Fundamental motor laws and dynamics of speech

A cumulative dissertation by

Stephan R. Kuberski

submitted for the degree of

Doctor of Philosophy

University of Potsdam
Faculty of Human Sciences

November 2019

Advisor: Prof. Dr. Admantios I. Gafos, University of Potsdam

Reviewers: PD Dr. Doris Mücke, University of Cologne
Prof. Dr. Adamantios I. Gafos, University of Potsdam

Defense: September 20, 2019

This work was funded with support from the Deutsche Forschungsgemeinschaft (Grant number 317633480, Sonderforschungsbereich 1287, Project C04) and the European Research Council (Grant number AdG 249440).

Published online at the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-43771
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-437714

Contents

Introduction

Speech is a highly complex motor task. It requires the precise positioning of a number of articulatory organs to perform specific gestures in a limited space inside the body. The vast amount of different conditions affecting individual motor performance makes speech a highly variable motor task too. For example, age, gender, loudness and tempo all contribute to the precise formation of gestures and their corresponding speech sounds. Speech has evolved to harness this complexity and variability for the purpose of communication. A remarkable fact about the resulting system is its robustness. What appears from the linguist's perspective as the same utterance can be conveyed under a wide range of conditions. Given such rampant variability, the identification of invariances characterizing properties of speech gestures has been seen as an important research goal. This is because it is hoped that the identification of invariances reduces the set of candidate theories about the nature of speech primitives and their organization.

In search for such invariances, a long line of work has been proceeded in documenting basic kinematic characteristics of speech. The basic kinematic parameters of speech are those of speed, amplitude and duration of gestures. A number of relations between these parameters are common knowledge today (e.g., Nelson, 1983; Ostry and Munhall, 1985; Vatikiotis-Bateson and Kelso, 1990). However, experimental constraints and other limitations specific to speech hindered progress towards the identification of further, non-kinematic relations along the lines pursued in the field of general motor control (see Fuchs and Perrier, 2005). The current work introduces a speech elicitation paradigm which as will be shown evades some of these speech-specific constraints and presents assessments of two of the perhaps most fundamental motor laws of (general) human motor behavior in the domain of speech. By means of an extensive and systematic speech rate manipulation, it is demonstrated that speech movements obey the same two fundamental laws just as movements from other domains of motor control do. Specifically, these are the power law relation between speed and curvature of movement and Fitts' law, the latter also referred to as a trade-off relation between speed and accuracy of target-directed movements. Beyond the assessment of these laws in speech, results indicate that both laws offer the potential to unveil the presence of qualitatively distinct spatio-temporal organizations of what appear, on the surface, to be the same sequences of phonemes. This is an entirely new result. The methods and datasets obtained in the two assessment above provide a rigorous foundation both for addressing implications for theories and models of speech as well as for better understanding the status of speech movements in the context of human movements in general.

According to all modern theories of language, the phonological message of an utterance is represented as a sequence of segments or phonemes which can in principle be mapped to spatio-temporal vocal tract action in the form of speech gestures (Chomsky and Halle, 1968; Browman and Goldstein, 1986). One of the major stumbling blocks in being able to systematically relate the observables of speech to phonological representations is the apparent incompatibility in the nature of the entities involved at the two different levels of description, phonemes and speech movements. Phonemes are discrete, qualitative, and context-invariant, while speech is continuous (in space and time) and highly context-dependent. Advances in understanding of the coordination and control of action, beginning with Bernstein (1967) and Turvey (1977), have provided a principled way of unifying these descriptions. Pioneering work on this frontier is best expressed in the writings of Fowler (1980) and Fowler, Rubin, et al. (1980), and was made more concrete in the development of Articulatory Phonology (Browman and Goldstein, 1986) and the Task Dynamics model (Saltzman, 1986; Saltzman and Munhall, 1989; Kelso, Saltzman, et al., 1986). It is now common to assume that the spatio-temporal organization of speech consists in a successive concatenation of target-directed movements, each with a target defined within some hypothesized state space characterizing the current phonemic (sub-)task (e.g., for /kat/, start from some vocal tract state characterizing the target of the phoneme /k/, move to another state for the vowel /a/ and finally move to yet another state for the consonant /t/). Contemporary dynamical models of speech differ with respect to assumptions about the required state space components (e.g., articulatory-only: Saltzman and Munhall, 1989; Perrier, Ostry, et al., 1996, acoustic-articulatory: Guenther, 1995) but largely agree on the hypothesis that speech targets in such space correspond to point attractors in an underlying dynamical system. In other words, the common conjecture about the principles of speech is that utterances conform to the concatenation of point attractor dynamics in a well-defined space of task states. Yet the existence of attractors in speech has not heretofore been rigorously assessed or verified by analysis of the phase space.

In its third component, thus, the present work examines the articulatory state space of speech for signatures of attractors and their corresponding topologies. The results indicate, by and large, conformity with the assumption that speech is regulated by (point) attractor dynamics. However, for the first time here, it is demonstrated that the mapping between the phonemic units of an utterance and the attractor topology is non-unique. Utterances of the same phonemic type are observed to be mapped to qualitatively distinct dynamical organizations with speech rate as the decisive factor. Moreover, as will be made clear in the forthcoming, this novel result from the third component of the presented work is in broad agreement, with the specific form of the results from the two preceding motor law assessments.

What follows now is a concise description of the principal experimental paradigm used by the three assessments in this work. As anticipated already above, the observation and study of speech action in the vocal tract meets certain limitations regarding the systematic control of the experimental conditions. For example, whereas in kinematic studies of limb movements the experimenter can control extrinsically defined quantities such as the extent of movement or the target size in a task where the participant is asked to look at or reach to a certain target, the same does not hold for speech. The classic studies by Paul Fitts which gave rise to an oceanic literature of what came to be called Fitts' law, had participants use a stylus to tap alternatively on two discs of a specified diameter (accuracy) placed at a certain distance from one another (extent of movement). However, in speech, it is infeasible to exactly produce speech movements of a specific (smaller or larger) accuracy or of a specific spatial extent.

The experimenter may of course instruct speakers to do so (via some encouragement to "speak accurately" or some such other instruction), but there is no guarantee that the scaling of whatever the speakers may be changing during their vocal tract actions (if they are scaling anything at all) corresponds to any parameter that is essential to an assessment of the basic laws in question. Determination of invariances in behavior crucially requires such systematic control of some parameters. Yet there are ways to gain implicit control over some of these parameters by use of another intermediary one. The proposed paradigm in this work utilizes extensive speech rate manipulations in order to gain indirect control over kinematic parameters (amplitude, speed, duration of movements) and non-kinematic properties crucially required for the assessments of the fundamental motor laws, that is, curvature of movement trajectory for the power law between speed and curvature and articulatory target size for Fitts' law.

Speech rate was controlled by means of an audible metronome with which participants were asked to utter repetitive sequences of elementary consonant-vowel syllables. The tempo of the metronome was systematically varied between extremely slow and extremely fast rates. For data registration, the latest generation electromagnetic articulography (EMA) was used, which allowed for three-dimensional tracking of articulatory movements at high spatial and temporal resolution (±0.3 mm at 1250 Hz). The method of EMA is based on the principle of electromagnetic induction. In preparation for data collection, small receiver coils are glued at specific locations on a participant's articulatory organs (e.g., tongue tip and tongue back as illustrated in the right-hand side of Figure A). During the experiment, a number of transmitter coils surrounding the participant's head are used to generate alternating magnetic fields of different frequencies. As a receiver coil moves through such field (during speech), an electric current is induced proportional to the cube of the distance from its transmitter. Hence, the observed composite current in any receiver encodes information about the distances from all installed transmitters and thus allows for their precise localization through subsequent triangulation. As a result, EMA provides an effective and pain-free method to register spatial and temporal properties of vocal tract action under fairly natural conditions. The left-hand side of Figure A gives a prototypical example of post-processed data from the proposed experimental paradigm. The figure shows

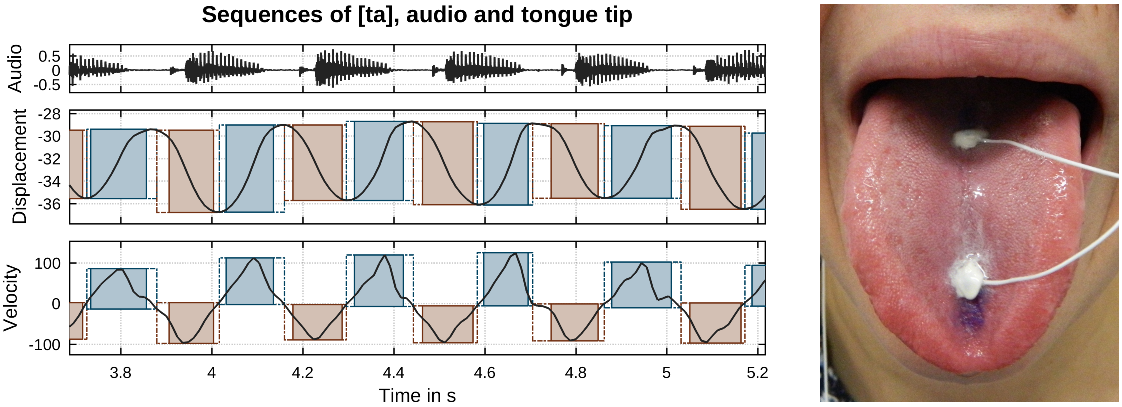

Figure A. Left: Time series of an example of repeated [ta] syllables at a metronome rate of 210 bpm. Top row shows the acoustic recording. Middle row shows the vertical displacement of the tongue tip articulator. Bottom row shows the articulator's velocity. Right: Picture of a participant's tongue with two receiver coils glued (tongue back and tip).

displacement and velocity time series of a receiver coil glued on a participant's tongue tip while producing repetitions of [ta] syllables (for reasons of presentation the figure shows only the vertical component of the three spatial dimensions). The phonologically relevant segments of the articulator's continuous motion are indicated by colored boxes. For the shown example, these are the closing (for the consonant [t]) and opening (for the vowel [a]) gestures highlighted by blue and red color respectively.

In processing and analysis of such data, crucial issues arise. One of them concerns the dimensionality of the data. Much work on (general) movement science and speech production limits evaluation of the fundamental laws to two dimensions. However, speech naturally happens in three dimensions. In what follows, evaluation of the fundamental laws takes place using all three dimensions. It will be explicitly argued that this in fact improves the conformity of the speech data to the laws. This will be shown both by explicit comparison of the results with results from previous work which limits data to two dimensions and also by explicit comparison of the results from our data when reduced to its two dimensions. Another crucial issue of processing data as collected for this work regards their numerical treatment. Electromagnetic articulography provides displacement time series of selected articulators at high temporal resolution (1250 Hz sampling rate). However, there is no direct experimental access to the velocities or high order derivatives (e.g., acceleration) of these. Accordingly, any access to these important properties of kinematics requires the application of numerical differentiation. Since numerical differentiation is highly sensitive to noise (the higher the order of the derivative, the more so), a robust and numerically stable algorithm is essential. With these strictures in mind, the present work utilized nine-point stencil differentiators for high precision numerical differentiation as well as single uniform quintic spline objects for a consistent analytical representation of displacement, velocity and acceleration time series at the same time.

In sum, the here proposed paradigm of repetitive speech with systematic speech rate control and advanced data representation will be shown to provide a reliable basis for an assessment of some of the so far difficult to access aspects of speech. Analysis of the data in their full three-dimensions adheres to the original space in which speech naturally happens. By its resulting variability of the crucial parameters of speech production, the extensive speech rate manipulation allows for a rigorous demonstration that speech movements obey the same fundamental motor laws just as movements from other domains do. In addition, the systematic rate control will reveal the existence of qualitatively distinct dynamical organizations in speech with rate as the major parameter of discrimination.

The remainder of this work consists in three submitted (or already published) manuscripts, in international peer-reviewed venues of the field of speech science, representing original research articles on the topics of fundamental motor laws in speech movement data registered using the paradigm above and the corresponding dynamics of these speech movements. The first article has been published in *PLOS One*, the second is being evaluated at *Phonetica* with its current status being under a second phase of evaluation requesting "minor revisions" and the third is under evaluation in *Language*:

- Kuberski, Stephan R. and Adamantios I. Gafos (2019). "The speed-curvature power law in tongue movements of repetitive speech". *PLOS One* 14(3). Public Library of Science. doi: 10.1371/journal.pone.0213851.
- Kuberski, Stephan R. and Adamantios I. Gafos (under review). "Fitts' law in tongue movements of repetitive speech". *Phonetica: International Journal of Phonetic Science*. Karger Publishers.

- Kuberski, Stephan R. and Adamantios I. Gafos (submitted). "Distinct phase space topologies of identical phonemic sequences". *Language*. Linguistic Society of America.

Each of these articles was coauthored by Adamantios I. Gafos, the doctoral advisor of the author of the present work. Beyond his gratefully acknowledged contributions in multiple, long discussions on all aspects of this work and in pointing out improvements in presentation and writing, Adamantios I. Gafos provided essential support to the author by the means of technical equipment (Carstens AG501 3D Electromagnetic Articulograph) and funding (Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 1287, project C04, grant number 317633480 and European Research Council, Advanced Grant 249440).

Research Article I

The speed-curvature power law in tongue movements of repetitive speech

PLOS One 14(3), Public Library of Science, doi: 10.1371/journal.pone.0213851
Submitted: October 15, 2018; *Accepted:* March 1, 2019; *Published:* March 18, 2019

Abstract

The speed-curvature power law is a celebrated law of motor control expressing a relation between the kinematic property of speed and the geometric property of curvature. We aimed to assess whether speech movements obey this law just as movements from other domains do. We describe a metronome-driven speech elicitation paradigm designed to cover a wide range of speeds. We recorded via electromagnetic articulometry speech movements in sequences of the form /CV.../ from nine speakers (five German, four English) speaking at eight distinct rates. First, we demonstrate that the paradigm of metronome-driven manipulations results in speech movement data consistent with earlier reports on the kinematics of speech production. Second, analysis of our data in their full three-dimensions and using advanced numerical differentiation methods offers stronger evidence for the law than that reported in previous studies devoted to its assessment. Finally, we demonstrate the presence of a clear rate dependency of the power law's parameters. The robustness of the speed-curvature relation in our datasets lends further support to the hypothesis that the power law is a general feature of human movement. We place our results in the context of other work in movement control and consider implications for models of speech production.

1. Introduction

Speech is perhaps "the most highly developed motor skill possessed by all of us" (Kelso, Tuller, et al., 1983). The continuous deformations of the vocal tract structuring the sound of speech involve the precise positioning of a number of articulatory organs as they form and release constrictions in a limited space inside the body. Speech has evolved to harness this complex activity for the purposes of communication. A remarkable fact about the robustness of the resulting system is that what a linguist considers to be the same utterance can be conveyed by different individuals under wildly different conditions. For example, age, gender, size, loudness, and speed all contribute to the formation of the speech sounds which are then recovered as an instance of e.g., [ta] or [ka]. Given the remarkable variability of conditions under which speech goals are achieved, the identification of invariances (at best, laws) in kinematic characteristics of speech movements has been seen as an imperative (Munhall, Ostry, et al., 1985; Turvey, 2007). The identification of such invariances offers crucial criteria for model evaluation. Any proposed model of speech production must conform to these. Furthermore, despite the relatively early influence of concepts and in some cases models from general motor control had on models of speech production (Fowler, Rubin, et al., 1980; Browman and Goldstein, 1986; Saltzman and Munhall, 1989; Guenther, 1995), our understanding of the extent to which speech movements conform to well-known laws from other areas of human movement is at its infancy (Nelson, 1983; Nelson, Perkell, et al., 1984; Ostry, Cooke, et al., 1987).

Results from speech hold potential for implications in the other direction, that is, from speech to theories of biological action in general. A recurring debate in motor control concerns the neurobiological bases of any given law. Opposing perspectives range from laws as consequences of direct cortical computations (Schwartz, 1994), to laws as consequences of coupling among neural and physical (limb) dynamics (Hatsopoulos and Warren Jr., 1996) and or viscoelastic muscle properties (Gribble and Ostry, 1996), and even sometimes to laws as mere computational artifacts (Marken and Shaffer, 2017). The human tongue is unlike the arm or the finger in that it is not supported by a skeletal system. Biomechanically, tongues, trunks of elephants and tentacles of cephalopods (Kier and Smith, 1985) are muscular hydrostats, that is, structures whose volumes remain constant across transformations in shape. Control of the hydrostat's shape is crucial to speech. Making a [ta] or a [ka] involves bending the muscular hydrostat by simultaneously contracting longitudinal and antagonistic circular, transverse or radial muscles. Overall, then, the control of limb versus tongue seems to be qualitatively different (Levine, Torcaso, et al., 2005). Knowledge of whether principles that organize and constrain the functioning of other effectors also apply to the tongue is thus likely to provide more clues on the bases of such principles. In the event that a law generalizes across tongues and limbs, this indicates that any proposed bases for that law must be general enough to encompass invariance with respect to the identity of the effectors and thus presumably the downstream specifics of their mode of operation.

The power law relation between speed of movement and curvature of effector trajectory is a celebrated law of human motor control. Consider drawing an ellipsis with a stylus or a finger. According to this law, as the stylus traces the trajectory segments of the ellipsis where curvature increases, movement speed systematically slows down. As the curvature of the traced segments decreases, speed increases. The law is called a power law because one quantity, here speed, varies as an exponential function, that is, a power of another quantity, here curvature. We aimed to assess whether speech movements obey this law. Explicit control of curvature in (speech) articulatory motion is not as feasible as in, for example, drawing tasks where the experimenter can prescribe at least the gross geometric features of the required movements such as in drawing a straight line or an ellipsis or a more complex, say, figure-eight-like pattern. In our assessment, thus, we capitalized on control of speed, the other essential quantity of the law in question. Using a metronome-driven speech elicitation paradigm, we aimed to elicit as great a range of movement speeds as possible. Specifically, building on the paradigm of repetitive speech (Kelso, Vatikiotis-Bateson, et al., 1985; Ostry, Cooke, et al., 1987; Patel, Löfqvist, et al., 1999), we recorded via three-dimensional electromagnetic articulometry (EMA), speech movements in sequences of the form /CV. . . / from nine adult speakers (five were native speakers of German and four were native of English) speaking at eight distinct rates ranging from extremely slow (30 beats per minute, bpm) to extremely fast (570 bpm). In the resulting dataset, we sought evidence of the sort that has provided support for the power law in other areas of motor control.

We show, first, that despite the extensive rate variation implemented in our experimental paradigm, the resulting data conforms to well known kinematic properties of speech. In turning to evaluate the speed-curvature power law, we take into account all three-dimensions of our registered data. Prior work on the law in non-speech domains and most work in the speech domain has so far either restricted movements in two dimensions via task-constraints, as in drawing on a plane where the relevant data are only those generated by contact between a stylus and a two-dimensional surface, or reduced what are naturally three-dimensional movements to two dimensions via method constraints by using previous generation two-dimensional electromagnetic articulometry systems. Our evaluation of the law is based on three-dimensions. Specifically, we demonstrate that using the full dimensionality of the movements and advanced numerical differentiation methods (as crucially required for the accurate estimation of curvature in 3D) substantially improves the significance of the findings in comparison to earlier reports while obviating the exclusion of data subsets with certain curvature values (as done in some of the earlier reports on the law). Furthermore, we demonstrate the presence of a clear speech rate dependency of the power law's parameters. The robustness of the law in our datasets lends further support to the hypothesis that the power law is a general feature of human movement. The modulation of the law's parameters by rate likely reflects geometric properties of effector trajectory shape. Such modulation is in agreement with other results from non-speech domains which have uncovered links between the geometry of a movement and the specific parameterization of the law as well as, under some proposals, between the geometry of a movement and the underlying control regime effecting that movement.

1.1 Background

Theoretically, the goal of moving an effector from one point in space to another can be achieved in an infinite number of ways. However, in biological action and human movement more specifically (Viviani, 2002), the motor system does not make use of the entire set of available paths in such point to point movements. Rather, the system seems to give preference to certain paths over others by exploiting invariances which effectively reduce the number of degrees of freedom. One such invariance is expressed by a relation between movement speed and trajectory curvature. When an effector traces a trajectory, the speed of the effector's movement decreases at the more curved parts of a trajectory and increases as the trajectory becomes straighter. This speed-curvature relation derives from observations on hand drawing and writing movements (Viviani and Terzuolo, 1982). In extension of this work to other settings, evidence for the law was reported also for movements of the oculomotor system evidenced in eye-tracking experiments (de'Sperati and Viviani, 1997) as well as for movements of the lower limb locomotion system (Hicheur, Vieilledent, et al., 2005); for a recent encompassing review see (Zago, Matic, et al., 2017). There is also evidence that the loss of function implicated in limb apraxia results in degradation of the relation between speed and curvature (Poizner, Mack, et al., 1990; Clark, Merians, et al., 1994; Poizner, Clark, et al., 1995; Rothi and Heilman, 1994). The law has been shown to hold not only during performance of various actions but also in perceiving those actions. When visual stimuli are designed to violate the law, perception of the portrayed motions is faulty or distorted in a direction that conforms to the law (Viviani, 2002). Furthermore, observing stimuli obeying the law elicits stronger activations in brain areas linked to visual processing, action production and action perception than stimuli violating the law (Dayan, Casile, et al., 2007).

As first proposed by Lacquaniti, Terzuolo, et al. (1983), the proposed relation between instantaneous speed of movement v and trajectory curvature κ reads

$$v = k\kappa^{-\beta} \tag{1}$$

with empirically determined exponent β and so-called velocity gain factor k. Thus, movement speed is a power function of trajectory curvature. Past work in non-speech domains suggests that, under certain assumptions about the class of shapes traced by the studied effector, the exponent β is generally close to a value of one third (hence, the law is often referred to as the one-third power law) and that k is constant within identifiable long segments of a trajectory but discontinuously changes at inter-segmental transitions (Viviani and Stucchi, 1992; Hicheur, Vieilledent, et al., 2005).

A few attempts have been made in assessing the presence of the speed-curvature power law in the domain of speech. Tasko and Westbury (2004) analyzed two-dimensional articulatory data from nine male and nine female American English speakers registered using an x-ray microbeam system. The participants were asked to read sentences at a self-selected speech rate. Movement speed and trajectory curvature were found to be related by a power law with an exponent near to the value of one third, with some indication of different articulators exhibiting variations in the precise expression of the law. Perrier and Fuchs (2008) investigated data registered using two-dimensional midsagittal electromagnetic articulometry from six speakers of different languages (French, German and Mandarin Chinese). The study's reading task contained several vowel-vowel and vowel-consonant-vowel combinations. In addition to these data, the authors analyzed data simulated by a biomechanical tongue model at three distinct speech rates (slow, normal and fast). Results showed that the power law offers a fair description of the global speed-curvature relation for all speakers and all languages as well as for the simulated data, however, with a significant inter-speaker variability of the power law's exponent β and velocity gain factor k. Neufeld and van Lieshout (2014) analyzed three-dimensional articulatory data of six native Canadian English speakers (three females, three males). Prioritizing naturalistic speech, participants were asked to read aloud from a book of short stories. The work reports that the speed-curvature power law holds with an exponent close to one third, though, with differences between articulators similar to the earlier results reported in Tasko and Westbury (2004). Most recently, Tomaschek, Arnold, et al. (2018) analyzed articulations of German words of distinct lexical frequencies. In two-dimensional data from 16 speakers (eight female, eight male), the study found support for the presence of the speed-curvature power law. However, it was also reported that for lower values of curvature the effect of curvature on speed levels off substantially. Hints of weakening of the law for low curvature values are also found in Tasko and Westbury (2004). Thus, both latter works excluded extremal values of curvature prior to the regression analyses used in assessing the law in their speech data.

In the background of this prior work, the key characteristic of our present study is the extensive rate manipulation using metronome rates from extremely slow (30 bpm) to extremely fast (570 bpm). This allowed us to assess the speed-curvature power law on the basis of data from a (substantially) wider range of speed conditions than in any prior work. It should be evident that because the law refers to speed, a thorough evaluation of its applicability in speech should aim to vary this parameter to the extent possible.

2. Methods

Five native speakers of German (three females, two males) and four native speakers of English (three females, one male) participated in the experiment. In the following, aliases G1–G5 refer to the German speakers of our study. Aliases E1–E4 refer to the English speakers. Data from two other English speakers was registered but had to be excluded due to an unnoticed hardware equipment failure while recording. The speakers were between 18 and 35 years old and without any present or past speech or hearing problems. They were recruited at the University of Potsdam and paid for their participation in the experiment. All procedures were performed in compliance with relevant laws and institutional guidelines and were approved by the ethics committee of the University of Potsdam. Written informed consent was obtained from all subjects.

During the experiment, participants were prompted on a computer screen to produce sequences of repeated [ta] or [ka] syllables in time with an audible metronome. The metronome served as an extrinsic index of the intended rate of syllable production. The participants were instructed to articulate their responses accurately and naturally. The rate of the metronome was set to the values of 30, 90, 150, 210, 300, 390, 480 and 570 bpm (corresponding to 0.5, 1.5, 2.5, 3.5, 5.0, 6.5, 8.0 and 9.5 Hz). At the start of each trial, the participant was exposed to the metronome beats and begun articulating the required response syllable at a point of their choice. Intuitively, we can describe this by saying that participants begun uttering repetitive syllables once they had listened to the metronome and internalized the intended rate. Once participants begun uttering, the length of each such trial of repetitive syllable productions consisted in a sequence of approx. 30 syllables. Starting with the slowest rate, a minimum of four trials at each of the above rates were recorded. Once this minimum of four trials was reached, recording proceeded to the next higher rate. The entire procedure was performed in two successive blocks, first for sequences of [ta] and then for sequences of [ka].

Articulatory data as well as acoustic data were registered from all participants. All recordings took place in our sound-attenuated booth using a Carstens AG501 3D Electromagnetic Articulograph for articulatory, and a YOGA Shotgun microphone EM-9600 attached to a TASCAM US-2x2 Audio interface for acoustic data registration. Three-dimensional electromagnetic articulography (EMA) allowed measurement of kinematic displacement data of selected articulators at a high precision. Along with some other auxiliary reference locations (upper and lower incisors, nose bridge, left and right mastoid processes), we tracked the positions of sensors attached to the tongue tip and tongue back articulators, the major effectors involved in the production of [ta] and [ka] respectively. All data and source code files used to produce the results presented below are fully available under doi: 10.5281/zenodo.2273898.

2.1 Data processing

Three-dimensional displacement data, provided by the AG501 device, was digitized at a sampling rate of 1250 Hz. In order to reduce storage and memory footprint as well as to improve further data processing performance, the sampling rate of all signals was decreased to a value of 83.33 Hz by means of a double-stage decimation procedure. This decimation procedure included two successive 30th order equi-ripple FIR lowpass filters (with an effective cutoff frequency of 41.67 Hz) eliminating most high frequency noise. Based on the downsampled, noise-reduced signals, spatial transformations of head movement correction and occlusal reference frame alignment were determined and applied by means of the method proposed by Horn (1987). Finally, a zero-delay Butterworth lowpass filter of fourth order with cutoff frequency of 25 Hz was utilized to eliminate any further noise potentially present (see Maoz, Portugaly, et al., 2006).

All signals were represented analytically by means of a quintic spline approximation. Spline approximation was carried out individually for each dimension at equidistant knots given by the constant sampling rate of 83.33 Hz. Quintic polynomials were determined by a least squares regression of data points and continuity constraints of derivatives up to the third order (see e.g., Ramsay and Silverman, 2005; Ruppert, 2011 for a general description of regression splines). This method allowed for the representation of an articulator's essential kinematic properties (displacement, velocity and acceleration) at the same time by a single analytic spline object. Numerical differentiation was carried out using nine-point finite differences with eighth order of accuracy (stencil coefficients were determined using the Python package; Baer, 2018).

The continuous three-dimensional motion of the tongue back and tongue tip articulators was segmented into separate, successive closing and opening movements. The basis for this segmentation was the first derivative (velocity) of the displacement's principal component (by means of a PCA, representing displacement along movement direction). Instants of zero-crossings in the PCA velocity were used as primary movement delimiters (see, e.g., Munhall, Ostry, et al., 1985). Movement onset and offset locations were determined as the points where velocity rose above (onset) or fell below (offset) 20% peak velocity. In cases where the value of 20% peak velocity was crossed more than twice in a single (closing or opening) movement (multi-peak velocity profiles), we chose the earliest crossing as the movement onset delimiter and the latest crossing as the offset delimiter. The 20% threshold was chosen in order to avoid potentially poorly defined transitions into or away from a quasi-steady-state phase (Hoole, Mooshammer, et al., 1994). Segmentation was carried out fully automatically. A final, manual task was performed by selecting only those movements directly involved in forming or releasing tongue-palate constrictions neglecting any other, non-targeted movements potentially present (movements between syllables which did not contribute to any acoustic output, primarily present at the slowest rate, were neglected). As an example of our data, Figure 1 shows a series of [ka] syllables produced by one of our participants at a metronome rate of 150 bpm. In total, we registered 12 738 movements in the [ta] case (6 284 closures and 6 454 openings) and 12 379 movements in the [ka] case (6 119 closures and 6 260 openings).

3. Results

In the context of previous work on the speed-curvature relation, the distinguishing property of our experimental design is the extensive range of rates under which participants spoke their utterances. Our participants spoke syllables at eight distinct rates ranging from extremely slow (30 bpm) to extremely fast (570 bpm). For comparison, in prior work on repetitive speech, Kelso, Vatikiotis-Bateson, et al. (1985)

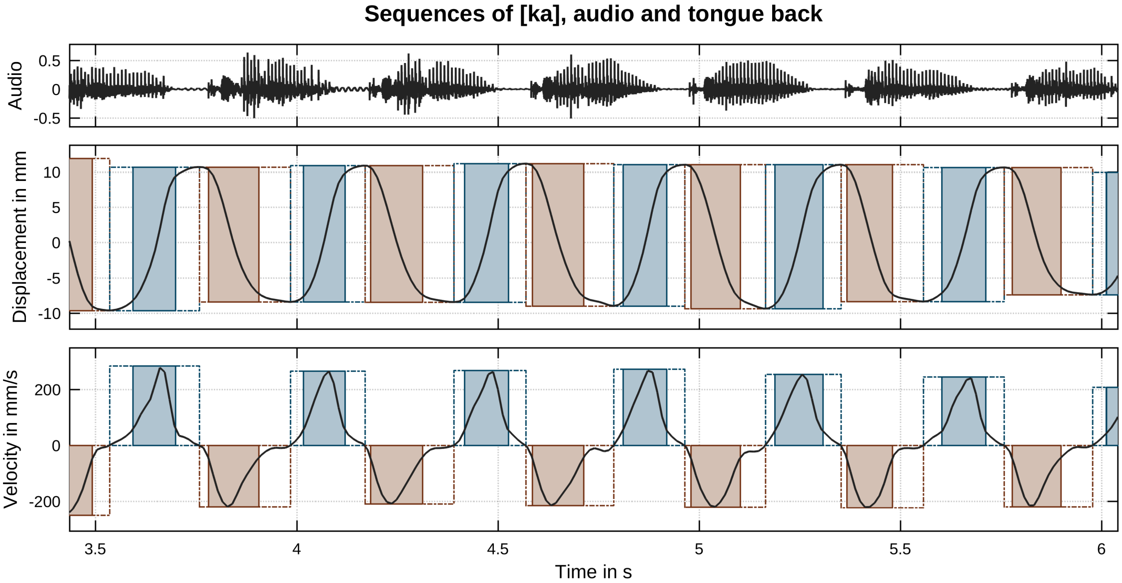

Figure 1. Section of a [ka] sequence at 150 bpm. (Top) Acoustic recording. (Middle) Principal component (PCA) of tongue back displacement. (Bottom) First derivative (velocity) of PCA. Segmented movements are indicated by shaded boxes (closing movements: blue, opening movements: brown). Dashed line rectangles correspond to movement delimiters (zero velocity) and solid line rectangles to movement onsets and offsets determined by 20% peak velocity crossings.

included two rates (speaker-selected conversational and fast), as did Ostry, Cooke, et al. (1987) (approx. one syllable and two syllables per second), and Patel, Löfqvist, et al. (1999) used a metronome to suggest a rate to their participants (which was 120 bpm) but there was no metronome during the actual registration of a participant's utterances. In previous work on the speed-curvature power law, Perrier and Fuchs (2008) considered three rates (slow, normal and fast) in target-to-target biomechanically simulated tongue movements. Tomaschek, Arnold, et al. (2018) investigated two speech rates (slow and fast) by control of inter-stimulus and stimulus presentation times. In light of our extensive speech rate manipulation, it is thus imperative to first ensure that our so-registered data are in conformity with what is known about speech movements from earlier work.

For any movement, the three basic kinematic properties are those of movement duration T, movement amplitude A and peak velocity v^* (Nelson, 1983). In our data, movement duration was determined as the time between movement onset and offset based on 20% peak velocity delimiters. Movement amplitude was computed as the length of the trajectory the effector moved in three-dimensional space. Peak velocity was determined as the maximum magnitude of tangential velocity

$$v = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}, \tag{2}$$

where $\dot{x}$, $\dot{y}$ and $\dot{z}$ denote the first derivatives of the horizontal, vertical and lateral components of the effector's position in space.

One empirically well-documented relation between the above kinematic parameters is the correlation between movement's peak velocity v^* and its amplitude A. It has been reported for a large variety of consonant-vowel (CV) and vowel-consonant (VC) sequences involving movements of tongue body, tongue tip, lips and jaw (Ostry, Keller, et al., 1983). The relationship has been described as an overall linear correlation (Ostry and Munhall, 1985) with A-v^* slopes steeper for faster than for slower speech rates (Vatikiotis-Bateson and Kelso, 1990, 1993) and decreasing covariation as durational variability increases (ibidem). Evidence for divergence from linearity at larger amplitudes is known from previous data and has been of concern in previous modeling studies (e.g., Sorensen and Gafos, 2016). Figure 2 and Figure 3 show per speaker A-v^* scatter plots of our data in sequences of [ta] and [ka]. For both sequences there is an overall correlation between peak velocity and movement amplitude with divergence from linearity at larger amplitudes. In addition, we observe substantially steeper correlation slopes of A and v^* for faster speech rates (indicated by darker shades) than for slower rates (indicated by brighter shades). Thus, the characteristics of the A-v^* relation in our data are in broad conformity with earlier reports across the entire range of induced speech rates in our paradigm.

Another prototypical feature of speech movements is the presence of a relation between all three kinematic parameters considered. It has been commonly reported that the ratio of peak velocity to amplitude v^*/A varies inversely with movement duration T (e.g., Munhall, Ostry, et al., 1985) across manipulations of stress, vowel and consonant identity (Ostry and Munhall, 1985; Fuchs, Perrier, et al., 2011). The relations found in our data are shown in Figure 4 and Figure 5 for both sequences of [ta] and [ka]. Regressions of the inverse correlation $v^*/A \propto 1/T$ are drawn in black color. Values of correlation slopes separated by

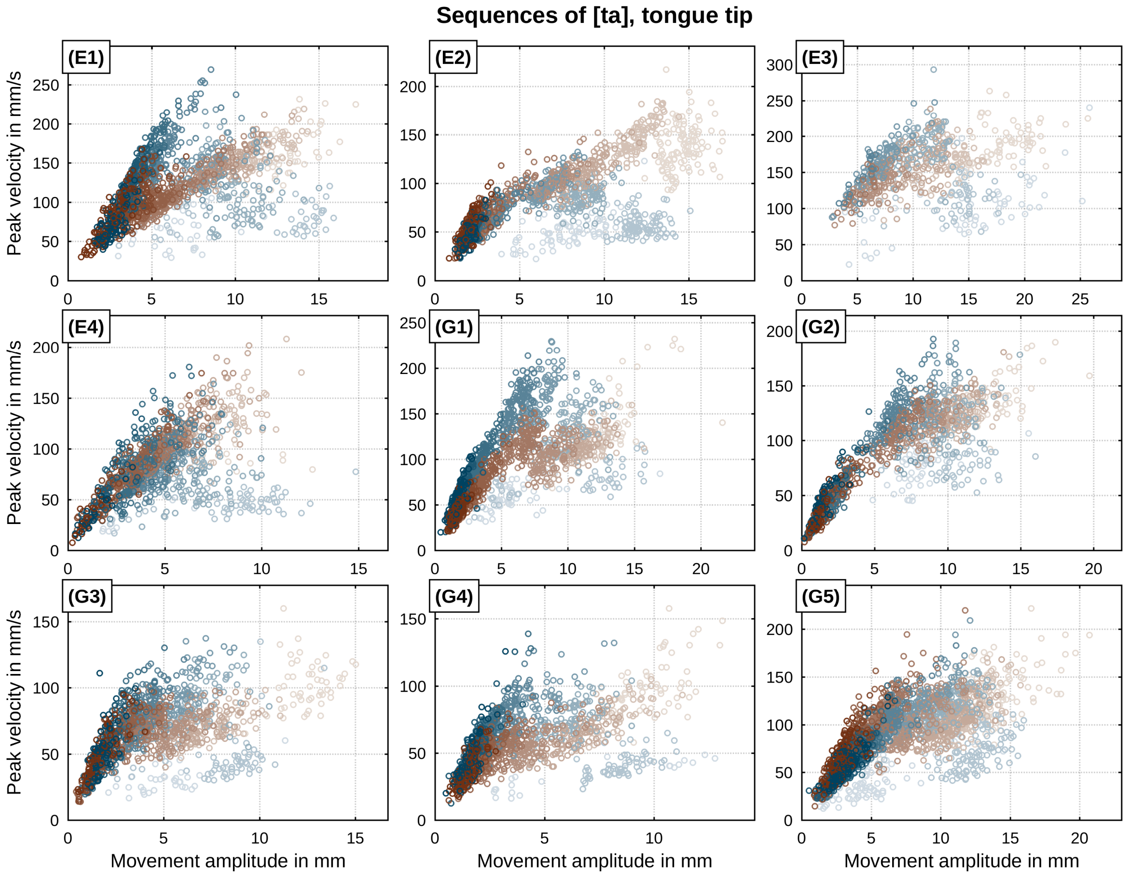

Figure 2. Per-speaker relation between peak velocity and movement amplitude in sequences of [ta]. Direction of movement is indicated by color (closing: blue, opening: brown). Metronome rate is indicated by gradual shade (slowest: bright, fastest: dark).

movement direction are listed in Table 1 (for direct comparison with the constant factor c in Equation (4) these values are divided by π). In conformity with past observations, our data exhibit a clear inverse correlation between the ratio of peak velocity to movement amplitude and movement duration. Deviations from the overall relationship at longer movement durations (slower metronome rates, indicated by brighter shades) have been previously reported by Ostry and Munhall (1985) and related to changes in articulator stiffness.

Both empirical relations discussed here are in good agreement with assumptions about the control system governing speech gestures. A long line of work has proceeded on the hypothesis that the units of action underlying the flow of speech movements are controlled by an organization similar to a mass-spring

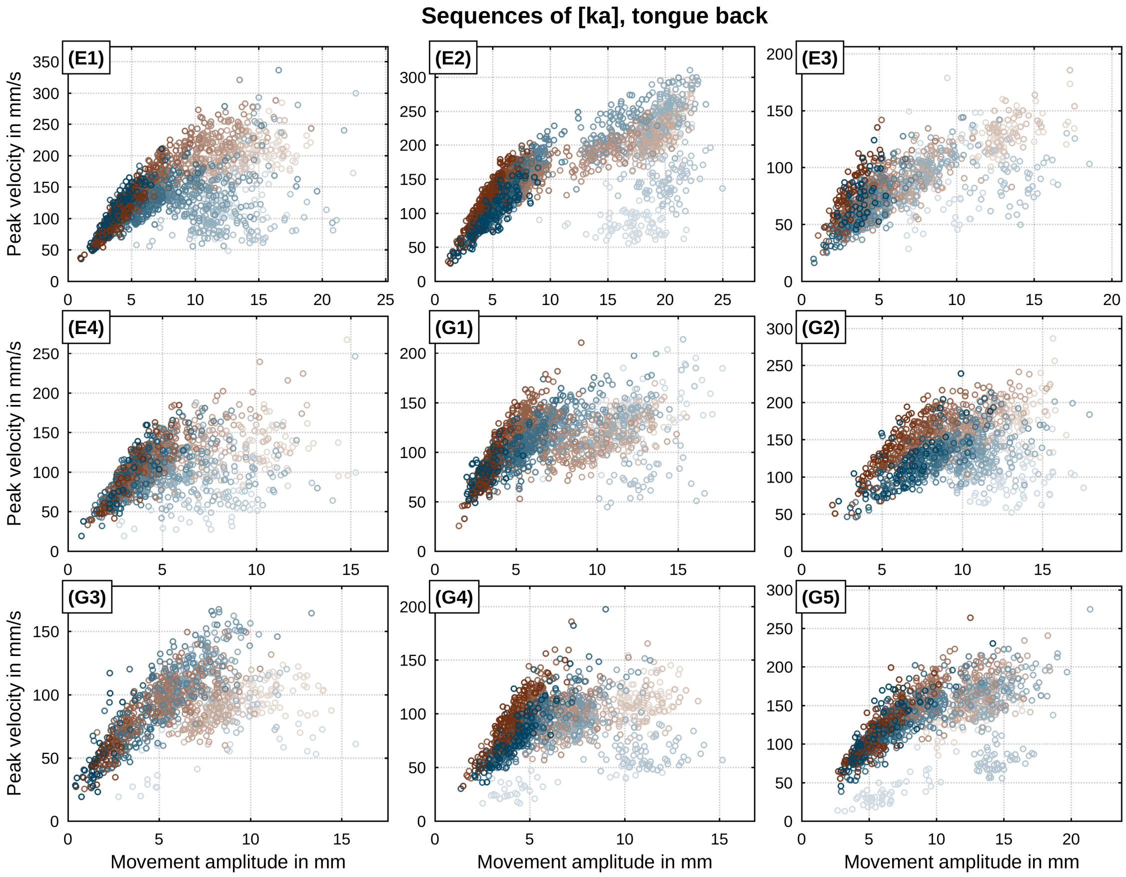

Figure 3. Per-speaker relation between peak velocity and movement amplitude in sequences of [ka]. Direction of movement is indicated by color (closing: blue, opening: brown). Metronome rate is indicated by gradual shade (slowest: bright, fastest: dark).

	E1	E2	E3	E4	G1	G2	G3	G4	G5
[ta] closing	0.515	0.515	0.516	0.538	0.507	0.499	0.508	0.487	0.484
[ta] opening	0.454	0.474	0.505	0.479	0.477	0.470	0.489	0.485	0.471
[ka] closing	0.461	0.471	0.467	0.482	0.451	0.458	0.493	0.462	0.462
[ka] opening	0.460	0.451	0.471	0.475	0.476	0.455	0.482	0.482	0.471

Table 1. Per-speaker regression slopes (in terms of π) of the inverse correlation between the ratio of peak velocity to amplitude and movement duration. Values are separated by sequence and movement direction.

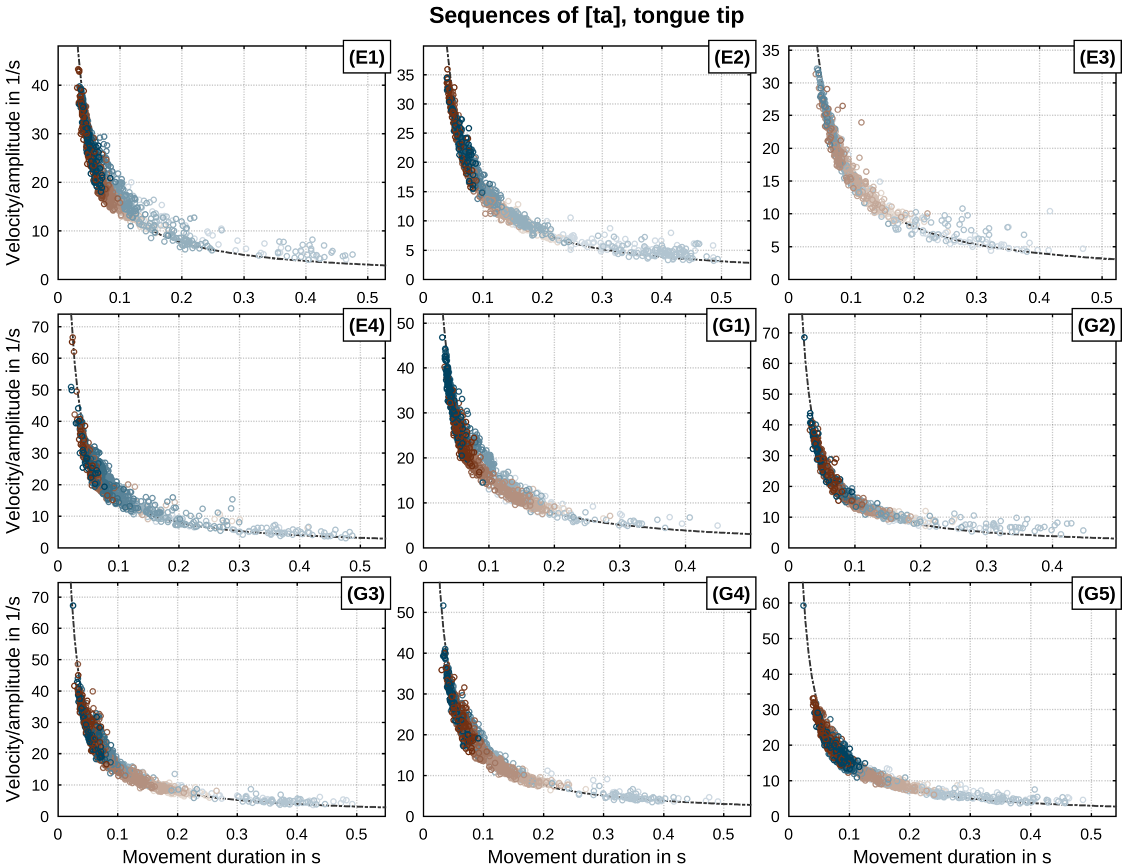

Figure 4. Per-speaker relation between ratio of peak velocity to movement amplitude and duration in sequences of [ta]. Direction of movement is indicated by color (closing: blue, opening: brown). Metronome rate is indicated by gradual shade (slowest: bright, fastest: dark). Regression lines according to $v^*/A \propto 1/T$ are drawn as black curves.

system (Fowler, Rubin, et al., 1980; Browman and Goldstein, 1986; Saltzman and Munhall, 1989). The standard model of the gesture is a special case of the damped linear oscillator with critical damping or more specifically

$$\ddot{x} = -kx - b\dot{x}, \qquad b = 2\zeta\sqrt{k}, \qquad \zeta = 1, \tag{3}$$

where stiffness k (not to be confused with the velocity gain factor in Equation (1) earlier) acts as the model control parameter. Critical damping is realized by a damping ratio of $\zeta = 1$ which also fixes the damping constant b of the model. Analytical solutions of the dynamical system in (3) can be derived by means of methods from the theory of ordinary differential equations. Analysis of these solutions

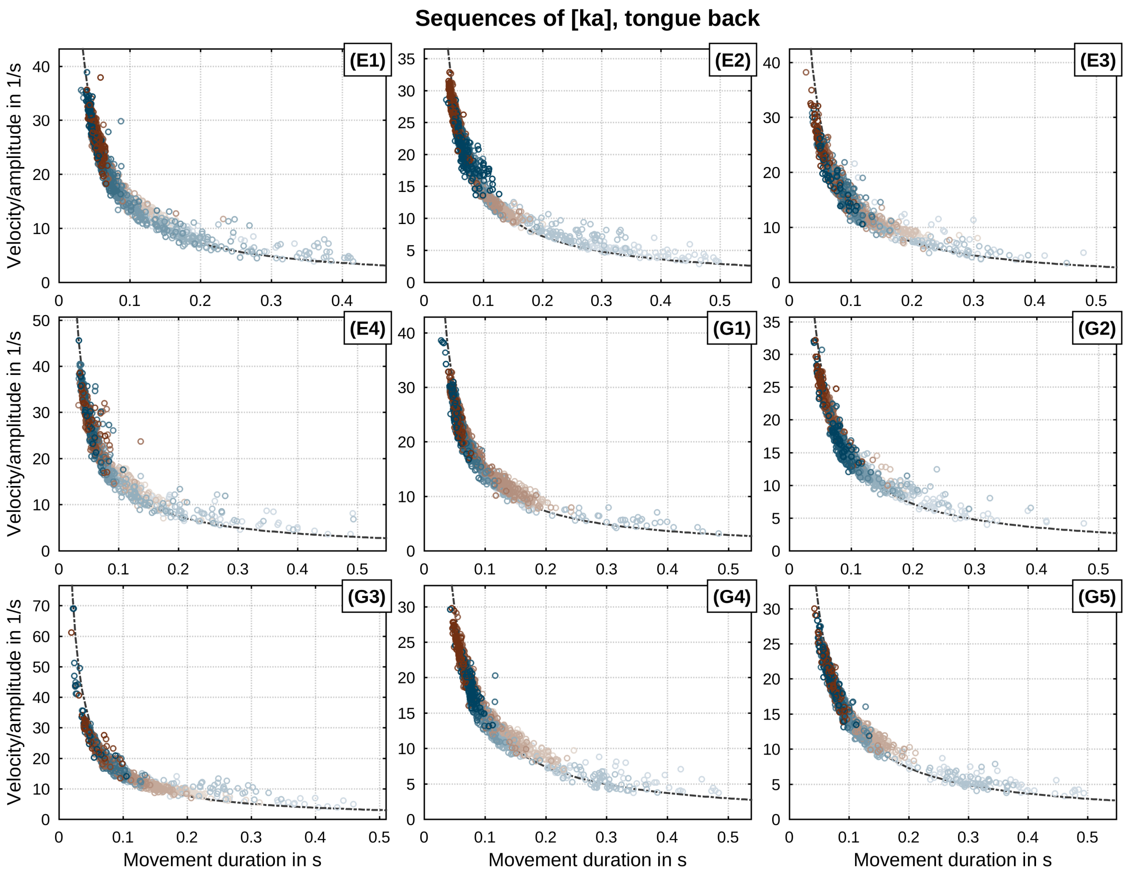

Figure 5. Per-speaker relation between ratio of peak velocity to movement amplitude and duration in sequences of [ka]. Direction of movement is indicated by color (closing: blue, opening: brown). Metronome rate is indicated by gradual shade (slowest: bright, fastest: dark). Regression lines according to $v^*/A \propto 1/T$ are drawn as black curves.

reveals the following relations between the kinematic properties of peak velocity v^*, movement amplitude A and movement duration T

$$v^* = c\sqrt{k}A \qquad \text{and} \qquad \frac{v^*}{A} = \frac{c\pi}{T}. \tag{4}$$

These are a linear correlation between peak velocity v^* and movement amplitude A (equation on the left) and an inverse relation between the ratio of peak velocity to amplitude and movement duration T (equation on the right) sharing the same constant factor c (if π is factored out as in the right equation).

It can be analytically proven that the factor c is upper-bounded by the value of one half (Fuchs, Perrier, et al., 2011). Table 1 lists the per-speaker determined factors c of our data. It is evident that there is general agreement between empirical observation and theoretical prediction. In other words, $c \lesssim 1/2$ for almost every observation. However, we also find differences between the two directions of movement (closing and opening movements). Factors c of closing movements generally attain values larger than those of opening movements, revealing a moderate asymmetry between the two types of direction. As we will show below, a parallel asymmetry exists with respect to properties of the conjectured power law.

3.1 Speed-curvature power law

We now turn to the assessment of the speed-curvature power law. In examining the relation between speed and curvature in our datasets, the full three-dimensional trajectories of the articulators were considered. Let us begin by reviewing the definitions of the two essential variables that play out in the speed-curvature power law. The instantaneous speed v of an effector moving in three-dimensional space is given by its tangential velocity expressed in Equation (2). According to the theory of the differential geometry of curves (e.g., do Carmo, 1976), the time-dependent curvature κ of a three-dimensional trajectory is given by

$$\kappa = \sqrt{\frac{(\dot{y}\ddot{z} - \ddot{y}\dot{z})^2 + (\dot{z}\ddot{x} - \ddot{z}\dot{x})^2 + (\dot{x}\ddot{y} - \ddot{x}\dot{y})^2}{(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^3}} \tag{5}$$

with $\ddot{x}$, $\ddot{y}$ and $\ddot{z}$ being the second derivatives (acceleration) of its displacement components. The curvature of a trajectory is a measure of how much a current segment of the trajectory deviates from a straight line. Large values of curvature indicate strongly curved segments whereas values close to zero imply close to straight segments. For the present analysis, we computed speed and curvature values by subsampling each movement's trajectory at five equidistant locations based on zero velocity onset and offset delimiters. It is usual practice that movement onset and offset delimiters are adjusted to coincide with the two instants of 20% peak velocity (or some such similar percentage). This adjustment is effectively a length reduction of movements yielding gaps in the sequence of successive movements (with gaps typically not included in any analysis). These gaps, the transitions between movements of opposite direction, are generally regions of high curvature. Hence, it is not advisable to suppress these gaps (by adjusting the delimiters) in the assessment of the speed-curvature power law. In total we computed approx. 125 000 pairs of speed and curvature values.

The majority of works on the speed-curvature power law in speech have so far considered only two-dimensional, mid-sagittal data. This is due to either method limitations (e.g., x-ray microbeam provided only two-dimensional displacements) or intentional projection of three-dimensional data onto the mid-sagittal plane. The same applies to many works from the field of general motor control which studied two-dimensional movements by physically restricting a participant's action to a plane. However, the speed-curvature power law itself is not restricted to planar movements (Pollick and Ishimura, 1996; Schaal and Sternad, 2001; Neufeld and van Lieshout, 2014). In our data, across all speakers and rates, angular dispersions away from the mid-sagittal plane are in the range of ±5–10° away from that plane and movement displacements perpendicular to the mid-sagittal plane are in the range of 5–10% of displacements along that plane. In sum, lateral components are relatively insignificant. Nevertheless, even insignificant movement in the lateral dimension affects the computation of curvature in 3D as given by Equation (5) specifically so that its value is different from that in 2D. Hence, exactness calls for usage of the proper three-dimensional expressions of speed and curvature as given by Equation (2) and (5). Furthermore, computation of curvature (either two-dimensional or three-dimensional) requires precise information about the second derivative of each component's displacement. High-order numerical differentiation is highly sensitive to noise (the higher the order of the derivative, the more so). Hence, a robust and numerically stable algorithm is required. In addition, expression (5) contains multiple products of first and second derivatives in the fraction's numerator. The sum of these products is related to the cross product $\dot{r} \times \ddot{r}$ of the vectors $\dot{r} = (\dot{x}, \dot{y}, \dot{z})$ and $\ddot{r} = (\ddot{x}, \ddot{y}, \ddot{z})$. Here too, then, to maintain computational accuracy it is crucial that the numerical derivative of $\dot{r}$ precisely approximates $\ddot{r}$. With these strictures in mind, in our data representation, we utilized nine-point stencil differentiators for high precision numerical differentiation as well as single uniform quintic spline objects for the consistent approximations of r, $\dot{r}$ and $\ddot{r}$ at the same time.

As seen in Equation (1), the speed-curvature law states that the speed of movement is a power function of curvature of the related trajectory. An equivalent formulation, in which Equation (1) has been transformed by taking the logarithm of both sides, is given by

$$\log v = \log k - \beta \log \kappa. \tag{6}$$

This is a straight line in the $\log \kappa$-$\log v$ plane whose slope corresponds to the negative exponent $-\beta$ and whose intercept corresponds to $\log k$, with β and k the two empirically-determined constants of the conjectured power law.

A first global overview of speed-curvature relations in our data is presented in Figure 6 and Figure 7: each panel is a per-speaker log-log scatter plot of the two essential parameters, speed and curvature. These depictions offer a global overview because, for now, we pool across our entire spectrum of rates within each speaker's data. In turn, this enables one to see that our experimental paradigm has provided more than ample variation in the two essential parameters whose relation is being assessed here. It is clear, in other words, that we are dealing with a very wide range of speeds and curvatures. It is also clear from this overview that, for each speaker, type of sequence ([ta] or [ka]) and direction of movement (closing or opening), there is evidence for a strong (log-log) correlation between movement speed and trajectory curvature. Individual regression lines according to (6) are drawn in black color. Determined values of the power law's exponent β are in the range of 0.40…0.46 (with correlation strengths $r^2 = 0.84 \ldots 0.93$) in case of [ta] and 0.40…0.48 ($r^2 = 0.78 \ldots 0.87$) in case of [ka]. Velocity gain factors k attain values in the range of 24.92…37.44 for sequences of [ta] and 30.68…45.61 for sequences of [ka]. All p-values reside below 0.0001 implying strongest significance. These results indicate the clear presence of a general power law relation between speed and curvature in our data. Recalling that $v = k\kappa^{-\beta}$, with instantaneous speed of movement v and trajectory curvature κ, a β of 0.46 implies that every time curvature increases by, for example, a factor of 10, the velocity decreases by a factor of roughly one third or $10^{-0.46} = 0.34$. The proportional change in velocity is thus independent of the value of curvature κ. This property of size-independence is the hallmark of a power law and implies that the principles organizing and constraining

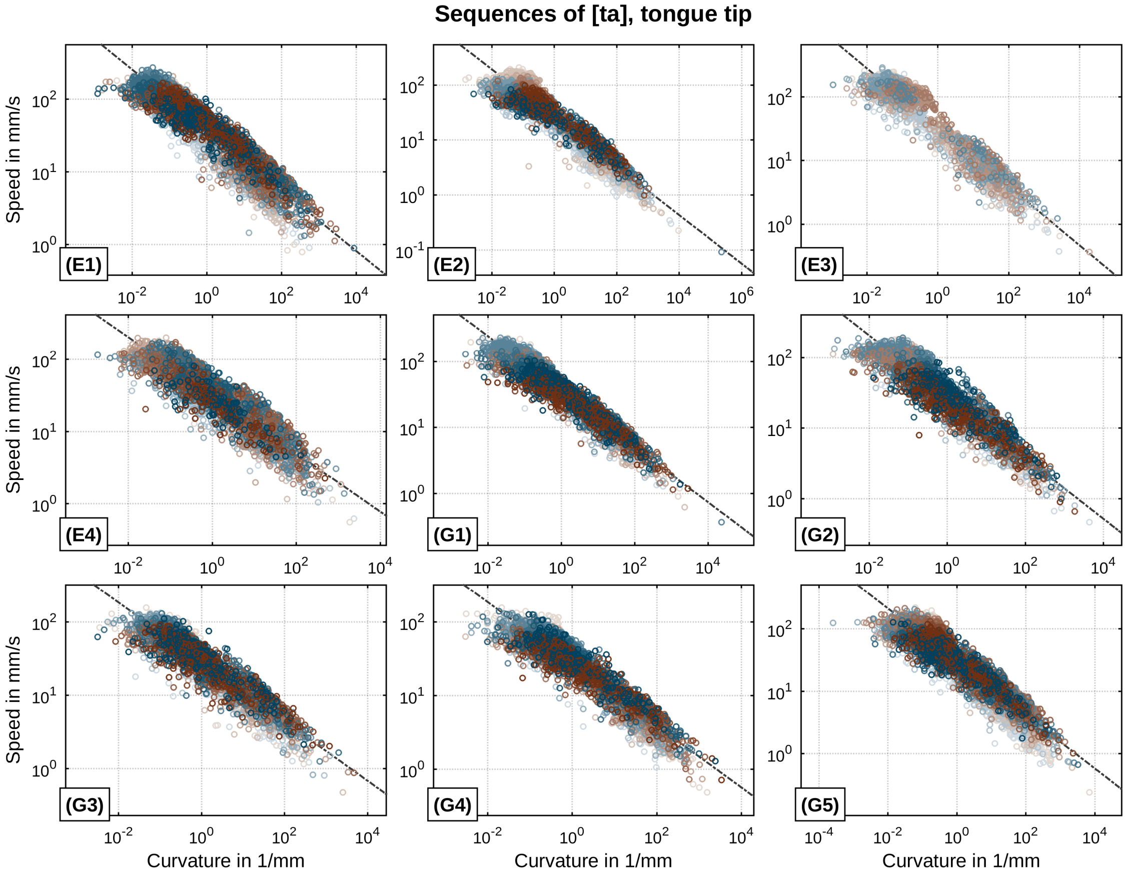

Figure 6. Per-speaker relation between speed and curvature of subsampled movement data in sequences of [ta]. Direction of movement is indicated by color (closing: blue, opening: brown). Metronome rate is indicated by gradual shade (slowest: bright, fastest: dark). Regression lines according to $\log v = \log k - \beta \log \kappa$ are drawn as black lines.

the speech motor system are invariant over variations in kinematic properties and the specifics of the effectors executing the speech movements.

We now proceed to a rate- and direction-specific analysis of our data. The same regression analysis as described above was performed separately for each metronome rate and movement direction. Figure 8 shows the so-determined power law exponents β. It is apparent that there is a substantial dependency of the exponent on the induced speech rate. For any type of movement ([ta] and [ka], closing and opening), the per-speaker values of the power law's exponents at the slowest rate (30 bpm) attain values significantly higher than the commonly reported value of one third but fairly approximate that value as metronome rate

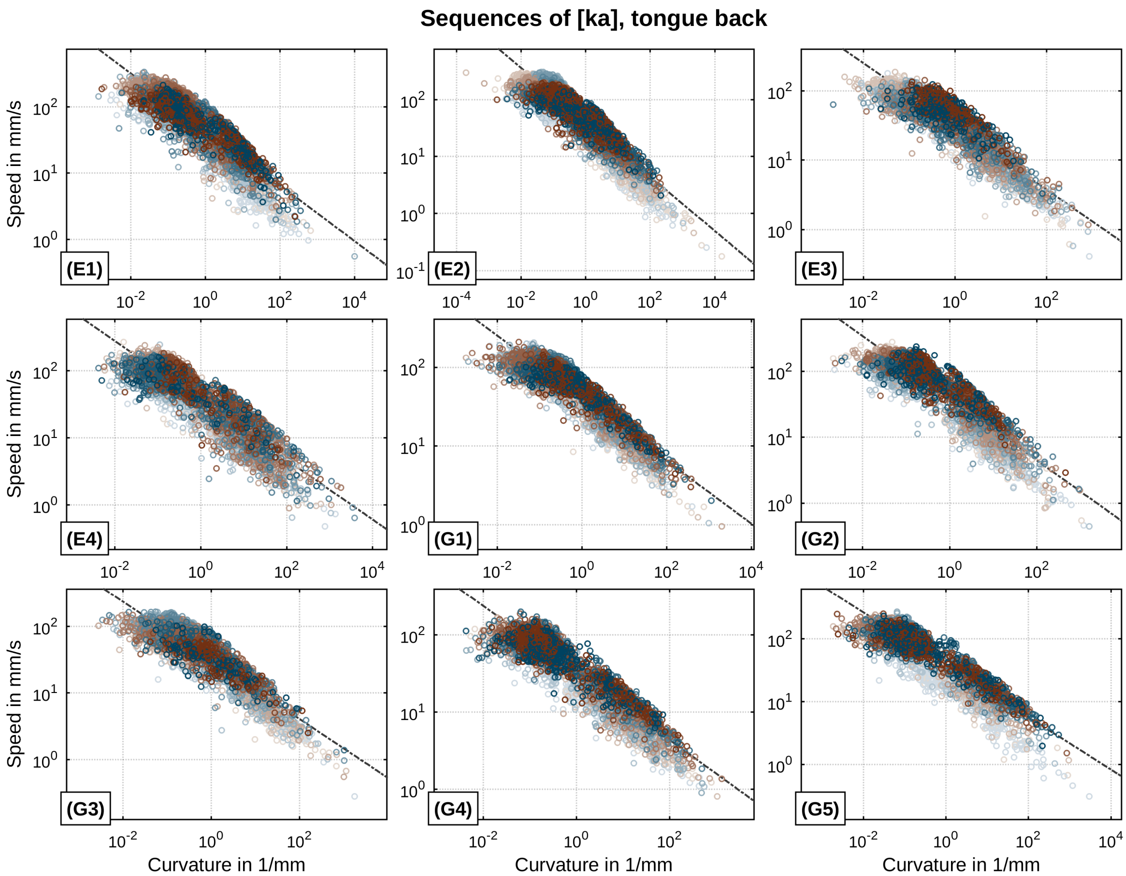

Figure 7. Per-speaker relation between speed and curvature of subsampled movement data in sequences of [ka]. Direction of movement is indicated by color (closing: blue, opening: brown). Metronome rate is indicated by gradual shade (slowest: bright, fastest: dark). Regression lines according to $\log v = \log k - \beta \log \kappa$ are drawn as black lines.

increases. There are minor differences with respect to movement direction and identity of the articulator: decreases of β values are slightly steeper for opening movements than for closing movements and exponents of [ka] sequences attain slightly larger values than [ta] sequences. Values of the per-speaker determined velocity gain factors, as shown in Figure 9, exhibit a clear rate dependency. That is, values of k increase with increasing metronome rate and saturate at faster rates. At faster rates, velocity gain factors for [ka] sequences attain values slightly larger than those for [ta] sequences. There are no notable differences between the two directions of movements in terms of k. Table 2 lists all determined correlation strengths r^2 of the power law. For any performed

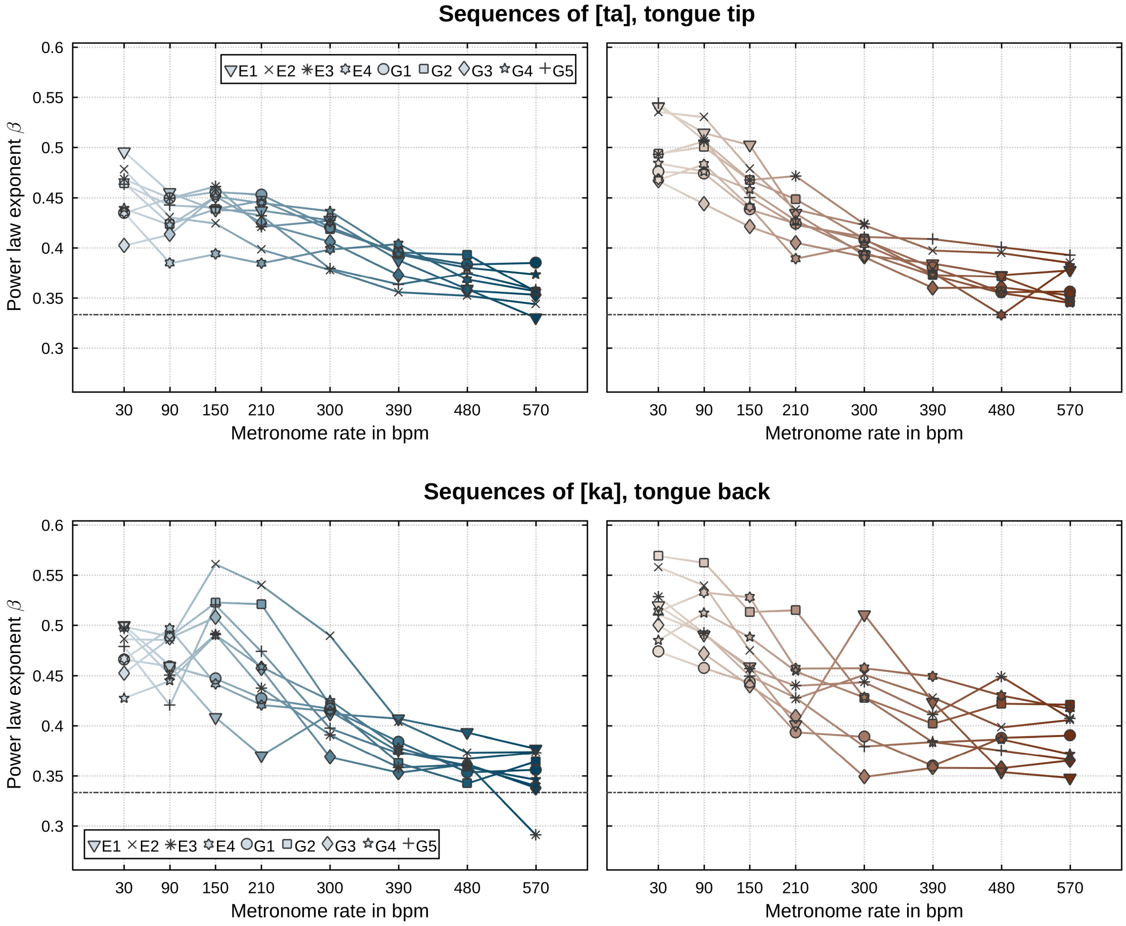

Figure 8. Power law exponents β per speaker (symbols) and metronome rate (abscissa). The commonly reported constant of one-third is indicated by a horizontal dashed line. (Top row) Sequences of [ta]. (Bottom row) Sequences of [ka]. (Left column) Closing movements. (Right column) Opening movements.

regression, the correlations' p-values reside below 0.0001 indicating strongest significance. The overall range of r^2 values is 0.76…0.97 (median: 0.92) in case of [ta] and 0.54…0.96 (median: 0.89) in case of [ka] sequences. There is a moderate trend for these values to decrease with induced speech rate. Sequences of [ta] attain r^2 values slightly larger than those of [ka].

These rate- and direction-specific results both support and sharpen previous results on the speed-curvature power law in a number of ways. Across all metronome rates and the two directions of movement, there exists a highly robust (log-log) correlation between movement speed and trajectory curvature: corre-

	E1	E2	E3	E4	G1	G2	G3	G4	G5
[ta] closing, 30 bpm	0.94	0.94	0.94	0.89	0.90	0.95	0.83	0.91	0.92
[ta] closing, 90 bpm	0.92	0.94	0.95	0.87	0.93	0.93	0.89	0.94	0.94
[ta] closing, 150 bpm	0.90	0.95	0.96	0.87	0.95	0.92	0.94	0.95	0.95
[ta] closing, 210 bpm	0.91	0.95	0.96	0.85	0.97	0.94	0.95	0.94	0.94
[ta] closing, 300 bpm	0.93	0.94	0.95	0.86	0.96	0.91	0.93	0.92	0.92
[ta] closing, 390 bpm	0.93	0.92	n/a	0.84	0.94	0.86	0.85	0.90	0.86
[ta] closing, 480 bpm	0.91	0.92	n/a	0.77	0.90	0.82	0.85	0.90	0.88
[ta] closing, 570 bpm	0.79	0.89	n/a	0.76	0.89	0.79	0.83	0.87	0.84
[ta] opening, 30 bpm	0.95	0.96	0.96	0.95	0.97	0.95	0.91	0.94	0.94
[ta] opening, 90 bpm	0.93	0.95	0.94	0.95	0.94	0.95	0.93	0.92	0.94
[ta] opening, 150 bpm	0.91	0.95	0.93	0.90	0.95	0.94	0.94	0.94	0.95
[ta] opening, 210 bpm	0.91	0.94	0.91	0.88	0.95	0.94	0.94	0.94	0.94
[ta] opening, 300 bpm	0.91	0.93	0.91	0.87	0.95	0.92	0.93	0.93	0.92
[ta] opening, 390 bpm	0.91	0.93	n/a	0.84	0.93	0.88	0.85	0.90	0.87
[ta] opening, 480 bpm	0.90	0.92	n/a	0.78	0.86	0.85	0.86	0.90	0.88
[ta] opening, 570 bpm	0.83	0.89	n/a	0.80	0.86	0.84	0.83	0.85	0.87
[ka] closing, 30 bpm	0.92	0.92	0.92	0.88	0.93	0.95	0.93	0.91	0.91
[ka] closing, 90 bpm	0.88	0.94	0.93	0.87	0.89	0.91	0.90	0.92	0.90
[ka] closing, 150 bpm	0.85	0.94	0.88	0.86	0.92	0.90	0.90	0.90	0.95
[ka] closing, 210 bpm	0.76	0.92	0.85	0.85	0.89	0.93	0.87	0.90	0.92
[ka] closing, 300 bpm	0.70	0.86	0.77	0.86	0.90	0.88	0.82	0.90	0.92
[ka] closing, 390 bpm	0.82	0.83	0.72	0.81	0.88	0.81	0.78	0.90	0.91
[ka] closing, 480 bmp	0.92	0.80	0.71	0.79	0.85	0.76	0.66	0.85	0.90
[ka] closing, 570 bpm	0.86	0.74	0.54	0.74	0.83	0.72	0.73	0.84	0.89
[ka] opening, 30 bpm	0.91	0.96	0.94	0.90	0.92	0.94	0.94	0.94	0.91
[ka] opening, 90 bpm	0.88	0.96	0.93	0.88	0.90	0.94	0.92	0.93	0.88
[ka] opening, 150 bpm	0.89	0.94	0.88	0.90	0.92	0.94	0.88	0.93	0.92
[ka] opening, 210 bpm	0.81	0.89	0.85	0.88	0.90	0.94	0.85	0.92	0.92
[ka] opening, 300 bpm	0.77	0.87	0.79	0.85	0.92	0.90	0.80	0.93	0.93
[ka] opening, 390 bpm	0.82	0.87	0.75	0.83	0.92	0.87	0.77	0.92	0.92
[ka] opening, 480 bpm	0.88	0.85	0.77	0.79	0.90	0.82	0.70	0.91	0.92
[ka] opening, 570 bpm	0.84	0.79	0.67	0.76	0.88	0.79	0.76	0.88	0.89

Table 2. Pearson correlation coefficients r^2 of full, three-dimensional data separated by speaker, movement direction and metronome rate. All p-values reside below 0.0001 indicating strongest significance.

	E1	E2	E3	E4	G1	G2	G3	G4	G5
[ta] closing, 30 bpm	0.87	0.89	0.87	0.84	0.85	0.92	0.79	0.90	0.86
[ta] closing, 90 bpm	0.85	0.90	0.90	0.80	0.87	0.87	0.86	0.91	0.91
[ta] closing, 150 bpm	0.83	0.89	0.93	0.82	0.90	0.88	0.87	0.90	0.91
[ta] closing, 210 bpm	0.86	0.88	0.92	0.78	0.93	0.91	0.90	0.89	0.89
[ta] closing, 300 bpm	0.88	0.88	0.89	0.80	0.92	0.84	0.88	0.86	0.87
[ta] closing, 390 bpm	0.89	0.85	n/a	0.77	0.90	0.81	0.79	0.84	0.81
[ta] closing, 480 bpm	0.87	0.86	n/a	0.70	0.86	0.72	0.81	0.84	0.80
[ta] closing, 570 bpm	0.67	0.82	n/a	0.67	0.81	0.71	0.78	0.82	0.76
[ta] opening, 30 bpm	0.91	0.92	0.93	0.84	0.92	0.94	0.85	0.89	0.91
[ta] opening, 90 bpm	0.88	0.93	0.87	0.89	0.89	0.91	0.91	0.88	0.91
[ta] opening, 150 bpm	0.84	0.90	0.87	0.85	0.92	0.89	0.89	0.85	0.91
[ta] opening, 210 bpm	0.87	0.88	0.86	0.83	0.90	0.92	0.89	0.88	0.89
[ta] opening, 300 bpm	0.88	0.88	0.89	0.80	0.92	0.87	0.88	0.86	0.87
[ta] opening, 390 bpm	0.87	0.87	n/a	0.76	0.88	0.81	0.76	0.83	0.82
[ta] opening, 480 bpm	0.85	0.87	n/a	0.70	0.81	0.78	0.82	0.81	0.82
[ta] opening, 570 bpm	0.73	0.82	n/a	0.67	0.77	0.80	0.79	0.80	0.79
[ka] closing, 30 bpm	0.88	0.88	0.90	0.82	0.85	0.90	0.85	0.81	0.85
[ka] closing, 90 bpm	0.83	0.89	0.86	0.78	0.82	0.87	0.82	0.85	0.82
[ka] closing, 150 bpm	0.80	0.94	0.80	0.83	0.86	0.84	0.82	0.85	0.92
[ka] closing, 210 bpm	0.70	0.88	0.77	0.83	0.81	0.88	0.80	0.85	0.88
[ka] closing, 300 bpm	0.59	0.80	0.69	0.80	0.81	0.83	0.74	0.85	0.87
[ka] closing, 390 bpm	0.72	0.78	0.64	0.77	0.80	0.76	0.67	0.84	0.86
[ka] closing, 480 bmp	0.86	0.73	0.64	0.69	0.79	0.66	0.57	0.77	0.85
[ka] closing, 570 bpm	0.79	0.65	0.47	0.61	0.75	0.64	0.57	0.77	0.82
[ka] opening, 30 bpm	0.87	0.91	0.90	0.84	0.86	0.91	0.89	0.91	0.86
[ka] opening, 90 bpm	0.80	0.89	0.85	0.80	0.80	0.85	0.81	0.90	0.81
[ka] opening, 150 bpm	0.84	0.90	0.76	0.85	0.86	0.88	0.80	0.90	0.82
[ka] opening, 210 bpm	0.77	0.82	0.76	0.85	0.81	0.89	0.79	0.88	0.84
[ka] opening, 300 bpm	0.69	0.82	0.71	0.81	0.84	0.88	0.71	0.87	0.88
[ka] opening, 390 bpm	0.73	0.83	0.67	0.78	0.85	0.84	0.70	0.85	0.86
[ka] opening, 480 bpm	0.84	0.77	0.73	0.66	0.84	0.73	0.55	0.84	0.87
[ka] opening, 570 bpm	0.77	0.66	0.62	0.67	0.85	0.75	0.57	0.80	0.79

Table 3. Pearson correlation coefficients r^2 of intentionally mid-sagittally projected data (2D) separated by speaker, movement direction and metronome rate. All p-values reside below 0.0001.

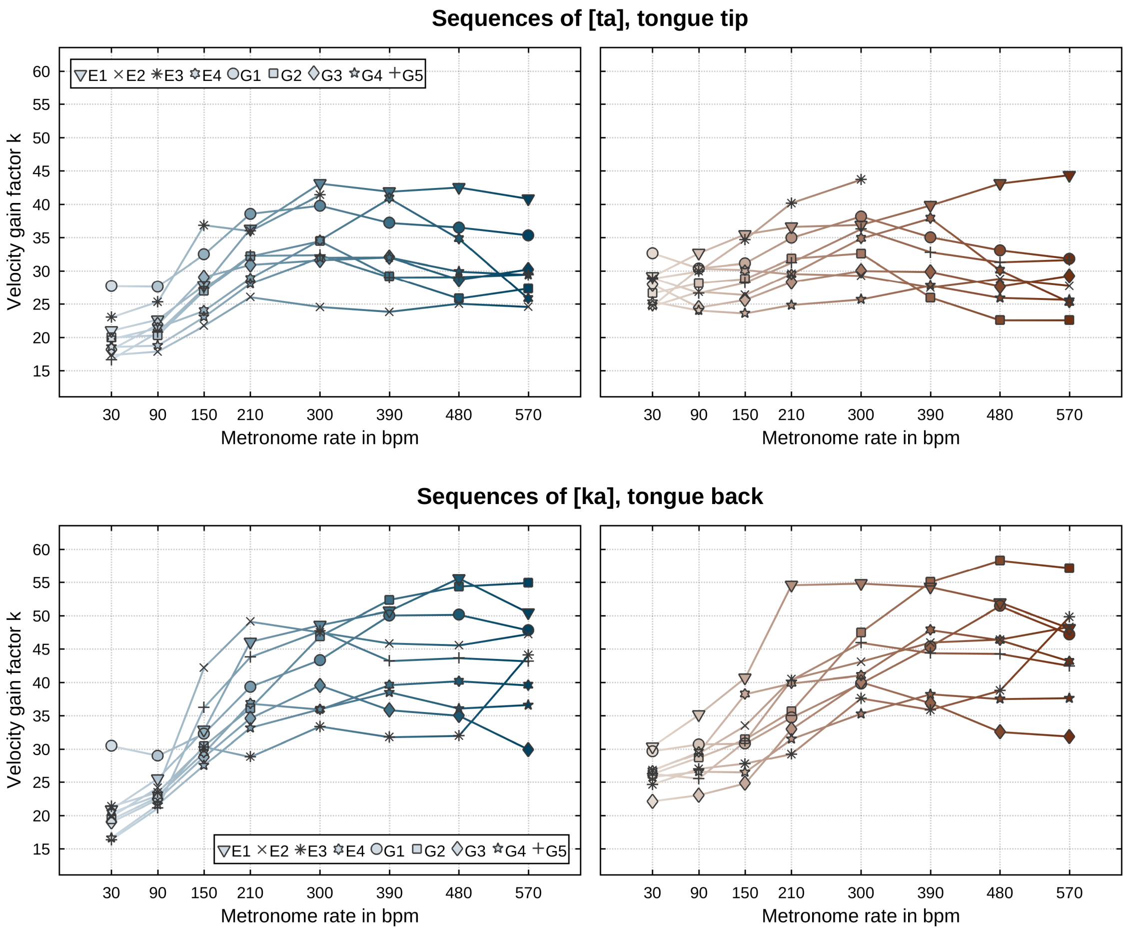

Figure 9. Velocity gain factors k per speaker (symbols) and metronome rate (abscissa). (Top row) Sequences of [ta]. (Bottom row) Sequences of [ka]. (Left column) Closing movements. (Right column) Opening movements.

lation strengths (see Table 2) attain values significantly higher than those reported in any previous work (Tasko & Westbury: 0.60–0.68, Perrier & Fuchs: 0.49–0.67, Neufeld & van Lieshout: 0.71–0.72, Tomaschek et al.: 0.68–0.73). This result seems, in one part, attributable to our enhanced data representation utilizing three-dimensional signals with advanced numerical differentiation. Regression results from our data intentionally projected to the mid-sagittal plane (2D) still show evidence for the presence of the conjectured power law, but with correlation strengths falling by 5–10 percentage points in comparison to those determined from the full, three-dimensional data; see Table 3 which shows r^2 values of mid-sagittally projected data in comparison to Table 2 of full, three-dimensional data.

Relatedly, in some previous works on the speed-curvature power law in speech, it has been reported that the (log-log) correlation between speed and curvature notably weakens for smaller curvature κ values. Thus, two previous studies on the speed-curvature relation, Tasko and Westbury (2004) and Tomaschek, Arnold, et al. (2018), excluded data points with extremal values of curvature from their analyses (Tasko & Westbury: $\kappa < 0.006\,\text{mm}^{-1}$, Tomaschek et al.: $\kappa < 10^{-4}\,\text{mm}^{-1}$ and $\kappa > 10^{3}\,\text{mm}^{-1}$). Tomaschek, Arnold, et al. (2018) even suggested a non-linear relation between (log) speed and (log) curvature. Visual inspection of the speed-curvature relation as shown in Figure 7 may appear to suggest for some speakers (e.g., G5, [ka]) a non-linear relation. However, segregation by rate, as pursued in our analysis, reveals a much more orderly picture. As we have demonstrated, exponents β exhibit a clear systematic dependency on metronome rate with strong significance. Different values of the power law's exponent β correspond to different slopes of the linear relation between log speed and log curvature. It is thus a consequence of this fact that the data, when pooled across multiple rates, may appear to show some tendency for nonlinearity. Due to the systematicity of the observed rate-dependency of β (small values of curvature correspond to high speeds which in turn are primarily realized at faster rates corresponding to smaller values of β), this tendency is an artifact of superposition of different slopes. When a rate-specific analysis is performed, small curvature values do not indicate any deviation from the (log-log) linearity between speed and curvature. Additionally, in computing curvature, numerical issues are amplified precisely at small curvature values. This is so due to instabilities in numerical estimation when $\dot{r}$ and $\ddot{r}$ in (5) are nearly linearly dependent vectors. Two vectors are said to be linearly dependent if they point in the same direction. The cross product of two vectors, which is implicated in computing curvature, vanishes or is very close to zero when the vectors are (nearly) linearly dependent and machine rounding issues become important. Hence, a robust algorithm along the lines described around (5) is called for. In sum, adoption of the natural dimensionality of the data (no reduction to 2D) and advanced numerical differentiation methods indicate no need to either exclude subsets of curvature values or to abandon linearity of the general law relating speed and curvature. As we have seen, in fact, stronger evidence than that reported by any previous work on the law is found when all data is included.

Previous work on the power law in speech sometimes reports differences in the values of exponents β and velocity gain factors k with respect to the identity of the articulator (e.g., Tasko and Westbury, 2004). However, these differences do not seem to hold robustly across studies. Thus, Perrier and Fuchs (2008) report that the direction of variation of β across effectors was speaker-dependent and that the significance of differences in k substantially reduces when segmented movements are considered (ibidem, p. 23). In our data, we find some evidence for articulator-related differences in β only at the slower rates and for k only at the faster rates (see Figure 8 and Figure 9). Specifically, at rates below approx. 300–390 bpm we observe that exponents of the tongue tip articulator (in sequences of [ta]) attain slightly smaller values than those of the tongue back articulator (in sequences of [ka]). For metronome rates above or equal to approx. 300–390 bpm, we find somewhat higher velocity gain factors for the tongue tip than the tongue back. One reason for the lack of agreement here may be that the results of Tasko and Westbury (2004) are based on a "broad form of analysis" (ibidem, p. 74) where movements of sensors placed on the orofacial structures were evaluated regardless of whether the effector (on which the sensor was attached to) was involved actively in the forming of a speech gesture. All speech gestures involve movement, but not all movements are part of a gesture. For example, the lower lip moves upward during the formation of the closure required for the first consonant in [ta], but that movement is a passive consequence of the jaw which rises (in unperturbed conditions) to aid in effecting the tongue tip constriction. In our work, we evaluate the law for effector movements in which that effector is actively engaged in the production of a speech gesture, e.g., the tongue body movement during the forming of the constriction and its release in [ka]. This is in keeping with the literature from other areas of motor control where, for instance, in evaluating the law for planar drawing tasks, it is the effector actively implicated in the act of drawing whose conformity to the law is evaluated. Our main result on the validity of the power law along with certain specifics, to be taken up in the next section, show very good agreement with other studies from other areas of motor control. Finally, in congruence with Perrier and Fuchs (2008) and Tomaschek, Arnold, et al. (2018), we demonstrate that values of the velocity gain factors k generally increase with increasing induced speech rate (see Figure 9). This finding is also in agreement with many reports from the field of general motor control (Viviani and Cenzato, 1985; Wann, Nimmo-Smith, et al., 1988; Viviani and Flash, 1995; Gribble and Ostry, 1996).

4. Discussion: from figural constraints to movement dynamics

The foregoing results and specifically the validity of the power law relation between speech and curvature under the rather wide range of rate conditions elicited in our paradigm implies the existence of general principles that organize and constrain the functioning of the speech motor system. In the present section, we turn to consider these results in the context of other work in movement science, considering both commonalities across domains as well as prospects for furthering our understanding of speech.

In the field of general motor control, the speed-curvature law is most often associated with the commonly reported exponent β of one third. A clear result from our data is that, in speech, an exponent of one third is approximated only at the faster rates. Slower rates exhibit significantly higher β values. As we discuss next, this result finds precursors in the literature on motor control. In turn, this indicates that, over and beyond the existence of the relation between speed and curvature in both speech and other domains, there exist further commonalities, in human movement, across effector systems and the particular modes of their operation. As we also take up next, various features of the performed actions and their lawful modulation as a function of task parameters (specifically, rate in our experiments) may offer potential entries into furthering our understanding of the control systems governing speech movements.

We begin with a brief summary of reports from the non-speech domain pointing to different sources of variability in the power law's parameters. In a three-dimensional drawing task, Schaal and Sternad (2001) reported increasing values of β for increasingly sized patterns. Significantly deviating from one third β values have also been reported for complex-shaped movements not limited to the class of ellipsoids from which initial evidence for the power law originally derived (Hicheur, Vieilledent, et al., 2005; Huh and Sejnowski, 2015). Specifically, Huh and Sejnowski (2015) have demonstrated that in planar drawing tasks complexity of movement shape is positively related to the value of the exponent β, i.e., more complex shapes correspond to higher values of β. A different source for the variability in the power law's exponent has been suggested in terms of constraints under which movements are performed. In a planar drawing task, Wann, Nimmo-Smith, et al. (1988) presented evidence for significantly smaller deviations of the exponent from one third when participants were forced to perform their movements at a faster rate compared to their self-chosen pace. Lastly, Pollick and Ishimura (1996) have demonstrated that three-dimensional target-to-target movements are accompanied by significantly larger than one third exponents and conjectured that these deviations from one third may be related to the control regime underlying the generation of movements. Specifically, that work raises the question whether the regime controlling movements for which only the endpoint target locations are specified is the same as in more constrained movements where effectors must follow specific paths.

Let us place our results in the context of these studies from the non-speech domain. As some of these studies refer to the shapes of trajectories, we present in Figure 10 (left-hand side) traces of tongue back and tongue tip movement patterns by one of our speakers for sequences of [ka] (top row) and [ta] (bottom row) at different metronome rates. By visual inspection of these traces, it is evident that there is a clear decrease in movement excursion extents with increasing metronome rate (from approx. 15 to 10 mm for [ka] and [ta]). In agreement with Schaal and Sternad (2001), this reduction in (what Schaal and Sternad, 2001 refers to as) pattern size is accompanied by a decrease in the power law's β values as demonstrated in our main results section. Furthermore, for both sequences of [ka] and [ta] (Figure 10, left-hand side, top and bottom row), movement patterns exhibit a transition from complex (non-elliptical) shaped traces to simpler elliptical shapes at the faster rates. This observation of a reduction in shape complexity accompanied by decreasing β values agrees with reports by Huh and Sejnowski (2015). However, in our data not all speakers exhibited a clear transition from non-simple to simple shapes (but all speakers did exhibit the decrease in β values). To wit, movement traces of [ka] and [ta] from another speaker are given in the right-hand side of Figure 10 (top and bottom rows). For this speaker, a transition from complex to simpler shapes is hardly visible (for both [ka] and [ta]). However, excursion extents for both articulators systematically decrease from approx. 15 mm

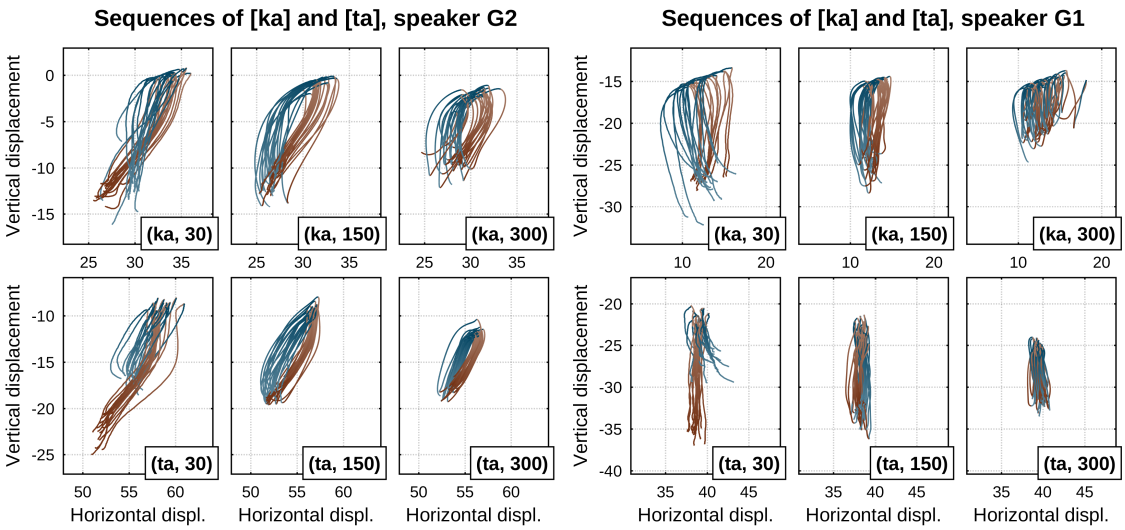

Figure 10. Movement traces of tongue back and tongue tip sensors in sequences of [ka] and [ta] from two speakers at different metronome rates. (Left-hand side) Speaker G2. (Right-hand side) Speaker G1. (Top row) Sequences of [ka]. (Bottom row) Sequences of [ta]. Direction of movement is indicated by color (closing: blue, opening: brown).

to 10 mm at the faster rates as with the other speaker. In sum, [ta] figural shapes are not identical to [ka] shapes (Figure 10, top vs. bottom rows) and they are individual-specific (Figure 10, left-hand vs. right-hand side). However, there is one rigorous generalization that clearly stands out across [ka] and [ta] and across all individuals: as demonstrated in our main results, there is a clear inverse correlation between metronome rate and the law exponent β (Figure 8). β decreases with rate. In other words, whereas shape generalization is not so evident across [ka] and [ta] and across speakers, generalization is seen in the systematicity of how the parameterization of the law changes as a function of rate across articulators and speakers.

Before elaborating on this relation between metronome rate and the law exponent β, let us consider reasons why finding shape generalizations in our data does not appear as straightforward as in Huh and Sejnowski (2015). First, whereas participants in Huh and Sejnowski (2015) were asked to trace planar movements with a single effector, two different effectors are involved in our paradigm, tongue tip and tongue back. Movements of the tongue body implicate deformation of a larger mass than that of the tongue tip. There are strong hints from other areas of biological action that body mass is crucially involved in how movements are organized and effected. For example, in studies of how speed affects locomotion across animals of different sizes, there is evidence that different modes of locomotion occur at similar speeds only when speed is scaled to body mass (Heglund, Taylor, et al., 1974). Second, unlike in drawing movements of the hand studied in Huh and Sejnowski (2015), the tongue moves in the oral cavity. Tongue tip movements for [t] are confined in a space different from that of tongue body movements for [k]. Thus, the presence of a more or less pronounced alveolar ridge and its more posterior palate morphology contribute both to the extent and the direction (horizontal, vertical, lateral) of movements (Stone, Faber, et al., 1992). It is conceivable that taking into account the geometry of the hard structures and in particular palate morphology and the ways in which tongue movements adapt to speaker-specific aspects of that morphology may be crucial in better understanding trajectory shape. For example, Neufeld and van Lieshout (2014) consider algorithms for transforming tongue sensor positions into a palate-relative coordinate space, by means of shortest paths to the surface of a speaker's palate. In the resulting methods, the way in which the trajectory for the tongue tip is transformed is different from that of the tongue back (because the palate has a different shape in the regions where tongue tip versus tongue body actions take place). Effectively, this implies that the geometric shape of tongue tip movements is dependent on the coordinate system chosen. It is thus plausible that using appropriate transformations of the different articulator trajectories may enable identification of shape transitions comparable to those we present in Figure 10 (left-hand side) also for speakers where Cartesian coordinates do not offer evidence for such transitions (right-hand side). Regardless, we emphasize again that even though shape generalizations are not so evident in our data as they may be in other domains (Huh and Sejnowski, 2015), there is a robust systematicity of how the law exponent β changes as a function of rate (Figure 8). This systematicity holds across articulators and speakers.

Let us, finally, consider our results in the context of Wann, Nimmo-Smith, et al. (1988) and the conjecture in Pollick and Ishimura (1996) regarding potential distinctions in the control regime or constraints governing the elicited movements. As demonstrated in our main results, there is a clear inverse correlation between metronome rate and the law exponent β (Figure 8) in agreement with the results of Wann, Nimmo-Smith, et al. (1988): β decreases with rate. In the context of Pollick and Ishimura (1996) who hint at the potential role of differences in the control regime underlying movement execution, it seems relatively uncontroversial that in our study movements at the slower rates can be considered as target-to-target movements with [k] (or [t]) and [a] as the oppositional consonant-vowel targets. In agreement with Pollick and Ishimura (1996), we find that these movements (at slower rates) show significantly larger exponents than one third. Just as Pollick and Ishimura (1996) have raised the question about the presence of a distinction between target-to-target and non-target-to-target movements in terms of the underlying nature of control and corresponding β values, we too are faced with the question whether movements at the faster rates with significantly smaller exponents belong to the same class as movements at the slower rates with higher exponents. This potential distinction between qualitative control regimes underlying what may be apparently similar movements has been pursued in work that has so far remained unrelated to the speed-curvature law. Specifically, there is evidence from other areas of motor control indicating that increasing rate may result in qualitative changes in the control regime underlying movements. In a unimanual finger flexion-extension task, Jirsa and Kelso (2005) and Huys, Studenka, et al. (2008) demonstrated that a dynamical system with distinct control regimes can model the transition from discrete target-to-target movements to rhythmic (cyclic) movements with movement rate as the bifurcation parameter. At the slow rates and specifically below a critical value of rate, finger movements are driven by a control regime governed by fixed point dynamics (with either one or two stable fixed points, depending on whether the motor system instantiates both a flexion-extremum target and an extension-extremum target or just one of these extrema as a target along with a transient excursion away from it). At the faster rates, above the critical value of rate, the control regime switches to a different dynamical organization constituting a limit cycle attractor. It seems plausible that a similar distinction between discrete target-to-target and cyclic (non-targeted) movements separated by a critical speech rate may apply to our data as well.

To assess the potential presence of evidence for distinct dynamical regimes governing movements in our data, we turn to consider the phase portraits of these movements as a function of speech rate. For a second-order dynamical system, as in the class of systems entertained in Jirsa and Kelso (2005) and Huys, Studenka, et al. (2008), phase portraits consist of trajectories in the x-$\dot{x}$-plane (displacement and velocity, the two dimensions of the system's phase space). Figure 11 shows examples of phase portraits of the tongue tip and tongue back articulator's principal component. These figures are spatio-temporal approximations of the phase space's density in the vicinity of an effector's trajectory. Darker shades (higher densities) correspond to more frequently visited phase space states, whereas brighter shades (lower densities) indicate rarely and shortly visited states of the phase space. The attractors of a dynamical system (such as fixed points or limit cycles which classify the governing regime) are structures in phase space towards which every trajectory ultimately evolves to, regardless of its starting point (the so-called initial conditions). It is plausible to assume that the phase space density (number of visiting trajectories) is much higher close to an attractor than far away from it. Thus, the topology of a region of high density can be exploited to identify the topology of the underlying dynamical regime. A localized region of high density (a dark point-like structure) suggests the presence of a fixed point. In contrast, dark ring-like structures suggest the presence of a periodic attractor (e.g., a limit cycle). By inspection of Figure 11, we see that evidence for localized structures indicating the presence of a fixed point can be found only at the slower rates (e.g., [ta] and [ka], speaker G2, 30–210/300 bpm). At faster rates, these structures disappear and dark ring-like structures of high density emerge indicating the presence of (non-fixed point) periodic attractors (e.g., [ta], speaker E2, 300–570 bpm). Furthermore, the same qualitative picture as shown in Figure 11 holds across all our speakers. Hence, there are promising indications for the existence of two distinct control regimes governing speech movements as a function of metronome rate. We plan to purse this evidence further in future work.

Overall, then, the suggestion is that the control regimes for speech may not be limited to target-to-target concatenation. In other words, fixed point dynamics as proposed by the standard model of speech gestures in Equation (3) may not be the only model available to the motor system in effecting speech action. Pursuing this suggestion would require rigorous assessment of candidate dynamical systems that can encompass more than a single control regime. Models with this property exist (e.g., Schöner, 1990; Jirsa and Kelso, 2005) but have not been explored in speech yet.

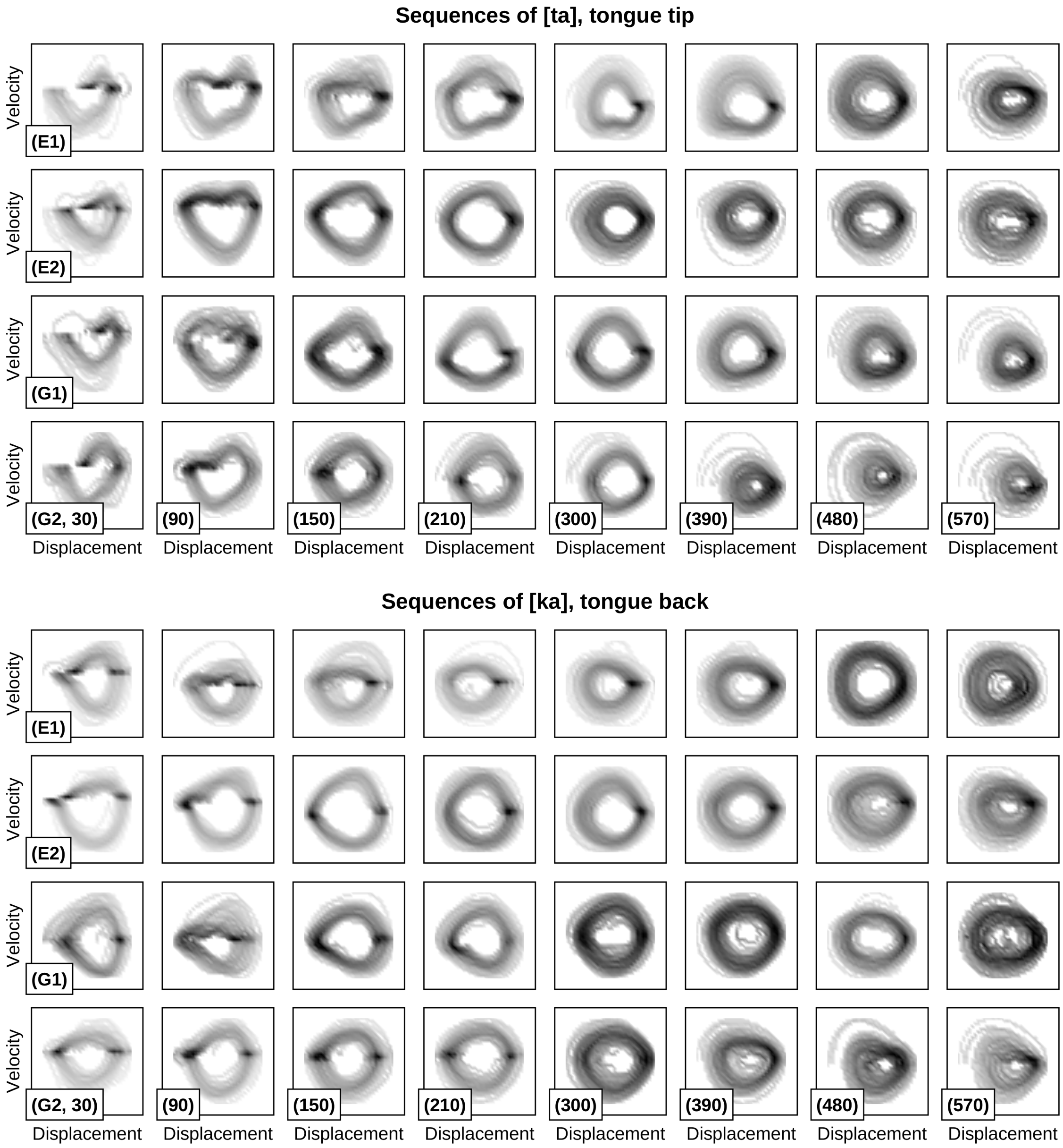

Figure 11. Phase portraits of tongue tip (top panel) and tongue back movements (bottom panel) from four speakers (E1, E2 of English and G1, G2 of German) at different metronome rates (columns). Darker shades indicate more visited states in the phase space. A localized region of frequent visits (dark point-like structures) hints at the presence of a fixed point.

5. Conclusion

The present article examined the speed-curvature power law in speech, using an experimental paradigm explicitly designed to elicit a wide range of speeds, from extremely slow to extremely fast. In this paradigm, we tracked the trajectories of sensors attached to the tongue tip and tongue back, the major lingual effectors involved in the production of the syllables [ta] and [ka] respectively. Analysis of our data in their full three-dimensions and using advanced numerical differentiation methods offers stronger evidence, as judged by correlation strengths, for the speed-curvature power law than that reported in previous studies devoted to its assessment. Furthermore, we demonstrate the presence of a clear speech rate dependency of the power law's instantiation in terms of its parameter values. Specifically, the often-sought or reported exponent of one third in the statement of the law is unique to a subclass of movements which in the case of speech correspond to just a small range of rates under which a particular utterance is produced. Comparison of our results to findings from other areas of motor control indicates that there is potential for lawful relations between rate and the geometric shape of speech movements to be revealed. It remains to be seen whether a thorough analysis of the movements' geometric properties in the domain of speech supports that indication and what lessons such analysis may offer for speech and non-speech biological action. Finally, our results and specifically the rate dependencies of the power law's exponent values hint at the possibility of a multiplicity or at least non-uniqueness in the control regimes controlling speech movements. We plan to pursue both of these indications in future work.

References

Baer, Matthias (2018). *A Python package for finite difference derivatives in any number of dimensions [Computer software]*. Retrieved from Python Package Index, https://pypi.org/project/findiff/.

Browman, Catherine P. and Louis M. Goldstein (1986). "Towards an articulatory phonology". *Phonology* 3(1), pp. 219–252. doi: 10.1017/S0952675700000658.

Clark, Mary Ann, Alma S. Merians, Abhay Kothari, Howard Poizner, Beth Macauley, Leslie J. Gonzalez Rothi, and Kenneth M. Heilman (1994). "Spatial Planning deficits in limb apraxia". *Brain* 117(5), pp. 1093–1106. doi: 10.1093/brain/117.5.1093.

Dayan, Eran, Antonino Casile, Nava Levit-Binnun, Martin A. Giese, Talma Hendler, and Tamar Flash (2007). "Neural representations of kinematic laws of motion: evidence for action-perception coupling". *Proceedings of the National Academy of Sciences of the United States of America* 104(51), pp. 20582–20587. doi: 10.1073/pnas.0710033104.

de'Sperati, Claudio and Paolo Viviani (1997). "The relationship between curvature and velocity in two-dimensional smooth pursuit eye movements". *The Journal of Neuroscience* 17(10), pp. 3932–3945.

Do Carmo, Manfredo Perdigão (1976). *Differential geometry of curves and surfaces*. New Jersey, United States: Prentice-Hall. isbn: 0132125897.

Fowler, Carol A., Philip Rubin, Robert E. Remez, and Michael T. Turvey (1980). "Implications for speech production of a general theory of action". *Language Production: Speech and Talk*. Ed. by Brian Butterworth. New York, United States: Academic Press, pp. 373–420.

Fuchs, Susanne, Pascal Perrier, and Mariam Hartinger (2011). "A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model". *Journal of Speech, Language, and Hearing Research* 54(4), pp. 1067–1076. doi: 10.1044/1092-4388(2010/10-0131).

Gribble, Paul L. and David J. Ostry (1996). "Origins of the power law relation between movement velocity and curvature: modeling the effects of muscle mechanics and limb dynamics". *Journal of Neurophysiology* 76(5), pp. 2853–2860. doi: 10.1152/jn.1996.76.5.2853.

Guenther, Frank H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production". *Psychological Review* 102(3), pp. 594–621. doi: 10.1037//0033-295X.102.3.594.

Hatsopoulos, Nicholas G. and William H. Warren Jr. (1996). "Resonance tuning in rhythmic arm movements". *Journal of Motor Behavior* 28(1), pp. 3–14. doi: 10.1080/00222895.1996.9941728.

Heglund, Norman C., C. Richard Taylor, and Thomas A. McMahon (1974). "Scaling stride frequency and gait to animal size: mice to horses". *Science* 186(4169), pp. 1112–1113. doi: 10.1126/science.186.4169.1112.

Hicheur, Halim, Stéphane Vieilledent, Magnus J. E. Richardson, Tamar Flash, and Alain Berthoz (2005). "Velocity and curvature in human locomotion along complex curved paths: a comparison with hand movements". *Experimental Brain Research* 162(2), pp. 145–154. doi: 10.1007/s00221-004-2122-8.

Hoole, Philip, Christine Mooshammer, and Hans G. Tillmann (1994). "Kinematic analysis of vowel production in German". *The 3rd International Conference on Spoken Language Processing* (Yokohama, Japan), pp. 53–56.

Horn, Berthold K. P. (1987). "Closed-form solution of absolute orientation using unit quaternions". *Journal of the Optical Society of America A* 4(4), pp. 629–642. doi: 10.1364/JOSAA.4.000629.

Huh, Dongsung and Terrence J. Sejnowski (2015). "Spectrum of power laws for curved hand movements". *Proceedings of the National Academy of Sciences of the United States of America* 112(29), E3950–E3958. doi: 10.1073/pnas.1510208112.

Huys, Raoul, Breanna E. Studenka, Nicole L. Rheaume, Howard N. Zelaznik, and Viktor K. Jirsa (2008). "Distinct timing mechanisms produce discrete and continuous movements". *PLoS Computational Biology* 4(4). doi: 10.1371/journal.pcbi.1000061.

Jirsa, Viktor K. and J. A. Scott Kelso (2005). "The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation". *Journal of Motor Behavior* 37(1), pp. 35–51. doi: 10.3200/JMBR.37.1.35-51.

Kelso, J. A. Scott, Betty Tuller, and Katherine S. Harris (1983). "A "dynamic pattern" perspective on the control and coordination of movement". *The Production of Speech*. Ed. by Peter F. MacNeilage. New York, United States: Springer, pp. 137–173. doi: 10.1007/978-1-4613-8202-7_7.

Kelso, J. A. Scott, Eric Vatikiotis-Bateson, Elliot L. Saltzman, and Bruce Kay (1985). "A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling". *The Journal of the Acoustical Society of America* 77(1), pp. 266–280. doi: 10.1121/1.392268.

Kier, William M. and Kathleen K. Smith (1985). "Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats". *Zoological Journal of the Linnean Society* 83(4), pp. 307–324. doi: 10.1111/j.1096-3642.1985.tb01178.x.

Lacquaniti, Francesco, Carlo Terzuolo, and Paolo Viviani (1983). "The law relating the kinematic and figural aspects of drawing movements". *Acta Psychologica* 54(1–3), pp. 115–130. doi: 10.1016/0001-6918(83)90027-6.

Levine, William S., Caroline Essex Torcaso, and Maureen Stone (2005). "Controlling the Shape of a Muscular Hydrostat: A Tongue or Tentacle". *New Directions and Applications in Control Theory*. Ed. by Wijesuriya P. Dayawansa, Anders Lindquist, and Yishao Zhou. Vol. 321. Lecture Notes in Control and Information Sciences. Springer, Berlin, Heidelberg. Chap. 13, pp. 207–222.

Maoz, Uri, Elon Portugaly, Tamar Flash, and Yair Weiss (2006). "Noise and the two-thirds power law". *Advances in Neural Information Processing Systems 18 NIPS* (Vancouver, Canada). Ed. by Yair Weiss, Bernhard Schölkopf, and John C. Platt, pp. 851–858.

Marken, Richard S. and Dennis M. Shaffer (2017). "The power law of movement: an example of a behavioral illusion". *Experimental Brain Research* 235(6), pp. 1835–1842. doi: 10.1007/s00221-017-4939-y.

Munhall, Kevin G., David J. Ostry, and Avraham Parush (1985). "Characteristics of velocity profiles of speech movements". *Journal of Experimental Psychology: Human Perception and Performance* 11(4), pp. 457–474. doi: 10.1037//0096-1523.11.4.457.

Nelson, Winston L. (1983). "Physical principles for economies of skilled movements". *Biological Cybernetics* 46(2), pp. 135–147. doi: 10.1007/BF00339982.

Nelson, Winston L., Joseph S. Perkell, and John R. Westbury (1984). "Mandible movements during increasingly rapid articulations of single syllables: preliminary observations". *The Journal of the Acoustical Society of America* 75(3), pp. 945–951. doi: 10.1121/1.390559.

Neufeld, Chris and Pascal van Lieshout (2014). "Tongue kinematics in palate relative coordinate spaces for electro-magnetic articulography". *The Journal of the Acoustical Society of America* 135(1), pp. 352–361. doi: 10.1121/1.4836515.

Ostry, David J., James D. Cooke, and Kevin G. Munhall (1987). "Velocity curves of human arm and speech movements". *Experimental Brain Research* 68(1), pp. 37–46. doi: 10.1007/BF00255232.

Ostry, David J., Eric Keller, and Avraham Parush (1983). "Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movement in speech". *Journal of Experimental Psychology: Human Perception and Performance* 9(4), pp. 622–636. doi: 10.1037//0096-1523.9.4.622.

Ostry, David J. and Kevin G. Munhall (1985). "Control of rate and duration of speech movements". *The Journal of the Acoustical Society of America* 77(2), pp. 640–648. doi: 10.1121/1.391882.

Patel, Aniruddh D., Anders Löfqvist, and Walter Naito (1999). "The acoustics and kinematics of regularly timed speech: a database and method for the study of the p-center problem". *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco, United States). Vol. 1, pp. 405–408.

Perrier, Pascal and Susanne Fuchs (2008). "Speed-curvature relations in speech production challenge the 1/3 power law". *Journal of Neurophysiology* 100(3), pp. 1171–1183. doi: 10.1152/jn.01116.2007.

Poizner, Howard, Mary Ann Clark, Alma S. Merians, Beth Macauley, Leslie J. Gonzalez Rothi, and Kenneth M. Heilman (1995). "Joint coordination deficits in limb apraxia". *Brain* 118(1), pp. 227–242. doi: 10.1093/brain/118.1.227.

Poizner, Howard, Linda Mack, Mieka Verfaellie, Leslie J. Gonzales Rothi, and Kenneth M. Heilman (1990). "Three-dimensional computergraphic analysis of apraxia". *Brain* 113(1), pp. 85–101. doi: 10.1093/brain/113.1.85.

Pollick, Frank E. and Gensel Ishimura (1996). "The three-dimensional curvature of straight-ahead movements". *Journal of Motor Behavior* 28(3), pp. 271–279. doi: 10.1080/00222895.1996.9941751.

Ramsay, James and Bernhard W. Silverman (2005). *Functional data analysis*. Springer Series in Statistics. New York, United States: Springer-Verlag. doi: 10.1007/b98888.

Rothi, Leslie J. Gonzalez and Kenneth M. Heilman (1994). *Apraxia. The neuropsychology of action*. Psychology Press.

Ruppert, David (2011). *Statistics and data analysis for financial engineering*. Springer Texts in Statistics. New York, United States: Springer-Verlag. doi: 10.1007/978-1-4419-7787-8.

Saltzman, Elliot L. and Kevin G. Munhall (1989). "A dynamical approach to gestural patterning in speech production". *Ecological Psychology* 1(4), pp. 333–382. doi: 10.1207/s15326969eco0104_2.

Schaal, Stefan and Dagmar Sternad (2001). "Origins and violations of the 2/3 power law in rhythmic three-dimensional arm movements". *Experimental Brain Research* 136(1), pp. 60–72. doi: 10.1007/s002210000505.

Schöner, Gregor (1990). "A dynamic theory of coordination of discrete movement". *Biological Cybernetics* 63(4), pp. 257–270. doi: 10.1007/BF00203449.

Schwartz, Andrew B. (1994). "Direct cortical representation of drawing". *Science* 265(5171), pp. 540–542. doi: 10.1126/science.8036499.

Sorensen, Tanner and Adamantios I. Gafos (2016). "The gesture as an autonomous nonlinear dynamical system". *Ecological Psychology* 28(4), pp. 188–215. doi: 10.1080/10407413.2016.1230368.

Stone, Meghan, A. Faber, L. J. Raphael, and T. H. Shawker (1992). "Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]". *Journal of Phonetics* 20(2), pp. 253–270.

Tasko, Stephen M. and John R. Westbury (2004). "Speed-curvature relations for speech-related articulatory movement". *Journal of Phonetics* 32(1), pp. 65–80. doi: 10.1016/S0095-4470(03)00006-8.

Tomaschek, Fabian, Dennis Arnold, Franziska Bröker, and R. Harald Baayen (2018). "Lexical frequency co-determines the speed-curvature relation in articulation". *Journal of Phonetics*, pp. 103–116. doi: 10.1016/j.wocn.2018.02.003.

Turvey, Michael T. (2007). "Action and perception at the level of synergies". *Human Movement Science* 26(4), pp. 657–697. doi: 10.1016/j.humov.2007.04.002.

Vatikiotis-Bateson, Eric and J. A. Scott Kelso (1990). "Linguistic structure and articulatory dynamics: a cross language study". *Haskins Laboratories Status Report in Speech Research* SR-103/104, pp. 67–94.

Vatikiotis-Bateson, Eric and J. A. Scott Kelso (1993). "Rhythm type and articulatory dynamics in English, French and Japanese". *Journal of Phonetics* 21, pp. 231–265.

Viviani, Paolo (2002). "Motor competence in the perception of dynamic events: a tutorial". *Common mechanisms in perception and action: attention and performance XIX*. Ed. by Wolfgang Prinz and Bernhard Hommel. Oxford, United Kingdom: Oxford University Press, pp. 406–442.

Viviani, Paolo and Marco Cenzato (1985). "Segmentation and coupling in complex movements". *Journal of Experimental Psychology: Human Perception and Performance* 11(6), pp. 828–845. doi: 10.1037/0096-1523.11.6.828.

Viviani, Paolo and Tamar Flash (1995). "Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning". *Journal of Experimental Psychology: Human Perception and Performance* 21(1), pp. 32–53. doi: 10.1037/0096-1523.21.1.32.

Viviani, Paolo and Natale Stucchi (1992). "Biological movements look uniform: evidence of motor-perceptual interactions". *Journal of Experimental Psychology: Human Perception and Performance* 18(3), pp. 603–623. doi: 10.1037/0096-1523.18.3.603.

Viviani, Paolo and Carlo Terzuolo (1982). "Trajectory determines movement dynamics". *Neuroscience* 7(2), pp. 431–437. doi: 10.1016/0306-4522(82)90277-9.

Wann, John, Ian Nimmo-Smith, and Alan M. Wing (1988). "Relation between velocity and curvature in movement: equivalence and divergence between a power law and a minimum-jerk model". *Journal of Experimental Psychology: Human Perception and Performance* 14(4), pp. 622–637. doi: 10.1037//0096-1523.14.4.622.

Zago, Myrka, Adam Matic, Tamar Flash, Alex Gomez-Marin, and Francesco Lacquaniti (2017). "The speed-curvature power law of movements: a reappraisal". *Experimental Brain Research* 236(1). doi: 10.1007/s00221-017-5108-z.

Research Article II

Fitts' law in tongue movements of repetitive speech

Phonetica: International Journal of Phonetic Science, Karger Publishers
Submitted: January 27, 2019; *Revise:* March 13, 2019

Abstract

Fitts' law, perhaps the most celebrated law of human motor control, expresses a relation between the kinematic property of speed and the non-kinematic, task-specific property of accuracy. We aimed to assess whether speech movements obey this law using a metronome-driven speech elicitation paradigm with a systematic speech rate control. Specifically, using the paradigm of repetitive speech, we recorded via electromagnetic articulometry (EMA) speech movement data in sequences of the form /CV.../ from six adult speakers. These sequences were spoken at eight distinct rates ranging from extremely slow to extremely fast. Our results demonstrate, first, that the present paradigm of extensive metronome-driven manipulations satisfies the crucial prerequisites for evaluating Fitts' law. Second, we uncover for the first time in speech evidence for Fitts' law at the faster rates and specifically beyond a participant-specific critical rate. Crucially, however, we find no evidence for Fitts' law at the slowest metronome rates. Finally, we discuss implications of these results for models of speech.

1. Introduction

Speech is perhaps "the most highly developed motor skill possessed by all of us" (Kelso, Tuller, et al., 1983, p. 137). The continuous deformations of the vocal tract structuring the sound of speech involve the precise positioning of a number of articulatory organs as they form and release constrictions in a limited space inside the body. Speech has evolved to harness this complex activity for the purposes of communication. A remarkable fact about the robustness of the resulting system is that what a linguist considers to be the same utterance can be conveyed by different individuals under wildly different conditions. For example, age, gender, size, loudness, and speed all contribute to the formation of the speech sounds which are then recovered as an exemplar of e.g., [ta] or [ka]. Given the remarkable variability of conditions under which speech goals are achieved, the identification of invariances (at best, laws) in kinematic characteristics of speech movements has been seen as an imperative (Munhall, Ostry, et al., 1985; Turvey, 2007). The identification of such invariances offers potentially crucial information for model evaluation. Any proposed model for speech must conform to such invariances. Furthermore, if there are invariances found in some areas of motor control but not in speech, this in turn informs the field of motor control in general in that it points to specificities of functional organization with respect to different types of movements and/or effectors. Despite the relatively early influence of concepts and in some cases models from general motor control has had on models of speech (Fowler, Rubin, et al., 1980; Browman and Goldstein, 1986; Saltzman and Munhall, 1989; Guenther, 1995), our understanding of the extent to which speech movements conform to well-known laws from other areas of human movement is at its infancy (Nelson, 1983; Nelson, Perkell, et al., 1984; Ostry, Cooke, et al., 1987).

Perhaps the most celebrated law of human motor control is Fitts' law (Fitts, 1954). This law expresses a relation between the kinematic property of movement speed and the non-kinematic, task-specific property of accuracy. In all its simplicity, this relation reads

$$T = a + b\,ID, \tag{1}$$

where movement duration T (a measure of speed) is a linear function of a task-specific index of difficulty ID, a quantity defined by the ratio of amplitude A (a measure of the excursion of some effector to reach a target) to width W (a measure of the target's size). More specifically, in its Shannon formulation (cf. I. S. MacKenzie, 1992, 2013), *ID* is defined by

$$ID = \log_2\left(\frac{A}{W} + 1\right). \tag{2}$$

The presence of the relation expressed by Fitts' law has been reported for a multiplicity of effector systems engaged in a variety of movement types (see Plamondon and Alimi, 1997; Schmidt and Lee, 2011 for extensive overviews).

The law is sometimes described as a trade-off between speed and accuracy of movement, reflecting the observation that the accuracy of spatially constrained, target-directed movements diminishes when speed becomes excessive. The study of this observation dates as back as the work of Woodworth (1899) on hand movements of line drawing tasks. However, it was Fitts (1954) and Fitts and Peterson (1964) who consolidated results with the two, now famous, stylus-tapping experiments, one using a reciprocal tapping protocol and another investigating discrete tapping movements.

Three primary considerations motivate seeking evidence for Fitts' law in speech. First, the law is effector-independent. Laws that have this property are good candidates for disclosing the abstractness of the principles that underwrite performance in some domain and potentially also the nature of these principles. Second, Fitts' law is the only law that expresses a relational invariance among kinematic (duration, amplitude) and non-kinematic (width) variables. All other relations so far studied in speech, such as that between peak velocity and amplitude or that between the ratio of peak velocity over amplitude as a function of duration, hold over kinematic-only variables. More specifically, the parameter W which enters into the expression of the law is a task property. Leading theoretical perspectives point to the thesis that speech goals are defined in task dimensions rather than individual effector dimensions (see Saltzman and Munhall, 1989 and Guenther, 1995 with important antecedents in understanding of coordination and control of action found in the work of Bernstein, 1967 and Turvey, 1977). Because Fitts' law expresses an abstract relation that involves both kinematic and task space coordinates, it captures a relational invariance which models of speech should be able to account for (assuming the law holds for speech).

Last but not least, Fitts' law has been shown to hold also in the perception of action. Grosjean, Shiffrar, et al. (2007) asked subjects to judge whether the movement times in a motion display of an arm moving between two targets (of specified width and amplitude or distance from one another) would be possible without missing the targets. The times reported by the participants as being possible were precisely those times that are predicted by Fitts' law.

We aimed to assess whether speech movements conform to this law using a metronome-driven speech elicitation paradigm. Specifically, using the paradigm of repetitive speech (cf. Kelso, Vatikiotis-Bateson, et al., 1985; Ostry, Cooke, et al., 1987; Patel, Löfqvist, et al., 1999), we recorded via electromagnetic articulometry (EMA) speech movement data in sequences of the form /CV. . . / from six adult speakers (five were native speakers of German and one was a native speaker of American English). These sequences were spoken at eight distinct rates ranging from extremely slow (30 beats per minute, bpm) to extremely fast (570 bpm). For comparison purposes, Kelso, Vatikiotis-Bateson, et al. (1985) included two rates as did Ostry, Cooke, et al. (1987) and Patel, Löfqvist, et al. (1999) used a metronome to suggest a rate to their participants (which was 120 bpm) but there was no metronome during the actual registration of a participant's utterances. In the resulting dataset, we sought evidence of the sort that has provided support for Fitts' law in other areas of motor control. This is a non-trivial undertaking because, unlike in other movement domains, direct control over the quantities A and W is infeasible in speech. Any paradigm aiming to assess the law in speech must fulfill certain prerequisites that ensure its compatibility to the original Fitts paradigm. In particular, as noted by Plamondon and Alimi (1997, p. 280), among others, the following two prerequisites must be met in order to (potentially) reveal a trade-off relation between speed and accuracy as found by Fitts. Firstly, movement amplitude A and target size W must be demonstrated to have varying values across different experimental conditions (experimental control). Secondly, the movements must be performed under time pressure (rapidness of movement). Specifically, Fitts instructed his subjects "to work at his maximum rate" (Fitts, 1954, p. 383).

A first attempt to assess the presence of Fitts' law in speech was made by Lammert, Shadle, et al. (2018). The data were real-time magnetic resonance imaging (rtMRI) recordings of five male and five female American English speakers from a reading task of the USC-TIMIT database. Amplitudes of movements and extents of articulatory targets were operationally-defined as elements of an approx. 50-dimensional vector space. Reported correlation strengths r^2 evaluating linearity between time and index of difficulty (as encoded in Fitts' law) were in the range of 0.03 to 0.52. The study reported methodological challenges in defining and measuring the Fitts' key variables in high dimensional real-time MRI data. Recall the two prerequisites for evaluating Fitts' law in any domain. The first requires that movement amplitude A and target size W have varying values across different experimental conditions. There is no explicit information on this for the Lammert, Shadle, et al. (2018) datasets. It is conceivable that the low r^2 values reported may have been due to insufficient coverage of the A-W space. The second prerequisite for evaluating Fitts' law is the requirement of temporal pressure. No explicit information about the speech rate of the rtMRI dataset was given. Presumably it was a moderate rate of common reading tasks (around 150–250 bpm). Overall, then, whereas this study is a first attempt at assessing the law in speech, challenges arising from mainly methodological limitations, as acknowledged by the authors, remain and call for additional studies (e.g., "Among these challenges are higher frame rate data, and exploring additional definitions of the key relevant quantities"; Lammert, Shadle, et al., 2018, p. 21).

In the present work, we pursued a metronome-driven paradigm which, we demonstrate, enabled us to successfully manipulate the variables of movement amplitude and target size essential to any Fitts-style analysis. Specifically with respect to the notion of target, our implementation utilizes a descriptive three-dimensional spatial articulatory target. Furthermore, the design of our paradigm includes the important aspect of temporal pressure.

We uncover for the first time evidence for Fitts' law at the faster rates and specifically beyond a participant-specific critical rate. Crucially, however, we find no evidence for Fitts' law at the slowest metronome rates. We discuss implications of these results for models of speech.

2. Methods

Five native speakers of German and one native speaker of American English (three females and three males in total) participated in the experiment. Data from another British English speaker was registered but had to be excluded due to an unnoticed hardware equipment failure while recording. The speakers were between 22 and 35 years old and without any present or past speech or hearing problems. They were recruited at the University of Potsdam and paid for their participation in the experiment. All procedures were performed in compliance with relevant laws and institutional guidelines and were approved by the ethics committee of the University of Potsdam. Written informed consent was obtained from all subjects.

During the experiment all participants were prompted on a computer screen to produce sequences of repeated [ta] or [ka] syllables in time with an audible metronome. The metronome served as an extrinsic index of the intended rate of syllable production. We did not require participants to aim for synchronizing any specific point of the sequence [ta] or [ka] with the metronome. As we demonstrate below, this procedure was adequate to induce sufficient scaling of kinematic quantities to the extent that makes assessment of the law we aim to assess feasible. The participants were instructed to articulate their responses accurately and naturally. The rate of the metronome was set to the values of 30, 90, 150, 210, 300, 390, 480 and 570 bpm (corresponding to 0.5, 1.5, 2.5, 3.5, 5.0, 6.5, 8.0 and 9.5 Hz). At the start of each trial, the participant was exposed to the metronome stimulus and begun articulating the required response syllable at a point of their choice. Starting with the slowest rate, a minimum of four trials at each rate were recorded. Once this minimum was reached, recording proceeded with the next higher rate. The duration of each trial (hence duration of the metronome stimulus) was timed such that the participant was able to adjust to the beat of the metronome and produce a coherent sequence of approx. 30 syllables. The entire procedure was performed in two successive blocks, first for sequences of [ta] and then for sequences of [ka].

Articulatory data as well as acoustic data were registered from all participants. All recordings took place in our sound-attenuated booth using a Carstens AG501 3D Electromagnetic Articulograph for articulatory, and a YOGA Shotgun microphone EM-9600 attached to a TASCAM US-2x2 Audio interface for acoustic data registration. Three-dimensional electromagnetic articulography (EMA) allowed measurement of kinematic displacement data of selected articulators at a high precision. Along with some other auxiliary reference locations (upper and lower incisors, nose bridge, left and right mastoids), we tracked the positions of sensors attached to the tongue tip and tongue back articulators, the major effectors involved in the production of [ta] and [ka] respectively.

2.1 Data processing

Three-dimensional displacement data, provided by the AG501 device, was digitized at a sampling rate of 1250 Hz. In order to reduce storage and memory footprint as well as to improve further data processing performance, the sampling rate of all signals was decreased to a value of 104.167 Hz (a twelfth of 1250 Hz). To avoid aliasing effects the decimation procedure implied an initial lowpass filtering using an eighth order Chebyshev type I filter with cutoff frequency of 46.875 Hz which also eliminated most high frequency noise. Based on these decimated signals, spatial transformations of head movement correction and occlusal reference frame alignment were determined and applied by means of the method proposed by Horn (1987). Finally, a zero-delay Chebyshev type II lowpass filter with cutoff frequency of 25 Hz and stop-band attenuation of 80 dB was utilized to eliminate any further noise potentially present.

The continuous motion of the tongue back and tongue tip articulators was segmented into separate, successive closing movements. The basis for this segmentation was the first derivative (velocity) of the displacement's principal component (PCA, representing displacement along movement direction). As an example of our data, Figure 1 shows a series of [ka] syllables produced by one of our participants at a metronome rate of 150 bpm. Instants of zero-crossings in the PCA velocity were used as movement delimiters (see, e.g., Munhall, Ostry, et al., 1985). In total, we registered 4 314 movements in the [ta] case and 3 991 movements in the [ka] case.

3. Results

In assessing Fitts' law in speech, it is imperative to demonstrate that our experimental design conforms with Fitts' original design. In his classic experiments with a reciprocal tapping apparatus, Fitts' participants had to strike alternately the center of each of two target plates of width W using a metal stylus (Fitts, 1954; Fitts and Peterson, 1964). The quantities of movement amplitude A, corresponding to the distance between the two plates, and target size W were under

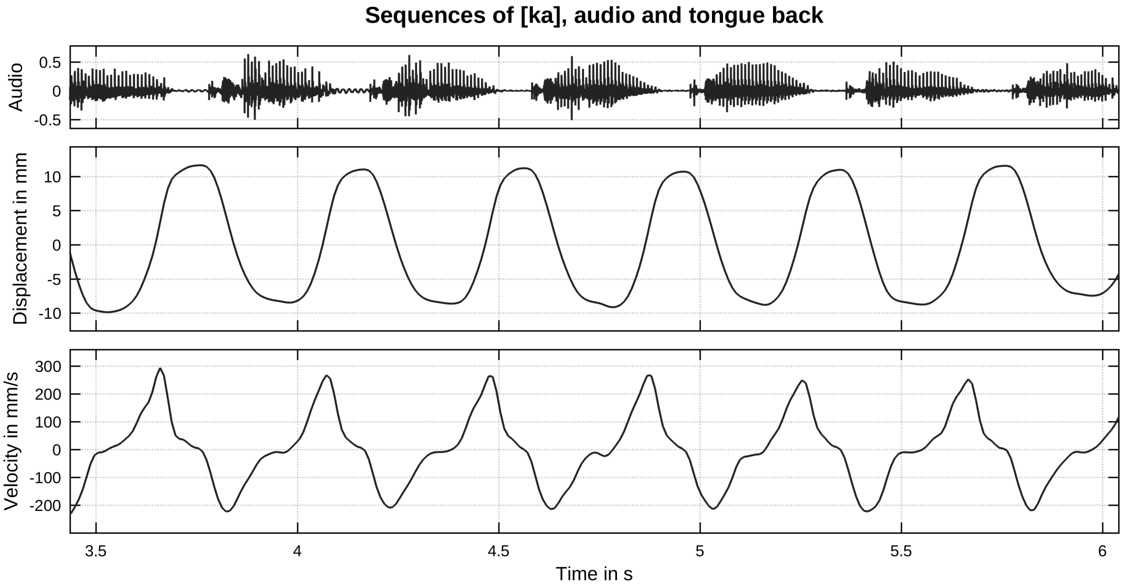

Figure 1. Section of a [ka] sequence at 150 bpm. Top: Acoustic recording. Middle: Principal component (PCA) of tongue back displacement. Bottom: First derivative (velocity) of PCA.

the direct control by the experimenter. These quantities were thus chosen to vary over a considerable range of values. Such variation is absolutely crucial to enable evaluation of the predicted linearity relating movement speed and index of difficulty $ID = \log_2(A/W + 1)$. In our domain, the stylus is the part of the tongue used for the formation of the consonant (tongue tip for [t], tongue body for [k]). However, A and W are not under our independent control. As one of the two crucial preconditions to be met in assessing Fitts' law (see the prerequisite referred to as "Experimental control" in the Introduction), it thus remains to be shown that these essential parameters visited a variegated set of values. In addition, because in our domain the notion of target size can only be determined *a posteriori*, it is essential to explicitly verify that our design resulted in sufficiently variegated ranges of W.

For plosive consonants like [t] and [k], it seems relatively uncontroversial that effectors such as the tongue tip and the tongue body form and release constrictions in characteristic regions of the vocal tract. One may thus operationalize a notion of target on the basis of spatial properties of these constrictions. However, unlike in Fitts' original design and as in many subsequent assessments of Fitts' law to other domains, the spatial dimensions of speech targets are not under direct control by the experimenter. In other words, there is no speech-task analogous to repetitively tapping a disk of some experimenter-specified diameter and systematically changing that diameter. For such cases, Welford (1968) proposed an *a posteriori* defined target size, derived from statistical properties of the data. This notion has been widely adopted in subsequent assessments in the Fitts' law literature. Specifically, for one-dimensional target extents the effective target width is defined by

$$W = \sqrt{2\pi \mathrm{e}}\,\sigma, \tag{3}$$

where σ is the common univariate standard deviation of movement endpoints. The scaling factor $\sqrt{2\pi e}$ corresponds to 96% of the standard normal distribution. In a three-dimensional extension, Wobbrock, Shinohara, et al. (2011) proposed the following replacement

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} \{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2\}}{N - 1}}, \tag{4}$$

where σ now denotes the trivariate deviation of N three-dimensional endpoints (x_i, y_i, z_i) from their centroid $(\bar{x}, \bar{y}, \bar{z})$. Spatial articulatory target widths W were computed in the way described above individually for each speaker and each metronome rate. For each movement, the amplitude A was determined as the three-dimensional Euclidean distance between its onset and offset endpoints.

Figure 2 shows scatter plots of the so-determined amplitudes A and effective target widths W for movements in sequences of [ta] and [ka]. It is evident that there is a variety of distinct values of target width W. Furthermore, for each of these there is a sufficient spread in the A quantity. This demonstrates that (in an indirect manner) our task successfully provided a sufficient range of amplitudes and target widths so as to make an assessment of Fitts' law feasible. We are not aware of any previous demonstration of this in speech.

3.1 *Presence of Fitts' law*

We now turn to assess the presence of Fitts' law on the registered speech movement data. Recall that evidence for Fitts' law would be demonstrated on the basis of a linear relation between movement speed (measured by duration T) and index

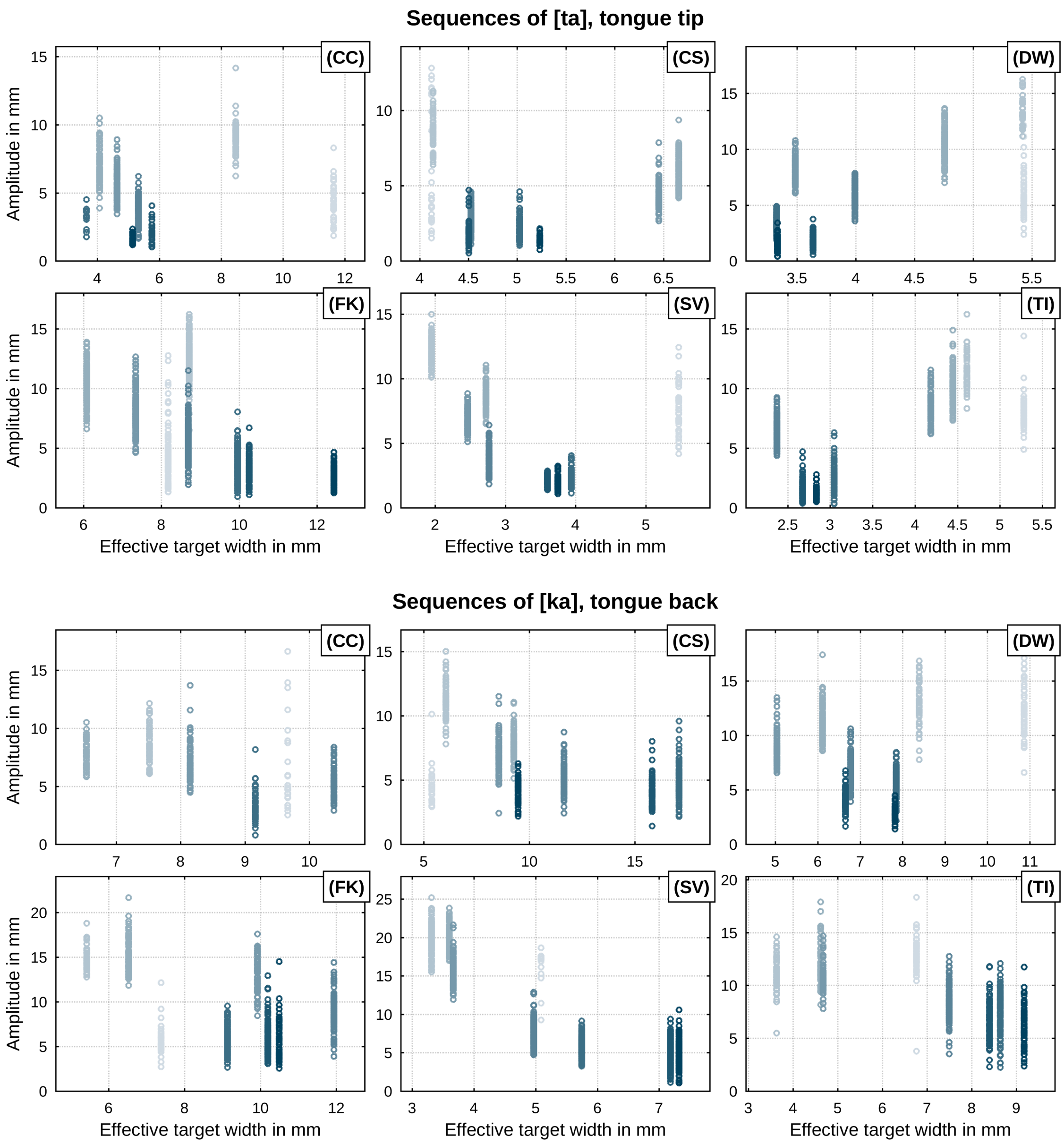

Figure 2. Range of movement amplitudes and effective target widths. Top: Sequences of [ta]. Bottom: Sequences of [ka]. Data are drawn separately for each speaker (sub-panels). Metronome rate is color-coded using shades from faint (slowest rate) to dark (fastest rate).

of difficulty $ID = \log_2(A/W + 1)$, as in

$$T = a + b\,ID, \tag{5}$$

where the constants a and b are empirically determined.

Figure 3 shows scatter plots of the two essential quantities of Fitts' law, duration T and index of difficulty ID. The drawn data are pooled across the entire range of metronome rates individually for each speaker. The first observation is that there is clearly no evidence for the law across the entirety of induced speech rates. That is, there is no obvious linearity across the whole range of data. Nevertheless, there appear to be identifiable regions of linearity as predicted by Fitts' law. These regions moreover do not seem to be random collections of data points across different conditions. Rather, they appear to be structured by metronome rate (recall that metronome rate is color-coded in the drawn data, with fainter shades for slower rates and darker shades for faster rates). Specifically, regions of linearity seem to hug the datapoints starting with the fastest rates and proceed downwards up to some slower rate where ultimately linearity degenerates or breaks down completely. In what follows, our aim is to identify these regions of linearity, hence, revealing evidence for the presence of Fitts' law in our speech data.

Recall that Fitts' paradigm concerned movements performed under temporal pressure; see the prerequisite referred to as "Rapidness of movement" in the Introduction. Fitts did not define the notion of temporal pressure. Instead, he instructed participants in ways that resulted in movements that were fast while still conforming to the demands of his tasks. For example, in his reciprocal tapping task where participants used a stylus to strike two plates of some specified width, the instruction was to "Score as many hits as you can" (Fitts, 1954, p. 384). In our task, not all sequences were produced under (the same) time pressure. In an extension to Fitts' dichotomous view (temporal pressure present or not), it seems reasonable to assume that temporal pressure in our task scales with increasing metronome rate. Conversely, this implies that as metronome rate slows down, there is a rate which may violate Fitts' paradigm (because of insufficient temporal pressure). Crucially, this means that once such a rate has been identified, no rate slower than that rate (with even less temporal pressure) satisfies

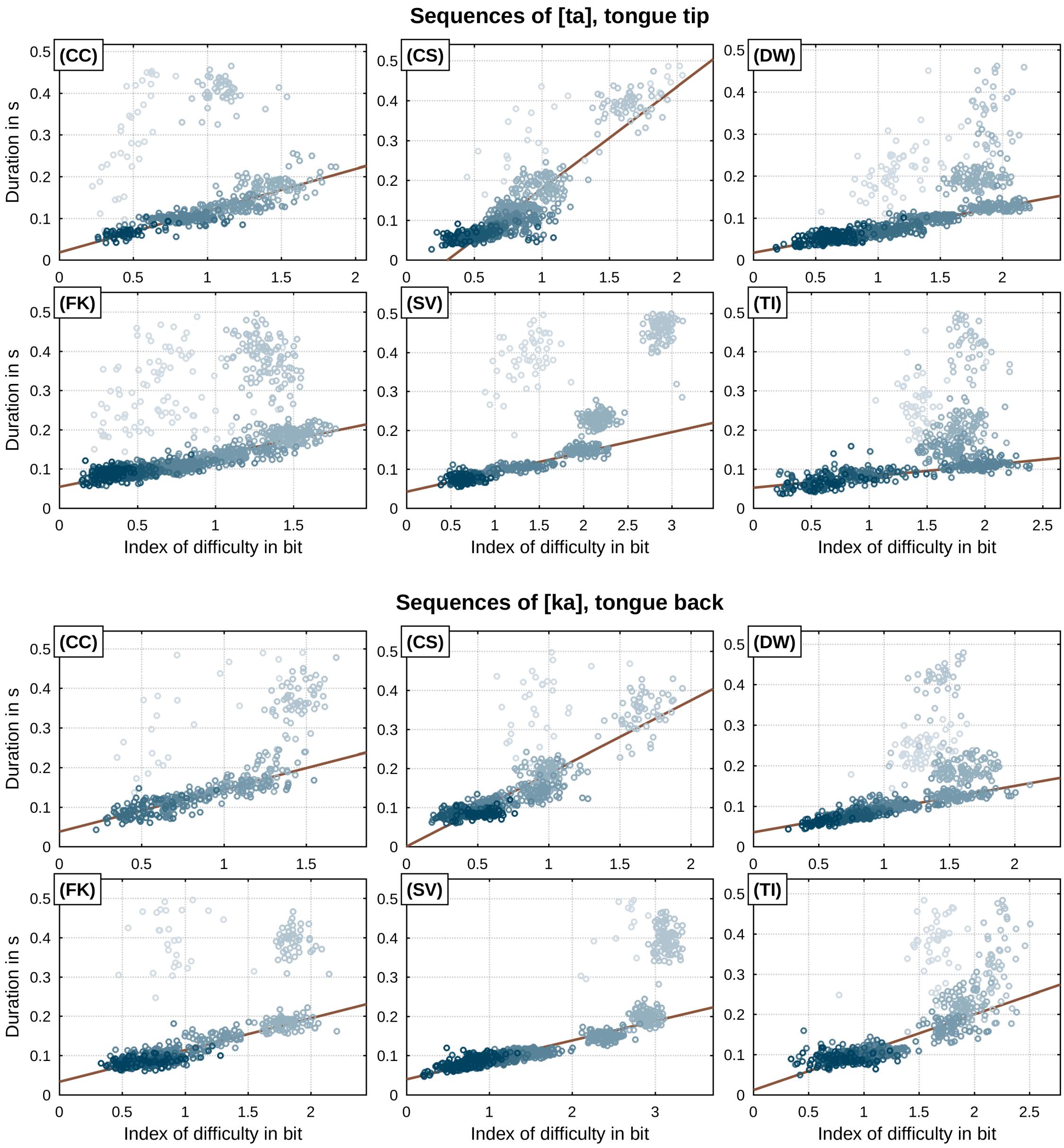

Figure 3. Relation between movement duration and index of difficulty. Top: Sequences of [ta]. Bottom: Sequences of [ka]. Data are drawn separately for each speaker (sub-panels). Metronome rate is color-coded using shades from faint (slowest rate) to dark (fastest rate). Linear regressions of contiguous Fitts-compliant rates are drawn as thick dashed lines. Linear regression lines are not meant to indicate fits to the entire dataset but only to a subset starting from a (speaker-specific) rate and including all higher rates. It is in these range of rates where our paradigm becomes compatible with the temporal pressure requirement in Fitts' paradigm. (See text for details.)

Fitts' paradigm. Hence, we can partition our rate continuum into a set of contiguous rates that conform to Fitts' paradigm and another set of contiguous rates that do not. This contiguity property reflects precisely the bias towards fast movements inherent to Fitts' paradigm, though in a gradual way as required in our task. In addition, it endorses a group-wise analysis of the available data. Data of individual, ungrouped rates rarely show a significant correlation as predicted by Fitts' law. This is so because considering any given rate by itself weakens considerably the required diversification of A and W.

Our aim thus is to identify the largest set of contiguous rates obeying Fitts' law such that any other larger set will show a lesser quality of linearity or not satisfy the preconditions of assessing linearity. Quality of linearity was judged using the classic metrics of correlation slopes and correlation strengths (in terms of Pearson's correlation coefficient r). For ease of presentation, we order metronome rates R_i ($i = 1 \ldots 8$) backwards, starting from the fastest rate $R_1 = 570$ bpm (highest temporal pressure) to the slowest rate $R_8 = 30$ bpm (lowest temporal pressure). Our procedure of determining the largest set of contiguous rates obeying linearity is as follows:

1. Construct the i-th set S_i of datapoints from contiguous rates, always starting with the fastest rate R_1 and proceeding to the slower rate R_i. Thus, whereas set S_1 consists of the datapoints from just set R_1, set S_2 includes those in S_1 plus the datapoints from R_2, set S_3 includes those in S_2 plus the datapoints from R_3, and so on. The larger the index i, the larger the constructed set, as more metronome rates are included. Compute the correlation strength r_i^2 and the correlation slope b_i (of the T-ID relation of the datapoints) for each S_i.

2. Determine the difference between the correlation slopes of set S_i and the next larger set S_{i+1}, which is the union of S_i and the datapoints from the next slower rate R_{i+1}. Slope differences are computed using a null hypothesis test for identical slopes (e.g., Cohen, 1983) resulting in the F-scores F_i. The higher the F-score F_i the more the slopes of S_i and S_{i+1} differ.

3. Consider any instance of increasing slope differences in the F-scores F_i as a function of index i. Such an increase $\Delta F_i = F_{i+1} - F_i$ indicates that, by in-

clusion of the next slower rate R_{i+1}, the correlation slope of the data rapidly changes (in the sense of an accelerated change given by the difference of differences ΔF_i) and thus quality of linearity decreases significantly.[1] Let each index i of increasing slope difference be a candidate to stop further inclusion of slower rates. For each such candidate index, there is a corresponding set S_i, rate R_i and correlation strength r_i^2. Among these candidates, choose the one which maximizes the correlation strength.[2] That chosen index identifies the sought maximal set of contiguous rates with the highest quality of linearity.

Table 1 shows values of slope differences F_i and correlation strengths r_i^2 obtained from our data by the above procedure. The top half of the table lists values for sequences of [ta] and the bottom half for sequences of [ka]. Each row corresponds to one of the constructed sets of rates S_i, starting from the smallest set $S_1 = \{570\,\text{bpm}\}$ proceeding to the largest set $S_8 = \{30 \ldots 570\,\text{bpm}\}$. Values of slope differences F_i in each row (except the bottom one for which there is no next row) were computed based on the two sets S_i (current row) and S_{i+1} (next row below). Large values of F-score in any given row of the table give a measure of the decrease in the quality of linearity that would be incurred if the next slower rate were to be added to the expanded set of rates. Increases of F-scores between successive rows ($\Delta F_i = F_{i+1} - F_i$) are taken to be candidates to stop further inclusion of slower rates. Out of these candidates of indicated increasing loss of quality of linearity, the set S_i which maximizes the value of correlation strength r_i^2 is chosen. This S_i is the sought largest set of contiguous rates.

---

[1] Consider a set S_i of slope b_i and a superset $S_{i+1} \supset S_i$ of slope b_{i+1} differing significantly from b_i. If the cardinality of the two sets is comparable, then set S_{i+1} will show (more) ill-shaped residuals in comparison to S_i. Hence, a linear model for S_{i+1} will not be of the same quality as for S_i.

[2] Note that this does not maximize correlation strength across the entirety of all possible sets S_i (that is, there might be another set with higher correlation strength than the determined one). Our approach is thus conservative as it does not solely optimize for high correlation strengths but also considers the quality of regression (by slope differences). High correlation strength is not necessarily associated with high quality of linearity, e.g., the distribution of residuals can differ dramatically for the same value of r^2 (Anscombe, 1973). An example from our domain is shown in Figure 4.

	CC		CS		DW		FK		SV		TI	
S_i	F_i	r_i^2	F_i	r_i^2	F_i	r_i^2	F_i	r_i^2	F_i	r_i^2	F_i	r_i^2
[ta] 570 bpm	0.17	0.53	0.40	0.07	0.48	0.31	3.09	0.17	0.47	0.00	0.09	0.26
[ta] ≥480 bpm	6.27	0.49	2.75	0.13	0.05	0.24	0.12	0.10	0.59	0.02	0.88	0.30
[ta] ≥390 bpm	8.51	0.52	17.59	0.24	26.90	0.48	2.61	0.13	17.92	0.06	3.74	0.31
[ta] ≥300 bpm	4.59	0.57	21.11	0.48	4.67	0.75	38.11	0.42	30.36	0.63	19.92	0.62
[ta] ≥210 bpm	45.26	0.72	57.66	0.50	98.14	0.87	133.67	0.70	176.85	0.88	29.32	0.54
[ta] ≥150 bpm	2.98	0.77	117.49	0.62	26.87	0.67	112.94	0.83	232.13	0.81	19.08	0.42
[ta] ≥90 bpm	9.32	0.18	0.39	0.84	0.15	0.52	4.29	0.44	0.90	0.79	0.01	0.28
[ta] ≥30 bpm	n/a	0.05	n/a	0.80	n/a	0.42	n/a	0.28	n/a	0.61	n/a	0.25
[ka] 570 bpm	n/a	n/a	0.44	0.09	0.59	0.40	1.18	0.18	0.04	0.45	0.86	0.00
[ka] ≥480 bpm	n/a	n/a	0.11	0.06	1.06	0.49	0.92	0.26	11.33	0.47	0.25	0.03
[ka] ≥390 bpm	0.60	0.33	15.46	0.07	1.69	0.47	11.39	0.18	0.23	0.43	7.28	0.05
[ka] ≥300 bpm	0.98	0.32	28.03	0.26	0.68	0.60	40.73	0.26	28.88	0.63	65.90	0.20
[ka] ≥210 bpm	11.77	0.68	63.55	0.63	113.18	0.77	1.70	0.61	134.37	0.84	70.31	0.67
[ka] ≥150 bpm	110.03	0.76	97.82	0.71	21.87	0.72	154.47	0.83	341.21	0.89	63.06	0.74
[ka] ≥90 bpm	1.00	0.67	2.81	0.86	0.89	0.36	0.96	0.64	1.38	0.70	6.03	0.72
[ka] ≥30 bpm	n/a	0.50	n/a	0.68	n/a	0.36	n/a	0.40	n/a	0.68	n/a	0.57

Table 1. Statistics of contiguous sets of rates per speaker. F-scores F_i of slope differences and correlation strength r_i^2. Index i increases with rows (separately for [ta] and [ka]). Per-speaker determined maximal sets of contiguous rates are indicated by shaded cells.

Let us walk through an example of how our procedure determines maximal sets of rates conforming to Fitts' law in our speech data. Consider the data from [ta] sequences by speaker CC. By visual inspection of Figure 3 (top panel, CC), there is clear evidence for a correlation between *ID* and *T* at metronome rates faster than the three slowest rates (faintest shades). F-scores F_i of slope differences increase by extending the rates of 570, 480, 390, 300, 210 bpm and 90 bpm (Table 1, bottom half, CC). Out of these candidate sets, the set with highest correlation strength is that with the slowest rate of 210 bpm (set S_5 with $r_5^2 = 0.72$). Hence, the determined slowest rate of the largest set of [ta] sequences by speaker CC is 210 bpm (indicated by a shaded cell in Table 1). Any slower rate included (potentially 150, 90 and 30 bpm) would reduce the quality of linearity of the sought maximal set of rates. Note that by inclusion of the next slower rate of 150 bpm correlation strength would attain a larger value of $r_6^2 = 0.77$. However, this gain would come at the cost of a lesser quality of linearity which can be seen in the residual plots in Figure 4 showing details of three consecutive sets considered in the determination of the maximal set of rates. These sets are S_4 (non-maximal set, [ta] ≥300 bpm), S_5 (determined maximal set, [ta] ≥210 bpm) and S_6 (rejected set, [ta] ≥150 bpm). Relations between movement duration and index of difficulty of these sets by way of linear regressions are drawn in the left-hand side of Figure 4 (S_4: dotted, S_5: solid and S_6: dashed line). The right-hand side of Figure 4 shows per-set detrended normal quantile-quantile plots of the corresponding regression residuals. It is evident that when set S_5 is expanded to the next larger set S_6 there is a clear loss of normality in the distribution of the regression residuals. This can be seen by the group of deviating (from the red horizontal dashed line) residuals in the right-hand side of the bottom right panel of Figure 4. Moreover, as can be inferred from the (faint) shades of these residuals, their deviation from normality is solely caused by the inclusion of the next slower rate of 150 bpm (recall that metronome rate is color-coded in the drawn data, with fainter shades for slower rates and darker shades for faster rates). In contrast, when set S_4 is expanded to the larger set of S_5, the residuals distribution is unaffected, perhaps even improves (as can be seen by considering Figure 4, right top vs. right middle panels). Normality of residuals is a crucial assumption of linear regression. A violation of this assumption strongly indicates the absence of a linear relation in the data. Hence, by its design, our method includes the rate of 210 bpm but excludes the rate of 150 bpm from the sought maximal set of rates and settles to S_5 as its output. In other words, we seek linearity but we do not impose linearity on our data.

Overall, then, our procedure derives the per-speaker largest set of contiguous rates exhibiting significance of the Fitts' law predicted linearity. Regressions of these determined sets of rates are drawn in Figure 3 and correspond well to the impressionistic view of where linearity resides in these datasets. Table 2 lists for every participant the slowest rate R, correlation strength r^2 and inverse correlation slope $1/b$ separately for [ta] and [ka]. Correlation strengths attain values in the range of 0.62 to 0.89, with all p-values below 0.0001, indicating very strong significance. The per-speaker slowest rates for which Fitts' law holds are in the range of 150–300 bpm, corresponding to slow to modest speech rates. These per-speaker slowest rates demarcate the lower end of the set of contiguous rates,

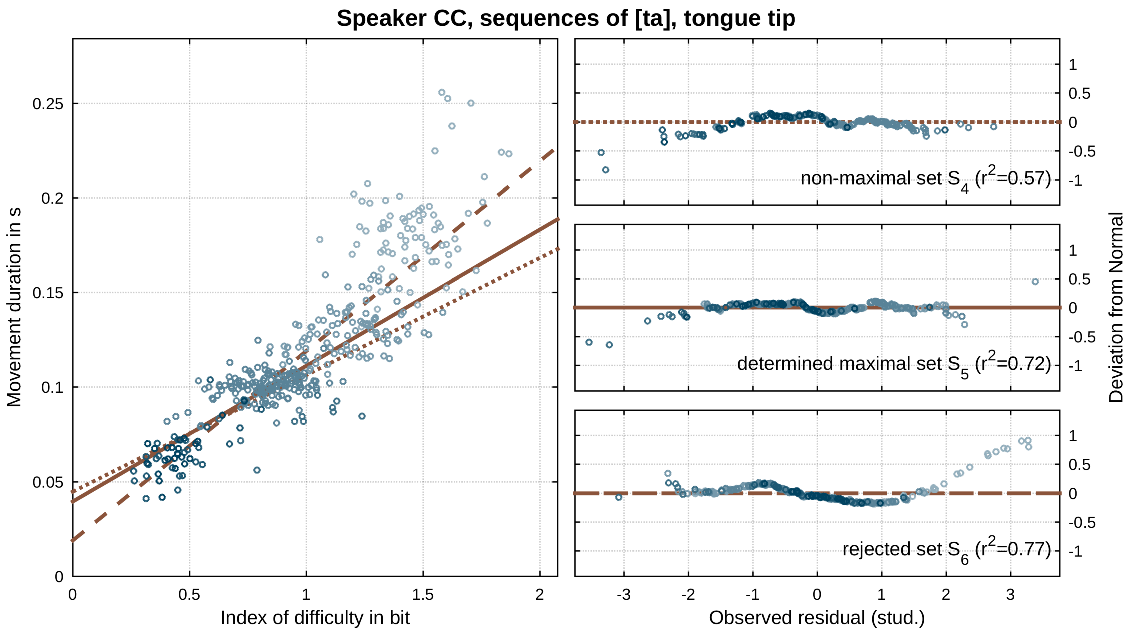

Figure 4. Details of three consecutive sets of rates ($S_4 \subset S_5 \subset S_6$; see text for specifics) considered in the determination of the maximal Fitts-compliant set of rates for sequences of [ta] of speaker CC. Metronome rate is color-coded using shades from faint (slowest rate) to dark (fastest rate). Left: Relation between movement duration and index of difficulty along with individual regression lines (S_4: dotted, S_5: solid and S_6: dashed). Right: Detrended normal quantile-quantile plots of (studentized) regression residuals for the individual sets of rates as constructed by our procedure (from top to bottom). Normality of the residuals for the determined maximal set of rates (S_5) has improved by expansion from S_4 (right, top panel) to S_5 (right, middle panel). When S_5 is expanded to the next larger set S_6 (right, bottom panel), residuals substantially deviate from normality and this deviation is solely caused by inclusion of the next slower rate (indicated by the faintest shaded data).

starting from the fastest and descending to these lower rates, showing evidence for the presence of Fitts' law. For full comparability with Fitts (1954), Table 2 also lists throughput values, given by the inverse slope $1/b$, which range from 5.8 to 34.7 bit/s for sequences of [ta] and from 7.5 to 20.1 bit/s for sequences of [ka].[3] These estimates, with a median of 17.7 bit/s in case of [ta] and 14.0 bit/s in case of [ka], attain values above but of the same order as in Fitts' original results (approx. 10 bit/s, $r^2 = 0.79$, $p < 0.05$) (Fitts, 1954, p. 385).

[3] Throughput expresses how much information per unit time the motor control system processes in achieving the task of the corresponding target-directed movement.

In sum, our results are twofold. We find no evidence for Fitts' law in the data below 150–300 bpm. For faster rates (equal to and above 150–300 bpm) there is evidence for the presence of Fitts' law. In these rates, that is, the Fitts' law expected linear correlation between ID and T is very strong ($r^2 = 0.62 \ldots 0.89$, $p < 0.0001$) and this linear relation holds regardless of the effector implicated in the task, tongue tip for [ta] or tongue body for [ka].

4. Discussion

In this section, we turn to consider our results in the context of work in both speech and other areas of movement science. We address first implications of our results for models of speech, moving on to prospects for extending this work to other classes of speech actions, and finally to commonalities across speech and other domains of human movement.

In modern approaches to speech, an utterance is a sequence of overlapping gestures. Each gesture is controlled by a dynamical system with fixed-point dynamics. That is, a gesture is a unit of action which specifies how, from an arbitrary initial value of a controlled task variable, the vocal tract stabilizes that task variable. A long line of work has proceeded on the hypothesis that the units of action underlying this flow of movements are controlled by an organization similar to a mass-spring system. The standard model of the gesture is a special case of the damped linear oscillator with critical damping (Saltzman and Munhall, 1989; Fowler, Rubin, et al., 1980). Specifically, the model utilizes the dynamical system $\ddot{x} = -kx - b\dot{x}$, with stiffness parameter k and damping $b = 2\zeta\sqrt{k}$. Critical damping is realized by

	CC	CS	DW	FK	SV	TI
[ta] Slowest rate R in bpm	210	150	210	210	210	300
[ta] Correlation strength r^2	0.72	0.62	0.87	0.70	0.88	0.62
[ta] Throughput $1/b$ in bit/s	13.9	5.8	18.2	17.2	19.6	34.7
[ka] Slowest rate R in bpm	150	150	210	150	150	210
[ka] Correlation strength r^2	0.76	0.71	0.77	0.83	0.89	0.67
[ka] Throughput $1/b$ in bit/s	9.3	7.5	17.5	12.4	20.1	15.5

Table 2. Properties of per-speaker determined Fitts-compliant regions. All p-values reside below 0.0001 (very strong significance).

a fixed damping ratio ζ equal to one (thus, neither ζ nor b act as variable control parameters of the model). It can be shown that for any $k > 0$ and $\zeta > 0$ solutions of the system are of the form $x(t) = \mathrm{e}^{-\gamma t}\tilde{x}(t)$, with some real-valued constant γ and some function $\tilde{x}$ not eliminating the exponential signature of x. Hence, the solutions of the standard model are of exponential form for any $k > 0$ and $\zeta > 0$.

In independent work, Crossman and Goodeve, first in a presentation in 1963 and later in published form (Crossman and Goodeve, 1983), as well as Card, Moran, et al. (1983) and Connelly (1984) have shown that Fitts' law holds true for any model dictating movement trajectories of an exponential form (i.e., functions of time that exponentially approach a steady state as time approaches infinity; Connelly, 1984, p. 625). For such models, it was proven analytically that movement time scales linearly with the logarithm of movement error (and thus accuracy). This linear relationship is identical to what Fitts' law predicts.

Consequently, any instantiation of the damped linear oscillator model for speech predicts that the data it describes must conform to Fitts' law (irrespective of the specific values of the parameters k and ζ). Recall now that in our data we have found evidence for Fitts' law only for speaker-specific rates above or equal to 150–210 bpm. The standard model of speech gestures thus fails with respect to the above prediction, that is, the model is not consistent with the *absence* of Fitts' law for every speaker at some rates. Moreover, this failure occurs at rates that are certainly rates which include normal speaking conditions.

Given these results, a validity test for any proposed model is that it must predict both the presence and absence of Fitts' law. One way that this may be accomplished is via some model parameter (or set of parameters) which reflects the presence or absence of sufficient temporal pressure or its gradual equivalent. In the standard model, the only such parameter is the control parameter of stiffness k, which may be considered as a proxy to speech rate by controlling the frequency of the oscillator $\omega = \sqrt{k}$ (cf. Kelso, Vatikiotis-Bateson, et al., 1985, also Fuchs, Perrier, et al., 2011). However, as shown above, manipulation of k does not alter the general exponential signature of the movement trajectories. Hence, another way to characterize the failure of the standard model on the Fitts' law test is to say that the model does not include a parameter or set of parameters which would express the same notion of temporal pressure as required by Fitts. Other candidate models for fixed-point dynamics exist (e.g., Kröger, Schröder, et al., 1995; Sorensen and Gafos, 2016) but have not been investigated yet with respect to their conformity to Fitts' law. It remains to be seen how these models fare in the face of the evidence from our results.

The above aim must proceed in tandem with elaborating and extending the empirical range of speech actions with respect to Fitts' law. Our assessment of the law focused on oral plosives. Plosives are produced by an occlusion in the mid-sagittal section of the vocal tract. This occlusion is achieved when an active articulator (e.g., the tongue tip or the tongue back) makes contact with a region on the palate along the longitudinal axis of the vocal tract. For [t] and [k], a position-based notion of target seems relatively uncontroversial. Our assessment shows that one can fruitfully follow rigorous data-derived methods for defining targets for plosives. This is one of the reasons we focused on this class of speech segments. We are aware that, especially when it comes to other segment classes, there are approaches to the notion of target which use combinations of orosensory parameters or also acoustic notions of target (e.g., Guenther, 1995). One other class of speech segments where Fitts' law also appears to be particularly relevant is fricatives. For fricative consonants (e.g., [f, v, s, z, ʃ, ʒ, x, ɣ]), the constriction is not full. Rather, a small channel is formed between the active articulator and some vocal tract region with the airstream passing through giving rise to turbulence generated either at the point of the constriction (channel turbulence) as in the velar fricative [x] or by the airstream hitting an obstacle anterior to the occlusion (wake turbulence) as in [s]. The cross-sectional area of this channel must be sufficiently small to generate turbulence but not too narrow to result in a complete constriction and not too wide to result in an approximant (Catford, 1977). Examples include, at the velar place of constriction, stop [k] vs. fricative [x] vs. approximant [ɰ] or, at the palatal place of constriction, [c] vs. fricative [ç] vs. approximant [j]. For these reasons, the articulatory postures of fricatives seem to require more precise control of the supralaryngeal configuration of the vocal tract than those for the corresponding plosives. Kinematic comparisons between plosives and fricatives appear consistent with this distinction.

One empirically well-documented kinematic relation is that between a movement's peak velocity and its amplitude. This relation has been described as an overall linear correlation (Ostry and Munhall, 1985) with velocity-amplitude slopes steeper for faster than for slower speech rates (Vatikiotis-Bateson and Kelso, 1990, 1993) and decreasing covariation as durational variability increases (ibidem). Most relevantly for our purposes, Kuehn and Moll (1976) observed higher velocity-amplitude relationship slopes for movements toward plosives than for movements toward fricatives (see also Guenther, 1995, p. 605). Moreover, such evidence from kinematics for plosives versus fricatives appears consistent with what is known from other human movement domains where precision requirements in some performed task have been linked to a number of kinematic manifestations. Thus, in discrete aiming tasks of the hand, C. L. MacKenzie, Marteniuk, et al. (1987) report lower peak velocities for smaller target sizes as well as modulations of velocity profile shape (that is, change in velocity over time) as a function of target size. The empirical evidence here concerning peak velocity and velocity profiles has been used as a testbed for dynamical models of movement in discrete aiming tasks (C. L. MacKenzie, Marteniuk, et al., 1987), reciprocal tapping tasks (Bootsma, Fernandez, et al., 2004), saccade eliciting tasks (Opstal and Gisbergen, 1987) and finally also in speech (Kröger, Schröder, et al., 1995; Sorensen and Gafos, 2016).

In sum, extending our understanding of speech actions with respect to Fitts' law would enable further elaboration of models of speech and clarification of potential connections between speech and other domains of human movement.

5. Conclusion

We asked whether speech movements abide to Fitts' law as (target-directed) movements from other domains of human motor control do. To address this question we registered movement data from [ta] and [ka] sequences spoken at eight distinct rates, ranging from extremely slow to extremely fast (30–570 bpm). In the resulting dataset, we sought evidence of the sort that has provided support for Fitts' law in other areas of motor control. We find that slow rates do not abide to Fitts' law. But, beyond a (participant) specific rate, the characteristic linearity of the relation between time and index of difficulty emerges. In sum, fast tongue movements of repetitive speech conform to Fitts' law; for slower movements, the relation expressed by this law seems to break down. In future work, we aim to pursue ways in which models of speech may account for our current results and to broaden the empirical basis wherein relations involving kinematic and task space coordinates are implicated.

References

Anscombe, Francis John (1973). "Graphs in statistical analysis". *The American Statistician* 27(1), pp. 17–21. doi: 10.2307/2682899.

Bernstein, Nikolai A. (1967). *The co-ordination and regulation of movements*. Oxford, United Kingdom: Pergamon Press.

Bootsma, R. J., L. Fernandez, and D.Mottet (2004). "Behind Fitts' law: kinematic patterns in goal-directed movements". *International Journal of Human-Computer Studies* 61(6), pp. 811–821. doi: 10.1016/j.ijhcs.2004.09.004.

Browman, Catherine P. and Louis M. Goldstein (1986). "Towards an articulatory phonology". *Phonology* 3(1), pp. 219–252. doi: 10.1017/S0952675700000658.

Card, Stuart K., Thomas P. Moran, and Allen Newell (1983). *The psychology of human-computer interaction*. Hillsday, United States: Lawrence Erlbaum Associates. isbn: 0898592437.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington, United Kingdom: Indiana University Press.

Cohen, Ayala (1983). "Comparing regression coefficients across subsamples. A study of the statistical test". *Sociological Methods & Research* 12(1), pp. 77–94. doi: 10.1177/0049124183012001003.

Connelly, Edward M. (1984). "A control model: interpretation of Fitts' law". *Twentieth Annual Conference on Manual Control* (Moffett Field, United States). Vol. 1, pp. 621–642.

Crossman, E. R. F. W. and P. J. Goodeve (1983). "Feedback control of hand-movement and Fitts' law". *Quarterly Journal of Experimental Psychology* 35(2). Paper presented at the meeting of the Experimental Psychology Society, Oxford, 1963, pp. 251–278. doi: 10.1080/14640748308402133.

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

Fitts, Paul M. and James R. Peterson (1964). "Information capacity of discrete motor responses". *Journal of Experimental Psychology* 67(2), pp. 103–112. doi: 10.1037/h0045689.

Fowler, Carol A., Philip Rubin, Robert E. Remez, and Michael T. Turvey (1980). "Implications for speech production of a general theory of action". *Language Production: Speech and Talk*. Ed. by Brian Butterworth. New York, United States: Academic Press, pp. 373–420.

Fuchs, Susanne, Pascal Perrier, and Mariam Hartinger (2011). "A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model". *Journal of Speech, Language, and Hearing Research* 54(4), pp. 1067–1076. doi: 10.1044/1092-4388(2010/10-0131).

Grosjean, Marc, Maggie Shiffrar, and Günther Knoblich (2007). "Fitts's Law Holds for Action Perception". *Psychological Science* 18(2), pp. 95–99. doi: 10.1111/j.1467-9280.2007.01854.x.

Guenther, Frank H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production". *Psychological Review* 102(3), pp. 594–621. doi: 10.1037//0033-295X.102.3.594.

Horn, Berthold K. P. (1987). "Closed-form solution of absolute orientation using unit quaternions". *Journal of the Optical Society of America A* 4(4), pp. 629–642. doi: 10.1364/JOSAA.4.000629.

Kelso, J. A. Scott, Betty Tuller, and Katherine S. Harris (1983). "A "dynamic pattern" perspective on the control and coordination of movement". *The Production of Speech*. Ed. by Peter F. MacNeilage. New York, United States: Springer, pp. 137–173. doi: 10.1007/978-1-4613-8202-7_7.

Kelso, J. A. Scott, Eric Vatikiotis-Bateson, Elliot L. Saltzman, and Bruce Kay (1985). "A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling". *The Journal of the Acoustical Society of America* 77(1), pp. 266–280. doi: 10.1121/1.392268.

Kröger, Bernd J., Georg Schröder, and Claudia Opgen-Rhein (1995). "A gesture-based dynamic model describing articulatory movement data". *The Journal of the Acoustical Society of America* 98(4), pp. 1878–1889. doi: 10.1121/1.413374.

Kuehn, David P. and Kenneth L. Moll (1976). "A cineradiographic study of VC and CV articulatory velocities". *Journal of Phonetics* 4(4), pp. 303–320.

Lammert, Adam C., Christine H. Shadle, Shrikanth S. Narayanan, and Thomas F. Quatieri (2018). "Speed-accuracy tradeoffs in human speech production". *PLOS One* 13(9). doi: 10.1371/journal.pone.0202180.

MacKenzie, C. L., R. G. Marteniuk, C. Dugas, D. Liske, and B. Eickmeier (1987). "Three-dimensional movement trajectories in Fitts' task: implications for control". *Quarterly Journal of Experimental Psychology* 39(4), pp. 629–647. doi: 10.1080/14640748708401806.

MacKenzie, Ian Scott (1992). "Fitts' law as a research and design tool in human-computer interaction". *Human-Computer Interaction* 7(1), pp. 91–139. doi: 10.1207/s15327051hci0701_3.

MacKenzie, Ian Scott (2013). "A note on the validity of the Shannon formulation for Fitts' index of difficulty". *Open Journal of Applied Sciences* 3(6), pp. 360–368. doi: 10.4236/ojapps.2013.36046.

Munhall, Kevin G., David J. Ostry, and Avraham Parush (1985). "Characteristics of velocity profiles of speech movements". *Journal of Experimental Psychology: Human Perception and Performance* 11(4), pp. 457–474. doi: 10.1037//0096-1523.11.4.457.

Nelson, Winston L. (1983). "Physical principles for economies of skilled movements". *Biological Cybernetics* 46(2), pp. 135–147. doi: 10.1007/BF00339982.

Nelson, Winston L., Joseph S. Perkell, and John R. Westbury (1984). "Mandible movements during increasingly rapid articulations of single syllables: preliminary observations". *The Journal of the Acoustical Society of America* 75(3), pp. 945–951. doi: 10.1121/1.390559.

Opstal, A. J. Van and J.A.M. Van Gisbergen (1987). "Skewness of saccadic velocity profiles: a unifying parameter for normal and slow saccades". *Vision Research* 27(5), pp. 731–745. doi: 10.1016/0042-6989(87)90071-X.

Ostry, David J., James D. Cooke, and Kevin G. Munhall (1987). "Velocity curves of human arm and speech movements". *Experimental Brain Research* 68(1), pp. 37–46. doi: 10.1007/BF00255232.

Ostry, David J. and Kevin G. Munhall (1985). "Control of rate and duration of speech movements". *The Journal of the Acoustical Society of America* 77(2), pp. 640–648. doi: 10.1121/1.391882.
Patel, Aniruddh D., Anders Löfqvist, and Walter Naito (1999). "The acoustics and kinematics of regularly timed speech: a database and method for the study of the p-center problem". *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco, United States). Vol. 1, pp. 405–408.
Plamondon, Réjean and Adel M. Alimi (1997). "Speed/accuracy trade-offs in target-directed movements". *Behavioral and Brain Sciences* 20(2), pp. 279–303. doi: 10.1017/S0140525X97001441.
Saltzman, Elliot L. and Kevin G. Munhall (1989). "A dynamical approach to gestural patterning in speech production". *Ecological Psychology* 1(4), pp. 333–382. doi: 10.1207/s15326969eco0104_2.
Schmidt, Richard and Tim Lee (2011). *Motor control and learning: a behavioral emphasis*. Champaign, United States: Human Kinetics. isbn: 0736079610.
Sorensen, Tanner and Adamantios I. Gafos (2016). "The gesture as an autonomous nonlinear dynamical system". *Ecological Psychology* 28(4), pp. 188–215. doi: 10.1080/10407413.2016.1230368.
Turvey, Michael T. (1977). "Preliminaries to a theory of action with reference to vision". *Perceiving, acting and knowing: toward an ecological Psychology*. Ed. by R. Shaw and J. Bransford. Lawrence Erlbaum Associates, pp. 211–265.
Turvey, Michael T. (2007). "Action and perception at the level of synergies". *Human Movement Science* 26(4), pp. 657–697. doi: 10.1016/j.humov.2007.04.002.
Vatikiotis-Bateson, Eric and J. A. Scott Kelso (1990). "Linguistic structure and articulatory dynamics: a cross language study". *Haskins Laboratories Status Report in Speech Research* SR-103/104, pp. 67–94.
Vatikiotis-Bateson, Eric and J. A. Scott Kelso (1993). "Rhythm type and articulatory dynamics in English, French and Japanese". *Journal of Phonetics* 21, pp. 231–265.
Welford, Alan Traviss (1968). *Fundamentals of skill*. Methuen's manuals of modern psychology. London, United Kingdom: Methuen. isbn: 0416030009.
Wobbrock, Jacob O., Kristen Shinohara, and Alex Jansen (2011). "The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, Canada), pp. 1639–1648. doi: 10.1145/1978942.1979181.
Woodworth, Robert S. (1899). "The accuracy of voluntary movement". *The Psychological Review*. Vol. III (3). Series of Monograph Supplements. doi: 10.1037/h0092992.

Research Article III

Distinct phase space topologies of identical phonemic sequences

Language, Linguistic Society of America
Submitted: April 10, 2019; *Reviewer invited:* April 24, 2019

Abstract

Using a metronome-driven speech elicitation paradigm, we reconstruct the phase space underlying speech movements in repeated syllables and study its topological structure in terms of the number and type of attractors present. We show that, at slow metronome rates, the topological structure underlying the observed action corresponds to a successive concatenation of closing (for the consonant) and opening (for the vowel) gestures, each with its own equilibrium point. This organization changes qualitatively at faster rates. First, as rate increases, the equilibrium of opening gestures dissolves leaving only a closing gesture equilibrium. Second, for some speakers, an additional organization is observed. For these speakers, at the fastest rates, our results indicate that both equilibria vanish in favor of a different, periodic attractor.

1. Introduction

A large class of models known as equilibrium point models propose that the organism accomplishes speech and other biological action by setting the location of a so-called point attractor (Abraham and Shaw, 1983) within some hypothesized state space characterizing the observed action (Feldman, 1980a,b; Kelso, 1977; Polit and Bizzi, 1978, 1979; in the speech community, Saltzman and Munhall, 1989; Browman and Goldstein, 1986; Perrier, Ostry, et al., 1996; Guenther, 1995). Consider, for example, a repeated /CVCV.../ sequence. Achieving the C involves setting up a target for the C. This must then be followed by the initiation of the opening gesture for the vowel, so as to go from the C to the V, which in turn must be followed by the closing gesture by setting up the target of the C again, and so on.

Any theory of speech units, including units corresponding to consonants and vowels but perhaps also larger constellations of such units, must necessarily make explicit how such units flexibly unfold in time and yet are stable enough to be reproducible in behavior. It is thus not surprising that the foundation for all modern models of speech and other biological action is dynamical systems theory, which concerns itself with recurrence in patterning over time. In a most concise way, a dynamical system is a formal model that expresses a rule for how the state of a system changes in time. The state of a dynamical system is defined in terms of a set of dimensions needed to describe the system. Each state is a point specified by values along these dimensions. The set of all possible states is called the phase space. Under the assumption of second order dynamics (Kelso and Tuller, 1984; Kay, Kelso, et al., 1987), the phase space is defined by the dimensions of displacement u and its first derivative $\dot{u}$. Thus, any state of such a system is unambiguously given by a pair of values $(u, \dot{u})$. As the state of the system changes, it traces out a path in phase space called a phase trajectory. The behavior of these trajectories reveals information about the topology of the underlying phase space (its number and type of attractors). Consider the phase space of the speech action in a sequence of repeated [ka] syllables. This sequence involves two types of gestures, closing (leaving [a] going toward [k]) and opening (leaving [k] going toward [a]). In other words, two targets [k] and [a] are involved which must correspond to two point attractors in phase space. In dynamical systems theory, point attractors are stable states which attract nearby passing phase trajectories. Given this distinctive property of attractors, there exists concepts and tools for their identification by inspection of the phase space. Yet the existence of attractors in speech has not heretofore been rigorously assessed or verified. Our aim is thus to inspect the phase space of our registered data for signatures of point attractors or any other type of attractor.

Specifically, in this contribution, we reconstruct the phase space underlying the speech movements in repeated syllables and study its topological structure in terms of the number and type of attractors present. In obtaining speech movements, we asked participants to repeat [ka] syllables in time with a metronome. The participants' speech movements were recorded with 3D electromagnetic articulometry (EMA). Metronome rate was set at 150, 210, 300, 390, 480 and 570 beats per minute (bpm; corresponding to 2.5, 3.5, 5.0, 6.5, 8.0 and 9.5 Hz). At the slower rates, our results indicate high conformity with the basic understanding of how speech is regulated. We show that, for slow metronome rates, the topological structures underlying the observed movements consist of a successive concatenation of closing and opening gestures, each with its own equilibrium point. As rate increases, opening and closing equilibria are not equally stable. Specifically, the equilibrium point of the opening gesture vanishes as rate increases leaving only a closing gesture equilibrium point. The organization switches qualitatively to one of forming the constriction then leaving it and repeat. For some speakers, an additional organization is observed. For these speakers, at the fastest rates, our results indicate that both equilibria vanish in favor of a different periodic attractor.

2. Methods

Five native speakers of German (three females and two males referred to as CC, DW, CS and TI and FK respectively) participated in the experiment. The speakers were between 23 and 30 years old, without any present or past speech or hearing problems. They were recruited at the University of Potsdam and paid for their participation in the experiment. All procedures were performed in compliance with relevant laws and institutional guidelines and were approved by the ethics committee of the University of Potsdam. Written informed consent was obtained from all subjects.

During the experiment, all participants were prompted on a computer screen to produce sequences of repeated [ka] syllables in time with an audible metronome. The metronome served as an extrinsic index of the intended rate of syllable production. We did not require participants to aim for synchronizing any specific point of the sequence with the metronome. The participants were instructed to articulate their responses accurately and naturally. The rate of the metronome was set to the values of 150, 210, 300, 390, 480 and 570 bpm (corresponding to 2.5, 3.5, 5.0, 6.5, 8.0 and 9.5 Hz). At the start of each trial, the participant was exposed to the metronome stimulus and begun articulating the required response syllable at a point of their choice. Starting with the slowest rate, a minimum of four trials at each rate were recorded. Once this minimum was reached, recording proceeded with the next higher rate. The duration of each trial (hence, duration of the metronome stimulus) was timed such that the participant was able to adjust to the beat of the metronome and produce a coherent sequence of approximately 30 syllables.

Articulatory data as well as acoustic data were registered. All recordings took place in our sound-attenuated booth using a Carstens AG501 3D Electromagnetic Articulograph for articulatory, and a YOGA Shotgun microphone EM-9600 attached to a TASCAM US-2x2 Audio interface for acoustic data registration. Three-dimensional electromagnetic articulography (EMA) allowed measurement of kinematic displacement data at a high precision. Along with some other auxiliary reference locations (upper and lower incisors, nose bridge, left and right mastoids), we tracked the position of the sensor attached to the tongue back articulator, the major effector involved in the production of [ka] syllables.

Three-dimensional displacement data (x, y, z), provided by the AG501 Articulograph, was digitized at a sampling rate of 1250 Hz. In order to eliminate high frequency noise, a zero-delay Butterworth lowpass filter of fourth order with cutoff frequency of 25 Hz was applied to all recorded signals. Based on these signals, the spatial transformations of head movement correction and occlusal reference frame alignment were determined and applied.

Prior to the registration of speech movement data, we recorded palate traces of each speaker. A sensor coil was duct-taped to the thumb tip of each participant's dominant hand. Following this, the participants were instructed to slide their thumbs in a zig-zag motion starting from their front teeth and going posteriorly as far as possible to the rear of their palate. Each participant was instructed to maintain proper contact of the sensor with the palate. This procedure was repeated three times. The registered palate traces were collectively used to approximate the speaker's palate shape in the subsequent analysis (see below).

Recall that a speech gesture specifies a target on task space coordinates rather than on raw positions of individual effectors (here, the tongue back). Hence, a proper evaluation of the hypothesis that the speech gestures underlying [ka] involve the presence of two equilibria must be carried out in a phase space defined on such task space coordinates. Based on the palate models, the three-dimensional spatial trajectories (x, y, z) of the tongue back displacement data were transformed into one-dimensional palate-relative signals u expressing the degree of constriction between the articulator and the speaker's palate. Specifically, instantaneous values of u were computed as the vertical distance between the tongue back sensor position (x, y, z) and the determined palate surface (see Beckman, Jung, et al., 1995; Hashi, Westbury, et al., 1998; Neufeld and van Lieshout, 2014 for details and technical issues of constriction degree determination). The first derivative of u (velocity $\dot{u}$) was determined by means of numerical differentiation using nine-point finite differences with eighth order of accuracy.

For each speaker, we reconstructed a palate model by means of surface regressions of the recorded palate trace data. In particular, we used a locally weighted scatterplot smoothing method (LOWESS) with a (qualitatively determined) smoothing parameter of 0.15 (see also Neufeld and van Lieshout, 2014). LOWESS regression fits simple linear models to localized subsets of the data and does not require the specification of a global function to fit all of the data (e.g., polynomials as in common nonlinear regression). Figure 1 shows the results of the regressions along with the underlying registered palate traces.

3. Phase space reconstruction

The phase space of a dynamical system of second order is the set of all possible states defined on the displacement u-velocity v plane (for easier reading we write $\dot{u} = v$ from here on). Accordingly, time series of displacement u paired with their first derivative v form trajectories in phase space (phase trajectories) describing the evolution of the system's state as time proceeds. For each speaker and metronome rate (henceforth "condition"), we normalized phase trajectories such that their mean displacement extrema (u_{min} and u_{max}) and mean velocity extrema (v_{min} and v_{max}) map to the interval $[-1, 1]$.

Figure 2 shows displacement u and velocity v time series (left and right panels) and how these combine to phase trajectories in phase space (four center panels) from two of our speakers (TI and FK) and two metronome rates (210 and 300 bpm). The left half of the figure (the four time series and the two phase planes)

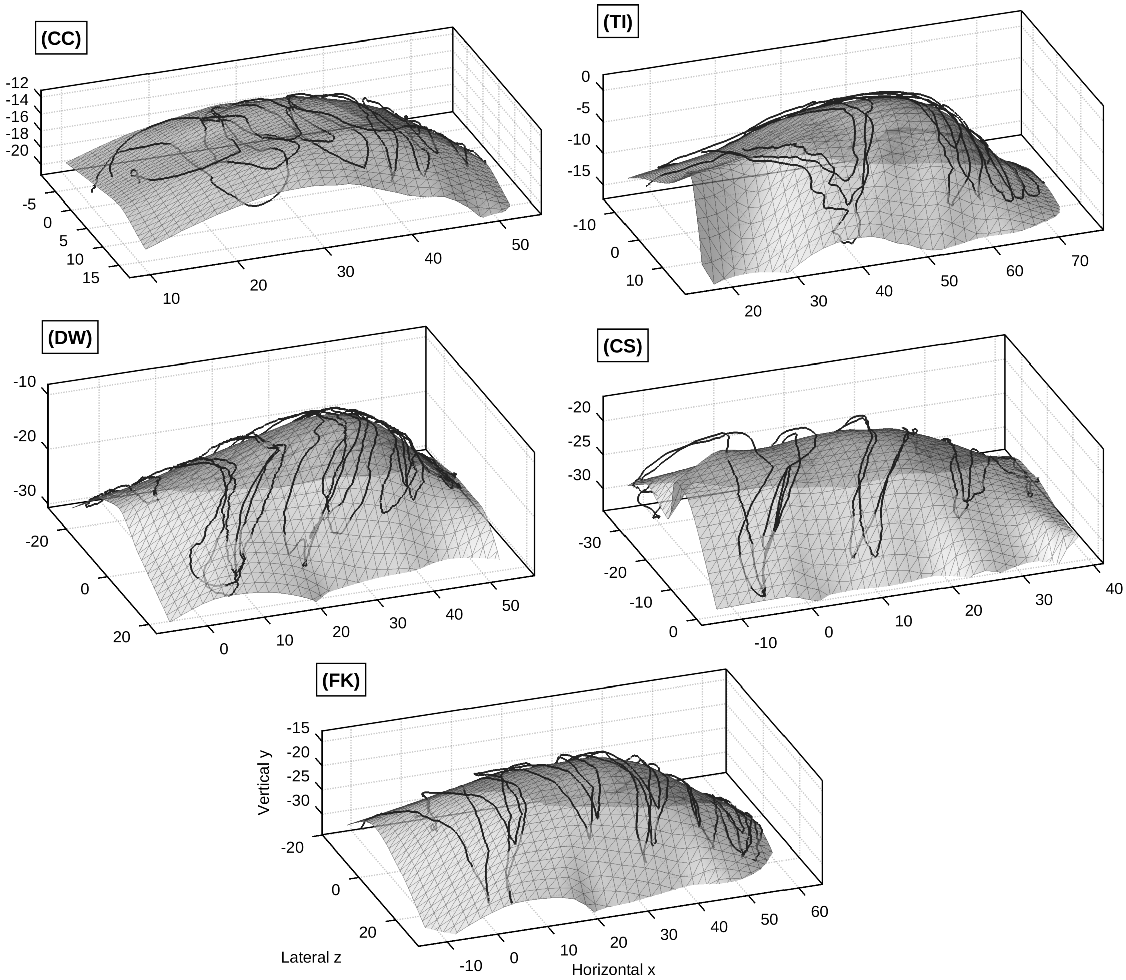

Figure 1. Palate models determined by surface regression (LOWESS, smoothing parameter 0.15). The registered sensor traces are superimposed as thick solid lines. These palate models were used to derive palate-relative signals expressing the degree of constriction between the tongue back sensor and the specific speaker's palate.

corresponds to the moderate metronome rate of 210 bpm. The right half of the figure (another four time series and two phase planes) corresponds to the faster metronome rate of 300 bpm. We choose these rates and speakers because they provide a minimal set for illustrating the following key point. Our examples of phase trajectories in Figure 2 clearly depict the repetitive nature of the task: form a tongue-palate constriction, release that, and repeat. Yet, mere inspection of phase trajectories and their time evolution does not serve to reveal the attractor topology of the phase space. In particular, as we demonstrate below, at the rate of 210 bpm (left half of Figure 2), the attractor topology of both speakers

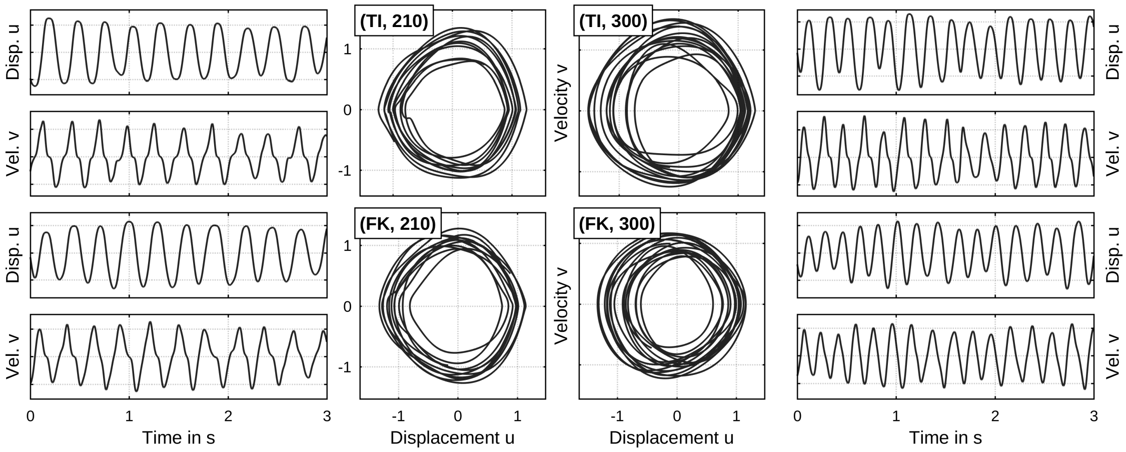

Figure 2. Time series of displacement u and velocity v (left and right panels) along with corresponding phase trajectories (four center panels) from two speakers (TI and FK) and two metronome rates (210 and 300 bpm). In the phase planes, time proceeds in clockwise direction as indicated by the two arrows. Formation of [k] constrictions belong to the upper arrow, following release and formation of [a] to the lower arrow.

is the same. Each speaker shows two equilibrium points, one for the closing and another for the opening gesture. In contrast, at the faster rate of 300 bpm (right half of Figure 2), our two speakers differ in ways that cannot be inferred by inspection of the phase trajectories: the phase space of speaker TI consists of a single equilibrium point for the closing gesture (no opening gesture equilibrium), whereas speaker FK's phase space comprises a periodic attractor (no equilibrium at all). This diversity in organization is hardly distinguishable in a representation like the one of Figure 2 and hence call for further inspection of the phase space.

In Figure 2, let us walk through the phase space of the action in the task of producing a sequence of repeated [ka] syllables. In this task, two types of gestures are involved, closing (leaving [a] going toward [k]) and opening (leaving [k] going toward [a]). In other words, two targets are involved, one for [k] with a maximal value of u (maximum degree of constriction) and another for [a] with a minimal value of of u (minimum degree of constriction). In the phase planes shown in Figure 2, these targets correspond to extrema on the horizontal axis of the displacement u. Since closing gestures change the system's state from a minimal to a maximal value of u, they are associated with trajectories with positive velocities $v > 0$ (i.e., increasing constriction degree). Similarly, opening gestures correspond to trajectories with negative velocities $v < 0$ (decreasing constriction degree). The targets of [k] and [a] are associated with states of zero velocity $v = 0$. Thus, the overall schematic picture of a phase trajectory of repeated [ka] syllables in the phase plane (u, v) is: starting from the first state $(u_{\min}, 0)$ (the vowel [a]), the trajectory follows a path through the upper half plane of the phase space $(v > 0)$ up to the second state $(u_{\max}, 0)$ (the consonant [k]). Subsequent to this closing gesture, the opening gesture of the vowel corresponds to the trajectory following a path through the lower half of the phase plane $(v < 0)$ back to the state $(u_{\min}, 0)$ of the vowel [a] from which the process repeats until the final [ka] syllable is reached.

Recall now the main problem raised at the outset: what are the principles that enable speech units to flexibly unfold (under a rather variegated set of conditions) and yet be reproducible in behavior? In any representation of speech as sequences of phonemes, this problem is implicit. The problem is made explicit in dynamical systems theory where concepts and tools exist that enable one to generate hypotheses and pursue these rigorously. The essential starting point is that organization and reproducibility imply the presence of recurrent, preferred modes in state space. Put in different words: not all states of the phase space are equiprobable. Integrated over time, some states are more likely to be visited than others. Specifically, states belonging to an attractor of the dynamical system will show higher probability than others. An attractor is a subset of the phase space which attracts all nearby passing phase trajectories. Hence, attractors continuously draw probability mass from their vicinities and accumulate it to a significant amount over time compared to other states. If there is any attractor present in the underlying dynamical system, the probability density function of the phase space can be exploited to reveal important information about its location and geometry. In particular, equilibrium points, as point attractors, will manifest themselves as localized peaks in the probability density function. Other potential attractors are conceivable, most notably periodic attractors such as limit cycles. Their signatures in the phase space probability density function would be that of closed, ring-like structures of high density.

The central hypothesis assessed is that sequences of repeated [ka]s correspond to an organization admitting two equilibrium points. Specifically, these should correspond to two states of maximum and minimum tongue-palate constriction, corresponding to [k] and [a] respectively. We now turn to examine the phase space in the vicinities of these states in order to uncover information about the underlying dynamical organization of the tongue movements under consideration here. For each condition (given by a particular speaker and metronome rate), we binned the phase space using an even rectangular grid of 25×25 cells (25 bins in each of the dimensions of displacement u and velocity v). For each bin, we counted the number of phase trajectory hits (given the sampling rate of 1250 Hz) and normalized by the total number of bin hits in the entire histogram. With this method we are able to estimate the probability densities of states in the phase space of each condition.

Figure 3 shows the probability density functions of our data separately for each condition (speaker and metronome rate). Therein, densities are color-coded using bright shades for low and dark shades for high density values. For reasons of presentation, we cropped the panels such that they only show regions of the phase space containing (at least) 96% of the probability mass. In each panel, the state of maximum constriction (maximal displacement u) associated with the consonant [k] is located at the right-hand side of the phase plane. The phase space state of minimal constriction (minimal displacement u) associated with the vowel [a] resides on the left-hand side of the phase plane. For all speakers and almost all metronome rates, there is strong indication for the presence of an equilibrium point associated with the consonant expressed by highly localized density peaks at the right-hand side of each phase plane. Furthermore, for all speakers, the same signature (localized density peaks, at the left-hand side of the phase plane) can be found for the slow rates of 150 and 210 bpm indicating the presence of an equilibrium point for the vowel there as well. At rates above or equal to 300 bpm, this vowel equilibrium point disappears leaving only an equilibrium for the consonant. For some speakers (CS and FK, conceivably DW), the density functions indicate a decay of these [k] equilibria too, giving rise to a periodic attractor expressed by high densities with closed, ring-like structure (CS, 300 bpm and above; FK, 480 and 570 bpm; conceivably DW at 570 bpm).

Let us sum up the main take away from these observations. What appears on the surface in the task our participants undergo is a monotony of reiterant [ka]s. What lies beneath and what our results indicate is this novel thesis: there is

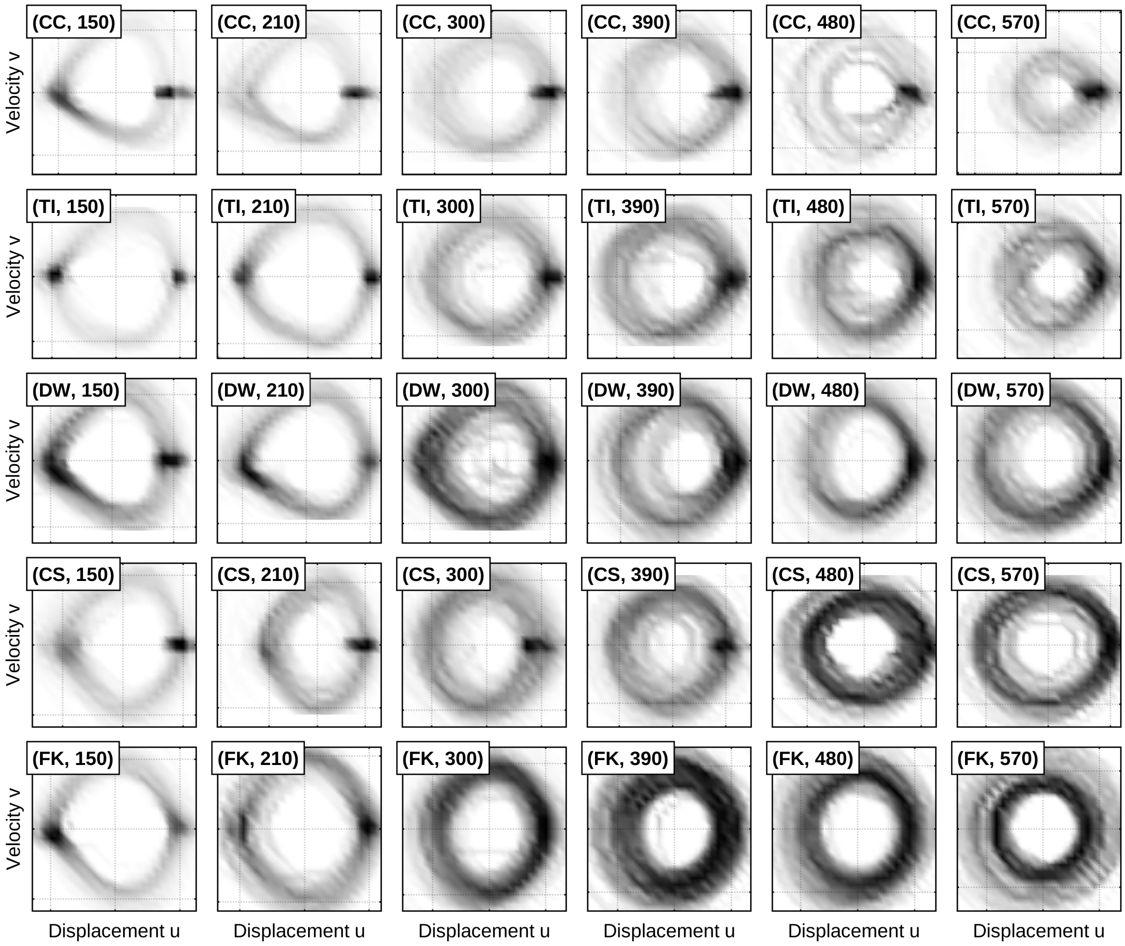

Figure 3. Probability densities of the phase plane (u, v). Densities are gradually color-coded using bright shades for low and dark shades for high values. Localized peaks in probability density indicate the presence of equilibrium points. Closed, ring-like structures of high density indicate periodic attractors. In each panel, the state of maximum constriction degree (corresponding to the consonant) is located at the right-hand side and the state of minimum constriction degree (corresponding to the vowel) at the left-hand side of the phase space.

no fixed isomorphism between a spoken utterance and its dynamical organization. Consider, for example, TI's and FK's utterances at 300 bpm. These utterances are perceived as the same phonemic type, a reiterant sequence of [ka]s. Yet, their dynamical organization is distinct: at 300 bpm, TI's phase space indicates the presence of a single equilibrium point for the closing gesture, whereas speaker FK's phase space indicates the presence of a periodic attractor. The same point can be illustrated within one speaker by looking at different rates. Thus, the phase space of speaker TI at 150 bpm indicates the presence of the two expected equilibria, one for the consonant and another for the vowel. But the phase space of speaker TI at 300 bpm indicates the presence of only one equilibrium point corresponding to the closing gesture for the [k]. Once again, organization is not fixed, here within a speaker. For all speakers, the organization underlying what is macroscopically a repeated sequence of [ka]s flexibly changes as rate is increased.

4. Conclusion

It is common to assume that speech involves the successive concatenation of attractors in some (appropriately defined) task space of speech goals. Yet the topological structure of speech action has not heretofore been rigorously assessed or verified. In a task where participants repeated syllables in time with a metronome, we reconstructed the layout of the attractors underlying the speech action and study their type and number as a function of speech rate. Movements were recorded with 3D electromagnetic articulometry. Metronome rate was set at 150, 210, 300, 390, 480 and 570 bpm (corresponding to 2.5, 3.5, 5.0, 6.5, 8.0 and 9.5 Hz). We find that at the slower rates, the topological structures underlying the observed movements can be described by successive concatenations of closing and opening gestures, each with its own equilibrium point. As rate increases, the equilibria of opening and closing gestures are not equally stable. Specifically, the equilibrium point of the opening gesture dissolves as metronome rate increases leaving only a closing gesture equilibrium point. That is, the organization has switched qualitatively to one of forming the constriction then leaving it and repeat. For some speakers, at the fastest rates, an additional organization is observed. For these speakers, also the remaining equilibrium of the consonant dissolves giving rise to a periodic attractor in the phase space. Our results suggest that the organization underlying simple utterances is non-unique and that qualitative changes in attractor topology result from increases in rate.

Let us place these results in the context of other work in speech and human movement. Pursuing observations by Stetson (1951), Tuller and Kelso (1991, 1990) provided evidence for a change from [ip ip ip...] to [pi pi pi...] by increasing rate (see also de Jong, 2001; de Jong, Lim, et al., 2004 for refinements) whereas we are concerned with changes in the same utterance (no reordering of phonemes).

In our case, the organization in the same utterance seems to change as a function of increasing rate. Potential distinctions in organization underlying what may be apparently similar sequences of actions have been pursued in work that has so far remained unrelated to speech. Specifically, there is evidence from other areas of human movement that increasing rate may result in qualitative changes in the organization underlying the movements. In a unimanual finger flexion-extension task, Jirsa and Kelso (2005) and Huys, Studenka, et al. (2008) provided evidence that, at slow rates, finger movements are driven by an organization governed by equilibrium point dynamics (with either one or two equilibria, depending on whether the motor system instantiates both a flexion-extremum target and an extension-extremum target or just one of these extrema as a target along with a transient excursion away from it). At the faster rates, the organization switches to a different regime constituting a limit cycle attractor. Finally, the indication in our data that the equilibrium point for the vowel dissolves in fast rates makes contact with reports that vowels delete in fast speech and/or under certain prosodic conditions such as in the so-called 'pretonic' schwa case where the vowel precedes a stressed syllable, e.g., *parade*, *Toronto*, [pʰred], [tʰranto] (Zwicky, 1972; Hooper, 1978; Dalby, 1986). In future work, we aim to scrutinize our present results by use of additional properties of the phase space (e.g., its vector field) and to extend our assessment of the topological structure of speech in other phonemic sequences involving different consonant-vowel combinations.

References

Abraham, Ralph H. and Christopher D. Shaw (1983). *Dynamics: the geometry of behavior*. Santa Cruz, United States: Aerial Press. isbn: 0942344022.

Beckman, Mary E., Tzyy-Ping Jung, Sook-hyang Lee, Kenneth J. de Jong, Ashok K. Krishnamurthy, Stanley C. Ahalt, K. Bretonnel Cohen, and Michael J. Collins (1995). "Variability in the production of quantal vowels revisited". *The Journal of the Acoustical Society of America* 97(1), pp. 471–490. doi: 10.1121/1.412945.

Browman, Catherine P. and Louis M. Goldstein (1986). "Towards an articulatory phonology". *Phonology* 3(1), pp. 219–252. doi: 10.1017/S0952675700000658.

Dalby, Jonathan (1986). "Phonetic structure of fast speech in American English". *Phonetics and Phonology*. Bloomington, United States: Indiana University.

De Jong, Kenneth J. (2001). "Rate-induced resyllabification revisited". *Language and Speech* 44(2), pp. 197–216. doi: 10.1177/00238309010440020401.

De Jong, Kenneth J., Byung-jin Lim, and Kyoko Nagao (2004). "The perception of syllable affiliation of singleton stops in repetitive speech". *Language and Speech* 47(3), pp. 241–266. doi: 10.1177/00238309040470030201.

Feldman, A. G. (1980a). "Superposition of motor programs—I. Rhythmic forearm movements in man". *Neuroscience* 5(1), pp. 81–90. doi: 10.1016/0306-4522(80)90073-1.

Feldman, A. G. (1980b). "Superposition of motor programs—II. Rapid forearm flexion in man". *Neuroscience* 5(1), pp. 91–95. doi: 10.1016/0306-4522(80)90074-3.

Guenther, Frank H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production". *Psychological Review* 102(3), pp. 594–621. doi: 10.1037//0033-295X.102.3.594.

Hashi, Michiko, John R. Westbury, and Kiyoshi Honda (1998). "Vowel posture normalization". *The Journal of the Acoustical Society of America* 104(4), pp. 2426–2437. doi: 10.1121/1.423750.

Hooper, Joan (1978). "Constraints on schwa-deletion in American English". *Recent Developments in Historical Phonology*. Ed. by Jacek Fisiak. The Hague, Netherlands: Mouton, pp. 183–207.

Huys, Raoul, Breanna E. Studenka, Nicole L. Rheaume, Howard N. Zelaznik, and Viktor K. Jirsa (2008). "Distinct timing mechanisms produce discrete and continuous movements". *PLoS Computational Biology* 4(4). doi: 10.1371/journal.pcbi.1000061.

Jirsa, Viktor K. and J. A. Scott Kelso (2005). "The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation". *Journal of Motor Behavior* 37(1), pp. 35–51. doi: 10.3200/JMBR.37.1.35-51.

Kay, B. A., J. A. Scott Kelso, Elliot L. Saltzman, and Gregor Schöner (1987). "Space-time behavior of single and bimanual rhythmical movements: data and limit cycle model". *Journal of Experimental Psychology: Human Perception & Performance* 13(2), pp. 178–192. doi: 10.1037//0096-1523.13.2.178.

Kelso, J. A. Scott (1977). "Motor control mechanisms underlying human movement reproduction". *Journal of Experimental Psychology: Human Perception and Performance* 3(4), pp. 529–543. doi: 10.1037/0096-1523.3.4.529.

Kelso, J. A. Scott and Betty Tuller (1984). "Converging evidence in support of common dynamical principles for speech and movement coordination". *The American journal of physiology* 246(6), R928–R935. doi: 10.1152/ajpregu.1984.246.6.R928.

Neufeld, Chris and Pascal van Lieshout (2014). "Tongue kinematics in palate relative coordinate spaces for electro-magnetic articulography". *The Journal of the Acoustical Society of America* 135(1), pp. 352–361. doi: 10.1121/1.4836515.

Perrier, Pascal, David J. Ostry, and Rafael Laboissière (1996). "The equilibrium point hypothesis and its application to speech motor control". *Journal of Speech, Language, and Hearing Research* 39(2), pp. 365–378. doi: 10.1044/jshr.3902.365.

Polit, Andres and Emilio Bizzi (1978). "Processes controlling arm movements in monkeys". *Science* 201(4362), pp. 1235–1237. doi: 10.1126/science.99813.

Polit, Andres and Emilio Bizzi (1979). "Characteristics of motor programs underlying arm movements in monkeys". *Journal of Neurophysiology* 42(1), pp. 183–194. doi: 10.1152/jn.1979.42.1.183.

Saltzman, Elliot L. and Kevin G. Munhall (1989). “A dynamical approach to gestural patterning in speech production”. *Ecological Psychology* 1(4), pp. 333–382. doi: 10.1207/s15326969eco0104_2.
Stetson, Raymond Herbert (1951). *Motor Phonetics: A Study of Speech Movements in Action*. 2nd ed. Amsterdam, Netherlands: North-Holland. isbn: 0942344022.
Tuller, Betty and J. A. Scott Kelso (1990). “Phase transitions in speech production and their perceptual consequences”. *Attention and Performance XIII: motor representation and control*. Ed. by Marc Jeannerod. Hillsday, United States: Lawrence Erlbaum, pp. 429–452.
Tuller, Betty and J. A. Scott Kelso (1991). “The production and perception of syllable structure”. *Journal of Speech, Language, and Hearing Research* 34(3), pp. 501–508. doi: 10.1044/jshr.3403.501.
Zwicky, Arnold (1972). “Note on a phonological hierarchy in English”. *Linguistic Change and Generative Theory*. Ed. by Robert P. Stockwell and Ronald K. Macaulay. Bloomington, United States: Indiana University, pp. 275–301.

Conclusion

The present work introduced a metronome-driven speech elicitation paradigm in which participants were asked to utter repetitive sequences of simple consonant-vowel syllables. This paradigm, explicitly designed to cover speech rates from a substantially wider range than has been explored so far in previous work, was demonstrated to satisfy the important prerequisites for assessing so far difficult to access aspects of speech. Specifically, the paradigm's extensive speech rate manipulation enabled elicitation of a great range of movement speeds (see Research Article I) as well as movement durations and excursions of the relevant effectors (see Research Artile II). The presence of such variation is a prerequisite to assessing whether invariant relations between these and other parameters exist and thus provides the foundation for a rigorous evaluation of the two laws examined in the first two contributions of this work.

In the data resulting from this paradigm, it was shown that speech movements obey the same fundamental laws just as movements from other domains of motor control do. In particular, it was demonstrated that speech strongly adheres to the power law relation between speed and curvature of movement with a clear speech rate dependency of the power law's exponent. The often-sought or reported exponent of one third in the statement of the law is unique to a subclass of movements which corresponds to the range of faster rates under which a particular utterance is produced. For slower rates, significantly larger values than one third were observed. Furthermore, for the first time in speech this work uncovered evidence for the presence of Fitts' law. It was shown that, beyond a speaker-specific speech rate, speech movements of the tongue clearly obey Fitts' law by emergence of its characteristic linear relation between movement time and index of difficulty. For slower speech rates (when temporal pressure is small), no such relation was observed. Beyond that, by the extensive speech rate manipulation in the herein used experimental paradigm it was demonstrated that speech exhibits clearly distinct dynamical organizations underlying the production of simple utterances (Research Article III). Specifically, at the slower rates, the dynamical organization underlying the repetitive production of elementary /CV/ syllables can be described by successive concatenations of closing and opening gestures, each with its own equilibrium point. As speech rate increases, the equilibria of opening and closing gestures are not equally stable yielding qualitatively different modes of organization with either a single equilibrium point of a combined opening-closing gesture or a periodic attractor unleashed by the disappearance of both equilibria. All in all, these results both support and question some of the previous assumptions about the principles organizing and constraining speech. What follows now is a discussion of these results in the context of other work in movement science, considering both commonalities across domains and prospects for furthering the understanding of speech.

All modern theories of language rely on a fundamental segmental hypothesis according to which the phonological message of an utterance is represented by a sequence of segments or phonemes (Chomsky and Halle, 1968). It is commonly assumed that each of these phonemes can be mapped to some unit of speech motor action, a so-called speech gesture. A long line of work has been proceeded to identify these discrete units of action in what naturally is observed as the continuous motion of the articulatory organs (see e.g., Fowler, Rubin, et al., 1980; Browman and Goldstein, 1986; Saltzman and Munhall, 1989 in their theory of Articulatory Phonology). Supported by technological advances over the last three decades enabling rigorous observation and quantification of vocal tract action, the success of these efforts in offering ways to characterize such units by considerations of the basic kinematic properties of speech has led to the continuation of an undisputed standing of the segmental hypothesis today. Accordingly, a large class of contemporary dynamical models of speech production describe speech as a sequence of concatenated speech tasks, each task with a target represented by a point attractor in some appropriately defined task state space (Saltzman and Munhall, 1989; Perrier, Ostry, et al., 1996; Guenther, 1995). In other words, speech movements are considered to be what in general movement science is referred to as discrete, point-to-point movements in a sequential order of targets reflected in the order of phonemes in the utterance. However, discrete movements are just one possible type of movement. Hogan and Sternad (2007) presented theoretical considerations for a taxonomy of human motion patterns with discrete and rhythmic behavior as the two major taxa (that is, rhythmic movements cannot be reduced to concatenations of discrete movements, and vice versa). Empirical evidence for such a clear dichotomy between discrete and rhythmic patterns has been reported by Schaal, Sternad, et al. (2004), in a neurophysiological study on arm motor control, and by Ahn and Hogan (2012) who argue on the basis of gait experiments with mechanical perturbations that human locomotion exhibits distinctive features of a nonlinear oscillator and is not organized as in e.g., discrete reaching. More recently, Hervault, Huys, et al. (2019) have shown that for finger and hand movements the motor inhibition processes involved in cancelling prepared discrete movements and stopping ongoing rhythmic movements are dissimilar. The question of whether speech movements adhere to the same or a similar classification scheme is imminent in face of the here presented evidence that speech obeys the same fundamental motor laws as movements of the upper and lower limbs.

The present work's assessment of the speed-curvature power law indicated evidence for the presence of non-discrete, non-point-to-point motor behavior in speech. Recall the assessment's demonstration of a systematic, rate-dependent variation of the power law's exponent with values close to one third for movements at the faster rates but significantly larger values at the slower rates. In Pollick and Ishimura (1996), larger than one third exponents were reported for discrete, point-to-point hand movements and attributed to differences in task constraints compared to the commonly studied rhythmic and path-constrained actions of hand drawing with an exponent close to one third. This raises then the possibility that target-constrained, point-to-point movements differ from rhythmic, path-constrained movements with respect to the precise instantiation of the speed-curvature power law (its exponent value). However, values of the power law's exponent increase smoothly with decreasing metronome rate. That is, for the present data, the exponent itself does not allow for a strict dissociation between these two kind of movements as generally conjectured by Hogan and Sternad (2007).

It remains to be seen whether there are other properties of the power law (e.g., the velocity gain factor) which would allow for a classification along the lines of the proposed taxonomy of human motor patterns.

The (novel) observation of non-point-to-point movements in speech raises the additional question about the mapping between the discrete phonological units and these units of motor action. As reviewed earlier, researchers in the fields of phonology and phonetics have sought to characterize the mapping between the discrete units of phonemes and some presumed to be discrete (despite being continuous in space and time) units of motor action and assumed that there exists a one-to-one relation between the two. This working hypothesis was motivated by the fact that, in most cases, for any given phoneme there is a corresponding task with an observed (physical) state of the vocal tract. For example, following traditional static descriptions of vocal tract action, the production of a [k] consonant involves a state in which the back of the tongue is in contact with the posterior region of the palate and held tightly enough to block airflow, whereas the production of an [a] vowel involves a state where the tongue has to be positioned as far as possible away from the palate. It thus seemed reasonable to identify the two phonemes of [k] and [a] with these two states of the vocal tract. Accordingly, contemporary dynamical attempts to model speech motor behavior of simple utterances, as in repeated /CV/ syllables, assume that there are two (in general unrelated) gestures to consider. These are a closing gesture forming the constriction associated with the consonant /C/ and an opening gesture for releasing the constriction and moving the involved articulators in a position required by the vowel /V/. Each of these gestures is taken to be governed by a dynamical system defined in some appropriate task space with an equilibrium point representing the gesture's target (e.g., the state of the vocal tract in case of the Task Dynamics by Saltzman and Munhall, 1989). As a result, a sequence of repeated /CV/ syllables is modeled by the succession of two concatenated point-to-point gestures where the two involved endpoints corresponds to equilibria in task space. For example, the repetition of [ka] syllables is modeled as a [k]-to-[a] transitions followed by a [a]-to-[k] transition, then repeat; [k] and [a] correspond to the two equilibria of the task space. Despite the popularity of equilibrium point models of speech, the existence of such equilibria has not heretofore been rigorously assessed or verified.

The third assessment of this work addressed the search for the presence of equilibrium points in the task space (as conceived by the most fleshed out approach to date, the Task Dynamics of Saltzman and Munhall, 1989) with speech production data from repeated [ka] syllables at different speech rates. It was shown that, at slower speech rates, the dynamical organization of the task can be described as a successive concatenation of opening and closing gestures with two equilibrium points corresponding to the phonemic targets [k] and [a]. This result supports the above sketched equilibrium point hypothesis to the extent that the simple phonemic sequence of repeated [ka] syllables can be decomposed into a succession of [k]s and [a]s with each of these phonemes being mapped to a point-to-point gesture (either [k]-to-[a] or [a]-to-[k]). Nonetheless, the results also revealed that the two observed equilibrium points of [k] and [a] are not equally stable. As speech rate increases, the equilibrium point associated with the vowel disappears leaving only an equilibrium point for the consonant. The observed action is best described as repetitive [k]-to-[k] transitions traversing the vocal tract state of [a]. This is so even though the auditory representation of such sequences corresponds to a repeated [ka] sequence. Yet the presented analysis shows that there is no equilibrium point for the vowel [a]. This observation is inconsistent with the current understanding of how phonemic sequences are mapped to the level of motor behavior (for each phoneme there is a distinct unit of action) as well as with contemporary modeling approaches (gestures are modeled by their phonemic endpoints; there is no [k]-to-[k] passing [a] gesture possible in such approaches). Moreover, for some speakers, also the remaining equilibrium point of the consonant dissolves as rate increases giving rise to a periodic attractor. The resulting organization can no longer be described in terms of the constructs of the current equilibrium point models; it is neither a [k]-to-[k] transition with an intermediate [a] nor a [a]-to-[a] transition with an intermediate [k]. Rather, the organization is best described as one cycle of repeated movement patterns that happens to pass through the states of [k] and [a]. Note that this non-uniqueness in description does not affect the perception of the utterance. These (fast) syllables are clearly perceived as ordinary repetitions of [ka]. That is, there is no change from [ka ka ka. . .] to [ak ak ak. . .] or some such similar change along the lines of the phenomena studied by Stetson (1951) and later Tuller and Kelso (1991, 1990) on changes from [ip ip ip…] to [pi pi pi…] by increasing rate of speech. Instead, the novel observation here is concerned with changes in the same utterance [ka ka ka…] without reordering of the phonemes.

This last observation, the non-uniqueness of the dynamical organization underlying what on the surface appear to be identical phonemic sequences, is an entirely new result in the domain of speech. Beyond that, the demonstration of periodic attractors in speech reveals that dynamical equilibrium point models do not account for all possible modes of speech motor behavior.

Bibliography

Abraham, Ralph H. and Christopher D. Shaw (1983). *Dynamics: the geometry of behavior*. Santa Cruz, United States: Aerial Press. isbn: 0942344022.

Ahn, Jooeun and Neville Hogan (2012). "Walking is not like reaching: evidence from periodic mechanical perturbations". *PLOS One* 7.3. doi: 10.1371/journal.pone.0031767.

Anscombe, Francis John (1973). "Graphs in statistical analysis". *The American Statistician* 27(1), pp. 17–21. doi: 10.2307/2682899.

Baer, Matthias (2018). *A Python package for finite difference derivatives in any number of dimensions [Computer software]*. Retrieved from Python Package Index, https://pypi.org/project/findiff/.

Beckman, Mary E., Tzyy-Ping Jung, Sook-hyang Lee, Kenneth J. de Jong, Ashok K. Krishnamurthy, Stanley C. Ahalt, K. Bretonnel Cohen, and Michael J. Collins (1995). "Variability in the production of quantal vowels revisited". *The Journal of the Acoustical Society of America* 97(1), pp. 471–490. doi: 10.1121/1.412945.

Bernstein, Nikolai A. (1967). *The co-ordination and regulation of movements*. Oxford, United Kingdom: Pergamon Press.

Bootsma, R. J., L. Fernandez, and D.Mottet (2004). "Behind Fitts' law: kinematic patterns in goal-directed movements". *International Journal of Human-Computer Studies* 61(6), pp. 811–821. doi: 10.1016/j.ijhcs.2004.09.004.

Browman, Catherine P. and Louis M. Goldstein (1986). "Towards an articulatory phonology". *Phonology* 3(1), pp. 219–252. doi: 10.1017/S0952675700000658.

Card, Stuart K., Thomas P. Moran, and Allen Newell (1983). *The psychology of human-computer interaction*. Hillsday, United States: Lawrence Erlbaum Associates. isbn: 0898592437.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington, United Kingdom: Indiana University Press.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. New York, United States: Harper & Row. isbn: 0262031795.

Clark, Mary Ann, Alma S. Merians, Abhay Kothari, Howard Poizner, Beth Macauley, Leslie J. Gonzalez Rothi, and Kenneth M. Heilman (1994). "Spatial Planning deficits in limb apraxia". *Brain* 117(5), pp. 1093–1106. doi: 10.1093/brain/117.5.1093.

Cohen, Ayala (1983). "Comparing regression coefficients across subsamples. A study of the statistical test". *Sociological Methods & Research* 12(1), pp. 77–94. doi: 10.1177/0049124183012001003.

Connelly, Edward M. (1984). “A control model: interpretation of Fitts’ law”. *Twentieth Annual Conference on Manual Control* (Moffett Field, United States). Vol. 1, pp. 621–642.

Crossman, E. R. F. W. and P. J. Goodeve (1983). “Feedback control of hand-movement and Fitts’ law”. *Quarterly Journal of Experimental Psychology* 35(2). Paper presented at the meeting of the Experimental Psychology Society, Oxford, 1963, pp. 251–278. doi: 10.1080/14640748308402133.

Dalby, Jonathan (1986). “Phonetic structure of fast speech in American English”. *Phonetics and Phonology*. Bloomington, United States: Indiana University.

Dayan, Eran, Antonino Casile, Nava Levit-Binnun, Martin A. Giese, Talma Hendler, and Tamar Flash (2007). “Neural representations of kinematic laws of motion: evidence for action-perception coupling”. *Proceedings of the National Academy of Sciences of the United States of America* 104(51), pp. 20582–20587. doi: 10.1073/pnas.0710033104.

De Jong, Kenneth J. (2001). “Rate-induced resyllabification revisited”. *Language and Speech* 44(2), pp. 197–216. doi: 10.1177/00238309010440020401.

De Jong, Kenneth J., Byung-jin Lim, and Kyoko Nagao (2004). “The perception of syllable affiliation of singleton stops in repetitive speech”. *Language and Speech* 47(3), pp. 241–266. doi: 10.1177/00238309040470030201.

de’Sperati, Claudio and Paolo Viviani (1997). “The relationship between curvature and velocity in two-dimensional smooth pursuit eye movements”. *The Journal of Neuroscience* 17(10), pp. 3932–3945.

Do Carmo, Manfredo Perdigão (1976). *Differential geometry of curves and surfaces*. New Jersey, United States: Prentice-Hall. isbn: 0132125897.

Feldman, A. G. (1980a). “Superposition of motor programs—I. Rhythmic forearm movements in man”. *Neuroscience* 5(1), pp. 81–90. doi: 10.1016/0306-4522(80)90073-1.

Feldman, A. G. (1980b). “Superposition of motor programs—II. Rapid forearm flexion in man”. *Neuroscience* 5(1), pp. 91–95. doi: 10.1016/0306-4522(80)90074-3.

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

Fitts, Paul M. and James R. Peterson (1964). “Information capacity of discrete motor responses”. *Journal of Experimental Psychology* 67(2), pp. 103–112. doi: 10.1037/h0045689.

Fowler, Carol A. (1980). “Coarticulation and theories of extrinsic timing”. *Journal of Phonetics* 8(1), pp. 113–133.

Fowler, Carol A., Philip Rubin, Robert E. Remez, and Michael T. Turvey (1980). “Implications for speech production of a general theory of action”. *Language Production: Speech and Talk*. Ed. by Brian Butterworth. New York, United States: Academic Press, pp. 373–420.

Fuchs, Susanne and Pascal Perrier (2005). “On the complex nature of speech kinematics”. *ZAS Papers in Linguistics* 42, pp. 137–165.

Fuchs, Susanne, Pascal Perrier, and Mariam Hartinger (2011). “A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model”. *Journal of Speech, Language, and Hearing Research* 54(4), pp. 1067–1076. doi: 10.1044/1092-4388(2010/10-0131).

Gribble, Paul L. and David J. Ostry (1996). "Origins of the power law relation between movement velocity and curvature: modeling the effects of muscle mechanics and limb dynamics". *Journal of Neurophysiology* 76(5), pp. 2853–2860. doi: 10.1152/jn.1996.76.5.2853.

Grosjean, Marc, Maggie Shiffrar, and Günther Knoblich (2007). "Fitts's Law Holds for Action Perception". *Psychological Science* 18(2), pp. 95–99. doi: 10.1111/j.1467-9280.2007.01854.x.

Guenther, Frank H. (1995). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production". *Psychological Review* 102(3), pp. 594–621. doi: 10.1037//0033-295X.102.3.594.

Hashi, Michiko, John R. Westbury, and Kiyoshi Honda (1998). "Vowel posture normalization". *The Journal of the Acoustical Society of America* 104(4), pp. 2426–2437. doi: 10.1121/1.423750.

Hatsopoulos, Nicholas G. and William H. Warren Jr. (1996). "Resonance tuning in rhythmic arm movements". *Journal of Motor Behavior* 28(1), pp. 3–14. doi: 10.1080/00222895.1996.9941728.

Heglund, Norman C., C. Richard Taylor, and Thomas A. McMahon (1974). "Scaling stride frequency and gait to animal size: mice to horses". *Science* 186(4169), pp. 1112–1113. doi: 10.1126/science.186.4169.1112.

Hervault, Mario, Raoul Huys, Chloe Farrer, Jean Christophe Buisson, and Pier-Giorgio Zanone (2019). "Cancelling discrete and stopping ongoing rhythmic movements: do they involve the same process of motor inhibition?" *Human Movement Science* 64, pp. 296–306. doi: 10.1016/j.humov.2019.02.010.

Hicheur, Halim, Stéphane Vieilledent, Magnus J. E. Richardson, Tamar Flash, and Alain Berthoz (2005). "Velocity and curvature in human locomotion along complex curved paths: a comparison with hand movements". *Experimental Brain Research* 162(2), pp. 145–154. doi: 10.1007/s00221-004-2122-8.

Hogan, Neville and Dagmar Sternad (2007). "On rhythmic and discrete movements: reflections, definitions and implications for motor control". *Experimental Brain Research* 181(1), pp. 13–30. doi: 10.1007/s00221-007-0899-y.

Hoole, Philip, Christine Mooshammer, and Hans G. Tillmann (1994). "Kinematic analysis of vowel production in German". *The 3rd International Conference on Spoken Language Processing* (Yokohama, Japan), pp. 53–56.

Hooper, Joan (1978). "Constraints on schwa-deletion in American English". *Recent Developments in Historical Phonology*. Ed. by Jacek Fisiak. The Hague, Netherlands: Mouton, pp. 183–207.

Horn, Berthold K. P. (1987). "Closed-form solution of absolute orientation using unit quaternions". *Journal of the Optical Society of America A* 4(4), pp. 629–642. doi: 10.1364/JOSAA.4.000629.

Huh, Dongsung and Terrence J. Sejnowski (2015). "Spectrum of power laws for curved hand movements". *Proceedings of the National Academy of Sciences of the United States of America* 112(29), E3950–E3958. doi: 10.1073/pnas.1510208112.

Huys, Raoul, Breanna E. Studenka, Nicole L. Rheaume, Howard N. Zelaznik, and Viktor K. Jirsa (2008). "Distinct timing mechanisms produce discrete and continuous movements". *PLoS Computational Biology* 4(4). doi: 10.1371/journal.pcbi.1000061.

Jirsa, Viktor K. and J. A. Scott Kelso (2005). “The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation”. *Journal of Motor Behavior* 37(1), pp. 35–51. doi: 10.3200/JMBR.37.1.35-51.

Kay, B. A., J. A. Scott Kelso, Elliot L. Saltzman, and Gregor Schöner (1987). “Space-time behavior of single and bimanual rhythmical movements: data and limit cycle model”. *Journal of Experimental Psychology: Human Perception & Performance* 13(2), pp. 178–192. doi: 10.1037//0096-1523.13.2.178.

Kelso, J. A. Scott (1977). “Motor control mechanisms underlying human movement reproduction”. *Journal of Experimental Psychology: Human Perception and Performance* 3(4), pp. 529–543. doi: 10.1037/0096-1523.3.4.529.

Kelso, J. A. Scott, Elliot L. Saltzman, and Betty Tuller (1986). “The dynamical perspective on speech production: data and theory”. *Journal of Phonetics* 14, pp. 29–59.

Kelso, J. A. Scott and Betty Tuller (1984). “Converging evidence in support of common dynamical principles for speech and movement coordination”. *The American journal of physiology* 246(6), R928–R935. doi: 10.1152/ajpregu.1984.246.6.R928.

Kelso, J. A. Scott, Betty Tuller, and Katherine S. Harris (1983). “A “dynamic pattern” perspective on the control and coordination of movement”. *The Production of Speech*. Ed. by Peter F. MacNeilage. New York, United States: Springer, pp. 137–173. doi: 10.1007/978-1-4613-8202-7_7.

Kelso, J. A. Scott, Eric Vatikiotis-Bateson, Elliot L. Saltzman, and Bruce Kay (1985). “A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling”. *The Journal of the Acoustical Society of America* 77(1), pp. 266–280. doi: 10.1121/1.392268.

Kier, William M. and Kathleen K. Smith (1985). “Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats”. *Zoological Journal of the Linnean Society* 83(4), pp. 307–324. doi: 10.1111/j.1096-3642.1985.tb01178.x.

Kröger, Bernd J., Georg Schröder, and Claudia Opgen-Rhein (1995). “A gesture-based dynamic model describing articulatory movement data”. *The Journal of the Acoustical Society of America* 98(4), pp. 1878–1889. doi: 10.1121/1.413374.

Kuehn, David P. and Kenneth L. Moll (1976). “A cineradiographic study of VC and CV articulatory velocities”. *Journal of Phonetics* 4(4), pp. 303–320.

Lacquaniti, Francesco, Carlo Terzuolo, and Paolo Viviani (1983). “The law relating the kinematic and figural aspects of drawing movements”. *Acta Psychologica* 54(1–3), pp. 115–130. doi: 10.1016/0001-6918(83)90027-6.

Lammert, Adam C., Christine H. Shadle, Shrikanth S. Narayanan, and Thomas F. Quatieri (2018). “Speed-accuracy tradeoffs in human speech production”. *PLOS One* 13(9). doi: 10.1371/journal.pone.0202180.

Levine, William S., Caroline Essex Torcaso, and Maureen Stone (2005). “Controlling the Shape of a Muscular Hydrostat: A Tongue or Tentacle”. *New Directions and Applications in Control Theory*. Ed. by Wijesuriya P. Dayawansa, Anders Lindquist, and Yishao Zhou. Vol. 321. Lecture Notes in Control and Information Sciences. Springer, Berlin, Heidelberg. Chap. 13, pp. 207–222.

MacKenzie, C. L., R. G. Marteniuk, C. Dugas, D. Liske, and B. Eickmeier (1987). "Three-dimensional movement trajectories in Fitts' task: implications for control". *Quarterly Journal of Experimental Psychology* 39(4), pp. 629–647. doi: 10.1080/14640748708401806.

MacKenzie, Ian Scott (1992). "Fitts' law as a research and design tool in human-computer interaction". *Human-Computer Interaction* 7(1), pp. 91–139. doi: 10.1207/s15327051hci0701_3.

MacKenzie, Ian Scott (2013). "A note on the validity of the Shannon formulation for Fitts' index of difficulty". *Open Journal of Applied Sciences* 3(6), pp. 360–368. doi: 10.4236/ojapps.2013.36046.

Maoz, Uri, Elon Portugaly, Tamar Flash, and Yair Weiss (2006). "Noise and the two-thirds power law". *Advances in Neural Information Processing Systems 18 NIPS* (Vancouver, Canada). Ed. by Yair Weiss, Bernhard Schölkopf, and John C. Platt, pp. 851–858.

Marken, Richard S. and Dennis M. Shaffer (2017). "The power law of movement: an example of a behavioral illusion". *Experimental Brain Research* 235(6), pp. 1835–1842. doi: 10.1007/s00221-017-4939-y.

Munhall, Kevin G., David J. Ostry, and Avraham Parush (1985). "Characteristics of velocity profiles of speech movements". *Journal of Experimental Psychology: Human Perception and Performance* 11(4), pp. 457–474. doi: 10.1037//0096-1523.11.4.457.

Nelson, Winston L. (1983). "Physical principles for economies of skilled movements". *Biological Cybernetics* 46(2), pp. 135–147. doi: 10.1007/BF00339982.

Nelson, Winston L., Joseph S. Perkell, and John R. Westbury (1984). "Mandible movements during increasingly rapid articulations of single syllables: preliminary observations". *The Journal of the Acoustical Society of America* 75(3), pp. 945–951. doi: 10.1121/1.390559.

Neufeld, Chris and Pascal van Lieshout (2014). "Tongue kinematics in palate relative coordinate spaces for electro-magnetic articulography". *The Journal of the Acoustical Society of America* 135(1), pp. 352–361. doi: 10.1121/1.4836515.

Opstal, A. J. Van and J.A.M. Van Gisbergen (1987). "Skewness of saccadic velocity profiles: a unifying parameter for normal and slow saccades". *Vision Research* 27(5), pp. 731–745. doi: 10.1016/0042-6989(87)90071-X.

Ostry, David J., James D. Cooke, and Kevin G. Munhall (1987). "Velocity curves of human arm and speech movements". *Experimental Brain Research* 68(1), pp. 37–46. doi: 10.1007/BF00255232.

Ostry, David J., Eric Keller, and Avraham Parush (1983). "Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movement in speech". *Journal of Experimental Psychology: Human Perception and Performance* 9(4), pp. 622–636. doi: 10.1037//0096-1523.9.4.622.

Ostry, David J. and Kevin G. Munhall (1985). "Control of rate and duration of speech movements". *The Journal of the Acoustical Society of America* 77(2), pp. 640–648. doi: 10.1121/1.391882.

Patel, Aniruddh D., Anders Löfqvist, and Walter Naito (1999). "The acoustics and kinematics of regularly timed speech: a database and method for the study of the p-center problem". *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco, United States). Vol. 1, pp. 405–408.

Perrier, Pascal and Susanne Fuchs (2008). "Speed-curvature relations in speech production challenge the 1/3 power law". *Journal of Neurophysiology* 100(3), pp. 1171–1183. doi: 10.1152/jn.01116.2007.

Perrier, Pascal, David J. Ostry, and Rafael Laboissière (1996). "The equilibrium point hypothesis and its application to speech motor control". *Journal of Speech, Language, and Hearing Research* 39(2), pp. 365–378. doi: 10.1044/jshr.3902.365.

Plamondon, Réjean and Adel M. Alimi (1997). "Speed/accuracy trade-offs in target-directed movements". *Behavioral and Brain Sciences* 20(2), pp. 279–303. doi: 10.1017/S0140525X97001441.

Poizner, Howard, Mary Ann Clark, Alma S. Merians, Beth Macauley, Leslie J. Gonzalez Rothi, and Kenneth M. Heilman (1995). "Joint coordination deficits in limb apraxia". *Brain* 118(1), pp. 227–242. doi: 10.1093/brain/118.1.227.

Poizner, Howard, Linda Mack, Mieka Verfaellie, Leslie J. Gonzales Rothi, and Kenneth M. Heilman (1990). "Three-dimensional computergraphic analysis of apraxia". *Brain* 113(1), pp. 85–101. doi: 10.1093/brain/113.1.85.

Polit, Andres and Emilio Bizzi (1978). "Processes controlling arm movements in monkeys". *Science* 201(4362), pp. 1235–1237. doi: 10.1126/science.99813.

Polit, Andres and Emilio Bizzi (1979). "Characteristics of motor programs underlying arm movements in monkeys". *Journal of Neurophysiology* 42(1), pp. 183–194. doi: 10.1152/jn.1979.42.1.183.

Pollick, Frank E. and Gensel Ishimura (1996). "The three-dimensional curvature of straight-ahead movements". *Journal of Motor Behavior* 28(3), pp. 271–279. doi: 10.1080/00222895.1996.9941751.

Ramsay, James and Bernhard W. Silverman (2005). *Functional data analysis*. Springer Series in Statistics. New York, United States: Springer-Verlag. doi: 10.1007/b98888.

Rothi, Leslie J. Gonzalez and Kenneth M. Heilman (1994). *Apraxia. The neuropsychology of action*. Psychology Press.

Ruppert, David (2011). *Statistics and data analysis for financial engineering*. Springer Texts in Statistics. New York, United States: Springer-Verlag. doi: 10.1007/978-1-4419-7787-8.

Saltzman, Elliot L. (1986). "Task dynamic coordination of the speech articulators: a preliminary model". *Generation and modulation of action patterns*. Ed. by Herbert Heuer and Christoph Fromm. Vol. 15. Experimental Brain Research Series. New York, United States: Springer, pp. 129–144.

Saltzman, Elliot L. and Kevin G. Munhall (1989). "A dynamical approach to gestural patterning in speech production". *Ecological Psychology* 1(4), pp. 333–382. doi: 10.1207/s15326969eco0104_2.

Schaal, Stefan and Dagmar Sternad (2001). "Origins and violations of the 2/3 power law in rhythmic three-dimensional arm movements". *Experimental Brain Research* 136(1), pp. 60–72. doi: 10.1007/s002210000505.

Schaal, Stefan, Dagmar Sternad, Rieko Osu, and Mitsuo Kawato (2004). "Rhythmic arm movement is not discrete". *Nature Neuroscience* 7(10), pp. 1136–1143. doi: 10.1038/nn1322.

Schmidt, Richard and Tim Lee (2011). *Motor control and learning: a behavioral emphasis*. Champaign, United States: Human Kinetics. isbn: 0736079610.

Schöner, Gregor (1990). "A dynamic theory of coordination of discrete movement". *Biological Cybernetics* 63(4), pp. 257–270. doi: 10.1007/BF00203449.

Schwartz, Andrew B. (1994). “Direct cortical representation of drawing”. *Science* 265(5171), pp. 540–542. doi: 10.1126/science.8036499.

Sorensen, Tanner and Adamantios I. Gafos (2016). “The gesture as an autonomous nonlinear dynamical system”. *Ecological Psychology* 28(4), pp. 188–215. doi: 10.1080/10407413.2016.1230368.

Stetson, Raymond Herbert (1951). *Motor Phonetics: A Study of Speech Movements in Action*. 2nd ed. Amsterdam, Netherlands: North-Holland. isbn: 0942344022.

Stone, Meghan, A. Faber, L. J. Raphael, and T. H. Shawker (1992). “Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]”. *Journal of Phonetics* 20(2), pp. 253–270.

Tasko, Stephen M. and John R. Westbury (2004). “Speed-curvature relations for speech-related articulatory movement”. *Journal of Phonetics* 32(1), pp. 65–80. doi: 10.1016/S0095-4470(03)00006-8.

Tomaschek, Fabian, Dennis Arnold, Franziska Bröker, and R. Harald Baayen (2018). “Lexical frequency co-determines the speed-curvature relation in articulation”. *Journal of Phonetics*, pp. 103–116. doi: 10.1016/j.wocn.2018.02.003.

Tuller, Betty and J. A. Scott Kelso (1990). “Phase transitions in speech production and their perceptual consequences”. *Attention and Performance XIII: motor representation and control*. Ed. by Marc Jeannerod. Hillsday, United States: Lawrence Erlbaum, pp. 429–452.

Tuller, Betty and J. A. Scott Kelso (1991). “The production and perception of syllable structure”. *Journal of Speech, Language, and Hearing Research* 34(3), pp. 501–508. doi: 10.1044/jshr.3403.501.

Turvey, Michael T. (1977). “Preliminaries to a theory of action with reference to vision”. *Perceiving, acting and knowing: toward an ecological Psychology*. Ed. by R. Shaw and J. Bransford. Lawrence Erlbaum Associates, pp. 211–265.

Turvey, Michael T. (2007). “Action and perception at the level of synergies”. *Human Movement Science* 26(4), pp. 657–697. doi: 10.1016/j.humov.2007.04.002.

Vatikiotis-Bateson, Eric and J. A. Scott Kelso (1990). “Linguistic structure and articulatory dynamics: a cross language study”. *Haskins Laboratories Status Report in Speech Research* SR-103/104, pp. 67–94.

Vatikiotis-Bateson, Eric and J. A. Scott Kelso (1993). “Rhythm type and articulatory dynamics in English, French and Japanese”. *Journal of Phonetics* 21, pp. 231–265.

Viviani, Paolo (2002). “Motor competence in the perception of dynamic events: a tutorial”. *Common mechanisms in perception and action: attention and performance XIX*. Ed. by Wolfgang Prinz and Bernhard Hommel. Oxford, United Kingdom: Oxford University Press, pp. 406–442.

Viviani, Paolo and Marco Cenzato (1985). “Segmentation and coupling in complex movements”. *Journal of Experimental Psychology: Human Perception and Performance* 11(6), pp. 828–845. doi: 10.1037/0096-1523.11.6.828.

Viviani, Paolo and Tamar Flash (1995). “Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning”. *Journal of Experimental Psychology: Human Perception and Performance* 21(1), pp. 32–53. doi: 10.1037/0096-1523.21.1.32.

Viviani, Paolo and Natale Stucchi (1992). “Biological movements look uniform: evidence of motor-perceptual interactions”. *Journal of Experimental Psychology: Human Perception and Performance* 18(3), pp. 603–623. doi: 10.1037/0096-1523.18.3.603.

Viviani, Paolo and Carlo Terzuolo (1982). "Trajectory determines movement dynamics". *Neuroscience* 7(2), pp. 431–437. doi: 10.1016/0306-4522(82)90277-9.

Wann, John, Ian Nimmo-Smith, and Alan M. Wing (1988). "Relation between velocity and curvature in movement: equivalence and divergence between a power law and a minimum-jerk model". *Journal of Experimental Psychology: Human Perception and Performance* 14(4), pp. 622–637. doi: 10.1037//0096-1523.14.4.622.

Welford, Alan Traviss (1968). *Fundamentals of skill*. Methuen's manuals of modern psychology. London, United Kingdom: Methuen. isbn: 0416030009.

Wobbrock, Jacob O., Kristen Shinohara, and Alex Jansen (2011). "The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, Canada), pp. 1639–1648. doi: 10.1145/1978942.1979181.

Woodworth, Robert S. (1899). "The accuracy of voluntary movement". *The Psychological Review*. Vol. III (3). Series of Monograph Supplements. doi: 10.1037/h0092992.

Zago, Myrka, Adam Matic, Tamar Flash, Alex Gomez-Marin, and Francesco Lacquaniti (2017). "The speed-curvature power law of movements: a reappraisal". *Experimental Brain Research* 236(1). doi: 10.1007/s00221-017-5108-z.

Zwicky, Arnold (1972). "Note on a phonological hierarchy in English". *Linguistic Change and Generative Theory*. Ed. by Robert P. Stockwell and Ronald K. Macaulay. Bloomington, United States: Indiana University, pp. 275–301.