addition to examining velic lowering in a CVN sequence, they examine lowering in sequences such as CVVN or CVVVN, etc. If the onset time of the movement associated with the target segment is found to vary in proportion to the number and/or duration of the neutral segments, this is taken as support for the look-ahead model. If, on the other hand, a consistent time interval intervenes between movement onset and some other stable landmark (such as the onset of the nasal murmur or the beginning of oral contact for the nasal consonant), this is taken as support for the coproduction model.

Assumptions of articulatory neutrality have often been questioned, however. It is well known, for instance, that English consonants such as /s/, /l/, and /ɹ/, although not phonologically contrastive for rounding, may be produced with a rounded configuration in unrounded environments (Delattre & Freeman, 1968; Leidner, 1973; Brown, 1981). In addition, numerous investigators have reported evidence that the velum has characteristic positions for different vowels (Moll, 1962; Ushijima & Sawashima, 1972; Bell-Berti, 1980; Bell-Berti, Baer, Harris & Niimi, 1981). In many studies (e.g., Benguerel & Cowan, 1974; Daniloff & Moll, 1968; Lubker, 1981; Bell-Berti & Harris, 1981; Perkell, 1986; Perkell & Chiang, 1986; Chiang, 1987), investigators attempted to characterize subjects for gross articulatory activity associated with the feature of interest during so-called neutral segments. Such activity, if found, would normally be ascribed to a sub-phonemic (i.e., non-contrastive) specification for that feature. But, as noted by Gelfer *et al.* (1989), investigators have not attempted a comprehensive analysis of the less obvious characteristics of segments included in their experimental corpora before drawing conclusions concerning coarticulation. That effects uncovered by such an analysis may have significant import can be seen in the following sets of data.

Figure 1, reprinted from Gelfer *et al.* (1989), shows EMG data in the form of ensemble-averaged orbicularis oris inferior (OOI) traces, as spoken by a native speaker of American English. This figure shows muscle activity associated with



**Figure 1.** Ensemble-averaged OOI EMG activity for 15–20 tokens of three pairs of utterances with increasing numbers of “neutral” consonants. The vertical line indicates the lineup point for averaging tokens, which was at the acoustic onset of the second vowel.

rounding in utterances of the structure  $V \# CV$ ,  $VC \# CV$ , and  $VCC \# CV$ , where the first vowel is unrounded, the second vowel rounded, and the consonants various combinations of /s/ and /t/. As noted above, it is normally assumed that the first instance of lip movement or orbicularis oris muscle activity indicates the onset of motor execution for the rounded vowel (see also Bell-Berti & Harris, 1981; Lubker, 1981). Thus, the fact that, in /i#tu/, /is#tu/, and /ist#tu/, OOI activity seems to begin between 200 and 400 ms before the acoustic onset of the rounded vowel might suggest that rounding for /u/ actually begins at these times. (Association of these EMG activity peaks with protrusion movements was confirmed by reference to simultaneously recorded lip movement data.) At first glance, these results appear to support the look-ahead model. Although the look-ahead model would not predict that the OOI trace would split into an early and a late peak of activity (as it does in all three cases), the time interval between the first peak of OOI activity and the acoustic onset of the /u/ was longer for words with greater numbers of intervocalic consonants, such as /ist#tu/ and /ist#stu/, suggesting that anticipation begins at the first “neutral” consonant.

As Gelfer *et al.* point out, however, this early movement and EMG activity is echoed in the traces for words where /u/ has been replaced with /i/, as in /is#ti/. Since no rounded vowel is present, the early activity must be attributed to sub-phonemic lip movement for either or both of the intervening consonant(s), and the two peaks of activity seen in, for example, /is#tu/, must be assigned separately to the intervocalic consonants and to the rounded vowel. The fact that EMG activity

is evident with the first consonant of the consonant string can no longer be viewed as unambiguous evidence of coarticulatory anticipation; rather, it reflects the characteristics of the intervocalic consonant(s). (Note that the second subject in this study showed similar patterns.)

A similar demonstration may be made in the case of what are known as "one-stage" and "two-stage" movement onsets for nasal and rounding coarticulation. Bladon & Al-Bamerni (1982) compared nasograph (Ohala, 1971) traces for various utterances incorporating one to three oral vowels followed by a nasal consonant. They found two patterns of velopharyngeal port opening, (a) a "one-stage" pattern beginning early in the vowel string and continuing smoothly to a maximum open position during the nasal consonant; and (b) a "two-stage" pattern, consisting of slight port opening during the vowel string followed by an abrupt high-velocity opening stage beginning just before the nasal consonant. Figure 2, adapted from Bladon & Al-Bamerni (1982), shows these two patterns, which occurred in different repetitions of the Kurdish sequence [tagu:tse:ɛ:ma:kha:]. The authors pointed out that timing for the "one-stage" pattern conformed to the predictions of the look-ahead model of coarticulation (i.e., port opening onset occurred at the acoustic onset of the neutral segment /e:/). However, with regard to the two-stage pattern, they argued that while the early stage conformed in timing to the predictions of the look-ahead model, the second stage conformed to the predictions of the coproduction model (i.e., the onset of the high-velocity stage occurred at a relatively fixed time in relation to the nasal murmur). Especially puzzling was the finding that, as Fig. 2 shows, a given speaker might produce either pattern during different repetitions of the same utterance. Bladon & Al-Bamerni were unable to discover any factors which would predict when the different patterns

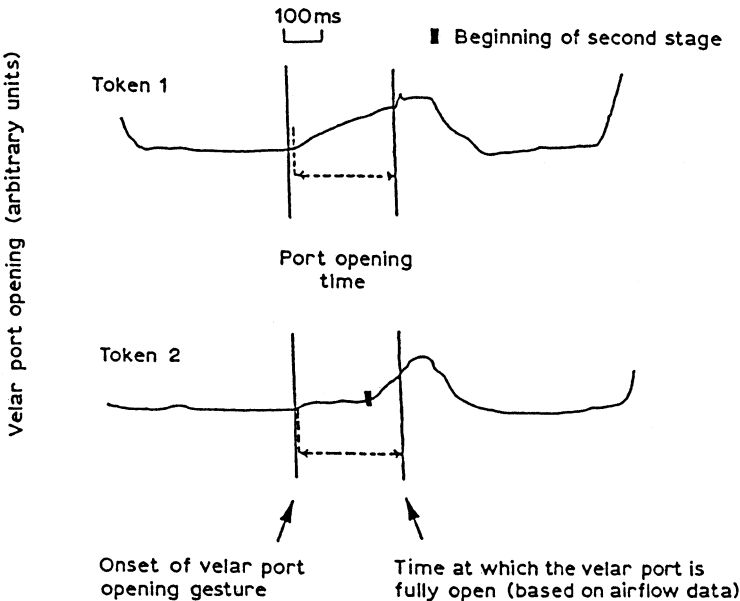
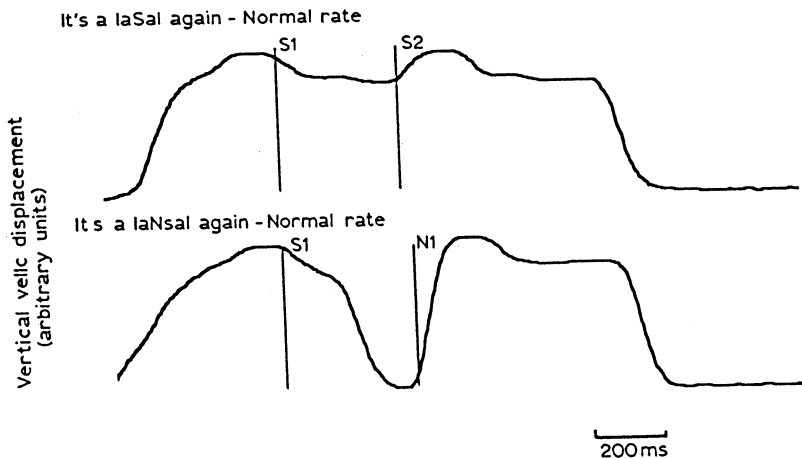


Figure 2. Nasograph traces for two tokens of [tagu:tse:ɛ:ma:kha:] as spoken by a native speaker of Kurdish. Adapted from Bladon & Al-Bamerni (1982).

occurred; accordingly, they suggested that the occurrence of one- and two-stage patterns might be random. Further, they proposed that findings in the literature supporting one model or the other were probably a function of measurement criteria emphasizing either the first or second stage onset of the two-stage pattern.

One possibility not explored in Bladon & Al-Bamerni's (1982) study was that early nasalization might be due to intrinsic, subphonemic velar position for these posited "neutral" segments. That the two-stage pattern owes its shape to a combination of intrinsic positions for the vowels as well as the nasal consonant has been argued by Bell-Berti & Krakow (1990). In what follows, we review some of their arguments. We will also suggest that at least some of the seemingly random alternations between single and multi-stage movement patterns are a function of suprasegmental variations leading to differences in the extent to which vocalic and consonantal gestures overlap in time.

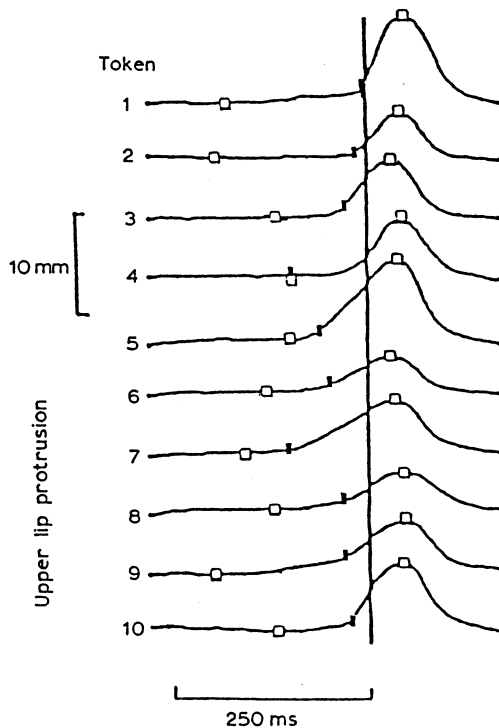
Figure 3 shows velic movement data, as recorded with the Velotrace (Horiguchi & Bell-Berti, 1987), for characteristic tokens of two utterances, /*uts* # *ʌ* # *lasal* # *ʌ*gen/ ("It's a /*lasal*/ again") and /*uts* # *ʌ* # *lansal* # *ʌ*gen/ ("It's a *lansal* again"), spoken by a native speaker of American English. The tokens were aligned at the acoustic release of the /*s*/ in "It's" for both utterances. (Note that for Velotrace data in this paper, a low position on the vertical scale indicates a low position of the velum. This is in contrast to the figure from Bladon and Al-Bamerni's study, in which a high position in the Nasograph trace indicates an open velopharyngeal port, and thus a low velic position.) As can be seen, the utterance with a nasal consonant—"It's a /*lansal*/ again"—appears to show two separate stages of movement leading up to the nasal, a pattern similar to that reported by Bladon & Al-Bamerni (1982). What appears to be the "first stage" of the velic lowering movement begins roughly at the release of the /*s*/ in "It's"; the "second stage" begins during the sequence /*ʌ*l $\alpha$ /. Looking at the trace for the oral utterance "It's a /*lasal*/ again", however, we see that it is nearly identical to that for "It's a /*lansal*/ again" up to the



**Figure 3.** Velotrace traces for oral and nasal versions of "It's a /*lɑ*(*n*)*səl*/ again", spoken at a normal conversational rate. Vertical lines labelled S1 indicate the acoustic offset of the /*s*/ frication in "It's", S2 indicates the onset of the /*s*/ frication in "/*lɑ*(*n*)*səl*/" and N1 indicates the onset of the nasal murmur for/*n*/. Tokens were aligned at S1.

point at which the so-called second stage begins. That is, "It's a lalsal again" shows a slight velic lowering movement that corresponds nearly exactly in time and amplitude with the early portion of the "two-stage" movement in "It's a lalsal again". Lowering of this type in the transition between consonants and vowels has been described previously for oral sequences (Kent *et al.* 1974). Thus, it seems highly likely that the first stage movement involved in "It's a lalsal again" is (a) characteristic of the sequence /sɔlɔ/, at least for this speaker, and (b) not related to anticipatory velic lowering for the nasal consonant. We conclude from this that the two-stage movement results, not from anticipation of velic lowering during an unspecified segment, but from combination of segments with consistent individual articulatory specifications for velic height.

The notion of "two-stage" anticipatory coarticulation was also examined by Perkell & Chiang (1986) using upper lip protrusion data for sequences such as /li#kut/, /li#sut/, /lis#kut/ with up to four medial consonants, as spoken by four native speakers of English. They partitioned each movement (whether or not it had an obvious "two-stage" shape) into two components: (a) a first stage, stretching between the first sign of protrusion (the beginning of positive velocity) preceding the acoustic onset of the /u/ and the point of maximum acceleration in that interval, and (b) a second stage, from the acceleration peak to the protrusion peak (end of positive velocity). (Typically, in a trace with a visually evident two-stage pattern,



**Figure 4.** Upper lip protrusion data obtained by strain-gauge for ten tokens of /li#sut/. The vertical line indicates acoustic onset of the second vowel. Adapted from Perkell & Chiang (1986). □: beginning and end of protrusion; ■: point of maximum acceleration.

peak acceleration tends to co-occur with the beginning of the second stage.) Figure 4, adapted from Perkell & Chiang (1986), illustrates upper lip protrusion traces from several tokens of the representative utterance /li#sut/, as spoken by a single speaker. As this figure shows, a wide range of variability was found in the onset of protrusion with regard to the acoustic onset of /u/. Timing between the onset of /u/ and the maximum acceleration of the lip protrusion movement was also variable. The authors concluded that, in spite of frequent similarities in the sizes and shapes of the second stage of the protrusion movement among the different tokens, there was still too much variability between the times of maximum acceleration and acoustic vowel onset, and between vowel onset and protrusion peak, to demonstrate the existence of a stable core gesture. At the same time, variability in the timing of the "first stage" onset *vis-à-vis* acoustic offset of the /i/ was greater than predicted by look-ahead models. The best predictor of movement patterns was the identity of the intervocalic consonant sequence, leading Perkell & Chiang to assign a greater role to the characteristics of supposedly "neutral" consonants than the look-ahead models had originally proposed.

Perkell & Chiang concluded from these findings that neither the look-ahead model nor the coproduction model could be correct in their strong form. Instead, they proposed a "hybrid" model, in which the characteristics of movements are determined in part by their segmental goals, and in part by token-specific interactions among acoustic, aerodynamic and biomechanical factors associated with neighboring segments. Implicit in this conclusion is the assumption that since such factors are minutely balanced from moment to moment during speech production, their effect on the timing and spatial characteristics of individual gestures cannot be precisely predicted. This conclusion was affirmed in a later study involving more subjects (Chiang, 1987).

Interestingly, in some respects Perkell & Chiang's hybrid model makes predictions of variability that are little different from those of the coproduction model. In the hybrid model, the point at which a gesture appears to start will change as a function of the constraints imposed by its context. In the coproduction model, the notion that gestures overlap and combine means that the point at which one gesture appears to dominate the signal is also dependent on its context. This is illustrated schematically in Fig. 5, which shows, in the left hand panel, two theoretical gestures associated with two overlapping segments, and, in the right hand panel, a smoothed trajectory representing their articulatory output as a "two-stage" pattern. The point at which the skirts of the two gestures overlap is, approximately, where we would expect to identify the beginning of the second stage.<sup>1</sup> If we think of the later gesture as protrusion for an /u/ vowel and of the earlier gesture as belonging to a preceding consonant, it is readily apparent that the point at which protrusion appears to begin will vary depending upon the size and shape of the consonant gesture, and its timing with respect to the vowel gesture. If the consonant gesture also involves protrusion, the apparent beginning of the "first stage" will depend on these characteristics of the individual consonant. Similarly, the point at which the "second stage" emerges from the "first stage" will depend on the characteristics of the overlap between the two

<sup>1</sup> Note that although the "two-stage" pattern is schematized in Fig. 5 as an interaction between two gestures, as a practical matter the multi-stage nature of the trajectory may not be apparent unless at least two "neutral" segments precede the target segment. Presumably, this is because the adjacent target and "neutral" segment gestures overlap in time so closely that their merger is not apparent.

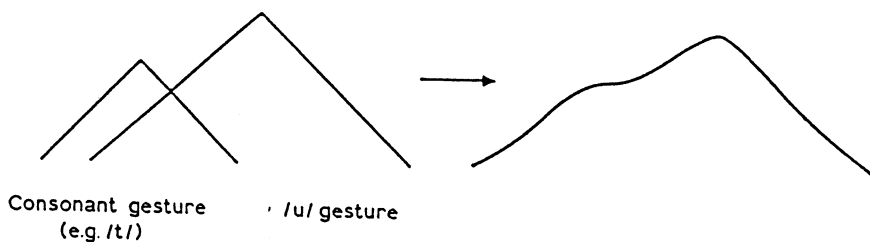


Figure 5. Schematized drawing of hypothesized consequences of gestural overlap.

gestures. Thus, Perkell & Chiang's conclusion that these aspects of the movement trajectory are constrained by context is also predicted by the coproduction model.<sup>2</sup>

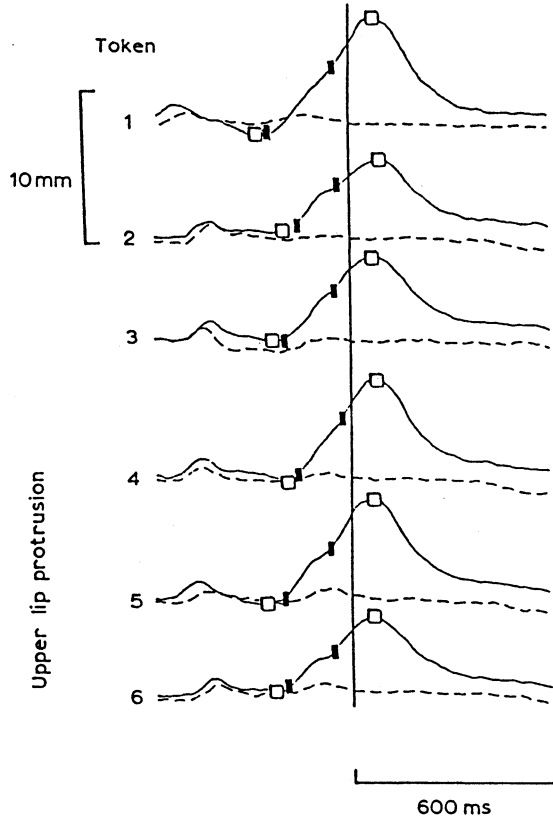
To show how variability of this type can be teased out from movement patterns, some data from Boyce (1988) are shown below. Figure 6 shows upper lip protrusion movement for six tokens of the nonsense words /kiktluk/ and /kiktlik/, embedded in the carrier phrase "It's a — again". The data were recorded from a native American English speaker using a modified Selspot system (Kay, Munhall, Vatikiotis-Bateson & Kelso, 1985). Tokens of /kiktluk/ and /kiktlik/, collected at different points during the experiment, are paired by similarity in shape. Both /kiktluk/ and /kiktlik/ tokens are lined up at the acoustic onset of the second vowel. To provide a parallel analysis with that of Perkell & Chiang (1986), the /kiktluk/ traces are marked for protrusion onset (onset of positive velocity), protrusion peak (onset of negative velocity), and for peak acceleration. There were typically two acceleration peaks of similar magnitude for /kiktluk/ tokens in the interval between the onset of protrusion and its peak, and both are shown.

In these data, the /kiktluk/ traces show two peaks of protrusion. The first peak occurs considerably before the acoustic offset of /i/ and may be related to the /s/ of "It's" in the carrier phrase. It is followed by a dip in the signal whose minimum occurs slightly prior to the acoustic offset of /i/ and generally coincides with the onset of positive velocity. The first maximum acceleration point occurs soon after. The main protrusion peak occurs approximately 100 ms after the acoustic onset of the /u/ vowel. Typically, the /u/ protrusion shows a second inflection between the protrusion onset and peak. The second maximum acceleration point occurs at this time.

Looking at the /kiktlik/ trace, however, we see that some portion of the protrusion movement in /kiktluk/ occurs in the absence of the rounded vowel. In particular, the inflection at the second maximum acceleration point, which identifies the beginning of the second stage, is often temporally aligned with a clear peak in the corresponding /kiktlik/ trace. In addition, protrusion onset for the /u/ in /kiktluk/, as determined by the calculation of positive velocity, appears to coincide with a parallel movement occurring in /kiktlik/. Probably, the minimum in the lip protrusion trace reflects a retraction movement for the /i/, whereas the following peak is associated with one or more of the intervening consonants. Thus, at least part of the first stage protrusion movement seen in /kiktluk/ tokens appears to be due to the consonant sequence found in both /kiktlik/ and /kiktluk/, rather than

<sup>2</sup> The issue of how overlapping gestures interact with one another is outside the scope of this paper. For further discussion of this issue, see Boyce (1988), Munhall & Löfqvist (1990), and Saltzman & Munhall (1989).





**Figure 6.** Upper lip protrusion traces, obtained by Selspot, for six pairs of tokens paired by similarities in shape and timing. The vertical line indicates acoustic onset of the second vowel. Baselines have been adjusted to facilitate comparison and do not necessarily reflect experimental baselines. —: kiktluk; ---: kiktlik; □: beginning and end of protrusion; ■: point of maximum acceleration.

being fully attributable to the rounding gesture for the /u/. (Similar demonstrations using upper and/or lower lip data can be made for the seven other subjects in Boyce's study.)

It is clear from these data that the time at which protrusion appears to begin cannot be assumed, *a priori*, to reflect the time course of the underlying protrusion gestures associated with /u/. Rather, it is necessary to identify the articulatory movements that are characteristic of each segment in a sequence before drawing any conclusion about the beginning of coarticulation. As in the velic movement data, the beginning of the second stage may reflect the boundary of interaction between gestures for adjacent segments.

One question that remains is how to account for the seemingly random incidence of one- *vs.* two-stage movements for different repetitions of the same utterance. In Bell-Berti & Krakow (1990), it was proposed that changes in speech rate might lead to alternations between simple and multi-stage movements. In what follows, we review their arguments and later we will suggest that, in addition to rate, stress may

influence the shape of articulatory movements such that they appear as “one”- or “two-stage” (cf. Krakow, 1986).

Figure 3 (above) shows movement traces from Bell-Berti & Krakow (1990) for tokens of / its# ʌ# lasal# ʌgen/ (“It’s a lasal again”) and / its# ʌ# lansal# ʌgen/ (“It’s a lansal again”), produced at a normal conversational rate. Figure 7 shows corresponding productions of the same utterances produced by the same subject at a self-selected “rapid rate”. As can be seen, the rapidly produced “It’s a lansal again” (seen in Fig. 7) contains only a single velic lowering movement in the vicinity of the nasal consonant; in contrast, the normally produced token (seen in Fig. 3) contains two component lowering movements. One way to account for this is to say that in the faster utterance, as in the slower utterance, there are independent (underlying) gestures for the vowel and the nasal consonant. They appear as a single movement in the faster rendition because there is not enough time for the two movements to occur in temporally separate space.

This interpretation is strengthened by the fact that the rapidly produced oral utterance, “It’s a lasal again” (seen in Fig. 7) shows a small lowering movement following the /s/ of “It’s”, suggesting that even in fast speech there is a small but independent velic gesture associated with the oral sequence. This evidence, together with the clear case of separate “oral” and “nasal” velic movements in the normal rate utterances of Fig. 3, suggests that the rapid utterance of “It’s a lansal again” contains two gestures combined in overlapping fashion. Additionally supportive is the fact that some of the more slowly produced rapid tokens of “It’s a lansal again” vocalic and consonantal gestures showed two-stage lowering patterns (Bell-Berti, 1990). While the rapid productions of this sequence contained some instances of one-stage movements and some of two-stage movements, the slower conversational rate productions always showed two-stage movement patterns.

One major similarity between the factors of speaking rate and number of “neutral” segments is the fact that both can be modeled as a continuous function of

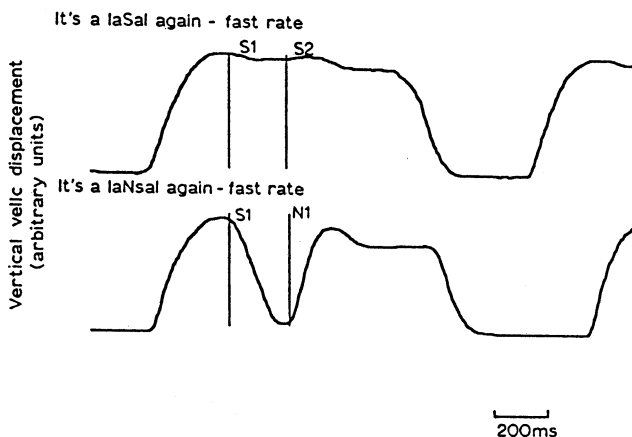
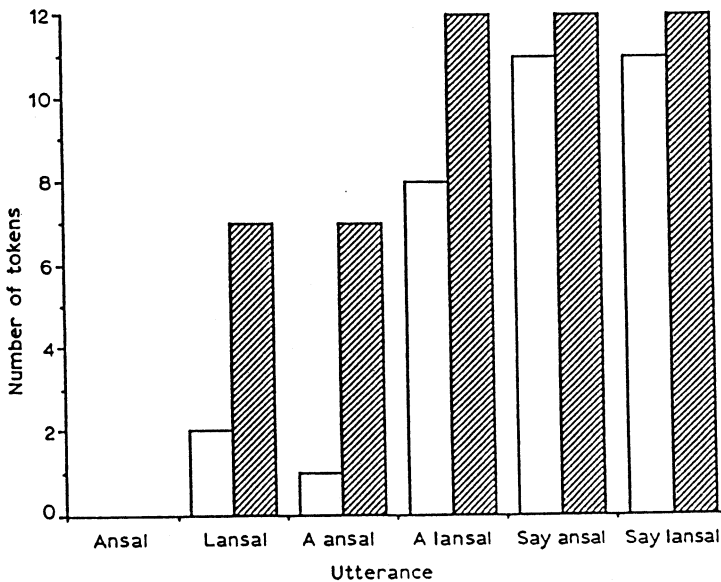


Figure 7. Velotrace traces for oral and nasal versions of “It’s a la(n)sal again”, spoken at a fast rate. Vertical lines labelled S1 indicate the acoustic offset of the /s/ frication in “It’s”, S2 indicates the onset of the /s/ frication in la(n)sal and N1 indicates the onset of the nasal murmur for /n/. Tokens were aligned at S1.

gestural overlap. As was pointed out in the discussion of Fig. 5, small variations in the way the gestures overlap may have significant consequences in terms of whether a movement appears to follow a "one"- or "two-stage" pattern. Bell-Berti & Krakow, in fact, presented additional evidence supporting the continuous nature of the "one"- vs. "two-stage" phenomenon as a function of speaking rate and number of "neutral" segments. This is shown in Fig. 8, which summarizes the results of manipulating the two factors, for the subject whose movement traces are shown in Fig. 7. Increasing the duration of the "neutral" string preceding the nasal consonant, whether by adding vocalic segments, or by slowing the rate, led to an increase in the incidence of multi-stage patterns. (See also Löfqvist (1989) and Munhall & Löfqvist (1990) for similar results using rate change in laryngeal opening/closing movements.)

Given these data, and additional data cited in Bell-Berti & Krakow (1990), we would like to suggest that variability in speech rate (whether due to fatigue, rushing, list effects, attitude, etc.) may lead to alternating patterns of one- vs. two-stage movements for different tokens of the same utterance containing either a nasal consonant or a rounded vowel. We would suggest, further, that such patterns can best be accounted for with the coproduction notion of overlap. In this model, even small variations of speaking rate within the categories of "fast" or "normal" would produce variation in the extent to which the gestures are overlapped, and thus variation in their apparent complexity of output, i.e., single- vs. multi-stage pattern of movement. (Note that, due to insufficient data, we are unable to determine if Bladon & Al-Bamerni's example of one- vs. two-stage movements is related to a difference in rate between tokens. That is, we were unable to consider either



**Figure 8.** Histogram showing number of tokens (out of a possible twelve) showing a multi-stage pattern, for fast and normal rate productions of five utterances containing nasal consonants preceded by "neutral" segments. Utterances are arranged left to right according to the number or duration of "neutral" segments. □: fast; ▨: normal.

acoustic or articulatory duration characteristics in detail since only those two tokens were shown and no acoustic traces or durations were provided.)

In the framework we present here, the addition of "neutral" segments and an decrease in speaking rate both favor the occurrence of multi-stage articulatory patterns, because they allow individual gestures to emerge as distinct entities. As a note of interest, evidence that stress may have a similar effect can be found in Krakow (1986). In her data, multi-stage velic raising movements were more often observed in stressed syllables containing a nasal consonant than in matched unstressed syllables. It may be that something about stress, perhaps the increased duration that characterizes stressed sequences, creates an environment in which discrete gestures are more likely to emerge.

To summarize, in this paper we have re-examined some of the previous conflicting evidence concerning the domain of coarticulation and its underlying mechanisms of control. We have attempted to show (a) that multi-stage patterns may be analyzed (at least to some extent) as the product of multiple independent gestures, and (b) that much of the variability in the incidence of single- and multi-stage patterns may proceed from variations in numbers of segments or from suprasegmental factors. This is not to say that we have accounted for all sources of variability in speech movement data. For example, Kent *et al.* (1974) reported that when they asked two speakers to produce a sentence at normal and fast rates, two different strategies for speeding up were employed. One subject showed a pattern which we would consider to be consistent with a simple strategy of greater gestural overlap with increased rate. That is, for fast productions, this speaker never attained as extreme high or low velic positions as he did in his slower productions. The other subject managed to attain the same extent of displacement in normal and fast productions by increasing the velocity of velic movement in the faster rendition. (This is consistent with a strategy of minimizing the degree of overlap.) In spite of the difference in strategy, however, both subjects evidenced a high degree of stability in their productions; when the fast and slow trajectories were overlaid in a time-normalized manner for each subject, the peaks (for the oral consonants) and the valleys (for the nasal consonants) coincided. Thus, the co-production model must allow for alternative stable gestural patterns.

In this paper, we have attempted to show how, given the assumption of overlapping core gestures, many of the inconsistencies in previously reported results, as well as more fundamental problems in speech production theory, may be accounted for within the coproduction framework. In particular, combining the coproduction framework with the notion of variability in the phasing of gestures (as might, for example, be induced by rate changes) provides an account in which single- and multi-stage gestures can be derived from the same underlying articulatory control pattern.

This research was supported by NIH grants NS-13617 and BRS RR-055996 to Haskins Laboratories and by NIH grant NS0-7040-15 to M.I.T. We would like to acknowledge the helpful comments of Catherine Browman, Jan Edwards, Marie Huffman, Sharon Manuel, Joe Perkell, and Elliot Saltzman, which greatly improved the quality and clarity of our paper.

### References

- Bell-Berti, F. (1980) Velopharyngeal function: A spatial-temporal model. In *Speech and language: advances in basic research and practice* (N. J. Lass, editor), vol. 4, pp. 291-316. New York: Academic Press

- Bell-Berti, F. & Harris, K. S. (1974) More on the motor organization of speech gestures, *Haskins Laboratories Status Report on Speech Research*, SR 37/38 pp. 73-77
- Bell-Berti, F. & Harris, K. S. (1979) Anticipatory coarticulation: some implications from a study of lip rounding, *Journal of the Acoustical Society of America*, 65, 1268-1270
- Bell-Berti, F. & Harris, K. S. (1981) A temporal model of speech production, *Phonetica*, 38, 9-20
- Bell-Berti, F. & Harris, K. S. (1982) Temporal patterns of coarticulation, *Journal of the Acoustical Society of America*, 71, 449-454
- Bell-Berti, F. & Krakow, R. A. (1990). Anticipatory velar lowering: a coproduction account. Haskins Laboratories Status Report on Speech Research. New Haven, CT.
- Bell-Berti, F., Baer, T. Harris, K. S. & Niimi, S. (1981) Coarticulatory effects of vowel quality on velar function, *Journal of Phonetics*, 36, 187-193
- Benguerel, A. P. & Cowan, H. A. (1974) Coarticulation of upper lip protrusion in French, *Phonetica*, 30, 41-55
- Bladon, R. A. W. & Al-Bamerni, A. (1982) One-stage and two-stage temporal patterns of velar coarticulation, *Journal of the Acoustical Society of America*, 72, S104(A)
- Boyce, S. E. (1988) The influence of phonological structure on articulatory organization in Turkish and English: vowel harmony and coarticulation, doctoral dissertation, Yale University Department of Linguistics
- Boyce, S. E., Krakow, R. A. & Bell-Berti, F. (1990) Phonological underspecification and speech motor organization, submitted to *Phonology*
- Browman, C. P. & Goldstein, L. M. (1986) Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252
- Brown, G. (1981) Consonant rounding in British English: The status of phonetic descriptions as historical data. In *Towards a history of phonetics* (Asher, R. E. and E. J. A. Henderson, editors), pp. 67-76. Edinburgh: Edinburgh University Press
- Chiang, C.-M. (1987) Models of coarticulation, B.S. thesis, M.I.T. Department of Engineering and Computer Science
- Daniloff, R. & Moll, K. (1968) Coarticulation of lip rounding. *Journal of Speech and Hearing Research*, 11, 707-721
- Delattre, P. & Freeman, D. (1968) A dialect study of American r's by x-ray motion picture, *Linguistics*, 44, 29-68
- Engstrand, O. (1981) Acoustic constraints of invariant input representation? An experimental study of selected articulatory movements and targets, *Reports from the Uppsala University Department of Linguistics*, 7, 67-94
- Fowler, C. A. (1980) Coarticulation and theories of extrinsic timing control, *Journal of Phonetics*, 8, 113-133
- Gelfer, C. E. Bell-Berti, F. & Harris, K. S. (1989) Determining the extent of coarticulation: effects of experimental design, *Journal of the Acoustical Society of America*, 86, 2443-2445
- Henke, W. L. (1966) Dynamic articulatory model of speech production using computer simulation, doctoral dissertation, M.I.T. Department of Electrical Engineering and Computer Science
- Horiguchi, S. & Bell-Berti, F. (1987) The Velotrace: A device for monitoring velar position. *Cleft Palate Journal*, 24, 104-111
- Kay, B. A., Munhall, K. G., Vatikiotis-Bateson, E. & Kelso, J. A. S. (1985). A note on processing kinematic data: Sampling, filtering and differentiation, *Haskins Laboratories Status Report on Speech Research*, SR 37/38, 73-77
- Keating, P. A. (1988) Underspecification in phonetics, *Phonology*, 5, 3-29
- Kent, R. D., Carney, P. J. & Severeid, L. R. (1974) Velar movement and timing: evaluation of a model for binary control, *Journal of Speech and Hearing Research*, 17, 470-488
- Krakow, R. A. (1986) Prosodic effects on velic movement. Paper presented at the meeting of the Linguistic Society of America, New York.
- Leidner, D. R. (1973) An electromyographic and acoustic study of American English liquids, doctoral dissertation, University of Connecticut
- Lindblom, B. (1983) The economy of speech gestures. In *The production of speech*, (P. F. MacNeilage, editor), pp. 217-245. New York: Springer-Verlag
- Löfqvist, A. (1989) Speech as audible gestures. Paper presented at the NATO ASI conference on speech production and speech modelling, Bonas, France
- Lubker, J. F. (1981) Temporal aspects of speech production, *Journal of Phonetics*, 38, 51-65
- Manuel, S. Y. & Krakow, R. A. (1984) Universal and language-particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, SR 77/78, 69-78
- McAllister, J. (1978) Temporal asymmetry in labial coarticulation, *Working Papers, Stockholm Institute of Linguistics*, 35, 1-29
- Moll, K. (1962) Velopharyngeal closure on vowels, *Journal of Speech and Hearing Research*, 5, 30-37
- Moll, K. & Daniloff, R. G. (1971) Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, 50, 678-684
- Munhall, K. & Löfqvist, A. (1990) Gestural aggregation in speech, submitted to *Journal of Phonetics*

- Ohala, J. J. (1971) Monitoring soft palate movements in speech, *Project on Linguistic Analysis*, 13, University of California Berkeley, JO1-JO15
- Perkell, J. S. (1986) Coarticulation strategies: preliminary implications of a detailed analysis of lower lip protrusion movements, *Speech Communication*, 5, 47-68
- Perkell, J. S. & Chiang, C.-M. (1986) Preliminary support for a "hybrid model" of anticipatory coarticulation. In *Proceedings of the 12th international congress of acoustics*, July, A3-6
- Saltzman, E. L. & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, 1, 333-382
- Sussman, H. M. & Westbury, J. R. (1981) The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation, *Journal of Speech and Hearing Research*, 46, 16-24
- Ushijima, T. & Sawashima, M. (1972) Fiberscopic examination of velar movements during speech, *Annual Bulletin of the Research Institute of Logopedics and Phoniatics*, 25-38