A new dynamics for prosodically-conditioned variation in articulation

Jason A. Shaw, Michael C. Stern
Department of Linguistics, Yale University (USA)

Overview: That prosodic structure manifests in articulatory kinematics is by now well-established, as are some patterns of variation across languages (e.g., Cho, 2016). Dynamical approaches have conceptualized prosody as a modulation of gesture dynamics, either by changing particular dynamical parameters (e.g., Beckman, Edwards, & Fletcher, 1992; Cho, 2006) or by trans-gestural modulation of time (Byrd, Krivokapić, & Lee, 2006; Byrd & Saltzman, 2003) and/or space (Katsika, Krivokapić, Mooshammer, Tiede, & Goldstein, 2014; Saltzman, Nam, Krivokapic, & Goldstein, 2008) within the scope of prosodic influence. These approaches typically assume a gesture governed by damped mass-spring dynamics with scaled activation (Byrd & Saltzman, 1998). We show here how a simpler dynamical model of the gesture can accurately characterize prosodically-conditioned temporal variation in articulatory kinematics. On this approach, a single control parameter, which can be estimated directly from data, captures articulatory kinematic variation across prosodic contexts. **Proposal:** Our dynamical system is defined in (1). x is the state of some dimension of phonological control; x_0 is the target state of that same dimension. For example, Lip Aperture (LA: the distance between the lips) is a dimension of phonological control for /b/ and /m/ and the target state is 0 (lips together) or even negative (e.g., Parrell, 2011). The variable λ modulates the relationship between velocity, $\dot{x}$, and the distance to target, $(x - x_0)$. For this system to accurately describe speech kinematics, λ must change over time. A constant λ would cause velocity to reach its maximum earlier than is observed in actual kinematic data. We propose that λ starts small (near zero) and increases non-linearly over time, according to the equation in (1b). The derivative of lambda, $\dot{\lambda}$, is equal to λ times a constant, r. This predicts a loglinear relationship over time between instantaneous velocity and distance to target, a prediction which we verify below. This means that r—one of two control parameters of the system, along with x_0—can be estimated for any trajectory as the slope of the natural logarithm of velocity divided by displacement over time. We show that this dynamics provides an excellent fit to data and that variation in r captures differences in the kinematics across prosodic environments. Note that the two equation system in (1) can be rewritten as a second order dynamical system in a single equation, as in (2). **Empirical validation**: Our data comes from Electromagnetic Articulography recordings of 1,977 tokens of CV syllables [ma], [mi], [ba], [bi] produced in real words of English (12 participants) and Mandarin (12 participants) in question-answer pairs designed to elicit variation in focus (data from Liu, Wang, Stern, Kramer, & Shaw, 2023). Liu et al. (2023) reported a significant effect of focus on lip closing, lip opening, and vowel (tongue body) constriction durations for English, replicating past work (Cho, 2006). For Mandarin, focus had a significant effect on vowel movement duration but not on the consonant movements: lip closing and lip opening. To validate our proposed dynamics, we first tested the main prediction that the relation between instantaneous velocity and distance is loglinear over time. For each trajectory, we fit a linear regression to the natural log of the relation between velocity and distance. The mean $\boldsymbol{R^2}$ of the linear fits was excellent: **0.95** <u>for lip closing movements</u>, **0.91** <u>for lip opening</u>; and **0.89** for <u>vowel constriction movements</u>. For reference, Kuberski & Gafos (2023) report mean R^2 values of between 0.68 and 0.86 for fits of the damped mass spring dynamics to /ta/ and /ka/ trajectories. Besides providing excellent fits to the kinematic trajectories, the distribution of r values captured the reported language-specific effects of focus (Figure 1). For English, r showed significant decrease under focus for all three gestures; for Mandarin, only the vowel gesture showed a significant decrease in r under focus. Our proposed dynamical system provides an excellent fit to EMA trajectories, and variation in a single control parameter, r, derives the complete range of focus-related kinematic variation in the data.

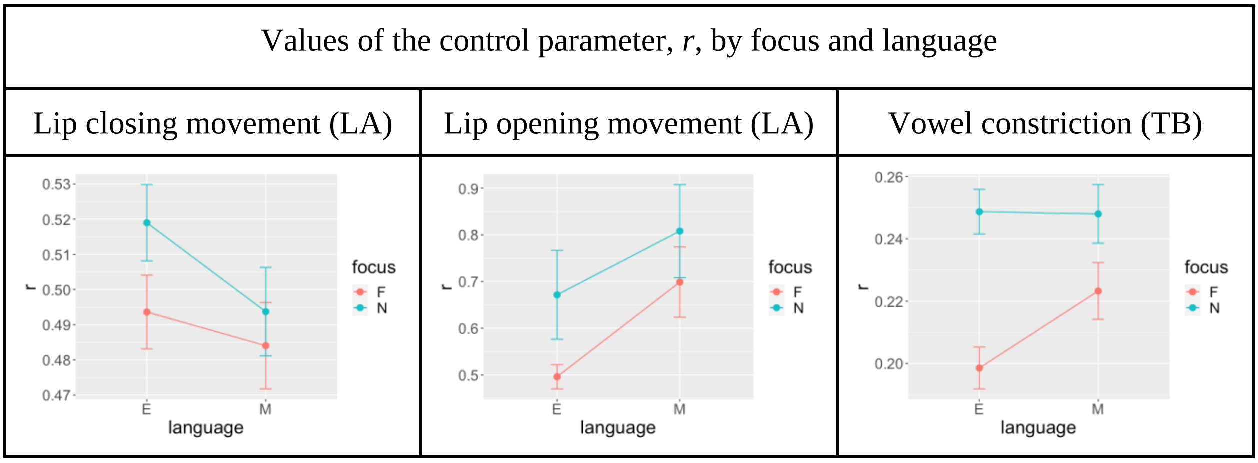
(1) Dynamical system (two equation rendering):

$$\dot{x} = -\lambda(x - x_0) \qquad (1a)$$

$$\dot{\lambda} = r\lambda \qquad (1b)$$

(2) Dynamical system (one equation rendering):

$$\dot{x} = (\ddot{x}/\dot{x} - r)(x - x_0)$$



Values of the control parameter, *r*, by focus and language

Lip closing movement (LA) | Lip opening movement (LA) | Vowel constriction (TB)

Figure 1. Means and 95% confidence intervals for values of *r* fit to movement trajectories across language {English (E), Mandarin (M)}, focus condition {focused (F), non-focused (N)}, and gesture: Lip Aperture (LA) for the closing movement of the consonant (left); LA for the release phase (opening) of the consonant (center); and vowel constriction movement (right). These movements were parsed from EMA trajectories—the lips for consonants; the tongue body for vowels— with reference to the velocity signal. Due to issues with parsing the release movement, far fewer data points contributed to this plot (N = 696) than to the lip closing (N = 1,977) and vowel constriction (N = 1,977) plots.

References

Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 68-86). Cambridge, England: Cambridge University Press.

Byrd, D., Krivokapić, J., & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *The Journal of the acoustical society of America, 120*(3), 1589-1599.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics, 26*, 173-199.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*, 149-180.

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory Phonology, 8*, 519-548.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*(3), 120-141.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics, 44*, 62-82.

Kuberski, S. R., & Gafos, A. I. (2023). How thresholding in segmentation affects the regression performance of the linear model. *JASA Express Letters, 3*(9).

Liu, Y., Wang, Y., Stern, M. C., Kramer, B. M., & Shaw, J. A. (2023). Temporal scope of articulatory slowdown under informational focus: data from English and Mandarin. In R. Skarnitzl & J. Volín (Eds.), *19th International Congress of Phonetics Sciences* (pp. 1691-1695). Prague, Czech Republic: Guarant International.

Parrell, B. (2011). Dynamical account of how /b,d,g/ differ from /p,t,k/ in Spanish: Evidence from labials. *Laboratory Phonology, 2*(2), 423-449.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic sctructure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *4th International Conference on Speech Prosody*. Campinas, Brazil.