Modeling trajectories of human speech articulators using general Tau theory

Citation for published version:
Elie, B, Lee, DN & Turk, A 2023, 'Modeling trajectories of human speech articulators using general Tau theory', *Speech Communication*, vol. 151, pp. 24-38. https://doi.org/10.1016/j.specom.2023.04.004

Digital Object Identifier (DOI):
10.1016/j.specom.2023.04.004

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Speech Communication

Speech Communication 151 (2023) 24–38

Contents lists available at ScienceDirect

Speech Communication

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

Modeling trajectories of human speech articulators using general Tau theory

Benjamin Elie[a,*], David N. Lee[b], Alice Turk[a]

[a] Linguistics and English Language; School of Philosophy, Psychology and Language Sciences; the University of Edinburgh, Edinburgh, Scotland, United Kingdom

[b] Psychology; School of Philosophy, Psychology and Language Sciences; the University of Edinburgh, Edinburgh, Scotland, United Kingdom

ARTICLE INFO

Dataset link: https://git.ecdf.ed.ac.uk/belie/tauspeech, https://datashare.ed.ac.uk/handle/10283/4495

Keywords:
Articulatory trajectories
General Tau theory
Optimal control theory

ABSTRACT

This paper presents an application of general Tau theory to the modeling and analysis of articulatory trajectories in speech. We evaluated the model using electromagnetic articulometry data from 12 native speakers of English reading a common text, where trajectories of the following sensors were fitted: lower and upper lips, jaw, and three tongue sensors. Additionally, we analyzed trajectories of the lip aperture signal. Our experiments show that the general Tau theory model gives a better fit than existing (i) methods based on critically damped oscillators, and (ii) a method based on sequential target approximation. These findings support the hypothesis of Tau-guided movements of articulators during speech production. In the second part of the paper, our Tau theory analysis shows that articulatory movements follow similar velocity profile distributions across speakers. In particular, the value of the shape parameter κ of the Tau theory equation is identically distributed across speakers, following a unimodal distribution. The statistical mode of the distribution corresponds to the value of κ that generates a symmetric velocity profile. The analysis of the statistical distribution of κ values also reveals that its variance decreases when greater articulatory effort is required, such that produced articulatory effort remains close to that predicted by the theoretical minimal cost function based on forces acting on the moving articulator. This provides new evidence that articulatory effort is optimized during speech production.

1. Introduction

The production of speech involves movements of articulators used to shape the geometry of the vocal tract. The temporal evolution of this geometry allows context-appropriate acoustic features of speech to be produced in order to convey information. How these movements are planned and executed by the speaker is still a subject of debate.

Models of speech articulatory planning usually contain a dynamical component aiming at computing and/or predicting articulatory trajectories generated by the speaker using generative models. The interest of such models are threefold, as they can be used for (i) generating articulatory data for articulatory speech synthesizers, (ii) analyzing articulatory movements in real speech using a small number of parameters, and (iii) assessing the validity of existing theories of speech production.

Many attempts have been made in the past to model the articulatory trajectories of speech. Statistical models provide a robust and efficient way to predict articulatory movements (Ling et al., 2010; Ribeiro et al., 2022), as they are based on observation and statistical learning. They are very useful for generating trajectories for articulatory speech synthesizers, but cannot be used effectively to analyze observed trajectories using a few parameters. For that purpose, articulatory trajectories are often analyzed via analytic and parametric models. Existing models of these types use either interpolation functions (Henke, 1966; Keating, 1990; Blackburn and Young, 2000; Okadome and Honda, 2001) or asymptotic approximations (Saltzman and Munhall, 1989; Kröger et al., 1995; Xu, 2004; Šimko and Cummins, 2010; Birkholz et al., 2010; Sorensen and Gafos, 2016). Asymptotic approximations are dominant for modeling articulatory trajectories of speech using parametric models and are at the core of the most common articulatory models of speech production. For instance, the Task Dynamics model (hereafter TD) (Saltzman, 1986; Saltzman and Munhall, 1989; Šimko and Cummins, 2010; Sorensen and Gafos, 2016) considers articulators as critically damped oscillators moving towards an asymptotic target without overshoot. In TD, articulator movements are achieved by a second-order dynamical system. This model is used in combination with Articulatory Phonology (hereafter AP) (Browman and Goldstein, 1986) to form AP/TD (Browman and Goldstein, 1995), one dominant model of speech production. Another asymptotic approach has been proposed by Birkholz et al. (2010), in which articulatory commands are modeled as a cascade of N first-order linear systems, with $N > 2$. All of these asymptotic approximation methods have been successfully applied to explain articulatory patterns (as for AP/TD, for instance), to model

[*] Corresponding author.
E-mail address: benjamin.elie@ed.ac.uk (B. Elie).

https://doi.org/10.1016/j.specom.2023.04.004
Received 3 August 2022; Received in revised form 12 April 2023; Accepted 22 April 2023
Available online 15 May 2023

and fit observed articulatory trajectories (Kröger et al., 1995; Birkholz et al., 2010; Birkholz and Hoole, 2012), and to generate trajectories for articulatory synthesizers (Birkholz, 2007; Prom-on et al., 2013; Xu et al., 2019; Alexander et al., 2019).

A potential limitation of asymptotic models is the implied temporal coordination mechanism. Because these models are asymptotic, i.e. the target is never reached, they usually assume that the temporal coordination of speech articulators occurs at the onset of articulatory gestures, which is in line with AP-based models (Browman and Goldstein, 1986). However, as discussed in Turk and Shattuck-Hufnagel (2020a), there is evidence that bodily movements are often coordinated to reach a target at a specific time point, *i.e.* the temporal coordination often happens at the goal-related movement offset, resulting in lower timing variability of movement endpoints. This has been shown, for instance, in typewriting (Gentner et al., 1980) and for periodic tapping (Spencer and Zelaznik, 2003). Perkell and Matthies (1992) observed lower variability of maximum protrusion in spoken /iCu/ sequences, as compared to the timing of a point after the movement onset. These findings suggest that models of articulatory planning should include explicit representations of movement endpoints, so that they can be timed with precision.

Recently, Turk and Shattuck-Hufnagel (2020a,b) have proposed to adapt general Tau theory of movement (Lee, 1998) to speech to tackle this issue. Lee's general Tau theory has been developed from previous work by Gibson (1966) and Bernstein (1966), and has been supported by various experiments on different kinds of bodily movements (Lee and Reddish, 1981; Lee et al., 1983; Craig and Lee, 1999; Schögler et al., 2008; Rodger et al., 2013). The basic assumption of the theory is that purposeful movements aim at closing gaps, e.g. a distance gap or an angle gap. The gap-closure function is defined such that the target is reached at the right time, *i.e.* for a distance gap, the gap is closed at the movement endpoint. Another interest for trajectory modeling is that Tau-guided movements exhibit single-peaked velocity profiles whose symmetry can be changed by adjusting a unique parameter, the Tau-coupling parameter κ.

Considering these features, general Tau theory is thus a good candidate to explain the production of articulatory trajectories in speech. This paper aims at evaluating its relevance for speech, both for the generation and analysis of articulatory trajectories. In order to do so, the paper compares the fit of the Tau theory equation to real articulatory trajectories, extracted from electromagnetic articulometry (EMA) data, with the fit of other methods based on asymptotic target approximations, namely two types of Critically Damped Oscillator models (Kröger et al., 1995; Sorensen and Gafos, 2016), and the Sequential Target Approximation (Birkholz et al., 2010) model. The comparison is carried over the trajectories of all sensor signals from a dataset from the DoubleTalk corpus (Scobbie et al., 2013; Geng et al., 2013), corresponding to EMA recordings of 12 native speakers of English reading an English text (Scobbie et al., 2013; Geng et al., 2013). This extensive comparison will show the model which is able to reproduce articulatory movement most accurately across speakers, and across varied prosodic contexts of spoken English.

This paper also presents a statistical investigation of the Tau equation parameters that provide the best fit to observed movement trajectories in the reading task part of the DoubleTalk corpus. This investigation of the readings of the same text by 12 different speakers will allow us to see if any speaker differences exist. In addition, they will allow us to see if the most commonly-used values of the shape parameter are those which are implicated in Tau-guided movements that are least effortful out of the set of possible Tau-guided movements.

The organization of the paper is as follows. The existing models of articulatory trajectories with which ours is compared are detailed in Section 2. General Tau theory is introduced in Section 3. Section 3 also provides analytical developments of general Tau theory that relate the mathematical properties of Tau-guided movements to known characteristics of velocity profiles of speech movements. The subsequent sections report experiments which introduce the application of Tau theory to speech. The aim of Section 4 is to assess the fit of the general Tau theory equations to real speech articulatory data, and consequently, their relevance for articulatory analysis and modeling. Section 5 provides an example of articulatory analysis using Tau theory. The aim of the experiment presented in Section 5 is to show that general Tau theory can be used to analyze speech movements: our findings highlight general tendencies related to velocity profiles. Finally, Section 6 provides an attempt to explain our observations from the preliminary analysis in Section 5, using general Tau theory as a basis. It investigates the relationship between the characteristics of observed articulatory movements and articulatory effort.

2. Trajectory models of speech articulators

In order to assess the relevance of applying general Tau theory to speech, this paper compares it with existing models. We chose to compare general Tau theory to Critically Damped Oscillator models (hereafter CDO) (Saltzman and Munhall, 1989; Kröger et al., 1995; Sorensen and Gafos, 2016) and the Sequential Target Approximation Model (hereafter STAM) (Birkholz et al., 2010). All of these are target approximation models. The choice of CDO is motivated by its wide use in speech production models, *i.e.* models based on Task Dynamics (Saltzman, 1986; Saltzman and Munhall, 1989; Kröger et al., 1995; Browman and Goldstein, 1995; Šimko and Cummins, 2010; Sorensen and Gafos, 2016; Sorensen et al., 2019), which makes it a dominant approach. The choice of STAM is motivated by the fact that it has been proved to be more accurate than CDO for fitting and generating articulatory trajectories of speech (Birkholz and Hoole, 2012; Birkholz et al., 2017), and is therefore a good candidate for being a relevant baseline. This section presents CDO-based models and STAM.

2.1. Critically damped oscillators

In Task-Dynamics based paradigms, the movements of speech articulators are modeled as those of oscillators for which the damping coefficient is set so that oscillators are critically damped. This is done to prevent overshoot and oscillation. The movement of such oscillators are driven by the classic second-order equation of motion

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

where x is the distance between the position of the articulator and the asymptotic target (*i.e.* the gap), m and k are the mass and the stiffness of the articulator, respectively, and $b = 2\sqrt{km}$ is the damping coefficient.

This model has the advantage of requiring a small set of parameters. Indeed, as the mass m is usually set to 1 and b is set according to k and m, the model only requires the knowledge of stiffness k and the target position, denoted x_T in this paper. However, the main drawback is the difficulty to obtain an accurate fit to movement trajectories due to a velocity profile that diverges from those observed in real articulatory movements, that is where the velocity peak is too early. In order to address this issue, some critically damped oscillation models include a gradual activation function to shape the velocity profile accordingly (Kröger et al., 1995; Byrd and Saltzman, 1998, 2003). With a gradual activation function $a(t)$, assuming $m = 1$, Eq. (1) becomes

$$\ddot{x} + a(t)\left[b\dot{x} + kx\right] = 0. \tag{2}$$

Note that by setting $a(t)$ as a step function, one gets Eq. (1). In this paper, we will adopt the activation function proposed by Kröger et al. (1995), defined as

$$a(t) = \begin{cases} 0, & t < t_0 \\ \sin\left[\frac{2\pi(t-t_0)}{4(t_1-t_0)}\right], & t_0 \leq t < t_1 \\ 1, & t_1 \leq t < t_2 \\ \sin\left[\frac{2\pi(t-T)}{4(t_2-T)}\right], & t_2 \leq t < T \\ 0, & t \geq T, \end{cases} \tag{3}$$

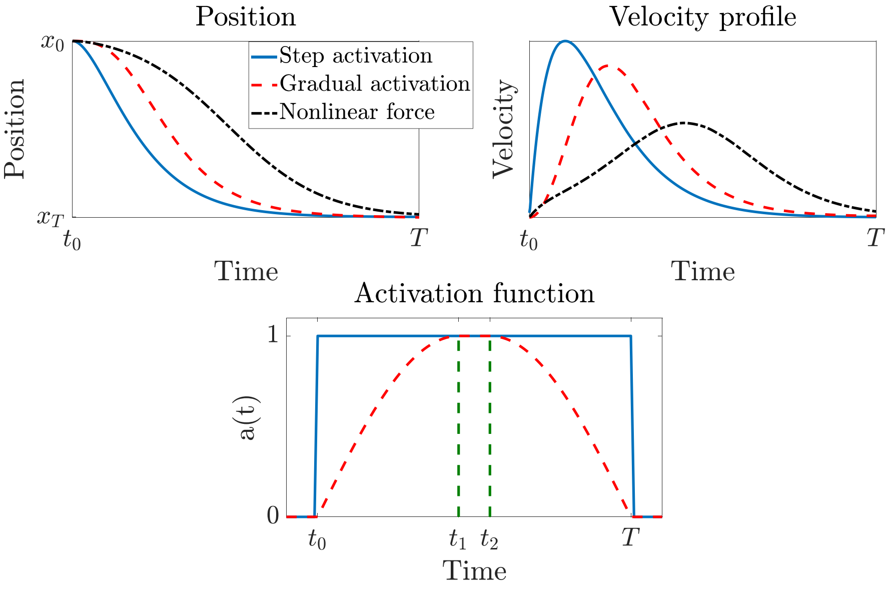

Fig. 1. An example of movements of critically damped oscillators with a step activation function (solid line –), a gradual activation function (dashed line - -), and a nonlinear restoring force (dash-dotted line -·-). The top left panel shows the trajectories, the top right shows the velocity profiles, and the bottom plot shows the activation functions. Note that the nonlinear force model uses a step activation function.

where t_0 and T denote the onset and the offset of the movement, respectively, and t_1 and t_2 are two time points defining the rise and fall intervals of the gradual function.

More recently, Sorensen and Gafos (2016) proposed to add a nonlinear restoring force to Eq. (1), as follows:

$$\ddot{x} + b\dot{x} + kx - dx^3 = 0, \tag{4}$$

where $d \in [0, k[$. According to the authors in Sorensen and Gafos (2016), this nonlinear version of CDO is able to reproduce the observed symmetrical velocity profile feature of speech articulatory movements.

Fig. 1 shows an example of movements and velocity profiles of critically damped oscillators with a step activation function (hereafter S-CDO), a gradual activation function (hereafter G-CDO), and a nonlinear restoring force (hereafter NL-S-CDO). It illustrates the modification of the velocity profile by applying the gradual activation function and the nonlinear restoring force. In this example, with a nonlinear restoring force, the proportional time to peak velocity, which corresponds to the time ratio between the peak velocity instant and the movement duration (a measure of velocity profile (a)symmetry) is around 0.45, which is closer to those observed in practice (close to 0.5, corresponding to a symmetric velocity profile) (Ostry et al., 1987; Byrd and Saltzman, 1998; Perkell and Zandipour, 2002).

Although the introduction of a gradual activation function or a nonlinear restoring force improves the fit to real movements, these methods suffer the drawback of adding more degrees of freedom for characterizing articulatory trajectories. This makes analysis of speech using these models more complex, and also slows down the optimization process that must be used to fit observed articulatory trajectories.

2.2. The sequential target approximation model

Birkholz et al. (2010) proposed to model the trajectories as a cascade of N identical first-order linear systems, each of these having the following transfer function $H(s)$:

$$H(s) = \frac{1}{(1 + s\tau)^N}, \tag{5}$$

where s is the complex frequency and τ is the time constant of the linear system.

In the time-domain, the system is then characterized by the following differential equation for $x(t)$:

$$\binom{N}{0}\tau^N x^{(N)} + \binom{N}{1}\tau^{N-1}x^{(N-1)} + \cdots + \binom{N}{N}x^{(0)} = x_T(t), \tag{6}$$

where $\binom{n}{k}$ is the binomial coefficient, $x^{(i)}$ is the ith derivative of $x(t)$, and $x_T(t)$ is the target function. The model allows $x_T(t)$ to be any function of time, but in practice, the target for each command is either fixed to a constant value (Birkholz et al., 2010), or to a linearly changing value (Xu, 2004; Birkholz and Hoole, 2012).

In the original paper Birkholz et al. (2010), the authors proposed to set the order of the system (*i.e.* the number of first-order linear systems) to $N = 10$. In a more recent study, N has been reduced down to 6 first-order linear systems (Birkholz and Hoole, 2012), as the authors consider it to be a good trade-off between a system that accurately fits the trajectories and one that does not induce unreasonable delays between the input command and the output trajectory. One feature of this method is that it conserves the system state from one command to the next, up to the Nth derivative. This has the advantage of preventing potential discontinuities in low-order derivatives that can occur for trajectories generated by models based on second-order linear systems (i.e. critically damped oscillators). However, this leads to a delay between the activation of the command and the moment from which the system moves towards the target.

Fig. 2 shows an example of an articulatory sequence modeled using STAM, for different time constants.

3. General Tau theory

This section introduces general Tau theory and its application to speech. General Tau theory has been applied to several kinds of movements, including plummeting gannets (Lee and Reddish, 1981), hitting a falling ball (Lee et al., 1983), suckling in babies (Craig and Lee, 1999), and musical performances (Schögler et al., 2008; Rodger et al., 2013). Its plausible application to speech articulation was suggested in Turk and Shattuck-Hufnagel (2020a,b). The current paper develops these ideas further via (i) a comparison of Tau theory with other models of speech trajectory formation, and (ii) a demonstration of speech analysis using this theory.

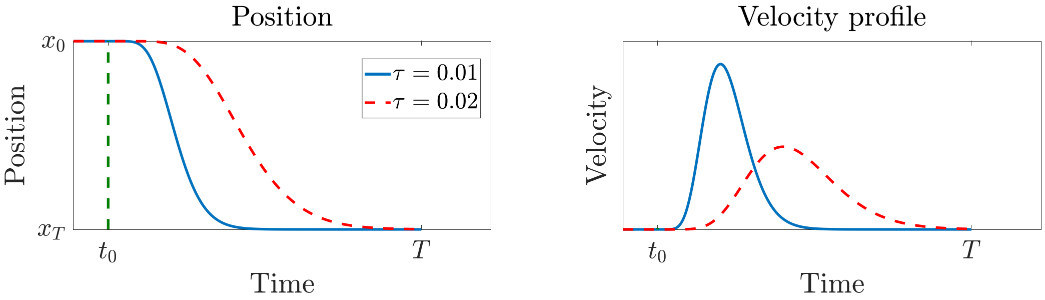

Fig. 2. An example of movements modeled using STAM with different time constants, namely $\tau = 0.01$ (solid line –) and $\tau = 0.02$ (dashed line - -). The left panel shows the trajectories and the right panel shows the velocity profiles.

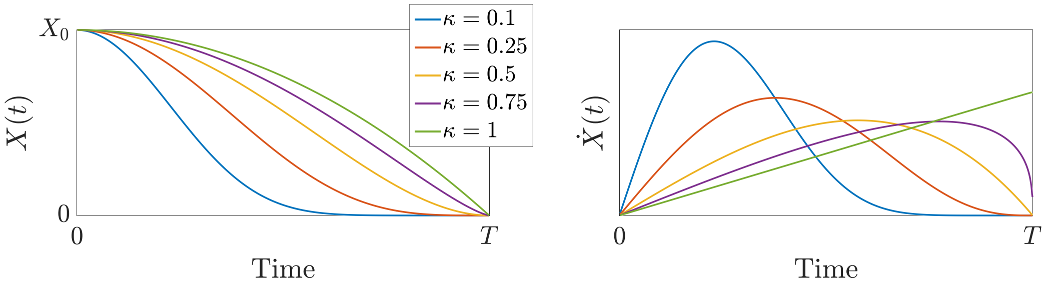

Fig. 3. Gap functions and velocity profiles of Tau-guided movements for various values of κ. The left plot displays Tau-guided movements for different values of κ. The right plot displays their corresponding velocity profiles.

3.1. Presentation

General Tau theory states that voluntary movements aim at closing a gap between the current state of an effector and its target state. One interesting feature of the theory is that, for a given movement of duration T and amplitude A, the timecourse of the movement is controlled via a single quantity, denoted by $\tau_X(t)$, which is defined as the gap function $X(t)$, divided by the gap-closing velocity $\dot{X}(t)$ (the derivative of the gap function $X(t)$ with respect to time). Coordination of movements that aim at closing several gaps (*e.g.* $X(t)$ and $Y(t)$) is ensured by *Tau-coupling*, namely the τs of the gap $X(t)$ and $Y(t)$ are kept at a constant ratio $\kappa_{X,Y}$, such that $\tau_X(t) = \kappa_{X,Y}\tau_Y(t)$. This ensures that gaps are closed simultaneously, regardless of the initial size of the gaps. In that case, $\tau_Y(t)$ is the Tau-guide of $\tau_X(t)$. When the planned movement involves only one gap to close, namely in the absence of an extrinsic guide, the movement is guided by an intrinsic Tau-guide, denoted $\tau_G(t)$, such that $\tau_X(t) = \kappa_{X,G}\tau_G(t)$. The Tau-guide function $\tau_G(t)$ is derived from Newton's law of motion as follows:

$$\tau_G(t) = \frac{1}{2}\left(t - \frac{T^2}{t}\right), \tag{7}$$

where T is the duration of the gap closure and t runs from 0 to T. Consequently, Tau-guided movements are governed by the following differential equation

$$\tau_X(t) = \frac{X(t)}{\dot{X}(t)} = \kappa_{X,G}\tau_G(t) = \frac{\kappa_{X,G}}{2}\left(t - \frac{T^2}{t}\right). \tag{8}$$

One can see that the Tau-guide τ_G ensures that the gap closes and reaches its target at the movement endpoint, namely when $X(T) = 0$. One other interesting feature, as shown by Eq. (8), is that, given an initial gap X_0 and gap-closure duration T, the gap-closing function depends on only one variable, namely the Tau-coupling parameter $\kappa_{X,G}$. Modifying the value of $\kappa_{X,G}$ will shape the velocity profile, as shown in Fig. 3, which displays Tau-guided movements and their corresponding velocity profiles for various values of $\kappa_{X,G}$. For the sake of simplicity, $\kappa_{X,G}$ is simply denoted κ in the rest of the paper.

3.2. Velocity profiles of Tau-guided movements

Like other practiced, voluntary movements, speech movements are characterized by smooth, single-peaked velocity profiles, which are often symmetric (Munhall et al., 1985; Ostry et al., 1987). These studies also observed that the ratio between the peak velocity, denoted $V_{\max}$ and the average velocity is relatively constant, hence the following relationship:

$$\frac{V_{\max}T}{X_0} = c, \tag{9}$$

with c a constant, X_0 is the amplitude of movement and T is the movement duration. Therefore, implementing a model of articulatory trajectory formation able to reproduce profiles with similar characteristics is essential for the relevance of the application of Tau theory to speech. This section analyzes the theoretical relationships between the parameters of Tau-guided movements and the characteristics of the resulting velocity profiles.

For that purpose, we consider a Tau-guided movement initiated at $t = 0$ and ending at $t = T$, starting at an initial gap $X(0) = X_0$, and having a shape parameter κ. A common descriptor of velocity profiles is the proportional time-to-peak velocity t_{ppv}. It is defined as the time t_p at which the peak velocity occurs divided by the movement duration, namely $t_{ppv} = \frac{t_p}{T}$.

First, let us solve the differential equation (8) to get the gap function $X(t)$. This yields

$$X(t) = X_0\left(1 - \frac{t^2}{T^2}\right)^{\frac{1}{\kappa}}. \tag{10}$$

The velocity and acceleration of the gap closure can then be derived from the gap function:

$$\dot{X}(t) = -\frac{2X_0 t}{\kappa T^2}\left(1 - \frac{t^2}{T^2}\right)^{\frac{1}{\kappa}-1}, \tag{11}$$

$$\ddot{X}(t) = \frac{4X_0 t^2\left(1 - \frac{t^2}{T^2}\right)^{\frac{1}{\kappa}-2}\left(\frac{1}{\kappa} - 1\right)}{\kappa T^4} - \frac{2X_0\left(1 - \frac{t^2}{T^2}\right)^{\frac{1}{\kappa}-1}}{\kappa T^2}. \tag{12}$$

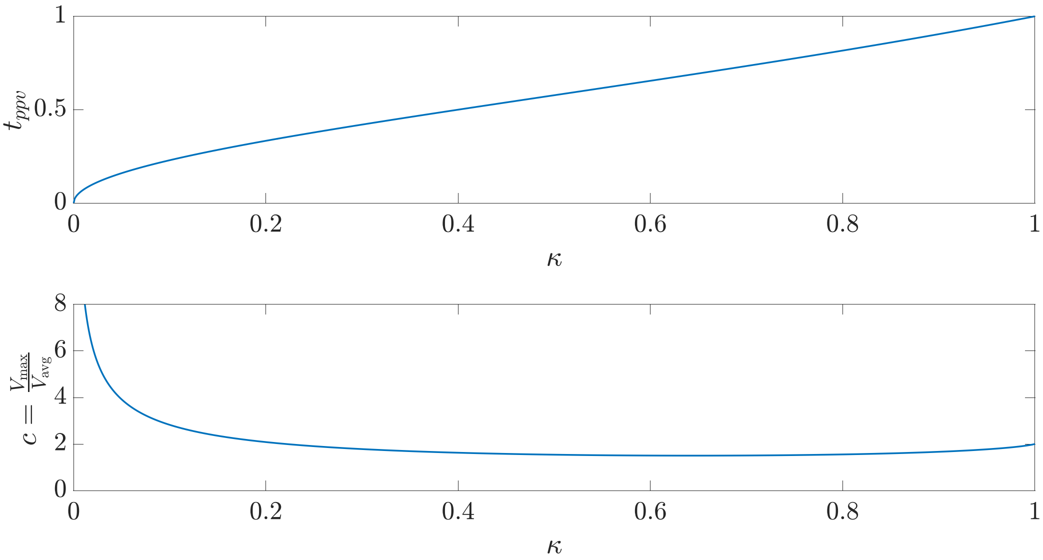

Fig. 4. Parameters of velocity profiles of Tau-guided movements as a function of κ. The top panel shows the proportional time-to-peak velocity. The bottom panel shows the parameter $c = \frac{V_{\max} T}{X_0}$.

Assuming $0 < \kappa < 1$, the velocity function in Eq. (11) reaches a local extremum when the acceleration function $\ddot{X}(t) = 0$, namely for $t_\mathrm{p} = T\sqrt{\frac{\kappa}{2-\kappa}}$. Consequently, the time-to-peak velocity is given by

$$t_{ppv} = \frac{t_p}{T} = \sqrt{\frac{\kappa}{2-\kappa}}. \tag{13}$$

This equation shows that the time to peak velocity of Tau-guided movements depends only on the shape parameter κ. The function $t_{ppv}(\kappa)$ is plotted in Fig. 4: it is monotonically increasing for $0 < \kappa < 1$, with a value of 0.5 reached for $\kappa = 0.4$. This means that Tau-guided movements having a κ-value lower than 0.4 exhibit an accelerating phase shorter than the decelerating phase, while movements with κ larger than 0.4 exhibit a longer accelerating phase. It also shows that for any $0 < t_{ppv} \leq 1$, there is a corresponding Tau-guided movement with $0 < \kappa \leq 1$.

The value of the peak velocity is obtained by substituting t_p into Eq. (11):

$$V_{\max} = |\dot{X}(t_p)| = \frac{2X_0}{T\sqrt{\kappa(2-\kappa)}}\left(1 - \frac{\kappa}{2-\kappa}\right)^{\frac{1}{\kappa}-1}. \tag{14}$$

For Tau-guided movements of duration T and shape parameter κ, peak velocity has a linear relationship with movement amplitude X_0, as observed in previous studies (Ostry and Munhall, 1985; Munhall et al., 1985; Ostry et al., 1987). Additionally, one can show that parameter c, defined in Ostry and Munhall (1985), Munhall et al. (1985) and Ostry et al. (1987) as the ratio between the peak velocity $V_{\max}$ and the average velocity ($\frac{X_0}{T}$) is a constant. Indeed, reorganizing Eq. (14) to fit Eq. (9) yields

$$c = \frac{2}{\sqrt{\kappa(2-\kappa)}}\left(1 - \frac{\kappa}{2-\kappa}\right)^{\frac{1}{\kappa}-1}, \tag{15}$$

which is a constant for all movements with a given κ.

Fig. 4 shows the relationship between c and κ for $0 < \kappa \leq 1$. The parameter c varies rapidly for small κ (≤ 0.2), and then varies slowly for $\kappa > 0.2$. It basically stays above $c = 1.5$, which is the minimum reached for $\kappa = 0.64$, and below $c = 2$. These values correspond to typical values observed in speech (*cf.* Munhall et al., 1985; Ostry et al., 1987 for instance).

These mathematical developments show that Tau-guided movements have a 1-to-1 relationship between the κ parameter and the skewness of the velocity profile (defined as the proportional time-to-peak velocity). In addition, we show that Tau-guided movements with a given κ have a constant c, consistent with findings in previous studies (Munhall et al., 1985; Ostry et al., 1987).

4. Experiments

This section presents experiments which evaluate the application of general Tau theory to articulatory movements in speech. We first fit Tau-guided gap functions to observed trajectories. In order to compare our model with existing models, the fitting errors of the Tau-guided functions are then compared with those obtained by fitting the CDO-based and STAM models to the same trajectories.

4.1. Data

Data used in these experiments come from a reading task in the DoubleTalk corpus (Scobbie et al., 2013; Geng et al., 2013). The corpus consists of synchronous EMA and audio recordings for each of 6 mixed-dialect pairs, resulting in a total of 12 native speakers of English. Five speakers have a Southern English accent (labeled SE), 5 have a Scottish accent (labeled SC), one has a Northern English accent (labeled NE), and one has a General American accent (labeled GA). We modified the labeling pattern of speakers from the DoubleTalk corpus to include information about accent, as shown in Table 1. The corpus includes several speech tasks, including spontaneous monologue, spontaneous conversation, repetition from memory, shadowing, and read speech. The experiments detailed in this paper were conducted solely on the read speech part of the corpus. The read speech task required speakers to read *Comma Gets a Cure* (Honorof et al., 2000), designed to showcase phonemes of English that exhibit significant phonetic variation across dialects. This story was adapted for Scottish English by Scobbie et al. (2013).

EMA data were collected using two synchronized Carstens AG500 electromagnetic articulometers at an acquisition rate of 200 Hz. Data consists of 3D positions and rotations of 12 sensors attached to the vermilion borders of the upper and lower lips, intra-orally, and on the head. Sensors attached on fixed parts of the head were used to correct for head movement, i.e. to remove low frequency head movements from the articulator sensor movements. Synchronized speech audio data was collected by means of Articulate Instruments Ltd. hardware.

Table 1
Modifications of the speaker labeling pattern from the DoubleTalk corpus.

In DoubleTalk	In this paper
R0020_cs5	1CS5NE
R0020_cs6	1CS6GA
R0033_cs5	2CS5SC
R0033_cs6	2CS6SE
R0034_cs5	3CS5SC
R0034_cs6	3CS6SE
R0035_cs5	4CS5SC
R0035_cs6	4CS6SE
R0036_cs5	5CS5SC
R0036_cs6	5CS6SE
R0039_cs5	6CS5SC
R0039_cs6	6CS6SE

4.2. Dealing with 2D movements: dimensionality reduction

In these experiments, we analyzed the trajectories of sensors attached to the lower and the upper lips, on the lower jaw, and on 3 points arranged midsagittally on the tongue. The tongue tip sensor was attached less than or equal to 1 cm from the tip. The tongue back sensor was attached as far back as was feasible, and the tongue mid sensor was approximately equidistant between the tongue tip and tongue back sensor. The tongue sensors were separated from each other by 1–2 cm. Additionally, in order to investigate the movement of task-related variables, we computed lip aperture, defined as the euclidean distance between the lower and the upper lips at every time sample. In order to align with the naming convention of existing articulatory models, the tongue back and tongue mid sensors will be labeled as *Tongue Dorsum* (TD) and *Tongue body* (TB) for the rest of the paper. We disregarded the x (lateral) dimension of the data, and used only the signals in the sagittal plane, namely in the y (sagittal) axis and in the z (vertical) axis. Since all of the investigated models of articulatory trajectories generate 1D signals, analysis using these models requires a reduction of dimensions to transform 3D or 2D position signals into a 1D signal. Birkholz et al. (2010) proposed to consider the analyzed 1D trajectory as the projection of 2D trajectories onto the first principal component, extracted from PCA. One alternative way of generating a 1D signal is to use parametric static articulatory models, namely models that define the geometry of the vocal tract using a few parameters, and then applying trajectory models to these parameters. For instance, CDO-based trajectory models are applied to the CASY articulatory model in TADA (Nam et al., 2004). Similarly, Šimko and Cummins (2010) proposed to use CDO-based models to generate the timecourse of the parameters of a simplified articulatory model. However, this alternative is less adapted for articulatory analysis as it requires the estimation of articulator parameters from the observation of their position (Toutios et al., 2011). Both methods add uncertainty about the real articulator position, due to the fact that the first principal component does not explain 100% of the variance for the PCA method, and due to limits on the precision of the articulatory model and to limits on the robustness of the inversion method for the method using an intermediate articulatory model. We chose to analyze the trajectories defined as the projection of articulator position data onto the first principal component extracted from Principal Component Analysis. PCA was applied to position data of each individual speaker and each individual sensor. Note that this was not applied to lip aperture, since it is already a 1D signal by definition. A 4-order 1D Gaussian filter was then applied to signals to smooth the trajectories. The aim of the filtering was to remove small fluctuations in the position signal that would perturb the articulatory segmentation as detailed in Section 4.3. For our data, using a Gaussian filtering of order 4 has been found to be a good trade-off between a good rejection of the fluctuations, and reasonable filtering that prevents overfitting and unrealistically long movements.

Table 2
Distribution of inter-pause intervals across speakers and sensors. TD, TB, and TT denote tongue dorsum, tongue body and tongue tip, respectively. LL and UL denote lower and upper lip, respectively, and LA denotes lip aperture.

Speakers	TD	TB	TT	Jaw	LL	UL	LA	All
1cs5NE	53	53	53	53	53	53	53	371
1cs6GA	24	24	24	24	24	0	0	120
2cs5SC	31	31	31	31	31	31	31	217
2cs6SE	42	42	31	42	42	42	42	283
3cs5SC	39	39	39	39	39	39	39	273
3cs6SE	36	36	34	36	36	36	36	250
4cs5SC	27	27	27	27	27	26	26	187
4cs6SE	50	50	50	50	50	44	44	338
5cs5SC	41	40	40	41	41	41	41	285
5cs6SE	36	40	40	40	40	40	40	276
6cs5SC	18	18	15	18	18	18	18	123
6cs6SE	47	47	47	47	47	23	23	281
All	444	447	431	448	448	393	393	3004

Table 3
Distribution of articulatory segments across speakers and sensors. TD, TB, and TT denote tongue dorsum, tongue body and tongue tip, respectively. LL and UL denote lower and upper lip, respectively, and LA denotes lip aperture.

Speakers	TD	TB	TT	Jaw	LL	UL	LA	All
1cs5NE	1183	1217	1384	1243	1332	1255	1296	8910
1cs6GA	723	726	732	711	769	0	0	3661
2cs5SC	1145	1140	1246	1252	1247	1050	1220	8300
2cs6SE	1419	1425	824	1329	1466	1243	1475	9181
3cs5SC	1034	1085	1171	1094	1223	1205	1204	8016
3cs6SE	1122	1149	1198	1163	1254	1227	1197	8310
4cs5SC	990	1004	1182	1050	1143	794	946	7109
4cs6SE	1124	1146	1302	1344	1387	918	1124	8345
5cs5SC	1149	1014	1120	1031	1230	1045	1249	7838
5cs6SE	1063	1181	1318	1195	1306	1162	1354	8579
6cs5SC	1029	1050	737	1086	1061	917	1057	6937
6cs6SE	1071	1108	1248	1081	1181	467	569	6725
All	13 052	13 245	13 462	13 579	14 599	11 283	12 691	91 911

4.3. Articulatory segmentation

The EMA signals were segmented at two levels. First, they were segmented into inter-pause intervals. Each inter-pause interval segment is then a sequence of articulatory movements between two pauses. Pauses at the onset and offset of inter-pause interval segments are included for the purpose of fitting only (as explained in Section 4.4). Due to experimental factors, e.g. sensor tracking difficulties, some inter-pause intervals were discarded from the analysis because they contained only noise. Finally, this resulted in a total of 3004 EMA inter-pause intervals to analyze. The second step consisted of segmenting inter-pause interval segments into articulatory movement units. For that purpose, we used a zero-crossing method. We defined movement units as time segments between two successive local extrema of the position signal. Local extrema correspond to zero-crossings of the time derivative of position, i.e. the velocity signal. Fig. 5 shows an example of segmentation using velocity zero-crossings.

The distribution of analyzed inter-pause intervals across speakers and sensor type is detailed in Table 2, while Table 3 details the distribution of analyzed articulatory segments across speakers and sensor types. Note that the signal from the upper lip sensor (UL) on speaker 1cs6GA was not valid due to a broken sensor, hence no inter-pause intervals have been analyzed on the upper lip for this speaker. The same applies for the lip aperture for this speaker, as it is derived from both the lower lip (LL) and upper lip (UL) signals.

4.4. Curve fitting

The comparison of studied models of articulatory trajectories is made using curve fitting. This section details the different methods for curve fitting to observed articulatory trajectories.

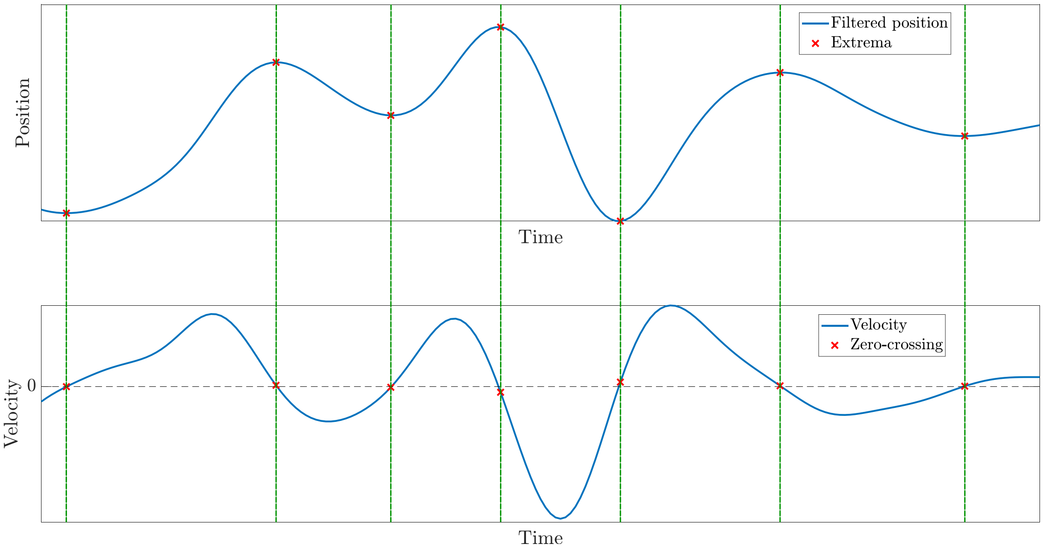

Fig. 5. Velocity Zero-Crossing segmentation. The top plot displays the filtered position signal and the found local extrema. The bottom plot displays the velocity and the estimated zero-crossing. The positions of the zero-velocity are marked by red crosses. Vertical dashed lines show the boundaries of individual segments.

All methods use an optimization algorithm which minimizes a defined cost function, *i.e.* a fitting error function. In order to allow fair comparison between methods, we used the same cost function and the same optimization algorithm as proposed in Birkholz et al. (2010) for all of them. The optimization procedure uses the Nelder–Mead simplex method (Nelder and Mead, 1965) to find the global minimum of a scalar objective function. This objective function represents the dissimilarities between the observed trajectory and the reproduced trajectory generated by the model. It is computed for each sequence s of M movement units (segmented following the method described in Section 4.3) inside the analyzed inter-pause interval. The cost function C is defined as

$$C(\theta) = \frac{\sqrt{\left[\sum_n w_n (s_n - \tilde{s}_n(\theta))^2\right] / \sum_n w_n}}{s_{\max} - s_{\min}}, \tag{16}$$

where $\theta \in \mathbb{R}^{lM}$ is the vector containing the l parameters to optimize for the M movement units, s_n is the observed sequence of movement units (at the inter-pause level) at sample n, $\tilde{s}_n(\theta)$ is the modeled trajectory of the inter-pause interval at sample n for the vector of parameters θ, $s_{\max}$ and $s_{\min}$ are the maximum and minimum of the observed inter-pause signal s, respectively, and

$$w_n = 1 + a\frac{v_n^2}{v_{\max}^2}, \tag{17}$$

is a weight signal used to account for change in velocity. Here, v_n is the velocity profile at sample n, $v_{\max}$ is the peak velocity, and $a = 5$ is a factor used to specify the relative importance of position versus velocity in the fitting process. This objective function is very similar to the one used in Birkholz et al. (2010). We chose to use the same cost function as a baseline method for all tested models in our paper to allow fair comparisons. Since this paper intends to compare fits for different speakers, signals, and phonetic contexts, we slightly modified the cost function to include a normalization. The cost function is normalized by the range of signal values within the inter-pause interval, namely $s_{\max} - s_{\min}$.

This optimization procedure requires an initial estimate, which strongly influences the final solution. Indeed, different initial estimates will give different solutions. Consequently, as suggested in the original paper (Birkholz et al., 2010), we ran a set of 100 optimization procedures for each sequence, with different random initial estimates, and we adopted the solution that provides the best fit as the solution to keep. Note that, as detailed in Section 4.4.3, the Tau-fitting method does not require multiple optimization processes, as the solution does not depend on the initial estimate.

4.4.1. *Fitting with the sequential target approximation model*

The fitting method using STAM is roughly the same as the one used in the original paper (Birkholz et al., 2010), including the changes detailed in Birkholz and Hoole (2012). The parameters to optimize are the onset time, the initial target, the time constant, and the slope of the target function of each command. The size of θ is then $4M$ (4 parameters to optimize for each movement unit).

The initial estimate parameters were set as follows. The initial targets for ascending movements were set to the maximum of the movement unit plus a random value between 0 and 1 cm (or the minimum minus a random value between 0 and 1 cm for descending trajectories). The initial onset times were set to a random value between 50 and 100 ms before the onset of the movement unit, to account for the delay induced by the cascade of first-order linear systems. The initial constant times were set to a random value between 0.01 and 0.02. Finally each initial slope for the target function was set to a random value between 0 and 50, multiplied by −1 for descending movements. Each random value was drawn from a uniform distribution.

4.4.2. *Fitting with critically damped oscillators*

Fitting with CDO-based techniques was done similarly to fitting with STAM. The optimization algorithm was also the Nelder–Mead simplex method (Nelder and Mead, 1965), and the objective function was similar to the one in Eq. (16). The differences were in the choice of the parameters to optimize. NL-S-CDO needs to optimize the target, the stiffness value, and the nonlinear constant value d; as a result the size of θ is $3M$, while G-CDO also needs to optimize the activation function parameters, namely the values of t_1 and t_2 of Eq. (3), relative to the movement duration T, in addition to the target and the stiffness value. Consequently, the size of θ is $4M$.

Similarly to STAM, a set of 100 optimization procedures with random initial estimates were run to obtain the best fit. Initial targets were set the same way as in the STAM fitting procedure. Initial stiffness values were taken as random numbers between 0 and 1000, and initial t_1 and t_2 were set to random numbers such that $0 < t_1 < t_2 < T$.

4.4.3. *Fitting with general Tau theory*

One advantage of curve fitting with the general Tau theory equation over STAM and CDO is that there is only one parameter to optimize, namely the Tau-coupling value κ. This is because the other parameters in the equation are directly observable from the EMA trajectories. The objective function of Eq. (16) can then be reduced to a one-dimensional

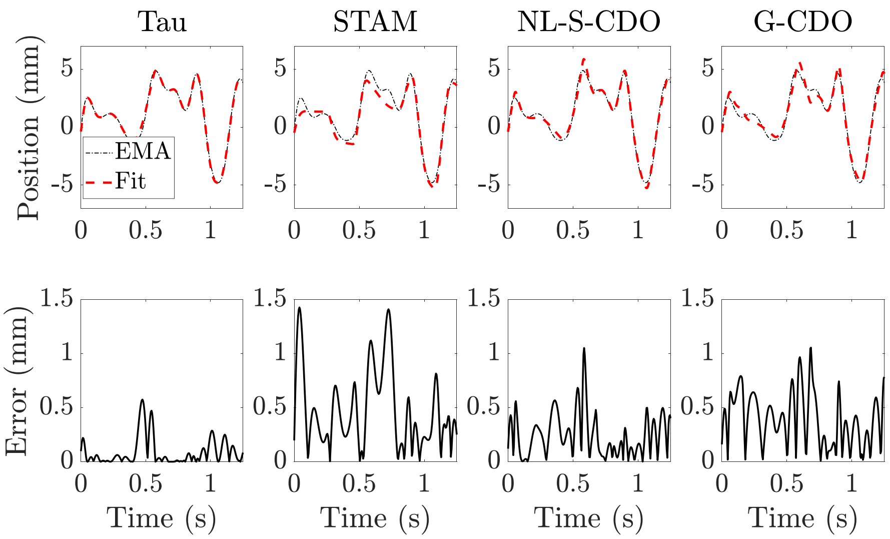

Fig. 6. Example of fits obtained with the different methods, namely Tau, STAM, NL-S-CDO, and G-CDO, from left to right. The fitted curve is a small section of an actual jaw position signal extracted from the analyzed corpus. The top plot shows the result of the fit and the bottom plot shows the absolute difference, in mm, between the fit and the EMA trajectory.

Table 4
Degrees of freedom (DOF), $F-$statistic and the corresponding $p-$value between parentheses for the 3 repeated-measures 3-way ANOVAs.

Paired-Test	DOF	Tau vs. STAM	Tau vs. NL-S-CDO	Tau vs. G-CDO
Sensors	6	51 (p = 1.1×10^{-61})	9.3 (p = 3.4×10^{-10})	17 (p = 5.6×10^{-20})
Speakers	11	30 (p = 2×10^{-61})	24 (p = 1.9×10^{-48})	33 (p = 3.8×10^{-68})
Methods	1	6.7e+03 (p = 0)	1.1e+04 (p = 0)	1e+04 (p = 0)
Sensors $*$ Speakers	64	1.5 (p = 0.0085)	2.8 (p = 1.9×10^{-12})	1.9 (p = 1.3×10^{-05})
Sensors $*$ Methods	6	63 (p = 2.2×10^{-76})	15 (p = 6.4×10^{-17})	23 (p = 7.3×10^{-27})
Speakers $*$ Methods	11	27 (p = 3.9×10^{-56})	21 (p = 1.8×10^{-43})	32 (p = 6.4×10^{-67})
Sensors $*$ Speakers $*$ Methods	64	0.96 (p = 0.57)	1.4 (p = 0.026)	1 (p = 0.37)

function, applied to each movement unit, with $\theta = \kappa$. In addition, $C(\kappa)$ is shown to admit only one minimum, which implies that the optimization algorithm always converges towards the same solution. Therefore, the solution does not depend on the initial estimate, hence there is no need to run several optimization processes. This feature significantly reduces the optimization burden, and the solution provides a global minimum of the objective function.

4.4.4. *Example of fit*

Fig. 6 shows example fits obtained with the different methods. The fitted curve is a small section of an actual jaw position signal extracted from the analyzed corpus. In this example, the Tau theory equation provides the best fit ($C = 0.017$), followed by NL-S-CDO ($C = 0.031$), G-CDO ($C = 0.043$), and STAM ($C = 0.054$). Tau Theory generates a good fit for all parts of the trajectory. Trajectories generated by other methods exhibit regions where fits clearly diverge from the observed trajectory. These regions correspond to parts of movement where the trajectory changes its direction, i.e. when the velocity is low. This is certainly due to the choice of the cost function C of Eq. (16), which gives less penalty in these regions. Note that, as explained previously in this section, this cost function has been adapted from Birkholz et al. (2010) to achieve the best fit possible with STAM, as suggested by the authors.

4.5. *Results*

Fits have been performed using the 4 models (Tau, STAM, NL-S-CDO, and G-CDO) on all inter-pause intervals and for every sensor. This section presents the results of the fitting errors when using these models.

4.5.1. *Statistical analysis*

We conducted 3 repeated-measures 3-way ANOVAs to assess the effect of the independent variables methods, speakers and sensors on the error fit. These analyses allow us to assess the following comparisons: (1) Tau vs. STAM, (2) Tau vs. S-CDO, and (3) Tau vs. G-CDO. The analyzed independent variables were speakers, sensors, and methods. Table 4 reports the $F-$statistic and the corresponding $p-$ value between parentheses. All main effects and interactions were significant, except for the Speaker*Sensor*Method interaction for all analyses. One possible explanation for why the fits change for different sensors and speakers with STAM and CDO could be that these methods show highly variable behavior, even for the same sensor used by the same speaker.

4.5.2. *Impact of the PCA on fitting error*

As discussed in Section 4.2, using PCA for dimensionality reduction adds uncertainty as the first principal component does not explain 100% of the variance. This is because articulatory sensors do not move along a straight axis, and the deviation of movement from the principal component depends on sensors and speakers.

In order to assess the effect of the explained variance on the fitting error, we computed Pearson correlation coefficients for each sensor and method. Results are shown in Fig. 7. Correlation coefficients show a weak linear correlation between the fitting error and explained variance ($|r| \leq 0.34$). For STAM and CDO based methods, the correlation coefficient is negative, meaning that the fitting error tends to slightly decrease when the explained variance increases. Surprisingly, Tau theory exhibits an inverse relationship, as correlation coefficients are mostly positive: the fitting error tends to increase with the explained variance. However, the correlation coefficient for Tau is very weak ($|r| \leq 0.11$). This observation suggests that the explained variance has a very marginal impact on the trajectory fitting using Tau.

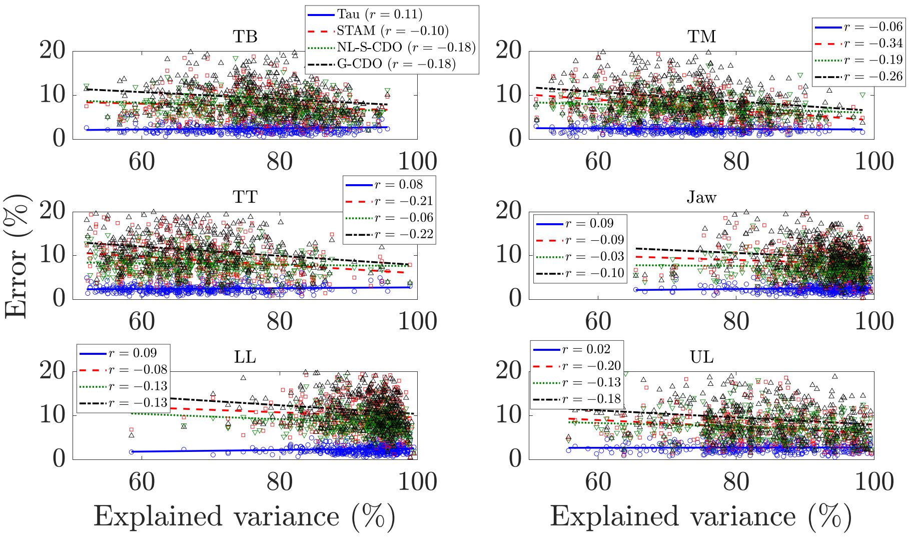

Fig. 7. Fitting error as a function of the explained variance for each articulatory sensor and each method. ('o' = Tau, '□' = STAM, '▽' = NL-S-CDO, and '△' = G-CDO). Lines represent the linear regression applied to the fitting vs. explained variance (Tau is represented by a solid line, STAM by a dashed line, NL-S-CDO by a dotted line, and G-CDO by a dash-dotted line). r denotes the Pearson correlation coefficient.

4.5.3. Presentation of results

Fig. 8 shows the distribution of fit errors for the 4 methods among speakers and sensors. The error is the minimum of the cost function in Eq. (16), expressed in %, returned by the optimization process.

Results show that the fit error is lowest for the Tau theory equation, followed by STAM and NL-S-CDO, which perform slightly better than G-CDO. When grouping all signals (all speakers and all sensors), the median fit error for Tau is 2.38%, 7.49% for STAM, and 7.48% and 9.64% for NL-S-CDO and G-CDO, respectively. This pattern is roughly the same across speakers and sensors. The median fitting error for Tau theory lies between 1.70% (lower lip sensor of speaker 3cs5SC) and 3.17% (lip aperture of speaker 6cs5SC). Curiously, fits on sensor signals from speaker 6cs5SC are almost always the least accurate. For instance, the errors with STAM are larger than 13% for each of the sensors, but only between 5.5% and 10% for other speakers, except for a few cases. These large median errors generally occur when the error variance is large, as is observed for some speakers and sensors. For two speakers (2cs5SC and 6cs5SC), errors using STAM show a much larger variance for lip and jaw sensors than for other sensors and speakers. One can also note a larger variance of STAM-fit errors for speaker 4cs5SC for the tongue-tip sensor. Since such a larger variance for these sensors and speakers is not clearly visible for other methods, this is probably due to the variability of the solution returned by the STAM optimization, as it requires more parameters to optimize. There is no qualitative difference in pattern among sensors. One can note, however, that the tongue tip, the jaw, and the lower lip sensors have larger fitting errors than other sensors for STAM and CDO-based methods, but this does not apply for Tau-fitting, and it is probably due to the aforementioned large variance observed for some speakers.

It is interesting to note that there are no qualitative differences across sensor types, despite being attached to very different articulators. Indeed, the upper lip is an end effector which does not depend on the movements of other analyzed sensor types, while some have position signals which are strongly inter-correlated between each other (*e.g.* tongue sensors). In the current experiments, speech articulation has been treated as a sequence of discrete, non-overlapping trajectories. However, it has been hypothesized in AP/TD that single articulators may sometimes be simultaneously governed by the activation of multiple abstract gestures. This is postulated to be the case for example, for /g/-vowel sequences, in which the tongue body gesture for the consonant partially overlaps with the tongue body gesture for the vowel. The activation functions in the current CDO analyses have not been optimized in a way that takes this type of partial overlap into account, and it is possible that better CDO fits could be achieved with a better model of coarticulation. However, the fact that Tau theory outperformed CDO models for articulators such as the lower lip, upper lip, tongue tip, and also the lip aperture signal, that are not often simultaneously governed by partially overlapping gestures, suggests that our finding of superior Tau theory performance is not exclusively due to our assumptions about independent, non-partially overlapping movements.

In addition, in our experiments, we have treated the movements of each sensor as being independent of movements of the other sensors, even though we know that the movement of the jaw will affect the movement of the tongue and lower lip, and each tongue sensor will be affected by the movements of the other tongue sensors. That is, for Tau theory, we have assumed that each sensor is Tau-guided independently, which is most certainly not the case. However, in spite of this assumption, we see good fits to the Tau theory equation for all sensors, and do not see a better fit for the upper lip sensor which is arguably most independent of other sensors. We therefore see no evidence that the interdependence of movement paths has affected the timecourse of movement, which appears to be Tau-guided for all sensors and speakers.

These results show that modeling articulatory trajectories with general Tau theory gives the most accurate representation of actual trajectories, as it provides the best fit, in comparison with other methods. Unlike other methods which require running many optimization procedures to approximate the global minimum of the cost function, Tau theory always provides the global minimum. This gives more confidence in the returned solution when analyzing articulatory trajectories. This also probably reduces the variance of fitting errors, as our results show that the latter is lower when fitting with Tau theory than for other methods. These observations suggest that Tau theory can be applied to model movements of any articulator, and also to task variables such as lip aperture, which gives a fit accuracy similar to fits on the positions of speech articulators. Our results also show that it applies similarly for each of the 12 analyzed speakers.

5. Speech analysis using general Tau theory

The previous section has provided support for the relevance of applying general Tau theory to speech, and more specifically for analyzing

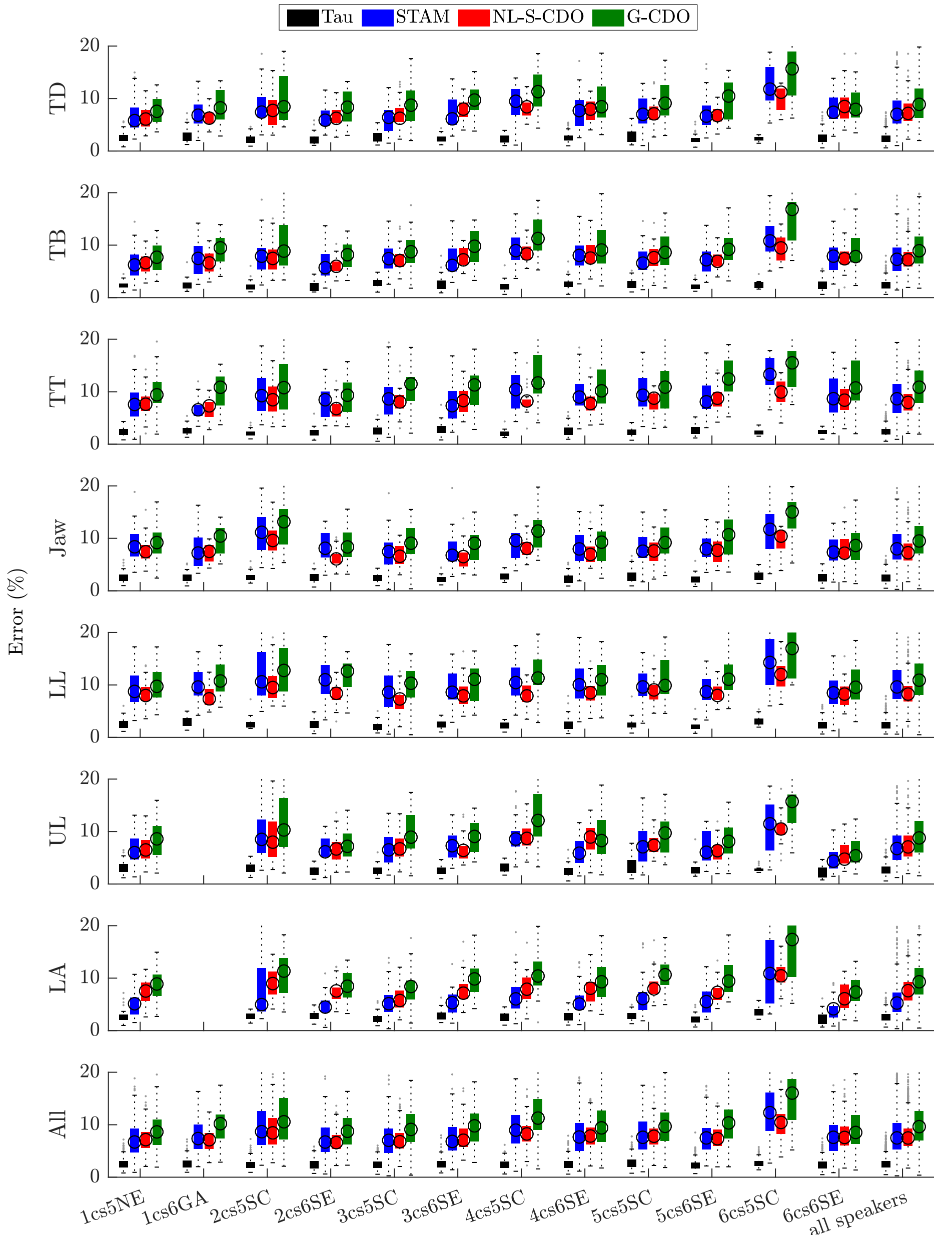

Fig. 8. Fitting errors obtained for the different methods for each speaker and each sensor. The error is the obtained minimum of the cost function in Eq. (16), expressed in %.

speech movement trajectories. Indeed, Tau theory provides a unique solution which fits the analyzed trajectories better than baseline methods. This seems to apply for any sensor or speaker. It also uses a reduced number of parameters to model and characterize a trajectory: initial and final position, onset and offset times, and κ, a shape parameter. In this section, we present a demonstration of speech analysis using general Tau theory: we provide a statistical analysis of the estimated κ-values produced by speakers in the *Comma Gets a Cure* reading task, under the assumption that their movements were Tau-guided.

5.1. Data and analysis

Data used for the analysis study were the same as those used in previous section. The shape parameter κ was estimated by fitting Tau-guided trajectories on observed trajectories. For the statistical analysis, we introduce a rejection criterion to discard shape parameters estimated from fits that are considered as non satisfying, namely fits for which $\kappa \leq 0$ or $\kappa \geq 1$. This is the case for 2.15% of the analyzed data.

5.2. Distribution of κ

Fig. 9 shows the individual statistical distributions of estimated κ for all speakers and sensors. All distributions are of similar shape, namely a slightly right-skewed unimodal distribution. Skewness, defined as the third statistical moment, is always positive, ranging from 0.29 (upper lip sensor of speaker 3cs5SC) to 0.82 (jaw sensor of speaker 4cs6SE). The excess kurtosis, defined as the fourth statistical moment, ranges from −0.43 (upper lip sensor of speaker 6cs5SC) to 1.6 (jaw sensor of speaker 4cs6SE). Most of the distributions (≃83%) are leptokurtic, namely with a positive excess kurtosis, suggesting a slight tendency for κ-values to exhibit a sharp distribution centered around its mode.

Modes and standard deviations of κ are displayed in Table 5. Distribution modes have been estimated using a Kernel density estimation. All distributions have very similar modes, ranging from 0.376 (tongue tip sensor of speaker 1cs6GA) to 0.439 (jaw sensor of speaker 6csSE), with most of the values (≃90%) being between 0.38 and 0.42. They also have similar standard deviations, ranging from 0.13 (jaw sensor of 4cs6SE) to 0.20 (upper lip sensor of speaker 6cs5SC).

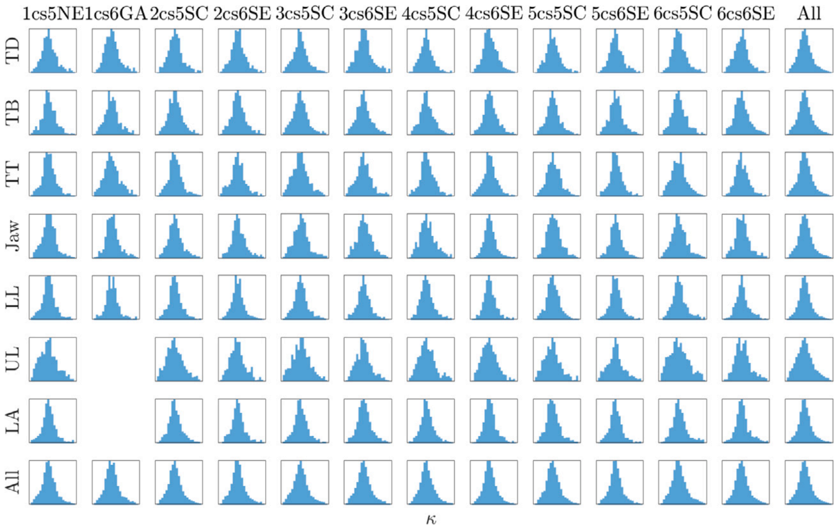

Fig. 9. Individual statistical distribution of estimated κ for all speakers (columns) and sensors (rows). The x-axes represent the κ values and the y-axes represent the normalized distribution (the peak is at 1).

Table 5
Modes and standard deviation (in parentheses) of κ estimated for each speaker and sensor.

Speakers	TD	TB	TT	Jaw	LL	UL	LA	All
1cs5NE	0.397 (0.16)	0.386 (0.15)	0.391 (0.15)	0.422 (0.15)	0.409 (0.15)	0.405 (0.18)	0.399 (0.15)	0.400 (0.16)
1cs6GA	0.403 (0.16)	0.391 (0.16)	0.376 (0.17)	0.413 (0.15)	0.370 (0.15)	–	–	0.381 (0.16)
2cs5SC	0.430 (0.16)	0.419 (0.15)	0.390 (0.15)	0.407 (0.16)	0.397 (0.14)	0.389 (0.18)	0.394 (0.15)	0.406 (0.15)
2cs6SE	0.417 (0.16)	0.395 (0.16)	0.399 (0.16)	0.401 (0.16)	0.390 (0.14)	0.386 (0.16)	0.408 (0.15)	0.396 (0.15)
3cs5SC	0.394 (0.16)	0.393 (0.16)	0.405 (0.17)	0.429 (0.16)	0.408 (0.15)	0.427 (0.18)	0.393 (0.15)	0.406 (0.16)
3cs6SE	0.414 (0.17)	0.406 (0.16)	0.396 (0.17)	0.403 (0.16)	0.408 (0.14)	0.403 (0.16)	0.403 (0.15)	0.405 (0.16)
4cs5SC	0.407 (0.15)	0.391 (0.16)	0.387 (0.15)	0.394 (0.17)	0.406 (0.15)	0.384 (0.18)	0.394 (0.15)	0.397 (0.16)
4cs6SE	0.385 (0.15)	0.407 (0.15)	0.400 (0.15)	0.404 (0.13)	0.403 (0.14)	0.386 (0.16)	0.403 (0.14)	0.405 (0.15)
5cs5SC	0.381 (0.15)	0.408 (0.15)	0.395 (0.15)	0.402 (0.15)	0.398 (0.14)	0.408 (0.18)	0.390 (0.15)	0.400 (0.15)
5cs6SE	0.395 (0.15)	0.395 (0.15)	0.403 (0.14)	0.404 (0.14)	0.384 (0.14)	0.409 (0.18)	0.393 (0.14)	0.395 (0.15)
6cs5SC	0.390 (0.15)	0.388 (0.16)	0.433 (0.16)	0.394 (0.16)	0.404 (0.16)	0.377 (0.20)	0.398 (0.16)	0.391 (0.16)
6cs6SE	0.431 (0.15)	0.398 (0.16)	0.378 (0.15)	0.439 (0.16)	0.412 (0.15)	0.384 (0.17)	0.406 (0.16)	0.407 (0.15)
All	0.405 (0.16)	0.397 (0.16)	0.389 (0.16)	0.406 (0.15)	0.410 (0.14)	0.405 (0.18)	0.396 (0.15)	0.401 (0.16)

These results show that the statistical distributions of κ are similar across speakers and sensors for this reading task: they conform to a slightly right-skewed closed-to-mesokurtic unimodal distribution. The median values of the statistical characteristics of the distribution are 0.400 for the statistical mode (standard deviation is 0.012), 0.155 for the standard deviation (standard deviation is 0.011), 0.55 for the skewness (standard deviation is 0.098), and 0.64 for the kurtosis (standard deviation is 0.36). The shape of this common distribution suggests that, under the hypothesis that articulatory trajectories are Tau-guided, articulators aim at coupling to a Tau-guide, with a Tau-coupling parameter κ of around 0.4. Still under this hypothesis, the speaker would solely have to choose the target position and the time of offset of the movement for a given articulation. The timecourse of the articulator to reach the target position at the offset time is then such that it follows the predefined Tau-guide, with a coupling constant of $\kappa \simeq 0.4$. Noise may occur in the production of the articulatory movement: the trajectory timecourse may not follow an ideal Tau-guide in these cases. Interestingly, according to Eq. (13), Tau-guides with a κ value of 0.4 presents a symmetrical velocity profile, namely with a peak velocity located exactly at half the duration of the movement. Further tests of this hypothesis will be required to see if the distribution of κ values is similar for other types of speech, including other speech styles or materials with other lexical content.

6. Articulatory effort of Tau-guided movements

The Tau-analysis in the previous section suggests that articulatory movements conform to the Tau equation, and suggests that most movements are coupled to a Tau-guide with a similar coupling constant κ. From these observations, we hypothesize that speakers most often choose a default κ value, which corresponds to the statistical mode of the κ distributions extracted from our experiments, namely $\kappa \simeq 0.4$. This section is an attempt to find a possible explanation for why speakers would plan Tau-guided articulations with this particular coupling constant.

Our intuitive solution to this problem is that Tau-guided movements planned with this particular κ value optimize an unknown performance objective. Indeed, besides satisfying primarily task objectives, skilled movements have been shown to satisfy energetic constraints, *i.e.* to satisfy an "economy of effort" constraint (Hoyt and Taylor, 1981; Nelson, 1983). Following this idea, speech articulatory movements should also be executed such that an effort-based cost function is optimized. This has been exploited, for instance, by Lindblom in his Hyper and Hypo continuum theory (also known as *H&H theory*) (Lindblom, 1990) and in *Emergent Phonology* (Lindblom, 1999). Articulatory effort is one of the constraints included in the cost function used in *Embodied Task Dynamics* model of speech production by Šimko and Cummins (2010). This section discusses the observed articulatory movements from the perspective of both general Tau theory and Optimal Control Theory (OCT). Note that this is a preliminary study. We do not claim to fully investigate the complex question of performance objectives applied to speech. Instead, this section attempts to provide a possible interpretation of our experimental observations about κ values from Section 5.2.

6.1. *Performance objectives*

Two dominant performance objectives are commonly used in OCT, namely minimum jerk (Flash and Hogan, 1985; Hoff and Arbib, 1993;

Sha et al., 2006), and minimum motor commands (Fagg et al., 2002; O'Sullivan et al., 2009; Shadmehr et al., 2010). Theories based on minimum jerk assume that movements reflect a minimized cost function J based on jerk (the derivative of acceleration):

$$J_{\text{jerk}} = \int_0^T |\dddot{x}(t)|^2 dt, \tag{18}$$

where $\dddot{x}(t)$ is the third derivative (jerk) of position with respect to time, and T is movement duration.

Theories based on minimum motor commands assume that motor commands are penalized, instead of jerk, hence

$$J_{\text{command}} = \int_0^T |u(t)|^2 dt, \tag{19}$$

where $u(t)$ is the motor command signal. In this preliminary study, we assume that the motor commands are defined as the resulting forces acting on the articulator, as proposed in Nelson (1983). This assumption is motivated by the fact that it has been successfully applied in previous literature on OCT-based speech production (Šimko and Cummins, 2010). Following Newton's second law of motion, the force acting on the articulator is $F(t) = m\ddot{x}(t)$, where m is the mass of the articulator. We will consider an arbitrary mass $m = 1$ for this study, hence the following cost function

$$J_{\text{force}} = \int_0^T |\ddot{x}(t)|^2 dt. \tag{20}$$

6.2. Theoretical minimal effort

This section investigates the influence of κ values on the effort required by Tau-guided movements and compares this with a theoretical minimal effort. In Flash and Hogan (1985), the authors derived the formula governing the linear 1D-movement which minimizes jerk, namely the solution $x(t)$ of Eq. (18). Assuming that both the velocity and acceleration are null at both the onset and offset of the movements ($t = 0$, and $t = T$, respectively), that the onset position is $x(0) = X_0$ and the offset position is $x(T) = 0$, one can find that the solution to Eq. (18) is

$$x_{\text{min}_{\text{jerk}}}(t) = X_0 \left[1 - 10\left(\frac{t}{T}\right)^3 + 15\left(\frac{t}{T}\right)^4 - 6\left(\frac{t}{T}\right)^5\right] \tag{21}$$

However, as shown in Eq (12), the acceleration of Tau-guided movements is not null at both end points. Consequently it is not appropriate to compare Tau-guide movements with minimal jerk movements since they do not have the same boundary conditions. However, following the same method as in Flash and Hogan (1985), without the need to impose boundary conditions for acceleration, one can find that the solution to Eq. (20) is

$$x_{\text{min}_{\text{force}}}(t) = X_0 \left[1 - 3\left(\frac{t}{T}\right)^2 + 2\left(\frac{t}{T}\right)^3\right] \tag{22}$$

It follows that the optimal acceleration is

$$\ddot{x}_{\text{min}_{\text{force}}}(t) = 6\frac{X_0}{T^2}\left(\frac{2t}{T} - 1\right). \tag{23}$$

Substituting Eq. (23) into Eq. (20) yields:

$$J_{\text{min}_{\text{force}}} = 12\frac{X_0^2}{T^3}. \tag{24}$$

For the purpose of comparing the effort required by Tau-guided movements with the theoretical minimal effort defined by Eq. (24), we introduce the effort ratio R_{force}, defined as

$$R_{\text{force}} = J_{\text{force}} / J_{\text{min}_{\text{force}}}, \tag{25}$$

where J_{force} and $J_{\text{min}_{\text{force}}}$ are defined as in Eqs. (20) and (24).

Fig. 10 displays the effort ratio R_{force} as defined in Eq. (25). It shows that R_{force} admits a minimum for $\kappa_{\text{min}} = 0.454$, where $R_{\text{force}}(0.454) = 1.034$. Interestingly, the effort ratio in the region where most of the mode of the statistical distribution of the κ estimated in our experiments (i.e. $\kappa \in [0.38, 0.42]$) is relatively close to 1. It is 1.10 and 1.04 for $\kappa = 0.38$ and $\kappa = 0.42$, respectively. This observation gives support to the hypothesis that, in addition to planning acoustic features for conveying information appropriate for each context, speakers aim at minimizing the articulatory effort required for their speech movements. In this case, it seems that they minimize the force acting on the articulators. Indeed, the modes of the probability density function of κ values in the analyzed data correspond to a κ-value of the Tau-guided movement for which the required effort is close to the minimal one.

6.3. Relationship between estimated κ and required effort

This section explores the possibility that speakers choose Tau-guided movements that are the least effortful. In particular, we investigate whether the effort of each movement unit impacts the distribution of κ. The statistical analysis of the κ distribution presented in Section 5.2 did not consider the effort, i.e. the resulting forces acting on the articulator, of each individual movement unit. Although Fig. 10 shows that the required effort of Tau-guide movements having a κ value equal to the statistical modes of the observed κ distribution is close to the minimal effort, the spread of the distribution shows that effort optimization is not always implemented. Our expectation is that the greater the effort, the more likely speakers will choose a κ value close to 0.4, namely the effort ratio R_{force} is close to 1.

Fig. 11 shows the estimated κ values as a function of the effort required to make the observed movement by the articulator, both theoretically (namely the theoretical minimal effort required) and observed (namely the observed effort). Effort has been normalized to allow for comparisons. The normalization has been performed by dividing the measured effort by the maximal effort, defined as the maximal measured effort among all movement units. Fig. 11 also shows the range of κ for which $R_{\text{force}} \in [1, 1.5]$. It shows that the values of κ span a wide range for movements that require low effort, while they tend to converge toward the optimal value for movements that require greater effort. That is, κ exhibits a smaller variance as the required effort increases. This is in agreement with our expectation for a smaller variance of κ for movements which require more effort, as well as convergence towards the optimal κ value. This observation supports our hypothesis that, for a given target position and duration of movement, the timecourse of articulatory movement is planned to minimize the force acting on the articulator.

7. Conclusion and discussion

This paper has presented an attempt to apply general Tau theory to speech, and more specifically to articulatory movements. This theory has been successfully applied in the past to other bodily movements (Lee and Reddish, 1981; Lee et al., 1983; Craig and Lee, 1999; Schögler et al., 2008; Rodger et al., 2013), hence is a good candidate to model the production of articulatory movements in speech. It proposes that the timecourse of articulators is controlled so that it is proportional to a guide function. This ensures that the position target is reached at the desired time.

Using EMA data from the DoubleTalk corpus (Scobbie et al., 2013; Geng et al., 2013), the paper has assessed the relevance of general Tau theory for articulatory movements in speech. The assessment has been done by comparing the accuracy of the Tau model to fit observed trajectories (for single articulators as well as lip aperture), with the accuracy of other widely-used parametric trajectory models. The models used for comparisons are the Critically Damped Oscillator (CDO) models (Saltzman, 1986), either with gradual activation functions (Kröger et al., 1995) or a non-linear restoring force (Sorensen and Gafos, 2016) (respectively G-CDO and NL-S-CDO), and the Sequential Target Approximation Model (STAM) (Birkholz et al., 2010). Our experiments show that the best fits are systematically obtained using general Tau

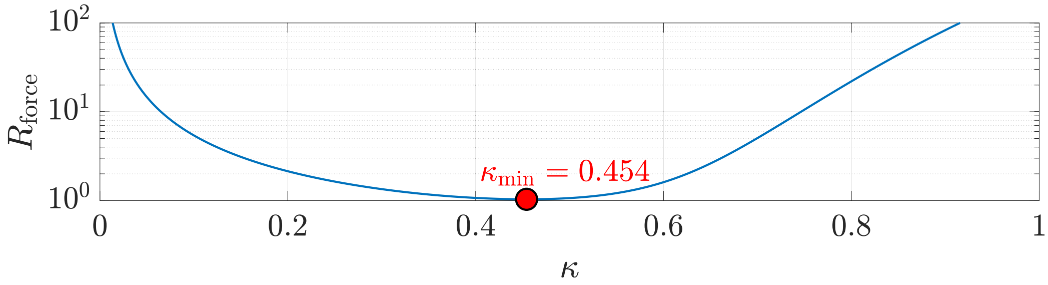

Fig. 10. Effort ratio R_{force} as a function of κ. The y-axis scale is logarithmic. The κ-value which correspond to local minima of the effort ratio is represented as a circle.

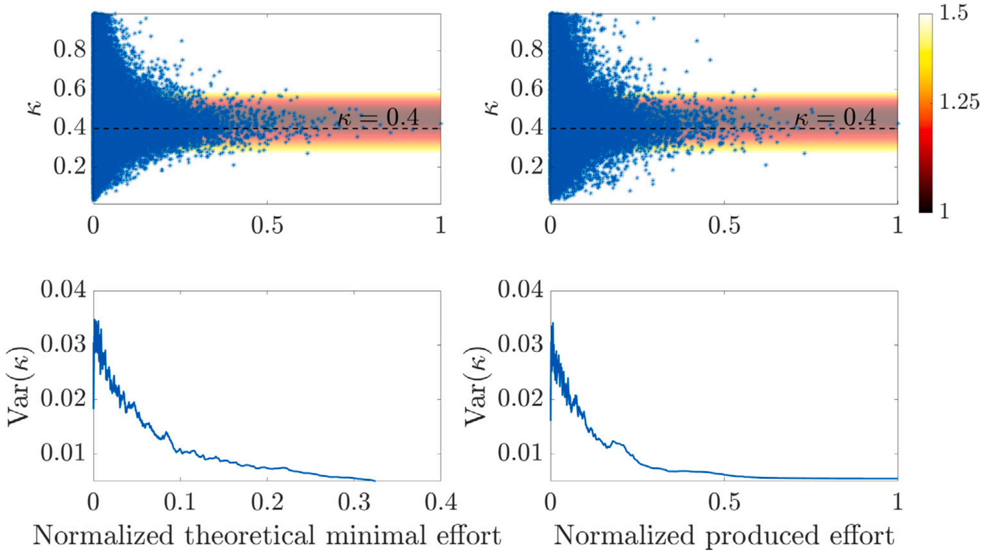

Fig. 11. The distribution of κ as a function of the effort required by the movement. The top plot shows the observed distribution as well as $R_{\mathrm{force}}(\kappa) \in [11.5]$ as a colormap; the bottom plot shows the variance of κ as the function of effort. The left plot shows the theoretical minimal effort required, normalized by the maximal effort. The right plot shows the effort estimated from observed data, normalized by the maximal effort. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

theory, independently of the analyzed articulator and/or speaker. These results suggest that general Tau theory may be a useful tool to analyze and generate speech articulatory movements. In addition, general Tau theory presents other practical and theoretical advantages when used for articulatory analysis. They include (i) a unique parameter to optimize, namely the Tau-coupling parameter κ, leading to a very fast computation time, as onset and offset times and movement distances can be directly estimated from observed data, and (ii) a unique solution which does not depend on the initial estimate. General Tau theory could also be used efficiently to generate articulatory trajectories for articulatory synthesis. Indeed, similarly to trajectory analysis, general Tau theory presents several advantages for generating trajectories. They include (i) direct and straightforward control of the skewness of the velocity profile by simply adjusting a unique parameter, κ, for a given amplitude and duration, as shown by Eqs. (13) to (15), and (ii) the possibility to control the timecourse of movement to achieve a target position on time, consistent with observations of less timing variability at movement endpoints. However, although this paper has shown that Tau-guided movements fit the projection of two-dimensional positions onto the first principal component very precisely, one would still need to implement a model to generate the actual two-dimensional position signal. One possible approach would be to use an intermediate articulatory model that describes the geometry of the vocal tract using a few articulator parameters, as proposed in TADA (Nam et al., 2004), in ETD (Šimko and Cummins, 2010), in VocalTractLab (Prom-on et al., 2013; Xu et al., 2019), or using Task-Dynamics (Alexander et al., 2019). This requires further investigation about how well general Tau theory applies to the timecourse of such articulator parameters as inferred from real speech.

The second part of the paper presented an analysis of articulatory movements from the same DoubleTalk EMA corpus using general Tau theory. This involved analyzing the distribution of the shape parameter κ which gave the best fit to analyzed movements. Our experiments show that the statistical distributions of κ are very similar, for all the speakers and all articulators. The typical distribution is a right-skewed unimodal distribution, having a peak at $\kappa = 0.4$ (± 0.01), a standard deviation of 0.155, with an excess kurtosis around 0.55. Under the hypothesis that articulatory movements are Tau-guided, these observations suggest that there is a coupling constant which is systematically favored for planning speech articulations. This target coupling constant is then the one which corresponds to the statistical mode of the κ-values extracted from our experiments, namely $\kappa \simeq 0.4$. This value corresponds to Tau-guides which exhibit a perfectly symmetrical velocity profile, namely for which the time-to-peak velocity ratio is 0.5. This preliminary study raises the question of the reason of this invariance of statistical distribution. One possible explanation, which has not been explored in this paper, would be that it is due to the unique type of analyzed speech data, namely read speech. In the future, it would be interesting to investigate the influence of the speech task on the statistical distribution of κ.

Finally, in the third part of the paper, we investigated our hypothesis that if such a target κ value exists, it is because resulting trajectories reflect the optimization of an objective function related to articulatory effort. Interestingly, our preliminary study shows that the peak region of the κ-distribution is close to the value of κ for which Tau-guided movements minimize the forces acting on the articulator for a given duration and amplitude of movement. In addition, our experiments also show that the variance of the estimated κ is reduced when the required effort increases. These results provide new evidence for a balance in articulatory gesture production between articulatory accuracy and minimal effort, as hypothesized in previous studies (Lindblom, 1990, 1999; Perkell and Zandipour, 2002; Šimko and Cummins, 2010).

At this stage, this paper simply proposes a new model for speech processing and does not claim to answer numerous questions related to speech articulatory planning. We would like this paper to constitute a basis for further investigations on speech production planning and control. It would be interesting to investigate several hypotheses that can emerge from our experiments, and which are beyond the scope of this paper. For instance, assuming the movements of the articulators are Tau-guided, the extent to which the speaker can control the value of the κ parameter is an important question. It is our hope that this paper will stimulate other researchers to explore issues like these.

CRediT authorship contribution statement

Benjamin Elie: Conceptualization, Methodology, Software, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization . **David N. Lee:** Conceptualization, Methodology, Writing – review & editing. **Alice Turk:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

Data availability

The source code for Tau analysis, as well as STAM and CDO fitting methods are freely available in a GitLab repository (Elie, 2022) (https://git.ecdf.ed.ac.uk/belie/tauspeech). The EMA data used in this paper are available at (https://datashare.ed.ac.uk/handle/10283/4495). Additional data are available on request from authors.

Acknowledgments

We gratefully acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (PlanArt: Planning the Articulation of Spoken Utterances; ERC Advanced Grant awarded to A. Turk; Grant agreement No. 101019847).

References

Alexander, R., Sorensen, T., Toutios, A., Narayanan, S., 2019. A modular architecture for articulatory synthesis from gestural specification. J. Acoust. Soc. Am. 146 (6), 4458–4471.

Bernstein, N., 1966. The co-ordination and regulation of movements. Pergamon Press, Oxford.

Birkholz, P., 2007. Control of an articulatory speech synthesizer based on dynamic approximation of spatial articulatory targets. In: Eighth Annual Conference of the International Speech Communication Association.

Birkholz, P., Hoole, P., 2012. Intrinsic velocity differences of lip and jaw movements: preliminary results. In: Thirteenth Annual Conference of the International Speech Communication Association.

Birkholz, P., Kröger, B.J., Neuschaefer-Rube, C., 2010. Model-based reproduction of articulatory trajectories for consonant–vowel sequences. IEEE Trans. Audio, Speech, Lang. Process. 19 (5), 1422–1433.

Birkholz, P., Martin, L., Xu, Y., Scherbaum, S., Neuschaefer-Rube, C., 2017. Manipulation of the prosodic features of vocal tract length, nasality and articulatory precision using articulatory synthesis. Comput. Speech Lang. 41, 116–127.

Blackburn, C.S., Young, S., 2000. A self-learning predictive model of articulator movements during speech production. J. Acoust. Soc. Am. 107 (3), 1659–1670.

Browman, C.P., Goldstein, L.M., 1986. Towards an articulatory phonology. Phonology 3, 219–252.

Browman, C.P., Goldstein, L., 1995. Dynamics and articulatory phonology. In: Mind as motion: Explorations in the dynamics of cognition. MIT press Cambridge, MA, Cambridge, pp. 175–194.

Byrd, D., Saltzman, E., 1998. Intragestural dynamics of multiple prosodic boundaries. J. Phonetics 26 (2), 173–199.

Byrd, D., Saltzman, E., 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. J. Phonetics 31 (2), 149–180.

Craig, C.M., Lee, D.N., 1999. Neonatal control of nutritive sucking pressure: evidence for an intrinsic τ-guide. Exp. Brain Res. 124, 371–382.

Elie, B., 2022. TauSpeech. https://git.ecdf.ed.ac.uk/belie/tauspeech.

Fagg, A.H., Shah, A., Barto, A.G., 2002. A computational model of muscle recruitment for wrist movements. J. Neurophysiol. 88 (6), 3348–3358.

Flash, T., Hogan, N., 1985. The coordination of arm movements: an experimentally confirmed mathematical model. J. Neurosci. 5 (7), 1688–1703.

Geng, C., Turk, A., Scobbie, J.M., Macmartin, C., Hoole, P., Richmond, K., Wrench, A., Pouplier, M., Bard, E.G., Campbell, Z., et al., 2013. Recording speech articulation in dialogue: Evaluating a synchronized double electromagnetic articulography setup. J. Phonetics 41 (6), 421–431.

Gentner, D.R., Grudin, J., Conway, E., 1980. Finger movements in transcription typing. tech. rep., California University San Diego La Jolla Center for Human Information Processing.

Gibson, J.J., 1966. Senses Considered as Perceptual Systems. Houghton Mifflin, Boston.

Henke, W.L., 1966. Dynamic articulatory model of speech production using computer simulation. tech. rep., Massachussetts Institue of Techonology, Cambridge.

Hoff, B., Arbib, M.A., 1993. Models of trajectory formation and temporal interaction of reach and grasp. J. Motor Behav. 25 (3), 175–192.

Honorof, D.N., McCullough, J., Somerville, B., 2000. Comma gets a cure. Diagn. Passage.

Hoyt, D.F., Taylor, C.R., 1981. Gait and the energetics of locomotion in horses. Nature 292, 239–240.

Keating, P.A., 1990. The window model of coarticulation: articulatory evidence. In: Papers in Laboratory Phonology. Cambridge University Press, Cambridge, pp. 451–470.

Kröger, B.J., Schröder, G., Opgen-Rhein, C., 1995. A gesture-based dynamic model describing articulatory movement data. J. Acoust. Soc. Am. 98 (4), 1878–1889.

Lee, D.N., 1998. Guiding movement by coupling taus. Ecol. Psychol. 10 (3–4), 221–250.

Lee, D.N., Reddish, P.E., 1981. Plummeting gannets: A paradigm of ecological optics. Nature 293, 293–294.

Lee, D., Young, D., Reddish, P., Lough, S., Clayton, T., 1983. Visual timing in hitting an accelerating ball. Q. J. Exp. Psychol. 35 (2), 333–346.

Lindblom, B., 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Speech Production and Speech Modelling. Springer, pp. 403–439.

Lindblom, B., 1999. Emergent phonology. In: Annual Meeting of the Berkeley Linguistics Society. 25, pp. 195–209.

Ling, Z.-H., Richmond, K., Yamagishi, J., 2010. An analysis of HMM-based prediction of articulatory movements. Speech Commun. 52 (10), 834–846.

Munhall, K.G., Ostry, D.J., Parush, A., 1985. Characteristics of velocity profiles of speech movements. J. Exp. Psychol.: Hum. Percept. Perform. 11 (4), 457–474.

Nam, H., Goldstein, L., Saltzman, E., Byrd, D., 2004. TADA: An enhanced, portable Task Dynamics model in MATLAB. J. Acoust. Soc. Am. 115 (5), 2430.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7 (4), 308–313.

Nelson, W.L., 1983. Physical principles for economies of skilled movements. Biol. Cybernet. 46, 135–147.

Okadome, T., Honda, M., 2001. Generation of articulatory movements by using a kinematic triphone model. J. Acoust. Soc. Am. 110 (1), 453–463.

Ostry, D.J., Cooke, J.D., Munhall, K.G., 1987. Velocity curves of human arm and speech movements. Exp. Brain Res. 68, 37–46.

Ostry, D.J., Munhall, K.G., 1985. Control of rate and duration of speech movements. J. Acoust. Soc. Am. 77 (2), 640–648.

O'Sullivan, I., Burdet, E., Diedrichsen, J., 2009. Dissociating variability and effort as determinants of coordination. PLoS Comput. Biol. 5 (4), e1000345.

Perkell, J.S., Matthies, M.L., 1992. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability. J. Acoust. Soc. Am. 91 (5), 2911–2925.

Perkell, J.S., Zandipour, M., 2002. Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements. J. Acoust. Soc. Am. 112 (4), 1642–1651.

Prom-on, S., Birkholz, P., Xu, Y., 2013. Training an articulatory synthesizer with continuous acoustic data.. In: INTERSPEECH. pp. 349–353.

Ribeiro, V., Isaieva, K., Leclere, J., Vuissoz, P.-A., Laprie, Y., 2022. Automatic generation of the complete vocal tract shape from the sequence of phonemes to be articulated. Speech Commun. 141, 1–13.

Rodger, M.W., O'Modhrain, S., Craig, C.M., 2013. Temporal guidance of musicians' performance movement is an acquired skill. Exp. Brain Res. 226, 221–230.

Saltzman, E., 1986. Task dynamic coordination of the speech articulators: A preliminary model. In: Generation and Modulation of Action Patterns (Experimental Brain Research Series). 15, Springer Berlin, Heidelberg, Berlin, pp. 129–144.

Saltzman, E.L., Munhall, K.G., 1989. A dynamical approach to gestural patterning in speech production. Ecol. Psychol. 1 (4), 333–382.

Schögler, B., Pepping, G.-J., Lee, D.N., 2008. TauG-guidance of transients in expressive musical performance. Exp. Brain Res. 189, 361–372.

Scobbie, J.M., Turk, A., Geng, C., King, S., Lickley, R., Richmond, K., 2013. The Edinburgh speech production facility DoubleTalk corpus. In: INTERSPEECH 2013: Proceedings of the 14th Annual Conference of the International Speech Communication Association (ISCA). International Speech Communication Association, pp. 764–766.

Sha, D., Patton, J.L., Mussa-Ivaldi, F.A., 2006. Minimum jerk reaching movements of human arm with mechanical constraints at endpoint.. Int. J. Comput. Syst. Signals 7 (1), 41–50.

Shadmehr, R., De Xivry, J.J.O., Xu-Wilson, M., Shih, T.-Y., 2010. Temporal discounting of reward and the cost of time in motor control. J. Neurosci. 30 (31), 10507–10516.

Sorensen, T., Gafos, A., 2016. The gesture as an autonomous nonlinear dynamical system. Ecol. Psychol. 28 (4), 188–215.

Sorensen, T., Toutios, A., Goldstein, L., Narayanan, S., 2019. Task-dependence of articulator synergies. J. Acoust. Soc. Am. 145 (3), 1504–1520.

Spencer, R.M., Zelaznik, H.N., 2003. Weber (slope) analyses of timing variability in tapping and drawing tasks. J. Motor Behav. 35 (4), 371–381.

Toutios, A., Ouni, S., Laprie, Y., 2011. Estimating the control parameters of an articulatory model from electromagnetic articulograph data. J. Acoust. Soc. Am. 129 (5), 3245–3257.

Turk, A., Shattuck-Hufnagel, S., 2020a. Speech timing: Implications for theories of phonology, speech production, and speech motor control. Oxford University Press, USA, pp. 238–263.

Turk, A., Shattuck-Hufnagel, S., 2020b. Timing evidence for symbolic phonological representations and phonology-extrinsic timing in speech production. Front. Psychol. 10:2952.

Šimko, J., Cummins, F., 2010. Embodied task dynamics. Psychol. Rev. 117 (4), 1229–1246.

Xu, Y., 2004. Transmitting tone and intonation simultaneously - The parallel encoding and target approximation (PENTA) model. In: International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages. pp. 215–220.

Xu, A., Birkholz, P., Xu, Y., 2019. Coarticulation as synchronized dimension-specific sequential target approximation: An articulatory synthesis simulation. In: Proceedings of the 19th International Congress of Phonetic Sciences.