

Articulatory Variability is Reduced by Repetition and Predictability

Language and Speech

1–27

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0023830920948552

journals.sagepub.com/home/las**Fabian Tomaschek** 

Department of General Linguistics, University of Tübingen, Germany

Denis Arnold

Institut für Deutsche Sprache, Mannheim, Germany

Konstantin Sering 

Department of General Linguistics, University of Tübingen, Germany

Benjamin V. Tucker 

Department of Linguistics, University of Alberta, Canada

Jacoline van Rij

Neurolinguistics and Language Development, University of Groningen, The Netherlands

Michael Ramscar

Department of General Linguistics, University of Tübingen, Germany

Abstract

Repeating the movements associated with activities such as drawing or sports typically leads to improvements in kinematic behavior: these movements become faster, smoother, and exhibit less variation. Likewise, practice has also been shown to lead to faster and smoother movement trajectories in speech articulation. However, little is known about its effect on articulatory variability. To address this, we investigate the extent to which repetition and predictability influence the articulation of the frequent German word “sie” [zi] (they). We find that articulatory variability is proportional to speaking rate and the duration of [zi], and that overall variability decreases as [zi] is repeated during the experiment. Lower variability is also observed as the conditional probability of [zi] increases, and the greatest reduction in variability occurs during the execution of the vocalic target of [i]. These results indicate that practice can produce observable differences in the articulation of even the most common gestures used in speech.

Corresponding author:

Fabian Tomaschek, Department of General Linguistics, University of Tübingen, Wilhelmstraße 19, Tübingen, 72074, Germany.

Email: fabian.tomaschek@uni-tuebingen.de

Keywords

Articulation, practice, predictability, variability, precision

Introduction

1.1 Improvements in kinematic behavior of hand movements

Kinematic behavior is inherently variable, such that repetitions of any single action, for example repeatedly making a particular hand gesture, will invariably exhibit variation in each instance. One source of this variability is noise in the central nervous system, which can affect sensory processes, and the activities associated with motion planning and movement execution (Bays & Wolpert, 2007; Churchland et al., 2006; Harris & Wolpert, 1998; van Beers et al., 2002). A further source of movement variability arises out of uncertainty about the target of a movement, and the degree to which the target location is predictable. The less predictable the target of a movement, for example because it shifts during testing, the greater will be the uncertainty about the target and the larger will be the variability of the movement (Georgopoulos et al., 1981; Pellizzer & Hedges, 2003).

An obvious way in which the uncertainty associated with any given movement can be moderated is through repetition and practice. In everyday life, the idea that the variability associated with a particular movement, for example swinging a golf club or selecting a note on a keyboard, can be reduced through practice is a familiar one. It is therefore not surprising that a considerable amount of laboratory work has been conducted to better understand the way that repetition influences kinematic behavior, particularly in hand movements. These studies have shown that practice of a particular movement serves to optimize the kinematic system, resulting in the execution of movement trajectories that are shorter and smoother, require less effort, and crucially, are less variable (Darling et al., 1988; Georgopoulos et al., 1981; Gribble & Ostry, 1996; Madison et al., 2013; Platz et al., 1998; Raeder et al., 2015; Segalowitz & Segalowitz, 1993; Sosnik et al., 2004; Viviani & Schneider, 1991). One way of explaining these findings is in terms of the way that practice reduces uncertainty along the trajectory of movements. As a movement is repeated, the various intermediate targets and gestures that comprise the movement as a whole become more predictable, resulting in a concomitant decrease in the uncertainty associated with each stage of movement. This raises a question. Playing instruments and swinging golf clubs are occasional pastimes. By contrast, barring a few exceptions, speech is a ubiquitous human behavior. Given that speech is effectively the most over-learned kinematic skill of all, can practice effects be observed in articulation?

1.2 Improvements in kinematic behavior of speech articulation

While there are many similarities between hand movements and articulation, there are of course differences. Beyond the fact that for most people speech will be the most practiced kinematic activity that they engage in, there are also basic structural differences in the underlying kinematic systems. Hand movements are effected with a relatively rigid body joint-angle system and usually have one target (Bourgeois & Hay, 2003). By contrast, the articulatory speech apparatus consists of a rigid bone structure (the jaw) and muscular hydrostats (the tongue, the lips) whose different parts are biomechanically joined and task-dynamically coupled (Bell-Berti & Harris, 1979; Fowler & Saltzman, 1993; Saltzman & Munhall, 1989). In speech, coupled articulators produce successive gestures, which continuously aim for articulatory, sensory and acoustic targets (Browman & Goldstein, 1986; Guenther, 1995; Johnson et al., 1993). The resultant articulatory gestures are submitted to tremendous contextual variability as targets can be partially competing. This partial

competition between targets results in a large overlap between adjacent gestures (Browman & Goldstein, 1986; Perkell & Nelson, 1982). Accordingly, articulatory trajectories for identical phones differ systematically depending on preceding and following phonemes due to coarticulation (Browman & Goldstein, 1986; Magen, 1997; Öhman, 1966).

These differences notwithstanding, there is some evidence to support the idea that speech is affected by practice. For example, the temporal characteristics of speech production have been shown to be influenced by repetition (Fowler, 1988; Fowler & Housum, 1987), which, by definition, will inevitably result in a decrease in uncertainty and a concomitant increase in predictability. Phones and words which are more predictable—either in context or because of their frequency of occurrence—tend to have significantly shorter acoustic durations than less predictable phones and words (Aylett & Turk, 2004; Bell et al., 2009; Cohen Priva, 2015; Gahl, 2008; Ramscar et al., 2014; Tremblay & Tucker, 2011; Whalen, 1991). These effects are mirrored in articulation. Repetitions and frequency are correlated with higher articulatory velocity, smoother gestural transitions and stronger anticipatory coarticulation, which are in turn mirrored by shorter execution times (Tiede et al., 2011; Tomaschek, Arnold, et al., 2018; Tomaschek, Tucker, Baayen, et al., 2018; Tomaschek, Tucker, Ramscar, et al., submitted; Tomaschek, Tucker, Wieling, et al., 2014; Tomaschek, Wieling, et al., 2013). Unsurprisingly, given the foregoing, the temporal coordination of gestural sequences is modulated by lifelong practice (Cychosz, 2020; Green et al., 2002; Noiray et al., 2013; Rubertus & Noiray, 2018, 2020; Tomaschek, Tucker, et al., 2018).

Given that practice does appear to result in faster and smoother kinematic behavior in speech production, the question thus arises whether articulation also exhibits a reduction in variability associated with practice. That is, as articulatory gestures become more predictable, such that the uncertainty associated with them is reduced, will we see the same reduction in variability in speech that has been observed in hand movement studies?

Some initial support for this hypothesis comes from studies showing a reduction of articulatory variability as a result of lifelong, long-term practice. Studies have consistently shown that, as compared to children, adults show smaller temporal variability during anticipatory coarticulation and smaller spatial variability during vowel and consonant production (Belmont, 2011; Goffman et al., 2008; Koenig et al., 2008; Zharkova et al., 2011, 2012). This suggests that, at least insofar as during the transition from novice to expert, practice may indeed reduce articulatory variability.

In what follows, we explore whether, even in mature adult speakers, articulatory variability is similarly reduced in linguistic contexts where articulatory gestures are more predictable—and hence, by definition, well-practiced—in contrast to linguistic contexts where they are less predictable—and hence less well-practiced.

1.3 The present study

In order to investigate effects of practice and repetition in hand movements, kinematic studies typically use well defined (quasi-) parabolic pathways that participants are asked to follow with a pen (e.g., Bourgeois & Hay, 2003; Sosnik et al., 2004; Viviani & Terzuolo, 1982). Practically, however, tasks that use precisely predefined trajectories like this are impossible to implement in studies of articulation. However, it is possible to approximate this methodology through the investigation of a specific well-practiced articulatory trajectory. In this case, our trajectory of interest will be a simple vertical articulatory gesture during which the tongue dorsum moves towards a vocalic target. Examples of this simple articulatory gesture can be found in many frequent monosyllabic CV words, in the case of the present experiment, in the German pronoun *sie* [zi] “Engl. they”.¹ Participants produced the gestures of interest during the articulation of “[zi] + verb” phrases. The kinematics of these gestures were recorded using electromagnetic articulography. This procedure

allowed us to examine multiple instances of—in the framework of articulatory phonology (Browman & Goldstein, 1986)—an identical articulatory gesture towards one invariant vocalic target (i.e., the one of /i/) while being produced in a range of contexts. These contexts in turn varied the predictability of the relationship between [zi] and the following verb, allowing the effect of predictability/uncertainty on the variability of the production of our gesture of interest to be observed.

In hand movements, the decrease of variability can occur along the trajectory (Sosnik et al., 2004) or at its end-point (Fitts, 1954). However, given the nature of speech production, articulatory gestures rarely have specifically defined end-points. Instead, articulatory trajectories are better characterized in terms of positions that the articulators consecutively aim for (Browman & Goldstein, 1986; Perkell & Nelson, 1982). Given that this means that articulatory movements tend to be continuous, we thus tested for changes in variability along the entire trajectory of the gestures of interest produced by our participants, including their vocalic targets.

2 Methods

2.1 Stimuli

The gesture of interest of this investigation is produced during the articulation of the German pronoun *sie* [zi] “they”. To get a measure of variability, we examined this articulation in conjunction with a total of 127 different verbs, such that the gesture of interest was always produced in a “[zi] + verb” phrase (a list of the stimuli can be found in the Appendix in Tables 3 to 6). Verbs were controlled for frequency of use, taken from the SDEWAC corpus (Faaß & Eckart, 2013; Shaoul & Tomaschek, 2013). In all of the verbs, /i:, ɪ, a:, a/ occurred as the nucleus in the first stressed syllable.²

Participants articulated [zi] 254 times in the course of the experiment. During the course of the study some trials produced no data due to mispronounced articulations or sensors that were not correctly tracked. As a result, the final data set contained on average 184 [zi] instances per participant (min = 107, max = 242). In the analyses that follow, only verbs with a single non-dorsal onset consonant were considered. In total, 3117 tokens were analyzed, consisting of 72 verb types beginning with a coronal consonant, and 55 verb types beginning with a labial consonant.

2.2 Speakers and recording

A total of 21 German native speakers were recorded and paid 15€ for their participation. All speakers provided informed consent before participating in the experiment, and speaker identity was anonymized. Four speakers had to be excluded from the analysis due a large proportion of missing data or faulty sensors. Of the remaining speakers in the analysis, nine were female and eight were male, with mean age 25.6 years ($SD = 3$ years).

All recordings were conducted in a sound-attenuated booth in the Department of Linguistics at the University of Tübingen. Speakers were instructed to read the stimuli ([zi] + verb) aloud after they appeared on a computer screen. The list of stimuli was pseudo-randomized for each participant and separated into three recording blocks. Each block was presented once in a *slow* (odd blocks, inter-stimulus-time: 600 ms; presentation-time: 800 ms) and once in a *fast* speaking condition (even blocks, inter-stimulus-time: 300 ms; presentation-time: 450 ms). In total, six blocks of stimuli were presented. Presentation order in *slow* and *fast* blocks was also pseudo-randomized. Crucially, each stimulus ([zi] + verb) was presented only twice, once in a *slow* and once in a *fast* condition.

Participants' tongue movements during articulation were recorded with an NDI wave articulo-graph at a sampling frequency of 100 Hz. Contemporaneous audio recordings of the articulations were also made (22.05 kHz, 16bit) and synchronized with the recordings from the articulo-graph. To correct for head movement and to define a local coordinate system, a special 6D reference sensor was attached to the speakers' forehead.

Before sensors were attached to a participant's tongue, a bite plate recording was made to determine the head's rotation in relation to the magnetic emitter. To the bite plate three sensors in a triangular configuration were attached, which represent a local reference for a standardized coordinate system. During the experiment, tongue movements were captured by three sensors: one slightly behind the tongue tip, one at the tongue middle and one at the tongue dorsum (distance between each sensor: approximately 0.5cm to 1cm).

The verbs recorded in the present study contain apical consonants that inhibit potential effects of variability on the tongue tip. Therefore, we investigated only tongue dorsum movements. Since horizontal tongue dorsum movements were minimal (< 1 mm) in the present data, we focused our investigation on movements in the vertical dimension.

2.3 Preprocessing

The effects of head movement on tongue movements were corrected for using an online procedure during recording. The recorded positions of the sensors were centered at the midpoint of the bite plate and rotated. As a result, the back-front direction of the tongue was aligned to the horizontal axis with more positive values towards the front of the mouth, and more positive vertical values towards the top of the oral cavity. Word boundaries were determined by automatically aligning the audio signal with phonetic transcriptions using a Hidden-Markov-Model-based forced aligner for German (Rapp, 1995), and manually verified and corrected where necessary.

3 Analysis

Having described the motivation, methods, and materials for the present study, we next discuss the variables that were used to predict articulatory movements, and lay out the statistical methods that were used in the analysis of these variables.

3.1 Movements across time

The very fact that the tongue movement will change its position as a function of time is inherent to speech production. It follows that any analysis of articulatory variability must inevitably be made in relation to changes in the tongue's position relative to time. To do this, we performed a time-course analysis of articulatory movements.

Given that articulatory rates are themselves variable, both within and across speakers—in the present dataset, durations for the gesture of interest ranged between 70 ms and 300 ms—it follows that any analysis of the variability of tongue movements in relation to specific gestures can never be a straightforward process. When trajectories from words with different durations are aligned at their onset, parts of the trajectory, such as the transition between [z] and [i] and the location of the maximum deflection of the tongue dorsum during [i], are located at different time points. This is illustrated in Figure 1 (a) for three different durations of [zi] (color coded), with time depicted at the x-axis and tongue height at the y-axis. Accordingly, articulatory trajectories from words with varying durations cannot be analyzed without mapping these different parts onto one another.

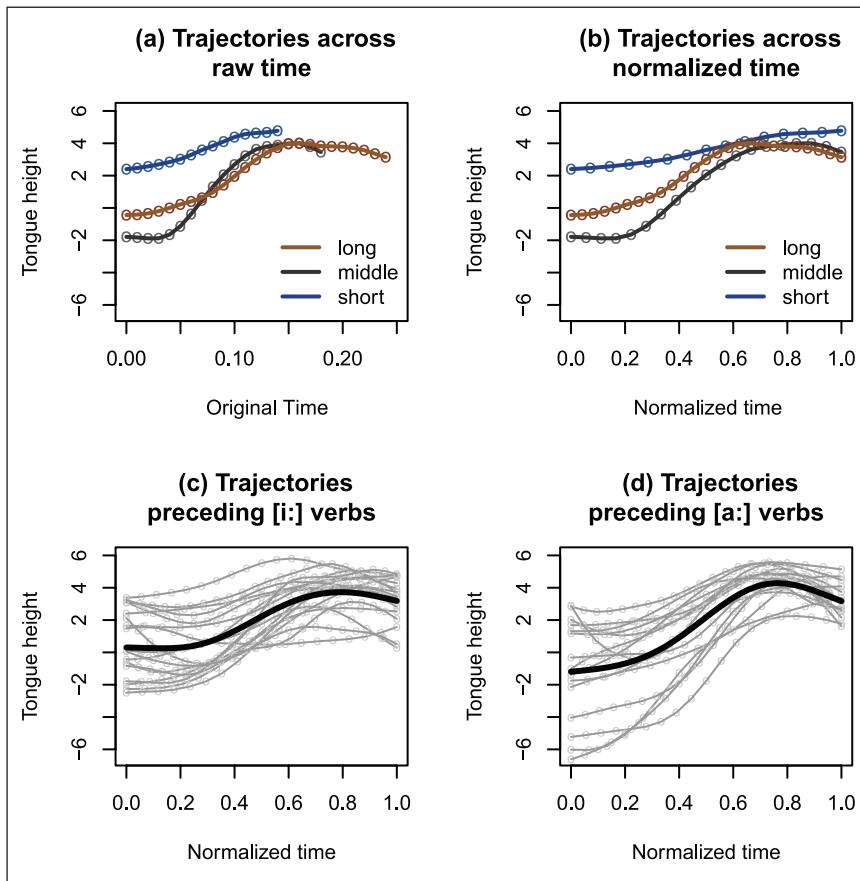


Figure 1. Illustrations of the articulatory trajectories of the tongue dorsum taken from [zi] utterances with three different durations (color coded) across original time (a) and across normalized time (b). (c–d) Illustrations of the data for one of the speakers in the slow speaking rate condition depending on the vowel in the following verb. Gray circles represent recorded positions, black lines represent average trajectories. Y-axis represents tongue dorsum heights, x-axis represents normalized times in the word [zi].

In order to achieve a mapping between different parts of the trajectories and to be able to control for duration differences between individual [zi] instances during statistical analysis, time normalization had to be performed. This was achieved by normalizing timestamps for the recorded articulatory positions to a [0, 1] interval, with 0 linked to vowel onset and 1 to vowel offset. This is illustrated in Figure 1 (b). In what follows, we refer to the timestamps for each recorded position normalized for the duration of [zi] as *Time*.

Figure 1 (c–d) illustrates how the recorded positions change across time for [zi] articulations preceding verbs with [i:] stem vowels (c) and verbs with [a:] stem vowels (d) across normalized time. The black lines in Figure 1 represent the average change in vertical position of the tongue dorsum across time. The tongue dorsum movement trajectory can be described as s-shaped across time, with the tongue dorsum located at a low position at the onset of [zi] and steadily rising towards the offset of [zi]. As can be seen, there are systematic differences in amplitude between the two upcoming vowels in the verbs: the articulatory trajectory during [zi] has a more pronounced

amplitude with a lower onset, and a higher peak and offset when preceding verbs with [a] than when preceding verbs with [i] as stem vowels.

The goal of the present study was to investigate how strongly articulatory trajectories vary across [zi] as a function of predictability and repetition within the experiment. In order to achieve this goal, we controlled for a range of intrinsic and extrinsic effects that are known to modulate average amplitude and shape of articulatory trajectories (represented with the black line in Figure 1). In the next section, we describe the predictors to capture these intrinsic and extrinsic effects as well as the predictors of interest. After this, we then introduce the statistical techniques that are used to analyze the data.

3.2 Predictors of interest

The participants in the present study articulated [zi] multiple times during the course of the experiment. This allowed us to investigate how the repetition of a highly practiced articulatory gesture within a relatively short time span might affect its variability. We operationalized repetition during our experiment by means of *WordRepetition* (z-scaled). We predicted that movement variability would become smaller due to short-term practice during the experiment. At the same time, this predictor was also used to control for the effects of repetition on the amplitude of the average articulatory trajectory in line with the findings by Tiede et al. (2011).

The contextual uncertainty of the articulatory gesture of [zi] was operationalized by the conditional probability of [zi] given the following verb, in other words, the inverse conditional probability of [zi]. Inverse conditional probability is an explicit measure of expectancy. By contrast, measures such as the verb's frequency or the phrase's bigram frequency alone are not (Baayen, 2001; Shannon, 1948). From a learning perspective, inverse conditional probability is also an index of the prediction error associated with the following verb given [zi] (Aizenberg et al., 2000; Daw et al., 2008; Dayan & Daw, 2008; Dechter, 1986; Hannun et al., 2014; Ng & Jordan, 2002; Schultz, 2006; Schultz et al., 1997; Sutton & Barto, 1981). Since prediction error of this kind is commonly used to explain learning at both, the behavioral and neural-biological level, it can be taken as an index of how our participants likely acquired their sensitivity to variance in the context of [zi]. This perspective is supported by studies in which conditional probability has been shown to account for behavior better than less informative measures (Arnon & Snider, 2010; Bannard & Matthews, 2008; Bell et al., 2009; Tremblay & Baayen, 2010). This is why the conditional probability measure was favored over the phrase's joint probability (i.e., the "[zi] + verb" bigram frequency). We calculated the conditional probability of [zi] given the following verb using Equation 1 (from now on *SieProbability*).

$$P(\text{sie} | \text{verb}) = \frac{C(\text{sie} + \text{verb})}{C(\text{verb})} \quad (1)$$

Because the SDEWAC corpus (Faaß & Eckart, 2013) did not contain all the phrases used in the current study, the number of Google search hits was used to obtain $C(\text{sie} + \text{verb})$ and $C(\text{verb})$. There are a number of reasons to suppose that Google hits can serve as a useful proxy for the actual experience of the participants in this study. First, Google hits for the verb have a Spearman's Rank correlation $\rho = 0.81$ with the frequency of occurrence in the SDEWAC corpus (for available verbs). Second, although less successfully than other frequency counts (Tucker et al., 2019), Google hits have already been shown to successfully predict cognitive behavior across a range of tasks (Baayen et al., 2017; Mather et al., 2014; Ramscar et al., 2017; Scharinger, 2006). For the purposes of the present analyses, *SieProbability* was log transformed and z-scaled to reduce any overly strong influences from outliers.

3.3 Control predictors

According to Fitts' Law (Fitts, 1954), movement variability is proportional to movement time. This effect has repeatedly been found for hand movements (Kim et al., 1999; Schmidt et al., 1979; Sosnik et al., 2004; Viviani & Schneider, 1991). Accordingly, we expected articulatory variability to be proportional to the duration of [zi], which we operationalized as the log-transformed and z-scaled measure *SieDurations*. Note that *SieDurations* and *Time* are independent of each other. *Time* predicts the time-course of the tongue's position, *SieDurations* predicts changes in the shape of the trajectory in relation to the total duration of the movement. Since our stimuli were recorded under two different speaking rate conditions, we also used the factorial predictor *SpeakingRateCondition*, with the levels fast and slow.

The current experimental set up also allowed for the investigation of an additional linguistic variable, namely anticipatory coarticulation. This is a consequence of the fact that [zi] was produced in an unstressed position of a large number of "[zi] + verb" phrases, where the verb contained either a high /i:/ and /i/ or a low /a:/ and /a/. Since unstressed syllables undergo anticipatory coarticulation of the upcoming syllable (Fowler & Brancazio, 2000; Hoole et al., 1993; Magen, 1997; Öhman, 1966; Recasens, 1984; Sziga, 1992; Tomaschek, Tucker, et al., 2018; Tomaschek et al., 2014; Tomaschek, Wieling, et al., 2013), we expected the tongue to change movement and variability patterns depending on the upcoming vowel.

Pilot analyses indicated that the average articulatory trajectory varied with the vowel category. Since the process of articulation is gradual rather than discrete, a gradient measure was used to account for the effect of the upcoming vowel.³ To do so, the measure *TargetDistance* was obtained by calculating the vertical difference between the tongue position in the center of [i] in [zi] and the tongue position in the center of the verb's stem vowel. Manual inspection indicated that this predictor exhibited a broad distribution of values, which allowed it to be included into the model. Model comparisons revealed that *TargetDistance* provided a better fit to the data (a reduction of the ML-score, $\Delta ML = -70.16$, and fewer degrees of freedom, $\Delta Edf. = -53$) than the factorial model differentiating between the different vowel classes, justifying its use.

Finally, Fitts' Law (Fitts, 1954) also predicts that movement variability is proportional to a limb's traveled distance. To account for a possible similar effect in tongue movements, we calculated the total distance that the tongue dorsum has traveled in the Euclidean space during the articulation of [zi] (*TraveledDistance*).

Given previous findings from studies of articulation, we expected that the average tongue trajectory across *Time* would be shallower in relation to greater *WordRepetition* as a result of fatigue effects, and shallower in relation to shorter *SieDurations* as a result of undershoot. We predicted a similar effect for *TargetDistance*, where average tongue trajectories were expected to be shallower when *TargetDistance* was greater, because this would result in increased anticipation of the vowel in the verb.⁴ Finally, we expected that standard deviations along the average tongue trajectory will be smaller in relation to greater *WordRepetition* due to practice during the experiment, smaller in the slow *SpeakingRateCondition*, smaller with smaller *TargetDistance*, and smaller in relation to smaller *TraveledDistance*.

3.4 Modeling strategy

To better understand any variability observed around the articulatory trajectory across [zi], we exploited the fact that standard regression models estimate the mean of the data and also provide the residuals, that is, the difference between the mean and the raw data. In Figure 1, residuals are the vertical differences between the black line and the gray circles. Studies by Chodroff and Wilson (2017), Sonderegger (2015) and Sonderegger et al. (2017) have modeled variability by fitting a

standard regression model, extracting the residuals from that model and subjecting them to a second regression analysis.

We used this approach in a top-down, bottom-up modeling strategy to obtain a final model for the average tongue trajectory and a final model for the residuals. These two models were subsequently merged and fitted simultaneously with a Generalized Additive Mixed-Effects model (GAMM, package **mgcv**, Version 1.8-31, Wood, 2006) using the “Gaussian location scale additive models” method (Wood et al., 2016). This method allowed us to fit the mean articulatory trajectory, controlling for any potential effects that might affect the time-course of the tongue dorsum, while simultaneously performing a statistical analysis of the variance around the average trajectory.⁵ To account for the exploratory approach used in the present study, we set the significance threshold to $p = 0.001$.

While linear mixed-effects models model linear functional relations between a response and a predictor, GAMMs model *non-linear* functional relations between a response and a numeric predictor. This is accomplished by means of *smooths*. In this way, GAMMs allow us to investigate the non-linear time-course of the average articulatory trajectory that can be seen in Figure 1. Further details of the mathematical underpinnings of GAMMs, as well as their use in analyzing time-dependent non-linear data can be found in Wieling et al. (2016) as well as in Tomaschek, Tucker, et al. (2018), Tomaschek et al. (2014), Tomaschek, Wieling, et al. (2013), Kryuchkova et al. (2012), Nixon, et al. (2016) and Tomaschek, Arnold, et al. (2018). Non-linear interactions between *Time* and potential covariates were modeled with *tensor product smooths*. Tensor product smooths explain the relation between the dependent variable and a non-linear interaction between two covariates by means of estimating a wiggly surface that best predicts the data.

We controlled for random effects on the average trajectory, using *random factor smooths*, that is, smooths for random effects that can be thought of as non-linear equivalents to a combination between random intercepts and random slopes in mixed-effects regression. These included random factor smooths across *Time* per *Speaker*, controlling for inter-speaker variability during the production of [zi] (see Fuchs et al., 2008; Tomaschek & Leeman, 2018; Weirich & Fuchs, 2006 for spatio-temporal articulatory variation due to physiological differences), and random factor smooths across *Time* per *Verb*, controlling for variation in tongue position due to near and far neighboring consonants (Fowler & Brancazio, 2000; Hoole et al., 1993; Magen, 1997; Öhman, 1966; Recasens, 1984; Sziga, 1992; Tomaschek et al., 2014; Tomaschek, Wieling, et al., 2013). Although *SieProbability* was not significant for tongue height in pilot analyses, we included a random factor smooth for *SieProbability* per *Speaker* in order to control for possible by-participant effects as a function of *SieProbability*.

The final model had $R^2 = 0.94$, in other words, the model explained a high amount of variance in the data. Note that a model lacking the standard deviation terms yielded $R^2 = 0.86$, in other words, the inclusion of the standard deviation terms increased the explained variance in the model. The current model explains a high proportion of the variance in the data, which is not usually the case in linguistic studies. Performing an exploratory analysis has revealed that roughly 64% of the variance in the data is explained by the random factor smooth for participants. This high R^2 in relation to random effects can be explained by the fact that all participants performed a very similar, sine-like, articulatory trajectory.

Having discussed the details of our analysis, we turn in the following section to the results. We first present the results of the model terms fitting the average time-course of the tongue dorsum. We describe how the GAMM model should be interpreted, as well as describe how GAMM plots should be read. Following this, we discuss the results of the model terms fitting the standard deviation around the average trajectory.

Table 1. Summary of partial effects in the GAMM model, fitting the **average tongue dorsum trajectory**. Random effect terms are prefixed with “RE”. “s()” terms represent smooths fitting tongue dorsum positions as a function of one numerical predictor, “ti()” terms represent partial tensors fitting tongue dorsum positions as a function of an interaction between “time” and another numerical predictor. Most of the smooths/tensors interact with “SpeakingRateCondition”. A term’s estimated degrees of freedom (“edf”) larger than 1 indicate non-linearity of the smooth/tensor when *p*-values are smaller than 0.001.

A) Parametric coefficients	Estimate	Std. error	z-value	p-value
(intercept)	9.661	1.257	7.690	< 0.001
B) Non-parametric effects	edf	Ref.df	F	p-value
s(Time):SpeakingRateCondition = slow	2.986	2.990	289.738	< 0.001
s(Time):SpeakingRateCondition = fast	2.984	2.989	281.217	< 0.001
s(SieDurations):SpeakingRateCondition = slow	2.739	2.943	288.712	< 0.001
s(SieDurations):SpeakingRateCondition = fast	2.699	2.932	51.372	< 0.001
s(TargetDistance):SpeakingRateCondition = slow	2.989	2.999	2551.481	< 0.001
s(TargetDistance):SpeakingRateCondition = fast	2.985	2.999	2145.550	< 0.001
s(WordRepetition)	2.939	2.996	325.464	< 0.001
ti(Time, SieDurations):SpeakingRateCondition = slow	7.278	8.247	1080.881	< 0.001
ti(Time, SieDurations):SpeakingRateCondition = fast	7.946	8.707	915.201	< 0.001
ti(Time, TargetDistance):SpeakingRateCondition = slow	6.739	7.835	204.681	< 0.001
ti(Time, TargetDistance):SpeakingRateCondition = fast	6.792	7.917	126.587	< 0.001
ti(Time, WordRepetition)	6.684	7.664	40.422	< 0.001
RE - s(Time, Participant)	64.687	68.000	20559.822	< 0.001
RE - s(SieProbability, Participant)	42.613	68.000	785.592	< 0.001
RE - s(Time, NextWord)	513.123	1142.000	13186.819	< 0.001

3.5 Effects of covariates on the average tongue dorsum trajectory

The average tongue trajectory was fitted by means of two three-way interactions. By means of the interaction *Time* × *SieDurations* × *SpeakingRateCondition*, we investigated how the shape of the tongue dorsum’s movement trajectory across time was modulated by differences in [zi] duration in slow and fast *SpeakingRateCondition*. By means of an *Time* × *TargetDistance* × *SpeakingRateCondition* interaction, we investigated how tongue movements across time varied due to anticipatory coarticulation in the two *SpeakingRateConditions*. By means of one two-way interaction, *Time* × *WordRepetition*, we investigated how tongue movements across time were modulated by *WordRepetition*. Pilot analyses indicated that *SieProbability* did not emerge as a significant predictor of the average articulatory trajectory. Also, we found that the average tongue height did not significantly differ between the fast and the slow *SpeakingRateCondition* ($\beta = 0.21$, $sd = 0.35$, $t = 0.62$), which is why it was not included as a main effect in the current model.

Table 1 shows the summary of the effects for the average trajectory in the final model. Part (A) reports the intercept (in mm) of the model. Part (B) reports the non-linear effects in the model. Interactions were fitted with combinations of main effects, using smooths “s()”, and partial tensor product smooths “ti()”. This method is equivalent to the $Y \sim X_1 + X_2 + X_1 : X_2$ approach in linear models. In contrast to linear models, non-parametric summaries do not report the estimated size or direction of the effect, but whether an effect was significantly non-linear. Size and direction of an effect have to be inferred from plots.

Estimated degrees of freedom (“edf”) of smooths for *Time*, *SieDurations*, *TargetDistance* in both *SpeakingRateConditions* are larger than 1, indicating that all these effects had a non-linear functional relationship with the position of the tongue dorsum. This was also the case for *WordRepetition*. Furthermore, the shape of the tongue movement changed across time due to the predictors, as indicated by the significant partial interactions between *Time* and *SieDurations*, *TargetDistance* and *WordRepetition*. All random factor smooths were significantly non-linear, as can be seen at the bottom of Table 1.

3.6 Estimated average trajectory

To understand the effects of variability, we first need to understand the time-course of the tongue dorsum sensor during [zi] as it was modulated by the interaction between the predictors *TargetDistance*, *SieDurations* and *WordRepetition*.

GAMM interactions are illustrated in a different way to standard regression plots. Estimates of the dependent variable obtained in a linear model are typically depicted on the y-axis. To illustrate interactions in a GAMM model, the shape of the estimated surface (in our case, tongue position across time modulated by another predictor) is illustrated by means of surface plots (Figure 2).

In Figure 1, the movement of the tongue was represented as a black line. By contrast, in a surface plot, movement can be represented in a way that is more akin to a geographical map, which often use color to represent the features of a terrain as a function of its geological coordinates. In the present surface plots, this color coding represents the estimated height of the tongue dorsum as function of time (depicted on the x-axis) in interaction with another covariate (depicted on the y-axis). Dark blue colors represent that the tongue dorsum is low at a given time point, as it is the case at the onset of [zi]; light yellow areas represent that the tongue dorsum is high, as it is the case in the offset of [zi]. Green represents tongue positions that are in between.

As can be seen in Figure 2, the movement of the tongue evolves from left to right, with the onset of the trajectory linked to the left edge, and its offset linked to the right edge. Modulations of tongue height by covariates are illustrated by changes in the color patterns. As in geographical maps, where areas of the same elevation are represented by one contour line, contour lines in the surface plots connect areas of the same tongue dorsum height on the estimated surface, irrespective of *Time* and the interacting predictor.

Figure 2 (a–b) illustrates how *TargetDistance* (depicted on the y-axis) modulated the tongue dorsum positions during the articulation of [zi] in both *SpeakingRateConditions*. The horizontal black dotted lines represent the average tongue height positions in the verbal stem vowel. When the vocalic targets in the stem of the following verb were higher, the gesture made by the tongue dorsum at the onset of [zi] was articulated at a lower position than when the vocalic targets in the stem were lower. This is illustrated by deeper blue colors at the onset, as *TargetDistance* increases. The deflection at the vocalic target (roughly at a time point of 0.8) was shallower when *TargetDistance* was reduced. Öhman (1966) observed a similar pattern of “anti-anticipation.” In his study, lowering of F1 frequencies of [y] became more pronounced preceding [a] than when preceding [y] (see also Magen (1997) for a similar finding in one of the speakers). One possible explanation for this finding might be that anticipation of the vowels in the verb is modulating the tongue’s palatal bracing during [zi] articulation. Palatal bracing has not only been observed in [i], but also in [z] (McAuliffe et al., 2001; McLeod et al., 2006; Stone & Lundberg, 1996). Studies of anticipatory coarticulation have also shown increased palate bracing for [i] relative to [a] (Recasens, 1984, 1990). Greater palate bracing results in stronger lowering of the tongue dorsum due to a more pronounced concave shape of the tongue. This suggests that palate bracing during [zi] was greater when the verb contained an [i], than when the verb contained an [a], although this hypothesis has yet to be tested empirically.

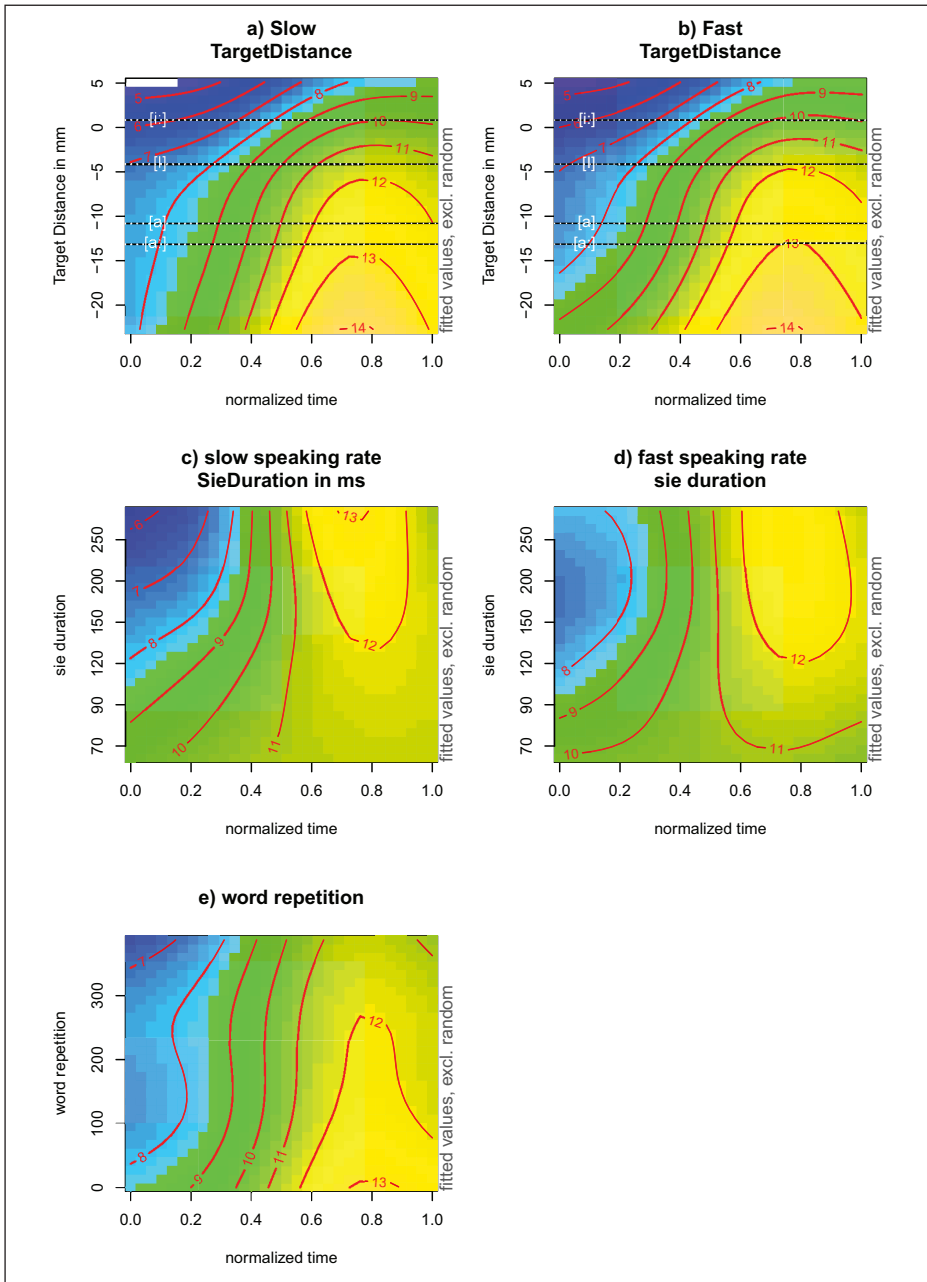


Figure 2. Estimated trajectories of the tongue dorsum across [zi] based on the generalized additive model. Random effects were excluded. X-axes represent normalized time, y-axes represent the modulating predictor. *SieDuration* values were back transformed to raw values for visualization. Contour lines connect areas of the same tongue dorsum height on the estimated surface. Values are in mm. Light yellow colors represent areas where the tongue is high, dark blue colors represent areas where the tongue is low (color online). For *TargetDistance*, the **horizontal lines** indicate the average tongue height of the stem vowel in the verb.

With regards to *SpeakingRateCondition*, the two conditions appear to have modulated the time point of the onset of the transition between [z] and [i]. It appears that this effect depends on *WordRepetition*, with the transition beginning earlier in the fast condition than in the slow condition.

The effect of *SieDuration* is illustrated in Figure 2 (c–d). The amplitude of the trajectory became larger and more distinct in longer words than in shorter words. Lower tongue dorsum positions were observed in the onset of [z], while higher tongue dorsum positions were seen in the [i]. Further, the tongue dorsum was articulated at lower positions in the fast than in the slow *SpeakingRateCondition*. This potentially reflects an effect of reduction due to a shorter articulation window during the experiment.

Figure 2 (e) illustrates the effect of *WordRepetition*. As can be seen, tongue dorsum positions across the entire [zi] articulation become lower—that is, more centralized—as the experiment progresses. This is likely to reflect a standard finding, namely fatigue due to repetition.

3.7 The variability around the average trajectory

Having discussed the control effects of speaking rate, word duration, anticipatory coarticulation, and repetition during the experiment on the average trajectory, we now turn our attention to our main hypothesis: examining whether the effects of practice are detectable in articulatory variability. Recall that we assessed the variability by means of the standard deviation term in the *gauss* GAMM, which represents the absolute deviation from the average trajectory.

The standard deviation in the model was fitted with a main effect for *SpeakingRateCondition*, and individual smooths for *WordRepetition*, *TraveledDistance* and *VowelProportions*. We included a main tensor product “te()” for a *Time* \times *SieProbability* interaction, fitting the main effects as well as the interaction simultaneously, and a partial tensor product “ti()” for a *Time* \times *TargetDistance* interaction.

The main effect for *TargetDistance* is missing, because in pilot analyses, it was found that it was not significantly predictive of standard deviation. The inclusion of *SieProbability* and the inclusion of the *Time* \times *SieProbability* interaction into the standard deviation model improved model fit, as supported by a decrease in AIC ($\Delta = 156.1 / 13.8$). Due to run time reasons, this was tested with a simpler model containing only a smooth for a *Time* \times *SpeakingRateCondition* for the average tongue dorsum trajectory and random factor smooths across *Time* by *Participants*. It is noteworthy that the effect of *SieProbability* was also present without the linguistic control predictors in both, the term fitting the average trajectory and the term fitting the standard deviations. We found that standard deviation around the average trajectory was significantly larger in the fast than in the slow *SpeakingRateCondition* (Part (A) of Table 2). This finding is in line with Fitts’ Law that movement variability is proportional to movement velocity (Fitts, 1954).

The non-linear effects of the standard deviation terms are shown in part (B) of Table 2. We found that standard deviation had a significant functional non-linear relation with *TraveledDistance* and *WordRepetition*. No significant interactions with time were found for the present data. This means that *TraveledDistance* and *WordRepetition* do correlate with variability during the entire movement trajectory, rather than only parts of it. The significant *Time* \times *SieProbability* interaction indicates that standard deviation around the average trajectory changed along the movement’s time-course in relation to *SieProbability*. The significant partial tensor smooth *Time* \times *TargetDistance* showed that this was also the case for *TargetDistance*. We also included random intercepts for *Speaker* and *Verb*. Figure 3 illustrates the estimated effects of standard deviation.

Standard deviations decreased for the first quantile of *TraveledDistance*, but increased steadily in the second to fourth quantile (Figure 3 (a)). This last effect mirrors the predictions of Fitts’ Law,

Table 2. Summary of partial effects in the GAMM model, fitting the **standard deviation** around the average tongue dorsum trajectory. Random effect terms are prefixed with “RE”. “s()” terms represent smooths fitting tongue dorsum positions as a function of one numerical predictor, “ti()” terms represent tensors fitting tongue dorsum positions as a function of a partial interaction between time and one numerical predictor; “te()” represents a main interaction. A term’s estimated degrees of freedom (“edf”) larger than 1 indicate non-linearity of the smooth/tensor when *p*-values are smaller than 0.001.

A) Parametric coefficients	Estimate	Std. error	z-value	p-value
(intercept)	0.340	0.081	4.225	< 0.001
SpeakingRateCondition = fast	0.164	0.010	15.891	< 0.001
B) Non-parametric effects	edf	Ref.df	F	p-value
s(TraveledDistance)	2.978	2.999	954.431	< 0.001
s(WordRepetition)	2.995	3.000	492.778	< 0.001
te(Time, SieProbability)	8.407	9.307	4162.077	< 0.001
ti(Time, TargetDistance)	7.398	8.374	252.802	< 0.001
RE - s(Participant)	15.959	16.000	13095.958	< 0.001
RE - s(NextWord)	117.987	125.000	2884.898	< 0.001

namely that variability increases with the distance between onset and target of the movement (Fitts, 1954). Note that *TraveledDistance* significantly correlated with *WordDuration* (Spearman’s rank correlation $\rho = 0.51$) and pilot analyses indicated that *WordDuration* had a similar effect on standard deviation.

Figure 3 (b) illustrates the partial effect of *WordRepetition*. The vertical gray dotted lines represent the average boundaries between the recording blocks. As can be seen, standard deviation decreases in the first two recording blocks, and then increases minimally in blocks three and four, before decreasing again in the last two blocks.⁶ Overall, this pattern indicates a reduction of variability across the experiment. This is consistent with our hypothesis that the repetition of one and the same tongue gesture would result in the same decrease in variance that has been observed in other kinematic studies (e.g., Sosnik et al., 2004). That is, consistent with what has been found for speed and smoothness (Tiede et al., 2011), short-term practice significantly also reduces articulatory variability. Given that we observed only a significant main effect for *WordRepetition*, these findings also seem to indicate that this decrease in articulatory variability as a result of practice occurs along the entire [zi] trajectory.

The effect on standard deviations of the interaction between *Time* and *SieProbability* are illustrated by means of a surface plot in Figure 3 (c). Changes in standard deviation are also color coded: dark blue colors represent a decrease of standard deviations, bright yellow colors represent an increase of standard deviations. It is possible that the exact position of the tongue dorsum during the production of [z] is less relevant than during the production of [i]. Accordingly, articulatory variability was greater at the onset of [zi], that is, during the [z] part, than in the center when the dorsum aimed for the vocalic center in [zi] (around 0.8 normalized time).

The observed reduction in variability was modulated by *SieProbability* as follows: In the onset of [zi], only a marginal effect of *SieProbability* can be observed. Towards the offset, as *SieProbability* increased, the variability around the average trajectory became significantly lower (mirrored by deeper shades of blue). The minimum of standard deviation across time coincides with the maximum deflection of the tongue dorsum movement, as observed in Figure 2. In other words: When

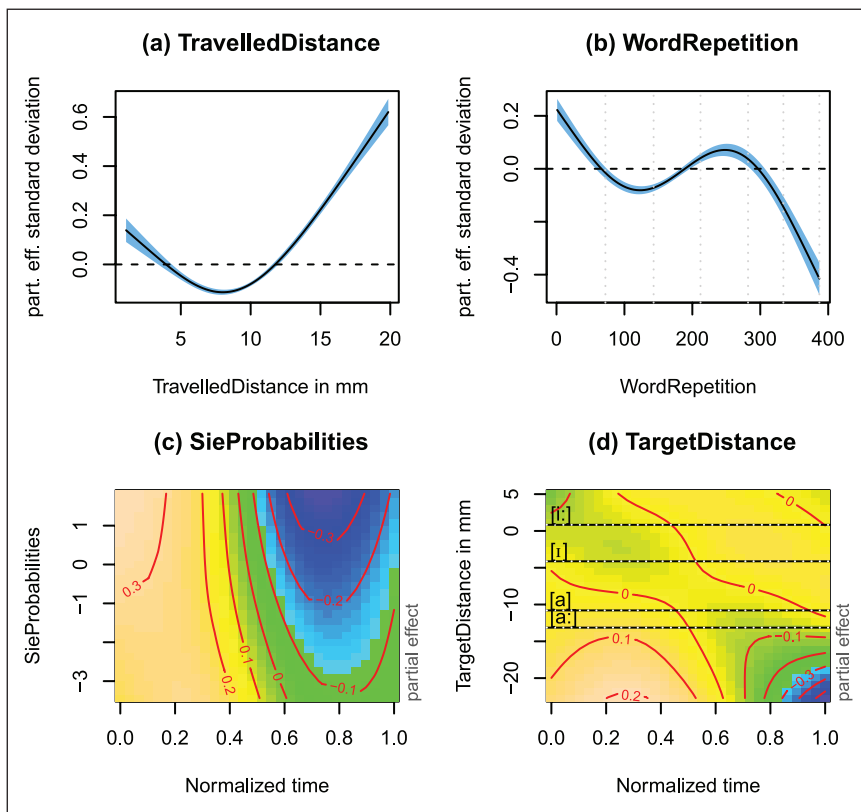


Figure 3. Partial effects for the **standard deviation** around the average trajectory of the tongue dorsum during [zi]. Top row: Smooths across (a) *TravelledDistance* and (b) *WordRepetition*, where x-axes represent the predictor and y-axes represent the partial effects for standard deviation. Dashed horizontal line represents zero. Bottom row: Partial effects for tensors, where x-axes represent normalized time, y-axes represent the (c) *SieProbabilities* and (d) *TargetDistance*. Light yellow colors represent areas of larger standard deviation, dark blue colors represent areas of smaller standard deviation.

the probability of [zi] was high, articulatory variability during the [i] target location was low and vice versa.

Finally, we turn our attention to the partial effect of *TargetDistance* (Figure 3 (d)). When [zi] was followed by verbs with high vowels, standard deviations decreased only slightly towards the onset of [zi] and increased only slightly towards the offset of [zi]. The effect was reversed when [zi] was followed by verbs with low vowels: standard deviations increase towards the onset of [zi] and decrease towards the offset of [zi]. One possible explanation of this difference in articulatory variability might be how the shape of the gesture's trajectory depends on the stem vowel in the upcoming verb. When the tongue body moves from [z] to [i] to a low vowel in the verb's stem, it describes an arch, whose peak is located at the time point of [zi]'s vocalic target (Iskarous, 2005). This is not the case with high vowels in the verb's stem. Instead, the tongue body describes a plateau when moving towards high vowels (see Figure 2 (a–b)). The downward movement towards low vowels potentially constraints the tongue's movement path which results in a reduction of variability at the offset of [zi]. By contrast, the plateau-like movement towards high vowels is less constrained, allowing for more variability.

4 Discussion

Practicing hand movements, that is, repeating a particular movement, makes the targets of the movement trajectory more predictable. Studies on movement kinematics have shown that this leads in turn to a reduction in the variability around the movement trajectory and in the target area (Georgopoulos et al., 1981; Pellizzer & Hedges, 2003; Sosnik et al., 2004).

Following from hand movement kinematics, we hypothesized that a corresponding effect of practice would be observed in articulatory variability. We operationalized practice both in terms of the repetition of gestures during an experiment, and their predictability based on the frequency with which speakers use articulatory gestures in speech. We tested our predictions by measuring movements of the tongue dorsum during the articulation of the German word *sie* [zi] “Engl. they” in multiple “[zi] + verb” phrases. Given that these predictions were supported by our results, we now turn to their theoretical implications.

The pronoun [zi] is one of the 50 most frequent German words (Arnold & Tomaschek, 2016; Faaß & Eckart, 2013; Shaoul & Tomaschek, 2013). It is thus clearly a highly practiced articulatory gesture for any native speaker of German. Therefore, one might reasonably ask whether the gestures associated with this word should even be susceptible to observable effects of predictability and repetition. There are a number of reasons why this might actually be the case.

Articulation has been shown to be influenced both by the words and gestures that a speaker previously articulated, and by the upcoming words and gestures that speakers will articulate. This is evidenced by findings on coarticulation (Katsika et al., 2015; Magen, 1997; Öhman, 1966; Whalen, 1990; Whalen et al., 2015) and systematic changes in the acoustic characteristics of words in relation to relative informativity of preceding words and the uncertainty associated with following words (Aylett & Turk, 2004, 2006; Bell et al., 2009; Tremblay & Tucker, 2011).

Learning offers a natural and parsimonious explanation of the effects of practice. In behavioral, neural, and even engineering models (Daw et al., 2008; Ng & Jordan, 2002; Rescorla & Wagner, 1972; Schultz, 2006; Schultz et al., 1997; Sutton & Barto, 1981), learning is normally formalized in terms of the way that experiences serve to influence the uncertainties associated about observable events (see discussions in Ramscar et al., 2010, 2013). These models treat learning as a process that is very sensitive to the success or failure of expectations, that is, whether or not the outcomes predicted by cues in earlier events actually occur.

Along with linguistic regularities that have been learned and encoded at higher levels of abstraction, from this perspective, words can be seen as acoustic cues which are informative to the speaker and the listener about the words that will be articulated after them. In addition to a word’s acoustics, the speaker can use its articulatory gestures as an additional set of cues informative about following words. As speakers conceptualize what they are going to say and plan the words that will be articulated, their abstract acoustic and gestural targets may also serve as cues for ongoing articulations. From this perspective, in our experiment, verbs serve as a source of uncertainty for [zi].

Since the frequencies of gestures, words, and other linguistic regularities vary systematically in natural speech (Linke & Ramscar, 2020; Zipf, 1935), it follows that the amount of information they provide will vary systematically as well. This kind of systematic variation can be seen in the identities and probabilities of the words that occur *before* the verb. Not only does German syntax allow word orders in which pronouns can occur in front of or after verbs, which goes along with high variability in the identity of words preceding the verb. Even in the “pronoun + verb” order, the verb forms used in the present experiment can also be preceded by the pronoun [viə] “Engl. we”.

It thus follows from this perspective that the uncertainty that occurs at the pronoun will vary considerably.

Thus, although [zi] is a highly frequent (i.e., probable) word—as are its gestural targets—when considered in isolation, in context its predictability will be subjected to a considerable amount of systematic variation. This variation will in turn affect the degree of uncertainty that is associated with the articulatory gestures that will have to be made in order to articulate it. This uncertainty will in turn be proportional to the amount of practice that a speaker gets with making any given gesture in context.

This then raises a question: what kind of mechanism might explain the relationship between practice and the precision with which gestures are articulated? According to articulatory phonology (Browman & Goldstein, 1986, 1989, 1992), articulatory gestures are defined by invariant tract variables which specify an articulator's degree of constriction, and its onset and offset timing relative to prosodic cycles. Contextual variation arises due to the phasing and blending between consecutive articulatory gestures. Effects of predictability and uncertainty such as those reported in the present study are not expected. In order to incorporate the current results into this framework, context sensitive constriction information, and potentially context sensitive timing information should be available in the lexicon.

In his DIVA model, Guenther (1995) proposed that articulators aim for sensory target areas. The size of target areas is proportional to speaking rate. In slow speech, target areas are smaller than in fast speech, resulting in smaller articulatory variability. A similar kind of mechanism could be responsible for the reduction of articulatory variability associated with practice and lower uncertainties about articulatory gestures.

Accordingly, it follows that from a learning perspective, when we observe the systematic patterns of articulatory variance even for highly practiced words like [zi], rather than being faced with accounting for how speakers acquire the knowledge that produces this variance, we must rather try to explain why it would be surprising if these systematically variable patterns of practice did *not* result in correspondingly systematic patterns in behavior (see e.g., Ramsar et al., 2014, 2017).

From the perspective we have so far described, learning is a discriminative process (Ng & Jordan, 2002) which serves to reduce the uncertainty associated with observations of and behavior in the world. The reduction of uncertainty emerges through dynamic interactions between learning, that is, the reinforcement of the relationships between cues and associated outcomes, and the unlearning of these relationships as the result of prediction error, that is, when the predicted outcome did not occur (Ramsar et al., 2010, 2013).

With regard to the current study, we can conceptualize the effects of learning in this way along the following lines: Whenever speakers articulate a “[zi] + verb” phrase in which the verb is highly likely to be preceded by [zi], the articulatory gestures of [zi] will be strongly reinforced in relation to the verb. Simultaneously, these particular articulatory gestures will be unlearned for all other contexts, especially in relation to verbs which are less likely to be preceded by [zi] (see Ramsar et al., 2013, for detailed description of this process). Whenever speakers articulate a “[zi] + verb” phrase in which the verb is very unlikely to be preceded by [zi], the reinforcement of the [zi] gestures will be relatively weak, as will be the unlearning for all the contexts in which [zi] is highly likely to occur. Critically, we are assuming that our participants' prior practice of the gestures they use to form [zi] did not occur in a vacuum, but rather that this practice occurred in and depended on its context. It also follows that this reasoning can be applied to similar contextual effects on the articulation of other words.

Seen like this, the contextual differences we actually observed were not merely unsurprising. They were in some sense inevitable, since these effects are predicted by virtually all behavioral and neuro-biological models of learning (Aizenberg et al., 2000; Daw et al., 2008; Dayan & Daw, 2008; Dechter, 1986; Hannun et al., 2014; Schultz, 2006; Schultz et al., 1997; Sutton & Barto,

1981). Moreover, this logic is also consistent with the results of numerous studies that have shown that speech production is not executed on a phone-by-phone and word-by-word basis but rather that anticipatory effects modulated by uncertainty can be found everywhere. For example, findings in studies of coarticulation (Katsika et al., 2015; Öhman, 1966; Ostry et al., 1996; Sziga, 1992; Tiede et al., 2011; Tomaschek, Tucker, et al., 2018; Whalen, 1990; Whalen et al., 2015), in studies of phonetic duration of single segments (Cohen Priva, 2015; Tomaschek, Plag, et al., 2019) and larger sequences (Bannard & Matthews, 2008; Shaoul et al., 2013; Tremblay & Tucker, 2011) are very likely to be reflecting the same underlying learning processes.

Our data clearly show that variability is proportional to uncertainty. As uncertainty decreases, variability decreases; as uncertainty increases, variability increases. Nevertheless, we should sound a note of caution with regard to the sources of increased variability. The nature of linguistic distributions guarantees that as the frequency of lexical events decreases, levels of individual practice will increasingly vary (Ramscar et al., 2014). To give a trivial example, the word “formant” is a common word in the vocabularies of most readers of this paper, yet it is likely that the majority of the population are entirely unfamiliar with this word. This means that the increased variability we see with lower probability events could reflect either the increased variability we would expect across individuals because of the distributional facts. Or else it might reflect increased variability within individuals as a result of diminished practice (as is the case at the beginning of our experiment). It is very likely that both forces are at work. We leave it to future studies to disentangle their relative contributions to the variability we observe in articulation.

Acknowledgements

We would like to thank the associate editor and two anonymous reviewers for their invaluable comments on previous versions of the manuscript that led to improvements in the exposition. Our thanks also go to our student assistants Mathias Müller from the University of Zürich and Josh Ring from the University of Tübingen for their inexhaustible willpower while correcting all instances of [zi].

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded in part by the Alexander von Humboldt Chair awarded to R. H. Baayen (grant 1141527) and by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 “Spoken Morphology,” Project “Articulation of morphologically complex words,” BA 3080/3-1).

ORCID iDs

Fabian Tomaschek  <https://orcid.org/0000-0003-2246-663X>

Benjamin V. Tucker  <https://orcid.org/0000-0001-8965-7890>

Konstantin Sering  <https://orcid.org/0000-0002-1178-7932>

Notes

1. [zi] typically has three meanings/uses in Standard German. It can serve as the polite second person pronoun, the singular female third person pronoun, and the third person plural pronoun. Only the latter use was employed in this study, which we will refer to henceforth by its transcription [zi] notwithstanding any possible reductions of the consonant or the vowel during actual articulations of [zi]. In Section “The vowel in [zi]” in the Appendix, we show that in fact [i] was not reduced in the articulations produced by participants in the current experiment.
2. A series of complementary investigations of the production of the same words targeting different gestures of interest and different effects are reported in Tomaschek et al. (2014) and reanalyzed in Tomaschek, Tucker, et al. (2018). The effects studied include frequency of use on the articulation of stem vowels in German verbs presented in the context of *sie* “they” or *ihr* “you, pl.”.

3. We would like to thank an anonymous reviewer for the suggestion to investigate the effect of the verbal stem vowel in this way.
4. Following the suggestion of an anonymous reviewer, we also tested how the shape of the average articulatory trajectory changed in relation to the place of articulation of the verb's onset consonant. No significant intercept difference was found between labial and coronal consonants ($\beta = 0.16663$, std. error = 0.170, z-value = 0.978, $p = 0.328$). Place of articulation showed a marginally significant difference of the smooths for *Time* when preceding coronal consonants and when preceding labial consonants (edf = 1.547, Ref.df = 1.693, Chi.sq = 6.832, $p = 0.015$). In spite of this finding, we refrained from an inclusion of place of articulation as a predictor to the model. First, the trajectory differed only by roughly 0.1 to 0.2 mm between labial and coronal, which is relatively small in comparison to the trajectory's amplitude of roughly 4 to 6 mm; second, its inclusion did not significantly improve the model fit, as indicated by AIC and ML-score comparison.
5. The model presented in the current paper is the result of exploratory pilot analyses following the two-step procedure as applied by Chodroff and Wilson (2017), Sonderegger (2015), and Sonderegger et al. (2017). We performed top-down GAMM analyses fitting only the average tongue dorsum trajectory and testing the inclusion of predictors and interaction with an χ^2 test (using the `compareML()` function in the **itsadug** package, (van Rij et al., 2015)). Following the procedure of Sonderegger and colleagues, we then submitted the residuals of that *mean* model to another round of analyses using GAMMs, testing the inclusion and exclusion of model terms. In both rounds, we accounted for autocorrelation in the residuals of lag 1 by an AR(1) parameter (cf. Wieling et al., 2015, 2016). This two-step procedure was applied because of the high calculation costs of the *gaulss* model. Although the *gaulss* family does not provide an AR(1) parameter, that is, a parameter that accounts for autocorrelation in the residuals, the estimates for both the average tongue dorsum trajectory and the standard deviations from the *gaulss* model correlated strongly with the estimates from the two-step procedure (Pearson's product-moment correlation $R = 0.99$, $p < 0.001$, for the average trajectory and $R = 0.96$, $p < 0.001$, for the standard deviations). All analyses performed for this study can be found in Supplementary Materials, available from <https://osf.io/r78mk/>.
6. The non-linear pattern in Figure 3 (b) has very likely multiple sources. As has been shown in Tomaschek and Leeman (2018), participants differ in the size of their oral cavities. Differences in oral cavities might affect the participants' reduction strategies. In addition, participants might differ in terms of how they cope with the slow/fast alternations between blocks and with fatigue during the experiment. Finally, participants might differ in their degree of learning during the experiment. The present data is not suited for investigating all these potential sources.

References

- Aizenberg, I., Aizenberg, N. N., & Vandewalle, J. P. (2000). *Multi-valued and universal binary neurons: Theory, learning and applications*. Springer Science & Business Media.
- Arnold, D., & Tomaschek, F. (2016). The Karl Eberhards Corpus of spontaneously spoken Southern German in dialogues—audio and articulatory recordings. In C. Draxler & F. Kleber (Eds.), *Tagungsband der 12. tagung phonetik und phonologie im deutschsprachigen raum. p und p12. 12. tagung phonetik und phonologie im deutschsprachigen raum, 12.–14. oktober 2016* (pp. 9–11). Ludwig-Maximilians-Universität München. <https://ids-pub.bsz-bw.de/frontdoor/index/index/year/2017/docId/5944>
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Language*, 62, 67–82.
- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56.
- Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119(5), 3048–3058.
- Baayen, R. H. (2001). *Word frequency distributions*. Kluwer Academic Publishers.
- Baayen, R. H., Tomaschek, F., Gahl, S., & Ramscar, M. (2017). The Ecclesiastes principle in language change. In M. Hundt, S. Mollin, & S. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives* (pp. 21–48). Cambridge University Press.

- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241–248.
- Bays, P. M., & Wolpert, D. M. (2007). Computational principles of sensorimotor control that minimize uncertainty and variability. *The Journal of Physiology*, 578(2), 387–396. <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.2006.120121>
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92–111.
- Bell-Berti, F., & Harris, K. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 65(5), 1268–1270.
- Belmont, A. (2011). *Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter across the lifespan: An ultrasound study*. Doctoral Dissertation, University of South Florida.
- Bourgeois, F., & Hay, L. (2003). Information processing and movement optimization during development: Kinematics of cyclical pointing in 5- to 11-year-old children. *Journal of Motor Behavior*, 35(2), 183–195.
- Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology*, 3, 219–252.
- Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3–4), 155–180.
- Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61, 30–47.
- Churchland, M. M., Afshar, A., & Shenoy, K. V. (2006). A central source of movement variability. *Neuron*, 52(6), 1085–1096. <http://www.sciencedirect.com/science/article/pii/S0896627306008713> <https://doi.org/10.1016/j.neuron.2006.10.034>
- Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, 6(2), 243–278.
- Cychosz, M. (2020). Practice and experience predict coarticulation in child speech. <https://doi.org/10.31234/osf.io/vwhtk>
- Darling, W., Cole, K., & Abbs, J. (1988). Kinematic variability of grasp movements as a function of practice and movement speed. *Experimental Brain Research*, 73(2), 225–235.
- Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. *The Probabilistic Mind*, 431–452.
- Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 429–453.
- Dechter, R. (1986). *Learning while searching in constraint-satisfaction problems*. University of California, Computer Science Department, Cognitive Systems.
- Faaß, G., & Eckart, K. (2013). Sdewac—a corpus of parsable sentences from the web. In I. Gurevych, C. Biemann, & T. Zesch (Eds.), *Language processing and knowledge in the web* (pp. 61–68). Springer.
- 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.
- Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 31(4), 307–319.
- Fowler, C. A., & Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, 42(1), 1–41.
- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26(5), 489–504.
- Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2–3), 171–195.
- Fuchs, S., Winkler, R., & Perrier, P. (2008). Do speakers' vocal tract geometries shape their articulatory vowel space? In *8th International Seminar on Speech Production, Strasbourg, France* (pp. 333–336).
- Gahl, S. (2008). Thyme and time are not homophones: Word durations in spontaneous speech. *Language*, 84(3), 474–496.
- Georgopoulos, A., Kalaska, J., & Massey, J. (1981). Spatial trajectories and reaction times of aimed movements: Effects of practice, uncertainty, and change in target location. *Journal of Neurophysiology*, 46(4), 725–743.

- Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The breadth of coarticulatory units in children and adults. *Journal of Speech, Language, and Hearing Research*, 51(6), 1424–1437.
- Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research*, 45(1), 66–79.
- Gribble, P. L., & Ostry, D. J. (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), 2853–2860. <http://jn.physiology.org/content/76/5/2853>
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Biological Cybernetics*, 72(1), 43–53.
- Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., Ng, A. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.
- Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394(6695), 780.
- Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1993). A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. *Language and Speech*, 36(2–3), 235–260.
- Iskarous, K. (2005). Patterns of tongue movement. *Journal of Phonetics*, 33(4), 363–381.
- Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 69(3), 505–528.
- Katsika, A., Whalen, D. H., Tiede, M., & King, H. (2015). Articulatory measures of planned and unplanned coarticulation. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS iixx)*, Glasgow.
- Kim, S., Carlton, L. G., Liu, Y., & Newell, K. M. (1999). Impulse and movement space–time variability. *Journal of Motor Behavior*, 31(4), 341–357.
- Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, 124(5), 3158–3170. <https://doi.org/10.1121/1.2981639>
- Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, 122(2), 81–91.
- Linke, M., & Ramscar, M. (2020). How the probabilistic structure of grammatical context shapes speech. *Entropy*, 22(1), 90.
- Madison, G., Karampela, O., Ullén, F., & Holm, F. (2013). Effects of practice on variability in an isochronous serial interval production task: Asymptotical levels of tapping variability after training are similar to those of musicians. *Acta Psychologica*, 143(1), 119–128.
- Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 25, 187–205.
- Mather, E., Jones, L. L., & Estes, Z. (2014). Priming by relational integration in perceptual identification and stroop colour naming. *Journal of Memory and Language*, 71(1), 57–70. <http://www.sciencedirect.com/science/article/pii/S0749596X13000983> <https://doi.org/10.1016/j.jml.2013.10.004>
- McAuliffe, M. J., Ward, E. C., & Murdoch, B. E. (2001). Tongue-to-palate contact patterns and variability of four English consonants in an/i/vowel environment. *Logopedics Phoniatrics Vocology*, 26(4), 165–178.
- McLeod, S., Roberts, A., & Sita, J. (2006). Tongue/palate contact for the production of/s/and/z. *Clinical Linguistics & Phonetics*, 20(1), 51–66.
- Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in Neural Information Processing Systems* (pp. 841–848).
- Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from cCantonese segment and tone perception. *Journal of Memory and Language*, 90, 103–125.
- Noiray, A., Ménard, L., & Iskarous, K. (2013, 01). The development of motor synergies in children: Ultrasound and acoustic measurements. *The Journal of the Acoustical Society of America*, 133, 444–452. <https://doi.org/10.1121/1.4763983>

- Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39(151), 151–168.
- Ostry, D., Gribble, P., & Gracco, V. (1996). Coarticulation of jaw movements in speech production: Is context sensitivity in speech kinematics centrally planned? *Journal of Neuroscience*, 16(4), 1570–1579.
- Pellizzer, G., & Hedges, J. H. (2003). Motor planning: Effect of directional uncertainty with discrete spatial cues. *Experimental Brain Research*, 150(3), 276–289. <https://doi.org/10.1007/s00221-003-1453-1>
- Perkell, J., & Nelson, W. (1982). Articulatory targets and speech motor control: A study of vowel production. In *Speech Motor Control* (pp. 187–204). Elsevier.
- Platz, T., Brown, R., & Marsden, C. (1998). Training improves the speed of aimed movements in Parkinson's disease. *Brain*, 121, 505–513.
- Raeder, C., Fernandez-Fernandez, J., & Ferrauti, A. (2015). Effects of six weeks of medicine ball training on throwing velocity, throwing precision, and isokinetic strength of shoulder rotators in female handball players. *Journal of Strength and Conditioning Research*, 29(7), 1904–1914.
- Ramscar, M., Dye, M., & McCauley, S. (2013). Error and expectation in language learning: The curious absence of “mouses” in adult speech. *Language*, 89(4), 760–793.
- Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, R. H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6(1), 5–42.
- Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, R. H. (2017). The mismeasurement of mind: Life-span changes in paired-associate-learning scores reflect the “cost” of learning, not cognitive decline. *Psychological Science*, 28(8), 1171–1179.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909–957.
- Rapp, S. (1995). Automatic phonemic transcription and linguistic annotation from known text with hidden Markov models / an aligner for German. In *Proceedings of Elsnets East and iMacs Workshop, Moscow*.
- Recasens, D. (1984). Vowel-to-vowel coarticulation in catalan vcv sequences. *The Journal of the Acoustical Society of America*, 76(6), 1624–1635.
- Recasens, D. (1990). An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, 101(102), 130–148.
- Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64–69). Appleton Century Crofts.
- Rubertus, E., & Noiray, A. (2018, 09). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PLOS ONE*, 13(9), 1–21. <https://doi.org/10.1371/journal.pone.0203562>
- Rubertus, E., & Noiray, A. (2020). Vocalic activation width decreases across childhood evidence from carryover coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 11(1), 1–27.
- Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333–382.
- Scharinger, M. (2006). *The representation of vocalic features in vowel alternations: Phonological, morphological and computational aspects* (Unpublished doctoral dissertation).
- Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., & Quinn, J. T. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review*, 86(5), 415.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87–115.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14(3), 369–385.
- Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Shaoul, C., & Tomaschek, F. (2013). *A phonological database based on celex and n-gram frequencies from the SDeWaC corpus*. https://fabiantomaschek.files.wordpress.com/2016/07/tomaschek_corpus_readme.pdf

- Shaoul, C., Westbury, C. F., & Baayen, H. R. (2013). The subjective frequency of word n-grams. *Psihologija*, 46(4), 497–537.
- Sonderegger, M. (2015). Trajectories of voice onset time in spontaneous speech on reality TV. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS ixx)*, Glasgow.
- Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, 93(3), 598–640.
- Sosnik, R., Hauptmann, B., Karni, A., & Flash, T. (2004). When practice leads to co-articulation: The evolution of geometrically defined movement primitives. *Experimental Brain Research*, 156, 422–438.
- Stone, M., & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*, 99(6), 3728–3737.
- Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88(2), 135.
- Sziga, E. C. (1992). Acoustic evidence for gestural overlap in consonant sequences. *Haskins Laboratories Status Report on Speech Research*, SR-111/112, 43–62.
- Tiede, M., Mooshammer, C., Goldstein, L., Shattuck-Hufnagel, S., & Perkell, J. (2011). Motor learning of articulator trajectories in production of novel utterances. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPHS ixx)*, Hong Kong.
- Tomaschek, F., Arnold, D., Broecker, F., & Baayen, R. H. (2018). Lexical frequency co-determines the speed-curvature relation in articulation. *Journal of Phonetics*, 68, 103–116.
- Tomaschek, F., & Leeman, A. (2018). The size of the tongue movement area affects the temporal coordination of consonants and vowels—a proof of concept on investigating speech rhythm. *The Journal of the Acoustical Society of America*, 144(5), EL410–EL416.
- Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2019). Modeling the duration of word-final s in English with naïve discriminative learning. *Journal of Linguistics*, 1–39.
- Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2013). Neural processing of acoustic duration and phonological German vowel length: Time courses of evoked fields in response to speech and nonspeech signals. *Brain and Language*, 124(1), 117–131. <http://dx.doi.org/10.1016/j.bandl.2012.11.011>
- Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2015). Discrimination sensitivities and identification patterns of vowel quality and duration in German /u/ and /o/ instances. In A. Leemann, M.-J. Kolly, S. Schmid, & V. Dellwo (Eds.), *Trends in phonetics and phonology: Studies from German speaking Europe*. Lang.
- Tomaschek, F., Tucker, B. V., Baayen, R. H., & Fasiolo, M. (2018). Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistic Vanguard*, 4(s2), 1–13.
- Tomaschek, F., Tucker, B. V., Ramscar, M., & Baayen, R. H. (submitted). Paradigmatic enhancement of stem vowels in regular English inflected verb forms.
- Tomaschek, F., Tucker, B. V., Wieling, M., & Baayen, R. H. (2014). Vowel articulation affected by word frequency. In *Proceedings of the 10th ISSP, Cologne* (pp. 425–428).
- Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience. In *Proceedings of Interspeech, Lyon*.
- Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication*. Continuum.
- Tremblay, A., & Tucker, B. V. (2011). The effects of n-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2), 302–324.
- Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadic, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior Research Methods*, 51(3), 1187–1204.
- van Beers, R. J., Baraduc, P., & Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 357(1424), 1137–1145. <https://doi.org/10.1098/rstb.2002.1101>
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). *ITSADUG: Interpreting time series, autocorrelated data using GAMMs*. (R package version 0.8)
- Viviani, P., & Schneider, R. (1991). A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, 17(1), 198–218.

- Viviani, P., & Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience*, 7(2), 431–437.
- Weirich, M., & Fuchs, S. (2006). Palatal morphology can influence speaker-specific realizations of phonemic contrasts. *Journal of Speech, Language, and Hearing Research*, 56, 1894–1908.
- Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18(1), 3–35.
- Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *The Journal of the Acoustical Society of America*, 90(4), 2311–2311.
- Whalen, D. H., Katsika, A., Tiede, M., & King, H. (2015). Acoustic measures of planned and unplanned coarticulation. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS iixx)*, Glasgow.
- Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., & Baayen, R. H. (2015). Investigating dialectal differences using articulatory data. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPHS iixx)*, Glasgow.
- Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S., Baayen, R. H. (2016). Investigating dialectal differences using articulatory data. *Journal of Phonetics*, 59, 122–143.
- Wood, S., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563. <https://doi.org/10.1080/01621459.2016.1180986>
- Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.
- Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15(1), 118–140.
- Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sv/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42(2), 193–208.
- Zipf, G. (1935). *The psycho-biology of language: An introduction to dynamic philology*. MIT Press.

Appendix

The vowel in [zi]

German pronouns are usually articulated with reduced vowels and durations (i.e., rather [ə] like), partially because they are high-frequency words. It is therefore possible that the vowel in [zi] was strongly reduced and resembles rather an [ə] than an [i] in the current data set. To inspect this possibility we fitted [i] durations from [zi] using a linear mixed-effect model analysis, controlling for effects of *SpeakingRateCondition* and *SieDurations*. Speakers and following verbs were included as random intercepts. Mean [i] duration in [zi] was 0.085 seconds when the verb contained a long, tense [i] and it was significantly longer when the verb contained the other three vowels ($\beta = 0.0052 / 0.0038 / 0.005, sd = 0.0016 / 0.0016 / 0.002, t = 3.16 / 2.33 / 2.48$ for [a], [ɪ] and [a:], respectively). Releveling the following vowel to [ɪ] indicated that there was no significant difference in [i] in [zi] depending on the latter three vowels. The average duration of [i] in [zi] is located at the category boundary between short and long vowels for German speakers (as found for [a] vs. [a:] Tomaschek, Truckenbrodt, & Hertrich, 2013, 2015). We also tested whether *TargetDistance* differed by vowel class in the following verb, using the same model structure testing duration differences.

VowelProportions did not yield any significant effect. We found that [i] in [zi] was not significantly different from [i:] in the stem (*Intercept* = 0.8 , *SDE* = 0.49 , *t* = 1.62), while all other vowels in the verb ([ɪ], [a] and [a:]) were significantly lower than [i] in [zi] ($\beta = -5.0 / -11.78 / -14.0, SDE = 0.42 / 0.42, 0.52$, *t* = -11.7 / -27.6 / -27.0 , respectively). This means that speakers indeed targeted a tense [i] in [zi] and did not reduce it to [ə], inspite of its relatively short duration, as is frequently the case for this pronoun. One possibility, why this was the case is that the experiment at hand targeted the plural, rather than the singular pronoun.

Stimuli

Table 3. Verbs used in the “sie+verb” phrases with [ɪ] as stem vowels. Counts represent google hits. The number of google hits for “sie” was 221000000.

Verb	Translation	Transcription	C(verb)	C(sie+verb)	P(sie verb)
binden	bind	[bɪndən]	16300000	124000	0.00761
bitten	ask	[bɪtən]	65700000	677000	0.01030
dingen	hire	[dɪŋən]	53500000	3340	0.00006
filmen	film	[fɪlmən]	49500000	33500	0.00068
filtern	filter	[fɪltən]	42400000	35400	0.00083
finden	find	[fɪndən]	636000000	13900000	0.02186
finger	finger	[fɪŋən]	6880000	14600	0.00212
fischen	fish	[fɪʃən]	9500000	21000	0.00221
liften	lift	[lɪftən]	2960000	399	0.00013
lindern	alleviate	[lɪndən]	1430000	21200	0.01483
lispeln	lisp	[lɪspələn]	112000	1360	0.01214
nippen	nip	[nɪpən]	273000	1380	0.00505
nisten	nest	[nɪstən]	449000	20800	0.04633
mildern	alleviate	[mɪldən]	526000	6350	0.01207
mischen	mix	[mɪʃən]	11900000	91400	0.00768
pinkeln	pee	[pɪŋkələn]	817000	24700	0.03023
pinseln	paint	[pɪnzələn]	443000	637	0.00144
schillern	shimmer	[ʃɪlən]	130000	2710	0.02085
schimmeln	mold	[ʃɪməln]	269000	3840	0.01428
schimmern	shimmer	[ʃɪmə]	443000	13400	0.03025
schimpfen	rant	[ʃɪmpfən]	605000	23600	0.03901
schinden	maltreat	[ʃɪndən]	339000	1630	0.00481
schippen	scoop	[ʃɪpən]	149000	1580	0.01060
schippern	sail	[ʃɪpən]	235000	2170	0.00923
sinnen	muse	[zɪnən]	5650000	5450	0.00096
singen	sing	[zɪŋən]	25100000	320000	0.01275
sinken	sink	[zɪŋkən]	6730000	42100	0.00626
sitzen	sit	[zɪtsən]	36100000	505000	0.01399
tilgen	repay	[tɪlgən]	479000	5590	0.01167
tingeln	travel	[tɪŋələn]	82900	1020	0.01230
tippeln	typen	[tɪpələn]	37700	347	0.00920
tippen	type	[tɪpən]	13200000	47900	0.00363
tischlern	carpent	[tɪʃlən]	127000	456	0.00359
widmen	dedicate	[vɪtmən]	6470000	42200	0.00652
wildern	poach	[vɪldən]	403000	1570	0.00390
windeln	change diapers	[vɪndələn]	1550000	23700	0.01529
winden	wind	[vɪndən]	5100000	9680	0.00190
winkeln	angle	[vɪŋkələn]	637000	1300	0.00204
winseln	whine	[vɪnzələn]	140000	2330	0.01664
wispern	whisper	[vɪspən]	132000	8170	0.06189
wissen	know	[vɪsən]	227000000	10400000	0.04581
witzeln	joke	[vɪtsələn]	82200	881	0.01072
zimmern	carpenter	[tsɪmə]	12800000	1510	0.00012
zischeln	hiss	[tsɪʃələn]	33100	885	0.02674
zittern	shiver	[tsɪtən]	3030000	29200	0.00964

Table 4. Verbs used in the “sie+verb” phrases with [i:] as stem vowels. Counts represent google hits. The number of google hits for “sie” was 221000000.

Verb	Translation	Transcription	C(verb)	C(sie+verb)	P(sie verb)
bieten	offer	[bi:tən]	184000000	12600000	0.06848
dienen	serve	[di:nən]	56500000	709000	0.01255
fiedeln	fiddle	[fi:dələn]	35900	1490	0.04150
lieben	love	[li:bən]	74800000	949000	0.01269
liefern	deliver	[li:fən]	51100000	427000	0.00836
mieten	rent	[mi:tən]	43300000	272000	0.00628
niesen	sneeze	[ni:zən]	696000	9640	0.01385
nieten	rivet	[ni:tən]	1530000	6390	0.00418
schieben	push	[ʃi:bən]	9310000	65600	0.00705
schielen	squint	[ʃi:lən]	462000	5330	0.01154
schienen	rail	[ʃi:nən]	10500000	156000	0.01486
sieben	sieve	[zi:bən]	56100000	150000	0.00267
siedeln	settle	[zi:dələn]	417000	12000	0.02878
sieden	boil	[zi:dən]	465000	5810	0.01249
siezen	address formally	[zi:tsən]	207000	3040	0.01469
wienern	polish	[vi:nən]	186000	457	0.00246
zielen	target	[tsi:lən]	8510000	75700	0.00890
ziemen	behave	[tsi:mən]	89100	507	0.00569

Table 5. Verbs used in the “sie+verb” phrases with [a:] as stem vowels. Counts represent google hits. The number of google hits for “sie” was 221000000.

Verb	Translation	Transcription	C(verb)	C(sie+verb)	P(sie verb)
baden	take a bath	[ba:d:n]	189000000	113000	0.00060
bahnen	channel	[ba:nən]	8620000	6460	0.00075
faseln	babble	[fa:zələn]	122000	3450	0.02828
labern	babble	[la:bən]	454000	11700	0.02577
lahmen	lame	[la:mən]	392000	1370	0.00349
latschen	slouch along	[la:tʃən]	442000	3370	0.00762
mahlen	grind	[ma:lən]	600000	11200	0.01867
malen	paint	[ma:lən]	25100000	90200	0.00359
mahnen	remind	[ma:nən]	476000	15900	0.03340
schaben	scrape	[ʃa:bən]	532000	2020	0.00380
schaden	damage	[ʃa:dən]	45300000	128000	0.00283
tadeln	criticize	[ta:dələn]	306000	9400	0.03072
tafeln	feast	[ta:fələn]	7420000	2210	0.00030
tapern	be insecure	[ta:pən]	27300	545	0.01996
waten	wade	[va:tən]	406000	3470	0.00855
zahlen	pay	[tsa:lən]	77300000	779000	0.01008
zahnen	teethe	[tsa:nən]	431000	2960	0.00687

Table 6. Verbs used in the “sie+verb” phrases with [a] as stem vowels. Counts represent google hits. The number of google hits for “sie” was 221000000.

Verb	Translation	Transcription	C(verb)	C(sie+verb)	P(sie verb)
ballen	bale	[balən]	14400000	3590	0.00025
ballern	shoot	[balən]	452000	2900	0.00642
balgen	scrap	[balgən]	308000	2030	0.00659
balzen	mate	[baltsən]	175000	2840	0.01623
bangen	fear	[baŋən]	844000	8320	0.00986
basteln	tinker	[bastələn]	23700000	37200	0.00157
danken	thank	[dan̩kən]	17300000	31400	0.00182
fallen	fall	[falən]	262000000	469000	0.00179
fangen	catch	[faŋən]	11200000	213000	0.01902
fassen	catch	[fasən]	12900000	75200	0.00583
fasten	fast	[fastən]	21100000	23500	0.00111
landen	land	[landən]	78000000	130000	0.00167
langen	long	[laŋən]	57800000	13400	0.00023
lallen	slur	[lalən]	296000	1940	0.00655
lassen	let	[lasən]	292000000	818000	0.00280
malmen	malmen	[malmən]	429000	320	0.00075
mangeln	lack	[maŋələn]	451000	2250	0.00499
naschen	nibble	[naʃən]	696000	7030	0.01010
paddeln	paddle	[padələn]	642000	8460	0.01318
panschen	panchen	[paŋʃən]	45400	760	0.01674
panzern	armour	[pant̩sən]	467000	452	0.00097
passen	fit	[pasən]	67000000	510000	0.00761
pataschen	patching	[patʃən]	160000	490	0.00306
salben	ointment	[zalbən]	748000	4290	0.00574
salzen	salts	[zaltsən]	878000	1440	0.00164
satteln	saddle	[zatələn]	264000	1800	0.00682
schalten	go to	[ʃaltən]	20200000	98800	0.00489
sabbern	drool	[zabən]	439000	5130	0.01169
sabbeln	slobbering	[zabələn]	59500	446	0.00750
sammeln	collect	[zamələn]	84600000	389000	0.00460
schaffen	create	[ʃafən]	54300000	454000	0.00836
schallen	resound	[ʃalən]	334000	1860	0.00557
tanken	refuel	[taŋkən]	27300000	21200	0.00078
tanzen	dancing	[tansən]	30800000	187000	0.00607
tappen	Tap	[tapən]	772000	6340	0.00821
tapsen	taps	[tapsən]	72500	810	0.01117
tasten	buttons	[tastən]	14100000	20100	0.00143
Wallen	boil	[valən]	5400000	3540	0.00066
Walten	govern	[valtən]	760000	6170	0.00812
Walzen	rolling	[valtsən]	878000	5560	0.00633
wandeln	convert	[vandələn]	2560000	66000	0.02578
wandern	hiking	[vandən]	32800000	177000	0.00540
Wanken	waver	[vaŋkən]	588000	5540	0.00942
Zanken	quarrel	[tsaŋkən]	249000	6950	0.02791
Zapfen	tap	[tsapfən]	429000	4250	0.00991
Zappeln	fidget	[tsapələn]	402000	9300	0.02313