

# Fitts' Law in Tongue Movements of Repetitive Speech

1957

Stephan R. Kuberski Adamantios I. Gafos

Department of Linguistics and Cognitive Sciences, University of Potsdam,  
Potsdam, Germany

---

### 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 speech movement data in sequences of the form /CV.../ from 6 adult speakers. These sequences were spoken at 8 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 in a subset of our elicited rates. 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. We find no evidence for Fitts' law at the slowest metronome rates. Finally, we discuss implications of these results for models of speech.

© 2019 S. Karger AG, Basel

### Introduction

Speech is perhaps “the most highly developed motor skill possessed by all of us” (Kelso 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 widely 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

All data and source code files used to produce the results presented here are uploaded to a general-purpose repository (DOI: 10.5281/zenodo.3247110) and will be made publicly available without restriction (CC-BY) on acceptance.

of, for example, [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 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 (Browman & Goldstein, 1986; Fowler et al., 1980; Guenther, 1995; Saltzman & Munhall, 1989), 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 et al., 1984; Ostry 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

$$T = a + b ID, \quad (1)$$

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). In its original formulation, by adopting a concept from signal and information theory, the index of difficulty was defined by  $ID = \log_2(2A/W)$ , in which the ratio of twice the movement amplitude  $A$  to target width  $W$  operates as the measure of accuracy (Fitts, 1954). This definition, a simplified form of the Shannon-Hartley theorem on the information capacity of a noisy channel, is convenient but only appropriate for conditions in which the signal-to-noise ratio is large (that is, when amplitude  $A$  is much larger than target width  $W$ ). Several variations of  $ID$  have been proposed over the years (see MacKenzie, 2013, for a comprehensive discussion and comparison of these). In the present work, we will use the unmodified Shannon-Hartley formulation of Fitts' index of difficulty given by

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

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 & Alimi, 1997, and Schmidt & 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 far 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 candi-

dates 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 (for any targeted movement, its three kinematic variables are its amplitude, duration, and speed). More specifically, the parameter  $W$  which enters into the expression of the law is a task (not a kinematic) property. Leading theoretical perspectives point to the thesis that speech goals are defined in task dimensions rather than individual effector dimensions (see Saltzman & 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 a relation that involves both kinematic (duration and amplitude) and task space coordinates ( $W$ ), it captures a relational invariance which may serve as a potential entry into the principles that underlie speech. Assuming the law holds for speech, any model of speech should be able to account for it. Last but not least, Fitts' law has been shown to hold also in the perception of action. Grosjean et al. (2007) asked participants 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 et al., 1985; Ostry et al., 1987; Patel et al., 1999), we recorded via electromagnetic articulometry speech movement data in sequences of the form /CV.../ from 6 adult speakers (5 were native speakers of German and 1 was a native speaker of American English). These sequences were spoken at 8 distinct rates ranging from extremely slow (30 beats per minute, bpm) to extremely fast (570 bpm). For comparison purposes, Kelso et al. (1985) included two rates, as did Ostry et al. (1987), and Patel 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 data set, 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 with 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. First, movement amplitude  $A$  and target size  $W$  must be demonstrated to have varying values across different experimental conditions (experimental control). Second, the movements must be performed under temporal pressure (rapidness of movement). Specifically, Fitts instructed each participant "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 Lamert et al. (2018). The data were real-time magnetic resonance imaging (MRI) recordings of 5 male and 5 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 approximately 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 0.03–0.52. 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 data sets by Lammert et al. (2018). 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 real-time MRI data set was given. Presumably it was a moderate rate of common reading tasks (around 250–300 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 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 an empirically derived three-dimensional spatial articulatory target fully faithful to the dimensionality of speech action. 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. We find no evidence for Fitts' law at the slowest metronome rates. We discuss implications of these results for models of speech.

## Methods

Five native speakers of German and 1 native speaker of American English (3 females and 3 males in total) participated in the experiment. Data from another British English speaker were 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 participants.

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 began articulating the required response syllable at a point of their choice. Starting with the slowest rate, a minimum of 4 trials at each rate was 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. The entire procedure was performed in 2 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 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.

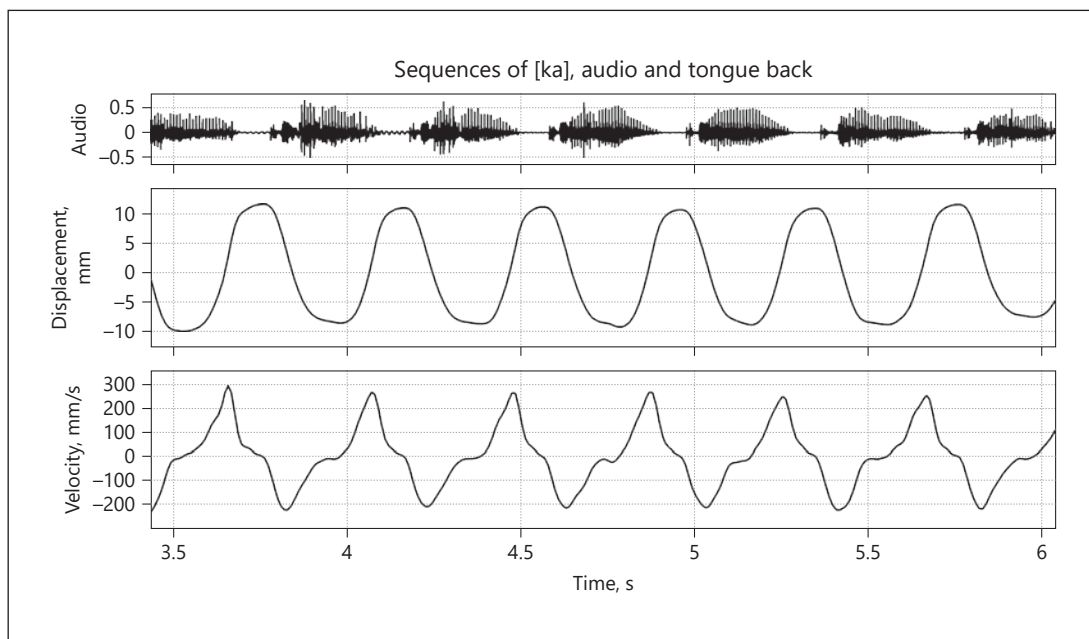
### *Data Processing*

Three-dimensional displacement data, provided by the AG501 device, were digitized at a sampling rate of 1,250 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 1,250 Hz). To avoid aliasing effects the decimation procedure implied an initial lowpass filtering using an eighth-order Chebyshev type I filter with a cut-off 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 a cut-off 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 analysis (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 et al., 1985). In total, we registered 4,314 movements in the [ta] case and 3,991 movements in the [ka] case. All data and source code files used to produce the results presented here are uploaded to a general-purpose repository (doi: 10.5281/zenodo.3247110).

## **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 centre of each of two target plates of width  $W$  using a metal stylus (Fitts, 1954; Fitts & Peterson, 1964). The quantities of movement amplitude  $A$ , corresponding to the distance between the two plates, and target size  $W$  were under 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 back 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



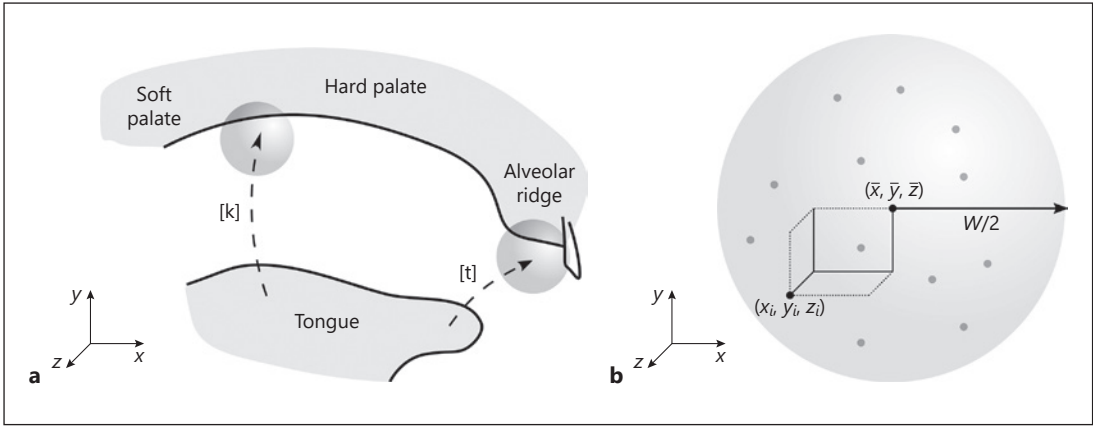
**Fig. 1.** Section of a [ka] sequence at 150 bpm. Top: acoustic recording. Middle: principal component (PCA) of tongue back displacement. Bottom: first derivative (velocity) of PCA.

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 back 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, citing an unpublished work by Crossman, p. 146) proposed an a posteriori defined target size, derived from statistical properties of the data (see MacKenzie, 1992, for an in-depth analysis of the information-theoretic background of this approach). This notion has been widely adopted in subsequent assessments in the Fitts' law literature (see e.g., Murata, 1999; Plamondon & Alimi, 1997; also, for a comparison of the effects of an effective and nominal target definition, Zhai et al., 2004). Furthermore, the usage of effective targets has been extended to assessments of other speed-accuracy trade-offs (Wright & Meyer, 1983). Specifically, for one-dimensional target extents the effective target width is defined by

$$W = \sqrt{2\pi\epsilon} \sigma, \quad (3)$$





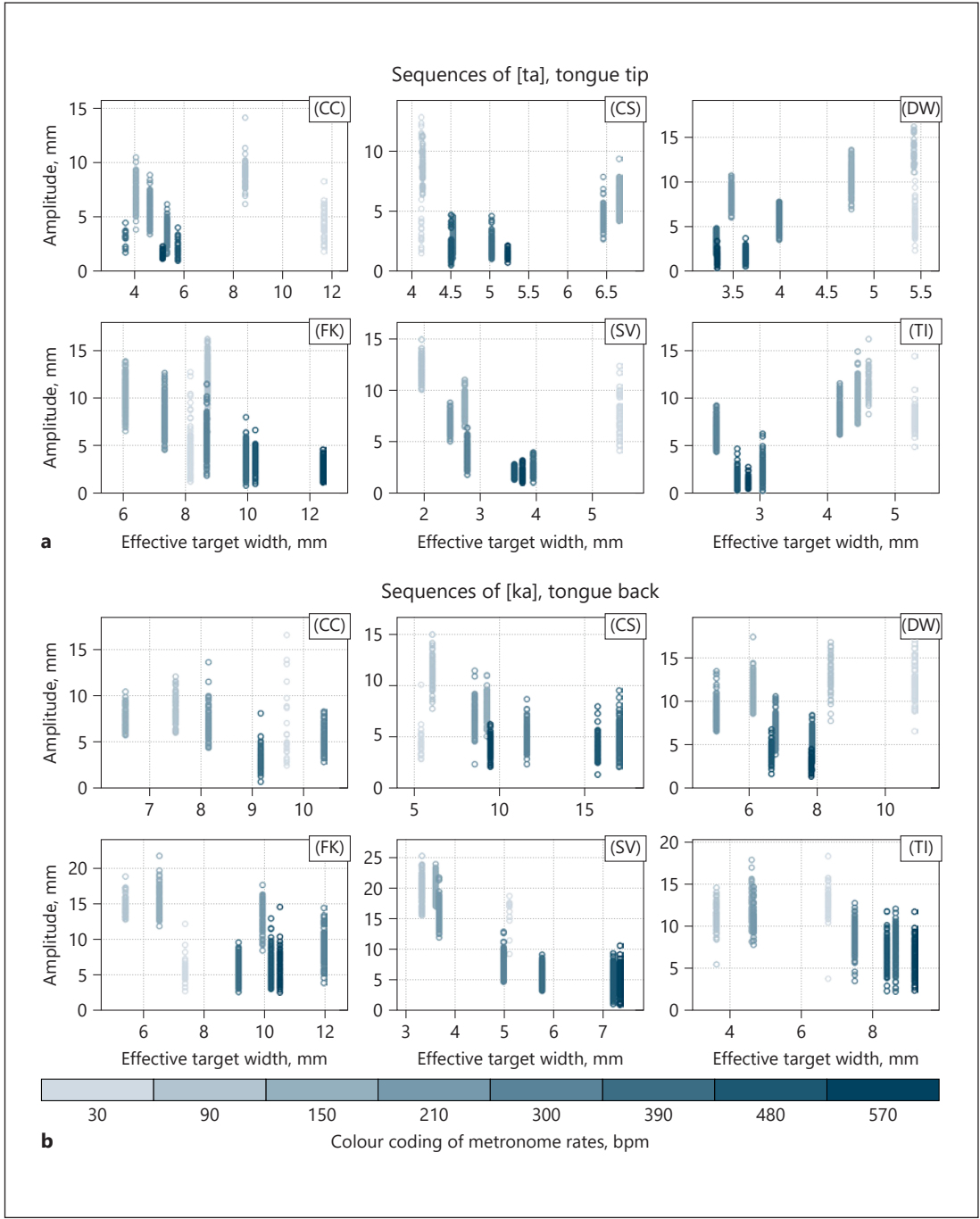
**Fig. 2.** Illustration of effective target width determination. **a** Schematic depiction of the oral cavity ( $x$ : horizontal,  $y$ : vertical,  $z$ : lateral axis). For production of the plosives [t] and [k], the tongue tip and tongue back effectors form constrictions in characteristic regions of the vocal tract, schematically indicated by the shaded spheres. For each articulatory movement (indicated by the dashed lines), movement amplitude is determined by the Euclidean distance between its onset and offset end points. **b** Schematic depiction of one of the two characteristic regions in the vocal tract. The set of all movement end points  $(x_i, y_i, z_i)$  associated with a certain plosive yields a distribution in three-dimensional space. The width  $W$  (twice the radius  $W/2$ ) of the distribution is determined by the trivariate deviation of end points around the centroid  $(\bar{x}, \bar{y}, \bar{z})$  (see text for details). Note that the trivariate deviation is not the spread of (Euclidean) distances from the centroid (which would be a univariate deviation of three-dimensional distances; see Wobbrock et al., 2011, for further details).

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

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

where  $\sigma$  now denotes the trivariate deviation of  $N$  three-dimensional end points  $(x_i, y_i, z_i)$  around their centroid  $(\bar{x}, \bar{y}, \bar{z})$ ; see Figure 2 for an illustration of this approach. 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 end points.

Figure 3 shows scatter plots of the so-determined amplitudes  $A$  and effective target widths  $W$  for movements in sequences of [ta] and [ka]. Recall the two prerequisites for an assessment of Fitts' law in any domain. One such prerequisite is that movements must be elicited under temporal pressure. This is undeniably satisfied in our paradigm, given the range of rates elicited. The other prerequisite, the one that is not straightforward to satisfy, is that movement amplitude and target size have vary-



**Fig. 3.** Range of movement amplitudes and effective target widths. **a** Sequences of [ta]. **b** Sequences of [ka]. Data are drawn separately for each speaker (subpanels). Metronome rate is colour coded with fainter shades for slower rates and darker shades for faster rates.



ing values across experimental conditions.<sup>1</sup> It is evident from Figure 3 that there is a variety of distinct values of target width  $W$ . Furthermore, for each of these there is a spread in the  $A$  quantity. Whether the range of variation in the  $A$  and  $W$  quantities is sufficient to allow for a robust assessment of the law (in the statistical sense) will be fully answered in the next section. Even though our statements about variation in  $A$  and  $W$  can only be taken as descriptive up to now, we are not aware of any previous demonstration from a corpus study or an experimental paradigm which has rendered such multiplicity in these quantities with speech data.

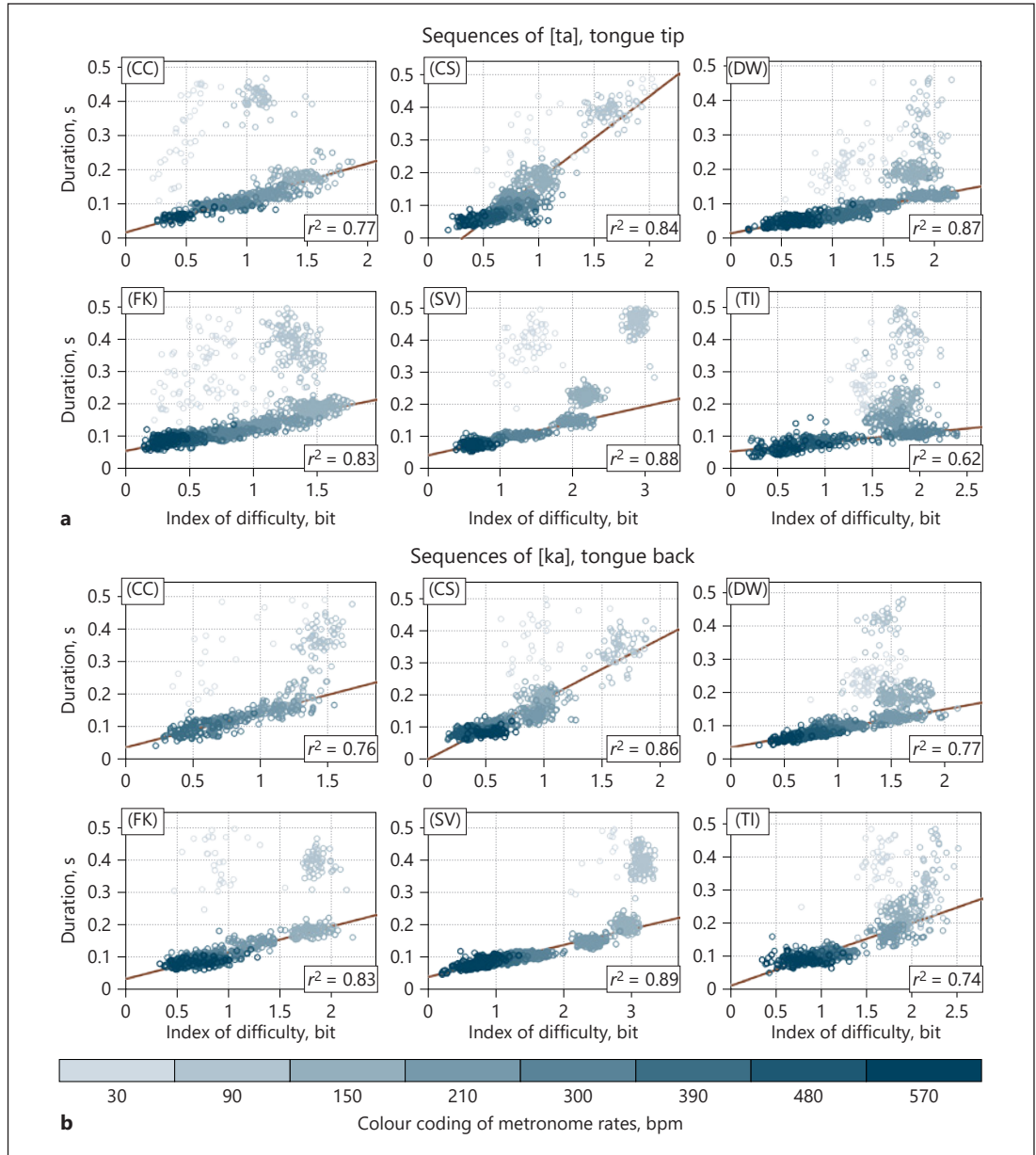
#### *Presence of Fitts' Law*

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 of difficulty  $ID = \log_2(A/W + 1)$ , as in  $T = a + b ID$ , where the constants  $a$  and  $b$  are empirically determined.

Figure 4 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 colour-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 data points 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) temporal 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 the 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 (with even less temporal pressure) satisfies Fitts' paradigm. Hence, we can partition our rate continuum into a set of

<sup>1</sup> In preliminary work with available speech corpora, we found that it is possible to mimic evidence for Fitts' law in ways that do not truly entail independent manipulations of  $A$  and  $W$  as in the original Fitts' design. For example, by artfully choosing amplitude differences, it is possible to demonstrate weak linear relations between time and amplitude while keeping  $W$  apparently constant.



**Fig. 4.** Relation between movement duration and index of difficulty. **a** Sequences of [ta]. **b** Sequences of [ka]. Data are drawn separately for each speaker (subpanels). Metronome rate is colour coded with fainter shades for slower rates and darker shades for faster rates. Linear regressions of contiguous Fitts-compliant rates are drawn as thick lines (corresponding  $r^2$  values are given in the bottom right corner of each panel). Linear regression lines are not meant to indicate fits to the entire data set but only to a subset starting from a (speaker-specific) rate and including all higher rates (see text for details).

**Table 1.** Statistics of contiguous sets of rates per speaker

$S_i$	CC		CS		DW		FK		SV		TI	
	$F_i$	$r_i^2$	$F_i$	$r_i^2$	$F_i$	$r_i^2$	$F_i$	$r_i^2$	$F_i$	$r_i^2$	$F_i$	$r_i^2$
[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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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] $\geq$ 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

*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.

contiguous rates that conform to Fitts' paradigm and another set of contiguous rates that do not. This contiguity property reflects precisely the requirement of temporal pressure 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 is thus 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  $R_i$  ( $i = 1 \dots 8$ ) 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 of  $S_i$  data points 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 data points from just set  $R_1$ , set  $S_2$  includes those in  $S_1$  plus the data points from  $R_2$ , set  $S_3$  includes those in  $S_2$  plus the data points 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 data points) for each  $S_i$ .

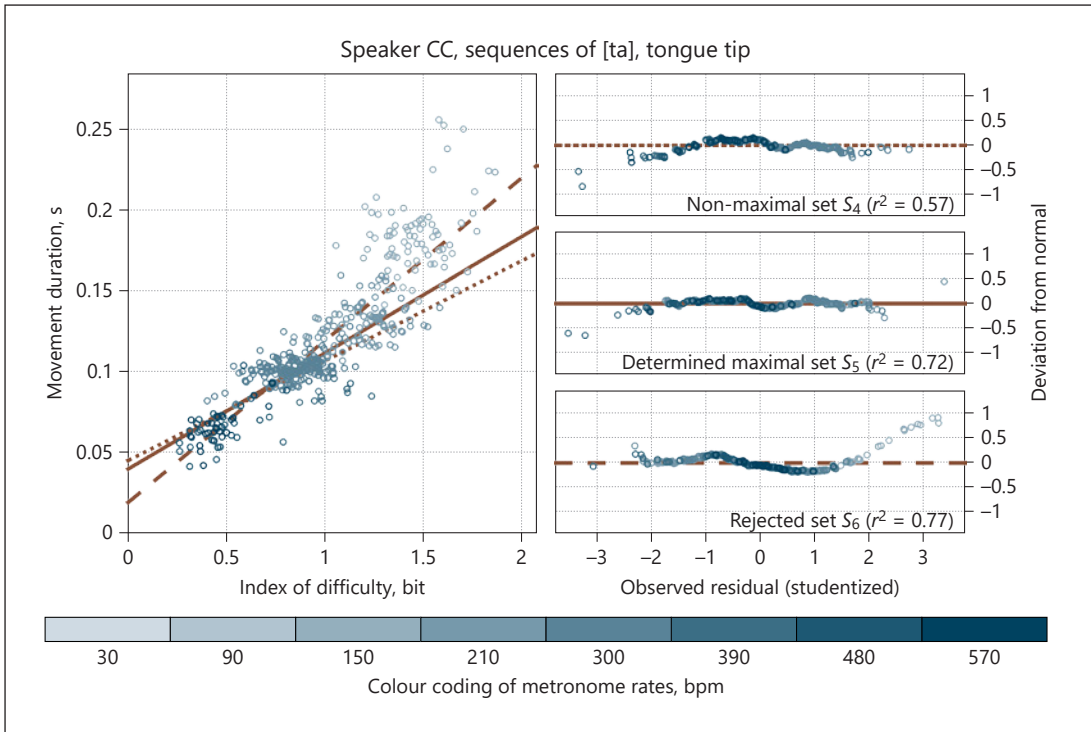
- 2 Determine the difference between the correlation slopes of set  $S_i$  and the next larger set  $S_{i+1}$ , which is the union of  $S_i$  and the data points from the next slower rate  $R_{i+1}$ . Slope differences are computed using a null hypothesis test for identical slopes (e.g., Cohen, 1983) 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 inclusion 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  $\Delta F_i$ ) and thus quality of linearity decreases significantly.<sup>2</sup> 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.<sup>3</sup> 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  $S_8 = \{30 \dots 570 \text{ bpm}\}$ . 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 in  $F$  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.

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, and 90 bpm (Table 1, top half, CC). Out of these candidate sets, the set with the 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

<sup>2</sup> 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  $S_i$ .

<sup>3</sup> 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  $r^2$  (Ancombe, 1973). An example from our domain is shown in Figure 5.



**Fig. 5.** 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 colour coded with fainter shades for slower rates and darker shades for faster rates. 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).

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 5 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]  $\geq 300$  bpm),  $S_5$  (determined maximal set, [ta]  $\geq 210$  bpm) and  $S_6$  (rejected set, [ta]  $\geq 150$  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 5 ( $S_4$ : dotted,  $S_5$ : solid, and  $S_6$ : dashed line). The right-hand side of Figure 5 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

**Table 2.** Properties of per-speaker determined Fitts-compliant regions

	CC	CS	DW	FK	SV	TI
[ta] slowest rate $R$ , 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 $1/b$ , bit/s	13.9	5.8	18.2	17.2	19.6	34.7
[ka] slowest rate $R$ , 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 $1/b$ , bit/s	9.3	7.5	17.5	12.4	20.1	15.5

All  $p$  values reside below 0.0001 (very strong significance).

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 5. 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 colour 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 5, 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 4 and correspond well to the impressionistic view of where linearity resides in these data sets. 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 0.62–0.89, 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 slower end of the set of contiguous rates, 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 reciprocal 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].<sup>4</sup> 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).

<sup>4</sup> 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 very strong 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 ( $r^2 = 0.62 \dots 0.89$ ,  $p < 0.0001$ ), and this linear relation holds regardless of the effector implicated in the task, tongue tip for [ta] or tongue back for [ka].

## 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, where each 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 (e.g., Browman & Goldstein, 1986; Fowler et al., 1980; Saltzman & Munhall, 1989). Accordingly, several contemporary (dynamical) approaches to the units of speech action assume that these units (speech gestures) are controlled by a dynamical system with fixed-point dynamics (e.g., Guenther, 1995; Perrier et al., 1996; Saltzman & Munhall, 1989). For example, task dynamics (Saltzman & Munhall, 1989), perhaps the most fleshed out approach to date, utilizes the dynamical system  $\ddot{x} = -kx - b\dot{x}$ , with stiffness parameter  $k$  and damping  $b = 2\zeta\sqrt{k}$ . Here, for the purpose of illustration, we restrict ourselves to the simplified, one-dimensional case with  $x$  describing only a scalar quantity of the task (e.g., tongue-palate constriction degree). Critical damping, as assumed by Task Dynamics, is realized by a fixed damping ratio  $\zeta$  equal to 1 (thus, neither  $\zeta$  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  $x(t) = e^{-\gamma t} \tilde{x}(t)$ , with some real-valued constant  $\gamma$  and some function  $\tilde{x}$  not eliminating the exponential signature of  $x$  (that is,  $\tilde{x}$  does not cancel the exponential factor  $e^{-\gamma t}$ ). Hence, the solutions of this 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 & Goodeve, 1983), as well as Card 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 and  $k$  and  $\zeta$ ). Recall now that in our data we have found evidence for Fitts' law only for speaker-specific rates above or equal to 150–300 bpm. Hence, the absence of Fitts' law for every speaker at some rates is outside the scope of



any instantiation of the damped linear oscillator model for speech (including that in Task Dynamics). Note that this inconsistency occurs at rates from the slower end of typical speaking conditions.<sup>5</sup>

Given these results, a validity test for any proposed model is that it must predict both the presence and the 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 Task Dynamics, 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 et al., 1985; also Fuchs et al., 2011). However, as shown above, manipulation of  $k$  does not alter the general exponential signature of the movement trajectories (nor does manipulation of  $\zeta$ ). Hence, another way to characterize the failure of any damped linear oscillator model of fixed-point dynamics 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., Guenther, 1995; Kröger et al., 1995; Perrier et al., 1996; Sorensen & 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 so as to result in a complete constriction and not too wide so as to result in an approximant (Catford, 1977). Examples include, at the velar place of constriction, stop [k] versus fricative [x] versus approximant [ɰ] or, at the palatal place of constriction, [ç] versus fricative [ç] versus approximant [j]. For these reasons, the articulatory postures of fricatives seem to require more precise control of the supralaryngeal configuration of the vocal tract

<sup>5</sup> Recent studies on German corpora place speech rates of typical speaking conditions in the range of 276 bpm (Gerstenberg et al., 2018), 300 bpm (Pellegrino et al., 2004), 357 bpm (Pellegrino et al., 2011), and 245–322 bpm (slow to normal; Dellwo & Wagner, 2003).

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 & Munhall, 1985) with velocity-amplitude slopes steeper for faster than for slower speech rates (Vatikiotis-Bateson & Kelso, 1990, 1993) and decreasing covariation as durational variability increases (Vatikiotis-Bateson & Kelso, 1990, 1993). Most relevantly for our purposes, Kuehn & 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, MacKenzie et al. (1987) report lower peak velocities for smaller target sizes as well as modulations of velocity profile shape (i.e., change in velocity over time) as a function of target size. Peak velocity magnitude and velocity profile shape have been used as a testbed for dynamical models of movement in discrete aiming tasks (MacKenzie et al., 1987), reciprocal tapping tasks (Bootsma et al., 2004), saccade-eliciting tasks (Van Opstal & Van Gisbergen, 1987), and finally also in speech (Kröger et al., 1995; Sorensen & 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. The here observed qualitative distinction between a set of fast target-directed movements obeying Fitts' law and another set of slow but likewise target-directed movements for which the law breaks down finds similarities in other areas of human movement science. Potential distinctions between qualitative control regimes underlying what may be apparently similar movements have been pursued in work that has so far remained unrelated to speech. In particular, there is evidence from limb motor control indicating that increasing movement rate may result in qualitative changes in the control regime underlying these movements (see e.g., Huys et al., 2008, and Jirsa & Kelso, 2005, which present evidence for bifurcations in finger movement data with movement rate as the bifurcation parameter). In parallel work, we are exploring the phase space of tongue movement data, including those considered here, and find evidence for the existence of distinct dynamical regimes with speech rate as the parameter whose scaling results in the change from one dynamical regime to another. It remains to be seen if (and how) distinctions in dynamical regimes of movements can be related to the here observed dichotomy of data sets under different speech rates conforming differently to Fitts' law.

## Conclusion

We asked whether speech movements abide by 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 8 distinct rates, ranging from extremely slow to extremely fast (30–570 bpm). In the resulting data set, 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 by 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 where-in relations involving kinematic and task space coordinates are implicated.

### Statement of Ethics

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 participants.

### Disclosure Statement

Both authors have contributed equally to this work and have no conflicts of interest to declare.

### Funding Sources

This work has been supported by the European Research Council (AdG 249440, <https://erc.europa.eu/>) and the Deutsche Forschungsgemeinschaft (grant No. 317633480, SFB 1287, Project C04, <http://www.dfg.de/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### References

- Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27, 17–21.
- Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Oxford, UK: Pergamon Press.
- Bootsma, R. J., Fernandez, L., & Mottet, D. (2004). Behind Fitts' law: Kinematic patterns in goal-directed movements. *International Journal of Human-Computer Studies*, 61(6), 811–821. <https://doi.org/10.1016/j.ijhcs.2004.09.004>
- Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3, 219–252.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum Associates.
- Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington, IN: Indiana University Press.
- Cohen, A. (1983). Comparing regression coefficients across subsamples. A study of the statistical test. *Sociological Methods & Research*, 12(1), 77–94. <https://doi.org/10.1177/0049124183012001003>
- Connelly EM (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., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' Law. *Quarterly Journal of Experimental Psychology*, 35(Pt 2), 251–278. <https://doi.org/10.1080/14640748308402133>
- Dellwo V, Wagner P (2003): Relations between language rhythm and speech rate. *15th International Congress of Phonetic Sciences*.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391. <https://doi.org/10.1037/h0055392>
- Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 67(2), 103–112. <https://doi.org/10.1037/h0045689>

- Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production: speech and talk* (pp. 373–420). New York, United States: Academic Press.
- Fuchs, S., Perrier, P., & Hartinger, M. (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: JSLHR*, *54*(4), 1067–1076. [https://doi.org/10.1044/1092-4388\(2010/10-0131\)](https://doi.org/10.1044/1092-4388(2010/10-0131))
- Gerstenberg A, Fuchs S, Kairret JM, Schröder J, Frankenberg C (2018): A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French. *Proc. 9th International Conference on Speech Prosody 2018*, pp. 211–215. <https://doi.org/10.21437/SpeechProsody.2018-43>
- Grosjean, M., Shiffrar, M., & Knoblich, G. (2007). Fitts' law holds for action perception. *Psychological Science*, *18*(2), 95–99. <https://doi.org/10.1111/j.1467-9280.2007.01854.x>
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594–621. <https://doi.org/10.1037/0033-295X.102.3.594>
- Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America. A, Optics and Image Science*, *4*(4), 629–642. <https://doi.org/10.1364/JOSA-A.4.000629>
- Huys, R., Studenka, B. E., Rheaume, N. L., Zelaznik, H. N., & Jirsa, V. K. (2008). Distinct timing mechanisms produce discrete and continuous movements. *PLoS Computational Biology*, *4*(4), e1000061. <https://doi.org/10.1371/journal.pcbi.1000061>
- Jirsa, V. K., & Kelso, J. A. S. (2005). The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation. *Journal of Motor Behavior*, *37*(1), 35–51. <https://doi.org/10.3200/JMBR.37.1.35-51>
- Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A “dynamic pattern” perspective on the control and coordination of movement. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 137–173). New York, United States: Springer. [https://doi.org/10.1007/978-1-4613-8202-7\\_7](https://doi.org/10.1007/978-1-4613-8202-7_7)
- Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (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), 266–280. <https://doi.org/10.1121/1.392268>
- Kröger, B. J., Schröder, G., & Oppen-Rhein, C. (1995). A gesture-based dynamic model describing articulatory movement data. *The Journal of the Acoustical Society of America*, *98*(4), 1878–1889. <https://doi.org/10.1121/1.413374>
- Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, *4*, 303–320.
- Lammert, A. C., Shadle, C. H., Narayanan, S. S., & Quatieri, T. F. (2018). Speed-accuracy tradeoffs in human speech production. *PLoS One*, *13*(9), e0202180. <https://doi.org/10.1371/journal.pone.0202180>
- MacKenzie, C. L., Marteniuk, R. G., Dugas, C., Liske, D., & Eickmeier, B. (1987). Three-dimensional movement trajectories in Fitts' task: implications for control. *The Quarterly Journal of Experimental Psychology*, *39*(4), 629–647. <https://doi.org/10.1080/14640748708401806>
- MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, *7*(1), 91–139. [https://doi.org/10.1207/s15327051hci0701\\_3](https://doi.org/10.1207/s15327051hci0701_3)
- MacKenzie, I. S. (2013). A note on the validity of the Shannon formulation for Fitts' index of difficulty. *Ocean Journal of Applied Sciences*, *3*(6), 360–368. <https://doi.org/10.4236/ojapps.2013.36046>
- Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology. Human Perception and Performance*, *11*(4), 457–474. <https://doi.org/10.1037/0096-1523.11.4.457>
- Murata, A. (1999). Extending effective target width in Fitts' law to a two-dimensional pointing task. *International Journal of Human-Computer Interaction*, *11*(2), 137–152. [https://doi.org/10.1207/S153275901102\\_4](https://doi.org/10.1207/S153275901102_4)
- Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, *46*(2), 135–147. <https://doi.org/10.1007/BF00339982>
- Nelson, W. L., Perkell, J. S., & Westbury, J. R. (1984). Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *The Journal of the Acoustical Society of America*, *75*(3), 945–951. <https://doi.org/10.1121/1.390559>
- Ostry, D. J., Cooke, J. D., & Munhall, K. G. (1987). Velocity curves of human arm and speech movements. *Experimental Brain Research*, *68*(1), 37–46. <https://doi.org/10.1007/BF00255232>

- Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America*, 77(2), 640–648. <https://doi.org/10.1121/1.391882>
- Patel, A.D., Löfqvist, A., & Naito, W. (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.
- Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87(3), 539–558. <https://doi.org/10.1353/lan.2011.0057>
- Pellegrino, F., Farinas, J., & Rouas, J. (2004). Automatic estimation of speaking rate in multilingual spontaneous speech. *International Conference on Speech Prosody 2004*.
- Perrier, P., Ostry, D. J., & Laboisière, R. (1996). The equilibrium point hypothesis and its application to speech motor control. *Journal of Speech, Language, and Hearing Research: JSLHR*, 39(2), 365–378. <https://doi.org/10.1044/jshr.3902.365>
- Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, 20(2), 279–303. <https://doi.org/10.1017/S0140525X97001441>
- Saltzman, E. L. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (Vol. 15, pp. 129–144). Experimental Brain Research Series New York, United States: Springer. [https://doi.org/10.1007/978-3-642-71476-4\\_10](https://doi.org/10.1007/978-3-642-71476-4_10)
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333–382. [https://doi.org/10.1207/s15326969eco0104\\_2](https://doi.org/10.1207/s15326969eco0104_2)
- Schmidt, R., & Lee, T. (2011). *Motor control and learning: a behavioral emphasis*. Champaign, United States: Human Kinetics.
- Sorensen, T., & Gafos, A. I. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology*, 28(4), 188–215. <https://doi.org/10.1080/10407413.2016.1230368>
- Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: toward an ecological psychology* (pp. 211–265). Lawrence Erlbaum Associates.
- Turvey, M. T. (2007). Action and perception at the level of synergies. *Human Movement Science*, 26(4), 657–697. <https://doi.org/10.1016/j.humov.2007.04.002>
- Van Opstal, A. J., & Van Gisbergen, J. A. (1987). Skewness of saccadic velocity profiles: a unifying parameter for normal and slow saccades. *Vision Research*, 27(5), 731–745. [https://doi.org/10.1016/0042-6989\(87\)90071-X](https://doi.org/10.1016/0042-6989(87)90071-X)
- Vatikiotis-Bateson, E., & Kelso, J. A. S. (1990). Linguistic structure and articulatory dynamics: a cross language study. *Haskins Laboratories Status Report in Speech Research, SR-103*(104), 67–94.
- Vatikiotis-Bateson, E., & Kelso, J. A. S. (1993). Rhythm type and articulatory dynamics in English, French and Japanese. *Journal of Phonetics*, 21, 231–265.
- Welford, A. T. (1968). *Fundamentals of skill. Methuen's manuals of modern psychology*. London, United Kingdom: Methuen.
- Wobbrock JO, Shinohara K, Jansen A (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. <https://doi.org/10.1145/1978942.1979181>
- Woodworth, R. S. (1899). The accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, 3(3), i-114.
- Wright, C. E., & Meyer, D. E. (1983). Conditions for a linear speed-accuracy trade-off in aimed movements. *The Quarterly Journal of Experimental Psychology Section A*, 35(Pt 2), 279–296. <https://doi.org/10.1080/14640748308402134>
- Zhai, S., Kong, J., & Ren, X. (2004). Speed-accuracy tradeoff in Fitts' law tasks—On the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies*, 61(6), 823–856. <https://doi.org/10.1016/j.ijhcs.2004.09.007>