Kinematic proficiency

Version: October 9, 2017

Practice makes perfect: The consequences of lexical proficiency for articulation

October 9, 2017

Abstract

Many studies report that frequent words have shorter acoustic durations, more co-articulation and reduced articulatory targets. This study calls attention to a factor ignored in discussions on the relation between lexical frequency and the phonetic detail, namely, that motor skills improve with experience. Since frequency of use is a measure of experience, it follows that frequent words should show increased articulatory proficiency. We used electromagnetic articulography to test this prediction against German inflected verbs with [a] stem vowel, focusing on vertical tongue movements. Medium-frequency words with a [t]-exponent revealed U-shaped trajectories that were more shallow and had a higher deflection, reflecting more co-articulation and reduced articulatory targets. Words with the [n]-exponent showed movements that varied with tongue region. Tongue tip sensors revealed U-shaped curves positioned higher in the mouth for frequent words, without being more shallow, however. Sensors further back on the tongue showed deeper and more long-lasting downwards trajectories infrequent words, in combination with stronger co-articulation. These results challenge the hypothesis that higher frequency of use necessarily comes with more co-articulation *and* more articulatory reduction. We argue that the observed patterns are best understood as arising from the opposing pressures of the communicative forces of predictability and discriminability.

Index Terms: electromagnetic articulography, frequency of use, quantile regression, generalized additive models, co-articulation

1 Introduction

How words are realized in speech varies substantially. A survey of English spontaneous conversations (Johnson 2004) indicates that some 5% of words are spoken with one syllable less than the citation form, and that roughly over 20% of words miss at least one phone. Several factors have been identified that co-determine the details of words' forms. One such factor is audience design. Speakers may articulate words more carefully when speaking to a large audience or under noisy conditions, but in conversations with familiar interlocutors, they may hypo-articulate (Lindblom 1990). On the cline from hyper-articulation to hypo-articulation, words tend to be realized with shorter durations, vowels are more centralized, articulatory gestures overlap to a greater extent, and segments as well as full syllables can be dropped (Moon and Lindblom 1989; Lindblom 1990; Junqua 1993; Browman and Goldstein 1986; Browman and Goldstein 1989; Liberman and Mattingly 1985).

A second factor influencing word form is frequency of use. High frequency words typically have fewer segments (Zipf 1935; Zipf 1949), but they also have shorter acoustic durations when other factors such as number of segments and syllables are controlled for (Bell et al. 2009; Gahl 2008). Frequency of occurrence can be understood as a de-contextualized measure of lexical probability. Probability measures conditioned on a word's context, such as the probability of the following word (Bell et al. 2009) or syntactic probabilities (Gahl and Garnsey 2004; Tily et al. 2009) have been found to explain additional variance in word durations over and above frequency. Several functional explanations have been forwarded for the negative correlation of frequency and length. For example, Zipf (1949) pointed out that longer words require more articulatory effort (see also Lebedev, Tsui, and

Van Gelder 2001, for hand movements), and that a general biological constraint to reduce the costs of speaking will over time drive frequent words to become shorter.

According to the smooth signal redundancy hypothesis (Aylett and Turk 2004), language production is affected by a preference to distribute information uniformly across the linguistic signal. As high probability meanings are less informative, the complexity of the acoustic signal encoding these meanings has to be reduced in order to maintain a uniform flow of information per time unit. From the perspective of audience design, speakers have been argued to articulate rare words more carefully in order to ensure intelligibility of words that listeners would otherwise find hard to understand (Galati and Brennan 2010).

A third factor that has been put forward which effects word duration is lexical retrieval. According to Bell et al. (2009), less frequent words are realized with longer durations as a consequence of having to maintain synchrony between higher level planning and articulation. Concretely, for rare words, the phonological word becomes available later in time; for frequent of words, it is available earlier. Varying the duration of higher vs. rare words then contributes to a smoother flow of speech.

The terminology in use for describing shorter variants — articulatory undershoot, hypo-articulation, reduction — reflects the normative status accorded to the citation form found in dictionaries and represented in alphabetic writing systems. This seemingly negative evaluation does not do justice to the rich communicative values of the shorter forms (see Hawkins 2003, for discussion). Furthermore, even though especially highly reduced forms are often unintelligible in isolation, in the proper context they are fully functional (Arnold et al. 2017; Ernestus, Baayen, and Schreuder 2002).

The goal of the present study is to call attention to a fourth factor that co-determines how words are articulated, namely, the increase in skilled execution of articulatory gestures with experience. The three factors discussed above paint a picture of shorter forms as forms that are impoverished, that are less well discriminated from other forms, that convey less information, and that are more redundant. In what follows, we show, on the basis of results obtained with electromagnetic articulography for German inflected verbs, that high-frequency forms can maintain optimal articulatory targets in combination with strong co-articulation. Before introducing our experiment, we first provide an introduction to some relevant results in related domain of inquiry, in hand movements. We conclude with a discussion of our findings and their implications.

2 Kinematic proficiency in hand movements

For a fixed proficiency level, consider the time required for a movement, t, the distance the movement needs to cover, d, and the width of the targeted endpoint w. A greater width allows for a greater variety of endpoint positions, and hence is a measure of the desired movement accuracy. Experimentally, movement precision is typically gauged by the magnitude of the error between the executed trajectory and the optimal trajectory as well as by the magnitude of the error between the endpoint of the movement and its optimal target. According to Fitts' law (Fitts 1954; Wright and Meyer 1983; Bertucco and Cesari 2010),

$$t = a + b \log(2d/w). \tag{1}$$

Although the linearity assumption underlying (1) is a simplification (Langolf, Chaffin, and Foulke 1976), the linear form clarifies that decreasing movement time t for a fixed distance d goes hand in hand with an increase in variability w. Also, movements towards a target that are executed at a speed that exceeds the current level of proficiency will be less accurate. When the level of proficiency increases, trajectories become less variable for fixed t and d (Darling, Cole, and Abbs 1988; Georgopoulos, Kalaska, and Massey 1981; Platz, Brown, and Marsden 1998; Madison et al. 2013). Furthermore, for fixed w and d, movement time t decreases with proficiency (Raeder, Fernandez-Fernandez, and Ferrauti 2015; Platz, Brown, and Marsden 1998). Furthermore, practice is associated with smoother transitions between two successive gestures, and upcoming gestures are anticipated earlier. As a result, the overal time required for executing a movement is reduced (Sosnik et al. 2004).

To make this more concrete, consider someone learning to play the violin. With increasing practice, the violin player will be able to perform a more demanding showpiece with a speed that better approximates the fast tempo envisioned by the composer. At the same time, as learning proceeds, the quality of execution becomes higher, and less variable. Movements, such as those required for the bow, becomes more economical with a reduction in the distance travelled by the hand. The crucial point here is that at a given stage of learning, greater speed, exceeding the optimal speed for the given level of proficiency, and hence shorter durations, necessarily go hand in hand with reduced, more variable, and less accurate movements. However, with practice, greater speed can be realized while maintaining accuracy and reducing variability.

3 Kinematic proficiency in articulation

Like playing the violin, articulation is a complex motor skill that takes years of practice to master. A word's frequency of use is an index of the amount of training a speaker has received for properly coordinating the movements of tongue, jaw, lips, lungs, velum, and larynx during articulation. Given the above simple principles of kinematics, we can expect the articulatory record to show that with increasing practice, i.e., with more frequent use, articulatory gestures of words become less variable, more complex articulatory gestures can be executed without requiring slower execution, and that upcoming gestures will be anticipated earlier, without lowering standards for articulatory targets.

To avoid misunderstanding, we do not claim that how words are articulated is determined only by articulatory proficiency. As discussed above, audience design, probability, and lexical retrieval are forces that co-determine articulation and may exert an influence on articulation opposite to that predicted from increasing articulatory proficiency. Of course, this raises the question of whether the consequences of learning for articulation are at all detectable at the level of individual words.

Several previous studies have addressed this question. Using electromagnetic articulography, Tiede et al. (2011) were able to show that repetition of novel sequences of common syllables leads to a reduction of the distances travelled by the articulators as well as to increased gestural overlap, resulting in overall shorter words. Goffman et al. (2008) compared the speech of children with the speech of adults and observed reduced temporal variation during anticipatory co-articulation for adults. There is some evidence that over the lifetime, as experience accumulates, the vowel space expands (Baayen, Tomaschek, et al. 2017; Gahl and Baayen 2017), allowing improved

discrimination of an increasingly complex vocabulary (Keuleers et al. 2015; Ramscar et al. 2014). These findings suggest that learning may indeed play a role in articulation.

The next section presents an experiment we carried out with electromagnetic articulography (henceforth EMA) that was designed to clarify the consequences of experience for the articulation of inflected words. For this, we reanalyze the data from (Tomaschek, Tucker, et al. 2014). Given the literature summarized above,

we investigated how kinematic practice, parameterized by a word's frequency of occurrence, shapes on the one hand the target of articulation and on the other hand balances against anticipatory coarticulation of inflectional exponents (Öhman 1966; Magen 1997). Data and scripts for the analyses can be downloaded from `https://osf.io/snuqd/`.

4 Methods

4.1 Participants

Seventeen participants (9 female, 8 male; mean age: 26, sd: 3) took part in the experiment. They were undergraduate students at the University of Tübingen, all native speakers of German, with no known language impairments. They were either paid 10 Euro for their participation, or received course credit.

4.2 Stimuli

Participants were asked to read out loud inflected forms of twenty-seven German verbs. All verbs were presented in a context (*sie* ...) requiring a

realization that in its canonical form is disyllabic (e.g., *sie zahlen* [ziːtsaːlən]). Nine verbs were also presented in a context eliciting a monosyllabic form (*ihr zahlt* [iːɐtsːalt]). The verbs were selected to cover a wide range of frequencies of occurrence, extracted from the SDEWAC corpus (Faaß and Eckart 2013; Shaoul and Tomaschek 2013).

Log-transformed frequency of occurrence was not a significant predictor of the acoustic durations of the word stimuli ($\beta = 0.0068, s.e. = 0.0035, t = 1.954, p = 0.0508$), and was also not predictive for the acoustic duration of the stem vowel ($\beta = 0.0006, s.e. = 0.0006, t = 0.982, p = 0.326$) in mixed models with random intercepts for subject and word.

4.3 Recording

Recordings took place in a sound proof booth at the Department of Linguistics of the University of Tübingen. Participants were instructed to read out loud stimuli presented to them on a computer screen. Each participant produced two tokens for each verb, one monosyllabic and one disyllabic. The order of the stimuli was randomized for each individual participant, and divided into three parts. Each part was presented first in a slow (inter-stimulus interval: 600 ms; presentation-time: 800 ms) and then in a fast speaking condition (inter-stimulus interval: 300 ms; presentation-time: 450 ms). There were small self-timed breaks between the blocks of 1 to 2 minutes. During these breaks, we made sure that sensors were still properly attached, and reattached them if required. In what follows, we discuss only the fast speaking rate.

Articulatory movements of the tongue were recorded with the NDI wave articulograph at a sampling frequency of 100 Hz. Simultaneously, the audio

signal was recorded (Sampling rate: 22.05 kHz, 16bit) and synchronized with the articulatory recordings. To correct for head movements and to define a local coordinate system, a reference sensor was attached to the subject's forehead. Before the tongue sensors were attached, 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 (TT), one at the tongue middle (TM) and one at the tongue body (TB; distance between each sensor: around 2cm).

4.4 Preprocessing

The recorded positions of the tongue sensors were centered at the midpoint of the bite plate and rotated in such a way that the front-back direction of the tongue was aligned to the x-axis, with more positive values towards the front of the mouth, and more positive z-values towards the top of the oral cavity. To determine segment boundaries, the audio signal was automatically aligned with phonetic transcriptions by means of a Hidden-Markov-Model-based forced aligner for German (Rapp 1995). Vowel alignments were manually verified and corrected where necessary. The analyses of the movement trajectories of the three tongue sensors were restricted to the period of time during which the stem vowel of the verbs was articulated.

4.5 Statistical analysis

We used quantile GAMs as implemented in the R package **qgam**, available at `https://github.com/mfasiolo/qgam`, to investigate how the positions of the tongue sensors changed over time, and how these articulatory trajec-

tories were modified by predictors such as frequency of use and inflectional exponent. Quantile GAMs (Fasiolo et al. 2017) integrate quantile regression (Koenker 2005) with the generalized additive model (GAM, Hastie and Tibshirani 1990; Wood 2006; Wood 2011; Wood 2013b; Wood 2013a).

GAMs provide spline-based smoothing functions for modeling nonlinear functional relations between a response and one or more covariates, thereby enabling the analyst to model wiggly curves as well as wiggly (hyper)surfaces. Wiggly curves were fitted with thin plate regression splines, and interactions of covariates with time were modeled with tensor product smooths (see Baayen, Vasishth, et al. 2017, for an introduction to spline smooths). Quantile GAMs (henceforth QGAMs) implement a distribution-free method for estimating the predicted values of a given quantile of the response distribution, together with confidence intervals. In our analyses, we investigated the median, but other quantiles can also be of theoretical interest (see, e.g., Schmidtke, Matsuki, and Kuperman 2017). The **qgam** package builds on the **mgcv** package (version 1.8-5) for R (Version 3.0.2, (R Core Team 2014)). We used the `itsadug` package (J. van Rij et al. 2015) (Version 2.2) for visualization.

The choice for modeling articulatory trajectories with quantile GAMs was motivated by the strong autocorrelations present in the residuals of the Gaussian GAMs that we initially fitted to our data. Timeseries of slowly changing tongue positions are characterized by strong correlations between the position at time t and that at $t-1$. Although the **mgcv** package makes it possible to include an AR(1) autoregressive model for the residuals, we were not able to fit a model to the data with residuals that were properly Gaussian and identically and independently distributed. Since qGAMs are distribution-free, they are a natural and powerful alternative for the analysis

of articulatory trajectories as registered with electromagnetic articulography.

5 Analysis

Speakers sometimes reduced the [ə] of the inflectional exponent, resulting in a form such as [ziː tsaːlən] being realized as [ziː tsaːln]. We therefore created a factor (using treatment dummy coding), inflectional EXPONENT, with three levels: stem+[t], stem+[n], and stem+[ən]. For inclusion in the group with the [ən] exponent, the duration of the [ə] had too exceed 50 ms. The reference level of EXPONENT was [ən]. We use the notation `exponent`(j) for the level of EXPONENT that is instantiated for word j.

Stem vowel duration was normalized between 0 and 1. In what follows, we refer to this normalized duration as TIME (abbreviated to t in model specifications and model summaries).

Vowels' articulatory trajectories are influenced by the contexts in which these vowels occur. As a consequence, for each verb (abstracting away from its inflectional exponents), the consonants flanking the vowel are expected to have their own specific effect on how the vowel is articulated. We therefore included by-verb factor smooths for TIME in our models. These factor smooths are the nonlinear equivalent of the combination of by-verb random intercepts and by-verb random slopes for TIME in the linear mixed model (see Baayen, Vasishth, et al. 2017, for detailed discussion). By including these factor smooths, we stack the cards against the hypothesis that words' frequency of occurrence also co-determines the articulatory trajectories. In our qGAMs, an effect of frequency has to establish itself over and above the co-articulatory effects of the vowels' context. Since the effect of the inflectional exponents on articulation is probed with the factor EXPONENT, the combi-

nation of the by-verb factor smooths and EXPONENT bring under statistical control all parts of the word forms that potentially co-determine articulation.

As average tongue height was expected to differ between participants, we also included by-participant random intercepts (b_i) in the model specification.

Given a vector of covariates $\boldsymbol{x}$, a qGAM minimizes the loss function

$$\mathrm{E}[\rho_\tau(y-\eta)|\boldsymbol{x}],$$

where ρ_τ is the pinball loss for quantile $\tau \in c(0,1)$. In this study, we consider only the median ($\tau = 0.5$). The analyses reported below assume that the linear predictor η for the vertical position of a sensor for speaker i and word j with exponent `exponent`(j) at time t can be approximated by

$$\eta_{i,j,t} = \beta_0 + b_i + \mathrm{fs}(t,j) + \alpha_{\texttt{exponent}(j)} + \mathrm{te}(t, \texttt{frequency}_j, \texttt{exponent}(j)), b_i \sim \mathcal{N}(0,\sigma).$$

No sensor data were available for the tongue tip sensor for 464 measurement points (data loss 6.5%). A total of 870 data points was lost for the tongue mid sensor (12.1%), the loss for the tongue body sensor was 370 measurement points (5.2%). Separate qGAMs were fitted to the remaining data points for each sensor. Table 1 presents the model summaries, Figure 1 presents the by-word factor smooths for time, and Figure 2 visualizes the partial effects of the smooths for the time by frequency by exponent interaction.

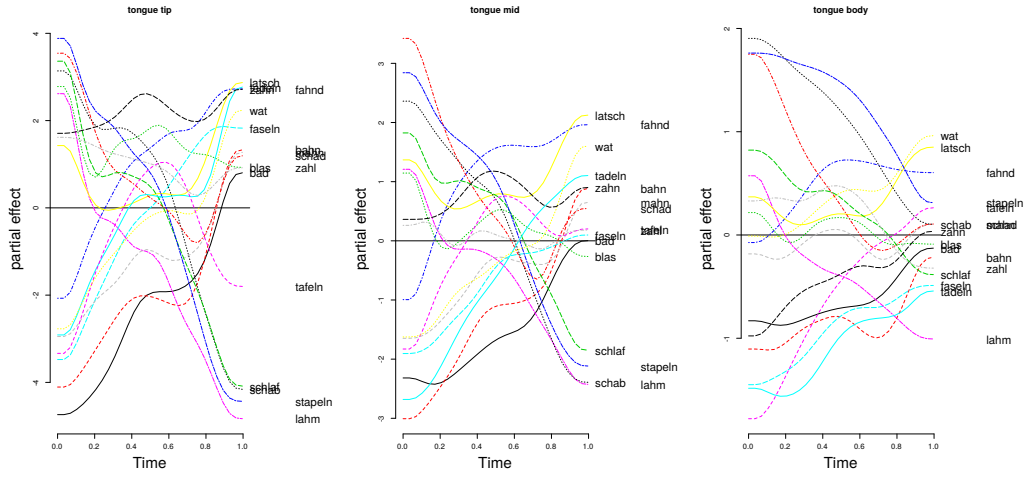

Figure 1: Partial effect of the by-word factor smooths for time, for tongue tip (left), tongue mid (center) and tongue body sensors (right). Each curve represents a word. Across sensor positions, the same words tend to show roughly the same trends, reflecting similar co-articulatory constraints with the consonants flanking the [a].

tongue tip sensor				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-6.5766	0.9925	-6.6263	< 0.0001
Exponent=t	0.9553	0.0828	11.5331	< 0.0001
Exponent=n	0.5544	0.0855	6.4863	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time, Frequency):Exponent=en	7.5679	7.8737	459.6735	< 0.0001
te(Time, Frequency):Exponent=t	7.5973	7.8898	434.1386	< 0.0001
te(Time, Frequency):Exponent=n	6.1565	6.7754	381.5733	< 0.0001
random intercepts Participant	15.9837	16	11709.6192	< 0.0001
factor smooth (Time,Word)	115.2519	152	5882.5003	< 0.0001
tongue mid sensor				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-3.3046	1.1901	-2.7768	0.0055
Exponent=t	0.5192	0.0666	7.7925	< 0.0001
Exponent=n	0.2714	0.0717	3.7849	0.0002
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time,Frequency):Exponent=en	4.9950	5.0048	533.0317	< 0.0001
te(Time,Frequency):Exponent=t	7.4413	7.8513	423.4764	< 0.0001
te(Time,Frequency):Exponent=n	7.3608	7.8227	490.8539	< 0.0001
random intercepts Participant	14.9942	15.0000	37470.4267	< 0.0001
factor smooth (Time,Word)	101.7264	152.0000	4557.1898	< 0.0001
tongue body sensor				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-0.3452	1.2854	-0.2686	0.7883
Exponent=t	0.3512	0.0666	5.2732	< 0.0001
Exponent=n	0.4276	0.0693	6.1665	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time, Frequency):Exponent=en	6.1863	6.7421	339.5993	< 0.0001
te(Time, Frequency):Exponent=t	7.5659	7.8606	273.5099	< 0.0001
te(Time, Frequency):Exponent=n	6.8855	7.4905	326.5897	< 0.0001
random intercepts Participant	15.9948	16	45916.4092	< 0.0001
factor smooth (Time, Word)	67.1645	152	6465.9622	< 0.0001

Table 1: QGAMs for the vertical position of the tongue tip, tongue mid, and tongue body sensor. te: tensor product smooth. The standard deviations for the by-participant random intercepts are for the tongue tip sensor: 3.72 (95% confidence interval (2.63,5.26); for the tongue mid sensor: 4.65 (95% confidence interval (3.25,6.65); and for the tongue body sensor 5.25, 95% confidence interval (3.71,7.42).

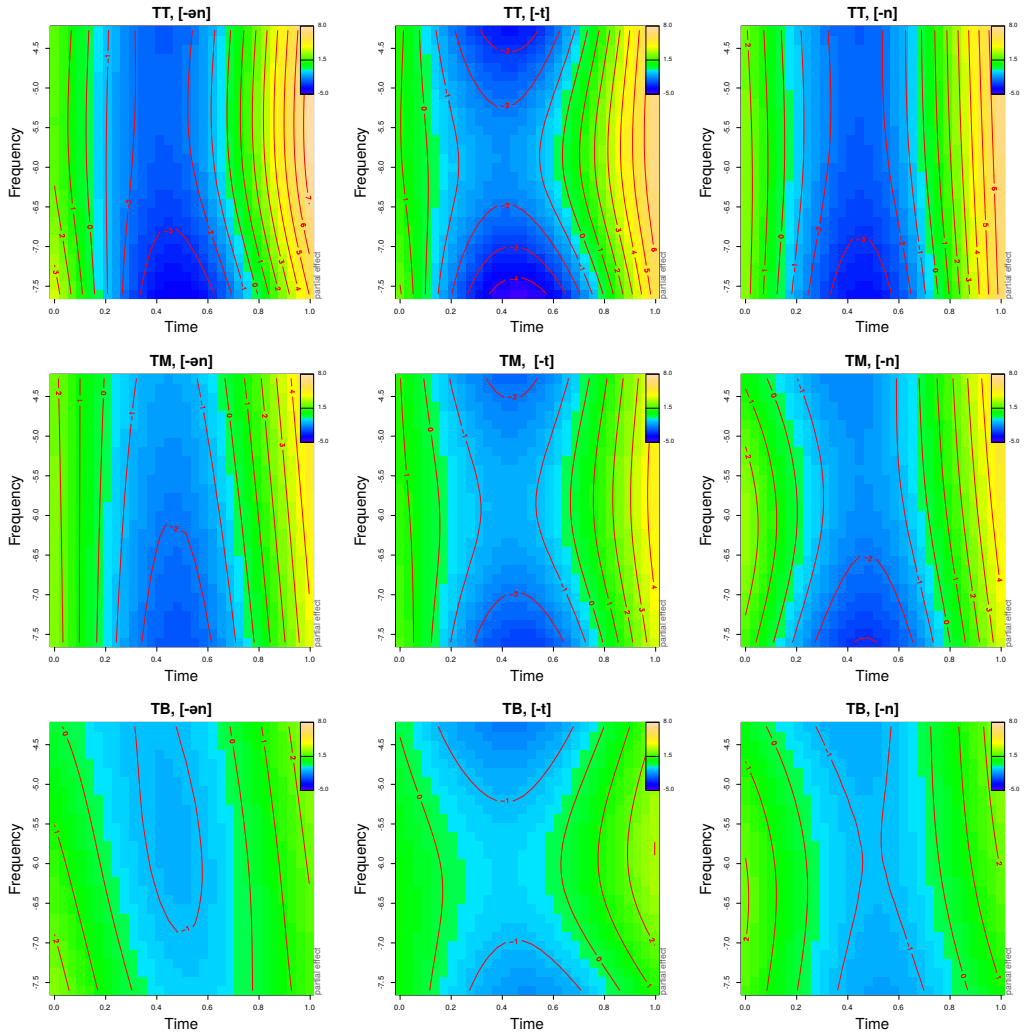

Figure 2: Partial effect of the interaction of TIME by FREQUENCY on the vertical position of the tongue tip sensor for the three exponents [-ən], [-t], and [-n]. Top panels: tongue tip sensor, center panels: tongue mid sensor, bottom panels: tongue body sensor. Deeper shades of blue indicate lower vertical positions, and darker shades of yellow indicate higher vertical positions.

We first consider the random-effects in this model. The by-word factor smooths (Figure 1) were included in order to statistically control for the consequences for the vowel's articulation of the preceeding and following consonants. Each curve represents one word (lemma). The curves for *stapeln* and *bad* in the left panel illustrate the very different consequences for articulatory trajectories of the place of articulation of the pre-vocalic consonants. For many words, the curves are roughly similar across sensor positions. For *stapeln*, we always find a downward trend, and for *bad*, an upward trend. Note that for the tongue tip and tongue mid sensors, variability is greatest at the edges of the time domain. For the tongue body sensor, by contrast, variability is reduced most towards the end of the vowel. As will become clear below, the tongue body sensor profiles itself differently from the tongue tip and tongue mid sensors.

The standard deviations for the by-participant random intercepts increased from tongue tip sensor: 3.72 (95% confidence interval (2.63,5.26) to tongue mid sensor: 4.65 (95% confidence interval (3.25,6.65), and further for the tongue body sensor 5.25, 95% confidence interval (3.71,7.42). Although the wide confidence intervals advise caution, the increase in variability from front to back suggests that sensors placed closer to the part of the tongue that is critically involved in articulating the [a] are more disturbing to the speaker, and lead to more perturbations of the articulatory trajectories. In addition, especially the tongue back sensor is the most difficult to attach, and here between-speaker differences in exact location are most likely.

Next, consider the main effect of EXPONENT (cf. Table 1). As expected, the intercept, representing (other things being equal) the group mean for [-ən], was lowest for the tongue tip sensor, intermediate for the tongue mid sensor, and largest for the tongue body sensor. These relative positions reflect

that the tongue tip tends to be close to the lower teeth when articulating the [a], and hence far below the tongue body sensor.

For tongue tip and tongue mid sensors, both the [-t] and [-n] exponents lead to higher sensor positions compared to [-ən]. The [-t] is articulated even further away from the [-ən] position than is the case for [-n], which may reflect both the total absence of a schwa or remnants thereof preceeding the [-t] exponent, the propensity of many speakers of German to articulate the [-n] with the tongue blade rather than the tongue tip, and anticipatory velar lowering for the [-n]. A difference in height between [-t] and [-n] is not present for the tongue body sensor.

Although the qGAM identified specific articulatory trajectories for each combination of lemma and exponent, as shown in Figure 1, an effect of word frequency received solid support as well. The rows of Figure 1 represent the partial effects for tongue tip (top), tongue mid (center) and tongue body sensor (bottom), and its columns the effects for [-ən] (left), [-t] (center) and [-n] (right). Across all panels, the blue areas represent the expected lower position of the sensors reached that are reached roughly midway through the vowel.

Note that colors are less bright as one moves down in the graph from tongue tip to tongue body sensor. This highlights that sensors further into the mouth show more reduced modulations of frequency. This pattern is consistent with the differences in intercepts, in that the greatest modulations over time are present for the tongue tip, the sensor with the lowest intercept. Thanks to its position in the front of the mouth, and furthest out on the jaw, the consequences of opening and closing the mouth during [aː] production are most visible here.

The U-shaped movement of the sensors is modified by frequency. If fre-

quency would have had no effect at all, all contour lines would be straight vertical lines. Model comparisons (not shown) pitting the present models against models without frequency provide strong support for the relevance of frequency as predictor.

The upper left panel shows that the [-ən] is articulated with the lowest tongue tip position for the lowest frequency words. Higher-frequency words not only have a lower minimum, but also, after having reached this minimum, they reach higher positions more quickly. In other words, higher frequency words show more co-articulation with the upcoming exponent, without dampening of the U-shaped articulatory trajectory. This patterning of the higher-frequency words is present across all sensors for both [-ən] and [-n].

The [-t] exponent shows an hour-glass pattern, with the lowest positions being reached for both the lowest and the highest-frequency words. This same pattern is also visible for the tongue body sensor for the [-n].

The tongue body sensor stands apart with an effect of frequency for the [-ən] exponent (lower left panel) that is the reverse of that observed for the other two sensors: the tongue body sensor reaches its lowest position for higher-frequency words. And whereas the deepest positions for the [-t] exponent are reached for the lowest frequency words for tongue tip and tongue mid sensors, at the tongue body sensor, the deepest position is more widespread for the highest frequency words. Thus, for the sensor best monitoring how the tongue movement, rather than jaw movement, shapes the [a], higher-frequency words show mastery of deeper articulation of [a] without sacrificing the benefits of a smooth co-articulation. Note that by the end of the time window, the tongue has reached a higher position ([-ən], [-n]) or a position nearly as high ([-t]) as that reached by words with intermediate or lower

frequencies.

6 Discussion

We used electromagnetic articulography to test whether articulation is also subject to the law that practice makes perfect. We investigated inflected German verbs with [a] as stem vowel with exponents [-t], [-ən], and [-n], focusing on the vertical trajectories of three sensors placed on the tongue. Articulatory trajectories were modulated by frequency of use. For the [-ən] exponent, the passive tongue positions (tongue tip and tongue mid) revealed U-shaped curves of similar shape that were shifted upwards for higher-frequency words. Here, co-articulation with the upcoming exponent resulted in an overall higher tongue position, while maintaining the amplitude of the U-shaped curve. For the tongue body sensor, which was closer to the part of the tongue involved in primary articulation of [a], higher-frequency words showed deeper and more long-lasting downwards curvature, in combination with earlier and stronger co-articulation.

Results for the [-n] exponent were similar to those for [-ən] for the tongue tip and tongue mid sensors. For the tongue body sensor, an hour-glass pattern was visible that also characterized the partial effects for the [-t] exponent. Here, words with intermediate frequency of use revealed U-shaped trajectories that were more shallow and at the same time had a higher minimum, reflecting more co-articulation in combination with a reduced articulatory target. In other words, the effect that the literature leads one to expect to be present for high frequency words — a muted articulatory trajectory with more co-articulation — is visible in part of our data, but for words of intermediate frequency. Higher-frequency words either show an overall higher

positioned but otherwise unchanged U-shaped trajectory, or they show strong co-articulation and at the same time a steeper U-shaped trajectory.

The hour-glass pattern that characterizes especially words with the [-t] exponent, such that the lowest positions are reached for words with extreme frequencies, is open to different interpretations. On the one hand, lemmas that are highly probable type-wise, i.e., lemmas that have common (log) frequencies close to the mean (log) frequency, may be articulated with less effort. Words that are (type-wise) unexpected in the experiment would then be articulated with more extreme vertical movements. On the other hand, if production of low-frequency words is dominated by assembly from phone-like units whereas high-frequency words are realized from larger planning units (see, e.g., Hickok 2014), then the medium frequency words are the ones for which not enough experience has accumulated to enable low target positions to be reached under the constraints of co-articulation. From the perspective of Blevins, Milin, and Ramscar (2015), the forms of low and high-frequency words may arise under the opposing pressures of the communicative forces of predictability and discriminability. Predictability enforces a clear articulatory target, with a strong U-shaped trajectory with a low minimum. Discriminability is served by co-articulation (see Kemps, Wurm, et al. 2005; Kemps, Ernestus, et al. 2005, for the discriminative role of durations of the stems of inflected words). Surprisingly, with the accumulation of experience, speakers apparently are able to optimize both constraints simultaneously: For [-ən], they do so by moving the tongue body down earlier, keeping it down longer, and then moving it higher.

The importance of frequency of use, as a measure of articulatory proficiency, for understanding articulatory gestures also emerged in several other studies. Tomaschek, Arnold, Bröker, et al. (under revision) showed that fre-

quency also modulates the speed-curvature relation. More frequent words, i.e., words that have received more practice, were articulated with reduced speed, reduced curvature, overall smoother movement trajectories, and less articulatory overshoot (see also Sosnik et al. 2004; Tiede et al. 2011). Tomaschek, Arnold, R. van Rij, et al. (under revision) showed furthermore that articulatory movements at more probable word boundaries were produced with less variability (cf. Goffman et al. 2008).

All these results were obtained for laboratory speech, using a registration technique that requires placement of sensors on tongue and lips. As a consequence, it is unclear whether the present results generalize to spontaneous speech (see, e.g., Gahl, Yao, and Johnson 2012, for potential differences in neighborhood effects in lab speech as compared to spontaneous conversation). Replication studies, ideally based on corpora of spontaneous speech with EMA or ultrasound registration, are essential for consolidating the results observed in the present study.

If the present results are pointing in the right direction, they have two important theoretical implications. First, the observation that frequency of occurence modulates the fine detail of how articulatory gestures are realized challenges the common assumption that articulation is planned post-lexically. This assumption is implemented in cognitive models for speech production, such as proposed by Dell (1986), Levelt, Roelofs, and A. S. Meyer (1999), and Goldstein et al. (2009), which assume that the representations driving articulation are assembled out of phonemes and morphemes, or gestural scores associated with these units. These models cannot straightforwardly accommodate the finding that experience at the level of individual words co-determines how articulatory trajectories are realized (see also Gahl 2008).

Second, higher-frequency forms are not necessarily more 'reduced'. We

note that the decrease in acoustic duration reported for higher-frequency words (van Bergem 1995; Aylett and Turk 2006; Schulz et al. 2016; Meunier and Espesser 2011) is not incompatible with the results reported here, as frequency of occurrence happened not to be predictive for the acoustic durations of our stimuli. Furthermore, the present data are also not necessarily incompatible with the increase in vowel centralization reported for high-frequency as opposed to low-frequency words (Aylett and Turk 2004). For instance, the tongue tip sensor revealed that higher-frequency words with the [-ən] exponent realize a very similar articulatory trajectory as lower-frequency words, but shifted to a slightly higher position in the mouth. Although we presently do not know how exactly the articulatory positions of the different parts of the tongue shape the average position of vowels in formant space, this kind of pattern could result in more vowel centralization.

Third, our results suggests that as speakers gain experience with individual words, the articulatory trajectories of these words move through three stages: an initial stage with a strongly profiled U-shaped curve and little co-articulation, followed by a second stage at which co-articulation dominates at the expense of a shallower trajectory, followed by a third stage at which an optimal solution is reached that respects both the necessity of profiling and the need for smooth co-articulation. Experiments with many more stimuli, covering a much wider range of frequencies, will be required for testing this hypothesis.

References

[1] D. Arnold et al. “Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algo-

rithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit." In: *PLOS ONE* (2017).

[2] M. Aylett and A. Turk. "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei." In: *Journal of the Acoustical Society of America* 119.5 (2006), 3048ff.

[3] M. Aylett and A. Turk. "The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech." In: *Language and Speech* 47.1 (2004), pp. 31–56.

[4] R. H. Baayen, F. Tomaschek, et al. "The Ecclesiastes principle in language change." In: *The changing English language: Psycholinguistic perspectives*. Ed. by M. Hundt, S. Mollin, and S. Pfenninger. Cambridge, UK: Cambridge University Press, 2017.

[5] R. H. Baayen, S. Vasishth, et al. "The cave of Shadows. Addressing the human factor with generalized additive mixed models." In: *Journal of Memory and Language* 94 (2017), pp. 206–234.

[6] Alan Bell et al. "Predictability effects on durations of content and function words in conversational English." In: *Journal of Memory and Language* 60.1 (2009), pp. 92–111. ISSN: 0749-596X.

[7] M. Bertucco and P. Cesari. "Does movement planning follow Fitts' law? Scaling anticipatory postural adjustments with movement speed and accuracy." In: *Neuroscience* 171.1 (2010), pp. 205–213.

[8] J. P. Blevins, P Milin, and M. Ramscar. "The Zipfian Paradigm Cell Filling Problem." In: *Morphological paradigms and functions*. Ed. by F. Kiefer, J. P. Blevins, and H. Bartos. Leiden: Brill, 2015.

[9] C. Browman and L. Goldstein. “Articulatory gestures as phonological units.” In: *Phonology* 6 (1989), pp. 201–251.

[10] C. Browman and L. Goldstein. “Towards an articulatory phonology.” In: *Phonology* 3 (May 1986), pp. 219–252.

[11] W.G. Darling, K.J. Cole, and J.H. Abbs. “Kinematic variability of grasp movements as a function of practice and movement speed.” English. In: *Experimental Brain Research* 73.2 (1988), pp. 225–235. ISSN: 0014-4819.

[12] G.S. Dell. “A spreading-activation theory of retrieval in sentence production.” In: *Psychological review* 93.3 (1986), pp. 283–321.

[13] M. Ernestus, R. H. Baayen, and R. Schreuder. “The Recognition of Reduced Word Forms.” In: *Brain and Language* 81.1–3 (2002), pp. 162–173. ISSN: 0093-934X.

[14] G. Faaß and K. Eckart. “SdeWaC - A Corpus of Parsable Sentences from the Web.” In: *Language Processing and Knowledge in the Web*. Ed. by I. Gurevych, C. Biemann, and T. Zesch. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, pp. 61–68.

[15] M. Fasiolo et al. “Fast calibrated additive quantile regression.” Manuscript, University of Bristol, 2017. URL: `https://github.com/mfasiolo/qgam`.

[16] Paul M. Fitts. “The information capacity of the human motor system in controlling the amplitude of movement.” In: *Journal of Experimental Psychology* 47.6 (1954), p. 381.

[17] S. Gahl. “”Thyme” and”Time” are not homophones. Word durations in spontaneous speech.” In: *Language* 84.3 (2008), pp. 474–496.

[18] S. Gahl and R. H. Baayen. "Twenty-eight years of vowels." In: *Manuscript submitted for publication* (2017).

[19] S. Gahl and S.M. Garnsey. "Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation." In: *Language* 80(4) (2004), pp. 748–774.

[20] S. Gahl, Y. Yao, and K. Johnson. "Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech." In: *Journal of Memory and Language* 66 (2012), pp. 789–806.

[21] A. Galati and S. E. Brennan. "Attenuating information in spoken communication: For the speaker, or for the addressee?" In: *Journal of Memory and Language* 62.1 (2010), pp. 35–51.

[22] A. Georgopoulos, J. Kalaska, and J. Massey. "Spatial trajectories and reaction times of aimed movements: Effects of practice, uncerntainty, and change in target location." In: *Journal of Neurophysiology* 46.4 (1981), 725ff.

[23] L. Goffman et al. "The Breadth of Coarticulatory Units in Children and Adults." In: *Journal of Speech, Language, and Hearing Research* 51.6 (2008), pp. 1424–1437.

[24] L. Goldstein et al. "Coupled oscillator planning model of speech timing and syllable structure." In: *Frontiers in phonetics and speech science* (2009), pp. 239–250.

[25] T.J. Hastie and R.J. Tibshirani. *Generalized Additive Models*. London: Chapman & Hall, 1990.

[26] S. Hawkins. "Roles and representations of systematic fine phonetic detail in speech understanding." In: *Journal of Phonetics* 31 (2003), pp. 373–405.

[27] G. Hickok. "The architecture of speech production and the role of the phoneme in speech processing." In: *Language, Cognition and Neuroscience* 29.1 (2014), pp. 2–20.

[28] K. Johnson. "Massive reduction in conversational American English." In: *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium.* The National International Institute for Japanese Language. Tokyo, Japan, 2004, pp. 29–54.

[29] J. C. Junqua. "The Lombard reflex and its role on human listeners and automatic speech recognizers." In: *The Journal of the Acoustical Society of America* 93.1 (1993), pp. 510–524.

[30] R. J. Kemps, M. Ernestus, et al. "Prosodic cues for morphological complexity: The case of Dutch plural nouns." In: *Memory & Cognition* 33.3 (2005), pp. 430–446. ISSN: 1532-5946.

[31] R. J. Kemps, Lee H. Wurm, et al. "Prosodic cues for morphological complexity in Dutch and English." In: *Language and Cognitive Processes* 20.1/2 (2005), pp. 43–73.

[32] E. Keuleers et al. "Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment." In: *The Quarterly Journal of Experimental Psychology* 8 (2015), pp. 1665–1692.

[33] R. Koenker. *Quantile regression.* Cambridge University Press, 2005.

[34] G. D. Langolf, D. B. Chaffin, and J. A. Foulke. "An investigation of Fitts' law using a wide range of movement amplitudes." In: *Journal of Motor Behavior* 8.2 (1976), pp. 113–128.

[35] S. Lebedev, W. H. Tsui, and P. Van Gelder. "Drawing movements as an outcome of the principle of least action." In: *Journal of mathematical psychology* 45 (2001), pp. 43–52.

[36] W. J. Levelt, A. Roelofs, and A. S. Meyer. "A theory of lexical access in speech production." In: *The Behavioral and brain sciences* 22.1 (Feb. 1999).

[37] A. M. Liberman and I. G. Mattingly. "The motor theory of speech perception revised." In: *Cognition* 21 (1985), pp. 1–36.

[38] B. Lindblom. "Explaining Phonetic Variation: A Sketch of the HH Theory." English. In: *Speech Production and Speech Modelling*. Ed. by WilliamJ. Hardcastle and Alain Marchal. Vol. 55. Springer Netherlands, 1990, pp. 403–439. ISBN: 978-94-010-7414-8.

[39] G. Madison et al. "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." In: *Acta Psychologica* 143.1 (2013), pp. 119–128.

[40] H. S. Magen. "The extent of vowel-to-vowel coarticulation in English." In: *Journal of Phonetics* 25 (1997), pp. 187–205.

[41] C. Meunier and R. Espesser. "Vowel reduction in conversational speech in French: The role of lexical factors." In: *Journal of Phonetics* 39.3 (2011). Speech Reduction, pp. 271–278. ISSN: 0095-4470.

[42] S-J. Moon and B. Lindblom. *Formant undershoot in clear and citation-form speech: a second progress report. STL-QPSR, Department of Speech Communication 1.* 1989.

[43] S.E.G. Öhman. "Coarticulation in VCV Utterances: Spectrographic Measurements." In: *Journal of the Acoustical Society of America* 39.151 (1966), pp. 151–168.

[44] T. Platz, R.G. Brown, and C.D. Marsden. "Training improves the speed of aimed movements in Parkinson's disease." In: *Brain* 121 (1998), pp. 505–513.

[45] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2014. URL: `http://www.R-project.org`.

[46] C. Raeder, J. Fernandez-Fernandez, and A. Ferrauti. "Effects of Six Weeks of Medicine Ball Training on Throwing Velocity, Throwing Precision, and Isokinetic Strength of Shoulder Rotators in Female Handball Players." In: *J Strength Cond Res.* 29.7 (2015), pp. 1904–14.

[47] M. Ramscar et al. "The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning." In: *Topics in Cognitive Science* 6.1 (2014), pp. 5–42.

[48] S. Rapp. "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, 1995.

[49] D. Schmidtke, K. Matsuki, and V. Kuperman. "Surviving blind decomposition: a distributional analysis of the time course of complex word recognition." In: *Journal of Experimental Psychology: Learning, Memory and Cognition* (2017).

[50] E. Schulz et al. "Impact of prosodic structure and information density on vowel space size." In: 2016, pp. 350–354.

[51] C. Shaoul and F. Tomaschek. "A phonological database based on CELEX and N-gram frequencies from the SDEWAC corpus." 2013. URL: `https:`

//fabiantomaschek.files.wordpress.com/2016/07/tomaschek_corpus_readme.pdf.

[52] R. Sosnik et al. "When practice leads to co-articulation: the evolution of geometrically defined movement primitives." In: *Exp Brain Res* 156 (2004), pp. 422–438.

[53] M. Tiede et al. "Motor learning of articulator trajectories in production of novel utterances." In: *Proceedings of the ICPHS XVII*. Hong Kong, 2011.

[54] H. Tily et al. "Syntactic probabilities affect pronunciation variation in spontaneous speech." In: *Language and Cognition* 1 (2009), pp. 147–165.

[55] F. Tomaschek, D. Arnold, Franziska Bröker, et al. "Lexical frequency co-determines the speed-curvature relation in articulation." In: *Journal of Phonetics* (under revision).

[56] F. Tomaschek, D. Arnold, R. van Rij, et al. "Proficiency effects on the movement precision during the execution of articulatory gestures." In: *The Journal of the Acoustical Society of America* (under revision).

[57] F. Tomaschek, B. V. Tucker, et al. "Vowel articulation affected by word frequency." In: *Proceedings of the 10th ISSP*. Cologne, 2014, pp. 425–428.

[58] D. R. van Bergem. "Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress." In: *Speech Communication* 16.4 (1995), pp. 329–358. ISSN: 0167-6393.

[59] J. van Rij et al. *itsadug: Interpreting Time Series, Autocorrelated Data Using GAMMs*. R package version 0.8. 2015.

[60] S. N. Wood. “A simple test for random effects in regression models.” In: *Biometrika* 100 (2013), pp. 1005–1010.

[61] S. N. Wood. “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.” In: *Journal of the Royal Statistical Society (B)* 73 (2011), pp. 3–36.

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

[63] S. N. Wood. “On p-values for smooth components of an extended generalized additive model.” In: *Biometrika* 100 (2013), pp. 221–228.

[64] C. E. Wright and Meyer. “Conditions for a linear speed-accuracy tradeoff in aimed movements.” In: *The Quarterly Journal of Experimental Psychology Section A* 35.2 (1983), pp. 279–296.

[65] G.K. Zipf. Cambridge, Massachusetts: Addison-Wesley Press, 1949.

[66] G.K. Zipf. *The Psycho-Biology of Language. An Introduction to Dynamic Philology.* Cambridge: MIT Press, 1935, xxii, 625 p.