Journal of Phonetics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Special Issue: Mechanisms of regulation in speech, eds. Mücke, Hermes & Cho

Exertive modulation of speech and articulatory phasing

Sam Tilsen

Dept of Linguistics, 203 Morrill Hall, Cornell University, Ithaca, NY 14853, United States

ARTICLE INFO

Article history:
Received 1 April 2016
Received in revised form 17 February 2017
Accepted 1 March 2017
Available online xxxx

Keywords:
Motor control
Articulation
Articulatory timing
Coordination
Coupled oscillators
Attention

ABSTRACT

An articulatory study was conducted to investigate whether fluctuations in exertive mechanisms (attention, effort, motivation, arousal, etc.) have a global effect on articulatory control systems. Participants in the experiment produced an articulatory pattern 400 times, attempting to do so as consistently as possible. Evidence for global exertive modulation was obtained in the form of widespread correlations between variables associated with biomechanically independent systems such as phonation, linguo-labial coordination, and head movement/posture. Analyses of movement timing autocorrelation showed evidence for random walk-like dynamics on short timescales and equilibrium dynamics on long timescales, along with evidence for low- and high-exertion states of production. An extension of the coupled oscillators model of articulatory coordination is presented to account for these phenomena.

1. Introduction

Speech involves a variety of systems which interact on a range of spatial and temporal scales. Most of the relevant systems—i.e. ensembles of neurons, individual speakers, social networks—are typically far from equilibrium. In other words, non-equilibrium states are the norm. Nonetheless, the interactions between non-equilibrium systems may produce emergent patterns which exhibit equilibrium-like behaviors. Whenever we observe hints of such patterns, it is important to study the relevant systems in detail, because they greatly constrain our models and provide insight into underlying control mechanisms.

This study focuses on one such pattern, the symmetric temporal displacement observed in the initiation of consonantal and vocalic articulatory movements. Unconditioned, spontaneous fluctuations in this timing pattern were collected in an experimental paradigm in which participants produced a single target response 400 times, attempting to do so as consistently as possible.

Two hypotheses were tested: an *exertive modulation hypothesis*, which holds that exertive mechanisms (e.g. attention, effort, arousal) induce random walk-like dynamics in a variety of independent speech motor control systems, and an *equilibration hypothesis*, which holds that equilibration mechanisms constrain articulatory control systems on long timescales, in effect confining the exertive random walk in a potential. The exertive modulation hypothesis was supported by positive lag-1 autocorrelations of response variables on short timescales, along with pervasive correlations between outputs of biomechanically independent motor systems. The equilibration hypothesis was supported for some participants by decreases in autocorrelation on longer timescales. It was also observed that variation in a model-derived proxy for exertive force was associated with differing profiles of variance and covariance in articulatory timing, suggestive of a contrast between high- and low-exertion regimes. These findings are important because they provide a new basis for understanding the mechanisms involved in speech production.

1.1. Equilibrium vs. random walk behavior of systems

The concept of equilibrium arises in many domains. For example a mechanical equilibrium refers to a situation in which the net force on an object is zero, a chemical equilibrium describes an equivalence of forward and reverse chemical reaction rates, and a population equilibrium describes a biological system with stable predator and prey populations. In all of these cases, if the equilibrium is stable, there is a cost (in energy or in some other quantity) for deviations from the equilibrium, with greater deviations being more costly. When

E-mail address: tilsen@cornell.edu

http://dx.doi.org/10.1016/j.wocn.2017.03.001
0095-4470/© 2017 Elsevier Ltd. All rights reserved.

fluctuations from the environment perturb a system from its equilibrium state, the system subsequently returns to the equilibrium.

In contrast, a system with random walk behavior does not have an equilibrium or steady state. Generically, in a random walk the state of the system changes at each time step with a random displacement from the previous state. In the absence of any external forces, there is no cost for changing states, and hence if we wait long enough, we can expect the system to be arbitrarily far from its starting point. Clearly the systems which are responsible for controlling speech production cannot be governed solely by random walk dynamics, but it does not follow that there are no random walk-like components to their behavior. One possible scenario is that speech control systems have equilibrium states, but fluctuations in the nervous system add a random walk-like component to system dynamics. Evidence for this scenario can only be found if the responses of systems to departures from equilibria are relatively slow. Hence to investigate these possibilities, we must examine speech patterns over an extended period of time.

A useful statistical approach to investigating these ideas is to estimate the autocorrelation functions of system outputs. A Gaussian white noise process exhibits zero autocorrelation at all lags, because any state of the system is independent from all previous states. This is a generic property of a system which quickly returns to its equilibrium when perturbed—as long as the return to equilibrium is faster than the measurement period, successive observations will be uncorrelated. In contrast, a random walk tends toward a lag-1 autocorrelation of one, because each state depends on the previous one. No equilibrium is present in a random walk. (See Box, Jenkins, Reinsel, & Ljung, 2015; Chatfield, 2016; Sethna, 2006 for introductions to time series analysis and random processes.)

Many systems of interest may have more complicated structure, such as a random walk in an external field or a random walk with an external noise source. To assess whether observations might be generated by such a system, it is useful to conduct analyses over a range of timescales by applying a coarse-graining procedure to the observation sequence. The coarse-graining used here involves averaging observations over non-overlapping windows of time, thus integrating out short-timescale fluctuations. The lag-1 autocorrelation *scaling function* shows how the lag-1 autocorrelation changes as a function of the size of the averaging window. Fig. 1 (right) shows mean and ±1 standard deviation for autocorrelation scaling functions of several different types of processes. Over all coarse-grain timescales (τ) the Gaussian noise process tends toward a lag-1 autocorrelation of zero, while the random walk tends toward a lag-1 autocorrelation of one (see Appendix A.1 for details).

Unlike Gaussian noise and a random walk, more complicated systems exhibit lag-1 autocorrelations that vary substantially with analysis timescale. For example, a random walk with an external Gaussian noise (e.g. measurement noise) converges to random walk-like autocorrelation when coarse-grained, but is less than one on short timescales. Alternatively, a random walk in a quadratic potential has a random walk-like autocorrelation on short timescales but eventually converges to Gaussian noise-like equilibrium dynamics on longer scales. Combining a random walk, external Gaussian noise, and a quadratic potential results in an autocorrelation scaling function which increases at short timescales, peaks at some intermediate scale, and decreases on longer timescales. We will see this same profile in the autocorrelation scaling functions of articulatory timing measures, which suggests that models of articulatory control require mechanisms for both random walk- and equilibrium-like dynamics.

1.2. Symmetric displacement in articulatory timing and the coupled oscillators model

In many languages a pattern of articulatory timing called the *C–center effect* is observed (Browman & Goldstein, 1988; Hermes, Mücke, & Grice, 2013; Marin & Pouplier, 2010; Tilsen et al., 2012), which involves the symmetric displacement of the initiations of consonantal movements from the initiation of a vocalic movement in syllables with complex onsets. As schematized in Fig. 2 (left), the initiations of C_1 and C_2 constriction gestures in a C_1C_2V syllable are equally displaced in

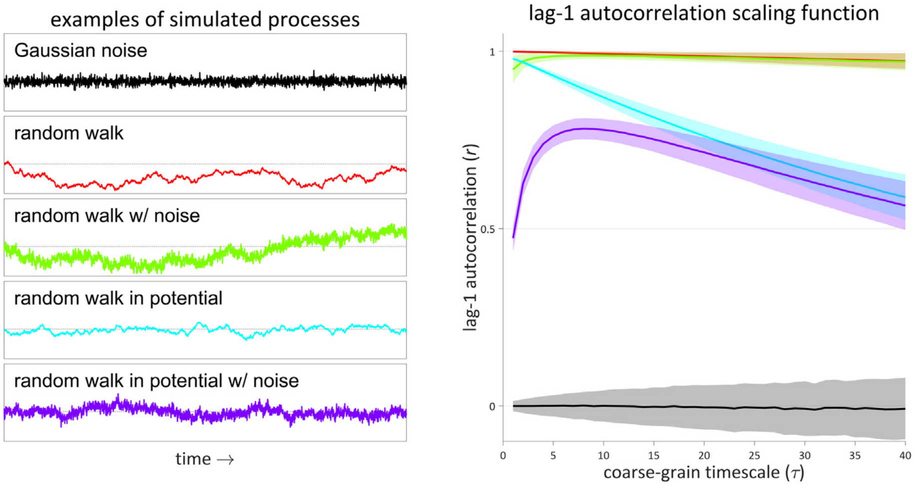

Fig. 1. Comparison of autocorrelation scaling functions of several random processes. (Left) example time series. (Right) Lag-1 autocorrelation scaling functions. Filled areas are ±1 s.d. of autocorrelation functions from 1000 simulations of 5000-sample observation sequences.

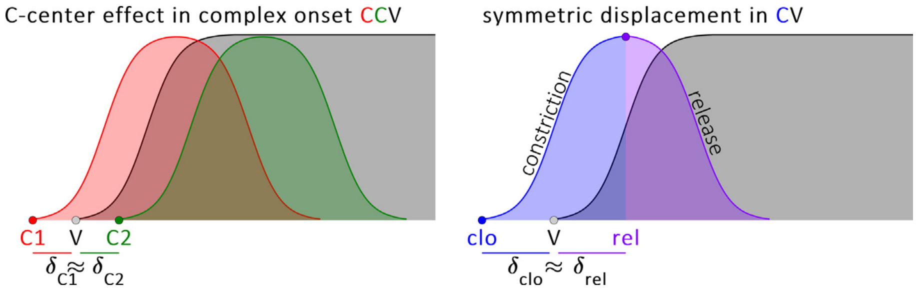
C-center effect in complex onset CCV
C1
V
C2
δC1 ≈ δC2
symmetric displacement in CV
constriction
release
clo
V
rel
δclo ≈ δrel

Fig. 2. Schematic illustrations of symmetric displacement patterns. (Left) CCV syllable C-center effect: initiations of consonantal gestures are displaced symmetrically from the initiation of the vocalic gesture. (Right) Symmetric displacement of consonantal closure and release gestures in a CV syllable.

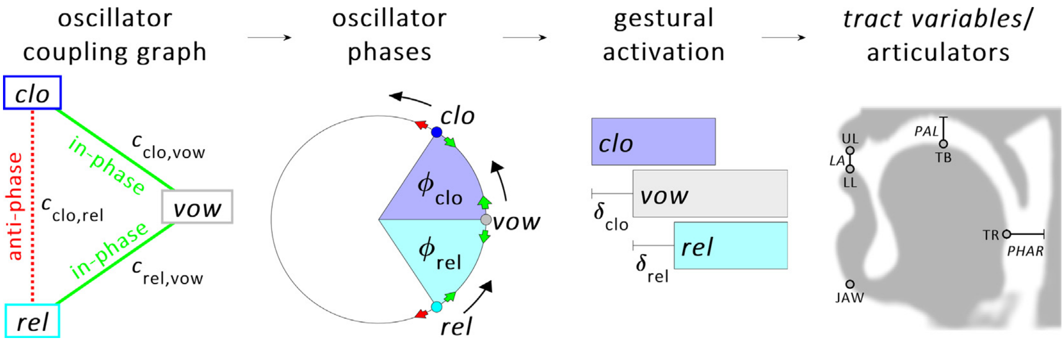

Fig. 3. Overview of the coupled oscillators model of articulatory phonology. A coupling graph specifies phase coupling with coupling strengths *c*. Equal in-phase coupling forces result in a stable relative phase configuration where $\phi_{\text{clo}} = \phi_{\text{rel}}$. The relative phases determine the relative timing of the activation of gestures that drive articulator synergies to achieve vocal tract targets.

opposite directions in time from the initiation of the vocalic gesture. This pattern provides evidence for the coupled oscillators extension to the theory of articulatory phonology (Browman & Goldstein, 2000; Saltzman & Munhall, 1989; Saltzman, Nam, Krivokapic, & Goldstein, 2008), and plays a role in accounts of cross-linguistic and developmental phonological patterns (Gafos & Goldstein, 2012; Goldstein, Byrd, & Saltzman, 2006; Shaw, Gafos, Hoole, & Zeroual, 2011; Tilsen, 2016).

The focus of the current study is the symmetric displacement pattern in simplex CV syllables, shown in Fig. 2 (right). The pattern applies to the initiations of consonantal constriction and release gestures relative to the initiation of the vocalic gesture. Importantly, the symmetry of the timing pattern is approximate and statistical in nature: any particular production is likely to deviate from symmetric displacement to some degree.

From a theoretical perspective (Goldstein et al., 2006; Nam, 2007; Nam & Saltzman, 2003; Tilsen, 2013), the timing pattern has been derived from equivalence of phase-coupling forces in a system of oscillators, such as those depicted for a CV syllable in Fig. 3. Here *clo*, *rel*, and *vow* refer to planning oscillators associated with the consonantal closure, consonantal release, and vocalic gesture. An anti-phase coupling force between *clo* and *rel* opposes in-phase coupling forces between *clo-vow* and *rel-vow*. When the strengths of $c_{\text{clo,vow}}$ and $c_{\text{rel,vow}}$ are equal, the stabilized relative phases of the oscillators, ϕ_{clo} and ϕ_{rel}, will also be equal, and hence the initiations of gestural activation intervals will be equally displaced in opposite directions in time from the vocalic initiation, resulting in the symmetric timing pattern $\delta_{\text{clo}} = \delta_{\text{rel}}$ (see Appendix A2 for model equations and further detail).

It is useful to estimate from empirical data both the common frequency ω of the oscillators and their relative phases ϕ_{clo} and ϕ_{rel}, because the frequency ω is hypothesized to reflect variation in exertive forces, and because the phases ϕ provide a frequency-normalized measure of deviation from symmetric displacement. However, ω and ϕ cannot be jointly estimated from observed timing intervals δ without imposing an additional constraint, since there are three unknown parameters $\{\phi_{\text{clo}}, \phi_{\text{rel}}, \omega\}$ but only two empirical measures $\{\delta_{\text{clo}}, \delta_{\text{rel}}\}$, an underdetermined problem. To address this problem we impose a *uniform coupling constraint*, which represents the working assumption that to a first approximation, in-phase and anti-phase coupling strengths are equal. Hence we can estimate ω and ϕ as in Eq. (1). Under the uniform coupling ansatz, the stable equilibrium relative phases are $\phi^* = \pi/3$. We can define an order parameter Φ that represents deviation from symmetric phasing (see Appendix A.2 for derivations and further detail).

$$\hat{\omega} = \frac{1}{3(\delta_{clo} + \delta_{rel})} \quad \hat{\phi} = 2\pi\hat{\omega}\delta \quad \Phi = \hat{\phi} - \frac{\pi}{3} \tag{1}$$

1.3. Exertive mechanisms and correlation of independent systems

As you read this paper, the actions of various autonomic systems (regulating respiration, heart rate, glucose levels, etc.) are changing over time. You may become drowsy, or perhaps excited, or switch more or less rapidly between the two. In all cases, what is certain is that the dynamics of exertive mechanisms—i.e. attention, focus, arousal, motivation, effort, etc.—are not entirely stationary. The same holds for a partici-

Please cite this article in press as: Tilsen, S. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.03.001

pant in an experiment, and so the fluctuations of systems which regulate exertion are of potential interest—especially if they alter the behavior of systems which control speech production.

Exertive systems have been studied in a wide variety of contexts. A recent theoretical approach of particular relevance holds that more stable coordinative patterns involve not only lower metabolic energy costs but also less activity in the central nervous system (Lay, Sparrow, & O'Dwyer, 2005; Temprado, Zanone, Monno, & Laurent, 1999; Zanone, Monno, Temprado, & Laurent, 2001). Indeed, Zanone et al. (2001) hypothesized that the same potential functions that describe the stability of movement coordination may describe levels of nervous system activity. Exertive mechanisms have also been theorized to play a critical role in modulating skill learning and memory consolidation (Carpenter & Grossberg, 1987). A likely consequence of fluctuations in exertive systems is chaotic variation in movement timing; such variation has been demonstrated in a number of studies: for example, inter-stride intervals exhibit correlations on multiple timescales (Hausdorff et al., 1996). It has been suggested that such variability is important for behavioral adaptability (Stergiou & Decker, 2011; Stergiou, Harbourne, & Cavanaugh, 2006) and may be associated with shifts from relatively more controlled to relatively more automatic processing (Paus et al., 1997).

The reader should keep in mind that the concepts of *exertive mechanisms* and *exertive force* are used here as heuristic devices. These concepts facilitate analyses of the effects of complex and not-well-defined cognitive systems associated with attention, effort, arousal, motivation, focus, etc., which presumably influence motor control in numerous ways. The heuristic simplification allows us to conflate the effects of exertive mechanisms into a single variable, conceptualized as a force acting on parameters of speech motor control. The estimated frequency ω of the planning oscillators is hypothesized to be a correlate of exertion. This follows from a microscopic *task ensembles* model in which planning oscillators are viewed as macroscopic descriptions of the coordinated spiking of neurons in premotor ensembles associated with speech tasks (see Section 4.1 for further detail).

Variation in exertive force should induce correlations between speech motor systems. However, correlations can also arise when there is a physical interaction between the systems which mediate control of outputs, such as with F0 and intensity, which are controlled by the same anatomical structures and which interact aeroacoustically (Tilsen, 2016). To conclusively test for the presence of exertive forces, only correlations between physically independent systems should be considered. It should be emphasized that although we hypothesize correlations between system *states*, the predictions relate to the observed *outputs* of those systems. Table 1 provides an overview of several broad categories of systems relevant to the current study, along with their degree of mechanical independence.

The outputs of phonation, head movement, and head posture control have negligible interactions with linguo-labial tasks and bilabial subtasks (cf. Section 2.4 for descriptions of variables in these categories). Phonatory correlations with vowel height/tongue posture have been observed in some studies (Ohala & Eukel, 1987; Whalen & Gick, 2001), yet these exhibit speaker- and language-specific variations in effect directions and are present only between high vs. low vowel categories; within a vowel height category such effects have not been reported. Articulatory tasks have been shown to interact with body posture (i.e. supine vs. upright) (Stone et al., 2007; Tiede, Masaki, & Vatikiotis-Bateson, 2000), but head posture variation in an upright position has not been shown to influence articulation. Head movements can be coordinated with speech movements in some contexts (Goldenberg, Tiede, Honorof, & Mooshammer, 2014; Tiede & Goldenberg, 2015), and can accompany pitch excursions (Ishi, Ishiguro, & Hagita, 2014; Krivokapić, 2014), but no studies have examined whether head movement has inertial consequences for the jaw or lips. Because of their relatively small masses, inertial forces on articulators are likely negligible compared to forces generated by muscle contraction. Conversely, the relative massiveness of the head renders head posture and head movement immune to inertial effects from articulators.

In sum, many of the outputs of systems measured in this experiment can be presumed to be nearly independent: the effects of their interactions are small relative to other sources of variance, and hence correlations that are observed between them implicate a non-mechanical source, such as an exertive force.

1.4. Hypotheses

There are two primary hypotheses tested in this study: the exertive modulation hypothesis and the equilibration hypothesis. These hypotheses are not mutually exclusive and make a variety of predictions detailed below. Some of these predictions are best viewed in relation to the autocorrelation scaling functions shown in Fig. 1. It is assumed that observations are subject to external Gaussian noise, which results from measurement procedures and/or non-exertive sources of noise in the nervous system.

Exertive modulation hypothesis: speech motor control systems are modulated globally by exertive forces. Exertive mechanisms are assumed to change slowly relative to the observation scale, and hence should induce random walk-like correlations between past and future states in independent systems. Analyses of autocorrelation scaling functions with

Table 1
Biomechanical interaction of systems.

	Linguo-labial tasks	Phonation (F0, I)	Head movement	Head posture
Phonation (F0, I)	Ø			
Head movement	Ø	?		
Head posture	Ø	?	?	
Bilabial subtasks	+	Ø	Ø	Ø

[Ø]: negligible interaction. [+]: strong interaction. [?]: unknown interaction.

Please cite this article in press as: Tilsen, S. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.03.001

regard to this hypothesis focus on asymmetry in movement phasing (i.e. Φ, cf. Section 1.2), because this most directly represents the interaction of planning oscillators theorized to be responsible for symmetric displacement. The following predictions are made:

(a) *Lag-1 autocorrelations of response parameters on the observation scale will have non-zero positive values. Lag-1 autocorrelation of the phase asymmetry order parameter Φ will increase over short analysis scales*. The increase is predicted because the coarse-graining procedure integrates out Gaussian fluctuations more effectively as the coarse-graining timescale is increased (see Fig. 1).

(b) *Correlations between outputs of mechanically independent systems will be observed.* Because exertive forces are global—i.e. they modulate all motor control systems—we predict correlations between the outputs of independent systems.

Equilibration hypothesis: mechanisms governing phase-coupling interactions between oscillators promote equilibrium-like behavior on long timescales, and variation in exertive force is reflected in the estimated planning oscillator frequency ω. This follows from a *task ensembles* interpretation of planning oscillators elaborated in Section 4.1. Fluctuations in exertive force are viewed as perturbations of planning systems. The following predictions are made:

(a) *Lag-1 autocorrelations will decrease on relatively long analysis scales*. This follows from viewing the long-timescale equilibration mechanism as a quadratic potential in which a random walk with Gaussian noise occurs (see Fig. 1).

(b) *Planning oscillator frequency ω will correlate with variability and covariability in articulatory timing*. This follows from viewing ω as a proxy for exertive force. A stochastic model of this phenomenon is presented in Section 4.1.

2. Method

Rather than examining *conditioned* variation, i.e. variation induced by experimentally manipulated factors, the current study investigates *unconditioned* variation, i.e. variation that emerges quasi-spontaneously. Speakers were encouraged by a variety of design features to produce a single target form (*ee-PAH*) exactly the same way throughout an entire session. The promotion of consistency and absence of conditioning manipulations serve the goals of testing the exertive modulation and equilibration hypotheses: long, uninterrupted observation sequences are desirable for investigating autocorrelation and system dynamics.

2.1. Participants and task

Data from six participants/sessions are presented here. All but one of the participants were native speakers of English, and none had any speech or hearing disorders. The author participated in one of the sessions. Note that four additional sessions with alternative designs were conducted, but are not reported. Participants in the study were told that they would say the nonword “ee-PAH”, [i'p^{h}a:], with a weak-strong stress pattern, on every trial of the experiment. They were to instructed to try to do this exactly the same way every time. In each session, 400 productions were collected.

Sessions were conducted in a quiet room in the Cornell Phonetics Laboratory. Participants were seated approximately 1.5 m from a computer monitor on which a response cue and occasional non-specific performance feedback were delivered. An electromagnetic articulograph (EMA) collected articulatory data (NDI Wave System, Berry 2011). Articulator sensors were adhered midsagittally on the lower and upper lip (LL, UL), on the lower incisors to capture jaw movement (JAW), and on the tongue body (TB) approximately 6–7 cm from the tongue tip. Reference sensors were positioned on the nasion (NAS) and left and right mastoid processes (MPR, MPL). A shotgun microphone was positioned about 1.25 m from the participant. Experimental sessions consisted of a setup phase (approximately 25 min.) and a data collection phase (approximately 1 h).

Each trial began with a response cue, the appearance of a green box on the stimulus monitor (85% of screen width and height). The cue remained on the screen for 2500 ms. Participants were told to produce the response when the cue appears, but *not* to try to respond as quickly as possible, i.e. not as an immediate reaction. They were also told to be sure to respond before the cue disappeared. A uniformly distributed random intertrial interval of 1500–7500 ms occurred before the next response cue. The randomization of the intertrial interval, along with its relatively large range, served the purpose of discouraging list-reading and rhythmic effects on the production of the response. Every 10 trials the participant received non-specific consistency feedback (see below), and after every block of 50 trials there was a 15–30 s break during which the experimenter told the participant to adjust their posture, if necessary. Note that participants often adjusted their posture between blocks or between trials without explicit instruction.

2.2. Response design

The target response [i'p^{h}a:] was designed to provide robust measures of the relative timing of coordinated movements, to diminish effects of confounding biomechanical interactions, and to preclude effects of anticipatory posturing on the measures of interest. The response contains four oral articulatory tasks, three of which are precisely coordinated: bilabial closure, pharyngeal constriction for [a], and bilabial release. These are preceded by a palatal constriction for [i] made with the tongue body. A disyllable rather than monosyllable was chosen because the initial vowel prevents pre-response anticipatory posturing from influencing the movements of interest (Tilsen et al., 2016).

The vowels [i] and [a] were chosen because they promote a relatively large magnitude high/front to low/back movement of the tongue body, which facilitates automated processing and provides relatively more space for variability in movement range and speed. A bilabial consonant was chosen because it does not specify a lingual movement, and therefore minimizes biomechanical interactions with the vowels. The bilabial release movement is robustly identifiable because the lips are opened quickly and widely in order to support the production of the vowel [a]. The trade-off for having an [i]–[a] transition is that the bilabial closure movement is smaller/lower in magnitude/velocity because the jaw is already raised for the [i]. This clo-

Table 2
Categorization and description of response variables.

Head movement		
Translational and rotational avg. speed	$\mathrm{HdMov}_{trans,rot}$	(mm/t), (rad/t): average magnitude of translation/rotation per second, calculated in a 500 ms peri-response window
Translational movement speed maximum	HdMov_{spd}	(mm/t): maximum RMS velocity of translational head movement in the peri-response window
Head posture		
Head posture and orientation	$\mathrm{HdPos}_{pos,ang}$	(mm), (rad): mean value of midsagittal angle and first principal component of head position
Intergestural timing		
Movement intervals	$\delta_{clo,rel}$	(ms) time intervals between the initiations of bilabial closure, vocalic movement, and bilabial release
Movement speed		
Maximal speed	$\mathrm{MovSpd}_{clo,vow,rel}$	(mm/s) maximum midsagittal RMS velocity of bilabial closure, vocalic, and bilabial release movements
Movement targets		
Target position	$\mathrm{MovTrgX}_{i,a}$ $\mathrm{MovTrgY}_{i,a}$	(mm) vertical and horizontal positions of TB at [a] and [i] target achievements
Bilabial subtasks		
Articulator contrib. to bilabial task	$\mathrm{AcClo}_{UL,LL,JAW}$ $\mathrm{AcRel}_{UL,LL,JAW}$	(proportion) contributions of vertical movement of UL, LL, and JAW to LA, in closure and release phases
Acoustic intensity		
Intensity	$\mathrm{I}_{,a,resp}$	(dB): RMS intensity, on a decibel scale, calculated for [i], [a], and the whole response
F0 and spectral tilt		
Fundamental frequency	$\mathrm{F0}_{avg,rng}$	(Hz): average and range of F0 over a 50 ms interval centered on maximum acoustic energy in [a]
Spectral tilt	$\mathrm{H12}_{avg,rng}$	Spectral tilt (dB): average and range of H1–H2 over a 50 ms interval centered on maximum acoustic energy in [a]

sure movement is nonetheless robustly identifiable in kinematic data.

2.3. Non-specific consistency feedback

Participants were periodically given non-specific feedback regarding the consistency of their responses. Every 10 trials the participant received a consistency score on a scale from 0 to 100, based on articulatory similarity of responses over the last 10 trials. The score was shown on the stimulus monitor for 2.5 s. Participants were given no specific information about the feedback score other than being told that it ranged from 0 to 100, represented consistency in recent responses, and that scores over 50 were good.

Feedback was designed to be uninformative (i.e. non-specific) with regard to the composition of responses, and to serve as motivation rather than to shape response patterns directly. In a previous study (Tilsen, 2015) presentation of feedback after every trial may have led participants to make arbitrary, unpredictable associations between their responses and feedback scores. Here feedback was provided every 10 trials in order to preclude arbitrary associations between scores and responses.

Consistency scores were derived from the average pairwise response distance over the preceding 10 responses, calculated as follows. The lag of maximal cross-correlation of articulator sensor position data (corrected for head movement) from a pair of trials was used for alignment. Then the Euclidean distance between sensor positions was computed over a 1000 ms window centered on the time of maximal speed in the movement of the tongue body from an [i] to [a] posture. This distance metric was averaged over all pairwise combinations of the past 10 trials. Consistency scores were obtained by calculating the z-score of the most recent average distance relative to all previous ones; the score was defined as the inverse normal cumulative distribution function of the z-score, rescaled from 0 to 100. The consistency score is thus a time-varying, participant-relative index of performance, rather than an absolute metric of consistency.

2.4. Data processing procedures

Measures analyzed in this study are summarized in Table 2. Sensor position data from the NDI Wave EMA (100 Hz) were smoothed with a 4th order Butterworth low-pass filter. A 5 Hz cutoff was used for reference sensors (NAS, MPR, MPL) and a 10 Hz cutoff for articulator sensors (UL, LL, JAW, TB). The positions of the articulator sensors were corrected for head movement. Acoustic data were collected at 22,050 Hz with a shotgun microphone. To facilitate automated data processing, responses were acoustically segmented with forced alignment using the Hidden Markov Model Toolkit (HTK) (Young et al., 1997). For each session, responses from 10 trials, selected randomly for 10 equal-size epochs from a session, were manually labeled for the purpose of training HMMs. From these training data 5-state HMMs of MFCC vectors (16 coefficients, 20 ms window) were estimated for each segment ([i], [p], [a]). These HMMs were subsequently used for forced alignment of the remaining trials.

Three metrics of head movement in the vicinity of the response were derived from the reference sensors. Head position was defined as the location in 3-dimensional transmitter coordinates of the centroid of the three reference sensors (NAS, MPR, MPL). Head orientation was defined as the angle of the MPR-MPL midpoint-to-NAS line on the triangular surface of the reference basis, in relation to the vertical plane of the transmitter (i.e. pitch). This is approximately the head angle in the mid-sagittal plane of the participant. Rotational movements in coronal and axial planes (i.e. roll and yaw) were not analyzed because these tend to be small. Translational and rotational movements of the head was measured in a peri-response period as the average change in position/angle per second. The peri-response period was a 500 ms window of time centered on the point of maximum velocity of the lower lip in the release. By restricting the estimation of head movement to this period of time, the measurements index head movement associated with response production, rather than head movement associated with sporadic postural adjustments occurring between responses. A maximum speed measure was estimated as the maximum translational RMS speed of the head in the peri-response period.

Two head posture measures were calculated for each response, one indexing head position, the other head angle. Head position principal components were calculated from the peri-response head centroid position data from each session. The mean value of the first principal component over the peri-response window was used to estimate head position; the mean value of the midsagittal angle in this period was used to estimate head orientation.

Two intergestural timing measures, δ_{clo} and δ_{rel}, were obtained by estimating movement initiation times (t_{clo}, t_{vow}, t_{rel}) from each trial. To facilitate automatic identification of these, articulator RMS velocity time-series were aligned iteratively using maximum cross-correlation with neighboring trials as a criterion. Closure, vowel, and release movement initiations were estimated as the times of maximal absolute acceleration in sigmoidal fits of the first principal components of LL and TB. Principal components were estimated with a 200 ms window centered on the velocity extrema associated with labial closure, release, and tongue body retraction/lowering. Examples of the first principal components of sensors and corresponding acceleration extrema are shown in Fig. 4. Target [i] and [a] positions were defined as the horizontal and vertical positions of the TB sensor when TB RMS velocity was at a minimum preceding/following the movement. Model quantities were estimated from δ_{clo} and δ_{rel} as described in Appendix A.2.

An articulator contribution index (ACI) was defined to characterize bilabial subtasks, i.e. the relative contributions of UL, LL, and JAW to the vertical lip aperture (LA) tasks of closure and release. Closure and release periods were identified in vertical LA acceleration extrema as the period from movement onset to target. Within each period, the ACI of an articulator was defined as the ratio of change in its vertical position, from start to end of the period, to the total change in LA (see Fig. 5). For LL the vertical component of JAW position was subtracted. Thus the UL, LL, and JAW ACIs always sum to one.

Acoustic intensity measures associated with [i], [a], and the whole response were extracted from each trial, by calculating RMS intensity in the following windows. For [i]: 100 ms before [i] voicing onset up to the [a] voicing onset. For [a]: [a] voicing onset until 350 ms later. For the whole response: from the start of the [i] window to the end of the [a] window. The robust automated pitch tracking algorithm (RAPT) implemented in the Matlab Voicebox toolbox (Brookes, 1997) was used to obtain estimates of the rate of vocal fold vibration during the segment [a]. This algorithm uses normalized cross-correlation (Talkin, 1995); the correlation window size was 7.5/15 ms for female/male participants. To avoid bias from microprosodic artifacts associated with voicing onset or offset, F0 averages and ranges were derived from a 50 ms window centered on the point of maximal acoustic energy in [a]. Spectral tilt averages and ranges were obtained from the same window as follows. The signal was split into 30 ms frames in 1 ms steps and a 4096-point power spectrum was estimated for each frame. Amplitudes of the first two harmonics (H1, H2) were extracted using the F0 estimate to constrain the identification of harmonic peaks.

2.5. Data analysis procedures

Section 3.1 examines lag-1 autocorrelations (r_1) in observation-scale time series of response variables. The maximal timescale (τ_{acmax}) of non-stationary features in each time series was measured by determining the trial lag at which a spline fit of the autocorrelation function fell below the upper limit of the 95% confidence interval for an uncorrelated white noise process, which is defined as $\sqrt{2}\mathrm{erf}^{-1}(0.95)(1/\sqrt{N})$, where N is the length of the time series and erf^{-1} is the inverse Gaussian error function. Correlations and autocorrelations reported in Sections 3.2 and 3.3 were obtained using a coarse-graining procedure in which observations were averaged within adjacent, non-overlapping observation windows.

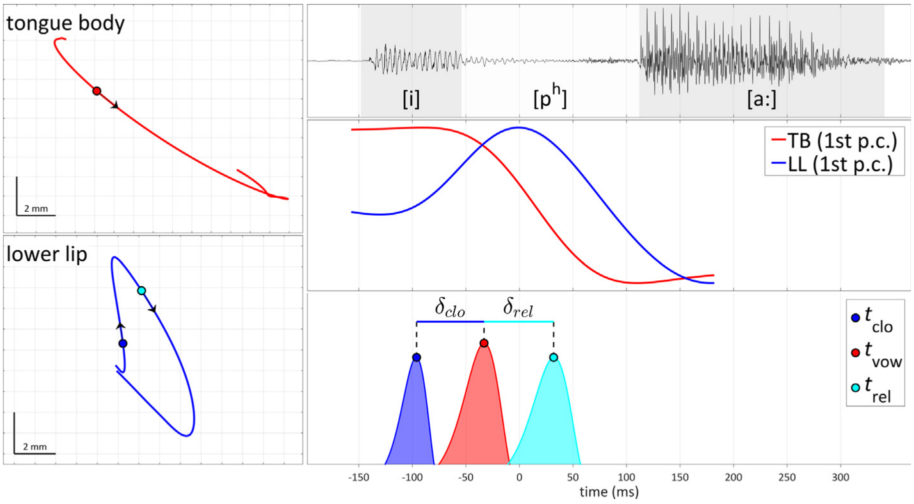

Fig. 4. Example production of *ee-PAH*. (Left): tongue body and lower lip trajectories in midsagittal plane. (Right, top to bottom): acoustic waveform; 1st principal components of TB and LL sensors; acceleration extrema.

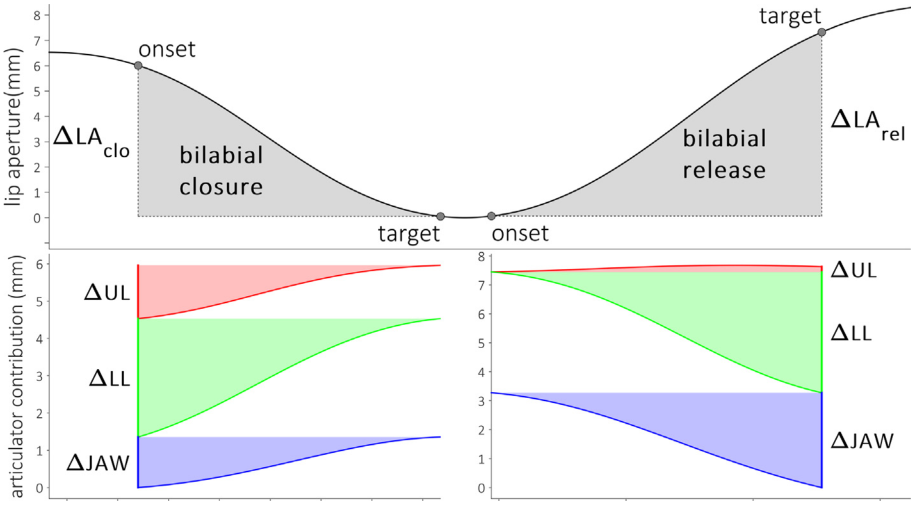

Fig. 5. Illustration of articulator contribution indices (ACIs). ACIs were defined for UL, LL, and JAW for both the closure and release gestures. The ACI is the ratio of change in the vertical position of an articulator (i.e. ΔUL, ΔLL, ΔJAW) to change in vertical lip aperture from gesture onset to target (ΔLA).

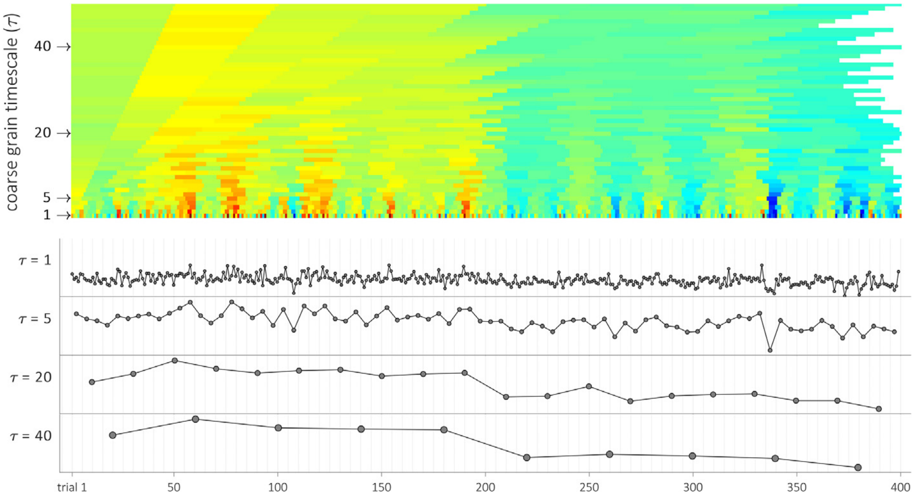

Fig. 6. Example of coarse-graining. Coarse-grained δ_{clo} on timescales of 1 (observation scale), 5, 20, and 40 are shown.

Fig. 6 illustrates the output of coarse-graining; note that the length of a coarse-grained time series is reduced by a factor of τ, i.e. the number of trials in the averaging window.

The correlation analyses conducted in Section 3.2 do not involve statistical inferences at the level of pairwise combinations of variables, but rather, treat correlations as samples from a population of correlations. A total of 27 variables, grouped into 8 categories (cf. Table 2), were included in correlation analyses. Correlations were calculated only for inter-category pairs. This resulted in the estimation of 313 unique correlations per session, and a total of 1878 correlations over the six sessions. Fig. 8 shows R^2 of correlations calculated from coarse-grained time-series with a timescale of $\tau = 20$, which was approximately half of the median autocorrelation timescale (τ_{acmax}). Qualitatively similar patterns of R^2 distributions were found for a fairly large range of τ, hence the analysis does not depend crucially on choice of scale. To obtain confidence limits for uncorrelated processes at the level of group-wise variable pairs, a Monte Carlo permutation procedure was used in which 95% confidence regions for R^2 were calculated from 1000 random permutations for each variable pair. The hierarchical clustering shown in Fig. 9 was obtained by using the mean between-group correlation strengths as a measure of similarity (the distance metric was 1-R^2, with R^2 averaged over sessions). Empirical autocorrelation scaling functions in Fig. 10 are smoothed with a window of $\tau = 3$ in order to emphasize the global shapes of the functions.

3. Results

3.1. Positive lag-1 autocorrelation of response variables

Many of the response variable time series obtained in this study exhibited positive lag-1 autocorrelations (r_1) on the observation timescale. This is consistent with the predictions of the exertive modulation hypothesis and implicates an exertive mechanism with a random walk-like influence on articulatory control systems.

Distributions of r_1 and τ_{acmax} from all sessions are shown in Fig. 7. About 92% (149/162) of all response variable time series had positive r_1 greater than chance levels (gray band, cf. Section 2.5). Fig. 7 also breaks down r_1 and τ_{acmax} by sessions

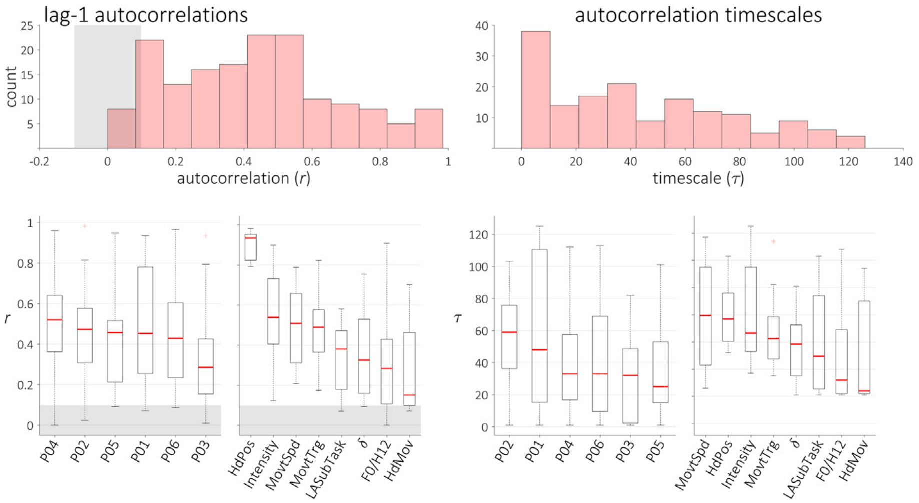

Fig. 7. Summary of lag-1 autocorrelations (r_1) and autocorrelation timescales (τ_{acmax}). (Top panels) Experiment-wide histograms of r_1 and τ_{acmax}. (Bottom panels) By-session and by-variable category distributions. Gray bands indicate 95% confidence intervals for the r_1 of stationary white noise.

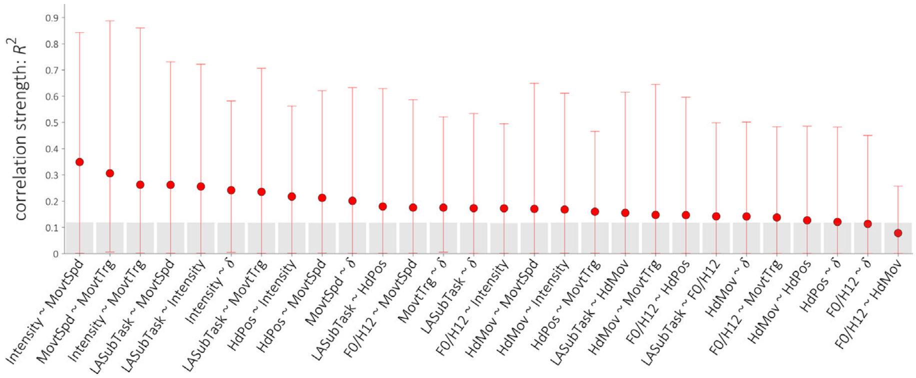

Fig. 8. Distributions of correlation strengths. The vertical axis represents R^2 from pairwise correlations of coarse-grained response variables (τ = 20). Red circles/bars show the median/5th–95th percentile range for R^2 from each combination of variable categories. Gray bands show 95% confidence regions for uncorrelated variables, estimated from random permutations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

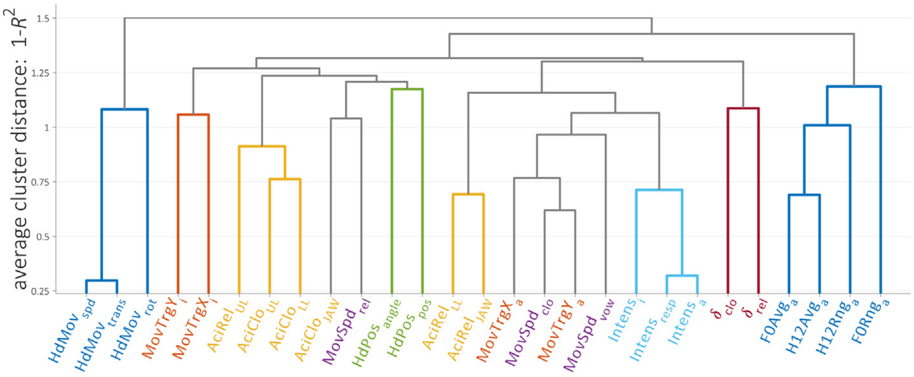

Fig. 9. Hierarchical clustering of response variables. Distance matrix derived from strengths of pairwise correlations of coarse-grained measures (τ = 20), with distance $1 - R^2$.

and variable categories. Within each session the median r_1 and even the 25th percentile r_1 were well above chance values. All variable categories exhibited positive r_1. Head position had especially high r_1. In contrast, peri-response head movement r_1 were closer to chance levels. Median τ_{acmax} by session ranged from 25 to 59 trials; the median across all variables/sessions was 37 trials. Head movement also had the lowest median τ_{acmax}.

3.2. Correlations between response variables

A key prediction of the exertive modulation hypothesis was upheld: response variables associated with independent systems were correlated above chance. Fig. 8 shows median and 5th–95th percentile ranges of R^2 values (red lines) from correlations calculated by session and grouped by variable categories. For all category pairs there were above-chance

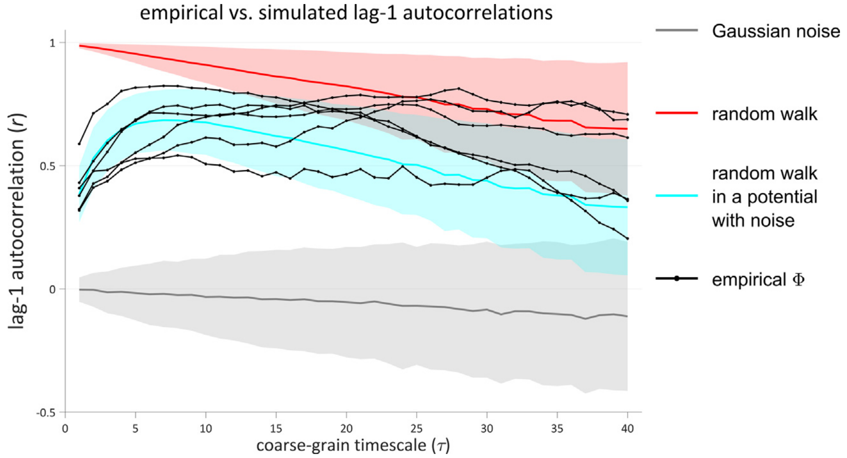

Fig. 10. Empirical vs. simulated lag-1 autocorrelations. Filled areas show ±1σ calculated from 10,000 random simulations.

correlations, and for all but one category pair the mean correlation strengths (red circles) were greater than expected by chance (gray band, estimated with a Monte Carlo procedure, cf. Section 2.5). Although not all category pairs in the analysis contain variables from mechanically independent systems, nearly all of the pairs that are independent (cf. Section 1.4) had a majority of correlations that were significant. In particular, the intensity measures and the F0 and H12 category of measures correlate with various categories of articulatory measures (MovtSpd, MovtTrg, δ, LASubTask).

Several interesting relations become evident when response variables are clustered hierarchically by their correlation strengths, as shown in Fig. 9. Cluster nodes formed lower down in the tree represent variables/sets of variables that are on average more strongly correlated. In general, response variables from the same category are grouped early on by the clustering algorithm, but there are some noteworthy patterns and exceptions.

Head movement variables and variables involving F0 and H12 are the least strongly correlated with other variables and hence are last to be merged into the tree. In contrast, head posture variables correlate more strongly with articulatory variables, specifically the speed of the release movement and the contribution of the jaw to the release. These findings are consistent with the disparity in lag-1 autocorrelation distributions between head movement and head posture variables, and suggest that exertive fluctuations manifest more strongly in head posture than in peri-response head movement. Intensity variables are more strongly correlated with a subset of vowel/-closure movement speed and [a] target variables than with other phonation-related variables involving F0 or H12. The contribution of the jaw to bilabial closure (AciClo$_{\text{JAW}}$) is more strongly correlated with the speed of the release movement than with other closure ACIs. Movement target variables for [i] are not strongly correlated with movement target variables for [a], the latter are instead correlated more strongly with vowel/closure movement speed.

3.3. Autocorrelation scaling functions

Empirical autocorrelation scaling functions are consistent with predictions of the exertive modulation hypothesis but provide mixed support for the equilibration hypothesis. The exertive modulation hypothesis predicted that intermediate positive r_1 in the order parameter Φ would increase on relatively short timescales. Indeed, as shown in Fig. 10, abrupt increases in r_1 at small analysis scales ($\tau_{\text{cg}} < 5$) were observed for all sessions. The cause of these abrupt increases is the smoothing effect of coarse graining on the Gaussian noise component of the process, which diminishes external noise and enhances the relative influence of the random walk component.

Three of the six sessions exhibited r_1 consistent with the equilibration hypothesis. This hypothesis predicted decreases of r_1 on long timescales, which would arise from long-term effects of forces promoting symmetric coupling. Three of the autocorrelation scaling functions do indeed have substantial decreases for $\tau > 20$; however the other three fail to exhibit this feature. Hence the data provide mixed support for the presence of forces promoting symmetric coupling.

There are two additional reasons why these results are not conclusive. First, finite sample length effects result in wide confidence intervals for autocorrelation scaling functions at long timescales, along with biases which reduce autocorrelation at long scales. Second, some aspects of the simulated scaling functions depend on parameters which are unknown (cf. Appendix A.1), and hence multiple models can produce fits of similar quality between empirical and simulated data. Further considerations in interpreting these data are discussed in Section 4.

3.4. Oscillator frequency and exertion regimes

Evidence for phase equilibration is less ambiguous with regard to the second prediction of the hypothesis, that oscillator frequency ω should correlate with variances and correlations of timing intervals δ. Indeed, there appear to be qualitatively different regimes of variation in δ_{clo} and δ_{rel}, which correlate with estimated ω. Fig. 11 shows heat maps of the joint distributions of δ_{clo} and δ_{rel}. In sessions P03, P05, and P06 there were moderate/weak negative correlations. In session P02 there was a weak positive correlation. These patterns are potentially misleading, however, because they may obscure transient fluctuations in correlation associated with exertive modulation.

Please cite this article in press as: Tilsen, S. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.03.001

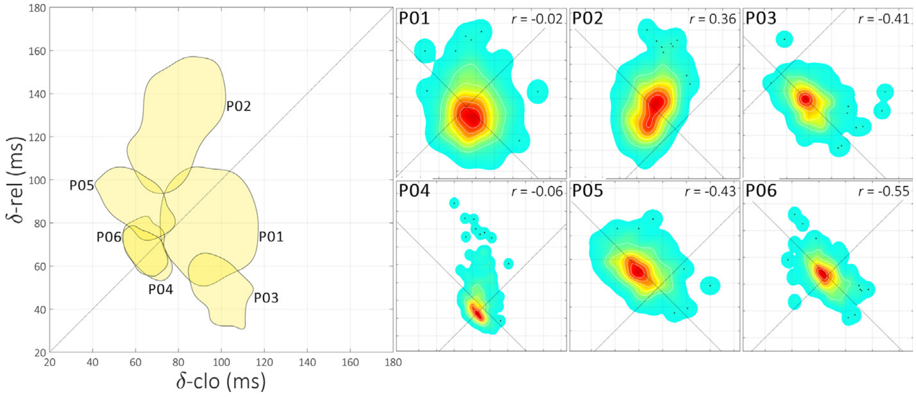

Fig. 11. Joint δ_{clo}, δ_{rel} distributions for each session. (Left): 50th percentile contours for all sessions. (Right): heat-maps of joint distributions by session. Diagonal lines represent positive and negative correlation.

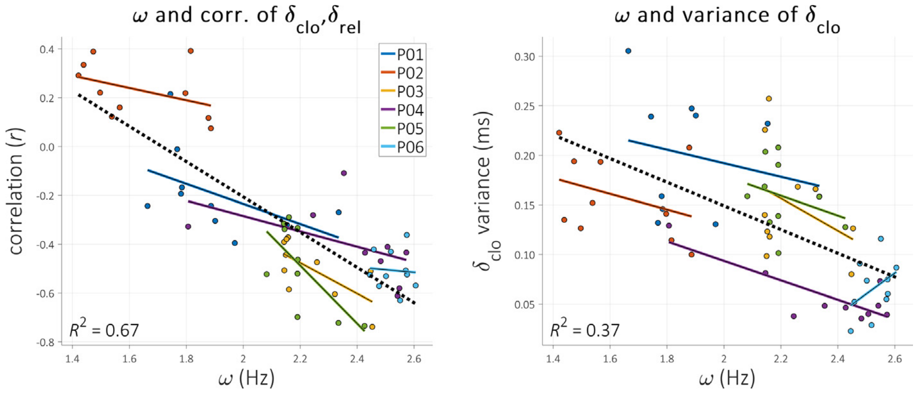

Fig. 12. Relations between ω and correlation and variance of timing intervals. (Left): relation between ω and the correlation of δ_{clo} and δ_{rel}. (Right): relation between ω and variance of δ_{clo}. Regression lines are shown for each session (colors) and across all sessions (dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The structure of variation in δ becomes more apparent when analyzed over time and in relation to the estimated frequency ω. Fig. 12 shows the relations between ω and the correlation of δ_{clo} and δ_{rel}, and between ω and the variance of δ_{clo}. The correlations and standard deviations were estimated in non-overlapping 40-trial windows. Oscillator frequency ω is a fairly good predictor of corr(δ) and δ variance: the ω-corr(δ) correlation had an R^2 of 0.67, the ω-variance(δ_{clo}) correlation R^2 was 0.37, and the ω-variance(δ_{rel}) correlation R^2 was 0.46.

Under the hypothesis that ω is associated with exertive force, the within- and between-speaker variation in Fig. 12 suggests that there were two qualitatively different exertion regimes. In P01 and P02, where ω was low, correlation and variances of δ were relatively high. In P03, P05, and P06, higher ω were associated with negative corr(δ) and low variance δ_{clo} and δ_{rel}. In two of the sessions, P01 and P04, regimes of both positive and negative correlation are observed. When examined temporally, it is evident that P01 began the session with moderate negative correlations, but after a while transitioned to a regime of weak/moderate positive correlation. P04 produced negative correlations through most of the session, but midway through entered into a transient epoch in which positive correlations were observed. In addition to the across-session/participant pattern, similar relations between ω and δ are observed within most of the sessions (cf. participant-specific regression lines in Fig. 12).

Overall, the ω–δ relations suggest a distinction between two regimes: a high-ω, high-exertion regime in which movements are tightly coordinated, and a low-ω, low-exertion regime with more variability. We consider this interpretation in further detail below and discuss the main results of the study.

4. Discussion

Support for the exertive modulation hypothesis was obtained in three forms: (i) positive lag-1 autocorrelation in response variables; (ii) an increasing autocorrelation scaling function for the phase asymmetry order parameter Φ on short timescales; and (iii) pervasive correlations between variables associated with independent motor systems. Mixed support for the equilibration hypothesis was observed: autocorrelation scaling functions for three of six sessions exhibited the predicted decrease in the autocorrelation at long timescales. The equilibration hypothesis was supported by correlations between estimated planning oscillator frequency ω and the variability and covariability of timing intervals $\delta_{\text{clo}}/\delta_{\text{rel}}$. Below we elaborate on the interpretation of these results, beginning with the presentation of a microscopic model of planning oscillators which elucidates the conceptual connection between frequency and exertion.

4.1. Exertion regimes in a task ensembles model of planning oscillators

To model exertion effects, we begin by associating each planning oscillator (closure, vowel, and release) with an ensemble of neurons, i.e. a *task ensemble*. The coupled oscillators model is based on the notion that there is an intrinsic timeframe or virtual cycle associated with planning an aperiodic movement (Fowler, 1980; Kelso & Tuller, 1987). In the microscopic task ensembles conception these cycles are interpreted as macroscopic (low-dimensional) approximations of the integrated spiking rates of large populations of neurons (cf. Fig. 13A). When excited by an external source (e.g. frontal systems which govern the intention to speak), the neurons in an ensemble exhibit a collective oscillation, which is possible because of reciprocal interactions with other systems and/or within-ensemble intrinsic dynamics associated with neuronal conduction delays (Izhikevich, 2006, 2007).

Prior to being in the excited, oscillatory state, the ensembles are near a critical point: a sufficient amount of external excitation induces a phase transition to the regime of collective oscillation. However, merely entering this excited state does not entail movement. A selection-thresholding process not directly modeled here determines whether the excited ensembles lead to movement. We will assume (1) that the three relevant task ensembles (*clo*, *vow*, *rel*) surpass the selection threshold during the same oscillatory cycle, and (2) that oscillator relative phases have stabilized prior to this. These assumptions are easy to motivate by consideration of the self-paced nature of the response along with empirical observations, and it follows from these assumptions that oscillator frequency and relative phases determine relative timing.

The coupling interactions between ensembles are associated with projections between ensembles. In-phase and anti-phase coupling modes derive from the balance of interensemble excitatory-to-excitatory and excitatory-to-inhibitory projections. In-phase coupling occurs when there are, in a pair of ensembles, a relatively large number of excitatory neurons in one ensemble that project to excitatory neurons in the other, and vice versa. In this case the oscillator phases that correspond to maximal spiking rate will tend to align closely in time.

In contrast, anti-phase coupling occurs when there are relatively many excitatory-to-inhibitory projections between ensembles. The maximally depolarized phase of one ensemble will tend to align with the minimally depolarized phase of the other. Note that inhibitory connections are assumed to be predominantly local, i.e. within ensemble. Thus when the excitatory neurons of ensemble A project to the inhibitory neurons in ensemble B, excitation of A transiently diminishes excitation of B.

There are several important postulates which connect the microscopic and macroscopic levels of description, and which allow us to relate coupling forces (c) to oscillator frequency (ω). These postulates are based on the idea that task ensembles can vary in *size*, i.e. in the number of excitatory neurons which participate in the ensemble.

> *Exertive force (E) and ensemble size (N):* exertive force is conceptualized as a global modulation of the sizes of task ensembles, as shown in Fig. 13B. Increasing *E* increases *N*.
> *Exertive force (E) and oscillator frequency (ω):* increasing E causes oscillator frequency to increase.
> *Ensemble size (N) and coupling strength (c):* increasing *N* increases *c*, because a larger ensemble will have more projections to other ensembles.

The postulates maintain that fluctuations in exertive force influence ensemble size and oscillator frequency. Because of the relation between ensemble size and coupling strength, increasing exertion also increases coupling strength. Furthermore, we can treat ω as a proxy for exertive force and model the effect of ω on coupling strengths.

Justification for modeling exertion as a global modulation is based on the empirical observations: the nonstationarity of response variables (i.e. positive lag-1 autocorrelations) and pervasive correlations between independent systems reported in Sections 3.1 and 3.2 argue in favor of a shared modulation rather than parallel, independent modulations. However, because the strengths of the correlations vary substantially across variable pairs, it is reasonable to infer that there is some variation in the extent to which ensembles respond to the global exertive force.

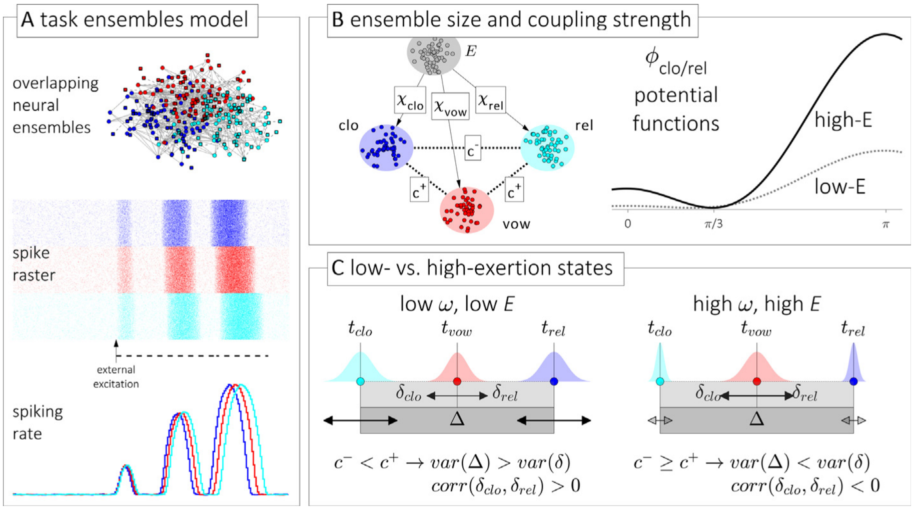
A task ensembles model
overlapping neural ensembles
spike raster
external excitation
spiking rate
B ensemble size and coupling strength
E
χ_{clo}
χ_{vow}
χ_{rel}
clo
rel
vow
c⁻
c⁺
c⁺
$\phi_{clo/rel}$ potential functions
high-E
low-E
0
π/3
π
C low- vs. high-exertion states
low ω, low E
high ω, high E
t_{clo}
t_{vow}
t_{rel}
δ_{clo}
δ_{rel}
Δ
$c^- < c^+ \rightarrow var(\Delta) > var(\delta)$
$corr(\delta_{clo}, \delta_{rel}) > 0$
$c^- \geq c^+ \rightarrow var(\Delta) < var(\delta)$
$corr(\delta_{clo}, \delta_{rel}) < 0$

Fig. 13. Exertion in the task ensembles model of planning oscillators. (A) Task-associated ensembles of excitatory and inhibitory neurons respond to external excitation with collective spiking that macroscopically resembles planning oscillations. (B) Schematization of macroscopic parameters: ensemble size, exertive force, susceptibility, and coupling strength. c^+ and c^- represent in-phase and anti-phase coupling forces, respectively. (C) Comparison of δ variance/covariance profiles in low exertion and high exertion states. $\Delta = \delta_{clo} + \delta_{rel}$.

We model variation in ensemble responses to exertive fluctuations with ensemble-specific susceptibilities χ. The susceptibility χ describes how much ensemble size N increases as the exertive force increases, i.e. $N = \chi E$. We allow for system-specific susceptibilities: χ_{clo}, χ_{vow}, χ_{rel}, and system-specific ensemble sizes: N_{clo}, N_{vow}, N_{rel}.

For conceptual simplicity, each ensemble is hypothesized to have a ground state ensemble size (N^0), i.e. an ensemble size associated with minimum E. Thus the lowest ω observed in the experiment (1.42 Hz) quantifies the value of E when ensembles are in the ground state. To assess the ability of this model to capture empirical patterns, model simulations were fit to the endpoints of across-subject linear regressions of the correlation and variance data from Fig. 12. There was a total of four parameters, but only two of these are critical to producing the qualitative relations between frequency and timing variability:

1. χ_{C}: the susceptibility of closure and release (χ_{V} was set to 1).
2. N°_{C}: the ground-state ensemble size of closure and release (N^0_{V} was set to 1).

The remaining two parameters scale the effect of ω on N and determine the amplitude of Gaussian noise in coupling strengths, allowing for more quantitatively precise fits of the empirical data (see Appendix A.3 for further detail).

In the absence of noise, the equality of in-phase coupling strengths results in symmetric temporal displacement. However, the model makes use of noise in coupling strengths to simulate the empirical correlation and variance patterns. Fig. 14 below shows the model fit of the empirical data, along with linear regressions. The model produces a fairly good match to both the correlation and variance data, although the variance of δ_{clo} appears to be slightly overestimated at low ω and underestimated at high ω.

The model accounts for the empirical correlation and variance as follows. Ground state ensemble size is relatively small for the consonantal ensembles, i.e. $N^{\circ}_{\mathrm{C}} < 1$, but the susceptibility of consonantal ensembles is relatively high, $\chi_{\mathrm{C}} > 1$. Differences in E result in different patterns of coupling strength, which impact correlation and variances in the following way:

Low exertion regime. At low ω, *clo* and *rel* ensembles are smaller than *vow*, i.e. $N_{\mathrm{C}} < N_{\mathrm{V}}$, and so anti-phase coupling between *clo* and *rel* is relatively weak compared to in-phase coupling between *clo-vow* and *rel-vow*. Hence Gaussian noise that mimics trial-to-trial fluctuations in coupling strengths results in var(Δ) > var(δ_{clo}) + var(δ_{rel}), and it follows that corr (δ_{clo}, δ_{rel}) > 0. (Note that $\Delta = \delta_{\mathrm{clo}} + \delta_{\mathrm{rel}}$, cf. Fig. 13C). This low-exertion regime corresponds to the empirical patterns from P02 and P01, shown in Fig. 12.

High exertion regime. At high ω, the greater susceptibility of consonantal ensembles entails that $N_{\mathrm{C}} > N_{\mathrm{V}}$, and so anti-phase coupling becomes relatively stronger than in-phase coupling. Coupling strength fluctuations then result in var(Δ) < var (δ_{clo}) + var(δ_{rel}), and it follows that corr(δ_{clo}, δ_{rel}) < 0. This corresponds to the patterns observed from P03 to P06, shown in Fig. 12.

The mechanism behind differences between the low- and high-exertion regimes involves the relative strength of the anti-phase and in-phase coupling forces. Strong anti-phase coupling is associated with high-ω and less variable/more stable Δ (=$\delta_{\mathrm{clo}} + \delta_{\mathrm{rel}}$), while weak anti-phase coupling is associated with low-ω and more variable/less stable Δ. These differences in stability are reflected in Fig. 13B in the comparison of high-ω and low-ω phase coupling potentials: the valley associated with the equilibrium ϕ is much steeper and narrower when in the high-ω regime with strong anti-phase coupling. Indeed, relations between the slopes of potential functions governing movement timing and variability of movement have been implicated in a variety of contexts (Goldstein et al., 2006; Haken, Kelso, & Bunz, 1985; Tilsen, 2009).

The mechanism that is indirectly responsible for differences between low- and high-exertion regimes is ensemble-specific susceptibility to exertive forces. Although consonantal ensembles are smaller at low exertion, they respond more than vocalic ensembles to fluctuations, and hence as exertion increases, the strength of the anti-phase interaction between consonantal planning systems grows more than the strength of the in-phase interactions.

The circumstance that anti-phase coupling is more influenced by exertion makes intuitive sense when considering perceptual recoverability in fast speech. In general, higher arousal might be associated with faster speech; accordingly, high values of E (and ω) correspond to shorter intervals between movement initiations. However, in faster speech there is a greater potential for gestural overlap to diminish the perceptual recoverability of linguistically relevant information in the signal (Chitoran & Goldstein, 2006). Augmented anti-phase coupling

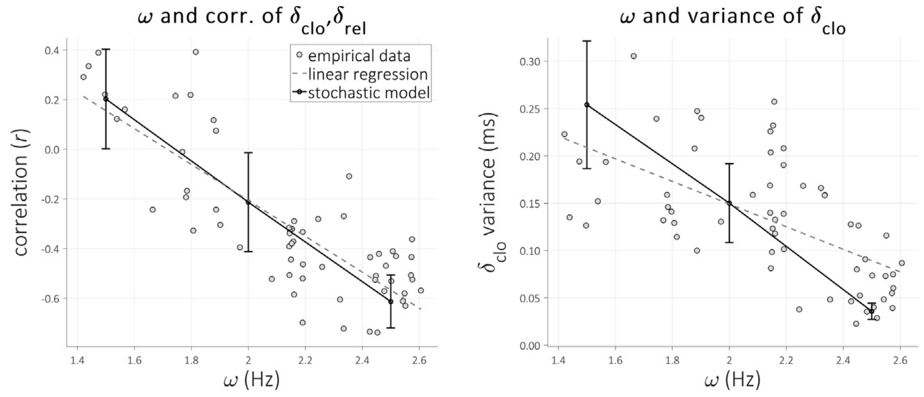

Fig. 14. Model fits of empirical data. (Left) Relation between ω and correlation of δ_{clo} and δ_{rel}. (Right) Relation between ω and variance of δ_{clo}. Confidence bars show ±1σ from 100 model simulations.

mitigates against this effect: stronger anti-phase coupling increases equilibrium ϕ; this makes timing intervals δ longer than they would be if coupling strengths and ω were independent. In other words, the relation between ω and anti-phase coupling is such that compression of δ is resisted as ω increases. Note that while ω may be correlated with speech rate, ω should not be viewed as a speech rate parameter: E and ω must to some degree be independent of speech rate because we can speak slowly in high-exertion states.

An alternative but problematic interpretation of the variability patterns involves a duration-variance relation. In general, variance in the performance and estimation of time intervals is positively correlated with the duration of those intervals. This relation could explain the low variability in Δ associated with high-exertion because Δ is shorter in this regime, but it does not account for why var(Δ) < var(δ_{clo}) + var(δ_{rel}); in other words, the δ intervals should also reflect the duration-variance relation, but they do not. Moreover, the interval durations under consideration here are well below the threshold where duration-variance relations are typically observed: there is a wealth of evidence that, at least in interval tapping and related manual tasks, variance is constant in performance of intervals below 250–300 ms (Peters, 1989; Wing & Kristofferson, 1973a, 1973b). This has been interpreted to indicate that short intervals are necessarily controlled automatically, i.e. coordinated. Hence the variability patterns cannot be interpreted as the consequence of a duration-variance relation.

The task ensembles model can be extended to account for empirically observed asymmetries in timing intervals. Although the across-participant central tendency is symmetric displacement, i.e. $\delta_{\text{clo}} \approx \delta_{\text{rel}}$, the joint δ distributions in Fig. 11 for P02, P03, and P05 exhibit deviations from perfect symmetric displacement. Although these deviations are not very large (they are on the order of 20 ms), an explanation is nonetheless desirable. Such deviations can be modeled straightforwardly by allowing for the ground-state ensemble sizes N^0_{Clo} and N^0_{rel} to differ. For example, P02 has a bias such that $\delta_{\text{clo}} < \delta_{\text{rel}}$. This bias can be modeled by positing that $N^0_{\text{rel}} < N^{\circ}_{\text{Clo}}$, which makes $c_{\text{clo-vow}} > c_{\text{rel-vow}}$. In other words, the closure ensemble is more strongly coupled to the vowel than the release ensemble because it is larger. Syllable stress might also play a role in the inequality of coupling forces.

It is worth pointing out that the model described above is just one of many possible models of the empirical patterns. There are alternative parameterizations that might be equally adept at modeling the data. For example, it may be possible to capture the patterns with a model in which consonantal and vocalic ensembles have equivalent sizes and susceptibilities but different ground-state frequencies (ω_0, see Appendix A.3). There are also alternative mechanisms that could be used to account for the patterns, such as exertion-dependent noise in ensemble sizes. The space of possible models is indeed very large, and restricting that space will require careful experimental and analytical methods.

4.2. Exertive fluctuations and assessment of the equilibration hypothesis

The results were somewhat ambiguous with regard to the equilibration hypothesis. While three of the six sessions exhibited a decrease toward zero lag-1 autocorrelation on long timescales, the other three did not. One reason why evidence for equilibration may be partly lacking is that, in those three sessions, the effects of an equilibration process were obscured by relatively large or more frequent exertive fluctuations. It should not be assumed that there is only one timescale on which exertive variation occurs: some fluctuations may occur rapidly while others occur more gradually, the latter being likely to obscure the effects of equilibration. Moreover, although variation in exertive force creates random walk-like autocorrelation on short analysis scales, it is important not to view exertion as a random walk process: exertive force is a conceptual integration of the effects of many complex metabolic and cognitive mechanisms. While exertive variation induces non-stationarity in behavior that is random-walk like to a first approximation, a more detailed model of exertive mechanisms could presumably generate non-stationarity through a collection of nonlinear interactions on multiple timescales.

Given the above considerations, the source of the ambiguity in assessing the equilibration hypothesis could simply be a methodological limitation: longer observation sequences may be necessary. Ideally, in order to conclusively detect equilibration using autocorrelation scaling analysis, the observation sequences should be substantially longer than the longest timescale on which exertive fluctuations occur. Obtaining sufficiently long, uninterrupted sequences may not be very practical.

An alternative strategy that could be pursued is to test equilibration by controlled perturbations of exertion. Imagine an experimental manipulation that transiently augmented exertion, perhaps the instruction: “You will receive an additional $10 in compensation if your response consistency score over the next 10 trials is above 50”. What would the equilibration dynamics look like subsequent to this perturbation? The response of articulatory control systems to such a perturbation should be phase-locked to the onset and offset of the augmented exertion epoch. Manipulations of this sort may shed light on the nature of interactions between exertion and articulatory control mechanisms.

Relatedly, an important question to consider regards how response consistency feedback in the experiment influences exertion. Although the feedback was designed not to induce dramatic changes in response behavior (cf. Section 2.3), exertive systems are presumably influenced by it to some extent. The nature of the feedback-exertion interaction may be quite complicated, since both negative and positive feedback might influence attention and effort. Post-hoc analyses of the relation between consistency scores and changes in δ_{clo} and δ_{rel} on post-feedback trials showed no clear pattern of correlation; indeed, the magnitudes of changes from pre-to-post-feedback trials were typically within the range of variation in trial-to-trial differences in the absence of feedback. These analyses suggest that the feedback scores did not strongly influence response behavior.

4.3. Further considerations

The theoretical perspective developed here can be considered in relation to the hyper- and hypo-articulation (H&H) theory (Lindblom, 1990). To some extent, the sort of variation

Please cite this article in press as: Tilsen, S. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.03.001

associated with low- and high-exertion might be mapped to a hypo-to-hyper-articulation continuum. However, the basis for variation between hypo- and hyper-articulation in the H&H framework is the involvement of social and communicative constraints, mostly related to listener perception of speech. The variation in the current context is not socially driven and therefore shows that other, system-internal factors can also induce variation. In addition, articulatory predictions discussed in the context of the H&H theory tend to focus on relations between segment duration and target undershoot, rather than timing of movement initiation as in the current study.

While there is undoubtedly some overlap between the hypo-to-hyper-speech continuum and the variation in exertion modeled here, the exertion model makes some specific predictions that the H&H framework does not. For example, the hypothesized relation between exertion and the oscillator frequency parameter ω was employed to predict patterns of variance and covariance in timing intervals. The exertive model also predicts that non-speech outputs of exertive systems such as pulse rate, respiration rate, eye movement, pupil dilation, skin conductivity, body posture and sub-cranial movement, etc. will correlate with variation in ω and accordingly with variation in speech motor system outputs.

One of the deeper implications of the analysis and interpretation of the experimental results here relates to linear vs. cyclic models of time. The order parameter used in the autocorrelation scaling analysis requires a concept of phase (i.e. phase angle), which derives from a cyclic metaphor for time. In contrast, the time periods we measure in experiments are derived from a linear metaphor. In the cyclic conception phases which differ by 2π radians (360°) are equivalent, and there is a maximal phase distance π. This maximal phase distance provides the basis for constraints that map from linear to cyclic time. Here the mapping was accomplished by adopting a uniform coupling constraint as a first approximation (see Appendix A.2), although other constraints might be considered.

A key question is why conduct analyses with phase, why use a cyclic metaphor for time in the first place? This question has been addressed in detail in work by Fowler (1980) and by Kelso and Tuller (1987), which laid out the foundations for the coupled oscillators approach. The most compelling argument in my view is that cyclic time is the conceptualization which most directly corresponds to cognitive representations of timing between coordinated movements—coordinative timing is the output of interacting collective oscillations. The concept of phase thus not only provides a deeper understanding of the motivation for symmetric displacement phenomena, but it is more consistent with the microscopic model in which gestural planning oscillators are instantiated as task-associated neural ensembles that collectively oscillate.

The implications of this conclusion are far reaching: in analyzing temporal phenomena, we must carefully consider which conception of time provides a more appropriate basis for the analysis. Only some temporal observations should be analyzed in a cyclic domain. Longer intervals of time associated with competitively selected movements are better analyzed in linear time, since these intervals derive from processes which are not fundamentally oscillatory (Tilsen, 2013, 2016). Shorter temporal intervals that are associated with coordinated movements (e.g. voice onset time, closure-release timing, consonantal constrictions in complex onsets, etc.) should be conducted in a cyclic time domain.

5. Conclusions

This study hypothesized that exertive mechanisms (attention, effort, focus, arousal, motivation, etc.) have a global modulatory effect on independent speech systems. Support for this hypothesis was obtained in the form of positive lag-1 autocorrelations in response variables and pervasive correlations between outputs of independent systems. The study also hypothesized that an equilibration process should promote symmetric displacement on long timescales. Partial support for this hypothesis was obtained, and evidence was found for high- and low-exertion states of production.

The results of this study show that a promising route to advancing our understanding of the organization and regulation of speech movements is to conduct large-scale investigations of spontaneous variation in speech behavior. The patterns identified in this study are unlikely to be detected in conventional paradigms where there are multiple response conditions. Each condition that an experimenter adds to a study not only reduces statistical power but introduces unavoidable context- and response-ordering confounds which can scramble important temporal dynamics in behavior. The unconditioned variation paradigm is inspired by experimental and theoretical approaches in statistical physics, where macroscopic dynamics of complex systems are measured repeatedly in carefully controlled conditions and models are studied using Monte Carlo methods. Although the dynamics of human behavior are far more chaotic than those occurring in most physical systems, we may gain new insights into speech behaviors by drastically reducing the complexity of both the behaviors we elicit experimentally and the contexts in which those behaviors are observed.

Acknowledgements

I would like to thank Abby Cohn, Draga Zec, and members of the Cornell Phonetics Lab for discussions of this research. Thanks also to Bruce McKee for assistance in conducting the experiment.

Appendix A.

A.1. Simulations of random processes

Means and standard deviations of lag-1 autocorrelation scaling functions for the random processes in Figs. 1 and 10 were calculated from 1000 simulations of each process. Time series in Fig. 1 had a length of 5000 samples; time series in Fig. 10 had a length of 400 samples, matching the empirical data. Random walks were simulated with the rule $x_{i+1} = x_i + \varepsilon$, where ε was a Gaussian-distributed step with standard deviation of 0.1. Gaussian noise components had a mean of 0 and standard deviation of 0.5. Potential functions were of the form $V(x) = 0.5ax^2$, and were incorporated in simulations by adding the opposite of their first derivative, $-dV(x)/dx = -ax$ at each time step. The scaling parameter a, which changes the width of the potential, was set to 0.02 for all simulations.

The reader should note that the simulated processes provide conceptual reference points but are not intended to model the data directly. The shapes of autocorrelation scaling functions for processes with Gaussian noise and a random walk component change substantially with changes in the relative amplitudes of these random noise components, and the potential scaling parameter *a* can have a substantial effect as well. The simulation lengths can also strongly influence the shapes of the autocorrelation scaling functions: comparison of scaling functions from the 5000-sample (Fig. 1) and 400-sample (Fig. 10) simulations illustrates how finite sample sizes induce a negative bias as the analysis scale increases.

A.2. Coupled oscillators model

Each articulatory gesture (indexed by *i*) is associated with a *planning oscillator*, which can be modeled in polar coordinates as a point with phase θ_i and amplitude r_i moving counterclockwise with phase velocity ω (Eq. (2)). Relative phases of planning oscillators are defined as in (Eq. (3)).

$$\dot{\theta}_i = 2\pi\omega, \quad \dot{r}_i = 0 \tag{2}$$

$$\phi_{ij} = \theta_i - \theta_j \tag{3}$$

$$\delta_{ij} = \left(\frac{1}{2\pi\omega}\right)\phi_{ij} \tag{4}$$

Prior to utterance initiation, relative phase coupling forces drive the system of planning oscillators toward a stable equilibrium relative phase configuration. After stabilization each gesture is initiated when its corresponding oscillator reaches an arbitrary initiation phase. The time period between the initiation of movements, δ, is determined by the relative phase and frequency of the planning oscillators associated with the equilibrium configuration (Eq. (4)). (Note that an alternative model from Tilsen (2013) derives timing of movement initiation from relative phases when oscillator amplitudes surpass a threshold).

By analogy to gravitational or electrostatic forces, in-phase and anti-phase potential energy functions V^+ and V^- are associated with ϕ-coupling forces (Eq. (5)), where force is proportional to $-dV/d\phi$ (Eq. (6)). The oscillator equations (Eq. (7)) thus include a term which incorporates the phase coupling, with coupling strengths c_{ij}.

$$V^+(\phi) = -\cos\phi, \quad V^-(\phi) = \cos\varphi \tag{5}$$

$$F(\phi) = -c\frac{dV(\phi)}{d\phi} \tag{6}$$

$$\dot{\theta}_i = 2\pi\omega + \sum_j c_{ij}\frac{-dV(\phi_{ij})}{d\phi_{ij}}, \quad C = \begin{bmatrix} 0 & c_{12} & c_{13} \\ c_{21} & 0 & c_{23} \\ c_{31} & c_{32} & 0 \end{bmatrix} \tag{7}$$

The steady state ϕ-configuration for a system of three oscillators is achieved when all coupling forces are balanced (Eq. (8)). The steady state equation has an analytic solution (Eq. (9)) when bidirectional coupling strength symmetry ($c_{ij} = c_{ji}$) is imposed on the coupling strength matrix *C*. Imposing the symmetric displacement constraint ($c_{12} = c_{23}$) entails that oscillators 1 and 3 are equally displaced from 2, i.e. $\phi_{12} = \phi_{23}$. Imposing the uniform coupling (i.e. $c_{12} = c_{23} = c_{13}$) entails a steady-state ϕ-configuration in which $\phi_{12} = \phi_{23} = \pi/3$.

$$\sum_{ij} c_{ij}\frac{-dV(\varphi_{ij})}{d\varphi_{ij}} = 0 \tag{8}$$

$$\varphi^*_{12} = \varphi^*_{23} = 2\tan^{-1}\left[\frac{2a-b}{2a+b}\right]^{\frac{1}{2}} \tag{9}$$

The uniform coupling constraint minimizes the potential energy cost of deviation in ϕ from the uniform coupling theoretical value of $\pi/3$. This occurs when the average in-phase coupling strength equals the magnitude of the anti-phase coupling strength, i.e. $(c_{12} + c_{23})/2 = |c_{13}|$. It follows that estimated relative phases (ϕ-hat) deviate equally from the uniform coupling equilibrium, $\pi/3$ (Eq. (10)), and leads to the expression for estimated frequency in (Eq. (11)).

$$\Phi = (\hat{\phi}_{12} - \phi^*_{12}) = -(\hat{\phi}_{23} - \phi^*_{23}), \quad \hat{\phi}_{12} = -\hat{\phi}_{23} + \frac{2\pi}{3} \tag{10}$$

$$\hat{\omega} = \frac{1}{3(\delta_{12} + \delta_{23})} \tag{11}$$

Note that the uniform coupling constraint makes the relative phases ϕ_{12} and ϕ_{23} redundant, since each deviates from $\pi/3$ the same amount in opposite directions. Hence no additional information is associated with one ϕ when the other one is known. This is desirable because we have reduced our description of the system to just two parameters: Φ and ω. The variable Φ is an order parameter of the system and describes the extent to which the estimated ϕ deviate from symmetric displacement.

A.3. Exertive force and susceptibility model

In order to test whether the exertive force model can reproduce the empirical correlation and variance patterns of δ_{clo} and δ_{rel}, a stochastic model was optimized to fit three datapoints of the linear regressions in Fig. 14 (also Fig. 12)—these points corresponded to the minimum ω, maximum ω, and the midpoint between these. A multistart optimization procedure was used, with the minimized cost being the sum of the normalized deviations from δ_{clo}, δ_{rel} correlation and variance of δ_{clo}. Because of the stochastic nature of the model, it is important to obtain multiple estimates of the output correlations and variances. Hence 50 repetitions of the simulation were produced for each frequency value, and the outputs were averaged by frequency value.

To simplify notation below, *clo* = 1, *vow* = 2, and *rel* = 3. Each simulation produced 40 trials of δ_{12} and δ_{23}, in order to match the window length used for the analysis in Fig. 12. The δ produced in each trial are influenced by random Gaussian noise that is added to each of the coupling strengths. The relative phases ϕ_{12} and ϕ_{23} were calculated using an analytical solution of the system of equations in (Eq. (12)) where the constraint $\theta_1 > \theta_2 > \theta_3$ was imposed.

$$\begin{aligned} c_{12}\sin\phi_{12} - c_{13}\sin\phi_{13} &= 0 \\ c_{23}\sin\phi_{23} - c_{13}\sin\phi_{13} &= 0 \\ c_{12}\sin\phi_{12} - c_{23}\sin\phi_{12} &= 0 \end{aligned} \tag{12}$$

Table A1
Exertive force model parameters.

Parameter	Value	Description
N^0_C	0.67	Consonantal ground-state ensemble size (vowel N^0 = 1)
χ_C	1.51	Consonantal susceptibility (vowel χ = 1)
β	2.69	Exponent of exertive force effect in Eq. (13)
σ	0.072	Standard deviation of Gaussian noise in Eq. (14)

The parameter ω, in combination with susceptibilities χ_i was used to calculate the ensemble sizes N_i, as in (Eq. (13)), where $N^0{}_i$ and ω^0 are the ground state ensemble size and frequency. The ground-state frequency ω_0 was set at 1.42 Hz, the lowest ω observed in the empirical data. Coupling strengths were determined in each trial from (Eq. (14)), where the first term is the geometric mean of the ensemble sizes and the second includes Gaussian noise with zero mean and standard deviation σ. Modulating this noise with the factor $[1 + (\omega - \omega_0)]$ improves the quantitative fit by making the coupling strength noise depend on ω, the proxy for exertion.

$$N_i = N_i^0 + \chi_i(\omega - \omega_0)^\beta \tag{13}$$

$$c_{ij} = (N_i N_j)^{\frac{1}{2}} + [1 + (\omega - \omega_0)]\varepsilon(0, \sigma) \tag{14}$$

Equality of consonantal gesture ensemble size was enforced by imposing the constraint: $N^0_1 = N^0_3 = N^0$C and $\chi_1 = \chi_3 = \chi_C$, i.e. the ground-state ensemble sizes and susceptibilities are the same for both consonantal gestures. The ground state N^0 for the vowel was fixed at 1, and the vowel χ was fixed at 1. The optimized parameter values are given in Table A1.

References

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Brookes, M. (1997). Voicebox: Speech processing toolbox for matlab. *Softw. Available Mar 2011*.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica, 45*(2–4), 140–155.

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

Carpenter, G. A., & Grossberg, S. (1987). Neural dynamics of category learning and recognition: Attention, memory consolidation, and amnesia. *Advances in Psychology, 42*, 239–286.

Chatfield, C. (2016). *The analysis of time series: An introduction*. CRC Press.

Chitoran, I. & Goldstein, L. (2006). Testing the phonological status of perceptual recoverability: Articulatory evidence from Georgian. In *Proc. of the 10th Conference on Laboratory Phonology, Paris, June 29th–July 1st, 2006*. pp. 69–70.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8*(1), 113–133.

Gafos, A., & Goldstein, L. (2012). Articulatory representation and organization. In A. Cohn, C. Fougeron, & M. K. Huffman (Eds.). *The handbook of laboratory phonology* (pp. 220–231). New York: Oxford University Press.

Goldenberg, D., Tiede, M., Honorof, D. N., & Mooshammer, C. (2014). Temporal alignment between head gesture and prosodic prominence in naturally occurring conversation: An electromagnetic articulometry study. *Journal of the Acoustical Society of America, 135*(4). 2294–2294.

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

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51*(5), 347–356.

Hausdorff, J. M., Purdon, P. L., Peng, C. K., Ladin, Z., Wei, J. Y., & Goldberger, A. L. (May 1996). Fractal dynamics of human gait: Stability of long-range correlations in stride interval fluctuations. *Journal of Applied Physiology, 80*(5), 1448–1457.

Hermes, A., Mücke, D., & Grice, M. (2013). Gestural coordination of Italian word-initial clusters: The case of 'impure s'. *Phonology, 30*(1), 1–25.

Ishi, C. T., Ishiguro, H., & Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication, 57*, 233–243.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation, 18*(2), 245–282.

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. MIT Press.

Kelso, J., & Tuller, B. (1987). Intrinsic time in speech production: Theory, methodology, and preliminary observations. *Sensory and motor processes in language* (vol. 203, pp. 222). Hilladale, NJ: Erlbaum.

Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 369*(1658), 20130397.

Lay, B. S., Sparrow, W. A., & O'Dwyer, N. J. (2005). The metabolic and cognitive energy costs of stabilising a high-energy interlimb coordination task. *Human Movement Science, 24*(5), 833–848.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403–439). Springer.

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control, 14*(3), 380–407.

Nam, H. (2007). Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In J. Cole & J. I. Hualde (Eds.). *Laboratory phonology* (Vol. 9, pp. 483–506). Berlin, New York: Walter de Gruyter.

Nam, H. & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th international conference on phonetic sciences, Barcelona, Spain*. pp. 2253–2256.

Ohala, J. J., & Eukel, B. W. (1987). Explaining the intrinsic pitch of vowels. *Honor Ilse Lehiste*, 207–215.

Paus, T., Zatorre, R., Hofle, N., Caramanos, Z., Gotman, J., Petrides, M., & Evans, A. (May 1997). Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *Journal of Cognitive Neuroscience, 9*(3), 392–408.

Peters, M. (1989). The relationship between variability of intertap intervals and interval duration. *Psychological Research Psychologische Forschung, 51*(1), 38–42.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the 4th international conference on speech prosody* (pp. 175–184). Brazil: Campinas.

Sethna, J. (2006). *Statistical mechanics: Entropy, order parameters, and complexity* (Vol. 14). Oxford University Press.

Shaw, J., Gafos, A. I., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology, 28*(3), 455–490.

Stergiou, N., & Decker, L. M. (2011). Human movement variability, nonlinear dynamics, and pathology: Is there a connection? *Human Movement Science, 30*(5), 869–888.

Stergiou, N., Harbourne, R. T., & Cavanaugh, J. T. (2006). Optimal movement variability: A new theoretical perspective for neurologic physical therapy. *Journal of Neurologic Physical Therapy, 30*(3), 120–129.

Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., ... Prince, J. (2007). Comparison of speech production in upright and supine position. *Journal of the Acoustical Society of America, 122*(1), 532–541.

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). *Speech Coding and Synthesis, 495*, 518.

Temprado, J.-J., Zanone, P.-G., Monno, A., & Laurent, M. (1999). Attentional load associated with performing and stabilizing preferred bimanual patterns. *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1579.

Tiede,M. K., Masaki, S., Vatikiotis-Bateson, E. (2000). Contrasts in speech articulation observed in sitting and supine conditions. In: *Proceedings of the 5th seminar on speech production, Kloster Seeon, Bavaria*. pp. 25–28.

Tiede, M., & Goldenberg, D. (2015). Dual electromagnetic articulometer observation of head movements coordinated with articulatory gestures for interacting talkers in synchronized speech tasks. *Journal of the Acoustical Society of America, 137*(4). 2302–2302.

Tilsen, S. (2015). "Structured nonstationarity in articulatory timing. *Proc. 18th Int. Congr. Phon. Sci.*

Tilsen, S. (2009). Toward a dynamical interpretation of hierarchical linguistic structure. *UC Berkeley Phonol. Lab Annu. Rep*. pp. 462–512.

Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speech planning. *PLoS ONE, 8*(4), e62800.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics, 55*, 53–77.

Tilsen, S. (2016). A shared control parameter for F0 and intensity. In *Proceedings of the 8th international conference on speech prosody, Boston, MA*.

Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS ONE, 11*(1), e0146813.

Tilsen, S., Zec, D., Bjorndahl, C., Butler, B., L'Esperance, M.-J., Fisher, A., ... Sanker, C. (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology, 2012*, 51–81.

Whalen, D. H., & Gick, B. (2001). Intrinsic F0 and tongue depth in ATR languages. *Journal of the Acoustical Society of America, 110*(5), 2761.

Wing, A. M., & Kristofferson, A. B. (1973a). The timing of interresponse intervals. *Perception & Psychophysics, 13*(3), 455–460.

Wing, A. M., & Kristofferson, A. B. (1973b). Response delays and the timing of discrete motor responses. *Perception & Psychophysics, 14*(1), 5–12.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., & Liu, X. (1997). *The HTK book* (Vol. 2). Cambridge: Entropic Cambridge Research Laboratory.

Zanone, P. G., Monno, A., Temprado, J. J., & Laurent, M. (2001). Shared dynamics of attentional cost and pattern stability. *Human Movement Science, 20*(6), 765–789.

Please cite this article in press as: Tilsen, S. Exertive modulation of speech and articulatory phasing. *Journal of Phonetics* (2017), http://dx.doi.org/10.1016/j.wocn.2017.03.001