Article:

Mücke, Doris, Roessig, Simon, Thies, Tabea et al. (2 more authors) (2024) Challenges with the kinematic analysis of neurotypical and impaired speech:measures and models. Journal of Phonetics. 101292. ISSN: 0095-4470

https://doi.org/10.1016/j.wocn.2023.101292

Journal of Phonetics 102 (2024) 101292

Contents lists available at ScienceDirect

Journal of Phonetics

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

Challenges with the kinematic analysis of neurotypical and impaired speech: Measures and models

Doris Mücke [a,*], Simon Roessig [b], Tabea Thies [a,c], Anne Hermes [d], Antje Mefferd [e]

[a] IfL – Phonetics, University of Cologne, Herbert-Lewinstr. 6, 50931 Cologne Germany
[b] Department of Language and Linguistic Science, University of York, York, YO10 5DD, UK
[c] Department of Neurology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Kerpener Str. 62, 50937 Cologne, Germany
[d] Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Sorbonne Nouvelle, ILPGA, 19 rue des Bernardins, 75005 Paris, France
[e] Hearing and Speech Sciences, Vanderbilt University Medical Center, MCE South Tower, 121521st Avenue South, Nashville, TN 37232 USA

ARTICLE INFO

Article history:
Received 14 September 2022
Received in revised form 21 December 2023
Accepted 27 December 2023

Keywords:
Speech kinematics
Motor speech disorders
Modeling
Dynamical systems
Stochastic noise
Articulatory Phonology/Task Dynamics
Articulatory gestures

ABSTRACT

A common goal of kinematic studies on disordered speech is the identification of speech motor impairments that negatively impact speech function. Although it is well-known that the kinematic contours of speakers with speech disorders often deviate considerably from those of neurotypical speakers, systematic quantitative assessments of these impairment-related movement disturbances remain challenging. Kinematic measurement approaches are commonly grounded in models and theories that have emerged exclusively from observations of healthy speakers. However, often these models cannot accommodate the deviant articulatory behaviors of speakers with speech motor impairment. In the present paper, we address this problem. By considering noise as a factor in Articulatory Phonology/Task Dynamics (AP/TD), we can account for articulatory behaviors that are known to occur in healthy speakers (e.g., during slow speech) as well as in speakers with motor speech impairments. In a proof of concept, we descriptively compare modeled articulatory behaviors that include noise at various levels with empirical data. We view such an extension of the AP/TD as a first step towards a more comprehensive speech production model that can serve as a theoretical framework to study the speech production mechanism in healthy speakers and speakers with motor speech impairments.

1. Introduction

A common goal of kinematic studies on disordered speech is the identification of speech motor impairments that negatively impact speech function (e.g., articulatory precision, speaking rate, speech intelligibility). Although it is well-known that the kinematic contours of speakers with motor speech impairments often deviate considerably from those of neurotypical speakers, systematic quantitative assessments of these impairment-related movement disturbances remain challenging. One major obstacle is that traditional kinematic measurement approaches are often grounded in models and theories that have emerged exclusively from observations of healthy speakers, which often cannot accommodate the deviant articulatory behaviors of speakers with speech motor impairment. Most importantly, speech pathologists and phoneticians currently cannot fully benefit from each other's work although their overall research goals do overlap considerably.

The large research field of speech pathology deals with many different types of speech motor impairments. Commonly, these investigations seek to answer questions such as "what articulatory behaviors are essential to produce intelligible speech?", "how do articulatory behaviors change in response to various speech demands (loudness, rate, speech clarity)?", "how do articulatory behaviors change over time due to speech motor development, motor learning, and/or degenerative disease processes?". Answers to these questions are critical to an understanding of the articulatory basis of speech intelligibility, changes due to maturation, therapeutic interventions, and neurological disorders. However, when trying to answer such questions, fundamental problems arise in how to characterize articulatory behaviors. For example, one major challenge is the continuous nature of articulatory movements during running speech. In addition, articulatory behavior is also highly variable

* Corresponding author.
E-mail addresses: doris.muecke@uni-koeln.de (D. Mücke), simon.roessig@york.ac.uk (S. Roessig), tabea.thies@uni-koeln.de (T. Thies), anne.hermes@sorbonne-nouvelle.fr (A. Hermes), antje.mefferd@vanderbilt.edu (A. Mefferd).

https://doi.org/10.1016/j.wocn.2023.101292
0095-4470/© 2024 The Authors. Published by Elsevier Ltd.

across individuals. Thus, it is difficult to identify a specific segment of articulatory behavior that can serve as a basic unit of speech and fulfil two important scientific research demands: 1) well-defined boundaries that can be reliably identified in the kinematic signal of neurotypical speakers as well as speakers with speech impairment and 2) a meaningful link to perceptual measures of speech (e.g., articulatory precision and speech intelligibility).

There are very powerful speech production models for describing speech motor control (Miller & Guenther, 2021): One of the most promising solutions to handling highly variable speech kinematics across speakers is the use of dynamical systems such as Articulatory Phonology based on Task Dynamics (henceforth AP/TD).[1] By defining the articulatory gesture as the basic unit of speech production, this theory fully integrates discrete symbolic representations and the continuous representations of the speech system (Browman & Goldstein, 1992; Gafos & Benus, 2006; Goldstein et al., 2006; Mücke et al., 2020; Nam et al., 2009; Shaw et al., 2011). However, models and theories like AP/TD have been predominantly applied to the kinematic analysis of healthy speakers, and their flexibility has been recently described as being insufficient in accounting for perturbed or highly noisy speech movement contours commonly observed in speakers with speech motor impairment (Parrell & Lammert, 2019). This incompatibility is particularly problematic when the goal of the kinematic investigation is to identify articulatory deficits that negatively impact perceptual aspects of speech.

Despite the previously described shortcomings of AP/TD for the analysis of disordered speech, we would like to argue in the present paper that AP/TD can serve as a useful model in the field of motor speech disorders/clinical linguistics. One major benefit of AP/TD is that it is a clearly defined mathematical model, which provides opportunities to test theoretically-grounded hypotheses about impaired as well as typical articulatory behaviors. However, we suggest that the gestural approach of AP/TD may require additional variables to be included in the model to more accurately represent the wide variety of articulatory behaviors that can be observed across typical speakers as well as speakers with speech impairment. As a first step, we focus on stochastic noise as a newly added variable within the gestural model. Stochastic noise is described as a parameter to model random fluctuations in movement patterns in gross motor control (Haken, 1977; Tilsen, 2022). Further, it is described as being a very useful parameter in speech motor control, responsible for natural variability observed in surface patterns: “An essential characteristic of natural speech is that it is unavoidably stochastic, and as a consequence, no two utterances are identical” (Tilsen, 2022, p. 15). In neurotypical speakers as well as speakers with speech impairment, the occurrence of noise is a well-accepted phenomenon at various levels of the motor control process (e.g., noise within the central nervous system, noise during execution in the periphery; Fitts, 1954). In the data of subjects with speech impairment, the level of variability is considerably higher. The question of which role stochastic noise should play in modelling this speech behavior is hence of particular interest when dealing with impaired speech. However, the AP/TD model in its current form does not explicitly consider it.

The current study incorporates some tutorial elements regarding speech as a dynamical system. It offers insights into dynamical models to current problems in clinical linguistics and the research of neurotypical speech. We illustrate the basics of dynamic modeling in terms of a gestural approach to show how the model can be used to make predictions that are related to the observations in our kinematic signals. In a proof of concept, we will demonstrate how the addition of stochastic noise to different levels of the dynamical system improves our ability to model deviant articulatory movement patterns that are observed in impaired speech. Specifically, we will show an interpretation of the dynamical control parameters (e.g., stiffness, target, and activation interval) as inherently noisy can produce patterns that qualitatively resemble those of speakers with speech motor impairment. We will also demonstrate that the amount of noise becomes an important characteristic on its own in this version of the model. Our goal is to extend the AP/TD so that it can be used to make predictions about articulatory behavior when dealing with different degrees of variability that can be found in impaired speech. This type of prediction can help to build linguistic and clinical hypotheses and to characterize the empirical results systematically. We will also highlight and discuss several challenges with the quantification of articulatory behaviors to motivate AP/TD model extensions. We conclude with a discussion of why we think that noise at a certain level of a dynamical system model is an appropriate approach for the study of motor speech disorders as well as a wide range of articulatory behaviors in neurotypical speakers that can occur under various speech conditions.

2. The gestural model

2.1. Introducing the gestural model

One of the most intriguing questions in linguistics is how the discrete, abstract categories of sound are related to the continuous, physical manifestation of speech, such as articulatory movements and sound waves. This question lies at the heart of the debate around the relation between phonology and phonetics, a debate that has inspired the development of numerous theoretical perspectives. One of these is the framework of AP/TD, that is built upon a fundamentally dynamical description of language and speech (Iskarous, 2017; Iskarous & Pouplier, 2022). The term “dynamical” is not an empty buzzword here, but refers to the fact that AP/TD uses dynamical systems as the basic theoretical building blocks. The introduction of dynamics into the conceptualization of phonetics and phonology is a groundbreaking solution to the phonetics-phonology divide, because dynamical systems combine discreteness and continuity in a non-dualistic description. This means, that any theory that conceptualizes on the one hand continuous phonetic patterns and on the other hand discrete cognitive representations of phonological forms must define a permissible range of speech output characteristics (Gafos

[1] Another important model is the (GO)DIVA model, which has been transformed from a somatosensory/auditory model into a neural imaging model over time (e.g., Miller & Guenther, 2021). In the present paper, we aim to model kinematic contours directly in a gestural account. It is not our aim to compare the gestural model of AP/TD with the (GO) DIVA model.

& Benus, 2006; Mücke et al., 2017; Roessig, 2021; Shaw et al., 2011). In this view, many phonological processes, such as assimilation and deletion, can be viewed as inherently continuous where categorical patterns emerge at the extremes of the continua. For example, assimilation may be conceptualized as gestural overlap with infinite intermediate stages and complete overlap as an extreme case (Browman & Goldstein, 1989).

An introduction to dynamical systems is beyond the scope of this paper. Readers are referred to Fuchs (2013) for a general introduction to dynamical systems theory and to Iskarous (2017) for an introduction to the application of dynamical systems in AP/TD. We will, however, briefly describe the basic concepts to create a logical flow and cohesiveness within this paper. To get a better understanding of dynamical systems, consider the simple differential equation $\dot{x} = -kx$. We say that the variable x is the state variable of the system. The equation lets us find $\dot{x}$, the change in time of the system for a given state x. In other words, it tells us for a given x, what the value of this variable x will be in the future. If we set k, we can solve the equation and observe how the system evolves over time. Fig. 1 illustrates this. We use k = 0.33 and k = 1.0. Regardless of the initial state and how we set k, the trajectories always converge to zero. This state is the attractor of the system. The parameter k governs how quickly the system converges the attractor. k and the attractor at zero are time-invariant, discrete parameters, but the trajectory is continuous in time.

AP/TD uses a very similar, only slightly more complicated dynamical system to model what is at the core of the theory: the gesture (see equation in (1)). This system is a second-order dynamical system, and hence not only contains the change of x, namely $\dot{x}$, but also the change of $\dot{x}$, namely $\ddot{x}$. We call $\dot{x}$ the velocity of the system and $\ddot{x}$ the acceleration.

$$m\ddot{x} + b\dot{x} + k(x - T) = 0 \tag{1}$$

The behavior of this system is that of a mass-spring system: A mass m is attached to a spring with a stiffness k. Depending on the damping b, the system will exhibit fundamentally different patterns of behaviors. If $b = 0$, the system will oscillate eternally. For values of $b > 0$, the system will converge to a resting position T, which is comparable with the target position. Articulatory gestures are not well described by oscillatory movements. That is why the damping parameter b is chosen so that the system is critically damped (the damping constant is $b = \sqrt{4mk}$). Critically damped systems approach the resting position without oscillations. The time history of the critically damped system with different initial values is shown in Fig. 2 (upper panel). All parameters are constant throughout the simulations for this figure: $m = 1, k = 1, b = \sqrt{4mk} = 2, T = 5$. Therefore, all trajectories converge to a single attractor at the target T = 5, regardless of the initial state.

The lower panel of Fig. 2 shows another important aspect of the model: Each gesture is associated with an activation interval. During this interval it influences the tract variable.[2] For example, the gesture responsible for the opening of the lips in the syllable [ba] acts on the tract variable lip aperture in the interval of its activation. The activation function that turns the gesture on and off can be integrated into the gestural equation as shown in equation (2) (cf. Saltzman & Munhall, 1989; Sorensen & Gafos, 2016). Note that we use the damped anharmonic oscillator put forward by Sorensen and Gafos (2016). The anharmonic oscillator is similar to the harmonic oscillator of the original AP/TD model; however, it offers more naturalistic velocity profiles of the gestures than the “standard” dynamical system, and hence, provides a solution for a long-standing problem in the modeling details of AP/TD.

Activations may be modeled using different functions, the simplest being a step function that abruptly changes between “0 = no activation” and “1 = full activation”. Other proposals include linear or sigmoidal ramping functions. In this example, the activation function is linearly ramped in the beginning and end of the activation interval. Note that we follow Tilsen (2018) in calling the target T instead of x_0 to avoid potential confusion with the initial state of x (at t = 0).

$$m\ddot{x} + A(t)(b\dot{x} + k(x - T)) = 0 \tag{2}$$

This system describes the continuous trajectory towards an invariant linguistic goal (moving to the point attractor at T) while at the same time incorporating context dependency (the different initial states). The initial states may be conceptualized as different gestural environments, for example, the tongue is lowered from /i/ to /e/ but raised from /a/ to /e/. In this way, the system is able to describe coarticulation in an elegant way (Browman & Goldstein, 1992; Hawkins, 1992).

2.2. Using the gestural model to make predictions about articulatory behavior

One of the main advantages of a mathematical model like the gestural model is that it can be used to make predictions about articulatory behavior. These predictions help to evaluate kinematic measures, to build linguistic and clinical hypotheses, and to characterize the empirical results in a systematic way. This section briefly demonstrates one way in which the model makes predictions. It is inspired by work on the interplay of articulation and prosody. The ideas outlined here are, however, far more general and extend beyond the domain of prosody research.

Research on prosody has shown that articulatory movements are influenced to a large degree by the position in the phrase or by accentuation (Cho, 2006; Fougeron, 2001; Krivokapić et al., 2017; Mücke & Grice, 2014; Roessig & Mücke, 2019). These effects are quantifiable in spatial and temporal features of the articulatory trajectories and thus to measurable effects in the acoustic signal. In the gestural model, these observations inspired the idea that prosody may be able to influence the parameters of a gesture, as illustrated in Fig. 3. For example, a gesture may be less stiff in certain contexts, e.g., at the end of a phrase (Byrd & Saltzman, 1998). In this case, the gestural model would include a smaller value for k (Fig. 3, middle panel). Another change could be that the target T changes, such that when a pitch accent is placed on a syllable, the gestures of that syllable exhibit larger amplitudes and more extreme target positions (Fig. 3, left panel). A further modification that has been described in this context is the modification of intergestural timing. In this case, the next

[2] In AP/TD, articulatory gestures determine the motions of *tract variables* involved in organ groups, rather than the movement of individual articulators (Browman & Goldstein, 1992). The organ groups are organized in coordinative structures, being expressed by the motion of the tract variable.

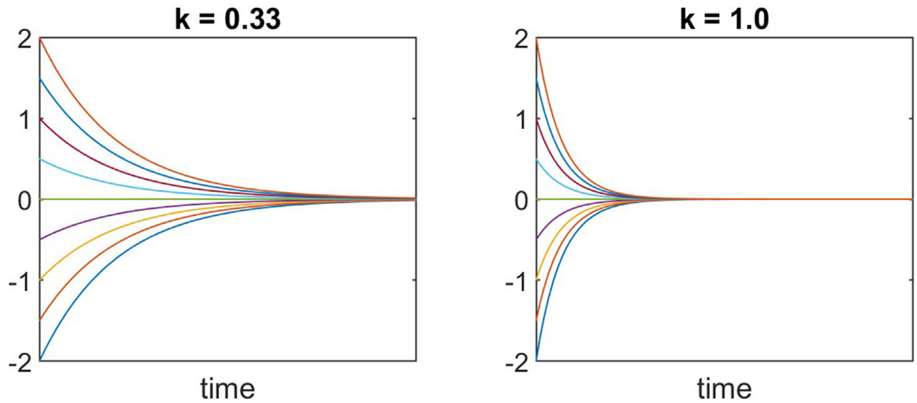

Fig. 1. Solutions over time for the simple dynamical system $\dot{x} = -kx$ with different initial states.

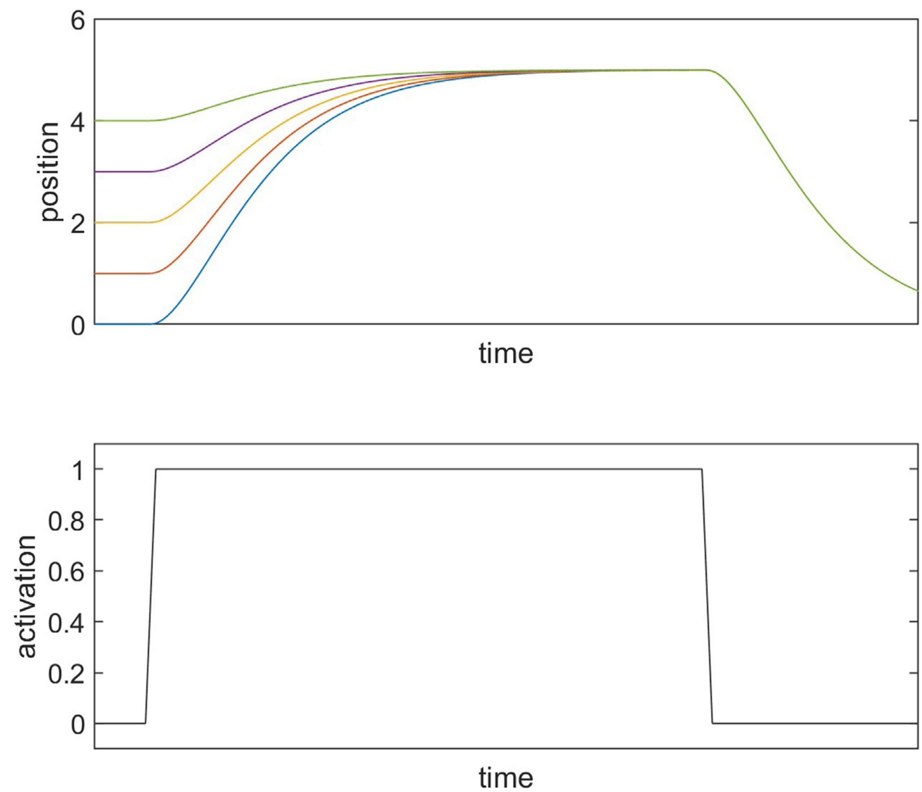

Fig. 2. Evolution of AP/TD gestures with the target $T = 5$ and different initial states (top). Activation interval for all gestures (bottom).

gesture starts earlier and truncates the gesture under investigation (Fig. 3, right panel). In terms of the AP/TD model, this change refers to the length of the gestural activation interval.[3]

In fact, gestural modifications are a critical component to understanding the articulatory basis of speech and speech changes that for example lead to intelligibility loss and gains in speakers with speech impairment. As the gestural model offers an elegant approach to separate phonetics from phonology, we argue that the basic concept of gestural modifications in typical speech can also be applied to describe impaired speech. Before we present examples of impaired speech and modeling approaches, we will illustrate the effects of parameter modifications in the model on gestures in more detail.

To better exemplify this point, let us consider how articulatory behavior changes when gestural parameters (i.e., stiffness, target, and duration of the gestural activation interval) are manipulated. The model provides a clear, structured way to test the predictions by using simulations. Target modifications lead to a more or less extreme target position (or displacement of the gesture). Increases in stiffness lead to an

[3] To modify the parameter specifications of a set of articulatory gestures in the temporal and/or spatial domain, Articulatory Phonology uses an approach involving abstract gestures modulating temporal and spatial properties of co-active consonantal and vocalic gestures in the domain where they are active (Byrd & Saltzman, 2003; Saltzman et al., 2008). Modulation gestures are not directly related to any specific vocal tract actions that are typical for the production of consonants and vowels. While the *prosodic gesture* (π-gesture) affects vocal tract gestures at prosodic boundaries, the *modulation gesture* (μ-gesture) modifies articulation under the influence of lexical stress.

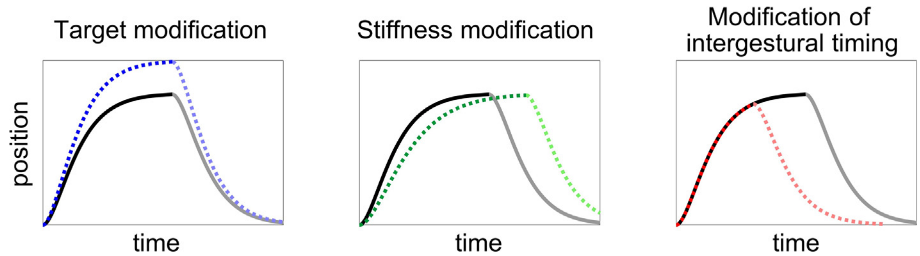

Fig. 3. Gestural modifications proposed for prosodic strengthening (cf. Cho, 2006). The dotted line shows the respective modifications: a more extreme target (left), lower stiffness (middle) and truncation (right).

earlier achievement of the target while decreases in stiffness result in a later achievement of the target. Consequently, durations from the trajectory onset to the target are shorter or longer, respectively. Lastly, for intergestural timing modification, the model suggests changes in duration as well as target position (Fig. 3, right panel). To differentiate between the specific parameter modifications, it is important to consider the target position and the duration between onset and target in context with each other rather than in isolation. Manipulating relevant parameters in the model can help us to better reflect the movement characteristics of the empirical data on the one hand. On the other hand, we can also predict empirical findings based on our knowledge about the impact of specific model parameters.

But how can parameter modifications be tested in a simulation process? In what follows, we present simulations in which a range of values for each of the three parameters mentioned above (stiffness, target, and activation duration) is tested. For each of the parameters, we tested 21 successive values on an increasing continuum. The code can be freely accessed on OSF (https://osf.io/qzcsf/?view_only=6f230fecbbe6438b85c 2bcbdc04d20f3). The procedure documents the effect of the parameter change on the following measurements:

All of these measurements in Fig. 4 are commonly used in articulatory research by phoneticians and speech pathologists when identifying a movement segment such as a bilabial closure of the lips during the production of /p/ in sequences like /apa/ or /ipi/. The annotation of the kinematic speech curves by using single landmarks is in a broader sense equivalent to a process of segmentation. At least in theory, single movement segments can be captured by the onset and target of the movement, the peak velocity and the related physical descriptions of displacement, duration, acceleration phase, and the ratio of displacement and peak velocity. The *peak velocity* refers to the maximum instantaneous velocity of the articulator while moving towards a target. The *displacement* refers to the movement amplitude between the onset and target of the movement, and it captures the way the articulator travels. The *time-to-peak velocity* refers to the time from the onset to the peak velocity of a movement. The *duration* refers to the temporal interval from the onset to the target of the movement, and it captures the time the articulator travels. The *ratio of peak velocity/displacement* captures the movement amplitude normalized by peak velocity. It is the relative speed of the movement (Munhall et al., 1985). In the AP/TD framework, the time *time-to-peak velocity* refers to the oscillation frequency of an articulatory movement, i.e., the stiffness of a gesture (Hawkins, 1992).[4]

Figs. 5–7 show the simulation results of gestures with varying values for the parameters target, stiffness, and the duration of the activation interval. The individual panels show how peak velocity, displacement (movement amplitude), time-to-peak velocity (from movement onset), duration, and the ratio of peak velocity / displacement change in response to each parameter manipulation. Note that the offset of the gesture (i.e., the “target” that can be measured) is defined as the point where velocity falls under threshold of 0.05. In the case of gestural activation, we manipulate the length of the activation interval by manipulating the end of it; the start remains constant.[5]

The simulation results are summarized in Table 1. This table lets us evaluate both directions: First, from parameter modifications to measurement changes; second, from measurement changes to parameter changes. When we choose the direction “parameter modifications to measurement changes”, we see that each parameter modification is associated with a unique set of changes in the measurements. When we choose the opposite direction, “measurement changes to parameter modifications”, there is no one-to-one relation between the measurement and the underlying parameter change. For example, the *duration* and *ratio of peak velocity/displacement* are both changed by manipulations of stiffness and activation interval duration. This observation is particularly interesting when comparing *time-to-peak velocity* to *ratio of peak velocity/displacement* since both have been proposed as assessments of stiffness. The model simulations here show that, in the context of the gestural model, *time-to-peak velocity* may perform as a better estimation of stiffness. This is because a change in *ratio displacement/peak velocity* may be caused by a change in stiffness or, alternatively, an earlier truncation of the gesture by a following gesture, i.e., a shorter gestural activation interval in the simulation.

[4] Stiffness is reported as an important dynamical control parameter for several temporal and prosodic modifications in speech. A decrease in stiffness is related to a decrease in the *global* articulation rate. Furthermore, stiffness is reported as a key property in prominence marking by initiating a *local* slowing down within the accented syllable (de Jong, 1995; Mücke & Grice, 2014) or to mark lexical stress (Beckman et al., 1992). In addition, Ackermann et al. (1995) suggested that stiffness is an important factor to differentiate between different types of speech disorders. And, last but not least, Roon et al. (2021) suggested stiffness as a predictor for overlap in consonant clusters.

[5] We tested 21 sequential values for each parameter: (i) Stiffness: [500, …, 1000], (ii) Target: [0.7, …, 1], and (iii) Activation end point: [0.15, …, 0.40]. For the gestural activation, we use a ramped function with sigmoidal increase / decrease of activation.

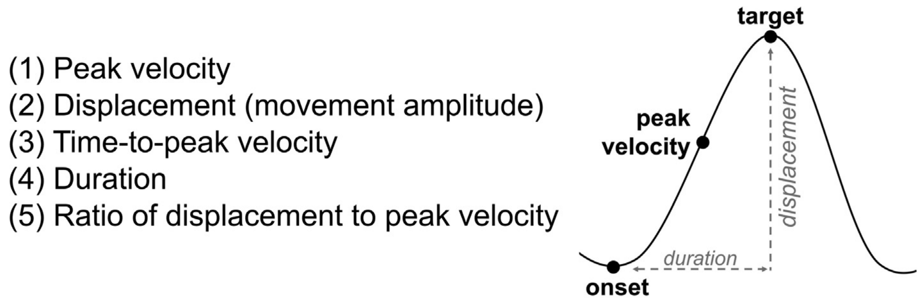

Fig. 4. Measurements for kinematic contours.

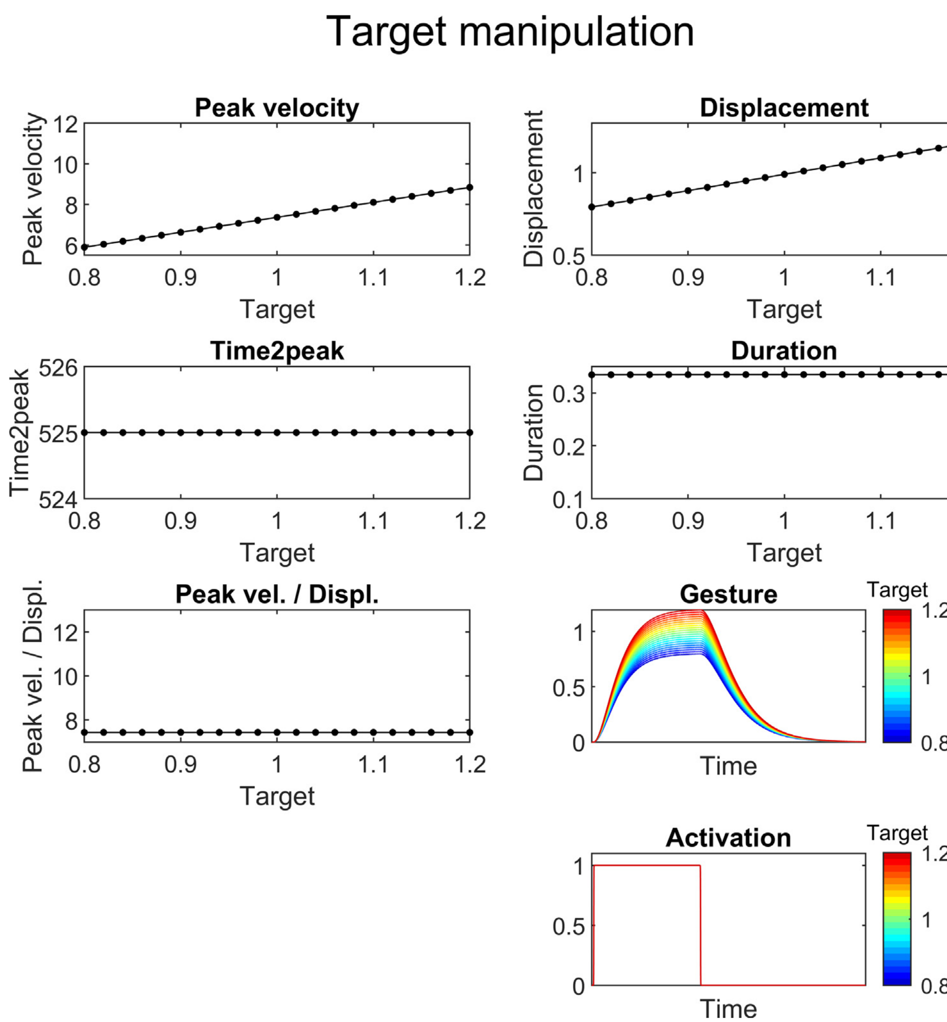

Fig. 5. Simulation results for target manipulation. Larger target values lead to higher peak velocities and larger displacements. The “gesture” panel gives the actual position of the gesture for the parameter range in the simulation. The “activation” panel shows the activation interval of the gesture over time.

This short demonstration of the correlation of model parameters and measurable effects in kinematics illustrates a major strength of employing a modeling approach in articulation research. A downside to this approach is, however, that we consider all parameters separately. Yet, they likely occur simultaneously in nature, which challenges the interpretation of empirical findings. Nevertheless, the ability to manipulate parameters in isolation and observe their impact on articulatory behaviors provides us with a starting point of a conceptual framework for the interpretation of the highly complex empirical findings that we observe in empirical data. The evaluation becomes more challenging when we consider the possibility that multiple parameters change at the same time. The main point, nevertheless, remains the same, it may even be strengthened in this scenario: The model puts us in a situation where we can make predictions about the data that we can use to better describe and classify our empirical findings with respect to neurotypical and impaired speech.

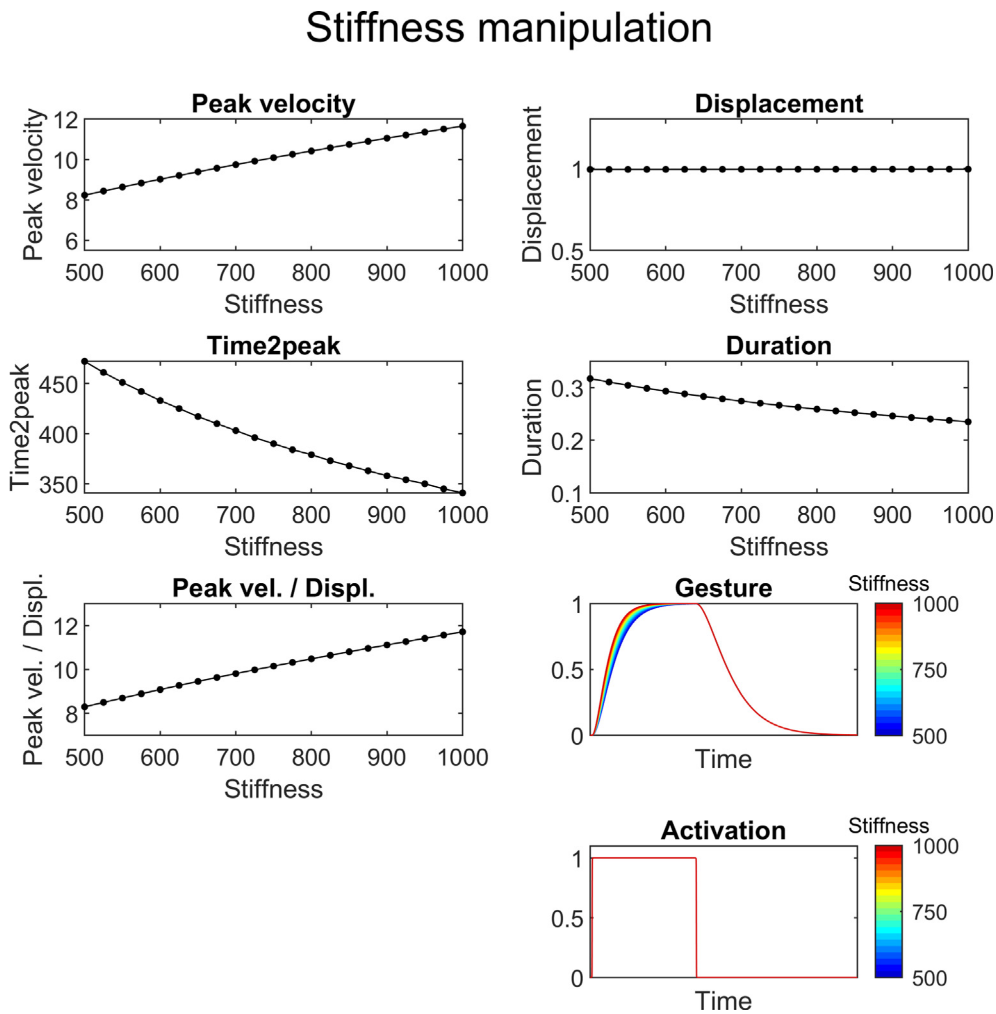

Fig. 6. Simulation results for stiffness manipulation. Higher stiffness values lead to higher peak velocities, shorter time-to-peak velocity lags (time2peak), higher ratios of peak velocity to displacement, as well as shorter durations. The "gesture" panel gives the actual position of the gesture for the parameter range in the simulation. The "activation" panel shows the activation interval of the gesture over time.

3. Stochasticity in the model

3.1. Using the gestural model with added noise

So far, we have treated the parameters of the model as scalar values that may be prone to systematic variation, for example, due to prosodic or other speech demands. However, it is also plausible that for a certain pattern of impairment or prosodic location, parameter values are fixed. In theory, this would result in the same movement trajectory for each realization. Of course, this is not what we observe empirically. Indeed, the concept of trial-to-trial variability is well-established in the literature. Yet, the underlying sources of this variability is still not fully understood. The notion that trial-to-trial variability can be attributed to noise is not new but has been established more than 60 years ago. That is, Fitts (1954) claimed that the motor behavior always encodes "the entire receptor-neural-effector system". In the speech motor control literature, however, the size and shape of distributions of a specific gestural control parameter (e.g., target) have rarely been studied (but see Mefferd, 2019; Perkell & Cohen, 1989; Whalen & Chen, 2019). As a first step, we can account for variability in our model by making the gestural parameters "noisy" in the simulations, by randomly drawing the parameter value from a Gaussian distribution.

The parameters (target, stiffness, activation duration, activation strength) were drawn from the Gaussian distributions for each pair of opening-closing gestures. The standard deviation of the distributions governs how much stochasticity is introduced. An increase in the standard deviation would lead to an increase in the deviation from the unperturbed contour (cf. Appendix 1). In an additional simulation, noise was added to the differential equation describing the acceleration of the gesture over time (cf. Appendix 1). Note that we use Gaussian distributions here as a first step; however, we acknowledge that this approach will require further refinements in the future as noise is likely not random but distributed in systematic ways (see Mefferd, 2019).

In this section, we use the basic unit of the AP/TD framework, namely the *gesture*, modeled as a critically damped harmonic oscillator to demonstrate some possible outcomes of

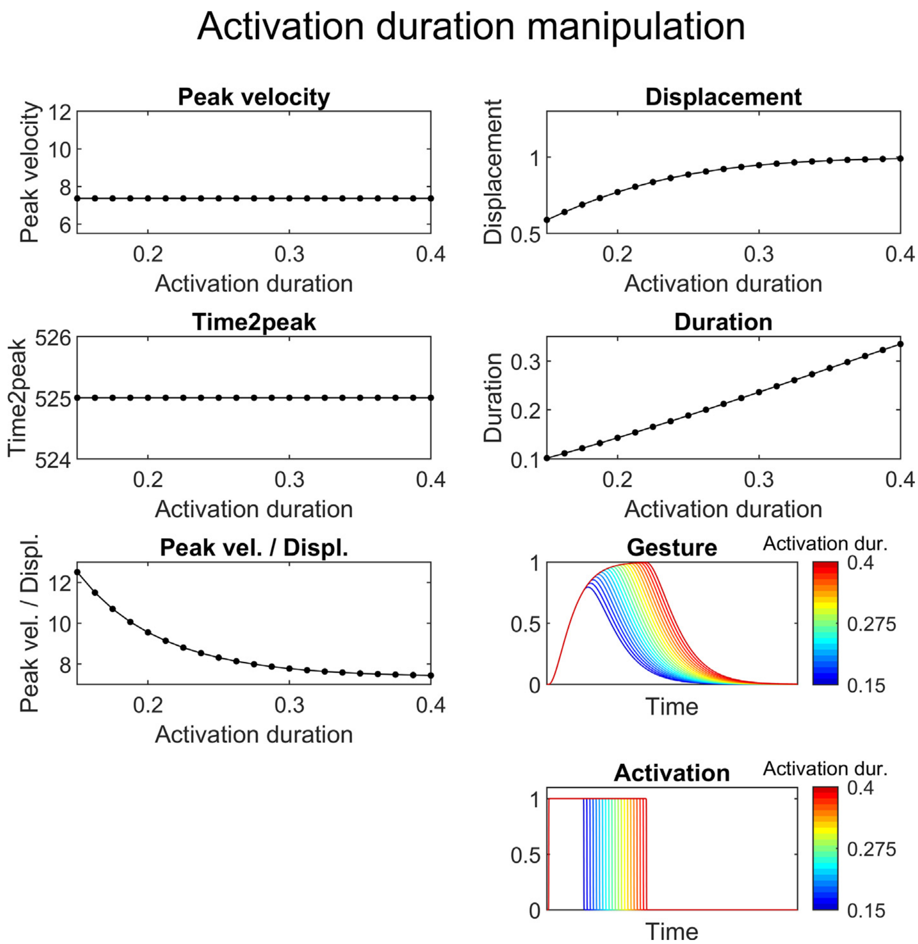

Fig. 7. Simulation results for activation duration manipulation. Longer activation intervals lead to larger displacements, longer durations, and lower ratios of peak velocity to displacement. The "gesture" panel gives the actual position of the gesture for the parameter range in the simulation. The "activation" panel shows the activation interval of the gesture over time.

the inclusion of stochasticity in this model. Fig. 8 shows simulations of a sequence of opening and closing gestures of the lips as they can be found in [papapa…]. The first four rows of the figure show simulations with noise on the parameters *target*, *stiffness*, *activation duration* and *activation strength.* In these simulations, we add noise to the parameters of each gesture *before* the simulation of that gesture starts. Once the gesture evolves, these parameters are not altered in the simulation anymore.

The fifth row from the top shows what happens if we add a noise in every step of the simulation of each gesture (which essentially renders the gestural system a non-deterministic system, see Appendix 1). Here, the parameters are not noisy. The sixth row shows the simulation with noise on all parts of the model considered here: noise on the parameters target, stiffness, activation duration, activation strength, and noise over time to the gestural equation. The Gaussian noise distributions used in the simulations are shown next to the simulation panels in the inner columns (center of the figure). Note that distributions for higher noise levels are flatter than those with moderate noise levels (i.e., more spread in the distributions). This means that the variance of the parameter will be greater when the noise level is higher. More details about the simulation can be found in Appendix 1. The last row, labelled "No noise", shows a simulation without any noise for comparison.

3.1.1. Kinematic contours: Simulations and speakers' behavior

In the application of dynamical systems to speech kinematic analysis, various challenges arise. This can be illustrated in fast syllable repetition tasks, also known as oral diadochokinetic tasks (DDKs), such as /papapa/ or /tatata/. These tasks elicit easily identifiable movement cycles, at least when produced by neurotypical speakers. DDKs have frequently been used in early speech kinematic research to investigate typical and impaired articulatory behaviors (Flanagan & Dembowski, 2002; Hertrich & Ackermann, 1997; Hirose et al., 2009; Hixon & Hardy, 1964; Jaeger et al., 2000; Nelson et al., 1984).

Presumably, articulatory behaviors elicited during DDK tasks should be best aligned with the kinematic description of speech movements because they inherit well-defined boundaries of a basic unit (e.g., movement cycle). However, comparisons of recorded DDK data from neurotypical speakers and speakers with motor speech impairment (Fig. 9) suggest that the contours generated by the model incorporating noise (Fig. 8) resemble more closely those of the patients for mild (Fig. 9b) and moderate (Fig. 9c) dysarthria severity levels[6]

[6] Note that these are just examples, for a more elaborate analysis see Mücke et al. (2018). The examples are taken from articulographic data, recorded at 1250 Hz, down sampled to 250 Hz, and smoothed with a 3-point moving average. Velocity is calculated after the smoothing process of the positional curves as first derivative of the respective x or y position of the smoothed contours.

Table 1
Summary of simulation results for parameter manipulations.

Measurement	Parameter		
	Stiffness	Target	Activation duration
Peak velocity	✓ less stiff ⟶ lower peak velocity stiffer ⟶ higher peak velocity	✓ lower target ⟶ lower peak velocity higher target ⟶ higher peak velocity	—
Displacement	—	✓ lower target ⟶ smaller displacement higher target ⟶ higher displacement	✓ shorter activation ⟶ smaller displacement longer activation ⟶ higher displacement
Time-to-peak velocity	✓ less stiff ⟶ longer time-to-peak vel. stiffer ⟶ shorter time-to-peak vel.	—	—
Duration	✓ less stiff ⟶ longer duration stiffer ⟶ shorter duration	—	✓ shorter activation ⟶ shorter duration longer activation ⟶ longer duration
Ratio peak velocity/displacement	✓ less stiff ⟶ larger ratio stiffer ⟶ smaller ratio	—	✓ shorter activation ⟶ larger ratio longer activation ⟶ smaller ratio

than the contours of the model without noise. For example, the scenario with high noise on all parameters depicts a pattern that is visible in Fig. 9c. Even the contours of the neurotypical control speakers show some irregularities (Fig. 9a). This is to be expected considering that noise is also present in neurotypical speakers. Smaller amounts of noise in general, or the application of noise only at the level of the parameters target, stiffness, and activation strength may be suited to model the production of these speakers. The simulations in Fig. 8 serve as a proof of concept. It shows that stochastic noise should be included in an extended version of the gestural model, which offers opportunities to simulate articulatory behavior more accurately and across a wide range including articulatory behaviors of speakers with speech impairment. While our focus is very often on the mean value, the variance of the distributions we find may give us an important characterization of the patterns we find. Adding Gaussian noise adds the amount of variance as a second parameter of interest alongside the mean.

3.1.2. Linking model output and empirical data

One key difference between the previously modeled articulatory behavior based on AP/TD and our proposed model output with noise is that the added noise can mimic well-known gesture-to-gesture variability in articulatory behavior. Thus, to determine the extent to which our proposed model with noise resembles articulatory movements produced by real speakers, we quantified and compared the cycle-to-cycle variability of the DDK movements. We predict that we will find a higher degree of variability in impaired speech compared to the speech of the control speakers. Likewise, we expect that the same is true for simulations with a high degree of noise vs a lower degree of noise.

Specifically, we selected ten DDK movement cycles of the lower lip (/papapa…./) produced by three essential tremor speakers with activated deep brain stimulation (mean age = 50 years, SD = 21), and three age-matched control speakers (mean age = 47 years, SD = 19), (cf. section 3.1.1). Every DDK cycle contained ten repetitions of the syllable /pa/.

We ran the simulations (as shown in Fig. 8) with fifteen iterations of six opening and closing gestures. For simulated data, we calculated cycle-to-cycle variability for two noise levels (medium, high) and six different noise conditions (see Figs. 5–8). As a first analysis step to compare simulated with empirical data, we focused exclusively on the closing movements of the vertical lower lip patterns.

For data extraction, we partially replicated a regular data processing pipeline as used in articulation research. That is, we automatically detected onsets and maximum targets of the closure gesture by using the zero crossings of the smoothed peak velocity trajectory (rolling mean). The consonantal closure duration for /p/ is the time from the onset of the closure gesture to the target of the gesture, and the consonantal closure displacement for /p/ is the difference between the position at the target and the position at the onset (cf. Fig. 4).

In a next step, we calculated the coefficient of variation (in percent) for the duration and the displacement for each simulation run, and averaged the coefficients for each condition. The coefficient of variation was calculated by dividing the standard deviation by the mean; this ratio was then multiplied by 100 to express it in percent: SD / Mean x 100. The results are presented in Table 2 for the empirical data, and in Table 3 for the simulated data. Note that we excluded the “no noise” scenario in Table 3 because it always yields a coefficient of 0%, i.e., all gestures are identical.

Based on visual inspection of our simulated and empirical data shown in section 3.1.1, we expected greater cycle-to-cycle variability for closing duration and closing displacement in our speakers with essential tremor than our control speakers (cf. Fig. 9). This observation is confirmed by the coefficient of variation in Table 2: While the control group shows an average coefficient of variation of 7.59% in duration and 15.73% in displacements for the closing movement, the essential tremor speakers produced 14.12% variability for duration and 23.43% for displacements. In addition, as expected, increasing the noise level from medium to high in the simulated data (cf. Table 3) led to greater variability in our simulated closure patterns for both, duration and displacement of the closing gesture. This is similar to the increased variability in speakers with speech impairments compared to the control speakers.

Examining the presence of noise across various parameters enables us to enhance our ability to predict the characteristics of control deficits associated with a particular speech motor disorder. Further, the modification of parameters regulat-

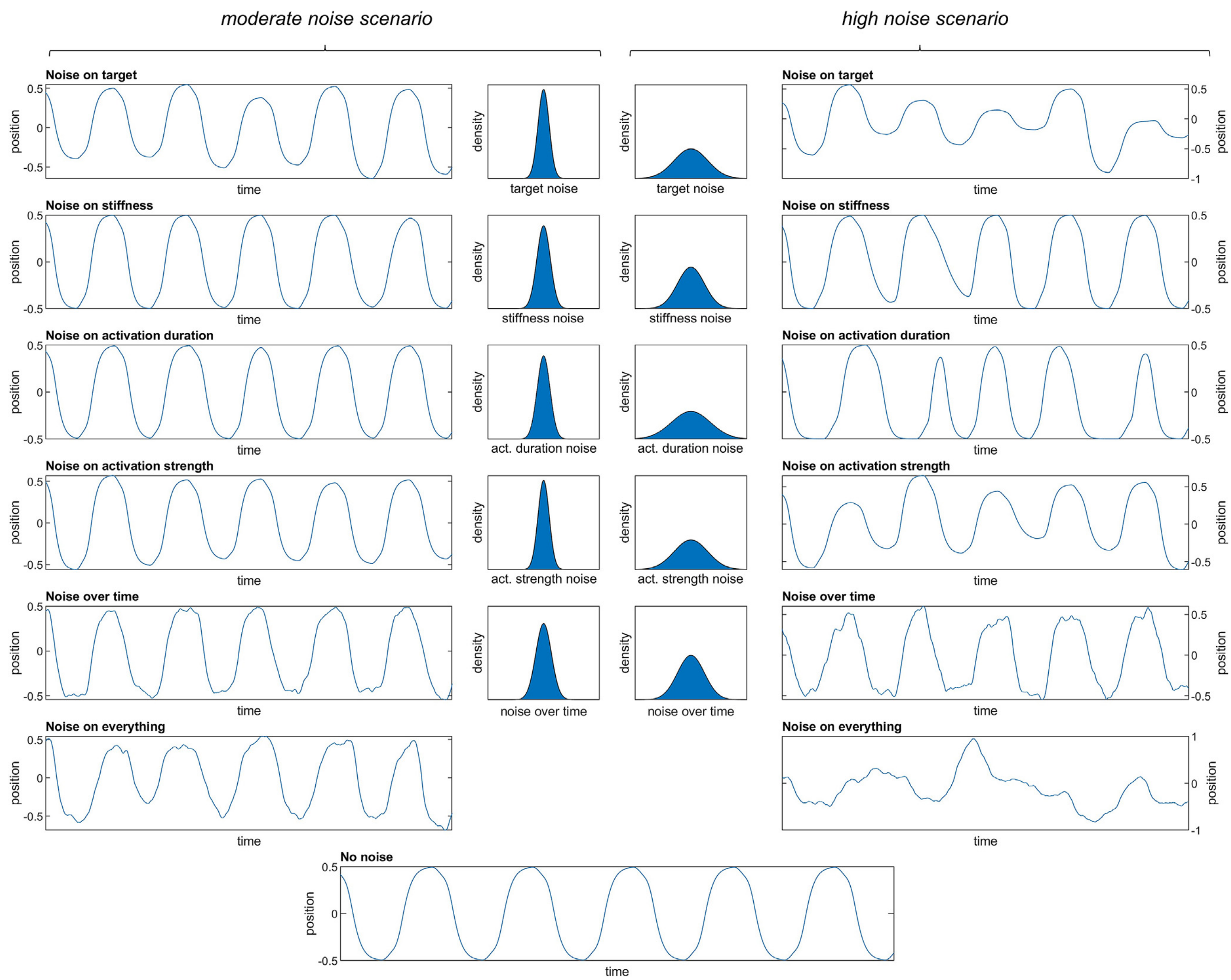

Fig. 8. Noise parameters on the AP/TD model. Left-most column: moderate amount of noise, right-most column: high amount of noise. Mid-columns: Gaussian distributions corresponding to the parameters in the simulations left and right.

ing speech motor control by introducing stochastic noise to the system can help us to get deeper insights into the nature of impaired speech patterns. In the example of Parrell et al. (2023), motor noise was added to the mobility-space acceleration of the TD simulations to model deviant speech patterns resembling changes in motor noise in speakers with amyotrophic lateral sclerosis (ALS). In our example of speakers with essential tremor with activated deep brain stimulation, we can think of durational and spatial variability resembled by, e.g., changes of activation duration and activation strength of the closing gesture as plausible parameters. However, more research on large-scale datasets is warranted to better understand the origin of the noise and, hence, the pathomechanisms of impaired articulatory behaviors.

3.2. Challenges with the kinematic analysis: Models and measures

3.2.1. Positional plateaus

In our simulation results (see Fig. 8 in section 3.1), positional plateaus occur when introducing noise on activation strength and noise over time in the “high noise scenario”. Such positional plateaus can be challenging for kinematic analyses. A movement plateau occurs due to minimal movement of the articulator (e.g., the speaker is holding the articulator steady for a certain period of time). Such an articulatory behavior is frequently observable when neurotypical speakers slow down their articulatory rate or speak as clearly as possible (Adams et al., 1993). It can also occur during prosodic modulations, such as prominence marking, to highlight an important word in an utterance or at prosodic edges, e.g., lengthening a final word (Beckman et al., 1992; Cho, 2006). Positional plateaus are also common in speakers with motor speech impairments, particularly those who exaggerate their movements or speak with an abnormally slow articulatory rate. When the analysis requires a clear velocity zero-crossing to index the time point of the articulatory target (e.g., when the articulator reaches its highest point) positional plateaus can be challenging. Although the target can be determined in the spatial domain, it is rather difficult to determine it in the temporal domain. The transition period (i.e., the time from movement onset to target) may be easily discernable, but the resulting segment duration does not consider the static portion of the articulatory gesture. Fig. 10 shows the irregular lower lip movements with the corresponding velocity signal during a fast syllable repeti-

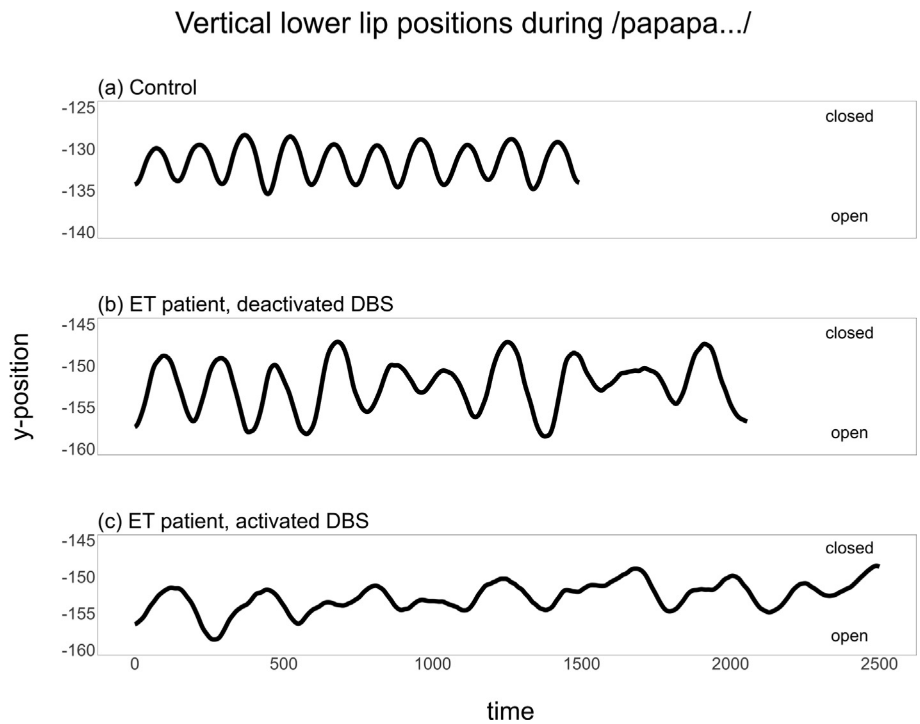

Fig. 9. Vertical lower lip movements elicited during the DDK task of /papapa. . ./, produced by a healthy speaker (a) and a speaker with essential tremor (ET) with deactivated (b) and activated (c) thalamic deep brain stimulation (DBS). Values on the y-axis are biteplane rotated data.

Table 2
Coefficient of variation of closure duration and displacement for empirical data: 3 healthy controls and 3 essential tremor speakers with activated deep brain stimulation.

Speaker group	Average coefficient of variation of closure duration (%)	Average coefficient of variation of closure displacement (%)
Control	7.59	15.73
Essential tremor	14.12	23.43

tion DDK task of /papapa/ related to impaired speech. The grey area displays an example of a positional plateau.

One solution is to use a threshold of local velocities to determine turning points of the movement. In many studies, the threshold is set to 20% of the peak velocity (Kroos et al., 1997; Tiede, 2005; Shaw & Chen, 2019), but different choices e.g., higher or lower thresholds are also possible. By using thresholds of velocity peaks, the positional plateau can be further divided into target achievement, maximum target (often midpoint of the plateau), and target release. Indeed, the threshold method improves robustness in the face of noise if the position maxima and minima are not well defined. There are two options to resolve this issue: Either the midpoint between the target achievement and release can operationally be defined as the time of target achievement to function as a reference point for the maximum target. Or the time from target achievement to target release should be investigated as an additional potentially interesting characterization of the movement. Indeed, some researchers use the term “formation duration” or similar expressions to encompass both the movement duration and the hold (plateau) duration when referring to this latter case (e.g., Katsika, 2016; Katsika & Tsai, 2021; Kim et al., in press). However, it might also be the case that thresholds are too unstable in highly irregular contours exhibiting

Table 3
Coefficient of variation of closure duration and displacement for simulated data by adding noise of different strength to the single parameters (target, stiffness, activation duration, activation strength, noise over time), and noise on everything.

Average coefficient of variation of closure duration (%)						
Noise strength	Noise on target	Noise on stiffness	Noise on activation duration	Noise on activation strength	Noise over time	Noise on everything
medium	0.41	4.14	19.54	1.03	84.83	86.55
high	16.75	9.37	38.69	4.79	90.69	103.34
Average coefficient of variation of closure displacement (%)						
Noise strength	**Noise on target**	**Noise on stiffness**	**Noise on activation duration**	**Noise on activation strength**	**Noise over time**	**Noise on everything**
medium	13.53	2.50	19.41	7.11	106.61	110.77
high	46.83	10.38	20.17	22.66	122.67	161.38

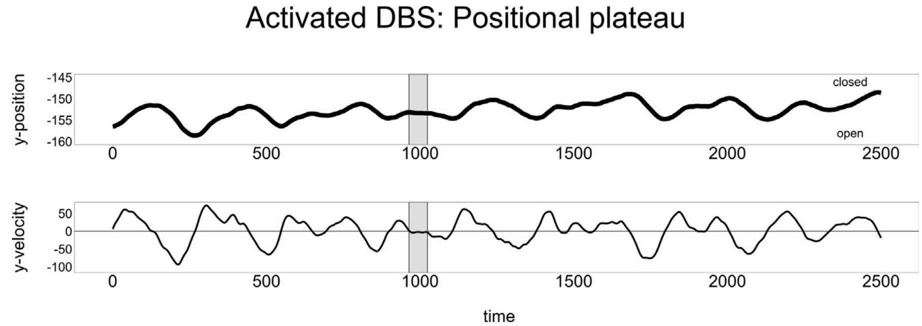

Fig. 10. Irregular lower lip movements (upper panel: vertical position; lower panel: respective velocity) during a fast syllable repetition task of /papapa. . ./, produced by a speaker with Essential Tremor with activated thalamic deep brain stimulation (DBS). The grey area indicates a positional plateau during a syllable cycle.

multiple peaks in the velocity contour; in these cases, defining local minima and maxima to identify onsets and targets of a movement can be the preferable option.

3.2.2. Multiple velocity peaks

Although the velocity signal of an articulatory gesture is typically characterized by a bell-shaped with a single peak (see Fig. 3 and Fig. 4 in section 2.2), multiple velocity peaks can occur, which also pose a challenge for the kinematic analysis. In the simulation results (section 3.1, Fig. 8), the multiple peak problem occurs when introducing noise to a combination of parameters, but not on a single noise parameter. Velocity-based landmarks are frequently used to identify positional targets. That is, local positional minima and maxima can be identified by zero-crossings in the velocity trace or by using thresholds of velocity peaks. However, irregularities in the velocity contours are quite frequent in impaired speech. A common scenario of multiple peaks in the velocity signals is shown in Fig. 11 during a DDK task of /papapa/ produced by a speaker with speech motor impairment. The grey areas display some examples of multiple velocity peaks in syllable cycles that can be identified in a single DDK task. The raising of the lower lip during one syllable cycle goes along with several velocity peaks. Thus, one has to decide which peak to select for the analysis, which subsequently impacts the durations of the acceleration and deceleration phases. If the first peak was taken as the relevant landmark, the time-to-peak velocity (acceleration phase) would be much shorter than if the second peak was taken. The deceleration phase would be either longer or shorter depending on which peak velocity was selected.

It has been reported in the literature for movements of the tongue, jaw and lips that the number of velocity peaks increases with a decrease in speaking rate resulting in more asymmetrical velocity profiles (Adams et al., 1993; Wieneke et al., 1987) 1993). These results are interpreted by Adams et al. (1993) as a change in motor control strategies due to speaking rate. While fast speaking rates seem to trigger pre-programmed unitary movements, slow speech might consist of multiple submovements that are shaped by several feedback mechanisms (Adams et al., 1993). They conclude that the occurrence of multiple velocity peaks in slow speech can be interpreted as a universal mechanism of rate control that can also be found in gross motor control.

The number of velocity peaks is an important aspect of articulatory behavior. Particularly the occurrence of multiple velocity peaks can be indicative of an impaired motor speech system. For example, differences in the smoothness of articulatory movements, also known as jerk, which is ultimately driven by multiple velocity peaks, can distinguish neurotypical speakers and speakers with progressive neurological conditions such as ALS (Bandini et al., 2018). Furthermore, a greater number of velocity peaks has been found in the articulatory movements of speakers who stutter when compared to those of speakers who do not stutter (Zmarich & Caldognetto, 1997).

Alternatively, occurrences of multiple velocity peaks may also be handled by hand-selecting only the highest of all velocity peaks within the gesture and ignoring all others, as suggested by a reviewer. Last but not least, one could determine a velocity threshold based on the observed maximum velocity and determine the time from exceeding 80% of peak velocity to time of falling below 80%. The resulting interval duration (or its

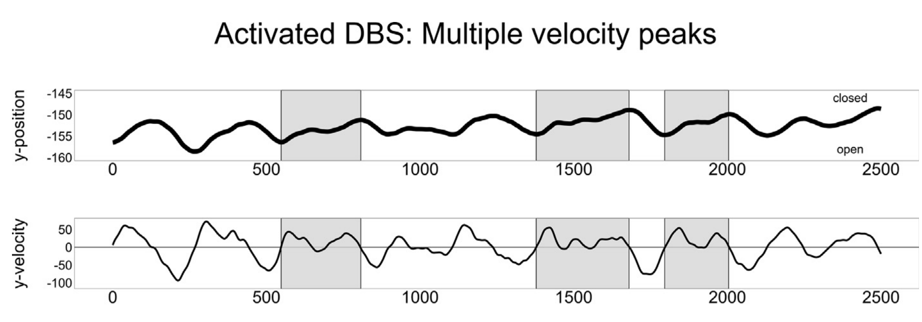

Fig. 11. Lower lip movements (upper panel: vertical position; lower panel: respective velocity) during a fast syllable repetition DDK task of /papapa. . ./, produced by a speaker with essential tremor with activated thalamic deep brain stimulation (DBS). The grey areas highlight some intervals of multiple velocity peaks during a syllable cycle.

variability) may be useful to characterize the transition period from acceleration to deceleration.

4. Discussion and conclusion

In this paper, we argued that the AP/TD approach presents promising opportunities to model and understand articulatory behavior in speakers with motor speech disorders as well as articulatory behavior of neurotypical speakers under various speech conditions. In AP/TD, speech can be decomposed into articulatory gestures that fulfil linguistic tasks. The speech gestures are governed by point-attractor dynamics allowing the description of the continuity and discreteness of speech in a single model. Especially for impaired speech, modulations of stiffness, target, and activation are important to model all degrees of temporal and spatial modulations. However, when analyzing kinematics in speakers with speech motor impairment, contours are often highly irregular. This poses challenges during the analysis, even in the steps of landmark annotations related to measures such as onset, peak velocity, and target of a movement. Given a phonological representation, it is currently not possible to derive a surface representation of disordered speech with available models. Thus, extensions of the model may be needed to better accommodate more deviant articulatory behaviors.

Introducing stochastic noise to various levels of simulations such as the level of parametrization (target, stiffness, activation duration, and activation strength) and over time to the gestural equation may be a meaningful extension to the AP/TD model. The use of noise as outlined in this paper presents a first proof of concept. Adding Gaussian noise essentially simulates the fact that we always deal with distributions of values in speech. In such a simulation, the variance of the Gaussian parameter distribution becomes a crucial parameter and it may give us an important characterization of the patterns we find, as these simulations show.

However, many questions arise from this extension to the gestural model. Our proposed approach does not address the origin of the noise. For example, we have only shown that the noisy modeled contours resemble those of speakers with speech motor impairment. Future studies are warranted to determine which levels of noise (parameters, over time to the differential equation, etc.) are most plausible and whether different types of impairments may be characterized by different degrees of noise on different levels (or even parameters) of the model. That is, which (neural) subsystems contribute which kind of noise? Furthermore, it remains unclear whether different types of pathological patterns are characterized by noise on different levels or parameters.

A thorough quantitative analysis of the kinematic data of speakers with speech impairment plus model simulations are needed to shed light on these questions. It should also be mentioned here that the AP/TD model is (explicitly) not a (detailed) biomechanical model. Therefore, it may not be possible to model all perturbations using noisy components. It may also be interesting to think about pathological patterns arising from noise during gestural selection (Tilsen, 2018). The selection process adds yet another potential level. Nevertheless, we think that the proposed extension of the gestural model is a starting point; therefore, its potential should be further evaluated. An important step in this direction is the recent work of Parrell et al. (2023). In a proof of concept, they used stochastic noise to model speech patterns in the AP/TD application resembling changes in motor noise in speakers with ALS. More specifically, they added noise to the mobility-space acceleration of the TD simulations. This is a level for the output of the control algorithm that is the final motor command to the plant and can be associated with motor noise in ALS speakers. Indeed, the simulated speech patterns may account for altered patterns of spatial variability in this population. This is an important step towards future directions for opening the AP/TD approach to give new insight into the analysis of impaired speech. Future work will consider all possible levels of the AP/TD framework on which stochasticity may play a role and evaluate their relative importance as well as their interaction with one another.

We would like to mention here that the provided discussion of the quantitative evaluation of kinematic analysis procedures in section 3.2 is limited to the acquisition of point parameterizations such as in electromagnetic articulography. There are other statistical approaches available such as functional data analysis and generalized additive models that can capture the contour more holistically without the pre-segmentation of the relevant movement unit. And, last but not least, other recording opportunities such as ultrasound are also available to analyze a wide range of speech kinematics, but they involve other challenges not discussed here.

We conclude that integrating the field of motor speech disorders and dynamical systems has the potential not only to improve our understanding of these disorders but also our understanding of speech production more broadly. It can help to start making connections between more abstract models and the neural basis of speech production. Such cross-talk between disciplines will be instrumental to refining and broadening current theories of speech motor control to account for a wide range of articulatory behavior – bridging clinical and linguistic approaches to speech.

CRediT authorship contribution statement

Doris Mücke: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing, Resources. **Simon Roessig:** Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition, Supervision, Validation. **Tabea Thies:** Conceptualization, Data curation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Anne Hermes:** Conceptualization, Funding acquisition, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Antje Mefferd:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table A1
Means and standard deviations of noisy parameters.

	Mean	Standard deviation moderate noise	Standard deviation high noise scenario
Target	−0.5 or 0.5 (depending on whether closing or opening gesture)	0.1	0.3
Stiffness	1	0.3	0.6
Activation duration	1000 time steps of the simulation	150	450
Activation strength	1	0.1	0.3

Acknowledgements

This work was supported by the German Research Foundation (DFG) as part of the SFB1252 “Prominence in Language” (Project-ID 281511265), project A04 “Dynamic modelling of prosodic prominence” at the University of Cologne, and project RO 6767/1-1. It was also supported by NIH NIDCD grant R01DC019648, by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex Empirical Foundations of Linguistics – EFL.

Appendix 1. Simulation of gestural noise

Lip aperture trajectories for a [papapapa…] sequence were simulated using the Euler-Maruyama method. The gestural target was alternating between a high (mean +0.5) and a low target (mean −0.5), simulating closing and opening gestures.

The parameters (target, stiffness, activation duration, activation strength) were drawn from the Gaussian distributions for each pair of opening-closing gestures. Table A.1 gives a list of their means and standard deviations.

In an additional simulation, noise was added to the differential equation describing the acceleration of the gesture over time:

$$\ddot{x} = -B(t)\dot{x} - K(t)(x - T(t)) + q\xi(t)$$

$K(t)$ and $T(t)$ are the stiffness and the target of the gesture. They are functions of time t, since they depend on the activation $A(t)$ of the gesture that changes over time: $K(t) = K \cdot A(t), T(t) = T \cdot A(t)$, where K and T are the values drawn from the Gaussian distributions described above. Where B(t) is the damping of the gesture, defined as $B(t) = 2\sqrt{K(t)}$. The mass of the gesture is fixed at 1. The noise is introduced by adding $q\xi(t)$, where $\xi(t)$ is a Wiener process and q is the noise strength (Schöner & Spencer, 2015). The noise strength parameter corresponds to the breadth of the Gaussian noise distribution. In the “moderate noise” scenario, this parameter was set to 7, while it was set to 12 in the “high noise” scenario. Note that in the simulation where noise was on the parameters only, this parameter was set to zero.

References

Ackermann, H., Hertrich, I., & Scharf, G. (1995). Kinematic analysis of lower lip movements in ataxic dysarthria. *Journal of Speech, Language, and Hearing Research, 38*(6), 1252–1259. https://doi.org/10.1044/jshr.3806.1252.

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech, Language, and Hearing Research, 36*(1), 1. https://doi.org/10.1044/jshr.3601.41.

Bandini, A., Green, J. R., Wang, J., Campbell, T. F., Zinman, L., & Yunusova, Y. (2018). Kinematic features of jaw and lips distinguish symptomatic from presymptomatic stages of bulbar decline in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 61*(5), 1118–1129. https://doi.org/10.1044/2018_JSLHR-S-17-0262.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Segment, gesture, prosody* (pp. 68–86). Cambridge University Press. https://doi.org/10.1017/CBO9780511519918.004.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units*. *Phonology, 6*(2), 201–251. https://doi.org/10.1017/S0952675700001019.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3–4), 155–180. https://doi.org/10.1159/000261913.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics, 26*(2), 173–199. https://doi.org/10.1006/jpho.1998.0071.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*, 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2.

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English. *Laboratory Phonology, 8*, 519–548. https://doi.org/10.1515/9783110197211.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America, 97*, 491. https://doi.org/10.1121/1.412275.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*(6), 381–391. https://doi.org/10.1037/h0055392.

Flanagan, K. P., & Dembowski, J. S. (2002). Kinematics of normal lingual diadokokinesis. *The Journal of the Acoustical Society of America, 111* (5_Supplement), 2476. https://doi.org/10.1121/1.1486357.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics, 29*(2), 109–135. https://doi.org/10.1006/jpho.2000.0114.

Fuchs, A. (2013). Nonlinear dynamics in complex systems. *Springer*. https://doi.org/10.1007/978-3-642-33552-5.

Gafos, A. I., & Benus, S. (2006). Dynamics of phonological cognition. *Cognitive Science, 30*(5), 905–943. https://doi.org/10.1207/s15516709cog0000_80.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge University Press. https://doi.org/10.1017/CBO9780511541599.008.

Haken, H. (1977). Synergetics. *Springer*. https://doi.org/10.1007/978-3-642-96363-6.

Hawkins, S. (1992). An introduction to task dynamics. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Segment, gesture, prosody* (pp. 9–25). Cambridge University Press. https://doi.org/10.1017/CBO9780511519918.002.

Hertrich, I., & Ackermann, H. (1997). Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *The Journal of the Acoustical Society of America, 102*(1), 523–536. https://doi.org/10.1121/1.419725.

Hirose, H., Kiritani, S., & Sawashima, M. (2009). Patterns of dysarthric movement in patients with amyotrophic lateral sclerosis and pseudobulbar palsy. *Folia Phoniatrica et Logopaedica, 34*(2), 106–112. https://doi.org/10.1159/000265636.

Hixon, T. J., & Hardy, J. C. (1964). Restricted motility of the speech articulators in cerebral palsy. *Journal of Speech and Hearing Disorders, 29*(3), 293–306. https://doi.org/10.1044/jshd.2903.293.

Iskarous, K. (2017). The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics, 64*, 8–20.

Iskarous, K., & Pouplier, M. (2022). Advancements of phonetics in the 21st century: A critical appraisal of time and space in Articulatory Phonology. *Journal of Phonetics, 95*. https://doi.org/10.1016/j.wocn.2022.101195 101195.

Jaeger, M., Hertrich, I., Stattrop, U., Schönle, P.-W., & Ackermann, H. (2000). Speech disorders following severe traumatic brain injury: Kinematic analysis of syllable repetitions using electromagnetic articulography. *Folia Phoniatrica et Logopaedica, 52*(4), 187–196. https://doi.org/10.1159/000021533.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics, 55*, 149–181. https://doi.org/10.1016/j.wocn.2015.12.003.

Katsika, A., & Tsai, K. (2021). The supralaryngeal articulation of stress and accent in Greek. *Journal of Phonetics, 88*. https://doi.org/10.1016/j.wocn.2021.101085 101085.

Kim, J. J., Kim, S., & Cho, T. (in press). Preboundary lengthening and articulatory strengthening in Korean as an edge-prominence language. *Laboratory Phonology*.

Krivokapić, J., Tiede, M. K., & Tyrone, M. E. (2017). A Kinematic study of prosodic structure in articulatory and manual gestures: Results from a novel method of data collection. *Laboratory Phonology, 8*(1), 3. https://doi.org/10.5334/labphon.75.

Kroos, C., Hoole, P., Kühnert, B., & Tillmann, H. G. (1997). *Phonetic evidence for the phonological status of the tense-lax distinction in German*:35 (pp.17–25). Forschungsberichte Institut für Phonetik und sprachliche Kommunikation der Universität München.

Mefferd, A. S. (2019). Effects of speaking rate, loudness, and clarity modifications on kinematic endpoint variability. *Clinical Linguistics & Phonetics, 33*(6), 570–585. https://doi.org/10.1080/02699206.2019.1566401.

Miller, H. E., & Guenther, F. H. (2021). Modelling speech motor programming and apraxia of speech in the DIVA/GODIVA neurocomputational framework. *Aphasiology, 35*(4), 424–441. https://doi.org/10.1080/02687038.2020.1765307.

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics, 44*, 47–61. https://doi.org/10.1016/j.wocn.2014.02.003.

Mücke, D., Hermes, A., & Cho, T. (2017). Mechanisms of regulation in speech: Linguistic structure and physical control system. *Journal of Phonetics, 64*, 1–7. https://doi.org/10.1016/j.wocn.2017.05.005.

Mücke, D., Hermes, A., Roettger, T. B., Becker, J., Niemann, H., Dembek, T. A., Timmermann, L., Visser-Vandewalle, V., Fink, G. R., Grice, M., & Barbe, M. T. (2018). The effects of Thalamic Deep Brain Stimulation on speech dynamics in patients with Essential Tremor: An articulographic study. *PLOS ONE, 13*(1), Article 1. https://doi.org/10.1371/journal.pone.0191359.

Mücke, D., Hermes, A., & Tilsen, S. (2020). Incongruencies between phonological theory and phonetic measurement. *Phonology, 37*(1), Article 1. https://doi.org/10.1017/S0952675720000068.

Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 457–474. https://doi.org/10.1037/0096-1523.11.4.457.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to phonological complexity* (pp. 299–328). Walter de Gruyter. https://doi.org/10.1515/9783110223958.297.

Nelson, W. L., Perkell, J. S., & Westbury, J. R. (1984). Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations. *The Journal of the Acoustical Society of America, 75*(3), 945–951. https://doi.org/10.1121/1.390559.

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.

Parrell, B., Mefferd, A., Harper, S., Roessig, S., & Mücke, D. (2023). Using computational models to characterize the role of motor noise in speech: the case of amyotrophic lateral sclerosis. In *Proceedings of 20th International Congress of Phonetic Sciences*. International Congress of Phonetic Sciences (ICPhS), Prague.

Perkell, J. S., & Cohen, M. H. (1989). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/. *Journal of Phonetics, 17*(1), 123–133. https://doi.org/10.1016/S0095-4470(19)31517-7.

Roessig, S. (2021). *Categoriality and continuity in prosodic prominence: Bd. Studies in Laboratory Phonology* (S. 176). Language Science Press.

Roessig, S., & Mücke, D. (2019). Modeling dimensions of prosodic prominence. *Frontiers in Communication, 4*. https://doi.org/10.3389/fcomm.2019.00044.

Roon, K. D., Hoole, P., Zeroual, C., Du, S., & Gafos, A. I. (2021). Stiffness and articulatory overlap in Moroccan Arabic consonant clusters. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 12*(1), Article 1. https://doi.org/10.5334/labphon.272.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2.

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Speech Prosody, 10*.

Schöner, G., & Spencer, J. (2015). *Dynamic Thinking: A Primer on Dynamic Field Theory* (1st edition). Oxford University Press.

Shaw, J. A., & Chen, W.-r. (2019). Spatially Conditioned Speech Timing: Evidence and Implications. *Frontiers in Psychology, 10*, 2726. https://doi.org/10.3389/fpsyg.2019.02726.

Shaw, J. A., 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. https://doi.org/10.1017/S0952675711000224.

Sorensen, T., & Gafos, A. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology, 28*(4). https://doi.org/10.1080/10407413.2016.1230368 4.

Tiede, M. (2005). *MVIEW: software for visualization and analysis of concurrently recorded movement data*. New Haven, CT: Haskins Laboratories.

Tilsen, S. (2018). Three mechanisms for modeling articulation: Selection, coordination, and intention. *Cornell Working Papers in Phonetics and Phonology*.

Tilsen, S. (2022). An informal logic of feedback-based temporal control. *Frontiers in Human Neuroscience, 16*. https://doi.org/10.3389/fnhum.2022.851991.

Whalen, D. H., & Chen, W.-R. (2019). Variability and central tendencies in speech production. *Frontiers Communication, 4*. https://doi.org/10.3389/fcomm.2019.00049.

Wieneke, G., Janssen, P., & Belderbos, H. (1987). The influence of speaking rate on the duration of jaw movements. *Journal of Phonetics, 15*(2), 111–126. https://doi.org/10.1016/S0095-4470(19)30547-9.

Zmarich, C., & Caldognetto, E. M. (1997). Analysis of lips and jaw multi-peaked velocity curve profiles in fluent speech by stutterers and nonstutterers. *Speech Production: Motor Control, Brain Research and Fluency Disorders*, 177–182.