Cognitive Science 33 (2009) 839–879

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01037.x

Multitimescale Dynamical Interactions Between Speech Rhythm and Gesture

Sam Tilsen

Department of Linguistics, University of California, Berkeley

Received 22 June 2008; received in revised form 28 December 2008; accepted 5 January 2009

Abstract

Temporal patterns in human movement, and in speech in particular, occur on multiple timescales. Regularities in such patterns have been observed between speech gestures, which are relatively quick movements of articulators (e.g., tongue fronting and lip protrusion), and also between rhythmic units (e.g., syllables and metrical feet), which occur more slowly. Previous work has shown that patterns in both domains can be usefully modeled with oscillatory dynamical systems. To investigate how rhythmic and gestural domains interact, an experiment was conducted in which speakers performed a phrase repetition task, and gestural kinematics were recorded using electromagnetic articulometry. Variance in relative timing of gestural movements was correlated with variance in rhythmic timing, indicating that gestural and rhythmic systems interact in the process of planning and producing speech. A model of rhythmic and gestural planning oscillators with multifrequency coupling is presented, which can simulate the observed covariability between rhythmic and gestural timing.

Keywords: Speech gesture; Motor coordination; Task dynamics; Coupled oscillators; Dynamical systems; Metrical theory; Speech rhythm

1. Introduction

Many human behaviors involve systems whose dynamics are associated with multiple timescales. In speech, these systems are normally referred to—from shortest to longest timescale—as gestures, moras, syllables, feet, and phrases. Gestural systems govern articulator movements, such as the fronting of the tongue for an [s] or the raising and protrusion of the lower lip for a [p]. Rhythmic systems coordinate groups of gestures or other rhythmic systems, for example, a phrase contains multiple metric feet, and a foot contains multiple syllables. Rhythmic systems exert their influence over longer periods of time than gestural

Correspondence should be sent to Sam Tilsen, Department of Linguistics, University of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 97201. E-mail: tilsen@berkeley.edu

systems. Both gestural and rhythmic systems have been described independently with dynamical models by various researchers (cf. Goldstein & Fowler, 2003; Saltzman, Nam, Krivokapic, & Goldstein, 2008; Van Lieshout, 2004), and a natural question to ask is whether these two domains interact dynamically.

This question is related to a more general line of dynamical systems-inspired research that has deeply explored interlimb coordination (Beek, Peper, & Daffertshofer, 2002) and which has broad applicability to domains such as music performance (Gabrielsson, 2003; Palmer, 1997), gesture and sign language (Goldin-Meadow, in press; So, Kita, & Goldin-Meadow, in press), social behavior (Castellano, Fortunato, & Loreto, 2007), biological rhythms (Winfree, 1980), and artificial intelligence (Beer, 1995). The results of this research point to general mechanisms that organize action and cognition on multiple timescales. Moreover, the study of motor coordination is becoming increasingly important in cognitive science due to the recent surge of evidence that motor simulation plays a key role in perception, thought, empathy, and the evolution of language (Arbib, 2005; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Oberman & Ramachandran, 2007).

A major innovation in the past several decades has been the application of concepts from theories of self-organization in physical systems to the understanding of various cognitive-behavioral patterns involving the coordinated movement of multiple effectors. Kelso (1981, 1984), to better conceptualize rhythmic interlimb coordination, employed the concept of an *order parameter* or *collective variable* that had been developed by physicist Hermann Haken. An order parameter is a low-dimensional variable that describes the collective behavior of a self-organized system composed of many individual systems driven far from thermal equilibrium (Haken, 1975, 1983; cf. Kelso, 1995, for an historical introduction to dynamical modeling of pattern formation processes and movement behavior). An oft-cited example is the Rayleigh–Benard experiment, in which oil heated in a pan forms convection rolls. The individual molecules in the system are enslaved to an orderly, coherent pattern of motion. Rather than needing to describe the motions of all these individual molecules, a collective variable can be used to describe the rolling motion, thereby using fewer degrees of freedom in the analysis of the system.

A central concept used in modeling the rhythmic coordination of individual effectors is that of *relative phase*. In general, relative phase is a collective variable that corresponds to the phase difference between oscillatory systems, each having its own well-defined individual phase that can typically be formulated as an angle in polar coordinates on the system's phase plane. In normal walking, the relative phase of the leg movements is approximately 180° (or π radians, or 0.5 unit phase), which is known as anti-phase coordination. In symmetric hopping, where both feet leave and return to the ground at the same time, the relative phase is 0, which is called in-phase coordination. If systems are coupled so as to mutually influence each other, and they exhibit some degree of stability in relative phase, the systems can be referred to as *synchronized*. One of the major insights into studies of coordinated movement has been that some relative phase relations in coupled systems are more stable than others. In a classic paper, Haken, Kelso, and Bunz (1985) reported that as movement frequency (a *control parameter*) is increased, anti-phase bimanual finger wagging movements become unstable; the relative phase of the system then undergoes a transition to the

more stable in-phase mode of finger wagging. Fig. 1A illustrates with simulated data a typical pattern of motion in a bimanual finger wagging experiment, and Fig. 1B shows how the relative phase (φ) undergoes a transition as movement frequency increases.

To represent the differential stability of coordinative modes of relative phase, the concept of a *potential* has also been borrowed from physics. In this context, the potential is a function that describes the force acting upon a collective variable, and its minima and maxima are the stable and unstable equilibria of the system. Haken, Kelso, and Bunz modeled the stability of relative phase in bimanual finger wagging using a potential function that is the superposition of two cosine waves with a 1:2 frequency ratio, that is, $V(\varphi) = -a\cos(\varphi) - b\cos(2\varphi)$. They hypothesized that the control parameter of movement frequency corresponds to the ratio of amplitudes of the components of the potential function (b/a). As the frequency increases, the ratio changes and the stable equilibrium corresponding to anti-phase coordination becomes unstable. Fig. 2 shows potential functions corresponding to several values of this ratio. The reader should observe how the anti-phase mode becomes unstable at a critical ratio of 0.25. The potential can be thought of as a force field that directs the change of relative phase, such that the relative phase velocity is the negative of the derivative of the potential with respect to relative phase [$\mathrm{d}\varphi/\mathrm{d}t = -\mathrm{d}V/\mathrm{d}t$]. This means that the force advances or delays the phases of component oscillators so that the relative phase moves toward the nearest local minimum in the potential function. An oft-used analogy to this is the motion of a ball (relative phase) rolling down a sticky hill (potential), which stops when it reaches the bottom (a local minimum).

Another conceptual advance in this framework has been the notion of *generalized relative phase*, which allows for a stable relative phase to be defined between systems of differing frequencies (cf. Keith & Rand, 1984; Pikovsky, Rosenblum, & Kurths, 2001; Saltzman

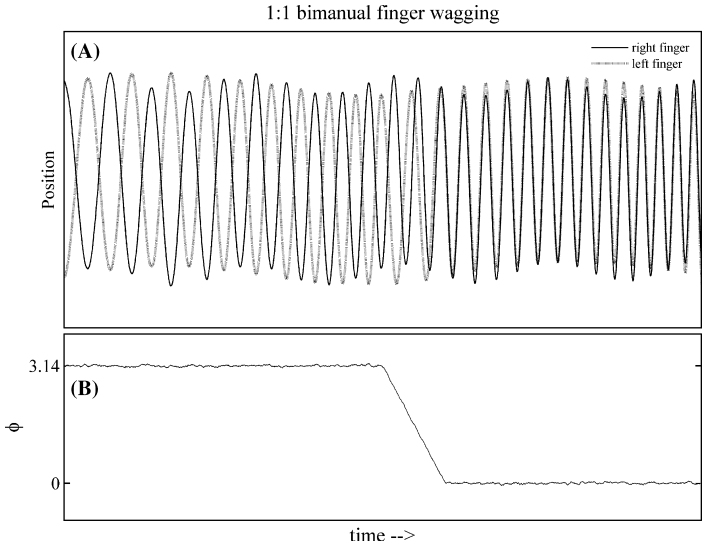

Fig. 1. Schematized phase transition from anti-phase to in-phase coordination in a bimanual finger wagging task, generated from a simulation with noise. (A) Displacements of the right and left index fingers from the midline. (B) Relative phase of the fingers.

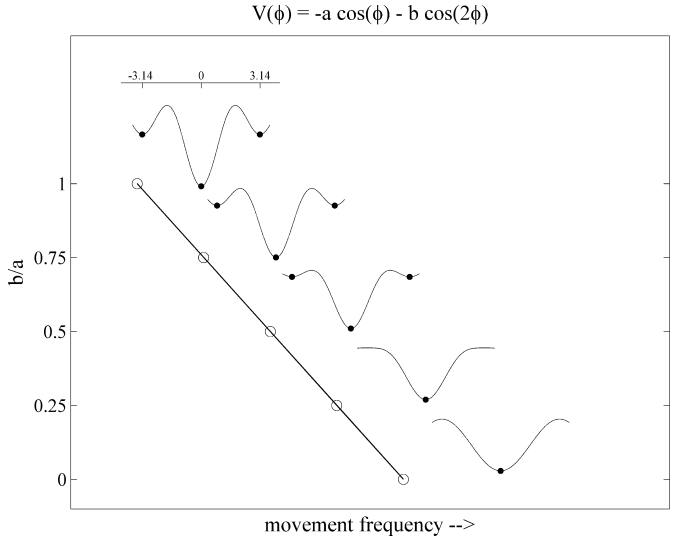

Fig. 2. Changes in the relative phase potential as movement frequency increases. Potential functions for five values of the amplitude ratio *b/a* are shown next to their corresponding values on the line, which relates movement frequency to the ratio *b/a*.

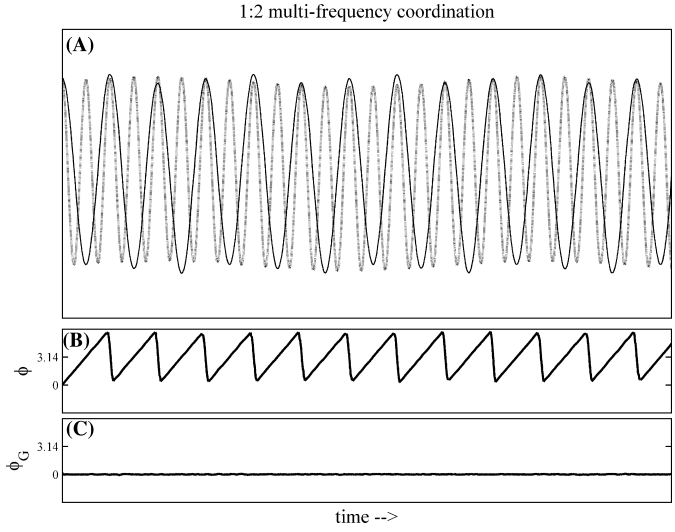

Fig. 3. A simulated example of 1:2 frequency-locking. (A) Position time-series of two hypothetical effectors. (B) Relative phase, which is periodic. (C) Generalized relative phase, which is constant.

& Byrd, 2000; Sternad, Turvey, & Saltzman, 1999). To see why this is necessary, consider Fig. 3A, which shows the motions of two hypothetical effectors in a 1:2 frequency-locked system. Fig. 3B shows that if no phase transformation is performed, then the relative phase is periodic when wrapped on the phase circle modulo 2π (or equivalently, drifts linearly over time when unwrapped). It is desirable for theoretic purposes to identify an order para-

meter whose constant value reflects a particular steady-state temporal relationship between two oscillations. For two frequency-locked oscillatory systems with individual phases θ_1 and θ_2, whose inherent frequencies ω_1 and ω_2 approximate an *m:n* ratio, the generalized relative phase $\varphi_G = \omega_1\theta_2 - \omega_2\theta_1$ is constant over time, as shown in Fig. 3C. For example, in the case of 1:2 frequency-locking, the phase of the slower oscillator is multiplied by the frequency of the faster oscillator prior to taking the phase difference.

The stability of various modes of frequency-locking can be described with potential functions of generalized relative phase variables (Saltzman & Byrd, 2000). A more general approach, in which the potential function is a superposition of different modes of frequency-locking with differing amplitudes, is described by Haken, Peper, Beek, and Daffertshofer (1996), who used this idea to model phase transitions between frequency-locked modes of drumming performed by skilled musicians. They observed that relatively high-order modes of frequency-locking such as 2:5 and 3:5 became unstable as movement frequency increased, resulting in phase transitions to nearby lower-order ratios such as 2:3 and 1:2. Their account posits that the amplitudes of higher-order frequency-locking terms in the coupling function decrease as movement frequency increases.

Experimental and theoretical work has further investigated these and other ideas, particularly in the domain of rhythmic interlimb coordination. Stochastic influences on coordination were considered by Schöner, Haken, and Kelso (1986). Handedness asymmetry has been shown to cause a slight phase-lead of the dominant limb in bimanual finger wagging (de Poel, Peper, & Beek, 2007), as well as asymmetry in coupling strength (Treffner & Turvey, 1995). The Haken et al. (1985) model predicted that coupling strength should increase with movement amplitude. This has been confirmed through perturbations of rhythmic movements with varying amplitudes (Peper, Boer, de Poel, & Beek, 2008), and effects of movement amplitude have been proposed to arise from neurophysiological time delays (Peper & Beek, 1998). Additionally, researchers have observed stability of anti-phase coupling and low-order frequency-locking in studying rhythmic coordination between nonhomologous limbs (Carson, Goodman, Kelso, & Elliot, 1995) and between people (Schmidt, Carello, & Turvey, 1990).

Speech coordination is related to such phenomena because it involves interacting systems associated with a hierarchy of timescales. What exactly are these timescales? The shortest one encompasses speech gestures, and informative work has been done in the task dynamic framework, in which target relative phases of gestures are derived from coupled systems (Browman & Goldstein, 1990; Nam & Saltzman, 2003; Saltzman, 1986; Saltzman & Byrd, 2000; Saltzman & Kelso, 1983, 1987; Saltzman & Munhall, 1989; Saltzman, Nam, Goldstein, & Byrd, 2006). Dynamical treatments of rhythmic units, such as moras, syllables, feet, and phrases, have also been proposed by several researchers (Barbosa, 2002; Cummins & Port, 1998; O'Dell & Nieminen, 1999; Port, 1986; Port, Cummins, & Gasser, 1995). A review of the application of dynamical systems theory to both gestural and rhythmic patterns can be found in Van Lieshout (2004), and a recent review of task-dynamic approaches can be found in Saltzman et al. (2008).

While previous work in the task-dynamic framework has modeled the dynamics of intergestural, syllable-foot, and foot-phrase levels independently, the current work is unique in two ways. First, it provides an empirical demonstration of linkage between these levels, and

second, it presents a computational model that simultaneously describes the dynamics at each of these levels, as well as interactions between them.

If short-timescale gestural systems and longer-timescale rhythmic systems do not interact with each other, then their dynamics can be understood independently and there should be nothing particularly interesting about the behavior of the system as a whole that cannot be learned from studying its parts. There is no a priori reason for such interactions to occur, although it seems intuitively obvious that some form of interaction must take place. Absence of an interaction should entail, for example, that the relative timing of tongue and lip gestures in an [sp] cluster in the word ''spa'' will be unaffected by the rhythmic context in which these gestures are articulated.

To the contrary, if rhythmic and gestural systems do interact, then intergestural timing should be influenced by the rhythmic context in which gestures are produced. Experimental perturbations of speech rhythm should influence the relative timing of movements associated with nearby gestures. One reason to suspect that rhythmic and gestural systems do interact is that correlations have been observed between patterns of deletion/reduction of speech gestures and the typological rhythmic classes of stress- and syllable-timed languages (Dauer, 1983; Ramus, Nespor, & Mehler, 1999). However, to date there has been no conclusive demonstration that rhythmic and gestural systems interact within a given utterance.

1.1. Rhythmic timing in speech

Multiscale dynamics involving rhythmic systems were discovered by Cummins and Port (1996, 1998), who used a *speech cycling task* (or phrase repetition task) to probe the relative timing between phrases and feet. In this task, subjects hear a high–low two-tone metronome pattern and repeat a phrase (e.g., *big for a duck*) so that the first and last stressed syllables of the phrase (i.e., *big* and *duck*) align with the metronome beats, as depicted in Fig. 4A. After a number of metronome pattern repetitions, the metronome fades out while subjects continue repeating the phrase, trying to maintain the metronome rhythm. This approach can be called a synchronization-continuation paradigm. The control parameter in this design is the target phase (Φ) of the second (L) metronome tone relative to the phrasal period defined by successive first (H) tones. On successive trials, target phases were varied uniformly and randomly in the interval (0.3, 0.7). In Cummins and Port (1998) the H to L tone interval was fixed at 700 ms and the phrasal period was varied accordingly. Fig. 4A shows a schematic representation of the design with target phases (0.3, 0.4, 0.5, 0.6, and 0.7) and presents simulated results for these targets.

Using the task described above, Cummins and Port discovered a *harmonic timing effect* whereby productions of the second stressed syllable were shifted toward phases of the phrasal period that were close to the low-order integer ratios 1/3, 1/2, and 2/3. Observed phases were measured by approximating the locations of p-centers, which are salient points in time believed to correspond to the ''beats'' of syllables (Allen, 1972, 1975; Howell, 1988; Morton, Marcus, & Frankish, 1976; Pompino-Marschall, 1989). In Fig. 4B, the observed relative phase φ is the ratio of the interval between p-centers of *big* and *duck* to the interval between successive p-centers of *big*. Cummins and Port found the distributions of observed

phases across the experiment were shifted toward the nearest low-order harmonic ratios. The same pattern was observed with and without the metronome tones present. If target phases are restricted to the ones shown in Fig. 4A, the distribution of produced phases would look similar to the simulated data in Fig. 4B.

Another important observation was that variance in produced phase was lower for target phases nearer to the low-order harmonic ratios; hence, variability in produced phase is highest for targets phases $\Phi_{0.4}$ and $\Phi_{0.6}$, moderate for $\Phi_{0.3}$ and $\Phi_{0.7}$, and lowest for $\Phi_{0.5}$. This suggests that the $\Phi_{0.5}$ rhythm is easiest to produce, while other target phases require more difficult rhythms.

Cummins and Port (1998) and Port (2003) suggested that the results of the speech cycling task can be modeled with phase-locked pulses from a multifrequency system of coupled oscillators, where harmonic phrase and foot oscillators are phase-locked and either 1:2 or 1:3 frequency-locked. Fig. 4A shows the oscillations of the foot oscillators corresponding to each target rhythm. The foot oscillator produces pulses at phase 0, and stressed syllable beats are attracted to the pulses. The motivation for this approach derives from the finding that low-order harmonic frequency ratios are more stable than higher-order ones (Haken et al., 1996). Correlations between target phase and variability may arise due to the relatively lesser stability of 1:3 frequency-locking compared to 1:2 locking.

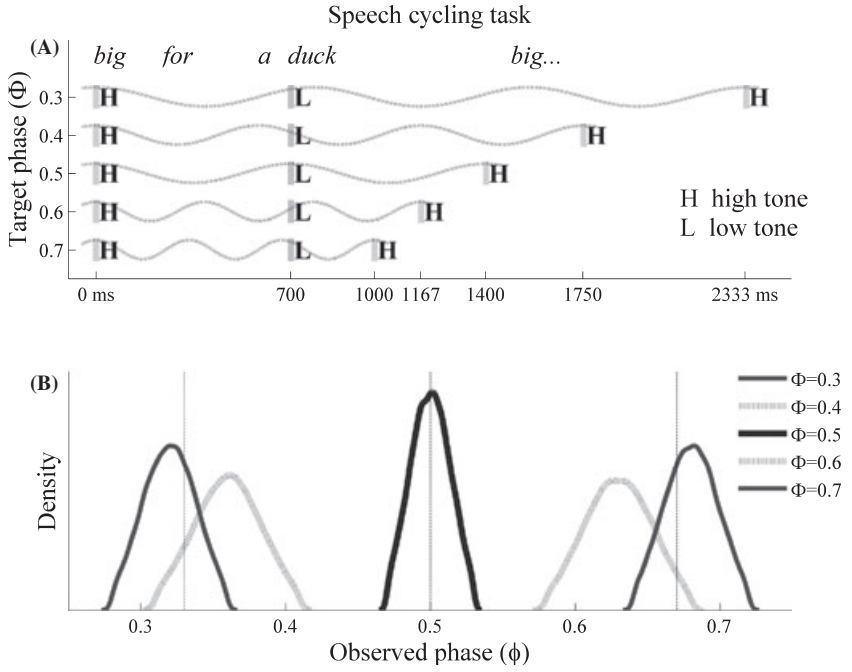

Fig. 4. Schematic illustrations of speech cycling task design and harmonic timing effect. (A) Intertone interval durations in each of the five target phase conditions, along with hypothesized foot system oscillations. (B) Relative phase distributions for each target phase condition (Φ). Shifts in observed phases toward harmonic ratios (vertical lines) are evident in the distributions, as well as increased variability in the higher-order target phase conditions. The same pattern is observed with or without metronome tones present. See text for further details.

The results of the speech-cycling task are interesting because they support the notion that there exist multitimescale dynamical interactions between units of the prosodic hierarchy, in this case feet and phrases. The Cummins and Port analysis reconceptualizes these hierarchically related units as frequency-locked coupled dynamical systems that are responsible for rhythmic coordination of speech. This constitutes a fairly radical departure from the mainstream linguistic understandings of metrical feet and their combinations that were developed in Liberman (1975) and Liberman and Prince (1977). In these approaches rhythmic units are organized either in a hierarchical branching structure or in a grid in which higher degrees of prominence are associated with higher-level units. As will be explained below, neither of these representational schemes makes any quantitative predictions about durations between stressed syllables, and so neither is well suited for understanding harmonic timing and rhythmic variability.

1.2. Intergestural timing

Focusing now on the smaller timescale appropriate for intergestural coordination, consider an important phenomenon known as the *c-center effect*, discovered by Browman and Goldstein (1988). This effect refers to timing patterns observed between multiple-onset consonant gestures and vowel gestures within a syllable, for example, in CCV syllables like *spa*. In syllables with a single-onset C, the gestures associated with the consonant and the vowel begin at approximately the same time (although the consonantal gestures are produced more quickly). However, in syllables with complex CC onsets, the beginnings of the consonant gestures are equally displaced in opposite directions from the start of the vocalic gestures, as schematized in Fig. 5C.

Browman and Goldstein (2000) argued that the c-center effect arises from competing lexical specifications of target relative phases between pairs of gestures. All consonant gestures in a syllable are specified to be in-phase coordinated with the initial vowel gestures of a syllable, as represented in Fig. 5A. In other words, the gestures belonging to each consonant in C_1C_2V have the same target relative phase specification as the gestures associated with the single consonant in CV. However, in C_1C_2V there is an additional anti-phase target between the consonants, as represented in Fig. 5B. Both of these relative phase targets cannot be simultaneously achieved. Instead they compete, the result being a compromise similar to the temporal configuration in Fig. 5C. In this case, the moment that ends up being in-phase synchronized with the beginning of the vowel gesture is the c-center, which refers to the point halfway between the beginnings of the C gestures.

Nam and Saltzman (2003) simulated the competitive coupling proposed by Browman and Goldstein (2000) using a system of coupled oscillators. Rather than treating the relative phases of gestural onsets as arbitrarily specified lexical information (as in Saltzman & Munhall, 1989), the newer model incorporates the idea that target relative phases are derived from the stable relative phases of a system of coupled oscillators. These oscillators can be conceptualized as *gestural planning systems*, whose properties are developed partly in Saltzman and Byrd (2000).

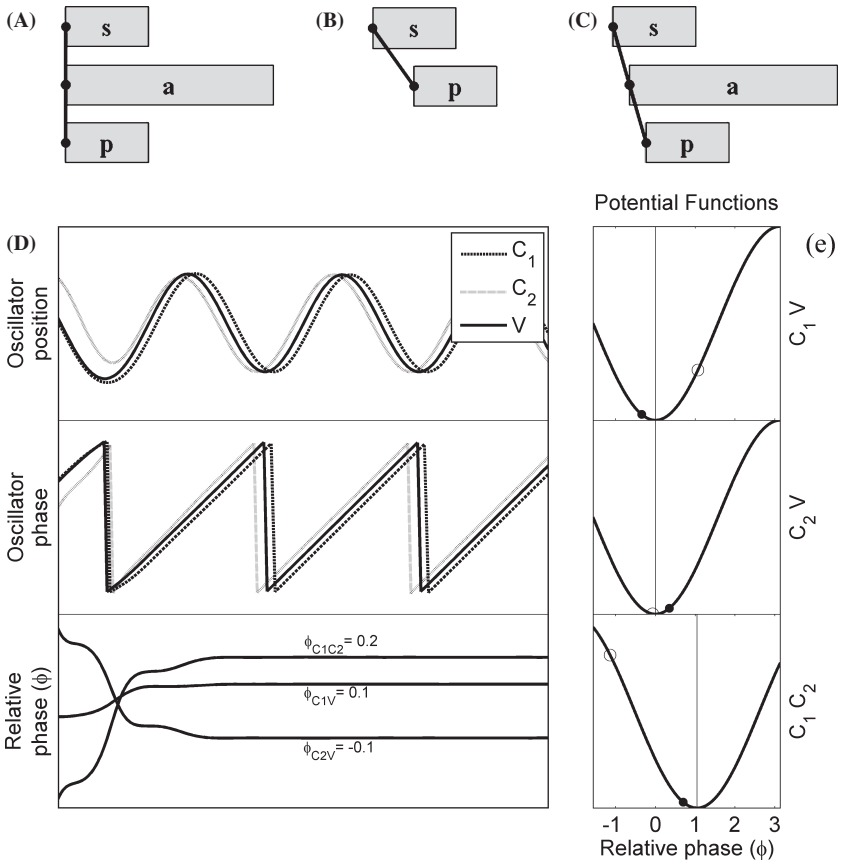

Fig. 5. Coupling graphs and simulation of c-center effect through competitive coupling. Potential functions depict relative phase targets with vertical lines, initial relative phases (randomly chosen) with open circles, and final relative phases with filled dots. Evolutions of oscillator positions, phases, and relative phases from random initial conditions are shown on the left. Note that initial transients die off quickly, and that once the system has stabilized, C_1 and C_2 are equally displaced from V in opposite directions.

There is an important distinction here between gestural systems and gestural planning systems. In the task dynamic framework, gestures are modeled as critically damped mass-spring systems that have fixed-point attractors. With parameter modifications, gestures can vary in speed, amplitude, and duration. The gestural planning oscillators are very different. They exhibit limit cycle dynamics and their relative phases are governed by potential functions. The planning oscillators bear an indirect relation to observable articulator movements. In the Nam and Saltzman (2003) model, the relative phases of the damped mass-spring gestural systems are determined by the stabilized relative phases of the planning oscillators.

Fig. 5D shows a simulation of the c-center effect in planning oscillators based upon the Nam and Saltzman model. For each pair of coupled oscillators (and there are three such pairs in a C_1C_2V system), the figure shows the potential function governing the evolution of the relative phase between the oscillators. Regardless of the initial relative phase, the system

settles into a stable configuration that corresponds to a balance between the competing forces associated with the potential functions. Following Nam and Saltzman (2003), the target relative phase of the consonants (Φ_{C1C2}) was approximately 1 radian (60°), rather than true anti-phasing (π radians or 180°). The minimization of potential energy is accomplished by equal displacement of C gestural phases from the phase of the V gesture, hence producing a c-center effect.

1.3. Coupling between gestural and rhythmic systems

Because dynamical coupling can usefully model empirically observed temporal patterns involving feet and phrases, as well as temporal patterns involving gestures, an obvious question to ask is whether evidence can be found for a more general model which accounts for patterns on both timescales. A synthesis of the two models requires bridging the gap between gestural timing and rhythmic timing. An obvious way to connect these disparate timescales is through generalized relative phase coupling. In addition to phrase-foot coupling and intergestural coupling, such a system presumably includes interactions between feet and syllables and/or moras.

A coupled-oscillators model of the relative timing of feet and syllables is described in O'Dell and Nieminen (1999). The time span of the metrical foot is generally reinterpreted in dynamical treatments as the *stress-foot*, which is an interval between stressed syllables. O'Dell and Nieminen model a foot with *n* syllables as a system composed of a stress and syllable oscillator exhibiting 1:*n* frequency-locking. A coupling strength parameter describes the relative strengths of the coupling forces between stress and syllable oscillators. Their model can predict durational attributes that correlate with the much-studied distinction between stress and syllable-timed languages (Abercrombie, 1967; Pike, 1945). In languages in which the syllable oscillator dominates, the interstress interval is temporally stretched away from its intrinsic period toward *n* times the intrinsic period of the syllable oscillator. In languages in which the stress oscillator dominates, however, it temporally compresses the period of the syllable oscillator when the number of syllables is greater than the ratio of the intrinsic foot period to syllable period, which results in an interstress duration attracted toward the intrinsic foot period. The predicted foot durations from this model accord with the analysis of Eriksson (1991), who argued that the distinction between syllable-timing and stress-timing results from additional duration endowed to stressed syllables in stress-timed languages.

The generality of coupling in dynamical models of speech gesture and rhythm begs for the development of a model integrating both domains. One would expect this model to cover the gamut of relevant timescales, from phrases (groups of feet), to feet (intervals between stressed syllables), to syllables, moras, and on the lowest level, gestures and effectors. Fig. 6 illustrates how prosodic units in the hierarchical branching model of prosodic and segmental structure can be related to oscillatory systems associated with a range of timescales.

Of course, not all languages exhibit evidence for systematic patterning on all of these scales; for example, languages that lack closed syllables and long vowels generally exhibit no phonological evidence for a moraic level of prosodic structure distinct from the syllabic level (e.g., Hua, Cayuvava; cf. Blevins, 1995), and many languages have no stressed sylla-

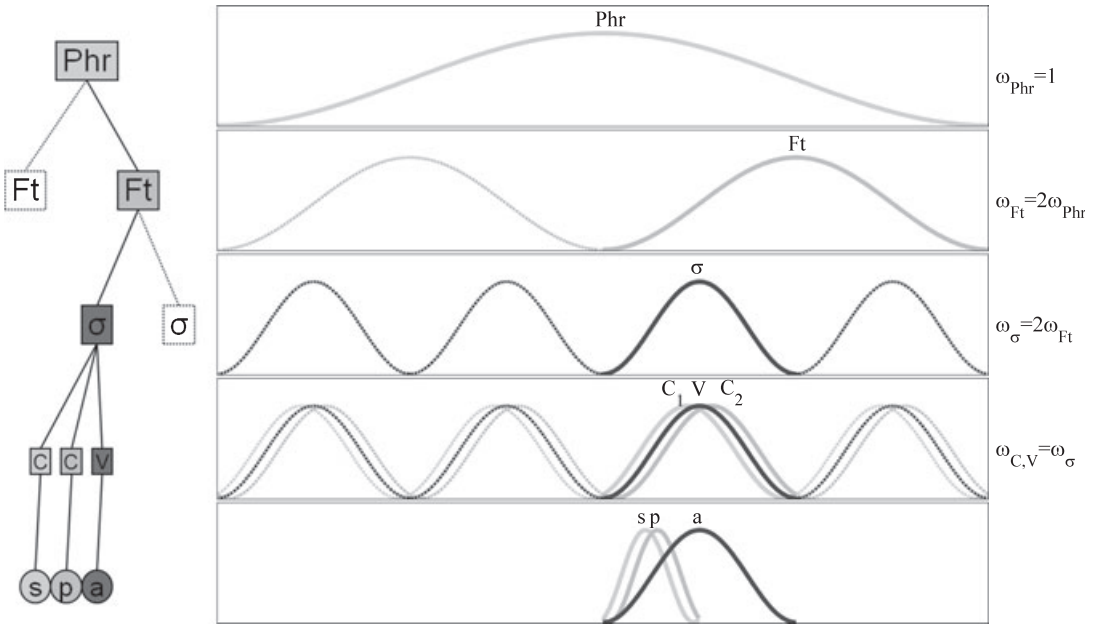

Fig. 6. Relations between the hierarchical model of prosodic and gestural units and a multiscale dynamical model. In the top four rows, prosodic and segmental units are conceptualized as oscillatory planning systems. Prosodic and segmental levels in the hierarchy are conceptualized as timescales, which correspond to the inherent frequencies of their associated planning oscillators, and which are related by integer ratios. In the bottom row, gestural constriction and release kinematics are schematized.

bles (e.g., Cantonese, Yoruba; cf. Hyman, 2006). Any satisfactory model should have parameters available to accommodate cross-linguistic variation, although such variation will not be our concern here. Further, the values of some model parameters may vary from speaker to speaker and from utterance to utterance. A long-term goal of any modeling enterprise should be the identification of intrinsic constraints on model parameters by thorough investigation of cross-linguistic, inter-speaker, and intra-speaker variation.

The multitimescale dynamical model reconceptualizes commonly accepted prosodic units such as syllables, feet, and phonological phrases (Nespor & Vogel, 1986) as planning oscillators, which can be described with differential equations. The ''levels'' (or types of units) in the traditional model become timescales, which correspond to the inherent frequencies of their associated oscillators. These frequencies are related by low-order ratios of integers, and connections between units are represented by coupling terms that can be specified by potential functions. Hence, the two most fundamental metaphors structuring the prosodic hierarchy are preserved: CONTAINMENT (of feet within phrases, syllables within feet, etc.) is implicit in the frequency relations between timescales, and CONNECTION (between systems) is a coupling interaction between planning oscillators. Note that gestural systems (as opposed to gestural planning systems) are not oscillatory; rather, these systems exhibit critically damped mass-spring dynamics, and the relative phases of their constriction formation and release movements are determined by the gestural planning systems.

There are many possibilities for exactly what form the equations describing a multiscale model can take, and in the end these should be resolved empirically. The issue of how to structure the coupling relations between planning oscillators will be considered in Section 4. Crucially, regardless of how the multiscale dynamical model is formulated, the model predicts the possibility of interactions between disparate timescales. For example, consider repeated productions of the phrase *take on a spa*, in which the words *take* and *spa* are stressed syllables. In addition to rhythmic temporal patterns between the phrase and feet (inter-stress intervals), there are intergestural patterns that involve the tongue movement associated with [s] and the lower lip movement associated with [p]. The multiscale dynamical model allows for the possibility of interactions between gestural timing and lower frequency foot and phrase timing. The reasoning behind this is as follows.

Increased variability found by Cummins and Port (1998) in the higher-order target phase conditions suggests that those rhythms are more difficult to produce. A likely explanation for this difficulty is that foot-phrase coupling and syllable-foot coupling are weaker when governed by oscillatory systems with higher order frequency-locking (cf. Haken et al., 1996). Now, assume that there are random noise-induced perturbations of planning oscillator phases at all timescales. It follows that if gestural systems are coupled to rhythmic systems, then intergestural timing should be less variable when rhythmic coupling is stronger, because more strongly coupled rhythmic systems will exert a more coherent, less variable stabilizing force upon the gestural systems, thereby counteracting noise-induced perturbations of intergestural relative phase. In the case of *take on a spa*, the relative timing of kinematic landmarks associated respectively with the tongue and lip gestures in [s] and [p] should be least variable when the phrase is spoken to a metronome target of $\Phi_{0.5}$ and should be more variable when the metronome target is $\Phi_{0.3}$, $\Phi_{0.6}$, etc.

Conventional, nondynamical models of speech production do not predict differences in temporal patterns due to interaction between rhythmic units and gestures. For example, the metrical grid proposed by Liberman (1975), which associates each syllable to a slot (container) in a grid (sequence of containers) and assigns degrees of prominence (height, a.k.a. accent) to each container, does not make even qualitative predictions about the timing of gestures within syllables. Likewise, ''metrical tree'' models (e.g., Halle & Vergnaud, 1978; Liberman & Prince, 1977), which utilize connections between feet, syllables, and segments, do not generate specific temporal predictions. Nor can such models be readily amended to account for rhythmic–gestural interaction: Without dynamic interaction, the intergestural timing in an [sp] onset cluster should exhibit the same amount of variability regardless of the metrical context in which it is produced.

The general problem with nondynamical approaches is that they provide no mechanism for predicting the temporal patterns of articulations. They treat gestures as sequential sets of actions that are *linked* or *connected* to higher-level and lower-level units (as in Fig. 6). Importantly, these connections are nondynamic: They do not incorporate coupling functions that influence relative phase—indeed, they *cannot* do so, because nondynamic models do not assign phase variables to the individual units in the representation. Nondynamical connections are purely representational and relational. In contrast, dynamical coupling can be treated analytically or numerically using standard mathematical techniques. The difference

between the two approaches boils down to whether the models make empirically testable predictions about temporal patterns on 10 and 100 millisecond scales of time.

The nondynamical models also do not account for previously observed rhythmic patterns. For one, the aforementioned harmonic timing effect does not follow intuitively from grid or tree representations. Further, nondynamical models must resort to rules or abstract constraints to account for phenomena such as the rhythm rule in English (Hayes, 1995). This pattern has been described as the avoidance of adjacent prominent syllables that is observed in primary stress–secondary stress alternations such as antíque armóir versus ántique sófa and háll fourtéen versus fóurteen hálls. Although widely used, rule, and constraint-based formulations of such patterns do not give insight into the deeper cognitive mechanisms from which the patterns arise.

To test for evidence of rhythmic–gestural interaction, a speech cycling task (as in Cummins & Port, 1998) was conducted using the phrase *take on a spa*, and movements of the jaw, lower lip, and tongue blade were recorded using electromagnetic articulometry. The experiment was designed to address the following hypothesis:

> *Hyp. Rhythmic-gestural covariability:* In higher-order target phase conditions of a phrase repetition task, rhythmic timing of take on a spa will be more variable, and intergestural timing between [s] and [p] will also be more variable.

2. Method

2.1. Task and participants

Five target rhythms were used (cf. Fig. 4), that is, target relative phases $\Phi = (0.3, 0.4, 0.5, 0.6, 0.7)$, in a synchronization-continuation phrase repetition task. Eight native speakers of English of ages 18–30 participated in the experiment. At the beginning of their first session, subjects were given instructions and practiced the $\Phi_{0.5}$ rhythmic condition. Sessions were organized in blocks of five trials; subjects performed the task with a particular target rhythm for five consecutive trials, after which they switched to a different target rhythm. Target phases from block to block always differed by more than 0.1, in order to reduce potential carryover effects from one block to another. Block orders were randomly assigned to subjects. After performing five blocks, subjects rested for several minutes, and then performed another five blocks in the same order. Thus, each subject produced 10 trials per target condition in each session.

Subjects were instructed to wait until the third repetition of the metronome pattern to begin on each trial and to produce the phrase so that the word *take* coincides with the first (H) tone and the word *spa* with the second (L) tone. The metronome tones were 50 ms long and offset by 700 ms in all target phase conditions, while the interval between H tones was varied to produce the aforementioned target phases. Assuming that the produced phrase is entrained to the metronome beats during the synchronization phase, the constant 700 ms interval between tones ensures that there are no tempo changes in the production of the

phrase. The H and L tones were at 1200 and 600 Hz and were windowed with a Tukey window (r = .5). The metronome pattern repeated 12 times from the start of each trial, and the last three pairs of tones were faded out by successive 50% decreases in amplitude. Subjects were instructed to continue repeating the phrase with the same rhythm after the metronome pattern faded out. This latter portion of the task constitutes the continuation phase. After a duration of time corresponding to 14 cycles of the metronome pattern, subjects were signaled to stop. They were told to speak with a normal volume and pitch, and not to tap their feet or hands, or to imagine doing so. They were instructed not to take shallow breaths between each repetition of the phrase, but rather, to take a deep breath when necessary and skip one cycle of the phrase. When taking a breath without the metronome, they were told to wait for the duration of approximately one cycle.

To reduce the influence of preceding and subsequent context on the gestural trajectories of [s] and [p], a number of measures were taken in the design of the carrier phrase, *take on a spa*. It was deemed important to employ a four-syllable phrase, because a three-syllable phrase is produced abnormally slowly to the 700 ms intertone interval employed by Cummins and Port (1998). A guiding principle behind this design was that the tongue and lip gestures entering and exiting [sp] should be of large magnitude, in order to make landmarks associated with their articulatory trajectories easily detectable. The unstressed vowel [ə] preceding [sp] encouraged subjects to pass through a relatively neutral vocal tract configuration immediately prior to the articulations of interest. A phrase with the fewest potential coarticulatory confounds would require no tongue raising or fronting, and no large magnitude lip closure, protrusion, or retraction gestures in the two stressless syllables preceding [sp]. However, this ideal design is not achievable with the set of English function words. The function word *on* was chosen because pilot investigations revealed that both the vowel and alveolar nasal in *on* [an] tended to be coarticulated with the preceding velar stop in *take* [teik]. Moreover, rather than a distinct tongue gesture for [n] occurring, a lengthened nasalized back vowel [ã:] was normally produced, resulting in the phonetic sequence [teikã:əspa]. The vowel following [sp] was chosen to be a low, central/back vowel [a] in order to maximize the speed and magnitude of the release gestures associated with [s] and [p]. No coda consonant followed the vowel, in order to further maximize movement amplitude and avoid confounds from anticipatory coarticulation.

2.2. Data collection

Kinematic data were collected at 200 Hz, using a Carstens Articulograph AG200 (an electromagnetic articulometer, EMA; cf. Hoole, 1996). Four transducer coils were used, all in the midsagittal plane, in the following locations: (a) on the forehead for reference, (b) on the lowermost projection of the jaw when the subject looked straight ahead, (c) on the outermost projection of the lower lip, and (d) on the blade of the tongue, approximately 1–1.5 cm from the tongue tip. In EMA studies, a bite plate is commonly used to discern the angle of the occlusal plane relative to the transmitter coils. However, unlike other EMA experiments, subjects were required to wear the (somewhat uncomfortable) transmitter helmet for a rather extended period of time (approximately an hour to an hour and a half during each session).

This necessitated occasional readjustment of the helmet. The weight of the helmet is balanced by a line above the head of the subject; this presents the dilemma that the more weight is taken off the helmet, the more susceptible to movement the helmet becomes, especially when subjects do not sit perfectly still. In addition, there is some variation in the helmet angle that subjects find most comfortable. Consequently, to estimate the occlusal plane throughout the experiment, one would have to recalibrate with a bite plate every time an adjustment is made, which is impractical.

However, the experimental hypothesis involves only within-subject comparisons of relative timing of gestures, rather than between-subject comparisons of articulator motions. For relative timing measurements, variation in the location of the occlusal plane is not problematic, and so no bite plate was used. Further, because gestures are understood as synergistic organizations of effector movements, measurements are derived from lip–jaw and tongue–jaw synergies. There is thus no need to decouple tongue and lip movement from jaw movement, which is more accurately accomplished when two jaw sensors have been used (Westbury, Lindstrom, & McClean, 2002). Strictly speaking, the bilabial gesture associated with [p] is a synergy of jaw, lower lip, and upper lip movement, the last of which was not recorded. Since upper lip movement is normally of relatively lower magnitude than lower lip movement, changes in lower lip position are fairly representative of the timing of the bilabial gesture. One issue that may arise in a study which attempts to investigate the relative contributions of individual articulators to synergies in a phrase repetition task is that these relative contributions are speaker-specific, task-specific, and vary within speaker and task across time (Alfonso & Van Lieshout, 1997).

It should be noted that the sensor coil on the tongue blade undoubtedly influences the observed articulatory patterns to some extent. Subjects may have produced abnormal alveolar closures to adjust for the presence of the sensor. It is unlikely, however, that this had a profound effect upon temporal relations between [s] and [p], or that this adjustment differed between rhythmic conditions.

Audio was recorded at 22,050 Hz using a microphone clipped onto the shirt of the subject. A secondary audio signal was collected for synchronization with the EMA using a table microphone. As the metronome tone would not be recorded along with the speech signal, subjects wore earbud headphones. Inspection of kinematic data with and without the earbuds revealed that the presence of the earbuds in the magnetic field generated by the EMA helmet produced no noticeable increase in noise or signal distortion.

2.3. Data analysis

To reduce signal noise, kinematic data were smoothed with an unweighted, 55 ms window moving-average filter. For each trial, the reference signal was taken as the origin, and the data were rotated in the X–Y plane so that the mean horizontal location of the jaw was on the Y-axis. This rotation, which was generally less than 20°, compensates for variation in helmet orientation and facilitates visual comparison of data across subjects, without significantly affecting relative timing measurements.

All horizontal and vertical minima (x, y), maxima (X, Y), velocity minima (dx, dy), and velocity maxima (dX, dY) in the neighborhoods of [sp] from *spa* and [t] from *take* were identified algorithmically in Matlab by looking for zero-crossings in difference vectors. In general, the horizontal motion (fronting) of the tongue from [ã:ð] to [s] was a higher amplitude and faster movement than the vertical motion (raising) of the tongue. For this reason, the maximum horizontal position of the tongue blade (TX) and the maximum horizontal velocity of the tongue blade (TdX) were deemed the most appropriate landmarks corresponding to the [s] gesture. For the lower lip, horizontal (LX) positional maxima were found to be appropriately representative landmarks of the [p] gesture for all but one subject, *s6*, for whom maximum lower lip vertical position (LY) was used. The appropriateness of these landmarks can be seen by qualitative inspection of 2D articulator trajectories. Fig. 7 (middle) shows that a rapid and relatively high amplitude fronting of the tongue (TdX) occurs just prior to the second metronome tone (M2).

Interval durations between the [s] and [p] gestural landmarks described above (i.e., TdX, TX, and LX) serve as the dependent variables in the intergestural timing analysis. The analysis is concerned with the variance of these durations, and compares within-subject variances in the higher-order Φ conditions to the variance in $\Phi_{0.5}$. These comparisons utilize ratios of variances, which are equivalent to F tests for significantly different variance. Intergestural interval durations are suitable measures for variability analysis as long as there are

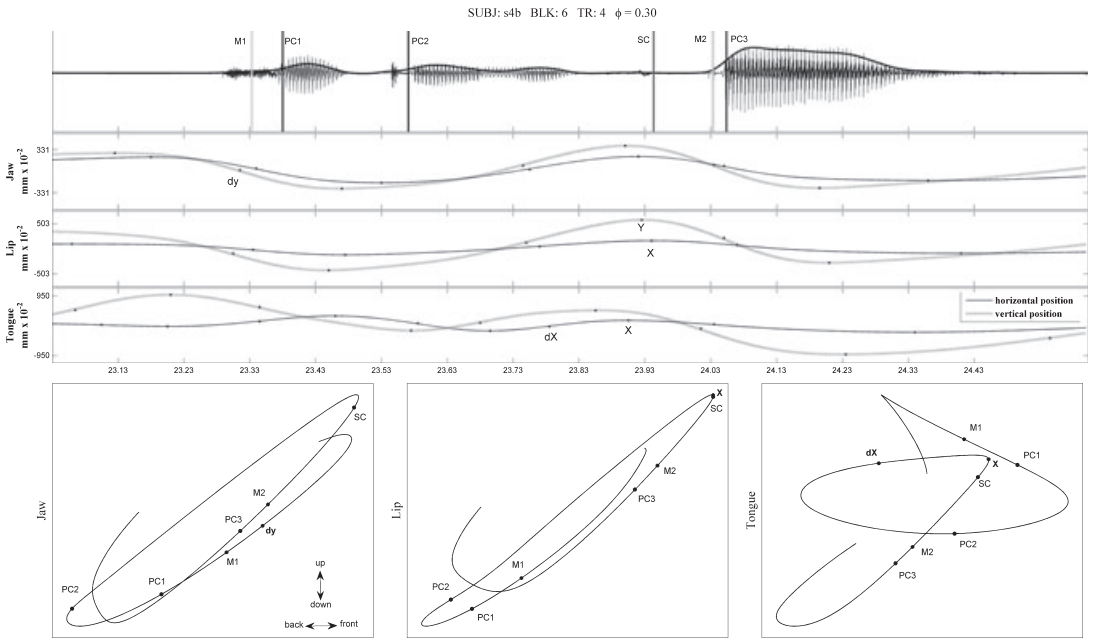

Fig. 7. Example of acoustic and kinematic data. (Top) Acoustic waveform, along with locations of metronome tones (M1, M2), acoustically estimated p-centers (PC1, PC2, PC3), and sibilance center (SC). (Middle) Time-aligned jaw, lower lip, and tongue blade horizontal and vertical positions. (Bottom) 2D trajectories of the jaw, lower lip, and tongue blade. See text for further details.

no substantial differences in the rate of articulation of [sp] clusters across the experimental conditions. With a couple of potential exceptions in $\Phi_{0.7}$ (cf. Section 3.2), this assumption held true. Measures of movement amplitude and absolute durations between gestural landmarks explain several anomalous patterns associated with the $\Phi_{0.7}$ condition. Amplitude was defined as the displacement in position from the point in time corresponding to TdX to the point in time corresponding to TX. This is a departure from the way that amplitude is usually defined—displacement from minimum to maximum—but is appropriate because measurement of the position of the preceding minimum is confounded by coarticulation with preceding gestures. Hence, the measure used is believed to be more representative of amplitude variation in the gestures of interest, even though it does not reflect the full range of movement.

Estimated p-centers of the syllables *take* and *spa* served as the bases for rhythmic timing analyses. The p-center is an approximate location of the ''beat'' of a syllable. In early work, p-centers were equivalent to the temporal location of a finger tap made in time with a stressed syllable (cf. Allen, 1972, 1975). More recently the p-center has become a ''perceptual-center,'' and its location in a single syllable can be determined using a dynamic rhythm-setting task in which the timing of a syllable relative to a repeating reference frame is adjusted with a knob (Morton et al., 1976). Numerous algorithms have been developed to estimate the locations of syllable p-centers (Howell, 1988; de Jong, 1994; Patel, Löfqvist, & Naito, 1999; Pompino-Marschall, 1989; Scott, 1993, 1998); however, there currently exists no consensus on how to predict p-center locations, nor on whether they correspond to gestural and/or acoustic events. While Cummins and Port (1998), following Scott (1993), used an estimation algorithm that treats the p-center as a rapid energy-rise in the acoustic signal, others have argued that the p-center corresponds to a gestural event (Fowler, 1979; Tuller & Fowler, 1980), such as a maximum in jaw opening speed, or even a composite of gestural and acoustic events (de Jong, 1994). Further, there are reasons to suspect that the locations of and information determining the perceived and produced beats of syllables may differ (Treffner & Peter, 2002).

In the purely acoustic algorithm used by Cummins and Port (1998), the speech signal is filtered with a passband of [700–1300] Hz, using a first-order Butterworth filter, which has gradual roll-offs in its response function. The effect of this filter is to greatly diminish spectral energy corresponding to F0 and higher-frequency fricative energy in the signal. The magnitude (absolute value) of the bandpass-filtered signal was then lowpass filtered using a fourth-order Butterworth filter with a 10 Hz cutoff. The result is a smoothly varying representation of mostly vocalic energy in the signal. P-centers are estimated to be the midpoints in time between the points when the signal amplitude is 10% above its local minimum and 10% below its local maximum. These estimates yield points that are near vocalic energy velocity maxima.

In the absence of a well-accepted algorithm that is robust to interspeaker and intersyllable variation, it was judged best to examine whether p-centers estimated from kinematic landmarks or from acoustic information were more closely timed to metronome tones. It was apparent from visual inspection and verified quantitatively that points of minimum vertical jaw velocity (Jdy), that is, when the jaw was opening most quickly, were more closely and less variably timed with the first metronome tone (M1) than acoustically estimated p-centers.

This can be seen in Fig. 7, where PC1 (acoustically estimated) lags behind the metronome beat by approximately 50 ms. In contrast, the acoustically estimated PC3 was generally more reliably timed with the second metronome tone (M2) than any kinematic landmark. Hence, the rhythmic measure used here will take Jdy as an estimate of PC1 and use the acoustic algorithm described above to locate PC3. The rhythmic φ variable for phrase n is

$$[\mathrm{PC3}_{\mathrm{n}}^{(\mathrm{acous})} - \mathrm{PC1}_{\mathrm{n}}^{(\mathrm{Jdy})}]/[\mathrm{PC1}_{\mathrm{n+1}}^{(\mathrm{Jdy})} - \mathrm{PC1}_{\mathrm{n}}^{(\mathrm{Jdy})}].$$

Using a purely acoustic definition of this variable does not result in major changes to variability patterns, but it does result in a shift of the observed phase distributions so that all distribution modes are earlier and less clearly reflect harmonic timing patterns. To illustrate φ_{PC3} distributions, Section 3.1 (cf. Fig. 8) presents histograms that were smoothed using a Gaussian kernel with a 0.015φ bandwidth over 60 bins. Also automatically detected in each phrase was the *s-center*, that is, the sibilance center, which is the point in time corresponding to maximum sibilance energy associated with the [s] in *spa*. In this case, a passband of [5000–7500] Hz was used.

To identify locations where subjects took a breath, the acoustic data from every trial were visually and auditorily inspected; in the course of this process, spurious and missing p-centers were corrected. The first trial of each session and the last phrase of each trial and inter-breath group (for which no φ is defined) were excluded from the analysis.

Each session yielded approximately 85–110 repetitions of the phrase in each target phase and metronome condition. Subject *s1* performed three sessions; all other subjects performed two sessions. One session from *s7* and one from *s8* were discarded due to equipment malfunction. In the remaining session performed by *s8*, kinematic landmarks failed to exhibit the same sort of consistency as those produced by the other subjects, making meaningful temporal analyses impossible—hence, data from *s8* were not analyzed. One session from *s5* was not analyzed because the subject suffered from fatigue and exhibited a lack of attention to the task. Several subjects attempted two or three trials in $\Phi_{0.7}$ before they were capable of performing a full trial without halting—these trials were excluded.

The dependent variable of rhythmic-timing to be analyzed is observed phase (φ), defined in the manner described above as the timing of the p-center of *spa* (PC3) relative to the preceding and following p-centers of *take* (PC1). The dependent variables of intergestural timing are LX–TdX and LX–TX, corresponding to intervals between lower lip horizontal/vertical maxima and tongue horizontal positional and velocity maxima. Section 3 presents intergestural results from only phrases produced without the metronome, although intergestural patterns were not substantially different with the metronome. Temporal patterns produced without the metronome are more theoretically appropriate tests of the covariability hypothesis, because movement patterns have been found to stabilize locally and globally with external perceptual stimuli present (Fink, Foo, Jirsa, & Kelso, 2000).

Approximately 2.0% of 12,645 eligible phrases were excluded because φ fell outside of the range 0.15 to 0.85 or was more than 2.5 *SD* from the mean phase for a given subject and target phase. Approximately 2.5% of phrases were excluded because one or more of the intergestural measures exceeded 3.0 *SD* from the mean value. Some of these outliers can be

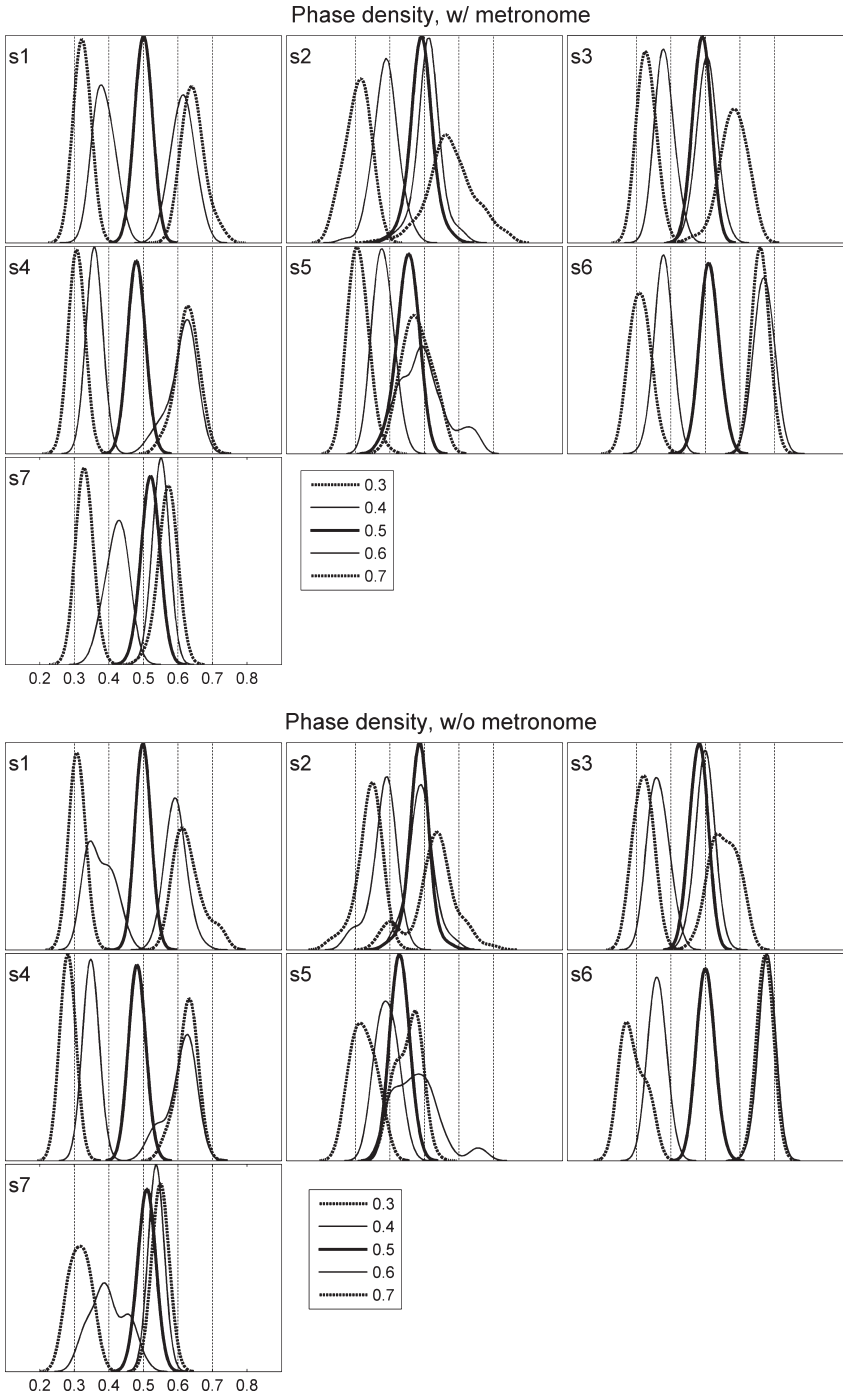

Fig. 8. Observed phase densities (φ) for each subject with (top) and without (bottom) the metronome showing harmonic timing effects and a tendency for lower variability in $\Phi_{0.5}$ compared with the other target Φ.

attributed to circumstances in which a subject misarticulated or hesitated abnormally. Overall, more than 95% of eligible phrases were retained in the analysis.

3. Results

3.1. Rhythmic timing variability

Cummins and Port's (1998) finding that less harmonic (higher-order) target rhythms are associated with increased rhythmic variability was generally replicated in the $\Phi_{0.3}$, $\Phi_{0.4}$, $\Phi_{0.6}$, and $\Phi_{0.7}$ conditions compared to the $\Phi_{0.5}$ condition. This was the case in both the synchronization phase (with the metronome) and in the continuation phase (without the metronome), albeit the effect was more consistent across subjects without the metronome. Harmonic timing effects were observed as well. Fig. 8 shows rhythmic φ smoothed histograms for each subject with the metronome (top) and without the metronome (bottom).

Examining target phase conditions $\Phi_{0.3}$ and $\Phi_{0.4}$, we can see that subjects *s1–s6* exhibited harmonic timing patterns in which one or both distributions were shifted toward 1/3. The distribution produced by *s7* is shifted toward 1/2 with the metronome and is bimodal without the metronome, exhibiting shifts toward 1/2 and 1/3.

In target conditions $\Phi_{0.6}$ and $\Phi_{0.7}$, three of the subjects exhibited shifts toward 2/3, while four subjects exhibited shifts toward 1/2. This intersubject variation is consistent with the observation of Cummins and Port (1998): individuals in their experiment differed with regard to whether they exhibited evidence of the 1/3 and/or 2/3 phase attractors. Such intersubject variation is also consistent with findings in the domain of polyrhythmic drumming (Haken et al., 1996). Additionally, subject *s5* exhibited a large deviation of the $\Phi_{0.5}$ distribution away from the expected 0.5; several other subjects exhibited similar deviations of smaller magnitude, and subject *s4* tended toward a lower phase in the $\Phi_{0.3}$ condition without the metronome. These deviations can most likely be attributed to inaccuracies in p-center estimation (cf. Section 2.3) resulting from intersubject variation in the gestural/acoustic events associated with syllable beats.

Without the metronome, φ variability was generally lowest in $\Phi_{0.5}$. This accords with previous findings that the lowest-order harmonic mode of frequency-locking is the most stable, and it is important for subsequent analyses in this study. With the metronome present, this variability pattern was less consistently observed. Fig. 9 compares phase variances within subjects across each Φ to the variance observed in $\Phi_{0.5}$, in phrases produced with the metronome (top) and phrases without the metronome (bottom). Ratios of variances, which are plotted on the vertical axis, are equivalent to F statistics used to test for significantly different variance. Approximate confidence intervals based upon a sample size of 85 phrases are shown.

With the metronome present, rhythmic variability patterns in $\Phi_{0.3}$ and $\Phi_{0.4}$ were not consistent across subjects: Some exhibited the expected greater variability, others were significantly less variable than in $\Phi_{0.5}$. A greater number of subjects produced the expected

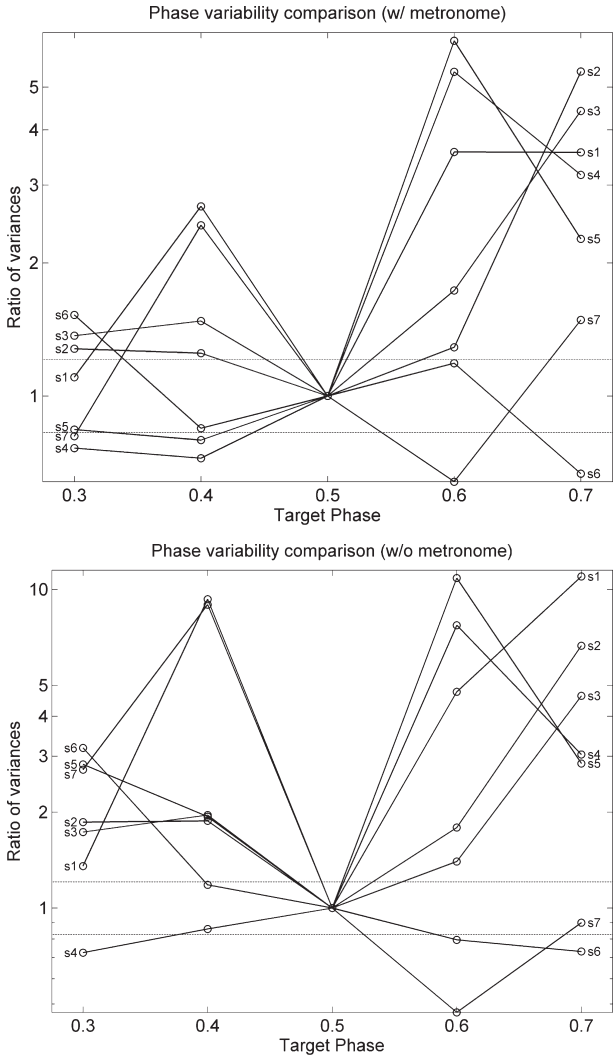

Fig. 9. Comparison of rhythmic timing variability between target phase conditions with (top) and without (bottom) the metronome. Ratios of variances in each Φ condition to $\Phi_{0.5}$ are plotted (on a logarithmic axis) for each subject. Horizontal lines show approximate confidence intervals for significantly different variance.

variability patterns with the metronome in $\Phi_{0.6}$ and $\Phi_{0.7}$, where there were only two instances of lesser variance.

Without the metronome, the majority of subjects were significantly more variable in the higher Φ compared to $\Phi_{0.5}$. This finding supports the hypothesis that rhythms corresponding to the higher-order target phases are more difficult to produce reliably, and it lends credence to the Cummins and Port (1998) and Port (2003) interpretation that such patterns result due to increased competition between relative phase potentials associated with 1:2 and 1:3 frequency-locked phrase and foot oscillators.

The cases of unexpected lower variance in higher order Φ may arise from an interaction between block order and changes in subject attention over the course of each session. Another possible explanation could be a difference between the method employed here and the method in Cummins and Port (1998): Here subjects produced only five different target phase patterns, and became well practiced in those specific patterns. In contrast, Cummins and Port (1998) varied the target phase in the interval from 0.3 to 0.7, using smaller step sizes and changing targets more often. It is possible that the smaller set of targets and greater extent of practice allowed some subjects to develop more consistent rhythmic performances in some of the higher order Φ.

3.2. Intergestural timing variability

As hypothesized, intergestural timing between [s] and [p] gestures was more variable in the higher-order Φ. The gestural landmarks of interest here are the maximum horizontal position and velocity of the tongue (TX and TdX), and maximum protrusion of the lip (LX), which are associated with [s] and [p], respectively (cf. Section 2.2). Fig. 10 compares variances in each Φ condition to the variance in $\Phi_{0.5}$ for each subject for the measures [LX–TdX] (top) and [LX–TX] (bottom).

Fig. 10 presents kinematic measures from phrases produced without the metronome. The without-metronome condition is considered more appropriate for analysis for several reasons. First, the expected rhythmic variability patterns were more reliably observed without the metronome (cf. Section 3.1). Second, the auditory stimulus adds an additional source of temporal regulation that is not normally present in speech. As observed by Fink et al. (2000), the metronome stimulus provides a local ''anchoring'' during which movements become less variable, and a doubly anchored movement pattern (i.e., two metronome tones) can increase the stability of the entire pattern. However, analyses conducted on phrases produced with the metronome revealed qualitatively similar variance patterns across Φ to phrases produced without the metronome. Note that because subject *s6* exhibited an abnormally high standard deviation and mean interval duration in LX for $\Phi_{0.6}$, LY was used for this subject instead.

Both [LX–TdX] and [LX–TX] show greater intergestural variability for most subjects in $\Phi_{0.3}$ relative to $\Phi_{0.5}$. This finding was less consistent across subjects for $\Phi_{0.4}$ and $\Phi_{0.6}$, but nonetheless the majority exhibited either comparable or greater variability, with only two instances of lesser variability associated with the higher-order Φ in [LX–TdX]. In $\Phi_{0.7}$ the hypothesized variability pattern was observed in [LX–TX] intervals more so than in [LX–TdX] intervals.

One possible explanation for the inconsistency in the relative variances between the higher target phase conditions $\Phi_{0.6,\ 0.7}$ and $\Phi_{0.5}$ is that some subjects may have substantially altered the amplitude and/or duration of their articulatory movements because of the proximity of the upcoming onset of the next phrase. In other words, in $\Phi_{0.6,\ 0.7}$ rhythms, having to produce the syllable *take* relatively quickly after *spa* may have led some subjects to decrease the magnitude of the [s] and [p] gestures. Subjects *s2* and *s5* showed relatively large reductions in gestural amplitude in $\Phi_{\ 0.7}$ compared to the other conditions, which may explain why these subjects did not exhibit the expected effects.

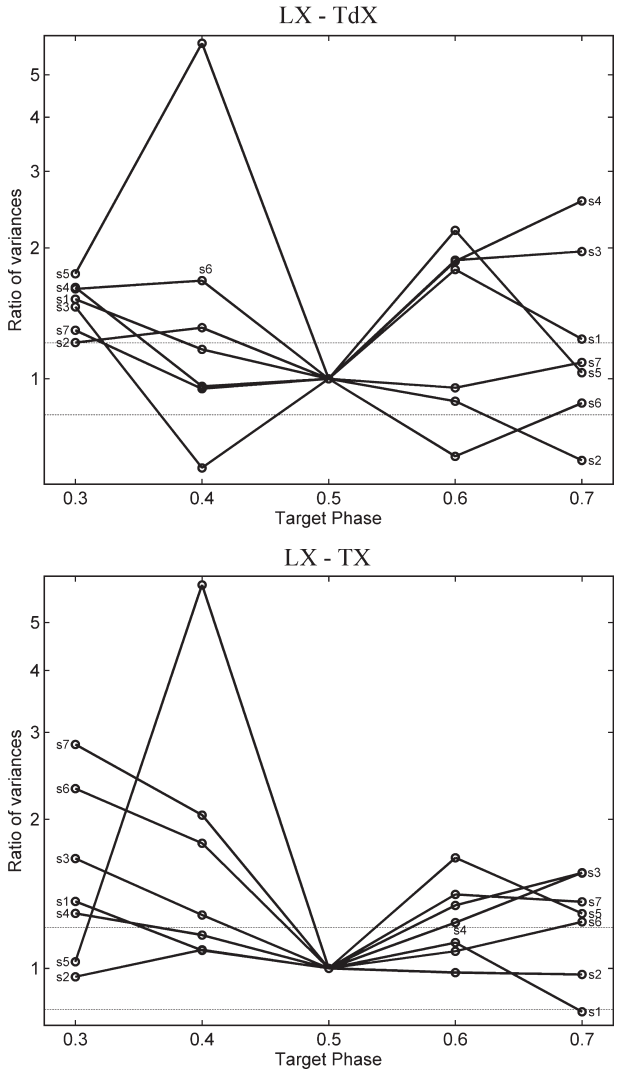

Fig. 10. Intergestural variability comparison. Ratios of variances in each Φ to $\Phi_{0.5}$ are shown for each subject, from phrases produced without the metronome. (Top) Ratios of variances for [LX–TX] interval duration. (Bottom) Ratios of variances for [LX–TdX] interval durations. Horizontal lines indicate approximate 95% confidence intervals.

The observation of reduced amplitude in $\Phi_{0.6,\ 0.7}$ raises the question of whether systematic changes in gestural amplitude and/or duration across target phases could have produced the observed intergestural variability patterns. This reasoning assumes that decreased intragestural movement amplitude will be correlated with decreased intragestural movement duration (less time to move a shorter distance), and hence decreased variance in intergestural interval durations. To wit, increased movement frequency in interlimb coordination has been observed to correlate with decreased intralimb movement amplitudes (Beek, Peper, &

Stegeman, 1995; Haken et al., 1996; Peper, de Boer, de Poel, & Beek, 2008). However, significant amplitude differences were not observed among the lower target phase conditions $\Phi_{0.3,\ 0.4,\ 0.5}$, and by-trial ANOVA with amplitude as a continuous predictor did not show $[X_{TdX} - X_{TX}]$ amplitude to be a significant source of variance in intergestural standard deviations ([LX–TdX]: $F = 1.3, p < .25$; [LX–TX]: $F = 0.57, p < .45$) for those target phases.

Examination of mean [LX–TX] and [LX–TdX] interval durations showed that several subjects (particularly, *s5* and *s7*) produced decreased interval durations in $\Phi_{0.7}$. Analyses of variance with duration as a continuous predictor confirmed that duration has an effect on interval variance. However, when analyses of variance were conducted on intergestural standard deviations only for $\Phi_{0.3,\ 0.4,\ 0.5}$, duration did not have significant effects ([LX–TdX]: $F = 0.32$, $p < .57$; [LX–TX]: $F = 1.13$, $p < .29$), nor were there significant interaction effects between duration and amplitude. This suggests that intergestural interval duration only had an effect in $\Phi_{0.6,\ 0.7}$. This effect opposed the predicted rhythmic-gestural covariability pattern and may explain why several subjects did not exhibit the predicted pattern in those conditions.

It should be noted that from a different perspective, one might expect decreased amplitude to be correlated with decreased stability and increased variability (Van Lieshout, 2004). This has been found to apply to coordinated systems, especially when undergoing a relative phase transition as movement frequency is increased, such as in the Haken–Kelso–Bunz model (cf. Section 1.1). Yet the same does not necessarily apply to intragestural movement amplitudes nor to the intergestural relation considered here, since no phase transition is observed between the gestures.

3.3. Rhythmic–gestural variability correlation

To determine whether variability in rhythmic timing and variability in intergestural timing are correlated, linear regression was performed on normalized variability measures. Normalized variability measures were computed within subjects by dividing the standard deviations of all rhythmic phase (φ) and intergestural interval measures in each target phase condition by the mean standard deviation of those measures across target conditions. Within-subject normalization is appropriate because of absolute differences in standard deviations between subjects. Fig. 11 shows the data points and regression lines for both of the variability measures presented in the preceding section.

Rhythmic and intergestural normalized variabilities were correlated significantly for [LX–TdX] ($R^2 = .26$, $F = 10.85$, $p < .005$) and for [LX–TX] ($R^2 = .20$, $F = 7.71$, $p < .01$). Three outliers (with residuals >95% of expected values, shown by open circles (o) in Fig. 11) were removed from regression analyses; including these outliers results in marginal significance ($p < .12$). A bootstrap analysis with 10,000 samples was conducted to further evaluate the significance of the correlations. For [LX–TdX], the correlation coefficient was $\rho = 0.51$, with a bootstrapped confidence interval of [0.29, 0.70], and for [LX–TX], the correlation coefficient was $\rho = 0.45$, with a bootstrapped confidence interval of [0.16, 0.67]. The R^2 values indicate that rhythmic variance accounts for around 20–25% of the variance in intergestural timing; this is not a very large amount, but it is surely indicative of a

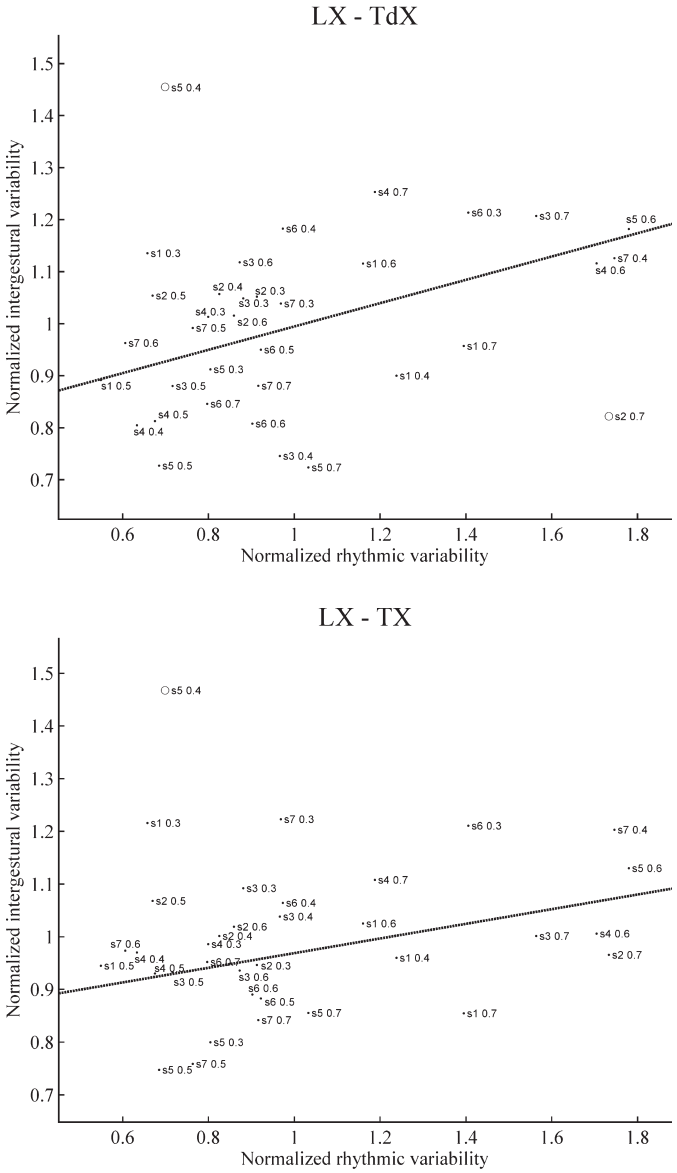

Fig. 11. Rhythmic intergestural normalized variability correlations. (Top) [LX–TdX] intergestural variability. (Bottom) [LX–TX] intergestural variability. Linear regression lines are shown; outliers (o) were excluded from regression.

functional relation. Moreover, one would not expect rhythmic variance to account for most of the intergestural variance, since there are a number other factors (gestural amplitude and duration, subject practice and attention, measurement noise, etc.) that are also important, along with a fair amount of intersubject variability. Even when interpreted conservatively, the regression analysis establishes that rhythmic and gestural temporal variation are

correlated, and it indicates that rhythmic and gestural systems interact in the planning and production of speech.

4. Discussion

The data confirm the hypothesis of rhythmic-gestural covariability: When rhythmic timing was more variable—which occurred generally in Φ conditions other than $\Phi_{0.5}$—intergestural timing between [s] and [p] was also more variable. Although there were several exceptions, the majority of subjects exhibited the hypothesized pattern. The following discussion is partly concerned with understanding how rhythmic–gestural covariability arises. It is shown that a model of coupled oscillators that integrates the rhythmic and gestural dynamical models described in the Introduction can simulate the observed covariability. The data speak to the existence of general principles regarding interactions between systems associated with different timescales, which may be applicable to a variety of human behaviors. It is proposed that activities such as dancing, music performance, and even more abstract cognition may be governed by the same set of general principles and may exhibit analogous multiscale interactions.

4.1. General issues

There are several findings that should be consistent with any proposed explanation for rhythmic–gestural covariability. For one, the harmonic timing observed between rhythmic systems in the phrase repetition task should be related in some manner to the mechanisms responsible for covariability. The relative stability of 1:2 and 1:3 frequency-locking not only accounts for harmonic timing effects but also for increased rhythmic variance in Φ conditions in which phrase and foot systems are related by 1:3 locking (or are subject to more interference between 1:2 and 1:3 frequency-locking). Further, a parsimonious understanding of rhythmic–gestural covariability might make use of this increased rhythmic variance in explaining intergestural variance.

For some subjects, gestural amplitude and intergestural interval duration were found to have effects on temporal variability patterns in $\Phi_{0.7}$. These effects should not be precluded by any model but will not be of primary concern in what follows. Because intergestural variability patterns in $\Phi_{0.3–0.5}$ were not significantly influenced by these factors, we should be able to understand the mechanism(s) behind the rhythmic–gestural covariability effect independently of amplitude and duration influences on variance. From a modeling perspective, not only would more data be required to better understand the effects of intersubject variation in intergestural duration and gestural amplitude, but changes to the data collection procedure would be necessary as well.

A difficult question here is whether phrasal tempo or frequency need be taken into account in understanding covariability. A cornerstone of the dynamical approach has been the idea that increased movement frequency corresponds to decreased coupling strength, particularly for higher-order frequency-locks. The crux of the difficulty here is that—in contrast to gestural and rhythmic interval durations—the hypothesized gestural and

rhythmic oscillatory systems are not directly observable. The interpretations of tempo or frequency variation in interval durations and in hypothesized oscillatory periods differ substantially. In the speech cycling task, the frequency/tempo of the phrasal period increases across the target phase conditions. But the foot durations do not change accordingly, since the first foot duration in the phrase remains close to 0.700 s in all conditions, and the second foot shortens to some extent only in the higher $\Phi_{0.6,0.7}$. Moreover, the frequencies of hypothesized oscillatory systems governing foot timing vary, but not in proportion to the phrasal frequency: The foot oscillator is actually faster in $\Phi_{0.4}$ than in $\Phi_{0.5}$, and only a little slower in $\Phi_{0.3}$ than in $\Phi_{0.5}$ (cf. Fig. 4). It is thus far from clear whether the speech-cycling task instantiates variation in tempo, and variation in frequency of hypothesized foot planning oscillators is not monotonically related to phrasal period.

Gestural durations do not increase in period all that much across conditions for most subjects, and are not oscillatory. In contrast, the hypothesized gestural planning oscillators have frequencies that are equivalent to the syllable oscillator frequency, which itself obtains a 2:1 or 3:1 ratio to the foot oscillator frequency. Thus, the frequency varies in the systems hypothesized to govern relative timing, but generally not in the observables of foot and gestural durations. Whether frequency and tempo have effects on rhythmic and gestural timing should be investigated by simultaneous variation of the phrasal repetition period and both foot intervals, but cannot be appropriately addressed in the present context, where the period of the first foot was constant across target rhythms.

Despite the subsequent focus on coupled oscillatory systems, it should be noted that other classes of models might make similar predictions, under the right assumptions. One possibility is that limits on attentional resources can explain rhythmic–gestural covariability. If attention is a limited resource, and some attention must be allocated to perform both rhythmic and intergestural timing with accuracy, then circumstances in which one type of timing is more difficult (and hence requires more resources) should result in fewer resources being allocated to the other. If more attentional resources are allocated to rhythmic timing in the higher-order Φ, then there should be a concomitant reduction in attention to intergestural timing, leading to increased intergestural variability. One drawback to this sort of explanation is that it does not relate the phenomenon of harmonic timing to covariability.

Models based upon hierarchically organized timekeepers (Ohala, 1975; Vorberg & Wing, 1996; Wing & Kristofferson, 1973) have been successful in explaining other sorts of variability patterns, such as negative lag-1 autocovariance. One might thus expect them to be similarly useful here. Can these models indeed simulate rhythmic–gestural covariability patterns through a group of timekeepers arranged hierarchically? The traditional assumption that a timekeeper is started by exactly one other timekeeper and then runs its course independently of that triggering event (Vorberg & Wing, 1996) seems to prohibit such covariability effects from arising. Timekeeper models do not normally provide a way for one process to continuously influence another (or for feedback between processes), nor a way for multiple processes to simultaneously exert an influence over the timing of a single lower-level process. Indeed, as will be clear below, contemporaneous and continuous influence of multiple rhythmic systems upon gestural systems is fundamental to a dynamical understanding of rhythmic–gestural covariability.

4.2. Dynamical model of rhythmic–gestural interaction

To account for rhythmic–gestural covariability in a dynamical framework, assume that all relevant prosodic units—phrase, foot, syllable—and gestural units C_1[s], C_2[p], and V[a] correspond to planning oscillators with phase variables (θ_{Phr}, θ_{Ft}, θ_{σ}, θ_{C1}, θ_{C2}, θ_{V},) and radial frequencies ($2\pi\omega_{Phr}$, $2\pi\omega_{Ft}$, etc.). These systems are presumed to correspond to the transient states of the averaged dynamics of distributed neural ensembles associated with speech planning, and although not currently observable, they should be observed in the future through more sophisticated analyses of EEG, MEG, or large-scale polycellular recordings.

In accordance with Cummins and Port (1998) and Port (2003), the ratio of frequencies ω_{Phr}:ω_{Ft} is 1:2 in $\Phi_{0.5}$, and either 1:2 or 1:3 in the higher-order Φ. One way to conceptualize the source of rhythmic variability is as a form of coupling strength–frequency order relation, whereby the amplitude of the relative phase potential is higher for systems with lower-order frequency-locking (i.e., 1:2 > 1:3 > 1:4, etc.). This is essentially the strategy used by Haken et al. (1996): higher-amplitude potentials in 1:2 coupling exert stronger forces to correct noise-induced perturbations in the relative phase. As for why such a relation occurs, one reason may be a form of ''interference'' between potentials corresponding to different frequency locks. In any given rhythmic condition, potential forces corresponding to both 1:2 and 1:3 frequency-locking are operative, although one mode dominates. Because 1:2 is of lower order, it exerts greater interference on 1:3 than the reverse. An alternative possibility is that faster oscillations are inherently noisier, that is, the level of noise in their phase velocity is frequency dependent. Both of these options will be considered below.

There are many possibilities for exactly what form the equations describing a group of coupled oscillatory systems can take, particularly with regard to the form of the oscillator equations, the coupling function, and how noise enters into the picture. Empirical data can inform these choices, although much remains to be explored. The model presented below draws heavily from Keith and Rand (1984), Kopell (1988) Saltzman and Byrd (1999, 2000), Nam & Saltzman (2003), and Haken et al. (1996), although for reasons described below, the model posits no amplitude dependence of coupling strength. The idea of a multitimescale dynamical model integrating rhythmic and gestural speech systems is not novel, but this is the first report of such a model being used to explore relations between rhythmic and gestural variability.

One of the key differences between the observables in the present experiment and those in studies of rhythmic interlimb coordination is the lack of a measurable amplitude of oscillation. With swinging legs or wagging fingers, the motions of the phase variables correspond to physical motions that are both cyclic and have easily measured positions in space. This is not so with the rhythmic measures derived from p-centers, which are spaceless points in time. By the Cummins and Port (1998) and Port (2003) hypotheses the locations of p-centers are attracted to pulses that are phase-locked to foot oscillations. Yet there is no known way to observe the amplitudes of these oscillations. With rhythmic interlimb coordination, observable characteristics of the limbs (such as amplitude) can be reasonably attributed to the oscillatory systems hypothesized to govern those movements. In contrast, it would be hasty

to endow speech–rhythmic oscillators with varying amplitudes in the absence of a directly related observable.

The same problem applies to the gestural systems. Physical correlates of gestures do not exhibit oscillatory amplitude variation, but rather behave like systems with point-attractors. Gestures (not gestural planning systems) are modeled with critically damped mass-spring systems in the task dynamic approach (Saltzman & Munhall, 1989). The target amplitude of a gesture is determined by the force exerted upon it when it becomes activated, which corresponds to defining a new equilibrium for the mass-spring system. However, the observable amplitude of a gesture will be influenced both by its parameterized target equilibrium and by the effects of contemporaneously activated gestures. Browman and Goldstein (2000) hypothesized that gestural planning oscillators determine the relative timing of gestural onsets, and Nam and Saltzman (2003) provided a computational model of this idea. However, the gestural planning oscillators do not govern the amplitudes of the gestural systems, nor can their amplitudes be observed. For these reasons, the model presented below does not attempt to account for the influence of oscillator amplitude on coupling strength (Peper et al., 2008). Instead, coupling strengths are based entirely on phase relations and all systems have unit radial amplitude. It may be possible to couple the gestural target equilibria to amplitude variation in gestural planning oscillators, although this possibility is not explored here.

Not only do both rhythmic and gestural oscillatory systems in speech lack observable amplitudes, but their phases are only observable indirectly and at limited points in time. Foot-systems, for example, can be stipulated to have phase 0 at stressed syllable p-centers, but their phases are not easily inferred at other points in time. Gestural planning system phases could be stipulated to correspond to special points of corresponding gestures, such as positional and velocity extrema, but these are not well-substantiated assumptions. Under the weakened assumption that the relative timing of gestural onsets is determined by gestural planning oscillator relative phases (Nam & Saltzman, 2003), then as long as gestural stiffness and target amplitude parameters (which together determine velocity; cf. Saltzman & Munhall, 1989) are unchanged, differences in intergestural relative phase translate into temporal differences in the timing of positional and velocity extrema associated with the gestures. In other words, we assume that variability in the gestural planning oscillators can be observed indirectly through measurement of kinematic landmarks.

Each oscillator in a multiscale model can be described in a simple fashion by differential equations governing its phase and amplitude, shown in Eq. (1):

$$\dot{\theta}_i = \omega_i(1+\eta_i) + \sum_j A_{ij}F(\phi_{ij}) \qquad (1)$$

$$\dot{r_i} = 1$$

$$\phi_{ij} = \theta_i - \theta_j \bmod 2\pi \qquad (2)$$

$$F(\phi_{ij}) = -\frac{\mathrm{d}V_{ij}}{\mathrm{d}\phi_{ij}} \qquad (3)$$

$$V = (\phi_{ij}) = \mathrm{COS}(\phi_{ij} - \Phi_{ij})$$

The equations in (1) show that the phase velocity of each oscillator is the inherent frequency (ω) of the oscillator modulated by Gaussian noise (η), plus the sum of phase changes due to coupling forces exerted on the oscillator (F). The coupling forces are weighted by corresponding coupling strengths (A). The polar amplitudes (radii) of the oscillators are constant, and so in the absence of noise and coupling, the equations in (1) correspond to a polar form of a harmonic oscillator. These systems can be visualized very readily as points moving around unit circles. More biologically plausible models usually incorporate some form of nonlinear damping, such as Van der Pol and Rayleigh damping. For current purposes, these nonlinear terms are unnecessary, because the entirety of the conceptual workload can be carried by the notions of phase coupling and synchronization, and because as explained above, the oscillations do not correspond directly to physical observables.

The coupling function F takes as its argument the difference between a target generalized relative phase (Φ_{ij}) and the current generalized relative phase (φ_{ij}) between a pair of oscillators i and j with phases θ_i and θ_j. The generalized relative phase is defined in Eq. (2) and is modulo 2π like the phase variable. Oscillator frequencies (ω) are restricted to integer multiples of the lowest frequency, which belongs to the phrase oscillator. The phrase oscillator can be defined to have an inherent frequency of 2π with the introduction of an appropriately normalized time variable. Following Haken (1983) and Saltzman and Byrd (2000), the coupling force between a pair of oscillators is the negative of the derivative of the potential function (V) governing their generalized relative phase, shown in Eq. (3). The potential force is zero when the relative phase between two oscillators is Φ_{ij}, that is, the target relative phase, which defines a stable equilibrium. The absolute value of the derivative of the potential corresponds to the speed with which relative phase changes and is greatest when $[\varphi_{ij} - \Phi_{ij} = \pm\frac{1}{2}\pi]$, halfway between the stable equilibrium $[\varphi_{ij} - \Phi_{ij} = 0]$ and unstable equilibrium $[\varphi_{ij} - \Phi_{ij} = \pi]$.

The coupling forces can either speed up or slow down the phase velocity of an oscillator. For any pair of oscillators, osc_i and osc_j, the strengths of the coupling forces exerted upon osc_i by osc_j and vice versa are represented separately in the coupling strength parameter matrix (A_{ij}). This allows for either oscillator to drive the other one, or for mutual interaction, which can be symmetric or asymmetric. General hypothetical parameter matrices relevant for rhythmic–gestural interaction in productions of the phrase *take on a spa* are shown in Table 1.

In the coupling structure in Table 1, the syllable [spa] belongs to its own foot, and interactions between this syllable and others are not considered. Moras have been omitted, as well as many other gestures involved. C[s] represents the tongue blade protrusion gesture associated with [s], C[p] the bilabial gesture, and V[a] represents a tongue body gesture associated with [a]. A target relative phase of 0 indicates in-phase synchronization. Although there are two Ft systems involved in production of the phrase, they are synchronized in-phase and collapsed into a single system in the present treatment. The only anti-phase specifications in the model adhere between the [s] and [p] gestures and are responsible for a c-center effect relative to the vowel (cf. Section 1.2).

The parameters instantiated in the coupling strength matrix correspond to a subset of all possible hypotheses for such a model. Parameters appearing on both sides of the diagonal

Table 1
Coupling strength and target relative phase parameter matrices relevant to the syllable *spa* in the phrase *take on a spa*

		Coupling Strengths *A*						Generalized Relative Phase Targets Φ					
		Phr	Ft	σ	$C_{[s]}$	$C_{[p]}$	$V_{[a]}$	Phr	Ft	σ	$C_{[s]}$	$C_{[p]}$	$V_{[a]}$
$\omega_{Phr} = 2\pi$	Phr		*a*						0				
$2\omega_{Phr}$, $3\omega_{Phr}$	Ft	*a*		*a*	*b*	*b*	*b*	0		0	0	0	0
$2\omega_{Ft}$, $3\omega_{Ft}$	$\sigma_{/spa/}$		*a*		*c*	*c*	*c*		0		0	0	0
ω_{σ}	$C_{[s]}$					*d*	*e*					–½π	0
ω_{σ}	$C_{[p]}$				*d*		*e*				½π		0
ω_{σ}	$V_{[a]}$				*f*	*f*					0	0	

Notes: Coupling strengths represent the influence of the system in the corresponding row upon the system in the corresponding column. ω, radial frequency. Phr, Ft, and σ indicate the rhythmic systems: phrase, foot, and syllable, and C[s], C[p], and V[a] indicate consonantal and vocalic gestural planning systems associated with the [s], [p], and [a] in the word ''spa.''

indicate either symmetric or asymmetric coupling. For example, the parameters *a* and *d* represent symmetric coupling, meaning that the phrase and foot systems, and also the consonantal gestures, mutually influence each other to the same extent. The parameters *e* and *f* represent the possibility of asymmetric coupling between consonants and vowels, so that one may influence the other to a greater extent. Where parameters appear only on one side of the main diagonal, this indicates a unidirectional coupling (often called *driving*), which means that the system in the corresponding row influences the system in the corresponding column, but not vice versa. In the coupling structure above, *b* and *c* represent foot and syllable driving influences upon gestural systems.

Covariability between rhythmic and gestural systems in the present experiment can be modeled parsimoniously with variation of coupling strength between rhythmic systems (*a*). Decrease in rhythmic coupling strength reflects changes in rhythmic difficulty or increased interference between 1:2 and 1:3 frequency-locking associated with the higher-order *Φ*. More difficult rhythms are associated with greater variance in rhythmic timing, and the model can reproduce this effect with a decrease in inter-rhythmic coupling.

Phase velocity noise plays a crucial role in how decreased inter-rhythmic coupling is translated to increased intergestural variability. Due to its Gaussian nature, the noise normally has only a small effect on oscillator phases (which translates into small effects on relative phases), but on occasion the noise has a larger effect. With relatively strong coupling forces between rhythmic systems, large noise-induced fluctuations in oscillator phases and relative phases are corrected more quickly, making rhythmic relative phases less variable. This in turn produces a more coherent, stronger overall force on gestural systems through the rhythmic–gestural coupling. When there is more noise, this effect is of larger magnitude.

An alternative way in which rhythmic–gestural covariability can arise is through a relation between the order of frequency-locking and noise. If higher-order frequency-locking is subject to more noise (perhaps due to interference from lower-order potentials),

then whenever the 1:3 ratio dominates the coupling between phrase and feet, variability is higher for both rhythmic and gestural systems.

4.3. Model simulations

Numerical simulations verified that a multifrequency system of coupled oscillators can replicate the experimentally observed rhythmic–gestural covariability. Fig. 12 presents an example simulation of the model. The top panel shows oscillator positions (i.e., cos θ_i) and selected relative phases over time. For both rhythmic and gestural systems, the oscillators begin out of phase but eventually synchronize. This synchronization can be observed in the stabilization of generalized relative phases (φ). Initial oscillator phases in this example were θ = [1.02, 5.75, 0.30, 5.40, 0.09, 2.20] radians, corresponding in degrees to [58°, 329°, 17°, 309°, 5°, 126°]. Regardless of the initial phases of the oscillators, the same relative phase configuration is eventually reached; initial phases only have an effect on how long relative phases take to stabilize. Following Saltzman and Nam (2003), Φ_{C1C2} was set lower than $-\pi$ (ideal anti-phasing), in this case $\Phi_{C1C2} = -1/2\pi$. Further details regarding simulation parameters can be found in the Appendix. For the Phr, Ft, and σ systems, the pattern is such that

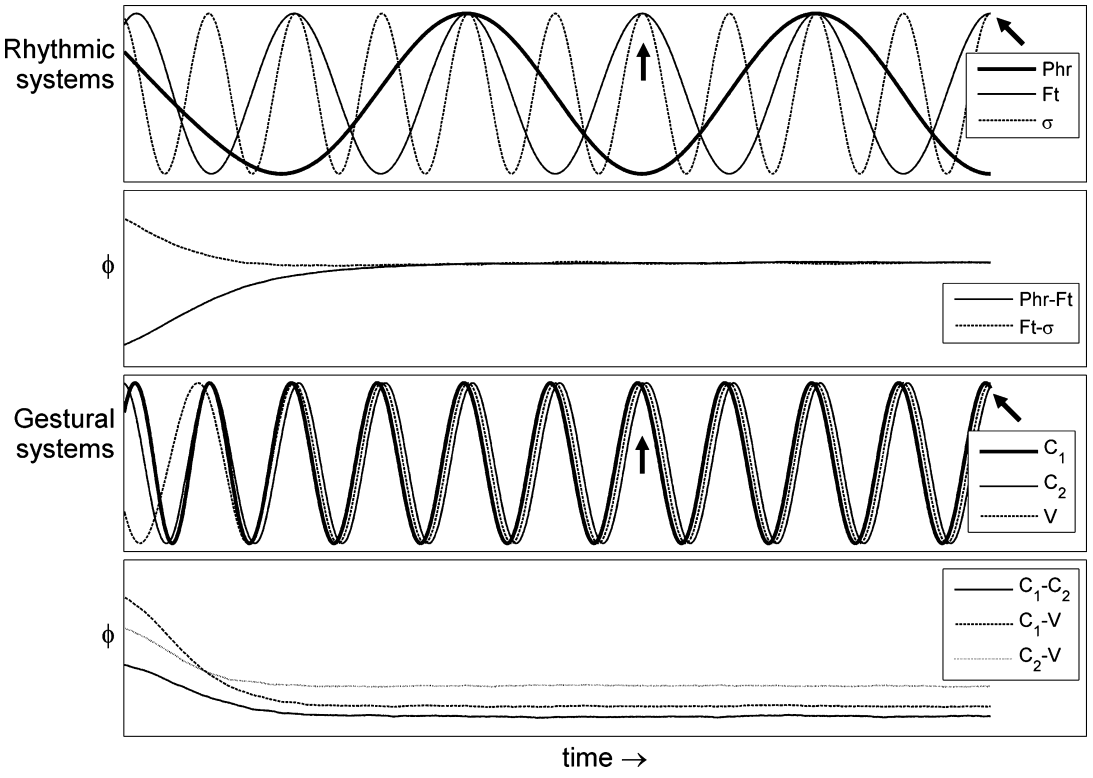

Fig. 12. Example simulation. Oscillator motions and relative phases are shown over time; arrows indicate points in time where relative phases are sampled.

the Ft repeats twice within the Phr period, the σ repeats four times, and both of these oscillators hit a peak (i.e., $\theta = 0 = 2\pi$) approximately when the Phr oscillator does. For the C_1, C_2, and V systems, the C_1 phase precedes V by the same amount that C_2 follows V—this exemplifies a c-center effect.

The simulation in Fig. 12 ends when Phr reaches a minimum ($\theta = \pi$), and thus the Ft, σ, and V oscillators are at a peak. Simulation results are presented below from four conditions: with 1:2 or 1:3 phrase-foot frequency ratios and with relatively high or low noise levels. Standard deviations of φ_{C1C2} were calculated from 1,500 samples of φ_{C1C2} in each condition (for further details, see Appendix). Fig. 13 shows that as inter-rhythmic coupling (a) is increased, intergestural relative phase φ_{C1C2} becomes less variable. In addition, 1:3 inter-rhythmic coupling incurs more intergestural variability than 1:2 coupling. This latter effect is a consequence of the modulation of inherent oscillator frequency by Gaussian noise: Faster systems will be subject to a greater degree of phase velocity noise.

To translate the standard deviations in Fig. 13 into quantities that can be more directly compared to the experimental data, some assumptions must be made. A reasonable possibility is that one period of the syllable oscillator corresponds to the average duration of the syllable ($\Delta\sigma$), and so simulated durations can be derived by assuming $\Delta\sigma = 2\pi$. Then, if $\Delta\sigma \approx 250$ ms, the standard deviations in Fig. 13 range from about 16 to 66 ms, which agrees remarkably well with the empirical range of standard deviations for all subjects: 13 to 60 ms. The changes in standard deviations observed within a given model as coupling strength is varied are smaller, on the order of 8 to 16 ms. These ranges are in line with the observations made for many subjects across target phase conditions, although there are some who exhibited larger changes in intergestural standard deviations between Φ conditions;

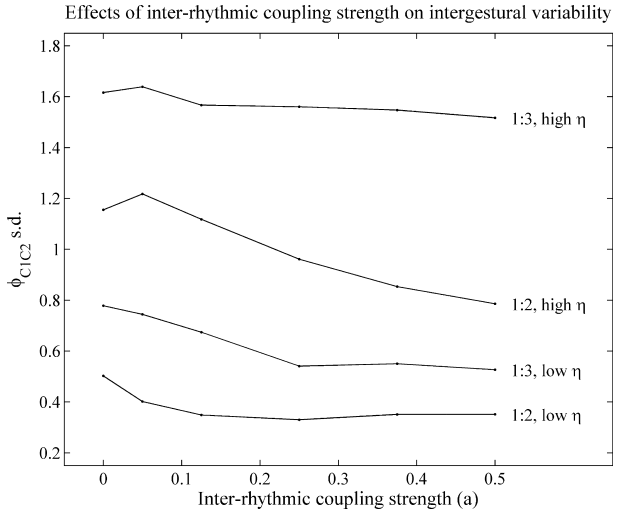

Fig. 13. Effects of inter-rhythmic coupling strength on intergestural variability: *SD* φ_{C1C2}. *SD*s of φ_{C1C2} are plotted for a range of a (inter-rhythmic coupling strengths) for 1:2 and 1:3 phrase-foot coupling in relatively high and low noise levels (η).

these may require additional mechanisms not captured by the model, or may be found in other regions of the model parameter space.

These simulations demonstrate that either one of two possible mechanisms may be responsible for rhythmic–gestural covariability. Weaker coupling between rhythmic systems leads to less coherent and weaker forces acting to stabilize the relative timing of gestural systems, resulting in increased intergestural variability; this effect is larger with greater noise levels. In addition, relatively greater noise levels associated with the higher-order 1:3 frequency-locking result in greater intergestural variability. Although the role of noise in the current simulations is entirely negative, it should be pointed out that noise can also be seen to play an important role in providing flexibility in coupling relations (Kelso & Fuchs, 1995; Mayer-Kress & Newell, 2003). Importantly, rhythmic–gestural coupling strengths were held constant in all the simulations ($b = c = 0.5$), so effects are attributable entirely to changes in inter-rhythmic coupling (a). If rhythmic-gestural coupling is similarly increased in proportion to inter-rhythmic coupling, then the covariability effects become even larger.

In the present implementation of the model, target relative phases of gestural onsets are derived from sampling gestural planning system relative phases, as in Nam & Saltzman (2003). However, an alternative is to couple linearly damped gestural oscillators to the limit cycle gestural planning oscillators, as has been proposed for interlimb coordination by Beek et al. (2002). Also, the implementation assumes unidirectional coupling from rhythmic systems to gestural systems, but a likely possibility is that this coupling is bidirectional. The effects of such bidirectional coupling have not yet been worked out, but they should depend greatly upon the strength with which the gestural systems influence the rhythmic ones. The observation that in many languages syllable weight (usually the presence of a coda or long vowel) influences the location of stress may constitute evidence for bidirectional coupling.

There are numerous ways in which the model presented above can be restructured or differently parameterized, some of which could be challenging—but not impossible—to distinguish experimentally. Regardless of which is correct, the importance of the model lies in its dynamical approach, particularly in its use of the concepts of stochastic influence, coupling, generalized relative phase, and multitimescale/multifrequency interaction. These concepts are general and will likely outlive major changes in specific instantiations of the model. They allow for a coherent understanding of an otherwise perplexing experimental observation: the correlation of variability in intergestural timing with variability in rhythmic timing.

4.4. Extensions

Given that multitimescale dynamical interactions occur in speech, it would be odd not to find them in other cognitively coordinated human behaviors. For example, consider the generic version of the salsa dance shown in Fig. 14A, in which Step 1 is a forward step with the left foot, Step 2 is a small step or weight shift with the right foot, and Step 3 moves the left foot back to its starting position. The same steps are then repeated with the opposite feet in the opposite direction. Fig. 14B depicts the musical beats that correspond to these steps as quarter notes followed by a quarter rest. This dance occurs in 4/4 time, meaning that there are four beats per measure. Each set of four beats is a sort of ‘‘phrase,’’ represented by a

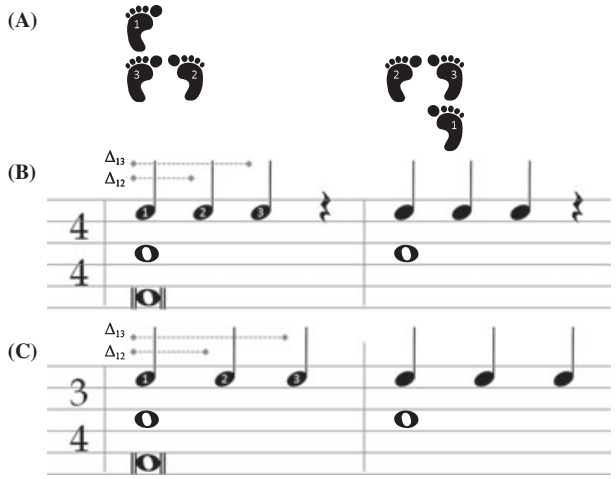

Fig. 14. Dance steps performed to salsa-like and waltz-like rhythms. (A) Representation of dance steps commonly used in salsa. (B) Musical rhythm corresponding to salsa, in 4/4 time. (C) Musical rhythm corresponding to waltz, in 3/4 time.

whole note in the figure. Because the pattern is repeated over two measures, the ''superphrase'' of the dance corresponds to eight beats, which is represented by the breve note on the bottom of the score. The same pattern of steps can be performed as a waltz (although usually with some stylistic differences). The waltz gets its characteristic rhythm from being performed in 3/4 time, meaning that there are only three beats per measure. Fig. 14C shows how the steps correspond to musical beats if they are performed as a waltz.

An interesting question to ask regards the timing of the second step (i.e., ''foot'') relative to the first and third steps. The variability of this timing could be measured as the ratio of the interval between the first and second steps (Δ_{12}) to the interval between the first and third steps (Δ_{13}). The question is whether there would be more variability in this relative phase measure in the salsa or the waltz. The prediction depends upon what is taken to be the relevant frequency ratio. If the salsa rhythm is governed by a 1:4 ratio and the waltz by a 1:3 ratio, then the variability should be higher in the salsa rhythm (assuming that the tempo is held constant and the dancer is equally well practiced in both rhythms, etc.). However, if the salsa rhythm is governed primarily by half notes (data not shown), which are analogous to ''subphrases'' and can occupy half of each measure in 4/4 time, then the system coordinating the dance steps would obtain a 1:2 frequency ratio and should be less variable than the waltz. However, if one takes into account that the salsa rhythm exhibits both 1:2, 1:4, and 1:8 ratios between superphrase, phrase, subphrase, and foot, and that the waltz exhibits 1:3 and 1:6 ratios between superphrase, phrase, and foot, then it is no longer clear what predictions should be made.

More to the point, if it is found that in one style the steps are more variable than in the other, an analogy with rhythmic–gestural covariability can be investigated by measuring the timing of additional flourishes that inhabit a faster timescale. For example, imagine that

some sort of two-movement shimmy or rapid hip twists occur on the second step (such flourishes are not uncommon in dancing). These faster movements correspond to eighth notes. The prediction that is made from analogy with rhythmic–gestural covariability in speech is that the relative timing of movements associated with the flourish will be more variable in the dance that is more variable rhythmically. Further, if these rhythms are being performed instrumentally—for example, on a piano—the same covariability relation should arise.

The dancing example is similar to speech in that there are movements associated with different timescales, and therefore the prediction of covariability effects in dancing is an obvious extension of the speech findings. Multitimescale dynamical interactions of the same sort are also likely to be observed in temporal relations between gesture and language, and in sign language (Goldin-Meadow, in press; So et al., in press). Speech is often accompanied by manual gestures that convey spatial and temporal information, and some of these involve repeated or iterated movements. For example, imagine that in describing the changing viewpoints of a political candidate, someone says that the candidate has ''gone back and forth, and back and forth on that issue...,'' and accompanies this statement with a repetitive hand motion. It is likely that the frequency and timing of the hand motion exhibits systematic regularities relative to the speech gestures in the utterance, and in some cases these may involve multifrequency relations. For another example, in sign languages, various sign gestures are made in rapid succession, but it remains to be determined whether the timing of these signs is governed by lower-frequency patterns involving phrases.

A tantalizing and even more profound possibility is that cognitive behaviors not involving movement can be usefully understood with multiscale dynamical interactions. Given recent evidence that motor simulation plays an important role in action recognition and conceptual reasoning (Arbib, 2005; Gallese et al., 1996), one might expect higher-level cognition to be governed by some of the same principles.

5. Conclusion

The experimental evidence reported here demonstrates that rhythmic and gestural systems in speech interact in a complex way. This evidence was seen in the correlation of intergestural temporal variability and rhythmic variability. This finding is important because conventional, nondynamical models of speech production do not predict such effects, and because multitimescale covariability may occur in many other cognitive domains. These variability patterns can be modeled with a multifrequency system of coupled planning oscillators, in which stronger coupling between lower-frequency oscillators makes the relative phases of higher-frequency oscillators less variable.

It is worth pointing out an inconsistency between current dynamical models and what is more realistically occurring in cognitive speech planning systems. In many instantiations to date, the dynamical descriptions of speech planning systems employ limit cycles with constant amplitude parameters, which entails that the systems oscillate indefinitely. This is obviously false: Oscillations must in reality be transient. Most models have used limiting

fixed points of the relative phase to make predictions about observables, but we must be open to the possibility that planning system amplitude variation occurs rapidly and that this has important consequences for what we observe. Moreover, in speech outside of the laboratory, the ''phrases'' used here as fundamental timescales are not coherent entities, and they are influenced by transiently active syntactic and semantic systems that presumably have their own complex dynamics.

Before future modeling efforts can address these issues, further investigation of the interaction between rhythmic and gestural timing should be conducted. Some individual variation was observed in the present experiment, and understanding the sources and consequences of this variation is important. There are many ways in which the experimental design can be altered to probe rhythmic–gestural interaction from different angles, and experimental paradigms that refine the speech cycling task are needed. It is my hope that this study will spark future research into phenomena of interaction between speech rhythm and gesture, as well as research into potentially related phenomena in a variety of cognitive domains.

Acknowledgments

This research would not have been possible without the generosity and guidance of Peter Watson and Ben Munson. Pascal van Lieshout, Ken de Jong, and Eliott Saltzman provided insightful reviews and editorial comments on earlier versions of this manuscript. Thanks to Louis Goldstein, Keith Johnson, Rich Ivry, and Sharon Inkelas for conversations and discussions throughout the design and analysis of this experiment, and thanks to Hosung Nam for help in numerical modeling.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.

Alfonso, P., & Van Lieshout, P. (1997). Spatial and temporal variability in obstruent gestural specification by stutterers and controls: Comparisons across sessions. In W. Hulstijn, H. Peters, & P. Van Lieshout (Eds.), *Speech production: Motor control, brain research, and fluency disorders* (pp. 151–160). Amsterdam: Elsevier.

Allen, G. (1972). The location of rhythmic stress beats in English: An experimental study, parts I and II. *Language and Speech*, *15*, 72–100, 179–195.

Allen, G. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*, 75–86.

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*, 105–167.

Barbosa, P. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. *Speech Prosody 2002 Proceedings* (pp. 163–166). Aix-en-Provence, France.

Beek, P., Peper, C., & Daffertshofer, A. (2002). Modeling rhythmic interlimb coordination: Beyond the Haken–Kelso–Bunz model. *Brain and Cognition*, *48*, 149–165.

Beek, P., Peper, C., & Stegeman, D. (1995). Dynamical models of movement coordination. *Human Movement Science*, *14*, 573–608.

Beer, R. (1995). A dynamical systems perspective on agent–environment interaction. *Artificial Intelligence*, *72*, 173–215.

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Cambridge, MA: Blackwell.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140–155.

Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I. Between the grammar and physics of speech* (pp. 341–376). Cambridge, England: Cambridge University Press.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlee*, *5*, 25–34.

Carson, R., Goodman, D., Kelso, J., & Elliot, S. (1995). Phase transitions and critical fluctuations in rhythmic coordination of ipsilateral hand and foot. *Journal of Motor Behavior*, *27*(33), 211–224.

Castellano, C., Fortunato, S., & Loreto, V. (2007). *Statistical physics of social dynamics*. arXiv:0710.3256.

Cummins, F., & Port, R. (1996). Rhythmic constraints on English stress timing. In H. T. Bunell & W. Idsardi (Eds.), *Proceedings of the fourth international conference on spoken language processing* (pp. 2036–2039). Wilmington, DE: Alfred duPont Institute.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.

Dauer, R. (1983). Stress timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.

Eriksson, A. (1991). *Aspects of Swedish speech rhythm*. Göteborg: University of Göteborg.

Fink, P., Foo, P., Jirsa, V., & Kelso, J. (2000). Local and global stabilization of coordination by sensory information. *Experimental Brain Research*, *134*, 9–20.

Fowler, C. (1979). Perceptual centers in speech production and perception. *Perception and Psychophysics*, *25*, 375–388.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music*, *31*(3), 221–272.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119* (2), 593–609.

Goldin-Meadow, S. (in press). Homesign: When gesture becomes language. In R. Pfau, M. Steinbach & B. Woll (Eds.), *Handbook on sign language linguistics*. Berlin: Mouton de Gruyter.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *From action to language: The mirror neuron system* (pp. 215–249). Cambridge, England: Cambridge University Press.

Goldstein, L., & Fowler, C. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and productions* (pp. 159–207). Berlin: Mouton de Gruyter.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Review of Modern Physics*, *47*, 67–121.

Haken, H. (1983). *Synergetics, an introduction: Nonequilibrium phase transitions and self-organization in physics, chemistry, and biology* (3rd ed.). New York: Springer-Verlag.

Haken, H., Kelso, J., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, *51*, 347–356.

Haken, H., Peper, C., Beek, P., & Daffertshofer, A. (1996). A model for phase transitions inhuman hand movements during multifrequency tapping. *Physica D*, *90*, 179–196.

Halle, M., & Vergnaud, J. (1978). *Metrical structures in phonology*. Unpublished manuscript, MIT.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hoole, P. (1996). Issues in the acquisition, processing, reduction and parameterization of articulographic data. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München*, *34*, 158–173.

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope. *Perceptual Psychophysics*, *43*(1), 90–93.

Hyman, L. (2006). Word-prosodic typology. *Phonology*, *23*(2), 225–258.

de Jong, K. (1994). The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics*, *56*(4), 447–460.

Keith, W., & Rand, R. (1984). 1:1 and 2:1 phase entrainment in a system of two coupled limit cycle oscillators. *Journal of Mathematical Biology*, *20*, 122–152.

Kelso, J. (1981). On the oscillatory basis of movement. *Bulletin of the Psychonomic Society*, *18*, 63.

Kelso, J. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology; Regulatory, Integrative, and Comparative*, *15*, R1000–R1004.

Kelso, J. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT.

Kelso, J. A. S., & Fuchs, A. (1995). Self-organizing dynamics of the human brain: Critical instabilities and sil'nikov chaos. *Chaos*, *5*(1), 64–69.

Kopell, N. (1988). Toward a theory of modeling central pattern generators. In A. Cohen, S. Rossignol, & S. Grillner (Eds.), *Neural control of rhythmic movement in vertebrates* (pp. 369–413). New York: John Wiley & Sons.

Liberman, M. (1975). *The intonational system of English*. PhD Dissertation. New York: MIT, Garland Publishing Co.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8* (3), 249–336.

Mayer-Kress, G. J., & Newell, K. M. (2003). Noise and chaos in motor behavior models. *Proceedings of SPIE – The International Society for Optical Engineering*, *5110*, 320–333.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405–408.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. *15th International Congress of Phonetic Sciences*, *3*, 2253–2256.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.

O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. *Proceedings of the XIV International Congress of Phonetic Sciences*, *2*, 1075–1078.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: The role of the mirror neuron system and simulation in social and communicative deficits of autism spectrum disorders. *Psychological Bulletin*, *133*(2), 310–327.

Ohala, J. (1975). The temporal regulation of speech. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and the perception of speech* (pp. 431–453). New York: Academic Press.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115–138.

Patel, A., Löfqvist, A., & Naito, W. (1999). The acoustics and kinematics of regularly timed speech: A database and method for the study of the p-center problem. *Proceedings of the 14th International Congress of Phonetic Sciences*, *1*, 405–408.

Peper, C., & Beek, P. (1998). Distinguishing between effects of frequency and amplitude on interlimb coupling in tapping a 2:3 polyrhythm. *Experimental Brain Research*, *118*, 78–92.

Peper, C., de Boer, B., de Poel, H., & Beek, P. (2008). Interlimb coupling strength scales with movement amplitude. *Neuroscience Letters*, *437*, 10–14.

Pike, K. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchronization: A universal concept in nonlinear sciences*. Cambridge: Cambridge Press.

de Poel, H., Peper, C., & Beek, P. (2007). Handedness-related asymmetry in coupling strength in bimanual coordination: Furthering theory and evidence. *Acta Psychology*, *124*, 209–237.

Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics*, *17, 175–192.*

Port, R. (1986). Translating linguistic symbols into time. *Research into phonetics and computational linguistics*. Report No. 5, Bloomington, IN: Department of Linguistics and Computational Linguistics, Indiana University.

Port, R. (2003). Meter and Speech. *Journal of Phonetics*, *31*, 599–611.

Port, R., Cummins, F., & Gasser, M. (1995). *A dynamic approach to rhythm in language: Toward a temporal phonology*. Technical Report No. 105. Bloomington, IN: Indiana University Cognitive Science Program.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Series*, *15*, 129–144.

Saltzman, E., & Byrd, D. (1999). Dynamical simulations of a phase window model of relative timing. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth international congress of phonetic sciences* (pp. 2275–2278). New York: American Institute of Physics.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, *19*, 499–526.

Saltzman, E., & Kelso, J. (1983). Skilled actions: A task dynamic approach. *Haskins Laboratories Status Report on Speech Research*, *SR-76*, 3–50.

Saltzman, E., & Kelso, J. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, *94*, 84–106.

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

Saltzman, E., Nam, H., Goldstein, L., & Byrd, D. (2006). The distinctions between state, parameter and graph dynamics in sensorimotor control and coordination. In M. L. Latash & F. Lestienne (Eds.), *Motor control and learning* (pp. 63–73). New York: Springer Publishing.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the 4th international conference on speech prosody (speech prosody 2008)*. Brazil: Campinas.

Schmidt, R., Carello, C., & Turvey, M. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology, Human Perception and Performance*, *16*(2), 227–247.

Schöner, G., Haken, H., & Kelso, J. (1986). A stochastic theory of phase transitions in human hand movement. *Biological Cybernetics*, *53*(44), 247–257.

Scott, S. (1993). *P-centers in speech: An acoustic analysis*. Unpublished Doctoral Dissertation, University College London.

Scott, S. (1998). The point of P-centres. *Psychological Research*, *61*, 4–11.

So, W. C., Kita, S., & Goldin-Meadow, S. (in press). Using the hands to keep track of who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*.

Sternad, D., Turvey, M., & Saltzman, E. (1999). Dynamics of 1:2 coordination: Sources of symmetry breaking. *Journal of Motor Behavior*, *31*(3), 224–235.

Treffner, P., & Peter, M. (2002). Intentional and attentional dynamics of speech-hand coordination. *Human Movement Science*, *21*, 641–697.

Treffner, P., & Turvey, M. (1995). Handedness and the asymmetric dynamics of bimanual rhythmic coordination. *Journal of Experimental Psychology, Human Perception and Performance*, *21*, 318–333.

Tuller, B., & Fowler, C. (1980). Some articulatory correlates of perceptual isochrony. *Perception and Psychophysics*, *27*, 277–283.

Van Lieshout, P. (2004). Dynamical systems theory and its application in speech. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech motor control in normal and disordered speech* (pp. 51–82). Oxford, England: Oxford University Press.

Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S. Keele (Eds.), *Handbook of perception and action, volume 2: Motor skills* (pp. 181–262). London: Harcourt Brace & Company.

Westbury, J., Lindstrom, M., & McClean, M. (2002). Tongues and lips without jaws: A comparison of methods for decoupling speech movements. *Journal of Speech, Language, and Hearing Research*, *45*(4), 651–662.

Winfree, A. T. (1980). *The geometry of biological time*. New York: Springer-Verlag.
Wing, A., & Kristofferson, A. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, *14*(1), 5–12.

Appendix

Numerical simulations were conducted using a fourth-order Runge-Kutta algorithm in Matlab. Each simulation ran for 100 s of model time, with 2^5 iterations per second. Initial phases of each oscillator were randomly taken from a uniform distribution on the interval $[0, 2\pi]$. Oscillator frequencies ω_i for the Phr, Ft, σ, C_1, C_2, V were $2\pi \times [1, 2, 4, 4, 4, 4]$ and $2\pi \times [1, 3, 6, 6, 6, 6]$. Phase and relative phase variables were wrapped on the unit circle; note that for clarity of conceptualization in Fig. 12 phases were depicted on the interval $[-\pi, \pi]$. All relative phase targets were 0, except for Φ_{C1C2}, which following Nam and Saltzman (2003) was set at $3\pi/2$, or equivalently, $-\pi/2$. Relative phases φ_{C1C2} were sampled from each foot-peak in each simulation after 10 s of model time had elapsed, which allows relative phases ample time to stabilize. The noise term η_i corresponds to the standard deviation of a Gaussian distribution that modulates oscillator frequency. Hence, the frequency terms in Eq. (1) can be rewritten $\omega_i\ (1 + \eta_i)$. For the high-noise condition $\eta = 2$ and for low-noise $\eta = 1$. These are both fairly high noise levels relative to the oscillator frequencies, but on average only produce low-amplitude relative phase perturbations. Coupling effects of consonant oscillators on the vowel oscillator were $e = 0.1$ and of the vowel on the consonants were $f = 0.2$. Coupling between consonants (d) was $d = 0.25$ (see Table 1 for definition of variables). The change in oscillator phase velocity due to each coupling term was taken simply as $2\pi \times$ the negative sine of the difference between generalized relative phase and target generalized relative phase, modulated by the coupling strength. It is common in more sophisticated models to use the inverse Jacobian to partition a change of generalized relative phase into changes of phase in component systems—although not strictly necessary, the Jacobian strategy is more theoretically attractive because it explicitly captures the geometry of the mapping between component-level changes of oscillator phase and task-level change of generalized relative phase.