In: Broe, M.B., & Pierrehumbert, J.B. (eds.) Papers in Laboratory Phonology V Acquistion and the Lexicon., Cambridge University Press, 2000.

6

'Glue' and 'clocks': intergestural cohesion and global timing

ELLIOT SALTZMAN, ANDERS LÖFQVIST AND SUBHOBRATA MITRA

6.1 Introduction

In this chapter, we present some recent experimental and simulation results within a dynamical systems framework and describe their implications for issues in laboratory phonology. The experimental work entails application of phase-resetting techniques in which mechanical perturbations are delivered to the articulators during speaking, and the resultant changes in the utterance's temporal structure are analyzed. This work is a subset of a more extensive data set described elsewhere (Saltzman, Löfqvist, Kay, Kinsella-Shaw & Rubin in press). These data are used to compare the degrees of cohesion among speech gestures—the strength of intergestural 'glue'—both within and between traditional segmental units of articulation. Following the description of these experimental results, we address preliminary simulation work that focuses on how linguistically conditioned modulations of speaking rate might be modeled within the task-dynamic model of gestural patterning (Saltzman & Munhall 1989). We describe the results of implementing a simple 'clocking' mechanism, and briefly review its implications for interpreting temporal variations in the articulation of syllable-sized units of speech.

6.2 'Glue'

In a dynamical systems framework (e.g. Browman & Goldstein 1990c, Saltzman & Munhall 1989), gestures are linguistically significant units of articulation that shape vocal tract activity over time. The postulation of segmental units in phonology implies that there is some degree of cohesion among the gestures that constitute these segments. In any given utterance, we expect that the cohesion among gestures within segmental units is stronger, in some sense, than the

1169

cohesion among gestures of different segmental units. However, the nature and origin of this intergestural 'glue' are issues that require empirical study. We have hypothesized from a dynamical systems approach that intergestural cohesion can be accounted for by coupling structures defined among gestural units (Saltzman & Munhall 1989). If so, evidence of such coupling should be experimentally observable.

We report results from a series of phase-resetting studies of speech production that investigate whether intergestural temporal cohesion is greater within segments than between segments. Phase-resetting techniques were pioneered in studies of the effects of perturbations on the temporal structure of general biological rhythms (e.g. Glass & Mackey 1988, Kawato 1981, Winfree 1980). In particular, such analyses are used to determine whether perturbations delivered during an ongoing rhythm have a permanent effect (i.e., phase shift) on the underlying temporal organization of the rhythm. Phase-resetting techniques have been used in many kinematic and neurophysiological studies of the control and coordination of rhythmic movements (e.g. Lennard & Hermanson 1985, Lee & Stein 1981). In such studies, what is measured is the amount of temporal shift introduced by the perturbation relative to the timing pattern that existed prior to the perturbation. This phase shift is measured after the transient, perturbation-induced distortions to the rhythm have subsided, and the system has returned to its pre-perturbation rhythmic state.

In our experiment, we focus on the coordination of bilabial closing and laryngeal devoicing gestures for /p/s both within and between successive syllable onsets in the repetitive sequence /...pæpæpæ.../. Using phase-resetting methods, we apply mechanical perturbations to the lips during these sequences and examine the intergestural phase shifts that result from the perturbations. We interpret these data in terms of models of speech production that posit a central timing network or 'clock' underlying the production of such sequences (e.g. Saltzman & Munhall 1989). Finding perturbation-induced phase-shifts would imply that the hypothesized central timekeeper could not simply drive the articulators unidirectionally. Rather, the hypothesized central timer and the peripheral articulators must influence one another bidirectionally, so that feedback from articulatory events can influence (i.e., phase-shift) the state of the timer.

6.2.1 Method

Two males with no history of language impairment were subjects, one a native speaker of American English (ES) and the other a native Swedish speaker (AL) fluent in American English. Each subject sat in an adjustable dental chair, with the head restrained in a fixed frame. A small paddle connected to a torque motor was placed on the lower lip with a tracking force of 3 gm, in order to deliver step pulses of downward force (50 gm) at random times during the experimental

6

'Glue' and 'clocks': intergestural cohesion and global timing

ELLIOT SALTZMAN, ANDERS LÖFQVIST AND SUBHOBRATA MITRA

6.1 Introduction

In this chapter, we present some recent experimental and simulation results within a dynamical systems framework and describe their implications for issues in laboratory phonology. The experimental work entails application of phase-resetting techniques in which mechanical perturbations are delivered to the articulators during speaking, and the resultant changes in the utterance's temporal structure are analyzed. This work is a subset of a more extensive data set described elsewhere (Saltzman, Löfqvist, Kay, Kinsella-Shaw & Rubin in press). These data are used to compare the degrees of cohesion among speech gestures—the strength of intergestural 'glue'—both within and between traditional segmental units of articulation. Following the description of these experimental results, we address preliminary simulation work that focuses on how linguistically conditioned modulations of speaking rate might be modeled within the task-dynamic model of gestural patterning (Saltzman & Munhall 1989). We describe the results of implementing a simple 'clocking' mechanism, and briefly review its implications for interpreting temporal variations in the articulation of syllable-sized units of speech.

6.2 'Glue'

In a dynamical systems framework (e.g. Browman & Goldstein 1990c, Saltzman & Munhall 1989), gestures are linguistically significant units of articulation that shape vocal tract activity over time. The postulation of segmental units in phonology implies that there is some degree of cohesion among the gestures that constitute these segments. In any given utterance, we expect that the cohesion among gestures within segmental units is stronger, in some sense, than the

cohesion among gestures of different segmental units. However, the nature and origin of this intergestural 'glue' are issues that require empirical study. We have hypothesized from a dynamical systems approach that intergestural cohesion can be accounted for by coupling structures defined among gestural units (Saltzman & Munhall 1989). If so, evidence of such coupling should be experimentally observable.

We report results from a series of phase-resetting studies of speech production that investigate whether intergestural temporal cohesion is greater within segments than between segments. Phase-resetting techniques were pioneered in studies of the effects of perturbations on the temporal structure of general biological rhythms (e.g. Glass & Mackey 1988, Kawato 1981, Winfree 1980). In particular, such analyses are used to determine whether perturbations delivered during an ongoing rhythm have a permanent effect (i.e., phase shift) on the underlying temporal organization of the rhythm. Phase-resetting techniques have been used in many kinematic and neurophysiological studies of the control and coordination of rhythmic movements (e.g. Lennard & Hermanson 1985, Lee & Stein 1981). In such studies, what is measured is the amount of temporal shift introduced by the perturbation relative to the timing pattern that existed prior to the perturbation. This phase shift is measured after the transient, perturbation-induced distortions to the rhythm have subsided, and the system has returned to its pre-perturbation rhythmic state.

In our experiment, we focus on the coordination of bilabial closing and laryngeal devoicing gestures for /p/s both within and between successive syllable onsets in the repetitive sequence /...pæpæpæ.../. Using phase-resetting methods, we apply mechanical perturbations to the lips during these sequences and examine the intergestural phase shifts that result from the perturbations. We interpret these data in terms of models of speech production that posit a central timing network or 'clock' underlying the production of such sequences (e.g. Saltzman & Munhall 1989). Finding perturbation-induced phase-shifts would imply that the hypothesized central timekeeper could not simply drive the articulators unidirectionally. Rather, the hypothesized central timer and the peripheral articulators must influence one another bidirectionally, so that feedback from articulatory events can influence (i.e., phase-shift) the state of the timer.

6.2.1 Method

Two males with no history of language impairment were subjects, one a native speaker of American English (ES) and the other a native Swedish speaker (AL) fluent in American English. Each subject sat in an adjustable dental chair, with the head restrained in a fixed frame. A small paddle connected to a torque motor was placed on the lower lip with a tracking force of 3 gm, in order to deliver step pulses of downward force (50 gm) at random times during the experimental

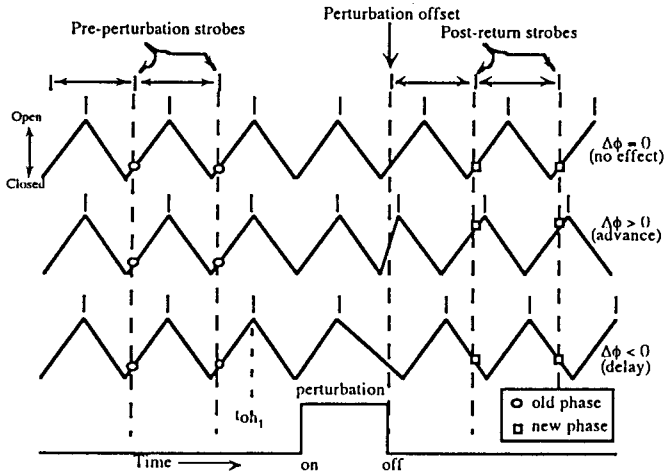

Figure 6.2 Schematic display of constriction (bilabial or glottal) aperture and perturbation trajectories for pre-perturbation, perturbation, and post-perturbation cycles. The offset of the perturbation is used as a temporal anchor point for strobing used to define old (open circles) and new (open squares) phases. Phase shift ($\Delta\phi$) is the average new phase minus the average old phase.

The relative phasing of the bilabial and laryngeal trajectories was computed as follows. Successive peak laryngeal openings were used to define 'strobe' events in each corresponding bilabial cycle. Relative phase for each laryngeally-strobed bilabial cycle could then be defined by:

$$\frac{[\text{(time of the laryngeal event)} - \text{(time of the preceding bilabial peak)}]}{\text{(period of the strobed bilabial cycle)}}$$

Average relative phase for each perturbed trial's strobed bilabial pre-perturbation cycles was defined computationally as [*bilabial* $\overline{\phi_{old}}$ – *laryngeal* $\overline{\phi_{old}}$]; average relative phase for the strobed bilabial post-return cycles was defined computationally as [*bilabial* $\overline{\phi_{new}}$ – *laryngeal* $\overline{\phi_{new}}$]. Shifts in relative phase for each trial were then defined by the average postreturn relative phase minus the average pre-perturbation relative phase.

The same cycle types and experimental measures were obtained for the control (no perturbation) trials, where calculations were anchored to the end of a randomly timed, but not delivered, 'phantom perturbation.'

For each perturbed trial, the measures of phase shift and shifts in relative phase were converted to (*experimental* – $\overline{control}$) difference scores, where $\overline{control}$ equals the session-specific control values computed for each measure averaged across all of the 'phantom perturbation' times. These difference scores were

partitioned into five bins and averaged according to a normalized measure of perturbation delivery time defined according to the events and cycle types shown in Figure 6.2: (*perton* – t_{on_1}) / $\overline{preperi}$, where *perton* = onset time of perturbation, t_{on_1} = onset time of the first-perturbation bilabial cycle, and $\overline{prepert}$ = the average duration of the bilabial pre-perturbation cycles for the trial. Thus, each bin includes a portion of the bilabial cycle during which perturbations occurred. For each difference measure (i.e., bilabial and laryngeal phase shifts; shifts in relative phase), separate *t*-tests were computed for each perturbation bin to test whether the measures differed from zero. To protect against an elevated rate of Type I errors (i.e., asserting that differences exist when in fact no such differences exist) due to multiple comparisons across perturbation bins, α-levels (significance levels) were selected by dividing .01 and .05 by the number of bins (i.e., 5).

Difference scores that significantly differed from zero indicated that the perturbations had reset an underlying central timing network for speech production. The amount of within-articulator resetting that occurred between successive syllable onsets (i.e., phase shifts) was interpreted as a measure of the strength of intergestural coupling between segments; the amount of resetting between articulators that occured within syllable onsets (i.e., shifts in relative phase) was interpreted as a measure of the intergestural coupling strength within segments.

6.2.3 Results

Figure 6.3 shows the results of the analyses of bilabial and laryngeal phase shifts, and shifts in relative phases, for subjects ES (left panel) and AL (right panel). For ES, protected *t*-tests showed that the bilabial and laryngeal rhythms were significantly phase advanced relative to the no-perturbation control trials only in the .11 and .99 time bins. The bilabial pattern replicates the phase-resetting results found in earlier studies of this speaker that focused only on bilabial behavior (Saltzman 1992, Saltzman, Kay, Rubin & Kinsella-Shaw 1991). Protected *t*-tests also showed a significant shift (negative) in the relative phasing of lips and larynx relative to control values only in the .99 time bin. In order to detect differences across the time bins in phase shifts and shifts of relative phase, one-way ANOVAs were performed separately for the bilabial, laryngeal, and relative phase data. There were significant main effects for the bilabial ($F(4,175) = 7.9$, $p < .001$) and laryngeal ($F(4,175) = 8.62$, $p < .001$) phase shifts. Post-hoc Tukey tests indicated that for this subject's bilabial and laryngeal data, phase shifts were greater in bins .11 and .99 than in bins .33, .55, and .77. There was no significant main effect of bin for shifts in relative phase ($F(4,175) = .54$, $p = .7$).

Schwartz 1991, Lathroum 1989) in current simulations focusing on the behavior of a simplified model system for producing gestural movement patterns. Our hybrid model combines a sequential neural network at the intergestural level with a task-dynamic model at the interarticulator level (Figure 6.4). In the hybrid model, the output of the sequential network represents the activation, a, of gestures defined in a single tract variable. This activation node acts only to inject a time-varying input into the tract-variable dynamics that represents the current value of the gestural target parameter, T. Thus, $T(t) = a(t)$, and in each simulation run a begins at a value of zero and changes over time to the desired gestural target value, T_d, where $-1 \leq T_d \leq 1$. The gestural stiffness, k, and damping, b, coefficients are fixed. The tract-variable is represented by a linear unit (the filled square in Figure 6.4 that receives input directly from the gestural activation node) whose inputs are current target value ($T = a$), tract-variable position (x), and tract-variable velocity ($\dot{x}$). The unit's output is current tract-variable acceleration ($\ddot{x}$), which is fed into a cascade of two linear, self-recurrent units that provide discrete-time, Euler integrations to generate the next tract-variable velocity and position, respectively. In each simulation run, tract-variable position and velocity are set to initial values of zero. Taken together, the tract-variable and integrator units represent a model of the *forward dynamics* from current tract-variable state (i.e., position & velocity) and target inputs to the next tract-variable state. Additionally, delay lines from the integrator units are used to feed back the current tract-variable state to the tract-variable acceleration unit and, via the task-dynamic state units, to the hidden units of the sequential network.

During the network training stage, *teaching vectors* are used to define the time intervals of desired target attainment for a given utterance. These intervals are called 'care' intervals; during the care intervals, output errors (desired target position, T_d, minus current tract-variable position, x) are measured. At all other times, the teaching vector defines 'don't care' conditions for which no errors are defined. During the 'care' intervals, however, the errors are propagated backward through the fixed tract-variable forward dynamics and applied to the sequential net's output units. From this point on, these back-propagated errors are used to train the inter-unit connection weights inside the sequential net. Because of the feedback implicit in the hybrid system's dynamics (e.g. the self-connections of the integrator units; the tract-variable state connections to the acceleration unit), the *back-propagation through time* training procedure (e.g. Rumelhart, Hinton & Williams 1986) is used. Network training and performance of a given sequence are done in the presence of a corresponding constant, biasing input from the plan units.

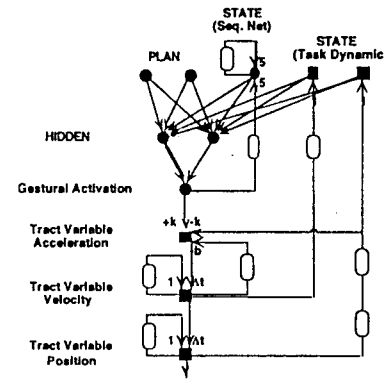

Figure 6.4 Architecture of the current hybrid network. Filled circles = sequential network elements; filled squares = task-dynamic elements; open 'lozenges' label delay lines; arrowheads = inter-element synapses; numbers and symbols denote the fixed weights assigned to some synapses.

6.3.1.1 Anticipatory interval of coarticulation

Our initial efforts focused on modeling a given gesture's anticipatory interval of coarticulation. This is defined operationally as the time from motion onset to the time of required target attainment. In a review of the literature, Fowler & Saltzman (1993) concluded that anticipatory coarticulation intervals are temporally constrained, and that they do not typically extend very far backward in time from the time of target attainment (but see Abry & Lallouache 1995, Lubker 1981, Perkell & Matthies 1992). This interpretation is consistent with that provided by Bell-Berti & Harris' (1981) *frame* model of coarticulation, and contrasts with the extensive degrees of anticipation that are possible in look-ahead models (e.g. Henke 1966; in fairness, it should be noted that the anticipatory feature-spreading used in Henke's model looked ahead only to the immediately following segment, although unlimited anticipation was allowed in principle).

Our hybrid model is capable of generating both unconstrained look-ahead and appropriately constrained, frame-like intervals of anticipatory coarticulation. In particular, we tested the hypothesis that 'frame-model-like' behavior requires the addition of side constraints to the sequential net's output units, uniformly applied during both the 'care' and 'don't-care' intervals of the network's training phase. Side constraints are typically chosen to reflect certain generic properties, such as effort minimization or maximization of movement smoothness, of the skilled activities being modeled (for detailed discussion of such constrained optimization methods, see Jordan 1992). We used a side constraint during training that penalized activity of the sequential network's output units (i.e., gestural activation units) and minimized gestural activation 'effort'. When this side constraint was used, the anticipatory intervals were constrained

As a consequence of Δt (clock) modulation, gestures with larger distances to travel tend to slow the global clock more than gestures with smaller distances to travel, and the overall system displays 'the farther the longer' behavior that has often been reported to be typical of both speech and limb movements (e.g. Kelso Kelso, Vatikiotis-Bateson, Saltzman & Kay 1985, Ostry & Munhall 1985). That is, durations are longer for simulated gestures with larger displacements from initial position to target (see Figure 6.5). These local variations in gestural duration are necessarily associated with concomitant variations in global utterance length (see Section 6.3.1.2).

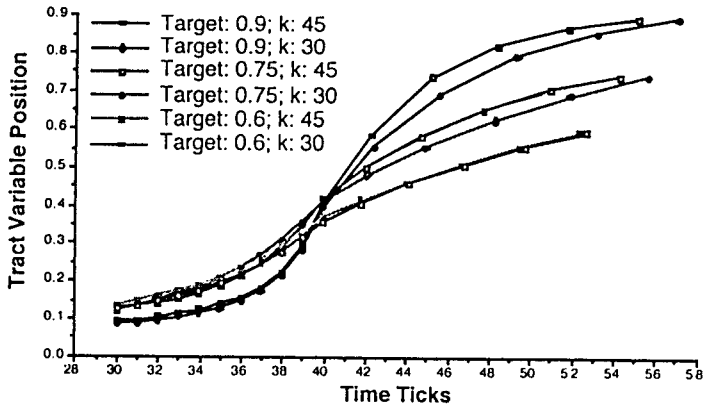

Figure 6.5 Clock modulation effects on tract-variable trajectories for single gestures that differ in target positions (T) for a set of three different gestural stiffnesses (k).

Such a modulable global clocking mechanism may provide hints for understanding the origins of a variety of speech timing phenomena (e.g. variations in rate and stress, final lengthening) through manipulation of a single variable—clock speed. For example, the voicing status of syllable-final consonants has durational consequences that are distributed temporally throughout the syllable (see, e.g. Lisker 1986). These results are consistent with the requirement for voicing/devoicing acting to modulate the rate of timeflow in an underlying clock whose period defines syllabic duration (e.g. Port & Cummins 1992). It has been proposed that such an hypothesized 'syllable clock' could be used to phase gestures within syllable-sized units, with each gesture temporally anchored to an associated critical phase angle (e.g. corresponding to onset, nucleus, coda) of the clock (see also Bailly, Laboissière, & Schwartz 1991, Lame 1990, Vatikiotis-Bateson, Hirayama, Honda & Kawato 1992). The global clocking subnetwork we have described will allow us to begin to explore these and other proposals of the effects of modulating speaking rate within the framework of our dynamical

model of gestural patterning (see also Byrd, Kaun, Narayanan & Saltzman, this volume, for a discussion of clock slowing in the context of prosodic lengthening).

6.4 Summary and conclusions

We have outlined some recent experimental and simulation results that address intergestural cohesion and temporal variation in articulatory gestures within a dynamical systems framework. In Section 6.2, we overviewed an experiment comparing the relative degrees of cohesion (i.e., the strength of intergestural 'glue') among speech gestures composing segments and between gestures in different segments. These results suggested (1) that gestures within segments demonstrate greater temporal cohesion than those between segments; and (2) that a central clock coordinating intergestural timing can be reset by peripheral events during the articulation of a gesture. In Section 6.3, we presented simulations incorporating a clocking mechanism into a hybrid dynamical model (sequential neural network + task dynamics) that begin to illuminate how linguistically conditioned modulations of speaking rate might be simulated within the task-dynamic model of gestural patterning. The model that we have investigated in these simulations yields several behaviors consistent with those characteristic of speech: (1) limited anticipatory look-ahead in coarticulation; (2) a relation between gestural displacement and duration that captures the fact that the farther a gesture must move the longer it tends to take; and (3) changes in overall utterance duration due to local changes in gestural duration. We are hopeful that these empirical and theoretical approaches to studying the temporal patterning of articulatory gestures can help to form a conceptual link between the underlying dynamics of articulatory control and coordination and the temporal structure of language.

Notes

This work was supported by NIH Grants DC-00121, DC-00865, DC-03663, NSF Grant DBS-9112198, and in part by Esprit-BR Project 6975-Speech Maps through NUTEK Grant P55-1. We are grateful to Dani Byrd and Abigail Kaun for discussion and helpful comments on drafts of this paper.

1 These trials were part of a more extensive experiment (Saltzman et al., in press) in which the temporal responses to perturbation in repetitive utterances were compared to those in discrete, more word-like utterances.