A Coupled Oscillator Planning Account of the Speech Articulatory Coordination Metric With Applications to Disordered Speech

Adam C. Lammert[a], James Williamson[b], Nadee Seneviratne[c], Carol Espy-Wilson[c] and Thomas F. Quatieri[b]
[a]Worcester Polytechnic Institute, [b]MIT Lincoln Laboratory, [c]University of Maryland, College Park

In healthy speech production, distinct actions of the vocal tract (i.e., *gestures*) are coordinated in potentially complex, but nonetheless reliable, temporal relationships. Maintaining these temporal relationships is critical for effective speech production because the temporal structure of gestures is informational[2], such that reducing either its complexity or reliability, can shift perception from one word to another or decrease intelligibility. Due to the challenge that maintaining such coordinative patterns is assumed to present to the nervous system, it has been hypothesized – e.g., by Williamson[6]– that changes in temporal coordination may act as a sensitive indicator of neurological change.

The Speech Articulatory Coordination (SAC) metric has been developed to quantify the temporal structure among gestures captured in time-series data, and to detect and discriminate between coordination patterns that are relatively simplified or variable, with applications to monitoring changes in neurological state. In that regard, the SAC metric has been successfully applied to assessment of a wide variety of neurological conditions, where it has been found, for example, that higher levels of depression severity are associated with simplified coordination patterns[6], while high altitude exposure is associated with more highly variable coordination patterns[5].

Though inspired by an understanding of gestural coordination, the foundations of the SAC metric in theories of speech motor planning have not been rigorously established. **The present work aims to** (1) develop a theoretical foundation for the SAC metric, based on coupled oscillator planning models of speech timing and syllable structure[2,3], and (2) demonstrate how the SAC metric changes as a function of complexity and reliability of gestural coupling in such models. Coupled oscillator planning models attempt to describe the temporal coordination of speech gestures as stemming from the entrainment of time-invariant systems of coupled oscillators. Such models have been used to explain complex sequencing between component gestures in healthy speech articulation, and can be extended to account for coordination changes in disordered speech, as developed here.

The SAC metric is based upon pair-wise comparisons between the factors in multivariate speech-relevant time series data, such as vocal tract constriction variables[1], formant frequencies, or mel-frequency cepstral coefficients[6]. Notably, consistent results have been observed in applying this metric to articulatory data and acoustic data[1]. Pair-wise covariance values are calculated at different time lags, and used to construct a time-embedded matrix, from which the eigenspectrum is computed. The eigenspectrum shape quantifies the complexity of the covariance matrix, and has been used as a metric of the coordinative patterns present in the original time-series data.

Differences in eigenspectrum shape – e.g., its slope as a function of eigenvalue index – may be explained, in the present account, through coupled oscillator models of speech motor planning[2,3]. In such models, the gestures that make up an utterance are each associated with corresponding planning oscillators, each of which may be coupled to other planning oscillators with specific phasing relationships. In a recent exposition of this idea by Parrell & Lammert[4], the equation governing each planning oscillator is given as:

$$m_i \dot{x}_k = \alpha_x + C_{kl}, \text{ where } C_{kl} = \alpha_{kl}\, sin([x_k - x_l] + \varphi_{kl}) \qquad (1)$$

for gesture k, with state x_k, coupled to gesture l with coupling strength α_{kl}. The variable φ_{kl} denotes the target relative phase between x_l and x_k, where their relative phase is defined as $x_k - x_l$. The parameters α_x determine the angular velocity of the oscillator and rate of convergence, respectively. It has been posited that the simplest, most common phasing relationship is in-phase ($\varphi = 0^{\circ}$), followed by anti-phase ($\varphi = 180^{\circ}$)[2,3]. Based on these couplings and phasing relationships, gestures entrain to each other during a motor planning phase, prior to initiation of action, determining their temporal structure.

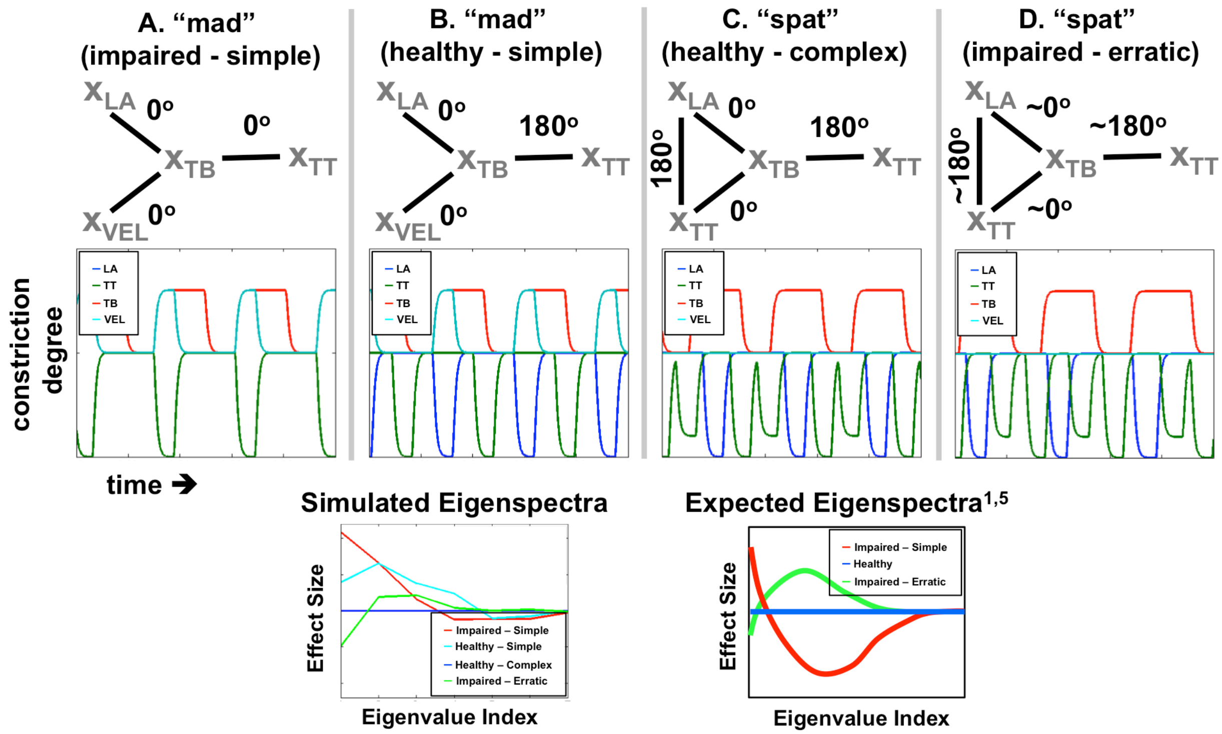

Figure 1: Example coupling configurations (A-D) among gestures (x) of various vocal tract variables (lip aperture, LA; tongue body, TB; tongue tip, TT; velum, VEL), along with associated vocal tract kinematics and SAC metric eigenspectra demonstrating changes in temporal structure. Configurations include: (B) the word "mad" under during healthy/intact conditions, with its simple onset, (A) the word "mad" under impaired conditions, with overly simplified, depression-like coupling, (C) the word "spat" under healthy/intact conditions, with its complex onset, (D) the word "spat" under impaired conditions, with erratic, hypoxic-like coupling. Multiple repetitions of each target utterance were simulated and are shown. Phasing relationships are indicated in degrees, with time-variant relationships indicated by "~".

The complexity of temporal structure depends upon the topology of the coupling graph (e. g., by adding or eliminating couplings) and the nature of the phasing relationships (e.g. , from 0° to 180°). Simpler coordination patterns are associated with fewer, in-phase couplings among gestures. For instance, whereas it has been proposed that gestures in simple syllable onsets (as in, e.g., “mad”) are coupled in-phase to the vowel nucleus, gestures in complex onsets (e.g., “spat”) also contain anti-phase couplings among themselves. The reliability of temporal structure depends upon the time-invariant nature of phasing relationships among gestures, and erratic behavior can result from allowing those relationships to change over time, for example according to a random walk.

Toward validating this theoretical account, we conduct simulation studies of changes in the SAC metric due to changes in the complexity and reliability of gestural coupling, using a custom implementation of coupled oscillator planning, and the associated Task Dynamics model of speech motor control to generate relevant speech kinematics. Selected examples of simulated utterances with healthy and pathological coordination patterns are shown in Figure 1. We further estimate SAC metric values as a general function of coupling topology, phasing relationships and time-variance of coupling relationships.

1 Espy-Wilson, C., Lammert, A. C., Seneviratne, N., & Quatieri, T. F. (2019). Assessing Neuromotor Coordination in Depression Using Inverted Vocal Tract Variables. Interspeech.

2. Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. In G. Fant, H. Fujisaki, & J. Shen (Eds.), Frontiers in phonetics and speech science.

3. Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. ICPhS.

4. Parrell, B., & Lammert, A.C. (2019). Bridging Dynamical Systems and Optimal Trajectory Approaches to Speech Motor Control With Dynamic Movement Primitives. Frontiers in Psychology, 10.

5. Williamson, J. R., Quatieri, T. F., Lammert, A. C., Mitchell, K., Finkelstein, K., Ekon, N., ... & Heaton, K. (2018). The Effect of Exposure to High Altitude and Heat on Speech Articulatory Coordination. Interspeech.

6. Williamson, J. R., Young, D., Nierenberg, A. A., Niemi, J., Helfer, B. S., & Quatieri, T. F. (2019). Tracking depression severity from audio and video based on speech articulatory coordination. CSL.