Precision in articulation
fabian.tomaschek@uni-tuebingen.de [1]
Version: January 13, 2018

Effects of predictability and practice on the movement precision during the execution of articulatory gestures

Fabian Tomaschek

Department of General Linguistics, University of Tübingen, Germany

Benjamin V. Tucker

Department of Linguistics, University of Alberta, Canada

Konstantin Sering

Department of General Linguistics, University of Tübingen, Germany

Denis Arnold

Institut für Deutsche Sprache, Mannheim, Germany

Jacolien van Rij

Neurolinguistics and Language Development, University of Groningen, The Netherlands

Michael Ramscar

Department of General Linguistics, University of Tübingen, Germany

---

Abstract

Typically, when people repeat the movements involved in playing instruments or sports – i.e. *practice* – their kinematic behavior improves: associated movements become more precise. Similar findings have been reported for articulation, with adults' articulatory gestures being more precisely executed than those of children. Given that practicing speech across the lifespan appears to improve production, it would seem to follow that the frequency with which articulatory movements are usually made ought to have a similar effect. To address this question, the present study investigated articulatory precision during the transition between the German word [ziː] 'they' and the following verb, while controlling for the vowel contained in the verb. The articulatory "practice" of [ziː] with each verb was operationalized as the conditional probability of the verb given [ziː], i.e. the degree to which experience of a verb is conditioned on [ziː] in context. Considering findings in relation to practice, we find that when the verb contained an [a], i.e. required a *different* vocalic gesture, higher conditional probabilities were associated with greater precision of articulation. By contrast, when the verb contained an [i], such that its production involved the *same* vocalic gesture, no correlation between conditional probability and precision was found. These results indicate that speakers tend to gravitate towards an optimal articulatory trajectory, and that depending on the amount of practice a speaker has with a given gesture, they will tend to follow its path with more or less precision.

---

1. Introduction

It is well know that kinematic movements show a speed-accuracy trade-off (Fitts, 1954; Wright and Meyer, 1983; Bertucco and Cesari, 2010): the same movement gesture, when executed faster or longer, will tend to be executed less precisely, i.e. with larger variability; slower or shorter movements will tend to be executed more precisely. Practice, i.e. the repetition of one and the same movement such as throwing a ball, will counteract this trade-off, with either increasing the speed while movement precision remains constant or increasing precision while speed stays constant (Georgopoulos et al., 1981; Darling et al., 1988; Platz et al., 1998; Sosnik et al., 2004; Madison et al., 2013; Raeder et al., 2015).

While speech production is itself the result of a set of kinematic movements, it is also clear that articulation and throwing a ball differ a great deal. Hand movements are effected with a relatively rigid body joint-angle system and usually have one target (cf. as throwing or catching a flying ball, fetching a cup or even hand writing (Bourgeois and Hay, 2003)). By contrast, the articulatory apparatus consists of a rigid bone structure (the jaw) and muscular hydrostats (the tongue). Whereas these are considered to be independent articulators, they are biomechanically joined and task-dynamically coupled (Bell-Berti and Harris, 1979; Saltzman and Munhall, 1989; Fowler and Saltzman, 1993) and produce successive gestures with their own, partially competing, targets which result in a strong overlap between adjacent gestures (Öhman, 1966; Browman and Goldstein, 1986; Magen, 1997).

This underlying kinematic complexity of speech production in contrast to hand movements raises the question how and to what extent practice increases precision during articulation. Studies of the timing of articulation indicate that this may be the case since practice has similar effects onto the timing of articulation like it has onto hand movements (Tiede et al., 2011; Tomaschek et al., under revision 1). In addition, several studies investigating articulation in children and adults show effects are consistent with the effects of practice.

Goffman et al. (2008) investigated anticipatory lip rounding of upcoming rounded vowels in children and adults. Lip rounding has been shown to be time-locked to the vowel's acoustic onset, independently of the number of consonants in the preceding cluster (Bell-Berti and Harris, 1979). Goffman et al. (2008) showed that children's anticipatory lip rounding was executed with smaller temporal precision than that of adults. Similar findings have been obtained from studies of the way that vowel production is affected by anticipatory coarticulation Zharkova et al. (2011, 2012); Belmont (2011). Further convergent findings come from

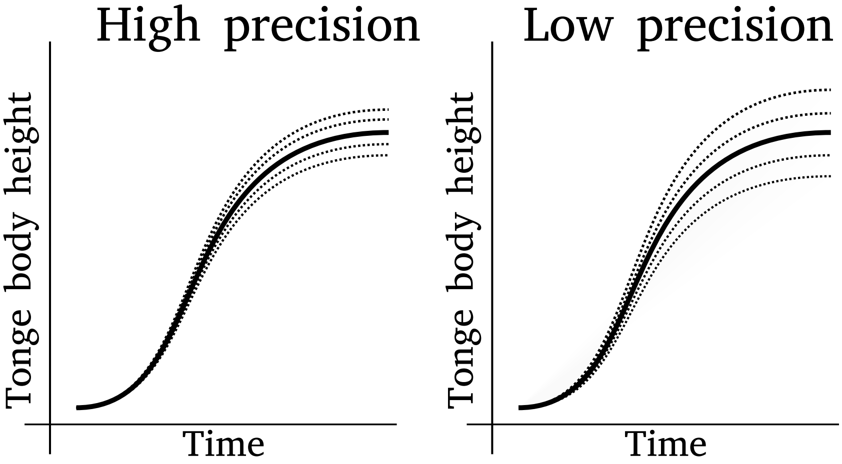

Figure 1: *Illustration of high/low precision during articulation of the vowel [iː] in the entire word [ziː] . Dark line represents a hypothetical optimal articulatory trajectory, dashed lines represent actual articulations. Low positions represent the position of the tongue body during the articulation of [z] , high positions represent the tongue body during the articulation of [iː] . Left: High precision, i.e. low variability around an optimal articulatory trajectory. Right: Low precision, i.e. high variability around an optimal articulatory trajectory.*

Heisler et al. (2010), who investigated how articulatory precision relates to lexical representations by examining the articulations of novel words, and found that articulatory patterns were more precise when a stronger lexical representation for a word had been established by means of a visual referent.

These findings support our hypothesis that practice with a word increases the precision with which an identical gesture will be executed during speech production.

All these findings all indicate that practice – whether as a result of repetition or as a result of age – serves to increase articulatory skills. However, none of these couple practice and articulatory precision in something as massively overlearned as adult speech, whose psychometric performance has been shown to be significantly predictable by frequencies of occurrence and conditional probabilities (Whalen, 1991; Aylett and Turk, 2004; Gahl, 2008; Bell et al., 2009; Tremblay and Tucker, 2011; Tomaschek et al., 2013, 2014; Ramscar et al., 2014; Priva, 2015; Tomaschek et al., under revision 2).

2. Articulation experiment

To examine the effects of practice operationalized by frequencies of occurrence and conditional probabilities, we investigated the precision of gestural execution during speech production recorded by means of electromagnetic articulography. In a similar fashion to kinematic studies, in which usually one movement pattern is investigated (e.g. Viviani and Terzuolo, 1982; Bourgeois and Hay, 2003), we controlled the shape of tongue movement trajectories by focusing on the articulation of a single German word – *sie* [ziː] 'engl. they'. In this way we were able to examine multiple instances of an identical articulatory gesture executed in differently practiced contexts. Given the foregoing, we expected increased articulatory precision, gauged by the size of the error around a hypothetically optimal trajectory, in more frequent ‘[ziː] + verb’ phrases, when the verb is more frequent, and also when [ziː] is more probable given the verb (see Figure 1, left vs. right). Importantly, the latter measure was favored because inverse conditional probability is an explicit measure of expectancy, i.e. practice of [ziː] given the following verb, whereas independently calculated measures such as the verb’s frequency or the phrase’s bigram frequency alone are not (Shannon, 1948; Baayen et al., 2013), and because empirically, conditional probability have been shown to account for behavior better than less informative measures (Arnon and Snider, 2010; Tremblay and Baayen, 2010).

Our method allowed for data on two additional linguistic measures to be collected: anticipatory coarticulation and speaking rate. The first measure is a consequence of the fact that [ziː] was produced in an unstressed position of a large number of ’sie + verb’ phrases, where the verb contained either a high [iː] or a low [a] . Since unstressed syllables are subjected to anticipatory coarticulation in the upcoming syllable (Öhman, 1966; Recasens, 1984; Sziga, 1992; Hoole et al., 1993; Magen, 1997; Fowler and Brancazio, 2000; Tomaschek et al., 2013, 2014) we expected the tongue to change movement patterns depending on whether a high vowel or a low vowel followed. Models of speech production (Bell-Berti and Harris, 1979; Browman and Goldstein, 1986; Saltzman and Munhall, 1989; Fowler and Saltzman, 1993) predict that the tongue will have only one vocalic target in the former case, but two in the latter. As two vocalic targets increase the complexity of the sequence – because two gestures need to be coordinated – articulatory variability ought to be larger in [ziː] + [a] verb phrases than in [ziː] + [iː] verb phrases.

The second linguistic measure afforded by our procedure was the manipulation of speaking rate: (*fast* vs. *slow*). Given Fitt’s Law (Fitts, 1954), we expected to find larger variability in the fast speaking rate condition than in the slow speaking condition, and additionally, given the natural temporal variations we could expect to encounter during speech production, we expected to find higher variability in the production of longer words and longer vowels.

```
## Error in setwd(DIR): cannot change working directory
```

2.1. Stimuli

The focal word examined in our investigation was the German pronoun *sie* [ziː] ’they’. (A series of complementary investigations of the production of the same words were reported in Tomaschek et al. (2014) – and reanalyzed in Tomaschek et al. (under revision 2) – studied the effect of frequency of use on the articulation of vowels in German verbs presented in the context of *sie* ’they’ or *ihr* ’you, pl.’.). In the present study, every participant articulated the word *sie* roughly 378 times during the experiment. Verbs were controlled for relative frequency of use, i.e. token count divided by corpus size, taken from the SDEWAC corpus (∼900 million entries, Shaoul and Tomaschek, 2013; Faaß and Eckart, 2013). Verbs contained the vowels [ɪ, iː, aː, a], respectively, in the first stressed syllable. For the present analysis, only verbs with a single non-dorsal onset consonant were considered (in total 63 [i] and 64 [a] verbs). 55 verbs began with a coronal consonant, 55 with a labial.

2.2. Speakers and Recording

A total of 21 German native speakers were recorded and paid 15€for their participation. All speakers signed an informed consent form agrreeing to their participation. Speaker identity was anonymized by means of experiment related identifiers. 4 speakers had to be excluded from the analysis due to missing data or faulty sensors. Of the remaining speakers in the analysis, 9 were females and 8 were males, with their mean age being 25.6 (sd= 3). All recordings were conducted in a sound-attenuated booth at the Department of Linguistics of the University of Tübingen. Speakers were instructed to read the stimuli ([ziː] + verb) aloud after they appeared on a computer screen. Each word in each context was presented once. The list was pseudo-randomized for each participant and divided into three parts. Each part 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, 6 blocks of stimuli were presented. Participants' tongue movements during articulation were recorded with an NDI wave articulograph at a sampling frequency of 100 Hz. Simultaneously, an audio signal of the sounds produced by these articulation was recorded (Sampling rate: 22.05 kHz, 16bit) and synchronized with the recordings from the articulograph. To correct for head movements 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 recording was made to determine the rotation from the local reference to a standardized coordinate system. The standardized coordinate system was defined by a bite plate to which three sensors in a triangular configuration were attached. 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 2cm). The analyses presented below focus on the sensor located near the tongue dorsum (the sensor located farthest back), which is considered least affected by non-dorsal consonants during the articulation of the vowel (Fowler, 1980; Öhman, 1966).

2.3. Preprocessing

Tongue movements were corrected for head-movements in an online procedure during recording. The recorded positions of the tongue sensors were centered at the midpoint of the bite plate and rotated in such a way that 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 by means of a Hidden-Markov-Model-based forced aligner for German (Rapp, 1995). Alignments were manually verified and corrected where necessary. Complete documentation and data are available at the Open Science Framework (https://osf.io/r78mk/, and will be downloadable upon publication).

Table 1: *Summary of generalized additive mixed-effects model investigating effects of Sie Duration and Vowel Proportion on vertical tongue dorsum movement. Effects in the bottom half represent random effects. Legend: SD = Sie Duration , VP = Vowel Proportion , SR = Speaking Rate Condition , NV = Next Vowel , SieProb= Sie Probabilities*

Effect	edf	F-value	p-value
te(Time,SD,VP):SRslow	50.81	13.84	<0.0001
te(Time,SD,VP):SRfast	71.49	11.11	<0.0001
te(Time,SD,VP):NVa	34.45	10.44	<0.0001
te(Time,SD,VP):NVi	31.42	6.92	<0.0001
s(Time,Participant)	138.31	63.27	<0.0001
s(SieProb,Participant)	121.72	4.38	<0.0001
s(Time,NextWord)	575.52	1.38	<0.0001
s(Time,Block)	15.40	13.09	<0.0001

3. Analysis

3.1. Assessing an optimal articulatory trajectory

In order to examine articulatory variability, we first needed to find an optimal articulatory trajectory for [ziː] against which any variation could be gauged. In obtaining this optimal trajectory, we controlled for known effects of temporal variation (i.e. word duration, vowel duration, Lindblom (1963); Gay (1978); Hertrich and Ackermann (2000)), for known effects of contextual variation due to neighboring consonants (Öhman, 1966; Sziga, 1992; Hoole et al., 1993) and form known effects of speaker variation Winkler et al. (2006); Fuchs et al. (2008); Brunner et al. (2009); Weirich and Fuchs (2006); Rudy and Yunusova (2013) in the random effects structure of our data. The use of Generalized Additive Mixed-effects Models (GAMM, package *mgcv*, Version 1.8-22 Wood (2006)) allowed us to do exactly this. GAMMs model non-linear functional relations between a response and one covariate – wiggly curves – by means of smooths and non-linear functional relations between a response and two or more covariates – (hyper)surfaces – by means of tensors. See Wieling et al. (2016) for a detailled description of GAMMs and their application to articulatory data.

The GAMM model was used to estimate an optimal articulatory trajectory, given the possible contextual, temporal and individual variation. From this the error around that optimal trajectory, i.e. the model's residuals, was then extracted to serve as a measure of variability around the optimal trajectory, with the residuals also being submitted to further statistical analysis (see Section 3.2.1 and Figure 1 for an illustration of variability around an optimal trajectory). (Although earlier studies of the variability in articulation simply report standard deviations in their analyses (Goffman et al., 2008; Heisler et al., 2010; Cho, 2001), approaches similar to that employed here have been reported in more recent studies: Sonderegger (2015), Sonderegger et al. (2017) and Chodroff and Wilson (2017) investigating the variation of VOTs by means of the residuals of a linear mixed-effects model.)

In exploratory pilot analyses, we first built our models using bottom-up and top-down fitting procedures and used the difference in maximum likelihood scores and edfs to test whether more complex models improved the model fit (van Rij et al., 2015). Improvement was further validated by means of contour plots. The following list presents the set of predictors we used in our GAMM analysis for predicting vertical tongue body position across time. All numeric variables were $\log_{10}$ transformed in order to obtain normally distributed data and centered and scaled in order to reduce possible correlations between random effects.

- Normalized time: ***Time*** In order to be able to match the articulatory trajectories in different [ziː] durations, durations were normalized into range between 0 and 1.
- Speaking rate condition: ***SpeakingRateCondition*** As described above, speech rates were binary categorized as either fast or slow.

  The vowel in the verb following [ziː]: ***NextVowels*** The vowel category in the following verb was classified as either [a] or [iː] . This predictor differentiates between those conditions in which the tongue body maintained its articulatory target (i.e. articulated a sequence of two [iː]) and those conditions in which the tongue body changed its articulatory target (i.e. articulated a sequence of [iː] and [a]).
- Acoustic durations of [ziː]: ***SieDurations*** In spite of the two speaking rate conditions, visual inspections showed that there was a high overlap in acoustic durations between the two conditions.
- Proportion of [iː] in [ziː] : ***VowelProportions*** Usually, another significant predictor of tongue position is vowel duration (Gay, 1978; Lindblom, 1963). However, vowel duration is strongly correlated with word duration and including it as a predictor would introduce collinearity into the model, increases the risk of Type 1 and Type 2 errors Tomaschek and Hendrix (under revision); Chatterjee and Hadi (2012, c.f.). We found that vowel proportions, with proportions calculated by dividing the vowel's duration by the [ziː] duration, did not introduce strong collinearity.
- We used Google counts of the verb form (*VerbFrequency*), Google bigram counts of the "sie + verb" phrase (*BigramFrequency*), and the conditional probability of "sie" given the following verb (*SieProbabilities*), calculated by dividing Google bigram counts by Google verb form counts. We will call these *lexical measures* in the remainder of the text to refere to them collectively.

Articulatory data constitute a time series with strong autocorrelation which results in strong autocorrelated residuals (i.e. one can predict the value at T+1 given the value at T). Residual autocorrelation results in anti-conservative p-values. We controlled autocorrelation in the residuals by an AR(1) parameter in our model (Wieling et al., 2015, 2016). We further controlled for the following random effects using *random factor smooths* (i.e. random smooths for factors) that can be regarded as non-linear equivalents of a combination between random intercepts and random slopes in mixed-effects regression:

1. Random factor smooths as a function of *Time* per speaker. This random factor smooth controls for speaker dependent variations during the production of [ziː] .

2. It is very well known that tongue positions during the entire vowel articulation are affected by near and far neighboring consonants (Öhman, 1966; Recasens, 1984; Sziga, 1992; Hoole et al., 1993; Magen, 1997; Fowler and Brancazio, 2000; Tomaschek et al., 2013, 2014). In order to control for this effect, we included random factor smooths as a function of *Time* per *verb*. Consequently, this random factor smooth controls for anticipatory coarticulation, which is not the focus of the current study.

3. Random factor smooths as a function of *Time* per *presentation block*. This random factor smooth was included in order to account for fatigue effects during the experiment.

4. Random factor smooths as a function of *Sie Probabilities* per *speaker*. Although our probabilistic measures were not significant in pilot analyses (see below for explanation), we included this random smooths in order to control for possible by participant effects as a function of *Sie Probabilities* .

GAMMs model how the *intercept* (in the parametric summary), i.e. the mean tongue body position, is changed by the fixed effects. We found that the tongue body position was significantly predicted by two four-way interactions (Table 1).

- a) between *Time*, *Sie Duration* , *Vowel Proportion* and *Speaking Rate Condition* ;
- b) between *Time*, *Sie Duration* , *Vowel Proportion* and *Next Vowel* .

Interactions of numeric predictors were modelled by means of a tensor product, including both, main effects as well as interactions (*te()*, similar to a $A \times B \times C$ formula in linear models). The tensor's interaction with the factorial predictors *Speaking Rate Condition* and *Next Vowel* is indicated by the colon. Edfs represent estimated degrees of freedom, indicating how many smoothing functions on average were needed to fit the data. P-values lower than 0.01 indicate significant non-linearity of the model term. No interaction between *Speaking Rate Condition* and *Next Vowel* as well as no effect of *Sie Probabilities* was found. The final model had $R^2 = 0.85$. The highly complex model was supported by means of ML-score comparison with a less complex model lacking the interactions.

Importantly, none of the lexical measures emerged as a significant predictor of articulatory trajectory in these preliminary models. One major reason is probably that [ziː] was highly repetitive in our experiment. Even though its conditional probability given the following word varied, the conditional probability of [ziː] in the experiment itself very soon asymptoted to roughly 0.5, as [ziː] occurred in roughly 50% of the trials at the beginning of the utterance. The estimated tongue body positions across time and their complex interaction with *Sie Duration* , *Vowel Proportion* , *Speaking Rate Condition* , *Next Vowel* can be inspected in the supplementary material. For illustrational purposes, we report here the estimated vertical tongue body positions across time in interaction with *Vowel Proportion* and *Next Vowel* for [ziː] instances of 220 ms. We chose *Vowel Proportion* , as this predictor turned out to be predictive for articulatory variability.

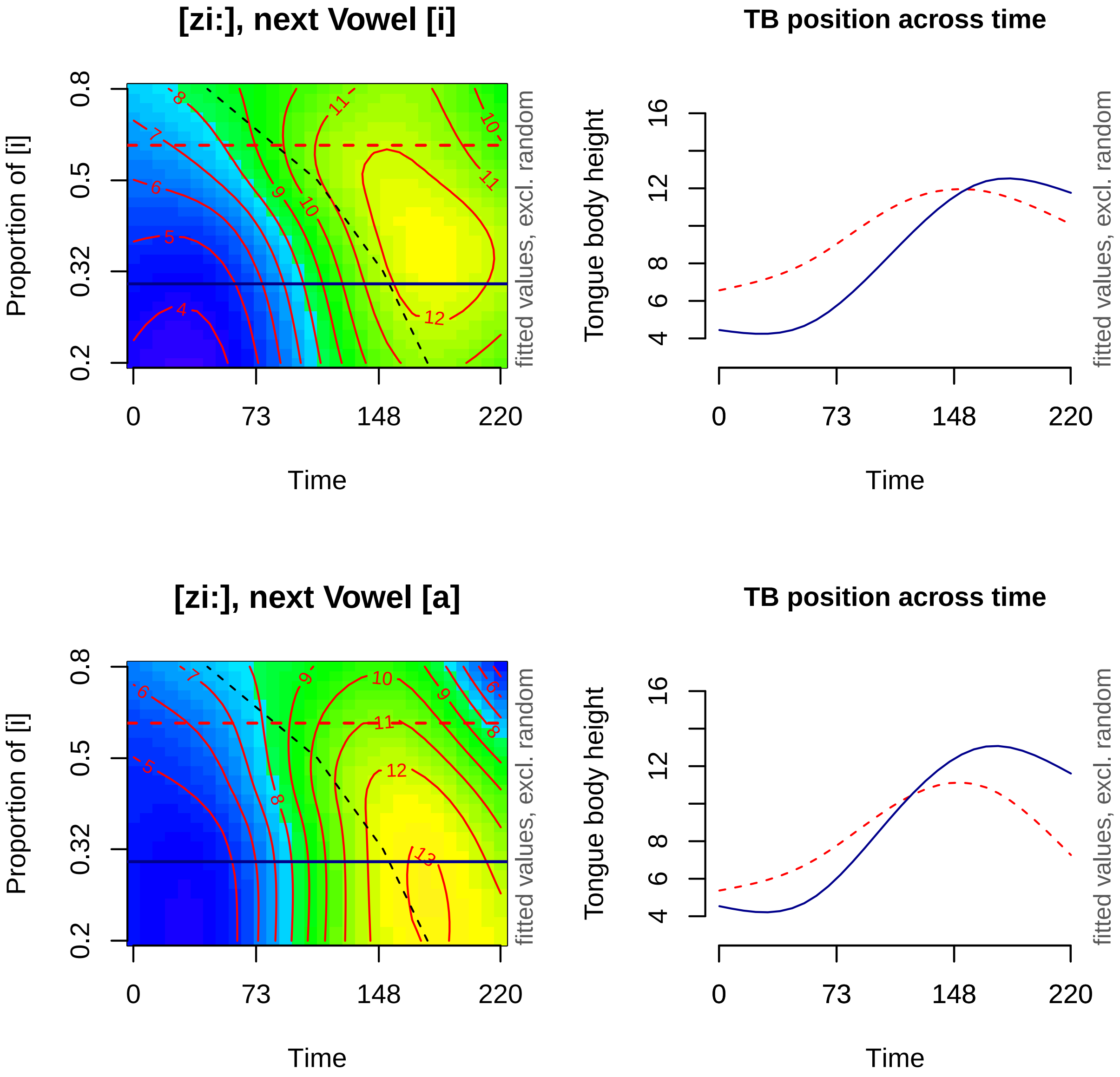

Figure 2: *Predictions for vertical tongue dorsum movement as a function of time based on the generalized additive model in Table 1. Random effects were excluded from the predictions. Rows represent different (WD). Columns represent Next Vowel / Speaking Rate Condition conditions. X-axis in plots represents normalized time, Y-axis represents the proportion of [iː]in [ziː]. Contour lines connect areas of the same tongue dorsum height. 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). The diagonal dashed line indicates the acoustic boundary between [z] and [iː]. All relevant values were back transformed to raw values for visualization.*

Before presenting the analysis of variability observed in this study, we decribe how articulatory trajectories for [ziː] varied, focusing in particular on the effects of *Sie Duration* , *Vowel*

Proportion and *Next Vowel* .
The top panel of Figure 2 illustrates the tongue body movements in those instances of [ziː] that are followed by verbs containing an [iː] . The x-axis of the contour plot (left column) represents time, the y-axis represents the proportion of the vowel in the entire [ziː] syllable (i.e. including the articulation of [z]). Areas with low tongue body positions are encoded in blue, areas with high tongue body positions are encoded in yellow, and intermediate positions are encoded in green; the contour lines represent positional distances of 1 mm. The quasi-diagonal black dashed line represents the acoustic boundary between [z] and [iː]. With increasing *Vowel Proportion* , it is shifted towards the onset of [ziː] . Examplarily, two slices from the *Vowel Proportion* continuum are illustrated in the right column, one for -1.5 and one for 1.5 standard deviations ($\sim$ 0.3, blue line / $\sim$ 0.6, red dashed line) away from the mean.

In the plots, higher tongue body positions at the onset of [ziː] are associated with higher *Vowel Proportion* (compare the blue and red lines in the right column). This difference is maintained until the last quarter of [ziː] . Furthermore, in words with high *Vowel Proportion* the highest deflection is reached earlier than in words with low *Vowel Proportion* . It appears that tongue body trajectories are smoother in words with longer acoustic vowel durations than in words with shorter acoustic vowel durations (i.e. high vs. low *Vowel Proportion*). In the latter, due to the shorter vocalic time window, the tongue body shows effects of an overshoot, probably due to higher movement velocities during rising (as can be seen in the steeper rising trajectory in the blue line). Consequently, there are higher tongue positions at [ziː] offset in words with lower *Vowel Proportion* than in words with higher *Vowel Proportion* , in spite of a lowering of the tongue body at the offset.

The movement pattern observed in instances of [ziː] that are followed by verbs containing an [a] looks similar (Figure 2, bottom panel), but differs in three respects as a result of anticipatory coarticulation. First, the difference in the first quarter of [ziː] between words with high *Vowel Proportion* and words with low *Vowel Proportion* is smaller. This is mainly because the articulatory trajectory in words with high *Vowel Proportion* is lower when [ziː] precedes words with [a] as compared to verbs with [iː] . Second, as a consequence, tongue height differs at the maximum deflection point. Surprisingly, it is higher in words with low *Vowel Proportion* when [ziː] precedes verbs with [a] than when it precedes words with [iː] . Third, anticipating the lower [a] , the tongue body is lowered in strong contrast to when [ziː] precedes verbs containing [iː] . However, this is only the case in instances of [ziː] with high *Vowel Proportion* , i.e. when the tongue has enough time during the acoustic production of the vowel the execute the movement. Overall, we also observe a high degree of systematic, complex structure during the production of [ziː] , as a result of both temporal variations in [ziː] and variation across the word boundary. It is these temporal changes that we can expect to affect articulatory variability at the transition between [ziː] and the following verb.

3.2. Analysis of articulatory variance

3.2.1. Preprocessing and modeling strategy

We used the GAMM model described above as a basis for the next set of analyses. Since we included random factor smooths as a function of time per speaker and per item as well as as a function of *Sie Probabilities* per speaker, we can expect to have covered a maximum of the variance in the data, a suggestion further supported by the model selection criterion. In line with Cho (2001), who analyzed errors around means, and Sonderegger (2015), Sonderegger et al. (2017) and Chodroff and Wilson (2017), who analyzed residuals from linear mixed-effect regression models, we used the residuals from the GAMM model, i.e. the error between the predicted optimal articulatory trajectory of [ziː] and the recorded articulatory trajectory in the following analysis of articulatory variance. In particular, because we were interested in absolute variation, we used the absolute residuals, which were log-transformed in order to obtain normally distributed data. In these analyses, high values of residuals will be interpreted as high variability, as they indicate that the articulated trajectory strongly deviates from an optimal trajectory (see Figure 1 for an illustration of variability around an optimal trajectory). In the remainder of the text, *smaller residuals* and *lower variability* will be used as synonyms, depending on the context and focus of the section.

Since we lack a precisely defined target for the articulatory movement of the tongue body, we focus instead on the tongue body's movement during anticipatory coarticulation in the last 10% of [ziː] , i.e. the transition between [ziː] and the following word. We fitted absolute logged residuals with linear mixed-effect regression (package *lme4*, Version 1.1-12, Bates et al. (2014)). We removed potential outliers by removing datapoints larger than 2.5 standard deviations away from the mean (N = 129, 2.5% of the data). During modeling, we used *Speaking Rate Condition* , *Sie Duration* , tongue body's vertical *movement velocity*, *Vowel Proportion* , *Next Vowel* and *next length* contrasting between short lax and long tense vowels as fixed effects. We included *Speakers* and *NextWord* as random intercepts.

VerbFrequency and *BigramFrequency* show high collinearity (r = 0.85), which is why they could not be included simultaneously into one model (Hadi, 1988; Tomaschek and Hendrix, under revision). Instead, we created additional versions of the model described above which included these predictors on an individual basis.

3.2.2. Results

Pilot modeling showed signficant interactions between *Next Vowel* and *Vowel Proportion* (β = 0.12, se = β = 0.03, t = β = 4.21) as well as between *Next Vowel* and *Sie Probabilities* (β = -0.08, se = β = 0.04, t = β = -2.15). In contrast to our hypothesis, no main effect of the following vowel onto variability was attested (β = 0.01, se = β = 0.04, t = β = 0.24).

To reduce the complexity of each model and highlight the effects of lexical measures, we analyzed each vowel with an individual model, fitting absolute residuals as a function of *Speaking Rate Condition* , *Vowel Proportion* , *Sie Duration* and *Sie Probabilities* . *Movement velocity* and *next length* failed to yield any significant effect in any of the models and were excluded from the analysis. It is however possible that *movement velocity* towards a target plays a more important role rather than away from the target, which occurred roughly before the offset of the vowel (cf. Figure 2).

Table 2: *Summary of partial effects in the linear mixed-effects models investigating effects on articulatory residuals in the transition between [ziː] and the following verb, containing [iː] (upper half) or [a] (lower half). Estimates for predictors marked by * are taken from a different model. See Section 3.1 for details on predictors.*

Partial effect	Next vowel	Estimate	Std.Error	t-value
(Intercept)	[i]	-0.045	0.097	-0.467
SpeakingRateConditionfast	[i]	0.147	0.042	3.489
VowelProportions	[i]	0.034	0.022	1.517
SieDurations	[i]	0.008	0.025	0.303
SieProbability	[i]	-0.005	0.029	-0.156
*VerbFrequency	[i]	-0.045	0.025	-1.811
*BigramFrequency	[i]	-0.043	0.024	-1.811
(Intercept)	[a]	-0.029	0.082	-0.350
SpeakingRateConditionfast	[a]	0.125	0.043	2.937
VowelProportions	[a]	0.161	0.022	7.481
SieDurations	[a]	0.055	0.025	2.176
SieProbability	[a]	-0.091	0.031	-2.970
*VerbFrequency	[a]	-0.001	0.033	-0.027
*BigramFrequency	[a]	-0.001	0.031	-0.027

According to Fitt's law, variability is proportional to movement speed and movement duration (Fitts, 1954; Wright and Meyer, 1983; Bertucco and Cesari, 2010). We observe an effect consistent with this in the present study: Absolute residuals were significantly larger in the fast speaking rate condition than in the slow speaking rate condition preceding verbs with both vowels (Table 2, cf. *Speaking Rate Condition*). No further interactions with speaking rate condition were found in these analyses.

When [ziː] preceded verbs with [iː] , all predictors apart from *Speaking Rate Condition* failed to yield significance (Table 2, upper half). If [iː] in [ziː] and [iː] in the following verb would have been represented by two rather than one vocalic target, executed by two gestures, one would expect that during the articulation of [ziː] temporal variation increases the variability during the interpolation from one to the second gesture. Since this is not the case, this null results indicate that [iː] in both words is represented by one gesture. Given that *next length* was not supported as a significant predictor, it is reasonable to assume that the distance in tongue height between the short lax and the long tense instance of [iː] was too small to account for a change in target location.

The temporal predictors *Speaking Rate Condition* , *Sie Duration* and *Vowel Proportion* were correlated positively with the absolute residuals only when [ziː] preceded verbs with [a] , (Table 2, lower half). Again, the result is perfectly consistent with Fitt's law: It appears that longer macroscopic and microscopic articulation times increase the probability that the tongue will deviate from an optimal trajectory, resulting in larger variability. Further, this indicates that articulatory movement variability is not affected by the global tongue body movement, but rather that it is only influenced by the duration of the relevant movement necessary to produce the acoustic signal. In other words, when the tongue body has more time to execute a movement, movements that are on the whole rather similar tend to vary more markedly.

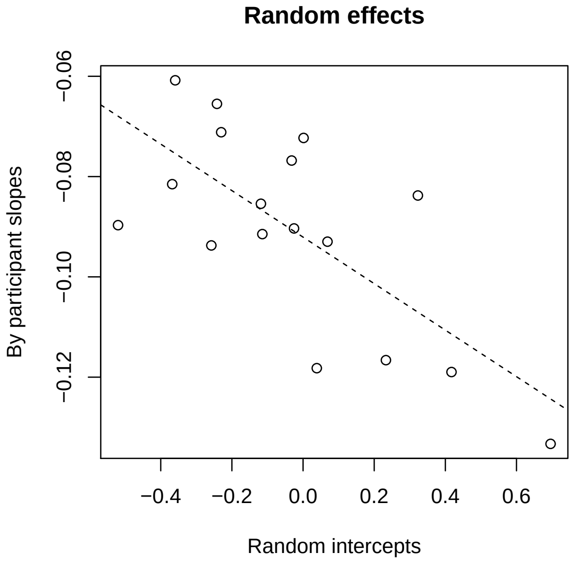

Figure 3: *By participant slopes for Sie Probabilities plotted against by participant random intercepts.*

One of the hallmarks of increasing kinematic skills due to practice is an increase in precision during a movement (Segalowitz and Segalowitz, 1993; Darling et al., 1988; Georgopoulos et al., 1981; Platz et al., 1998; Madison et al., 2013). We examined three ways of gauging practice: of these, *VerbFrequency* and *BigramFrequency* did not emerge as significant predictors for absolute residuals (Table 2, c.f. effects marked by stars).

However, and importantly, the inclusion of *Sie Probabilities* results in a better model fit, as indicated by a χ^2-test ($\chi^2_{(9,10)}$ = 7.81, p = 0.01). This indicates that the more probable [ziː] given the following word, i.e. the easier it is to plan it, the better the tongue follows a hypothetical, optimal spatial pathway. This was the case across all speakers, as can be seen in Figure 3, which illustrates the by-participant slopes (y-axis) plotted against by-participant random intercepts (x-axis). All speakers show negative adjustments of random slopes, i.e. reduction of variability across *Sie Probabilities* . Speakers with a greater articulatory variability benefited more from higher *Sie Probabilities* (R = -0.71, t_{15} = -3.88, p = 0.001).

This result is in line with findings that frequencies of occurrence are less informative for behavior than conditional probabilities (Shannon, 1948; Arnon and Snider, 2010; Tremblay and Baayen, 2010; Baayen et al., 2013). This is perhaps surprising. However, unlike the other frequency measures, the conditional probability measure captures the variance in amount of practice that speakers can be expected to have of [ziː] in the context of a given verb (i.e., it represents the amount of practice associated with any given instance of articulatory planing). Given the experimental set up, this measure was thus always likely to be more informative than the simple counts.

4. Summary and conclusion

Kinematic studies of hand movements repeatedly indicate that practice, i.e. the repetition of the same movement, increases precision associated with later performance of that movement (Darling et al., 1988; Georgopoulos et al., 1981; Platz et al., 1998; Bourgeois and Hay, 2003; Madison et al., 2013; Raeder et al., 2015), and in speech, a similar pattern of effects has been observed in articulation as speaker age increases (Goffman et al., 2008; Zharkova et al., 2011, 2012; Belmont, 2011). Here, we sought to see whether practice operationalized in terms of the probability of one identical gesture given another also reflected the effects of practice on the precision in an identical gesture executed in different environments (cf. Figure 1).

The effect of practice was dependent on the stem vowel in the verbs. When [ziː] preceded verbs with [iː] , no effects of practice were found; rather, these effects only emerged when [ziː] preceded [a] . Given that only labial and coronal intervocalic consonants were present in the data set at hand, a possible explanation for this difference is that the two acoustic instances of [iː] were represented by one vocalic target, articulated by a single tongue body gesture (Browman and Goldstein, 1986; Saltzman and Munhall, 1989). This would suggest that no temporal and articulatory coordination, as well as no interpolation between the two [iː] vowels was necessary in this case, and that speakers' uncertainty about the articulatory gesture in the verb's stem vowel was minimal during their productions of [ziː] . If this is correct, then it follows that no local effects of practice could arise. By contrast, producing the acoustic sequence [iː] and [a] involves two consecutive gestures, with the resulting trajectory being an interpolation between the two targets. In this case, it follows that the more practiced that trajectory is, the more precisely the tongue body will be able to follow a hypothetically optimal trajectory.

Within frameworks in which higher conditional probabilities and frequencies of occurrence are associated with faster lexical access (Bell et al., 2009; Gahl, 2008), the current findings could be explained by the additional time speech production gains from the faster access. With a longer preparatory window articulation can be planned in more detail, resulting in a less variable trajectory. From the perspective of the task-dynamic model (Saltzman and Munhall, 1989), which regards the executed articulatory trajectories to be interpolations between gestures and targets, a less variable trajectory at the word boundary probably indicates a smoother transition between two consecutive gestures, and possibly a stage, in which the two gestures are merged into a single one. Similar findings have also been presented in kinematic studies (Sosnik et al., 2004).

Previous studies have shown that frequencies of occurrence and conditional probabilities are associated with strong changes of spatio-temporal coordination of articulatory movements and their resulting phonetic signal (Aylett and Turk, 2004; Bell et al., 2009; Tremblay and Tucker, 2011; Priva, 2015; Tomaschek et al., under revision 1, 2013, 2014, under revision 2). Taken together with the current results, these findings indicate that practice, as gauged by lexical measures, serves to change fine articulatory details in speech production suggesting that the fine grained kinematic skills employed by speakers during articulation constantly improve with practice. Consequently, like other psychological behavior, speech production is submitted to an ongoing fine tuning and mirrors the dynamics of life-long learning (Ramscar et al., 2014).

5. Acknowledgements

We would like to thank 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 'sie'. This paper was funded in part by the Alexander von Humboldt Chair awarded to R. H. Baayen (grant 1141527) and in part by a collaborative grant from the Deutsche Forschungsgemeinschaft (BA 3080/3-1).

References

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.

Baayen, R. H., Hendrix, P., Ramscar, M., 2013. Sidestepping the Combinatorial Explosion: An Explanatio of n-gram Frequency Effects Based on Naive Discriminative Learning. Language and Speech 56 (3), 329–347.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4.
URL `http://CRAN.R-project.org/package=lme4`

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.

Bertucco, M., Cesari, P., 2010. Does movement planning follow fitts' law? scaling anticipatory postural adjustments with movement speed and accuracy. Neuroscience 171 (1), 205 – 213.

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., May 1986. Towards an articulatory phonology. Phonology 3, 219–252.

Brunner, J., Fuchs, S., Perrier, P., 2009. On the relationship between palate shape and articulatory behavior. The Journal of the Acoustical Society of America 125 (6), 3936–3949.

Chatterjee, S., Hadi, A. S., 2012. Regression Analysis by Example. Fifth Edition. Wiley.

Cho, T., 2001. Effects of morpheme boundaries on intergestural timing: Evidence from korean. Phonetica 58, 129–162.

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.
URL `http://www.sciencedirect.com/science/article/pii/S0095447017300013`

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.

Faaß, G., Eckart, K., 2013. Sdewac - a corpus of parsable sentences from the web. In: Gurevych, I., Biemann, C., Zesch, T. (Eds.), Language Processing and Knowledge in the Web. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 61 – 68.

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., 1980. Coarticulation and theories of extrinsic timing. Journal of Phonetics (8), 113–133.

Fowler, C., 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., Saltzman, E., 1993. Coordination and coarticulation in speech production. Language and Speech 36 (2-3), 171–195.
URL `http://dx.doi.org/10.1177/002383099303600304`

Fuchs, S., Winkler, R., Perrier, P., 2008. Do speakers' vocal tract geometries shape their articulatory vowel space? pp. 333–336.

Gahl, S., 2008. "thyme" and"time" are not homophones. word durations in spontaneous speech. Language 84 (3), 474–496.

Gay, T., 1978. Effect of speaking rate on vowel formant movements. The Journal of the Acoustical Society of America 63 (1), 223–230.

Georgopoulos, A., Kalaska, J., Massey, J., 1981. Spatial trajectories and reaction times of aimed movements: Effects of practice, uncerntainty, and change in target location. Journal of Neurophysiology 46 (4), 725ff.

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.

Hadi, A. S., 1988. Diagnosing collinearity-influential observations. Computational Statistics Data Analysis 7 (2), 143 – 159.
URL `http://www.sciencedirect.com/science/article/pii/0167947388900898`

Heisler, L., Goffman, L., Younger, B., 2010. Lexical and articulatory interactions in children's language production. Developmental science 13 (5), 722–730.

Hertrich, I., Ackermann, H., 2000. Lip–jaw and tongue–jaw coordination during rate-controlled syllable repetitions. The Journal of the Acoustical Society of America 107 (4), 2236–2247.
URL `http://dx.doi.org/10.1121/1.428504`

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.

Lindblom, B., 1963. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35, 1773–1781.

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.

Öhman, S., 1966. Coarticulation in vcv utterances: Spectrographic measurements. Journal of the Acoustical Society of America 39 (151), 151–168.

Platz, T., Brown, R., Marsden, C., 1998. Training improves the speed of aimed movements in parkinson's disease. Brain 121, 505–513.

Priva, U. C., 2015. Informativity affects consonant duration and deletion rates. Laboratory Phonology 6 (2), 243–278.

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. J Strength Cond Res. 29 (7), 1904–14.

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.

Rapp, S., 1995. Automatic phonemic transcription and linguistic annotation from known text with hidden markov models / an aligner for german. In: Proceedings of ELSNET goes east and IMACS Workshop. Moscow.

Recasens, D., 1984. Voweltovowel coarticulation in catalan vcv sequences. The Journal of the Acoustical Society of America 76 (6), 1624–1635.
URL `http://scitation.aip.org/content/asa/journal/jasa/76/6/10.1121/1.391609`

Rudy, K., Yunusova, Y., 2013. The effect of anatomic factors on tongue position variability during consonants. Journal of Speech Language and Hearing Research 56, 137–149.

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

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 (03), 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.
URL https://fabiantomaschek.files.wordpress.com/2016/07/tomaschek$_c orpus_r eadme.pdf$

Sonderegger, M., 2015. Trajectories of voice inset time in spontaneous speech on reality tv. In: Proceedings of the 18th ICPHS. Glasgow.

Sonderegger, M., Bane, M., Graff, P., 2017. The medium-term dynamics of accents on reality television.

Language January.

Sosnik, R., Hauptmann, B., Karni, A., Flash, T., 2004. When practice leads to co-articulation: the evolution of geometrically defined movement primitives. Exp Brain Res 156, 422–438.

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 ICPHS XVII. Hong Kong.

Tomaschek, F., Arnold, D., Bröker, F., R. Baayen, R. H., under revision 1. Lexical frequency co-determines the speed-curvature relation in articulation. Journal of Phonetics.

Tomaschek, F., Hendrix, P., under revision. Coping with collinearity in linguistic regression analysis. Journal of Phonetics.

Tomaschek, F., Tucker, B. V., Baayen, R. H., under revision 2. The consequences of lexical proficiency for articulation. Linguistic Vanguard.

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 the 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: Wood, D. (Ed.), Perspectives on Formulaic Language: Acquisition and communication. The Continuum International Publishing Group, London, UK.

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.

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., 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., 1991. Infrequent words are longer in duration than frequent words. The Journal of the Acoustical Society of America 90 (4), 2311–2311.
URL `http://scitation.aip.org/content/asa/journal/jasa/90/4/10.1121/1.401072`

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Baayen, R. H., 2015. Investigating dialectal differences using articulography. In: Proceedings of the 18th ICPHS. Glasgow.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., Baayen, R. H., 2016. Investigating dialectal differences using articulography. Journal of Phonetics.

Winkler, R., Fuchs, S., Perrier, P., 2006. The relation between differences in vocal tract geometry and articulatory control strategies in the production of french vowels: Evidence from mri and modeling. pp. 509–516.

Wood, S. N., 2006. Generalized additive models: an introduction with R. Chapman and Hall/CRC, Boca Raton, Florida, U. S. A.

Wright, C. E., Meyer, 1983. Conditions for a linear speed-accuracy trade-off in aimed movements. The Quarterly Journal of Experimental Psychology Section A 35 (2), 279–296.

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.