

COORDINATION AND COARTICULATION IN SPEECH PRODUCTION*

CAROL A. FOWLER

and

ELLIOT SALTZMAN

Haskins Laboratories and University of Connecticut

In this article, we consider the concepts of coordination and coarticulation in speech production in the context of a task-dynamic model. Coordination reflects the transient establishment of constrained relationships among articulators that jointly produce linguistically significant actions of the vocal tract – that is phonetic gestures – in a flexible, context-sensitive manner. We ascribe the need for these constraints in part to the requirement of coarticulatory overlap in speech production. Coarticulation reflects temporally staggered activation of coordinative constraints for different phonetic gestures. We suggest that the anticipatory coarticulatory field for a gesture is more limited than look-ahead models have suggested, consistent with the idea that anticipatory coarticulation is the onset of activation of coordinative constraints for a forthcoming gesture. Finally, we ascribe much of the context-sensitivity in the anticipatory or carryover fields of a gesture (variation due to “coarticulation resistance”) to low-level (below the speech plan) interactions among the coordinative constraints for temporally overlapping gestures.

Key words: coarticulation, coordination, task-dynamic model, articulatory gestures

INTRODUCTION

Our topic is the relation between coarticulation and coordination in speech. We will suggest that the two theoretical constructs reflect different perspectives on the same characteristic organization of the vocal tract for speech. In elaborating this idea we will depend on three central concepts that we define next: phonetic gesture, coordination and coarticulation.

Phonetic gesture

Such terms as “phonetic gesture” (Liberman and Mattingly, 1985; Mattingly, 1990),

* We would like to acknowledge grant support from the following sources: NICHD Grant #HD-01994 (Nature and Acquisition of the Speech Code and Reading), NIH Grant #DC-00121 (Dynamics of Speech Articulation), and NSF Grant #BNS-88-20099 (Phonetic Structure Using Articulatory Dynamics) to Haskins Laboratories. We thank Rena Krakow for helpful comments and feedback on an earlier version of this article.

“dynamically-defined articulatory gesture” (Browman and Goldstein, 1990) or more simply “gesture” (Browman and Goldstein, 1986, 1991; Fowler, in press; Saltzman and Munhall, 1989) represent a set of closely-related concepts. Essentially, phonetic gestures are linguistically significant actions of structures of the vocal tract;¹ alternatively, the term is restricted to the control structures that generate those actions. Linguistically, gestures are hypothesized to be the primitives of a language-user’s phonological system (see, especially, the research by Browman and Goldstein, cited above). Physically, they are (or, under the alternative definition, they generate) coordinated movements of the vocal tract that achieve a phonetically significant goal. In this paper, we adopt the definition proposed by Saltzman and Munhall (1989) and use the term gesture to refer to “a member of a family of functionally equivalent articulatory movement patterns that are *actively* controlled with reference to a given speech-relevant goal (e.g., a bilabial closure)” (p. 334). According to this usage, gesture and movement have different meanings: Although gestures are composed of articulatory movements, not all movements can be interpreted as gestures or gestural components. For example, when the vertical distance between the upper and lower lips changes due to the active coordination of the lips and jaw to produce a bilabial closure, the resultant movement pattern is considered to be a gesture. However, when the interlip distance changes as the passive consequence of the jaw’s active participation in a different gesture (e.g., an alveolar gesture), the bilabial movement pattern would not be called a gesture.

Coordination

Pattee (e.g., 1976) suggests that, in living organisms, coordination creates macroscopic (i.e., coarser-grained) order in systems composed of components at a finer-grained level of analysis. Coordination is achieved by the implementation of “constraints” linking the finer-grained components and thereby creating dependencies among them. Because of the dependencies they create, constraints reduce the overall degrees of freedom of the system, but they reduce them selectively so that the remaining freedoms constitute a select subset of those possible in the absence of constraint. The order observable at a macroscopic scale of analysis of the system is shaped by this subset.

Consider a familiar example in a (nonbiological) human artifact, the automobile. Cars are built with one steering wheel that controls the turning of both front wheels and, therefore, of the car as a whole. There are constraints linking the two front wheels and the steering wheel. The linkages prevent independent turning of the two wheels. Cars could be built differently, with one steering wheel per front wheel, so that the wheels could be controlled independently. This would restore degrees of freedom at the finer-grained level of analysis of the car that the linkages eliminated, but it would have counter-functional utility at the coarser-grained level of analysis in which goal-directed movements of the car as a whole are specified. The selective loss of freedom

¹ We will use the term “vocal tract” more encompassingly than is commonplace to refer not only to the supralaryngeal articulators, but also to the larynx and the respiratory system.

at the finer-grain permits only the subset of movements of the two front wheels that ensures macroscopic functional utility in car motion.

In speech, as we suggest in more detail below, as in other kinds of action, the coordinative constraints are not built into the body permanently as linkages are built into a car; instead, the constraints are marshaled transiently (e.g., Turvey, 1990). We propose that coordinative constraints implement phonetic gestures by establishing dependencies among articulators. The dependencies reduce the degrees of freedom of those articulators that are relevant to realization of the gesture by restricting the articulators to the subset of possible movements that will jointly achieve macroscopic order in the form of a phonetic gesture. Returning to the bilabial closing gesture, then, we suggest that the lips close during speech production because, in a sense to be elaborated shortly, speakers establish coordinative constraints among the jaw, the upper lip and the lower lip that restrict those articulators to just that subset of movements that eventuate in bilabial closure.

Coarticulation

Coarticulation refers to the fact that at any given point during an utterance, the influences of gestures associated with several adjacent or near-adjacent segments can generally be observed in the acoustic or articulatory patterns of speech. This central and very important fact of speech production has received considerable attention in the literature. However, there is no consensus on what coarticulation is. A common idea is that it is an influence – largely assimilative – of one phonetic segment on another (Daniloff and Hammarberg, 1973). According to this view, coarticulated segments somehow change one another in a context-sensitive manner at the level of underlying linguistic feature values. A different idea is that phonetic segments in a sequence do not change one another in their linguistic essence. Rather their underlying linguistic identities are defined context-independently, but there is temporal overlap (*coproduction*) in the implementation of the coordinative constraints that instantiate the segments in the vocal tract (Fowler, 1977; Hardcastle, 1985; Marchal, 1988). In this view, context-sensitivity in acoustic or articulatory flows arises primarily from the dynamics of gestural interaction during coproduction. Thus, understanding coarticulation entails addressing the following two questions: a) What processes govern the time courses of coordinative constraints for the gestures in a given utterance? b) When there is temporal overlap or coproduction in these gestural time courses, how do their influences blend in shaping articulatory and acoustic flows?²

² Whereas coarticulation generally has been taken to refer to the relative timing or phasing of gestures for different phonetic segments, we will not distinguish here between phasing of gestures for one segment (for example, phasing of bilabial closing and laryngeal devoicing for /p/) and phasing of gestures for different segments.

INTERARTICULATORY COORDINATION DURING SINGLE GESTURES

As noted already, the coordinations that create dependencies among articulators in speech are not hardwired; rather, they are established transiently in the service of phonetic goals. The strongest evidence favoring this claim is obtained from studies of speech production in which an articulator is unexpectedly perturbed (e.g., Abbs and Gracco, 1983; Kelso, Tuller, Vatikiotis-Bateson, and Fowler, 1984; Munhall and Kelso, 1985; Munhall, Löfqvist, and Kelso, 1986; Shaiman, 1989; Shaiman and Abbs, 1987). For example, Kelso *et al.* (1984) applied a sudden downward force to the jaw of a speaker on a randomly-selected 20% of his productions of a target syllable /bæb/ or /bæz/. The force was applied as the jaw began to raise for the final consonant in the target syllable, and thus opposed jaw raising. As a consequence, the jaw attained a lower position during closure for the final consonant than it attained on nonperturbed trials. Within 20 msec of onset of the perturbation, a compensatory response began to be observed. When the target utterance was /bæb/, Kelso *et al.* found a lower position of the upper lip on perturbed than on unperturbed trials. The lips did close on perturbed trials, despite the considerably lowered jaw position, and the increased excursion of the upper lip contributed to the successful achievement of closing. On /bæz/ trials, a short-latency compensatory response occurred in the genioglossus muscle of the tongue, and /z/s were perceptually normal. These results suggest that the extra tongue activation permitted the tongue to raise more than on unperturbed trials and to achieve the close constriction to the palate necessary for /z/ frication despite the unusually low positioning of the jaw.

Two aspects of these findings are especially pertinent. First, responses to the perturbation are compensatory in relation to the goals of the perturbed phonetic gesture. Extra downward movement of the upper lip fosters the lip closing essential for the production of bilabial closure; it would not foster the alveolar constriction essential to production of /z/, and it does not occur when /z/ is perturbed. Extra activation of the genioglossus muscle fosters achievement of an alveolar constriction when the jaw is forced to a more open position; it would not foster lip closing, and it does not occur when the final /b/ of /bæb/ is perturbed. Other evidence that compensation is functionally specific to the goals of a phonetic gesture occur when the upper lip is perturbed upward by a "lip paddle" during production of /p/ or /t/ (Shaiman and Abbs, 1987; Shaiman, 1989). When /p/ is perturbed, extra upward displacement of the lower lip compensates for the unusually raised position of the upper lip. When the upper lip is perturbed in /t/, in which it is not an articulator that contributes to achievement of the labiodental constriction, no extra displacement of the lower lip is found. The second important observation about these compensations to perturbations is that they occur at very short latencies. In the research of Kelso *et al.* (1984), compensations occurred with latencies that ranged between 15 and 35 msec. This suggests that such compensatory articulations are not consequences of any cognitive replanning of gestural trajectories. Rather, the compensations appear to reflect the ongoing state of coordinative constraints that serve to establish gesture-specific patterns of coupling or gating among the articulators.

We conclude, then, that coordinative constraints are established transiently to implement phonetic gestures in speech. The constraints permit context-sensitive, flexible, achievement by the articulators of their conjoint phonetic goal: lip closure for bilabial consonants, a close alveolar constriction for /z/, etc. Although the particular way in which the relevant articulators participate to achieve that goal can vary, the constraints ensure that the goal achievement itself is relatively invariant. How might such constraints be realized in the vocal tract? In the following section, we discuss the implementation of interarticulatory coordinative constraints for the production of single gestures in the context of a particular model of speech production, the *task-dynamic* model (e.g., Saltzman, 1986, 1991; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989). In later sections, we will also use this model to illustrate several issues pertaining to intergestural coordination and the shaping of coarticulatory movement patterns.

Task-dynamic modeling: Mechanical perturbations

One of the major tasks in speech is to create and release constrictions locally in different regions of the vocal tract, e.g., at the lips for bilabial consonants, or between the tongue dorsum and palate for some vowels. In the task-dynamic model, such constrictions are controlled by a dynamical system with two functionally distinct but interacting levels. The level of *intergestural* coordination is defined according to a set of *activation* coordinates; the level of *interarticulator* coordination is defined according to both *model articulator* and *tract-variable* coordinates (see Figure 1). The architectural relationships among these coordinates are illustrated in Figure 2. The linguistic identity of each gestural unit is defined context-independently by its activation coordinate and set of tract-variable and model articulator coordinates, and by its set of dynamical parameters (e.g., constriction target position). The current value of each gesture's activation coordinate defines the strength with which the gesture "attempts" to shape vocal tract movements at any given point in time according to its own phonetic goals. The tract variables and model articulators associated with each gesture specify the particular vocal-tract constriction (e.g., bilabial) and set of articulators (e.g., lips and jaw) whose behaviors are affected directly by the gesture's activation (see Figure 3). The intergestural level accounts for patterns of relative timing and cohesion among activation intervals for the gestures participating in a given utterance, e.g., for tongue-dorsum and bilabial gestures in a vowel-bilabial-vowel sequence. The interarticulator level accounts for the coordination among articulators that exists at a given point in time due to the currently active set of gestures, e.g., the coordination among lips, jaw, and tongue during periods of vocalic and bilabial gestural coproduction.

In task-dynamic simulations, each constriction type (e.g., bilabial) is associated with a pair (typically) of tract variables, one that refers to the location of the constriction along the longitudinal axis of the vocal tract, and one that refers to the degree of constriction measured perpendicularly to the longitudinal axis in the sagittal plane. Since constrictions are defined in the sagittal plane of the vocal tract only, they are at most two-dimensional. The reason for this is that the simulations use the articulatory geometry represented in the Haskins Laboratories software articulatory synthesizer (Rubin, Baer, and Mermelstein, 1981). This synthesizer is based on a midsagittal view

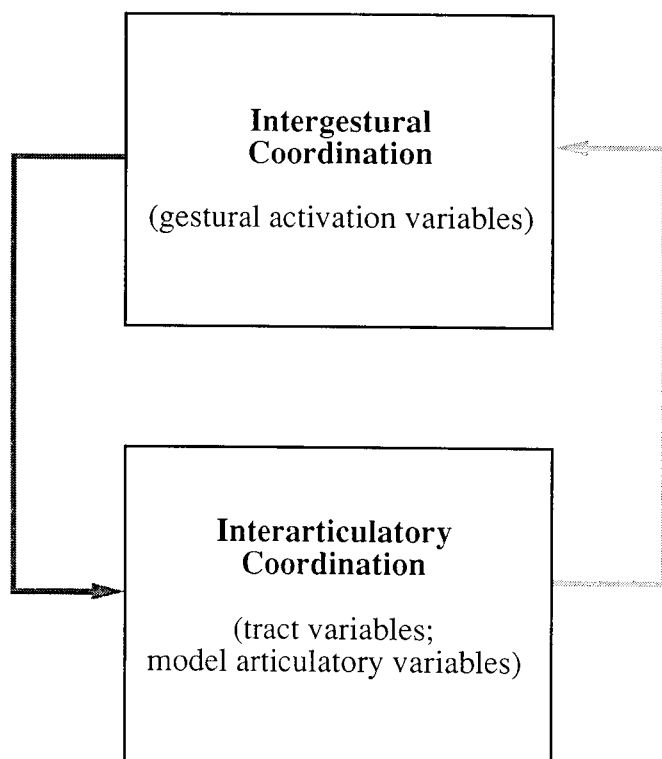


Fig. 1. Schematic illustration of the two-level dynamical model for speech production, with associated coordinate systems indicated. The darker arrow from the intergestural to the interarticulator level denotes the feedforward flow of gestural activation. The lighter arrow indicates feedback of ongoing tract-variable and model articulator state information to the intergestural level. (From Saltzman and Munhall, 1989; reprinted by permission.)

of the vocal tract and transforms a given articulatory configuration in the sagittal plane to a sagittal outline of the vocal tract, a three-dimensional tube shape, and, finally, with the addition of appropriate voice source information, an acoustic waveform. Modeling work has been performed in cooperation with several of our colleagues at Haskins Laboratories as part of an ongoing project focused on the development of a gesturally-based, computational model of linguistic structures (e.g., Browman and Goldstein, 1986, 1991; Kelso, Saltzman, and Tuller, 1986a, 1986b; Kelso, Vatikiotis-Bateson, Saltzman, and Kay, 1985; Saltzman, 1986, 1991; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989). For recent reviews, related work, and critiques, see also

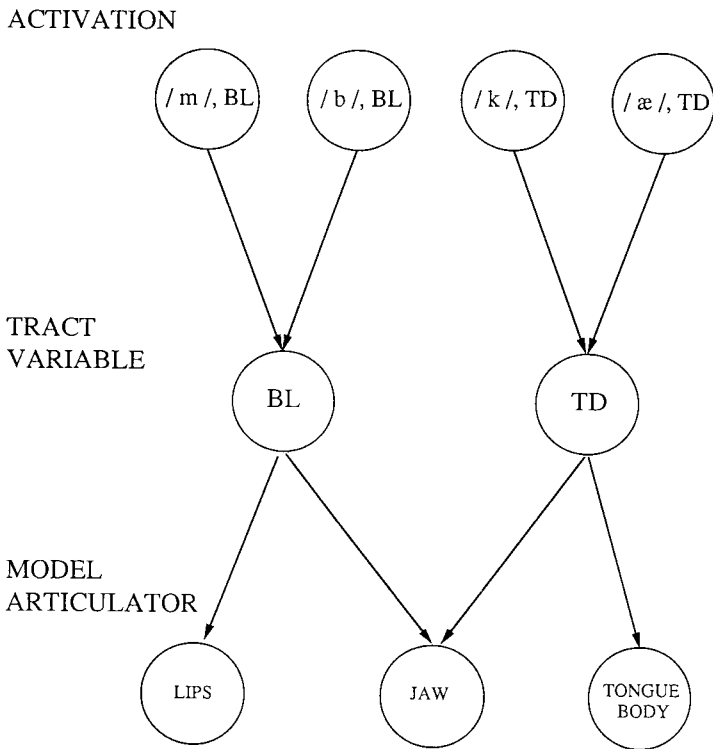


Fig. 2. Example of the architectural relationships defined among model-articulator, tract-variable, and activation coordinate systems. BL and TD denote tract variables associated with bilabial and tongue-dorsum constrictions, respectively. Gestures at the activation level are labeled in terms of both linguistic identity (e.g., $/k/$ and tract-variable affiliation (e.g., TD). (From Saltzman, 1991; reprinted by permission.)

de Jong (1991), Edwards, Beckman, and Fletcher (1991), Hawkins (1992), Jordan and Rosenbaum (1989), Perkell (1991), and Vatikiotis-Bateson (1988).

Compensatory articulation

The task-dynamic model displays gesture-specific patterns of compensatory articulation to simulated mechanical perturbations delivered to the model articulators. These patterns mirror those found in the experimental data described in the preceding section. In particular, simulations were performed on perturbed and unperturbed bilabial closing gestures (Saltzman, 1986; Kelso *et al.*, 1986a, 1986b). When the simulated jaw was “frozen” in place during the closing gesture, the system: a) attained the same final

Tract variables		Model articulators
LP	lip protrusion	upper & lower lips
LA	lip aperture	upper & lower lips, jaw
TDCL	tongue dorsum constrict location	tongue body, jaw
TDCD	tongue dorsum constrict degree	tongue body, jaw
LTH	lower tooth height	jaw
TTCL	tongue tip constrict location	tongue tip,body, jaw
TTCD	tongue tip constrict degree	tongue tip,body, jaw
VEL	velic aperture	velum
GLO	glottal aperture	glottis

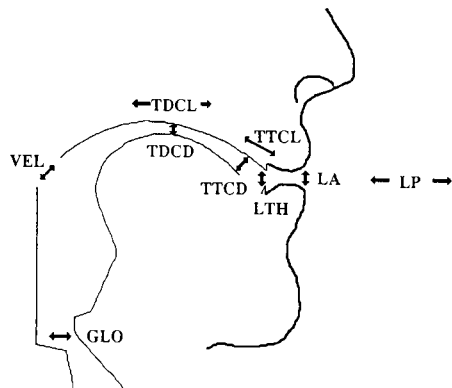


Fig. 3. Top: Table showing the relationship between tract variables and model articulators; Bottom: Schematic midsagittal vocal tract outline, with tract-variable degrees of freedom indicated by arrows. (From Saltzman and Munhall, 1989; reprinted by permission.)

degree of bilabial closure in both the perturbed and unperturbed cases, although with different final configurations for the articulators; and b) showed immediate “on-line” compensatory responses in the upper and lower lips to the jaw perturbation, in the sense that the system did not require replanning or reparameterization in order to compensate. Compensation occurred through the automatic and rapid redistribution of activity over the entire articulatory ensemble in a gesture-specific fashion, and the processes of control and coordination were exactly the same during both perturbed and unperturbed simulated gestures.

INTERARTICULATORY COORDINATION DURING COPRODUCTION

Given the rarity of perturbing jaw brakes and lip paddles outside the laboratory, however, we must ask why speakers establish coordinative constraints among articulators that permit flexible achievement of phonetic goals rather than simply choosing one, context-free, route to the goals. One important reason, we suggest, is the necessity of coarticulation. For example, during the intended utterance /bæb/, when the syllable-initial consonant /b/ is released, a more sonorant segment is produced necessarily as the vocal tract opens up. The only way to ensure that the more sonorant segment produced is the intended one, /æ/, is to move toward the /æ/ during closing and/or closure for the initial /b/. Coarticulation cannot be avoided, with the result that there are competing demands made on the jaw during the bilabial closing and/or closure. For example, as the jaw closes for the /b/, the /æ/ will tend to oppose this motion; a following higher vowel, /i/, will oppose it less, and a following lower vowel, /a/, will oppose it more. Therefore, bilabial closure must be achievable with context-sensitive contributions of the jaw, the upper lip, and the lower lip during everyday coarticulated speech. By hypothesis, the coordinative constraints allow the talker to select the same /b/ in every context rather than having to select a different, context-sensitive /b/ to fit each context. Coarticulation, then, can be viewed as a natural source of articulatory perturbations, analogous to those supplied artificially by lip paddles and jaw brakes in the laboratory. Context-sensitive compensation for perturbations are handled by the same processes of articulatory coordination and control, regardless of the source of perturbing influences.

Coarticulation resistance

Coarticulatory effects are themselves context-sensitive (e.g., Bladon and Al-Bamerni, 1976; Recasens, 1984a, 1984b; Keating, 1990), and it is clear why they must be. If the goals of a given intended phonetic gesture are to be invariantly achieved, then coarticulatory encroachments by nearby gestures must be limited to articulators that: a) are not involved in producing the intended gesture, or b) are involved but will not interfere with the achievement of gestural goals, due to coordinative compensation by other nonperturbed articulators. In cases where the encroached-upon articulators are centrally involved in producing the intended gesture, the intergestural interference must either be eliminated or kept within tolerable (articulatory or perceptual) limits.

The ability of a given gesture to resist potentially disruptive encroachments by nearby gestures has been termed "coarticulation resistance" by Bladon and Al-Bamerni (1976). Systematic investigation of this idea has been carried out largely by Recasens (1984a, 1984b, 1987, 1991; see also Farnetani, 1990). For example, in one investigation, Recasens (1984a) studied vowel-to-consonant coarticulatory influences in VCV utterances with vowels /i/, /a/, and /u/, the dorsopalatal consonant /j/, alveolo-palatal /ɲ/ and /ʎ/, and alveolar /n/. Using electropalatography, Recasens found that the consonants require decreasing contact of the tongue dorsum with the palate in the series as listed. He found also that the degree of anticipatory and carryover coarticulation of the vowels with the consonants varied inversely with the consonants' differential

demands on the tongue dorsum, also a primary articulator for the vowels. Compatibly, Farnetani (1990) showed a strong inverse relationship between degree of tongue/palate contact for singleton and geminate consonants of Italian and degree of coarticulatory influence from neighboring vowels. In her review of the literature on coarticulation resistance, Farnetani (1990) observed that strong coarticulation resistance in the production of a gesture was also associated with "coarticulatory aggression" in the influence of that gesture on preceding and following ones: "[T]he sounds that block or reduce the coarticulatory effects of the neighbour sounds are also exerting the strongest effects, in other words they exhibit the least contextual variation and induce the greatest" (p. 106).

These findings are just as expected if, in some way, talkers protect phonetic gestures against coarticulatory influences that would interfere with achievement of the gestures' phonetic goals. However, there are at least two alternative means by which that protection might be implemented. It might be achieved intentionally in a "speech plan" by adjusting the time courses of coordinative constraints for coarticulating gestures to their context, in order to prevent those gestures from interfering with the achievement of an ongoing gesture's phonetic goals. For example, talkers might choose to delay or weaken the anticipatory onset of an upcoming gesture to minimize coarticulatory encroachment upon the ongoing gesture. Alternatively, the plan for a gesture and its phasing with respect to others may be relatively context-free; the contextual variation observed in articulatory and acoustic flows would be shaped primarily by the manner in which the coarticulating gesture blends its influences on the vocal tract with those of an on-going one. In other words, the attempted influence of a gesture on the vocal tract would be relatively invariant, but it would meet differential resistance, depending on the demands placed on vocal tract structures by an on-going phonetic gesture. A simple analogy might make this second idea clearer. Imagine an actor planning an invariant squeezing gesture of the hand. On different occasions, the closing fingers encounter a rock, a solid rubber ball, a foam rubber ball, or nothing at all. The extent to which the fingers close for the same planned squeezing event will vary depending both on the intended strength of squeezing and on the resistance encountered by the fingers at the surface. Similarly, whether a coarticulating gesture exhibits strong or weak influences on vocal tract shape depends both on its own degree of coarticulatory aggression, and on the degree of coarticulatory resistance displayed by the on-going gestures.

These two alternatives are not mutually exclusive. However, there is one indication in the literature that at least some sources of context-sensitivity in coarticulation are not explicitly planned. In an extension of his research on coarticulation resistance in vowel-consonant production, Recasens (1984b) looked at electropalatographic evidence of transconsonantal vowel-to-vowel coarticulatory influences in the same VCV set. He reported instances in which vowel-to-consonant coarticulatory influences were blocked both in an anticipatory and a carryover direction by a consonant, / λ /, that places strong constraints on the tongue dorsum; in the instances in question, however, statistically significant evidence of anticipatory and carryover vowel-to-vowel coarticulation was observed. That is, coarticulatory influences of the vowels were discontinuous, jumping over the consonant, as it were.³ It seems to us implausible that talkers plan for

³ To our knowledge, Recasens' findings are unique, and he did not encounter them everywhere that he might have (in the context of /j/, for example). We mention the finding only because it is suggestive and needs to be pursued further.

discontinuous coarticulatory influences. That is, in the anticipatory direction, why start producing a vowel, stop, then start again; or, even less plausibly in the carryover direction, if a consonant puts an end to carryover coarticulation, why should the talker revive the vowel thereafter? It appears more plausible that talkers plan to phase a given vowel in a way that is relatively insensitive to its context, and the context-sensitivity arises primarily when the influences on the vocal tract from the vowel encounter other influences on the vocal tract from other ongoing segments. In fact, the task-dynamic model provides an account of coproduction that is consistent with this viewpoint.

Task-dynamic modeling: Intergestural blending

In the task-dynamic model, the articulatory effects of gestural coproduction vary as a function of the degree of spatial overlap on the gestures involved, i.e., the degree to which the gestures share articulators. As long as the spatial overlap is incomplete, there is minimal interference among the coproduced gestures and each can attain its own phonetic goals. In the model, this situation is one in which the gestures are defined along different sets of tract variables, and the gestures have no, or some but not all, articulators in common (see Figure 3). Figure 4A illustrates the operation of the model for two VCV sequences in which symmetric flanking vowels, /i/ and /æ/, vary across sequences, the medial consonant is the alveolar /d/ in both sequences, and the time courses of vowel and consonant activation waves are identical in both sequences. Here, the vowels are defined along the tract variables of tongue-dorsum constriction location and degree, which are associated with the jaw and tongue-body articulators; the alveolar is defined along the tract variables of tongue-tip constriction location and degree, which are associated with the jaw, tongue-body, and tongue tip articulators (see Figure 3). Hence, the vowel and consonant gestures share some but not all articulators, and the alveolar's tongue-tip constriction goals are identically met in both cases, with contextual differences in articulatory positions induced by corresponding differences in flanking vowel identities. (See Figure 4C for comparisons with the tract shapes associated with simulated steady-state production of /i/ and /æ/, unperturbed by consonant superposition.)

In cases where coproduced gestures are defined along the same sets of tract variables, however, all articulators are shared, and the potential for mutual interference in attaining completing phonetic goals is created. Figure 4B illustrates the operation of the model for two symmetric VCV sequences that are identical to those shown in Figure 4A, except that the medial consonant is the velar /g/. Here, the consonant and vowels are defined along the same tongue-dorsum tract variables and are associated with the jaw and tongue-body articulators. Hence, the gestures compete for control of tongue-dorsum motion during periods of coproduction. The result is that contextual variation is now seen even in the attainment of the constriction target for /g/. The velar's place of constriction is altered by the identity of the flanking vowels, although the degree of constriction is not. Importantly, the behavior displayed by the model in both Figures 4A and 4B mirrors the patterns observed during actual speech production (Öhman, 1967).

Within-tract-variable conflict is resolved in the model by a process of intergestural blending that is specified according to competitive-dynamics network equations (see

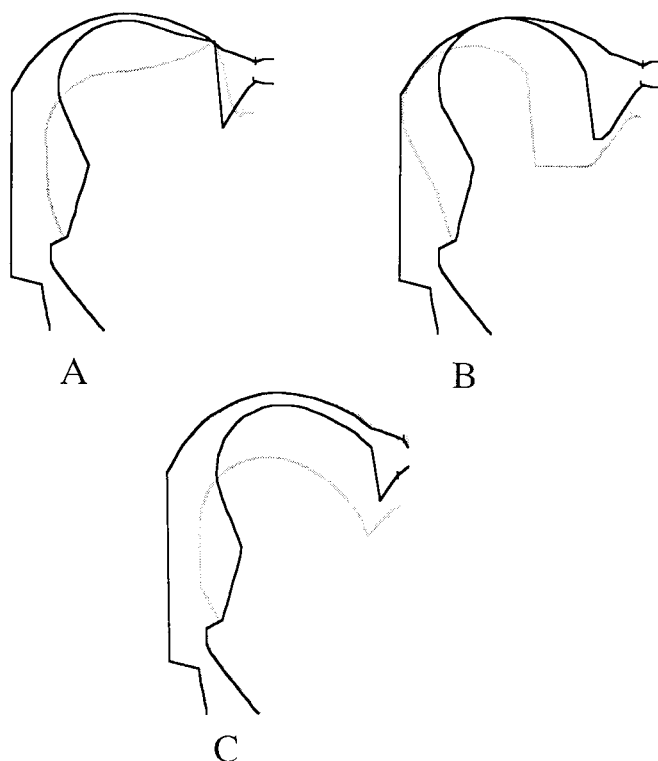


Fig. 4. Simulated vocal tract shapes. A: First contact of tongue-tip and upper tract wall during symmetric vowel-alveolar-vowel sequences; B: First contact of tongue-dorsum and upper tract wall during symmetric vowel-velar-vowel sequences; C: Corresponding steady-state vowel productions. In all cases, /æ/ and /i/ shapes are shown using grey and black outlines, respectively. (From Saltzman, 1991; reprinted with permission.)

Saltzman and Munhall, 1989, for details). The tract-variable components of a given gesture are endowed with their own characteristic degrees of blending strength, according to the context-independent sets of blending parameters assigned to these gestural components. Roughly speaking, blending strength varies inversely with gestural sonority — stops are strongest and vowels are weakest. Blending of gestures with similar strengths, e.g., two vowels, results in a simple averaging. This result is consistent with data on laboratory-induced speech errors during vowel production (Laver, 1980), in which blended vowel forms were produced that were intermediate between canonical forms. On the other hand, when the tract-variable components of coproduced gestures differ in blending strength, the stronger gestural component suppresses the weaker component. This is the case previously discussed and shown in Figure 4B, in which the stronger

velar consonant suppresses the weaker vowel, with the degree of suppression being stronger for constriction-degree than for constriction location. Blending strength, therefore, captures in a formal sense both the coarticulatory “resistance” and “aggression” of phonetic gestures observed in actual data.

Finally, it should be noted that many coproduced gestures seem to display a simple, additive form of blending (e.g., Löfqvist, 1990). Interestingly, additive blending seems to be displayed only by gestures that would be characterized in the task-dynamic model by a single tract variable. For example, Boyce (1988, 1990) showed that for three of four English speakers, lip protrusion movements in utterances such as /kuktluk/ can be closely approximated by adding the lip movement patterns from /kiktuk/ and /kuktik/, and then subtracting the pattern from /kiktik/ to eliminate /i/ influences. Relatedly, Bell-Berti and Krakow (1991) found that velum lowering in CVN sequences is the sum – with temporal overlap varying with rate of speaking – of lowering for the vowel and lowering for the consonant. Lastly, Munhall and Löfqvist (1992) reported that the glottal devoicing gestures associated with /s/ and /t/ in the utterance “Kiss Ted” spoken at various rates could be simulated successfully by assuming two independent devoicing gestures that overlapped to various degrees and summed.

INTERGESTURAL COORDINATION: THE PATTERNING OF GESTURAL ACTIVATION WAVES

In a previous section (*Task-dynamic modeling: Mechanical perturbations*), we defined the current value of a gesture’s activation coordinate as the strength with which the gesture is currently “attempting” to affect vocal tract shape in accordance with its own phonetic goals. Here we identify explicitly the time courses of these activation values with the time courses of gestural coordinative constraints. This allows us to define the “speech plan” for a given utterance formally as the set of activation waveforms for the gestures in the utterance. Understanding the processes that determine the speech plan then becomes a problem of understanding the processes that govern the time courses of gestural activation waves. Here, we address several issues related to the patterns displayed by an utterance’s activation waves. The first issue pertains to the shape of the waveforms associated with individual gestures, e.g., are they sharply defined step-functions or more smoothly defined? The second issue deals with waveform duration.

The shape of gestural activation waves

Beginning with Joos (1948), many investigators (e.g., Fowler and Smith, 1986; Löfqvist, 1990; Mattingly, 1981; Öhman, 1966; Saltzman and Munhall, 1989) have proposed that the influence of a phonetic segment (here, phonetic gesture) on the vocal tract waxes and wanes gradually. There is an interval of time in which the gesture’s attempt to influence the shape of the vocal tract is maximal, and in which the attempted influences of other gestures – at least of gestures for other phonetic segments – are weaker. However, before that time interval and after it are intervals in which an influence of the gesture on the vocal tract may be detected, but it is submaximal, and the

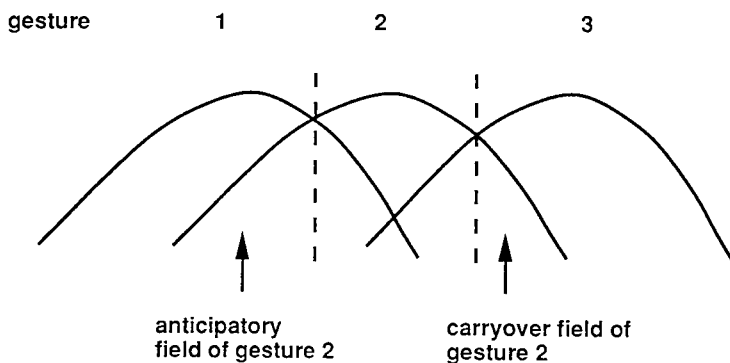


Fig. 5. Schematic representation of prominence or activation waves for three overlapping phonetic gestures, with anticipatory and carryover coarticulatory fields indicated.

influences of another gesture (or of other gestures) are closer to peak (see Figure 5). These proposals are consistent with the hypothesis that gestural activation waves are smoothly shaped, and that each gesture will display gradual implementation and relaxation phases. Intervals of coarticulation by a given gesture would be defined by the time courses of implementation (anticipatory field of coarticulation) or the relaxation (carryover field of coarticulation) of the gesture's activation wave. Thus, during any given instance of coproduction, it is expected that the net relative contributions of the participating gestures to ongoing vocal tract motion will be determined both by the relative intrinsic blending strengths (see the above section *Task-dynamic modeling: Gestural blending*) and by the relative activation levels of the gestures.

The hypothesis of smoothly graded activation waves is also consistent with recent electromyographic data reported by Mowrey and MacKay (1990) on sublexical speech errors that were induced during the production of tongue twisters. For example, during intended productions of "Bob flew by Bligh Bay", the investigators monitored EMG output from the lingual transversus/verticalis complex (T/V). In trials that were normal both electromyographically and acoustically, they found that T/V activity appeared in two distinct bursts accompanying the production of the lateral /l/s in "flew" and "Bligh". However, in trials that were anomalous both electromyographically and acoustically, they observed two types of error patterns. One type of error was classified as an *intrusion* error, consisting of increased T/V activity during the intended production of "Bay". In this case, the error sounded like "Blay". The second type of error was classified as an *exchange* error, consisting of both a decrease of T/V activity for "Bligh", and an intrusive increase of T/V activity for "Bay" and sometimes "by". In these cases, the errors sounded like "Bigh", "Blay", and "bly". Significantly, the T/V errors did not appear in an all or none manner, but were graded continuously between levels that were appropriate for normal and fully anomalous utterances. These smoothly graded EMG

patterns are consistent with the smoothly graded waveshapes proposed for gestural activation waves.

The duration of gestural activation waves

What is the duration of a gesture's activation wave? This question has been at the heart of a large body of research in speech science, particularly in the (sometimes) heatedly controversial literature on anticipatory coarticulation. In this section, we review a portion of this literature, focusing on the issue of the temporal extent of a gesture's anticipatory field. We develop the argument, as have others before us, that anticipatory coarticulation effects are temporally limited, and do not typically extend very far backward in time from the period of a gesture's own predominant interval. This interpretation is consistent with that provided by *frame* models of coarticulation (e.g., Bell-Berti and Harris, 1974, 1981), and contrasts with the extensive degrees of anticipation allowed by *look-ahead* models (e.g., Henke, 1966).⁴ Since data on anticipatory lip rounding and velum lowering have provided major testing grounds for distinguishing between these two general accounts of coarticulation, we begin by reviewing research in these areas. We end this section by examining data on trans-consonantal vowel-to-vowel coarticulation, and conclude that these, too, comprise evidence supporting the frame-model.

Anticipatory lip rounding. Look-ahead models rely on the hypothesized linguistic process of assimilative feature-spreading (e.g., Henke, 1966). In these models, a feature is frequently hypothesized to attach itself in an anticipatory direction to as many phonetic segments preceding the phonetic-segmental source of the feature as are "unspecified" for that feature. In the case of lip rounding in English, vowels are linguistically specified for rounding in that they must be marked as rounded or unrounded. Changing the rounding feature of a vowel changes its identity. In particular, unrounding /ɔ/ (approximately) gives /a/. In contrast, English consonants are unspecified for rounding, because rounding a consonant, for example /t/, does not change its identity. In look-ahead models, the rounded feature of a planned /u/ can be spread in an anticipatory direction to as many consonants as precede the /u/ in the speech plan. Anticipatory feature-spreading is blocked at the first unrounded vowel encountered in the planned sequence, thereby defining the onset time for the rounding gesture. Findings of extensive anticipatory rounding have been reported for English and other languages, and have been reported to be generally consistent with predictions of look-ahead models (e.g., Benguerel and Cowan, 1974; Daniloff and Moll, 1968).

Other investigators, most notably Bell-Berti and Harris (1974, 1981) have reported contrasting evidence that supports a much more limited anticipatory field for rounding gestures than would be predicted by look-ahead models. In particular, they reported

⁴ Actually, Henke's (1966) model incorporated an anticipatory feature spreading procedure that looked ahead only to the immediately following segment. He noted, however, that the model was capable in principle of more extensive anticipation, if this were deemed justifiable on the basis of actual speech production data.

that rounding gestures were initiated within a relatively invariant time frame prior to the acoustically-defined onset of rounded vowels,⁵ and accordingly proposed a frame model of gestural timing as an alternative to feature-spreading, look-ahead models.

A possible reconciliation of these apparently inconsistent findings was offered by Perkell and Chiang (1986), following earlier work by Bladon and Al-Bamerni (1982) on velum lowering. For lip rounding, Perkell and Chiang frequently observed two-phase rounding gestures: An earlier gradual onset of rounding typically beginning near the onset of the first consonant in a consonant string preceding the rounded vowel and a second phase (whose onset was identified by an acceleration maximum in the rounding gesture) having an abrupt increase in the rounding movement that was time-locked to acoustically-defined rounded-vowel onset. Perkell and Chiang propose a *hybrid* model (see also Perkell, 1990) of coarticulation that includes both a look-ahead component and a time-locked component for an anticipated gesture.

As appealing diplomatically as this solution may have seemed, it was incorrect, as Perkell and Matthies (1992) appear to agree. Following an earlier insight of Bell-Berti (1980), Gelfer, Bell-Berti, and Harris (1989) included control utterances in their stimulus set to isolate true coarticulatory influences of a rounded vowel from other sources of lip-muscle activity. Each control utterance was matched to a test utterance having a rounded vowel and was identical to the test utterance except that, in the control utterance, /i/ was substituted where test utterance had the rounded vowel, /u/. Looking at electromyographic activity of the orbicularis oris inferior muscle of the lower lip on test utterances only, Gelfer *et al.* might have concluded in favor of the hybrid theory of Perkell and Chiang. In utterances such as /ist#tu/, two bursts of muscle activity were detected, one early in the consonant sequence preceding /u/ and the other closer to the acoustically-defined onset of /u/. However, the control utterances, which had no rounded vowel in them, had muscle activity in the same place as the test utterances' early peak of activity. Clearly that activity was associated with production of the consonants, not with the rounded vowel. With that activity eliminated from consideration, findings were that activity associated with rounding for the vowel was, approximately, time-locked to the acoustically-defined vowel onset. Compatible findings have also been reported by Boyce (1990; see also Boyce, 1988) when lip protrusion movements are examined rather than muscle activity.

Finally, in a recent extension of the work by Perkell and Chiang, Perkell and Matthies (1992) agree that "evidence which was originally interpreted as support for the hybrid model (Perkell and Chiang, 1986) appears to stem, at least in part, from consonant effects . . . (also Gelfer *et al.*, 1989; Boyce, Krakow, Bell-Berti, and Gelfer, 1990)" (p. 2923). Thus, evidence for a long-duration anticipatory field for a rounded vowel has, in the recent literature, been reinterpreted as evidence that some consonants have

⁵ Anticipatory rounding was initiated an invariant interval before acoustically-defined /u/ onset in VC_nu sequences where n, the number of consonants preceding /u/, was two or more. In sequences where n equaled one, rounding began closer to /u/ onset due, Bell-Berti and Harris (1982) hypothesized, to an antagonistic influence from the unrounded V.

associated lip activity and protrusion movements. When they are eliminated from consideration, the remaining evidence suggests that the anticipatory field for a vowel rounding gesture is relatively short in duration.

Anticipatory velum lowering for a nasal consonant. The history of research on nasalization is like that for lip rounding. In English, consonants are specified for nasality, whereas vowels are not. Accordingly, look-ahead models predict that the feature [+nasal] will spread in an anticipatory direction from a nasal consonant through every preceding vowel up to the first occurrence of an oral consonant, which will block further spreading. Some early findings appeared consistent with the prediction (Kent, Carney, and Severeid, 1974; Moll and Daniloff, 1971). However, Bell-Berti (1980; see also Fritzell, 1969; Henderson, 1984; Moll, 1962; Ushijima and Sawashima, 1972) showed that vowels and oral consonants are themselves associated with characteristic velum heights. For our purposes the relevant finding is that vowels are associated with generally lower velum heights than are oral consonants. Accordingly, findings that the velum lowers in the first vowel after an oral consonant in a string of vowels preceding a nasal consonant do not support the look-ahead theory, because the velum would lower even in the absence of a nasal consonant. It lowers because vowels have lower velum heights than oral consonants. With that lowering eliminated from consideration, Bell-Berti (1980) reported that anticipatory lowering of the velum for a nasal consonant began a stable interval before the acoustically-defined onset of the nasal – a result consistent with the hypothesis that the anticipatory fields of gestural activation waves are relatively short in duration.

Bladon and Al-Bamerni (1982), however, reported finding two patterns of velum lowering in their subjects. In one pattern, velum lowering was monophasic and occurred at vowel onset, consistent with look-ahead theory. In the other, lowering was biphasic with one phase beginning at vowel onset and the other time-locked to the nasal consonant itself, consistent with Bell-Berti's frame model (Bell-Berti and Harris, 1981) of gestural timing. Bladon and Al-Bamerni were unable to discover variables that determined whether monophasic or biphasic lowering would be observed; recently, Bell-Berti and Krakow (1991) were able to do so.

In contrast to earlier investigations of anticipatory coarticulating of nasality, the experiment of Bell-Berti and Krakow included oral control utterances matched to nasal test utterances. That is, whereas test utterances consisted of an oral consonant followed by a variable number of vowels followed by a nasal consonant, control utterances consisted of an oral consonant followed by a variable number of vowels followed by an oral consonant. These investigators found one and two stage velum lowering patterns in their test utterances, just as Bladon and Al-Bamerni (1982) had reported. However, the first phase of the biphasic patterns and a piece of the only phase of the monophasic patterns occurred also in the control utterances, and hence could not be ascribed to the nasal consonant. Rather, these lowerings reflected lowering of the velum from the oral consonant to the first vowel in the sequence. With those influences on the velum eliminated from consideration, the lowering pattern was as predicted by the frame model. Anticipation did not spread far, and it appeared time-locked to the nasal consonant, and not to the preceding vowels.

Bell-Berti and Krakow (1991) were also able to predict when lowering would be monophasic or biphasic. The single-stage lowering occurred when the vowel sequence was short in duration, either because it included just one or two vowels or because the talker's speech rate was fast. These investigators interpreted this result as evidence that, when the vowel sequence is short, the velum lowering gesture for the nasal completely overlaps the vowel-related lowering. When the vowel sequence is longer, the two sources of velum lowering do not overlap, and they are manifested as a biphasic lowering movement.

In short, the evidence from investigations of velum lowering is wholly consistent with that from investigations of lip rounding in suggesting that the anticipatory coarticulatory field of a gesture is not very extensive, and in supporting the predictions of the frame model.

Transconsonantal vowel-to-vowel coarticulation. The look-ahead theory of coarticulation should predict no transconsonantal vowel-to-vowel coarticulation (except, perhaps for influences on the reduced vowel, schwa), because vowels are all specified for those features that reflect the vowels' height, frontness and lip configuration. However, such coarticulatory effects are typically found (e.g., Butcher and Weiher, 1976; Carney and Moll, 1971; Fowler, 1980; Öhman, 1966). The findings are generally consistent with Öhman's (1966) idea, echoed by Perkell (1969) and Fowler (1977), that vowels are produced more-or-less continuously ("diphthongally" according to Öhman) in speech, with the talker "superimposing" the consonantal gestures onto ongoing vowel-to-vowel gestures.

One idea as to why vowel production might be continuous (if it is) is that anticipatory transconsonantal vowel-to-vowel coarticulation may be necessary for the same reason offered above regarding the necessity of anticipatory vowel-to-consonant coarticulation (see section *Interarticulatory coordination during coproduction*). When a syllable-initial consonant is released, a sonorant sound, let us say a vowel, will be produced. To ensure that the vowel is the intended one requires that the coordinative constraints for that vowel be implemented well before consonant release. Consonantal constrictions, being short, may not in themselves provide the requisite lead time.

There has been less systematic attention paid to the question of the durational extent of vowel-to-vowel coarticulation than has been paid to the analogous question applied to lip rounding and velum lowering. In the latter domains, Bell-Berti and Harris (1981) estimated that muscle activation for lip rounding anticipates the acoustically-defined rounded vowel onset by about 200 msec at a normal rate of speech; similarly, Bell-Berti and Krakow (1991) estimated an anticipation of about 250 msec for the velum, based on their findings and those of Bell-Berti (1980).

Rather dramatic vowel-to-vowel effects were reported by Magen (1989) in two out of four speakers of English. These speakers produced /bV₁bəbV₂b/ nonsense trisyllables with vowels /i/ and /a/, and showed anticipatory influences of V₂ on V₁ through the intervening /bə/ syllable. Both speakers showed a significant influence on F2 at V₁ midpoint; one speaker showed an effect at V₁ onset if V₁ was /a/. Acoustical durational data were provided (Magen, 1989, p. 71) that permit the estimation of the durational extent of these anticipatory effects; in particular, Magen reported whole-word durations

and vowel durations. For the speaker showing influences only at V_1 midpoint, the average whole-word duration, measured from initial /b/ release through final /b/ closure, was 441 msec. The duration of V_2 accounted for 180 msec of that duration leaving less than 261 msec of the word (that is, 261 msec minus the unreported duration of the final /b/ closure) before the acoustically-defined onset of V_2 . The first vowel was 79 msec long on average. Accordingly, for this speaker, the midpoint of V_1 would be less than 222 msec from V_2 onset – a value close to those found by Bell-Berti and colleagues to characterize anticipatory lip rounding and anticipatory velum lowering. An analogous estimate for the speaker who showed anticipation at V_1 onset when V_1 was /a/ is 257 msec. Possibly, then, vowel-to-vowel coarticulation in English has about the same coarticulatory extent as lip rounding and velum lowering; durationally, none precede the predominant interval of their “own” gesture by a very long duration. Possibly, too, 200–250 msec is an estimate of the prototypical duration of the rising portion of a gesture’s coordinative constraints in the vocal tract for speech produced at a comfortable rate.

Task-dynamic modeling: Gestural activation waves

In the task-dynamic model, gestural activation waves are currently specified at the intergestural coordination level of the model (see Figures 1 and 2) as simple step functions, whose values change discretely between zero (the gesture is inactive) and one (the gesture is maximally active). The set of activation waves for a particular utterance is specified currently according to a *linguistic gestural model* that embodies the rules of Browman and Goldstein’s (1986, 1991) *articulatory phonology*, and incorporates knowledge concerning contrastive gestures for English and intergestural cohesion into larger linguistic units. The time courses of gestural activations are specified in the form of a *gestural score*, which plays the role of a “speech plan” in the present paper.

The rules of the linguistic gestural model generate activation waves with appropriately limited anticipatory fields, consistent with those predicted by the frame model of coarticulation. Roughly, the reason is that each gesture is associated with its own set of tract-variable dynamic parameters (e.g., stiffness, damping) that serve to define an effective time constant for the gesture. This means that if a gesture’s activation is switched from zero to one and remains at one for an extended period of time, the gesture displays a characteristic *settling time* in traveling to within a criterion percentage of the distance from its initial tract-variable position to its target position. The onset of a given activation wave is specified in order to create an activation step whose anticipatory field has a duration on the order of the settling time of the activated gesture (see also Coker, 1976).

There are (at least) two shortcomings of the present model concerning the shaping of activation waves. First, it appears that these waves should be defined in a smoothly graded form, rather than as step functions (see section *The shape of gestural activation waves*); and second, the gestural activation variables have no intrinsic intergestural-level dynamics that are comparable to the dynamics intrinsic to the interarticulatory coordination level of the model. For example, once the activation wave “plan” is

specified in the gestural score, the intergestural timing pattern is impervious to simulated online perturbations delivered to the model articulators or tract variables, in contrast to existing evidence in the experimental literature (e.g., Gracco and Abbs, 1989; Saltzman, Kay, Rubin, and Kinsella-Shaw, 1991).

Work is currently in progress to incorporate the dynamics of connectionist networks (e.g., Bailly, Laboissiere, and Schwartz, 1991; Grossberg, 1986; Jordan, 1986, 1990, in press; Kawato, 1989) at the intergestural level, in order to shape activation trajectories intrinsically and to allow for adaptive online interactions with the interarticulatory level of the model. In particular, we are adopting the recurrent, sequential network architecture of Jordan (1986, 1990, in press). Each output node of the network will represent a corresponding gestural activation coordinate in the current model, and the values of these output nodes will range continuously from zero to one. This will allow each gesture's attempted influence over the vocal tract to wax and wane in a smoothly graded manner. Additionally, the ongoing tract-variable state will be fed back into the sequential net, providing an informational basis for online modification of activation timing patterns to simulated perturbations delivered to the model articulator or tract-variable coordinates. Thus, rather than being explicitly and inflexibly determined prior to the onset of the simulated utterance, the activation waves will "emerge" online during the utterance as implicit consequences of the dynamics of the entire multilevel (intergestural and interarticulatory) system. Finally, given appropriate sets of side constraints during the network's learning or "programming" phase (see Jordan, 1990, for detailed discussion of these constrained optimization methods), the anticipatory fields of activation waves should be limited to durations on the order of the settling times of the activated gestures, similar to the results obtained in the present version of the model.

CONCLUDING REMARKS

We have outlined an account of speech production in which coarticulatory effects in a given utterance result primarily from context-specific interactions among invariant gestural units during periods of coproduction, rather than from context-specific alterations in the intrinsic linguistic identities of these units. In particular, we argued that context-sensitivity in acoustic or articulatory flows arises from two sources: the time courses of the activation waves for the gestures in the utterance (the "speech plan"), and the manner in which the coordinative constraints of temporally overlapping (coproduced) gestures blend or interact with one another. Although our focus was primarily on the latter source of contextual variation (e.g., in accounting for the phenomena of coarticulation resistance and aggression), we do not mean to slight the former source. For example, the relative phasings among gestures in English speech sequences are influenced by the segmental compositions of the sequences, as well as by their stress patterns, speaking rates, and syllable structures (e.g., Browman and Goldstein, 1988; Lubker, 1986; Nittrouer, 1991; Nittrouer, Munhall, Kelso, Tuller, and Harris, 1988). These results can be interpreted as reflecting corresponding variations in the relative phasings of the activation waves underlying these sequences. Additionally, cross-language variations in the degree of vowel-to-vowel coarticulation have been reported

that are functions of the relative numbers of vowels in the languages' phonetic repertoires – languages with more vowels show less extensive vowel-to-vowel effects than languages with fewer vowels (e.g., Manuel, 1987, 1990; Manuel and Krakow, 1984; but *cf.* Flege, 1989). The interpretation of these results has been that the more crowded the vowel space, the more care must be taken during production to keep the vowels perceptually distinct. In the context of the approach we have advocated in this paper, enhancing vocalic distinctiveness entails reducing the amount of overlap among vocalic activation waves. Such reduced overlap would reduce the chances that a listener would mistakenly ascribe acoustic consequences of the coarticulating vowel to the on-going vowel. This type of perceptual failure would cause the currently produced vowel to be heard as a different vowel (*cf.* Ohala, 1981; Manuel, 1987, 1990). Assuming that these differences in overlap are not simply related to corresponding language-specific differences in overall speaking rate (i.e., faster rates producing more overlap), these data suggest that the anticipatory and carryover fields of a vowel's activation wave are longer and show more overlap in languages with smaller vowel inventories. One possibility that we find most intriguing is that the patterning of activation waves, either on the fast (online) time scales of speech production or the slower (offline) time scales of language learning and sound change, may be interpreted as resulting from the dynamics of "learning" implicit in recurrent, sequential, connectionist networks, a field under intense study at this time (e.g., Jordan's presentation at this conference; see also Jordan, 1986, 1990, *in press*). A major challenge for speech research in the coming years will be to develop paradigms and techniques, both experimental and theoretical, to examine these (and possibly other) potential sources of coarticulatory variation.

REFERENCES

- ABBS, J.H., and GRACCO, V.L. (1983). Sensorimotor actions in the control of multimovement speech gestures. *Trends in Neuroscience*, **6**, 393–395.
- BAILLY, G., LABOISSIERE, R., and SCHWARTZ, J.L. (1991). Formant trajectories as audible gestures: An alternative for speech synthesis. *Journal of Phonetics*, **19**, 9–23.
- BELL-BERTI, F. (1980). Velopharyngeal function: A spatio-temporal model. In N. Lass (ed.), *Speech and Language: Advances in Basic Research and Practice* (pp. 291–316). New York: Academic Press.
- BELL-BERTI, F., and HARRIS, K. (1974). More on the motor organization of speech gestures. *Haskins Laboratories Status Reports on Speech Research*, **SR37–38**, 9–20.
- BELL-BERTI, F., and HARRIS, K. (1981). A temporal model of speech production. *Phonetica*, **38**, 9–20.
- BELL-BERTI, F., and KRAKOW, R. (1991). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America*, **90**, 112–123.
- BENGUEREL, A., and COWAN, H. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, **30**, 41–55.
- BLADON, A., and AL-BAMERNI, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, **4**, 137–150.
- BLADON, A., and AL-BAMERNI, A. (1982). One-stage and two-stage temporal patterns of coarticulation. *Journal of the Acoustical Society of America*, **72**, S104.

- BOYCE, S. (1988). *The Influence of Phonological Structure on Articulatory Organization in Turkish and English: Vowel Harmony and Coarticulation*. Ph.D. dissertation, Yale University, New Haven, CT.
- BOYCE, S. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America*, **88**, 2584–2595.
- BOYCE, S., KRAKOW, R., BELL-BERTI, F., and GELFER, C. (1990). Converging sources of evidence for dissecting articulatory movements into gestures. *Journal of Phonetics*, **18**, 173–188.
- BROWMAN, C.P., and GOLDSTEIN, L. (1986). Towards an articulatory phonology. In C. Ewan and J. Anderson (eds.), *Phonology Yearbook 3* (pp. 219–252). Cambridge, U.K.: Cambridge University Press.
- BROWMAN, C.P., and GOLDSTEIN, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, **45**, 140–155.
- BROWMAN, C.P., and GOLDSTEIN, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, 299–320.
- BROWMAN, C.P., and GOLDSTEIN, L. (1991). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology. I: Between the Grammar and the Physics of Speech* (pp. 341–376). Cambridge, U.K.: Cambridge University Press.
- BUTCHER, A., and WEIHER, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, **4**, 59–74.
- CARNEY, P., and MOLL, K. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, **23**, 193–202.
- COKER, C.H. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE*, **64**, 452–460.
- DANILOFF, R., and HAMMARBERG, R. (1973). On defining coarticulation. *Journal of Phonetics*, **1**, 239–248.
- DANILOFF, R., and MOLL, K. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research*, **11**, 707–721.
- DE JONG, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica*, **48**, 1–17.
- EDWARDS, J., BECKMAN, M.E., and FLETCHER, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, **89**, 369–382.
- FARNETANI, E. (1990). V–C–V lingual coarticulation and its spatiotemporal domain. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling* (pp. 93–130). Dordrecht: Kluwer Academic Publishers.
- FLEGE, J.E. (1989). Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech*, **32**, 123–147.
- FOWLER, C.A. (1977). *Timing Control in Speech Production*. Bloomington, Indiana: Indiana University Linguistics Club.
- FOWLER, C.A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, **8**, 113–133.
- FOWLER, C.A. (in press). Phonological and articulatory characteristics of spoken language. In G. Blanken, J. Dittman, H. Grimm, J. Marshall, and C.-W. Wallesch (eds.), *Linguistic Disorders and Pathologies: An International Handbook*. Berlin: Walter de Gruyter and Co.
- FOWLER, C.A., and SMITH, M. (1986). Speech perception as “vector analysis”: An approach to the problems of segmentation and invariance. In J. Perkell, and D. Klatt (eds.), *Invariance and Variability of Speech Processes* (pp. 123–136). Hillsdale, NJ: Lawrence Erlbaum Associates.
- FRITZELL, B. (1969). The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. *Acta Otolaryngologica*, Suppl. 250.
- GELFER, C., BELL-BERTI, F., and HARRIS, K. (1989). Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America*, **86**, 2443–2445.

- GRACCO, V.L., and ABBS, J.H. (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research*, **75**, 586–598.
- GROSSBERG, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E.C. Schwab and H.C. Nusbaum (eds.), *Pattern Recognition by Humans and Machines* (Vol. 1). New York: Academic Press.
- HARDCASTLE, W.J. (1985). Some phonetic and syntactic constraints on lingual coarticulation. *Speech Communication*, **4**, 247–263.
- HAWKINS, S. (1992). An introduction to task dynamics. In G.J. Docherty and D.R. Ladd (eds.), *Papers in Laboratory Phonology, II: Gesture, Segment and Prosody* (pp. 9–25). Cambridge, U.K.: Cambridge University Press.
- HENDERSON, J.B. (1984). *Velopharyngeal Function in Oral and Nasal Vowels: A Cross-Language Study*. Ph.D. dissertation, University of Connecticut, Storrs, CT.
- HENKE, W. (1966). *Dynamic Articulatory Models of Speech Production Using Computer Simulation*. Ph.D. dissertation, MIT, Cambridge, MA.
- JOOS, M. (1948). Acoustic phonetics. *Language*, **24**, 1–136.
- JORDAN, M.I. (1986). *Serial Order in Behavior: A Parallel Distributed Processing Approach* (Tech. Rep. No. 8604). San Diego: University of California, Institute for Cognitive Science.
- JORDAN, M.I. (1990). Motor learning and the degrees of freedom problem. In M. Jeannerod (ed.), *Attention and Performance XIII* (pp. 796–836). Hillsdale, NJ: Erlbaum.
- JORDAN, M.I. (in press). Serial order: A parallel distributed processing approach. In J.L. Elman and D.E. Rumelhart (eds.), *Advances in Connectionist Theory: Speech*. Hillsdale, NJ: Erlbaum.
- JORDAN, M.I., and ROSENBAUM, D.A. (1989). Action. In M.I. Posner, (ed.), *Foundations of Cognitive Science* (pp. 727–767). Cambridge, MA: MIT Press.
- KAWATO, M. (1989). Motor theory of speech perception revisited from minimum torque-change neural network model. *Proceedings of the 8th Symposium on Future Electron Devices, October 30–31, Tokyo, Japan* (pp. 141–150).
- KEATING, P. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology. I: Between the Grammar and the Physics of Speech* (pp. 451–470). Cambridge, U.K.: Cambridge University Press.
- KELSO, J.A.S., SALTZMAN, E.L., and TULLER, B. (1986a). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, **14**, 29–60.
- KELSO, J.A.S., SALTZMAN, E.L., and TULLER, B. (1986b). Intentional contents, communicative context, and task dynamics: A reply to the commentators. *Journal of Phonetics*, **14**, 171–196.
- KELSO, J.A.S., TULLER, B., VATIKIOTIS-BATESON, E., and FOWLER, C.A. (1984). Functionally-specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 812–832.
- KELSO, J.A.S., VATIKIOTIS-BATESON, E., SALTZMAN, E.L., and KAY, B. (1985). A qualitative dynamic analysis of reiteration speech production: Phase portraits, kinematics and dynamic modeling. *Journal of the Acoustical Society of America*, **77**, 266–280.
- KENT, R., CARNEY, P., and SEVEREID, L. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research*, **17**, 470–488.
- LAVER, J. (1980). Slips of the tongue as neuromuscular evidence for a model of speech production. In H.W. Dechert and M. Raupach (eds.), *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton.
- LIBERMAN, A., and MATTINGLY, I. (1985). The motor theory revised. *Cognition*, **21**, 1–36.
- LUBKER, J. (1986). Articulatory timing and the concept of phase. *Journal of Phonetics*, **14**, 133–137.
- LÖFQVIST, A. (1990). Speech as audible gestures. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling* (pp. 289–332). Dordrecht: Kluwer Academic Publishers.
- MAGEN, H.S. (1989). *An Acoustic Study of Vowel-to-Vowel Coarticulation in English*. Ph.D. dissertation, Yale University, New Haven, CT.

- MANUEL, S. (1987). *Acoustical and Perceptual Consequences of Vowel-to-Vowel Coarticulation in Three Bantu Languages*. Ph.D. dissertation, Yale University, New Haven, CT.
- MANUEL, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, **88**, 1286–1298.
- MANUEL, S.Y., and KRAKOW, R.A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Reports on Speech Research*, **SR-77/78**, 69–78.
- MARCHAL, A. (1988). Coproduction: Evidence from EPG data. *Speech Communication*, **7**, 287–295.
- MATTINGLY, I.G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, and J. Anderson (eds.), *The Cognitive Representation of Speech* (pp. 415–420). Amsterdam: North-Holland.
- MATTINGLY, I.G. (1990). The global character of phonetic gestures. *Journal of Phonetics*, **18**, 445–452.
- MOLL, K. (1962). Velopharyngeal closure in vowels. *Journal of Speech and Hearing Research*, **36**, 30–77.
- MOLL, K., and DANILOFF, R. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, **50**, 678–684.
- MOWREY, R.A., and MACKAY, I.R.A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, **88**, 1299–1312.
- MUNHALL, K.G., and KELSO, J.A.S. (1985). Phase-dependent sensitivity to perturbation reveals the nature of speech coordinative structures. *Journal of the Acoustical Society of America*, **78**, Suppl. 1, S38.
- MUNHALL, K., and LÖFQVIST, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, **20**, 111–126.
- MUNHALL, K.G., LÖFQVIST, A., and KELSO, J.A.S. (1986). Laryngeal compensation following sudden oral perturbation. *Journal of the Acoustical Society of America*, **80**, Suppl. 1, S109.
- NITTROUER, S. (1991). Phase relations of jaw and tongue tip movements in the production of VCV utterances. *Journal of the Acoustical Society of America*, **90**, 1806–1815.
- NITTROUER, S., MUNHALL, K., KELSO, J.A.S., TULLER, B., and HARRIS, K.S. (1988). Patterns of interarticulator phasing and their relation to linguistic structure. *Journal of the Acoustical Society of America*, **84**, 1653–1661.
- OHALA, J. (1981). The listener as a source of sound change. In C. Marek, R. Hendrick and M. Miller (eds.), *Papers from the Parasession on Language and Behavior* (pp. 178–203). Chicago: Chicago Linguistics Society.
- ÖHMAN, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, **39**, 151–168.
- ÖHMAN, S. (1967). Numerical models of coarticulation. *Journal of the Acoustical Society of America*, **41**, 310–320.
- PATTEE, H.H. (1976). Physical theories of biological coordination. In M. Grene and E. Mendelsohn (eds.), *Readings in the Philosophy of Biology* (pp. 153–173). Dordrecht: Reidel.
- PERKELL, J.S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, MA: MIT Press.
- PERKELL, J.S. (1990). Testing models of speech production: Implications of some detailed analysis of variable articulatory data. In W. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling* (pp. 263–288). Dordrecht: Kluwer Academic Publishers.
- PERKELL, J.S. (1991). Models, theory, and data in speech production. *Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 1* (pp. 182–191). Aix-en-Provence, France: Université de Provence Service des Publications.
- PERKELL, J.S., and CHIANG, C. (1986). Preliminary support for a 'hybrid model' of anticipatory coarticulation. *Proceedings of the 12th International Congress of Acoustics* (pp. A3–A6).

- PERKELL, J.S., and MATTHIES, M.L. (1992). Temporal measures of labial coarticulation for the vowel /u/: Within- and cross-subject variability. *Journal of the Acoustical Society of America*, **91**, 2911–2925.
- RECASENS, D. (1984a). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. *Journal of Phonetics*, **12**, 61–73.
- RECASENS, D. (1984b). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, **76**, 1624–1635.
- RECASENS, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, **15**, 299–312.
- RECASENS, D. (1991). An electropalatal and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics*, **19**, 177–192.
- RUBIN, P.E., BAER, T., and MERMELSTEIN, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, **70**, 321–328.
- SALTZMAN, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer and C. Fromm (eds.), *Generation and Modeling of Action Patterns* (pp. 129–144). New York: Springer-Verlag.
- SALTZMAN, E. (1991). The task dynamic model in speech production. In H.F.M. Peters, W. Hulstijn, and C.W. Starkweather (eds.), *Speech Motor Control and Stuttering* (pp. 37–52). Amsterdam: Excerpta Medica.
- SALTZMAN, E., KAY, B., RUBIN, P., and KINSELLA-SHAW, J. (1991). Dynamics of intergestural timing. *Perilus XIV*, Institute of Linguistics, University of Stockholm, Stockholm, Sweden (pp. 47–55).
- SALTZMAN, E., and KELSO, J.A.S. (1987). Skilled action: A task-dynamic approach. *Psychological Review*, **94**, 84–106.
- SALTZMAN, E., and MUNHALL, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, **1**, 333–382.
- SHAIMAN, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America*, **86**, 78–88.
- SHAIMAN, S., and ABBS, J.H. (1987). *Phonetic Task-specific Utilization of Sensorimotor Actions*. Paper presented at annual convention of the American Speech-Language-Hearing Association. November. New Orleans, LA.
- TURVEY, M.T. (1990). Coordination. *American Psychologist*, **45**, 938–953.
- USHIJIMA, T., and SAWASHIMA, M. (1972). Fiberscopic examination of velar movements during speech. *Annual Bulletin of the Research Institute of Logopaedics and Phoniatrics*, 25–38.
- VATIKIOTIS-BATESON, E. (1988). *Linguistic Structure and Articulatory Dynamics*. Ph.D. dissertation, Indiana University, Bloomington, IN (distributed by the Indiana University Linguistics Club, Bloomington, IN).