ISSP 2024 - 13th International Seminar on Speech Production
13-17 May 2024, Autrans, France

Predicting Articulatory Landmarks with Critically-Damped Oscillators and General Tau Theory

Christopher Geissler[1], Jyothiraditya Nellakra[1]

[1]*Carleton College, USA*
cgeissler@carleton.edu, nellakraj@carleton.edu

Abstract

Dynamical systems are useful for bridging discrete and continuous aspects of speech. In this paper, we compare the ability of two models, critically-damped oscillators and General Tau Theory, to predict gestural landmarks.

Predictions of the two models were compared with each other and with original kinematic data. The data consisted of electromagnetic articulography recordings of Tibetan collected as part of Geissler (2021). In addition to the landmarks-based approach, this study also uses a a language with a different phonological typology.

As compared to results from kinematic thresholds, the critically-damped oscillator model tended to predict that landmarks would take place earlier in time and closer to the target. The General Tau model generally predicted that landmarks would take place later and farther from the target. These results highlight the differences in, and invite further comparison on, the trajectory shape generated by the two models.

Keywords: articulation, articulatory phonology, gestures

1. Introduction

The mathematics of dynamical systems has proven to be a fruitful way to relate continuous and discrete properties of speech (Iskarous 2017; Mücke, Hermes, and Tilsen 2020). In this paper, we compare the ability of two models, critically-damped oscillators and General Tau Theory, to predict individual points in kinematic data.

Articulatory movements have been modeled as critically-damped mass-spring oscillators by Saltzman and Munhall (1989) in Task Dynamics. Among the benefits of this approach is the ability to describe intergestural timing in terms of phase, and to coordinate gestures by coupling the oscillators, as in Nam and Saltzman (2003).

More recently, Elie, Lee, and Turk (2023) have applied General Tau Theory to speech. This model, adapted from work on non-speech motor control, is based on the time-to-closure of "gaps" rather than mass-spring systems. Elie, Lee, and Turk (2023) found that a Tau-based approach compared favorably to coupled-oscillator implementations when fitting kinematic data. That study globally compared the fit of several models to a corpus of electromagnetic articulography (EMA) data of English speech.

The present study instead focuses on experimental stimuli collected to study gestural timing, and uses a typologically-different language, Tibetan. We test coupled-oscillator and General Tau models by fitting each to articulatory trajectories, then comparing their predictions for specific points that are commonly used as landmarks for characterizing articulatory gestures. Our findings highlight advantages of each model, and demonstrate how differences in the curves translate to differences at salient kinematic landmarks.

2. Methods

Predictions of the two models–critically-damped oscillators and General Tau Theory–were compared with each other and with original kinematic data. Data and code are available on OSF: https://osf.io/x34sa/

2.1. Kinematic data

The data consisted of electromagnetic articulography recordings collected as part of Geissler (2021). Six native speakers (four female) of Tibetan living in and around New York City participated in the experiment. All speakers were multilingual, and all speakers were raised in Tibetan diaspora communities in India and Nepal.

Stimuli consistent of Tibetan words elicited in a carrier sentence, presented on a screen in the Tibetan orthography. Target syllables were word-initial and consistent of /m/ followed by the vack vowels /u o a/. The target words were preceded by the vowel /i/ in the carrier sentence in order to facilitate identification of vowel retraction. Target syllables were balanced to include both high and low tone, presence/absence of a coda consonant, and occured either in one-syllable words or as the first syllable in a two-syllable word.

EMA sensors were placed on the upper and lower lips, lower incisor, tongue tip, dorsum, and blade. Consonant gestures were identified as the closing of the lips, and the vowel gesture was identified as the retraction of the tongue dorsum. Gestural landmarks, depicted in Figure 1, were calculated in *Mview* (Tiede 2005), and the position, velocity, and timestamp of each landmark was recorded. In the closure phase, **Gestural Onset** and **Nuclear Onset** were defined as the points at which 20% of peak velocity were achieved in acceleration and deceleration toward the target. Likewise, **Nuclear Offset** and **Gestural Offset** were defined as points with 20% of peak velocity in movement away from the target. These timestamps, along with the point of **Maximum Constriction**, were the focus of analysis.

2.2. Simulations

Parameters for each model were set using certain landmarks, then used to predict the spatio-temporal coordinates at other landmarks. Both models took as inputs the displacement (change in position) of a gesture; the critically-damped oscillator model used the peak velocity and the point at which this was achieved (PVEL and PVEL2), while the General Tau model also used the duration of the movement.

Both models could then calculate the position at any given

10.21437/issp.2024-44

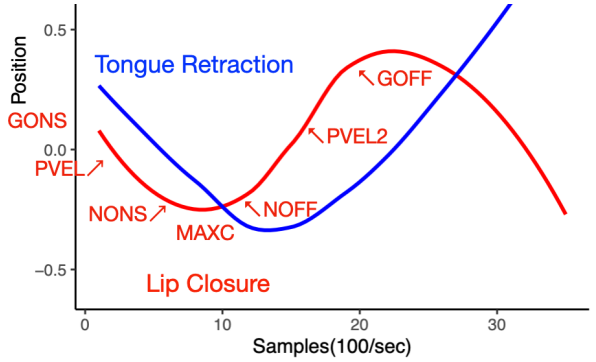

Figure 1: Gesutral landmarks in the lip closure gesture of [ma]. GONS = gesture onset; PVEL = peak velocity of closure; NONS = nuclear onset; MAXC = maximum constriction; NOFF = nuclear offset; PVEL2 = peak velocity of release; GOFF = gesture offset

time duration the gesture, and were used to identify the position and timestamps for the gesture-internal landmarks PVEL, NONS, and PVEL2.

2.2.1. *Critically-damped oscillator model*

In the critically-damped oscillator model, it is possible to calculate the position from a timestamp (or vice versa) using two parameters: the displacement and the natural frequency of the oscillator. The displacement was calculated as the distance from gestural onset to maximum constriction for the closure phase, and the distance from nuclear offset to gestural offset for the release phase. The natural frequency, ω_0, can be calculated from the position and velocity of the system at the point of peak velocity. (1) shows the general equation for a mass-spring system, which can be restated as (2) for a critically-damped oscillator.

$$m\ddot{x} + b\dot{x} + kx = 0, \tag{1}$$

$$\ddot{x} + 2\omega_0\dot{x} + \omega_0^2 x = 0. \tag{2}$$

At the point of peak velocity, this simplifies to (3), since the acceleration $\ddot{x} = 0$. Note that, since the oscillator returns to an equilibrium point $x = 0$, the velocity will always have a sign opposite to the displacement, which ensures that the value of ω_0 must be positive.

$$\omega_0^2 x_{\text{PVEL}} = -2\omega_0 \dot{x}_{\text{PVEL}} \implies \omega_0 = -2\left(\frac{\dot{x}_{\text{PVEL}}}{x_{\text{PVEL}}}\right) \tag{3}$$

Thus, by knowing the displacement x_0, peak velocity, and position at peak velocity, the position x can be calculated as a function of time t using (4):

$$x(t) = x_0\left(e^{-\omega_0 t} + \omega_0 t e^{-\omega_0 t}\right), \tag{4}$$

2.2.2. *General Tau model*

For the Tau model, we used the following equation from Elie, Lee, and Turk (2023), which is derived from Lee (1998). This gives the position of an articulator at a given time t from its starting position x_0 and T, the time at which the target ($x = 0$) is to be achieved.

The only additional parameter is κ, which is analogous to stiffness in that it determines the shape of the velocity profile. $\kappa = 0.4$ was used, following the observation by Elie, Lee, and Turk (2023) that this value held across speakers and articulators; this is also the value at which velocity profiles are symmetrical.

$$x(t) = x_0\left(1 - \frac{t^2}{T^2}\right)^{\frac{1}{\kappa}} \tag{5}$$

The Tau model thus requires the displacement and duration of a movement in order to predict the points in between. The displacement was the same as for the critically-damped oscillator model: the distance from gestural onset to maximum constriction for the closure phase, and the distance from nuclear offset to gestural offset for the release phase. The timestamps of these points were used as the durations.

3. Results

A comparison of predicted with actual data is presented in the figures below: **Figure 2** for position data, and **Figure 3** for temporal data. Analysis and model comparison with linear mixed-effects models confirmed that interactions between landmark and model/data type were significant for both time and distance.

As compared to results from kinematic thresholds, the critically-damped oscillator model tended to predict that landmarks would take place earlier in time and closer to the target. The General Tau model generally predicted that landmarks would take place later and closer to the target. These patterns broadly held for both closure and release landmarks, and for both the consonantal lip gesture and the vocalic tongue dorsum gesture.

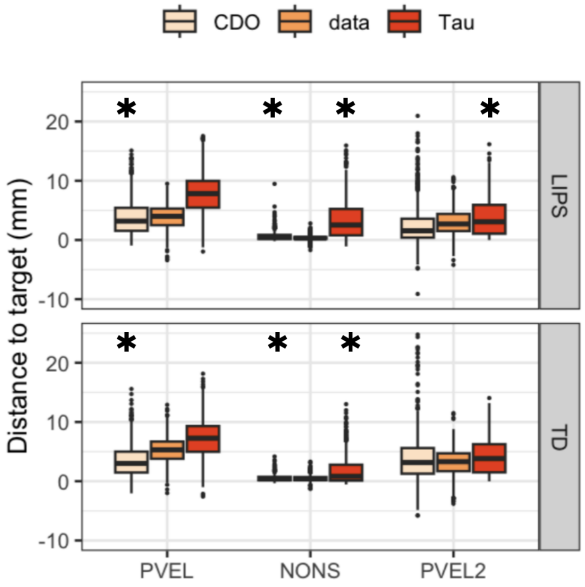

Figure 2: *Predicted position for landmarks in Tibetan /mV/ sequences. Asterisks indicate significant differences between predicted and observed data. CDO = critically-damped oscillator; data = kinematically-defined landmarks; Tau = General Tau model. PVEL/PVEL2 = point of peak velocity toward/away from target; NONS = (gestural) nucleus onset*

We performed a linear mixed-effects analysis on the relationship between these data and their source (kinematic data, oscillator model, Tau model) using the *lme4* package in R.

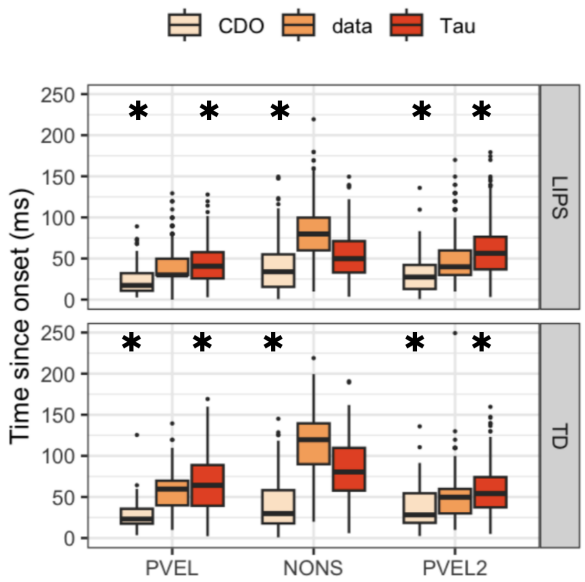

Figure 3: *Predicted time for landmarks in Tibetan /mV/ sequences. Abbreviations as in 2*

We fit two models: one for the position data, and one for the time data. For each, we entered as fixed effects the landmark (PVEL, NONS, PVEL2), articulator (lips or tongue dorsum), and source, as well as random effects of speaker and word. These models were compared to another pair of models that also included an interaction between landmark and source. **Table 1** reports this model comparison, which supports the model that includes an interaction.

Table 1: *Comparison of baseline and interaction models, showing improved fit with interaction.*

Position Model	AIC	BIC	logLik
baseline	184609	184687	-92295
interaction	182684	182797	-91329
Time Model	AIC	BIC	logLik
baseline	278535	278609	-139258
interaction	277350	277457	-138662

We conducted a post-hoc analysis using the *emmeans* package to identify pairwise differences between levels of the models. Specifically, we noted where there were significant differences between oscillator- or Tau-predicted data and the observed kinematic data. These are indicated in Figs. 2 and 3.

Predictions of the oscillator model were significantly different from kinematic data in 10 of 12 cases, while the predictions of the Tau model were significantly different in 7 cases. Interestingly, the Tau model achieved closer values than the oscillator model on the peak-velocity landmarks (PVEL and PVEL2) despite the fact that the oscillator model used these points as inputs.

The direction of the divergence between models is also noteworthy. In the spatial domain, the oscillator model tended to predict that landmarks would occur slightly closer to the target than was identified in the kinematics, while the Tau model predicted landmarks occurring slightly farther from the target. In the temporal domain, the oscillator model predicted landmarks occurring earlier than in the kinematics, while the Tau-predicted landmarks occurred around the same time as, or after, their kinematic equivalents.

4. Discussion

This study compared the ability of two models to predict the spatial and temporal points at which kinematically-defined gestural landmarks would occur. Both the critically-damped oscillator model and the General Tau model predicted landmarks with a fair degree of accuracy, but with some systematic differences. Oscillator-predicted landmarks fell sooner and closer to the target, while the opposite was the case for the Tau-predicted landmarks.

These results highlight the differences in the shapes of the trajectories generated by each model. Critically-damped oscillators move rapidly, then slow to asymptotically approach the target; Tau-derived trajectories unfold gradually (and, when $\kappa = 0.4$, symmetrically), and reach the target at a known point in space and time. We encourage further research on the models to address not only overall fit to data, but also how the details of particular shapes.

The use of data from a less-commonly studied language, Tibetan, is an important part of creating models that more accurately capture the diversity of human speech. It is noteworthy that the value of κ obtained from English speech by Elie, Lee, and Turk (2023) worked reasonably well for the Tibetan data. Further study is needed on the ways κ might vary across languages, speakers, natural classes, articulators, and contexts, parallel to similar work on stiffness in Task Dynamics.

Constructing these models also called attention to the importance of careful definitions for the start and end of a gesture. Both oscillator and Tau models required kinematic landmarks: the oscillator model used the onset of the gesture (along with the peak velocity), while the Tau model used both beginning and end of each gesture. Using different values, such as the point of maximum constriction rather than nuclear onset for the Tau model, leads to different results. Careful consideration for the use of particular landmarks is crucial to accurately comparing models.

This study was limited by the range of materials and the relatively simple versions of the models used. For example, we would expect to find better-fitting curves had the oscillator model used gradient activation like that of Sorensen and Gafos (2016). Nevertheless, the results demonstrate that generating predictions for specific points allows for models to be tested against each other and against speech data.

5. Acknowledgements

This research was supported by a Student Research Partnership grant from the Humanities Center at Carleton College. Original data collection was supported by Yale University, and conducted with the help of Jason Shaw and Muye Zhang.

6. References

Elie, Benjamin, David N. Lee, and Alice Turk (June 2023). "Modeling trajectories of human speech articulators using general Tau theory". en. In: *Speech Communication* 151, pp. 24–38. DOI: 10.1016/j.specom.2023.04.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167639323000614 (visited on 06/19/2023).

Geissler, Christopher (2021). "Temporal articulatory stability, phonological variation, and lexical contrast preservation in diaspora Tibetan". PhD thesis. Yale University. URL: https://elischolar.library.yale.edu/gsas_dissertations/52.

Iskarous, Khalil (Sept. 2017). "The relation between the continuous and the discrete: A note on the first principles of speech dynamics". en. In: *Journal of Phonetics* 64, pp. 8–20. DOI: 10.1016/j.wocn.2017.05.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0095447017301006 (visited on 10/17/2020).

Lee, David N. (Sept. 1998). "Guiding Movement by Coupling Taus". In: *Ecological Psychology* 10.3/4, p. 221. DOI: 10.1080/10407413.1998.9652683.

Mücke, Doris, Anne Hermes, and Sam Tilsen (Feb. 2020). "Incongruencies between phonological theory and phonetic measurement". en. In: *Phonology* 37.1, pp. 133–170. DOI: 10.1017/S0952675720000068. URL: https://www.cambridge.org/core/product/identifier/S0952675720000068/type/journal_article (visited on 07/20/2020).

Nam, Hosung and Elliot Saltzman (2003). "A competitive, coupled oscillator model of syllable structure". In: *Proceedings of the 15th International Congress of the Phonetic Sciences*.

Saltzman, Elliot and Kevin Munhall (Dec. 1989). "A Dynamical Approach to Gestural Patterning in Speech Production". en. In: *Ecological Psychology* 1.4, pp. 333–382.

Sorensen, Tanner and Adamantios Gafos (Oct. 2016). "The Gesture as an Autonomous Nonlinear Dynamical System". en. In: *Ecological Psychology* 28.4, pp. 188–215. DOI: 10.1080/10407413.2016.1230368. URL: https://www.tandfonline.com/doi/full/10.1080/10407413.2016.1230368 (visited on 12/15/2023).

Tiede, Mark (2005). *Mview: software for visualization and analysis of concurrently recorded movement data.*