SEPTEMBER 06 2023

How thresholding in segmentation affects the regression performance of the linear model

Stephan R. Kuberski 💿 ; Adamantios I. Gafos 💿

() Check for updates

JASA Express Lett. 3, 095202 (2023) https://doi.org/10.1121/10.0020815







Advance your science and career as a member of the Acoustical Society of America



asa.scitation.org/journal/jel

CrossMark



How thresholding in segmentation affects the regression performance of the linear model

Stephan R. Kuberski^a (D) and Adamantios I. Gafos (D) Department of Linguistics and Cognitive Sciences, University of Potsdam, Potsdam, Germany

kuberski@uni-potsdam.de, gafos@uni-potsdam.de

Abstract: Evaluating any model underlying the control of speech requires segmenting the continuous flow of speech effectors into sequences of movements. A virtually universal practice in this segmentation is to use a velocity-based threshold which identifies a movement onset or offset as the time at which the velocity of the relevant effector breaches some threshold percentage of the maximal velocity. Depending on the threshold choice, more or less of the movement's trajectory is left in for model regression. This paper makes explicit how the choice of this threshold modulates the regression performance of a dynamical model hypothesized to govern speech movements. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

[Editor: Rachel M. Theodore]

https://doi.org/10.1121/10.0020815

Received: 15 May 2023 Accepted: 16 August 2023 Published Online: 6 September 2023

1. Introduction

Modern linguistic theory and all models of speech agree that linguistic messages such as syllables or whole sentences are constructed out of combinations or sequences of discrete units (Chomsky, 2000; von Humboldt, 1836) or, in spoken languages, phonemes actuated as vocal tract gestures. Even though a sequence of units suggests that all properties of these units are located at the same point in time, the sequences generated in biological movements, be it of the vocal tract, the face, or the limbs in other activities, have an important characteristic. The units composing these sequences are coarticulated (Öhman, 1966) or interwoven in a continuous flow of smooth transitions from one unit to the next. This continuity burdens experimenters and modelers concerned with recovering the dynamics of individual units with the task of segmenting the continuous flow of effectors into sequences of movements (corresponding to the hypothesized units).

Whether speaking, pointing, running, or dancing, the continuous flow of effectors is typically captured by a number of sensors attached to the human body whose position is tracked over time by an external device. In the resulting data, the movement of a single effector is said to initiate, when the velocity (i.e., the rate of change of position) of the sensor attached to that effector exceeds some minimal value, and likewise the movement of the effector is said to end when the sensor's velocity falls below some (potentially different) minimal value (Hogan and Sternad, 2007). Accordingly, it has become standard to consider velocity as the primary source of information in carrying out the segmentation, with the virtually universal practice of using a velocity-based threshold parameter for identifying the onset or offset of a movement as the time at which the instantaneous velocity of a relevant effector breaches some threshold percentage of its maximal velocity. Notwithstanding the universality of this practice, a review of the literature indicates widespread quantitative differences regarding the value of the threshold parameter, with the consequences of these differences for model evaluation remaining, to our knowledge, so far unexplored.

In speech, an often used value for the segmentation threshold is 20% (e.g., Bombien *et al.*, 2013; Hoole *et al.*, 1994; Parrell and Narayanan, 2014; Shaw and Chen, 2019; Shaw *et al.*, 2011), with deviations from this value not uncommon. To give some examples: Pouplier *et al.* (2020) uses $20 \pm 5\%$ (as appropriate) for tongue movements in German and Georgian consonant-consonant-vowel sequences, Chitoran *et al.* (2008) makes use of 15% or 20% for tongue and lip movements in Georgian stop-stop sequences, and Lee *et al.* (2015) reduces the threshold to 10% for tongue movements in Korean liquids. Yet others (e.g., Kuberski and Gafos, 2019; Munhall *et al.*, 1985; Mücke *et al.*, 2012) use a threshold of 0% (no threshold), conceivably, to not cut away relevant information contained in the regions near the movement endpoints. For cogent discussion of potential problems with the threshold method and a proposal for using different sources of information than velocity for the process of segmentation, see Liu *et al.* (2022).

A similar variety in the choice of thresholds is evident in the fields of general motor control. For example, Hug (2011) calls a range of 15%–25% (or 1, 2, or 3 standard deviations) typical for onset detection in kinematic muscle activity data (EMG) and Baum and Li (2003) makes use of 10% (alternatively, 20% if more appropriate after visual inspection) in

^{a)}Author to whom correspondence should be addressed.



EMG data of the lower extremities in cycling. Coderre *et al.* (2010) mentions 5% to be typical for arm movements in reaching tasks and Buchanan *et al.* (2003) likewise uses 5% for data of hand tapping and elbow flexion-extension motions. In a study explicitly focusing on transitions between individual point-to-point arm movements, Sternad *et al.* (2013) considered threshold values of 1%, 3%, and 5%.

Despite the variety in segmentation practices, with threshold values ranging from 0% to 25% within and across fields, what is undoubtedly common in all these studies is the ultimate aim of understanding the underlying model controlling the so-segmented movements.¹ It is to this aim that our presented study is devoted. More specifically, our aim is to make explicit how the performance of the dynamical model in Task Dynamics (Saltzman, 1986; Saltzman and Kelso, 1987; Saltzman and Munhall, 1989) depends on the threshold value used in segmentation. Among other dynamical model candidates (see, e.g., Parrell et al., 2019, for an overview), we chose here the model proposed in Task Dynamics, which is that of the damped linear oscillator, $\ddot{x} = -k(x - x_0) - b\dot{x}$, because of its explicitness on how to relate movement data to model parameters and because of its time honored presence in the field of speech. As the first explicit dynamical formulation seeking to characterize some of the most important general aspects of the control and coordination of movements (e.g., task-specific invariance of effector trajectories, equifinality of movements, self-adjustment under perturbation; Fowler et al., 1980; Kelso et al., 1983; Saltzman and Kelso, 1987), the linear model is still considered a plausible candidate in increasing our understanding of speech. Recently, for example, on the basis of high-density intracranial electrocorticography, (ECoG) signals (Chartier et al., 2018, p. 1048) estimated speech effector positions from the recorded neural activity in the human sensorimotor cortex and, by considering characteristics of the estimated data in the position-velocity plane (also known as phase plane, see Sec. 4) as well as the linearity in the peak velocity-displacement relation concluded that the so-estimated movements exhibited damped oscillatory dynamics in accordance with the linear model in Task Dynamics.

In the following three sections, we first describe our approach of registering articulatory movement data (Sec. 2) to be segmented by six different threshold values (0%, 5%, 10%, 15%, 20%, and 25%) drawn from the entire range of values found in the literature. We then turn to a statistical assessment (Sec. 3) of the linear model's fit to the so-segmented data. Finally, we explain (Sec. 4) why thresholding modulates model performance in the ways revealed in our results, by considering dynamical tools of analysis, and outline implications for future work.

2. Methods

Five native speakers of German (three female, two male) and five native speakers of English (three female, two male) were recruited at the authors' institution to participate in an experiment of repeated syllable production. The speakers produced repeated /ka/ and /ta/ syllables at a rate indicated by the beats of an audible metronome. In blocks of four trials for each rate, the speakers were exposed to the beat of a metronome (with a duration allowing for the production of 30 consecutive syllables) and begun articulation at a point of their choice. The rate of the metronome was set to 90, 150, 210, 300, 390, and 480 beats per minute (bpm), covering slow, normal, and fast speech rates. Data registration was conducted using Electromagnetic Articulography (EMA; Carstens AG501, Carstens Medizinelektronik GmbH, Bovenden, Germany). Relevant for the purposes of this work were the movements (after head movement correction) of a sensor placed on the tongue body and a sensor placed on the tip of the tongue, the two primary effectors involved in the formation and release of constrictions in /ka/ and /ta/ syllables. The registered raw data were processed using a heptic-order spline smoothing approach with a fixed predicted mean square error, obtained by a FORTRAN-to-MATLAB software port of Woltring's classical spline smoothing code (Kuberski, 2023; Woltring, 1986).

The resulting tongue body (for /ka/) and tongue tip (for /ta/) signals were then segmented into individual movements by means of local tangential velocity extrema identified using the full three dimensions. As per its definition, tangential velocity was computed by the square root of the sum of squares of the horizontal, vertical, and lateral velocity components in three dimensions. Timestamps at which the tangential velocity of a relevant effector was minimal served as (pre-threshold) locations for a movement's endpoints. The maximal value of tangential velocity between any two consecutive minima was next used to apply the segmentation threshold and adjust the threshold-specific onset and offset locations accordingly. In particular, the difference between the maximal and minimal value multiplied by the threshold value was used as the offset in velocity an effector had to breach (exceed or fall below) to be regarded as the onset or offset of movement. Overall, for each considered threshold value, the segmentation process yielded speech movement data of about 5000 /ka/ and 5000 /ta/ syllables, across all speakers and metronome rates.

In the context of our ultimate aim, which is dynamical model evaluation, it is important to note that the soobtained movement data represented motion in three-dimensional space. Yet the formulation of the linear model takes place in a lower-dimensional so-called task space, consisting of only one dimension along the major line of action, referred to as constriction degree (Saltzman, 1986; Saltzman and Munhall, 1989) or reaching axis (Saltzman and Kelso, 1987).² Thus, to evaluate the performance of the model, we projected the segmented three-dimensional effector trajectories onto their individual constriction degree axes as indicated by the large gray arrows in Fig. 1. In accordance with the framework of Task Dynamics in which the linear model is formulated, we determined the constriction degree axes (i.e., the major lines of action) by the end point-to-end point vectors of the movements (*ibidem*).



Fig. 1. A midsagittal superposition of the movements of two sensors attached to the tongue body and tongue tip, the two effectors implicated in the production of two sequences of /ka/ and /ta/ syllables, produced by one of our speakers at slow speech rate (150 bpm metronome). Shown are two task-spatial coordinate systems exemplifying the major lines of action along the tongue body (TB, for /ka/) and tongue tip (TT, for /ta/) constriction degree axes. Spatial projection of the three-dimensional movement data onto these axes (indicated by the large gray arrows) yields a one-dimensional representation of effector motion as required by the formulation of any one-dimensional model such as that in Task Dynamics.

3. Results

We now turn to evaluate the linear model, $\ddot{x} = f(x, \dot{x}; x_0)$, where $f(x, \dot{x}; x_0) = -k(x - x_0) - b\dot{x}$, with an eye to assessing how different segmentation thresholds affect the regression performance of that model. To do so, we first obtained time series for each movement in position x_i , velocity \dot{x}_i , and acceleration \ddot{x}_i , with i = 1, ..., N, using N = 250 Chebyshev nodes in time. Chebyshev sampling (e.g., Mason and Handscomb, 2002; Rivlin, 2020) is typically used in polynomial regressions to minimize the effect of Runge's phenomenon because of its characteristic, nonuniform spacing of samples with an increased resolution at the edges of the sampling interval. In addition, the use of Chebyshev nodes yields time series approximately evenly spaced along the phase space trajectories of movements to be regressed.

Fitting data to the linear model requires choosing a value for the parameter x_0 in the equation of that model; this x_0 represents the target of the movement. For plosive consonants, like [k] and [t], it is relatively uncontroversial that effectors such as the tongue body and tongue tip must be controlled by some dynamical system responsible for the formation of their respective constrictions or so-called closing movements (from the previous vowel to the consonant). For these closing movements, it is also relatively uncontroversial to set the target parameter x_0 to these movements' offset position x_N , so that the overall regression model for each movement reads (with the remaining parameters, stiffness k and damping b, let free to vary)

$$\begin{pmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \vdots \\ \ddot{x}_N \end{pmatrix} = -k \begin{pmatrix} x_1 - x_N \\ x_2 - x_N \\ \vdots \\ x_N - x_N \end{pmatrix} - b \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_N \end{pmatrix}.$$

Though it is conceivable that the same dynamical model considered here may be in control also for opening movements (from the consonant to the next vowel), when it comes to vowels, there are approaches to the notion of target which use combinations of orosensory parameters or (also) acoustic dimensions (e.g., Guenther, 1995). In other words, for these opening movements, the task may not be specified (solely) in terms of body space positional coordinates. Hence, we chose to focus our assessment of the linear model to closing movements only, because of the relatively uncontroversial position that if the linear model applies to speech movements, then it ought to apply to (at least) formations of the closing movements for plosives, and because of the equally relatively uncontroversial assumption about what constitutes a target for these movements.

In applying a conventional statistical measure for the performance of a regression, we computed R-squared values of the regression results $\hat{f}(x, \dot{x}; x_0)$ via $R^2 = 1 - RSS/TSS$, with the residual sum of squares $RSS = \sum_i^N [\ddot{x}_i - \hat{f}(x_i, \dot{x}_i; x_0)]^2$ and the total sum of squares $TSS = \sum_i^N [\ddot{x}_i - \overline{\ddot{x}}]^2$, where $\ddot{\bar{x}} = 1/N \sum_i^N \ddot{x}_i$ was the average of the response variable (in our case, acceleration). Grand mean performances of the ten speakers were computed by first averaging the different experimental conditions for each speaker, then averaging across the entire group of speakers.³

Let us first evaluate the overall regression performance of the linear model by inspection of the so-obtained grand mean performance values given in Table 1. The table lists average R-squared values for both syllables /ka/ and /ta/,



/ka/	0%	5%	10%	15%	20%	25%
90 bpm	0.427	0.478	0.512	0.546	0.583	0.624
150 bpm	0.526	0.574	0.606	0.637	0.667	0.699
210 bpm	0.655	0.690	0.711	0.733	0.756	0.780
300 bpm	0.784	0.822	0.841	0.858	0.875	0.890
390 bpm	0.799	0.851	0.875	0.894	0.911	0.926
480 bpm	0.779	0.843	0.871	0.894	0.913	0.929
All rates	0.682	0.737	0.766	0.792	0.817	0.841
/ta/	0%	5%	10%	15%	20%	25%
90 bpm	0.477	0.537	0.569	0.597	0.625	0.668
150 bpm	0.636	0.665	0.688	0.712	0.737	0.763
210 bpm	0.715	0.760	0.786	0.809	0.830	0.851
300 bpm	0.793	0.835	0.857	0.877	0.894	0.909
390 bpm	0.807	0.848	0.869	0.887	0.903	0.918
480 bpm	0.763	0.818	0.845	0.868	0.889	0.907
All rates	0.713	0.761	0.788	0.812	0.834	0.856

Table 1. R-squared values of the linear model regression of /ka/ and /ta/ syllables separated by segmentation thresholds (columns) and experimental conditions (metronome rate, rows).

separated by segmentation thresholds (columns) and experimental conditions (metronome rates, rows). It turns out that, across all experimental conditions, regressions of the linear model performed fairly well, with R-squared values in the range of 0.682 to 0.841 for /ka/ and between 0.713 and 0.856 for /ta/ syllables (see the final row in Table 1). Interpretation of these performance ranges as proportions of the explained variance yields ranges of about 68%–84% for /ka/ and 71%–86% for /ta/ syllables, both indicating that a considerable amount of variance in the empirical data is explained by the linear model.

We focus next on the individual segmentation thresholds expressed by the different columns of Table 1. Inspection of the R-squared values in the table makes clear that, for any fixed experimental condition (metronome rate, rows), increases in the threshold value (i.e., moving rightwards in the table) come along with increases in regression performance. In other words, there is a systematic gain in the regression performance of the linear model as a function of increasing the segmentation threshold. In a way that would speak to researchers (ultimately) interested in assessing their dynamical model assumptions given speech movement data, we thus proceeded to evaluate this gain by comparing R-squared values of the different thresholds in relation to the pairwise increments in their values. Specifically, we aimed at providing an answer to the question: to what extent and significance does increasing the segmentation threshold by 5% (e.g., using 25% instead of 20%, or 20% instead of 15%, etc.), or any other such percentage, affect the performance of the hypothesized underlying model? Figure 2 visualizes the answers to this question by showing the two characteristics of



Fig. 2. Performance gain in regressions of the linear model (right ordinates, in percentage point of the explained variance) and their statistical significance (left ordinates, in standard *z*-score) as a function of metronome rate (abscissae) and increments in segmentation thresholding (0%, 5%, 10%, 15%, 20%, and 25%, color-coded). The *z*-scores larger than 1.96 (dotted lines) indicate statistical significance at an alpha-level of 0.05 (95% confidence).

concern as a function of all possible threshold increments in our study (0%, 5%, 10%, 15%, 20%, and 25%, color-coded) as well as all experimental conditions (metronome rate, abscissae). In particular, these two characteristics are, first, the amount of performance gain in percentage point of the explained variance (right ordinates) and second, the statistical significance of the gain in standard z-score (left ordinates).⁴ To give an example of the observed effect of thresholding on model performance, let us chose a critical z-score of 1.96 for an alpha-level of 0.05 (i.e., 95% confidence) indicated by the dotted lines in Fig. 2. For this conventional choice, increments in the segmentation threshold larger than about 15% (e.g., choosing a threshold of 15% over one of 0%, but also 20% over 0%, or 25% over 0%) generally lead to a significant difference in the regression performance. This can be inferred from the figure by inspecting which of the differently colorcoded threshold contours tend to reside above the critical z-score, which are the three darkest contour lines associated with the three highest threshold increments of 15%, 20%, and 25%. Likewise, the performance gains of these three highest threshold increments reach substantial values: expressed in proportions of the explained variance, they range from about 3% to 24% point. That is, a regression performance of (for example) 50% explained variance, determined by a 0%threshold segmentation, can be significantly boosted up to about 75% explained variance just by increasing the threshold to 25%. This observation holds true for both syllables (/ka/ and /ta/) and all experimental conditions (metronome rates). Thus, overall, there is no question that changes from a lower to a higher threshold come with an increase in regression performance. Moreover, the larger the threshold increment is, the larger is the gain in model performance as well as its statistical significance.

4. Discussion

Any evaluation of a dynamical model underlying the control and organization of movements requires segmenting a continuous flow of some effector into a sequence of movements. Since the beginnings of the availability of techniques like xray, magnetometer, etc., for obtaining quantitative records of articulatory motion, a virtually universal practice in speech is to use a velocity-based threshold which enables the researcher to declare the occurrence of a movement onset or offset as the first time at which the velocity of the relevant effector breaches some threshold percentage of the maximal velocity. Depending on the value of the threshold, more or less of the movement's trajectory is left in for fitting to the model, which is assumed to be underlying. Here, using one dynamical model, the linear model, hypothesized to govern speech, we have made explicit how the choice of this threshold modulates the resulting regression performance.

The systematicity in the patterning of our results (significant gains in the regression performance when using higher and higher segmentation thresholds) demands an explanation. An understanding of this systematicity can be obtained by considering what information in the movement trajectories is eliminated as one progressively increases the segmentation threshold. Figure 3 shows an example of unsegmented data of the tongue body effector during production of three consecutive /ka/ syllables at normal speech rate. The left panel of the figure shows the so-called phase plane of the data, that is, the trajectory of the effector's velocity \dot{x} as a function of its position x. In this representation, positive peak positions of the trajectory correspond to the consonant closure /k/ and negative peaks to the maximum opening for the vowel /a/. The arrows in the panel denote the direction of motion (time) in the plane. Thus, one cycle in the phase plane corresponds to one syllable in the repetition of many, with the closing movements (from /a/ to /k/) in the top and the opening movements (from /k/ to /a/) in the bottom half plane. In the same panel, we superimposed indications of how the trajectory data are affected by the different thresholds: regions shaded in gray cover those parts of the trajectory which



Fig. 3. A typical sequence of /ka/ syllables, produced by one of our speakers at normal speech rate (210 bpm metronome), portrayed in the phase plane (left) and in the Hooke plane (right). Arrows indicate the direction of time (closings: from /a/ to /k/, openings: from /k/ to /a/). Regions shaded in gray schematically indicate those parts of the data which are affected by velocity thresholding: the darker the shade (the higher the threshold), the larger the amount of data excluded from the trajectory.

are cut off during the segmentations process, with darker shades corresponding to higher thresholds (bigger cuts) and fainter shades to lower thresholds (smaller cuts). This illustration seems to suggest that the different segmentation thresholds can have only a marginal effect on the performance of the linear model: no matter the amount of data excised by a particular threshold, the overall shape of the movements remains intact, which appears to predict comparable regression performances across different thresholds.

However, this picture changes entirely, if one considers another diagnostic regularly entertained in the fields of dynamics. The right panel of Fig. 3 shows data of the same three /ka/ syllables in the so-called Hooke plane, a representation which renders the effector's acceleration \ddot{x} as a function of its position x. In this panel, as in the phase plane panel before, positive peak positions of the trajectory correspond to the consonant closure /k/ and negative peaks to the maximum opening for the vowel /a/, with arrows indicating the direction of time. Representing the data in the Hooke plane makes apparent a clear non-linear relation between the values of acceleration and position; the shape of the Hooke trajectory is similar to that of the letter N, thus, strongly indicating the presence of a cubic term (in x) in the underlying dynamics. [Note that the linear model, $\ddot{x} = -k(x - x_0) - b\dot{x}$, by definition, renders acceleration \ddot{x} as a linear function in position x.] Comparable N-shaped Hooke trajectories have been repeatedly observed in the past, in speech and non-speech movements, and brought into focus in, for example, Kelso et al. (1985), Kelso (1986), Mottet and Bootsma (1999), Buchanan et al. (2003), and Sorensen and Gafos (2016), all of which eventually considered additional, non-linear terms in the dynamical model governing these movements. Crucially here, the indicators of non-linearity (the two bends in the Hooke trajectory, one close to the /k/ and another one close to the /a/ target), brought out in the Hooke plane but not in the phase plane, are located precisely in those (gray-shaded) regions eliminated by segmentation. This, in turn, explains the significant gain in the regression performance of the linear model when using higher and higher segmentation thresholds: the higher the threshold (i.e., the darker the shade in the figure), the greater the non-linear portion of the Hooke trajectory excluded by segmentation, yielding an almost exact linear relation between acceleration and position at the highest threshold of 25%. It is precisely at this value of threshold where the linear model reaches maximal performance (cf. Table 1, final column).

Overall, our results and their attendant explanation disclose the extent to which thresholding has an effect on the regression performance of the linear model hypothesized to govern speech in Task Dynamics. The present approach can be applied in principle to other experimental speech paradigms as well as to other models that are sufficiently explicit to allow for evaluation using movement data. In the analysis of our repetitive speech movement data, we applied a conventional to non-repetitive speech segmentation procedure; thus, it is likely that the results presented here would generalize to data from non-repetitive tasks. Regarding extending our approach beyond the linear model, other promising models for the control and organization of human movements (e.g., Birkholz et al., 2011; Sorensen and Gafos, 2016, in speech; Beek et al., 1995; Huys et al., 2008; Jirsa and Kelso, 2005; Kay et al., 1987; Mottet and Bootsma, 1999; Schöner, 1990, in non-speech motor control) should be considered. Common to these alternative models is that they extend the linear model equation in Task Dynamics to more advanced equations by inclusion of additional, non-linear terms. For example, in speech, Sorensen and Gafos (2016) propose to include the cubic term $(x - x_0)^3$ to the model equation, improving conformity between model predictions and kinematic properties of observed speech movements and resulting in a more accurate modeling of trajectory shapes in the Hooke plane. In general motor control, Schöner (1990) includes several non-linear (both in x and \dot{x}) terms which likewise improve model predictions but also allow one to generate movements governed by qualitatively distinct dynamics (not just fixed point but also limit cycle dynamics). It remains to be seen how such revised models that admit non-linearities perform under different segmentation thresholds.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Project ID 317633480, SFB 1287, Project C04. Data collection has been partially funded by the European Research Council (ERC) under Grant agreement ID 249440.

References

¹Though the aforementioned studies are not necessarily concerned with evaluating a specific dynamical model, any study of gestures and their spatio-temporal organization employs movement data assumed to be the surface consequences of an underlying dynamical model giving rise to that data. This means that quantified landmarks (e.g., gestural onset or target) or intervals (e.g., time to peak velocity, closing phase duration), as widely employed in all these studies, depend crucially on the nature of the model assumed to govern the observed movements. Hence, even though not explicitly stated or assessed in these works, the model underlying the quantified aspects of the movements in these studies is essential. It is in this sense that we consider an evaluation of the relation between segmentation practice and underlying model performance to be of key importance and ultimately of relevance to these studies.

²The linear model is hypothesized to control change also in another task space dimension (orthogonal to that of the constriction degree as depicted in Fig. 1) and referred to as constriction location. This paper assesses the effect of thresholding only for the constriction degree task space variable.

 3 Averaging correlations is not as straightforward as averaging other statistical quantities. This is because of the skewed sample distribution of *R*'s, generally leading to an underestimation of the ground truth population mean (the higher the R-values, the more so). Taking into



account the skewed statistics of correlations, statistical advice recommends using Fisher's *z*-transformation when averaging correlations (e.g., Corey *et al.*, 1998; Silver and Dunlap, 1987). The *z*-transformation is given by $z = \operatorname{atanh} R$, with a back-transformation of $R = \operatorname{tanh} z$ (Fisher, 1921). Following this practice, we transformed R-values of the individual regressions into *z*-coordinates before computing their average and back-transformed the result to obtain an unbiased estimate of the mean.

⁴In comparing correlation performance, Fisher (1992) recommends a significance test for the difference of two correlations, *R* and *S*, of the same underlying sample size *N*, by means of the *z*-score (atanh R – atanh S)/ $\sqrt{2/(N-3)}$. This *z*-score is normal with unit variance and can be read as the result of a standard *z*-test: e.g., *z*-score values larger than 1.96 indicate significance at the 0.05 level (95% confidence), etc. Similarly, using the same transformation, one can obtain an unbiased estimate of the difference between two correlations via tanh(atanh *R* – atanh *S*). The square of this expression quantifies the amount of difference in the proportions of the explained variance of the two correlations, that is, the amount of performance gain between *R* and *S*.

- Baum, B. S., and Li, L. (2003). "Lower extremity muscle activities during cycling are influenced by load and frequency," J. Electromyogr. Kinesiol. 13(2), 181–190.
- Beek, P. J., Schmidt, R. C., Morris, A. W., Sim, M.-Y., and Turvey, M. T. (1995). "Linear and nonlinear stiffness and friction in biological rhythmic movements," Biol. Cybernet. 73(6), 499–507.
- Birkholz, P., Kröger, B. J., and Neuschaefer-Rube, C. (2011). "Model-based reproduction of articulatory trajectories for consonant-vowel sequences," IEEE Trans. Audio. Speech. Lang. Process. 19(5), 1422–1433.
- Bombien, L., Mooshammer, C., and Hoole, P. (2013). "Articulatory coordination in word-initial clusters of German," J. Phon. 41(6), 546-561.
- Buchanan, J. J., Park, J.-H., Ryu, Y. U., and Shea, C. H. (2003). "Discrete and cyclical units of action in a mixed target pair aiming task," Exp. Brain Res. 150, 473–489.
- Chartier, J., Anumanchipalli, G. K., Johnson, K., and Chang, E. F. (2018). "Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex," Neuron 98(5), 1042–1054.
- Chitoran, I., Goldstein, L. M., and Byrd, D. (**2008**). "Gestural overlap and recoverability: Articulatory evidence from Georgian," in *Laboratory Phonology 7* (De Gruyter Mouton, Berlin, Germany).
- Chomsky, N. (2000). New Horizons in the Study of Language and Mind (Cambridge University Press, Cambridge, UK).
- Coderre, A. M., Zeid, A. A., Dukelow, S. P., Demmer, M. J., Moore, K. D., Demers, M. J., Bretzke, H., Herter, T. M., Glasgow, J. I., Norman, K. E., Bagg, S. D., and Scott, S. H. (2010). "Assessment of upper-limb sensorimotor function of subacute stroke patients using visually guided reaching," Neurorehabil. Neural Repair 24(6), 528–541.
- Corey, D. M., Dunlap, W. P., and Burke, M. J. (1998). "Averaging correlations: expected values and bias in combined Pearson rs and Fisher's z transformations," J. Gen. Psychol. 125(3), 245–261.
- Fisher, R. A. (1921). "On the 'probable error' of a coefficient of correlation deduced from a small sample," Metron 1, 3-32.
- Fisher, R. A. (**1992**). "Statistical methods for research workers," in *Breakthroughs in Statistics: Methodology and Distribution*, edited by S. Kotz and N. L. Johnson (Springer, New York).
- Fowler, C. A., Rubin, P., Remez, R. E., and Turvey, M. T. (**1980**). "Implications for speech production of a general theory of action," in *Language Production*, edited by B. Butterworth (Academic Press, London).
- Guenther, F. H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. 102(3), 594–621.
- Hogan, N., and Sternad, D. (2007). "On rhythmic and discrete movements: Reflections, definitions and implications for motor control," Exp. Brain Res. 181(1), 13–30.
- Hoole, P., Mooshammer, C., and Tillmann, H. G. (1994). "Kinematic analysis of vowel production in German," in *Third International Conference on Spoken Language Processing*, September 18–22, Yokohama, Japan.
- Hug, F. (2011). "Can muscle coordination be precisely studied by surface electromyography?," J. Electromyogr. Kinesiol. 21(1), 1–12.
- Huys, R., Studenkaa, B. E., Rheaume, N. L., Zelaznik, H. N., and Jirsa, V. K. (2008). "Distinct timing mechanisms produce discrete and continuous movements," PLoS Comput. Biol. 4(4), e1000061.
- Jirsa, V., and Kelso, J. A. S. (2005). "The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation," J. Motor Behav. 37(1), 35–51.
- Kay, B. A., Kelso, J. A. S., Saltzman, E. L., and Schöner, G. (1987). "Space-time behavior of single and bimanual rhythmical movements: Data and limit cycle model," J. Exp. Psychol. Human Percept. Perform. 13(2), 178–192.
- Kelso, J. A. S. (1986). "Pattern formation in speech and limb movements involving many degrees of freedom," in *Generation and Modulation* of Action Patterns, edited by H. Heuer and C. Fromm (Springer, Berlin-Heidelberg).
- Kelso, J. A. S., Tuller, B., and Harris, K. S. (1983). "A 'dynamic pattern' perspective on the control and coordination of movement," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York).
- Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., and Kay, B. (1985). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling," J. Acoust. Soc. Am. 77(1), 266–280.

Kuberski, S. R. (2023). "GCV spline smoothing in Matlab," github.com/srkuberski/gcvspl (Last viewed August 30, 2023).

Kuberski, S. R., and Gafos, A. I. (2019). "The speed-curvature power law in tongue movements of repetitive speech," PLoS One 14(3), e0213851.

Lee, Y.-J., Goldstein, L. M., and Narayanan, S. S. (2015). "Systematic variation in the articulation of the Korean liquid across prosodic positions," in *Proceedings of ICPhS*, August 10–14, Glasgow, UK.

Liu, Z., Xu, Y., and Hsieh, F.-F. (2022). "Coarticulation as synchronised CV co-onset—Parallel evidence from articulation and acoustics," J. Phon. 90, 101116.

Mason, J. C., and Handscomb, D. C. (2002). Chebyshev Polynomials (Chapman and Hall, New York).

Mottet, D., and Bootsma, R. J. (1999). "The dynamics of goal-directed rhythmical aiming," Biol. Cybern. 80(4), 235-245.





- Mücke, D., Nam, H., Hermes, A., and Goldstein, L. (2012). "Coupling of tone and constriction gestures in pitch accents," in *Consonant Clusters and Structural Complexity*, edited by P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, and B. Kühnert (De Gruyter Mouton, Berlin).
- Munhall, K. G., Ostry, D. J., and Parush, A. (1985). "Characteristics of velocity profiles of speech movements," J. Exp. Psychol. Human Percept. Perform. 11(4), 457–474.

Öhman, S. E. (1966). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. 39(1), 151-168.

Parrell, B., Lammert, A. C., Ciccarelli, G., and Quatieri, T. F. (2019). "Current models of speech motor control: A control-theoretic overview of architectures and properties," J. Acoust. Soc. Am. 145(3), 1456–1481.

- Parrell, B., and Narayanan, S. S. (2014). "Interaction between general prosodic factors and languagespecific articulatory patterns underlies divergent outcomes of coronal stop reduction," in *Proceedings of the International Seminar on Speech Production (ISSP)*, May 5–8, Cologne, Germany.
- Pouplier, M., Lentz, T., Chitoran, I., and Hoole, P. (2020). "The imitation of coarticulatory timing patterns in consonant clusters for phonotactically familiar and unfamiliar sequences," J. Lab. Phonol. 11(1), 1.

Rivlin, T. A. (2020). Chebyshev Polynomials (Dover, New York).

Saltzman, E. L. (1986). "Task dynamic coordination of the speech articulators: A preliminary model," in *Generation and Modulation of Action Patterns*, edited by H. Heuer and C. Fromm (Springer, Berlin).

Saltzman, E., and Kelso, J. A. (1987). "Skilled actions: A task-dynamic approach," Psychol. Rev. 94(1), 84-106.

Saltzman, E., and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecol. Psychol. **1**(4), 333–382. Schöner, G. (**1990**). "A dynamic theory of coordination of discrete movement," Biol. Cybernet. **63**, 257–270.

Shaw, J. A., and Chen, W.-R. (2019). "Spatially conditioned speech timing: Evidence and implications," Front. Psychol. 10, 2726.

- Shaw, J. A., Gafos, A. I., Hoole, P., and 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.
- Sorensen, T., and Gafos, A. I. (2016). "The gesture as an autonomous nonlinear dynamical system," Ecol. Psychol. 28(4), 188–215.
- Silver, N. C., and Dunlap, W. P. (1987). "Averaging correlation coefficients: should Fisher's z transformation be used?," J. Appl. Psychol. 72(1), 146–148.
- Sternad, D., Marino, H., Charles, S., Duarte, M., Dipietro, L., and Hogan, N. (2013). "Transitions between discrete and rhythmic primitives in a unimanual task," Front. Comput. Neurosci. 7, 90.
- von Humboldt, W. (1836). Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts (On the Diversity of Human Language Construction and its Influence on the Mental Development of the Human Species) (Druckerei der königlichen Akademie der Wissenschaften, Berlin, Germany).
- Woltring, H. J. (1986). "A Fortran package for generalized, cross-validatory spline smoothing and differentiation," Adv. Eng. Softw. 8(2), 104–113.