University of Tübingen, University of Alberta, University of Bristol Kinematic proficiency

Version: April 5, 2018

Practice makes perfect: The consequences of lexical proficiency for articulation

Fabian Tomaschek, Benjamin V. Tucker, Matteo Fasiolo, R. Harald Baayen

April 5, 2018

Abstract

Many studies report shorter acoustic durations, more co-articulation and reduced articulatory targets for frequent words. This study investigates a factor ignored in discussions on the relation between frequency and phonetic detail, namely, that motor skills improve with experience. Since frequency is a measure of experience, it follows that frequent words should show increased articulatory proficiency. We used EMA to test this prediction on German inflected verbs with [a] as stem vowels. Modeling median vertical tongue positions with quantile regression, we observed significant modulation by frequency of the U-shaped trajectory characterizing the articulation of the [a:]. These modulations reflect two constraints, one favoring smooth trajectories through anticipatory co-articulation, and one favoring clear articulation by realizing lower minima. The predominant pattern across sensors, exponents, and speech rate suggests that the constraint of clarity dominates for low-frequency words. For medium-frequency words, the smoothness constraint leads to a raising of the trajectory. For the higher-frequency words, both constraints are met simultaneously, resulting in low minima and stronger co-articulation. These consequences of motor practice for articulation challenge both the common view that a higher frequency of use comes with more articulatory reduction, and cognitive models of speech production positing that articulation is post-lexical.

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

1 Introduction

How words are realized in speech varies substantially. A survey of spontaneous conversations (Johnson 2004) indicates that roughly 20% of words miss at least one phone (see also Turnbull 2018). Several factors have been identified which co-determine the words' phonetic details. One is audience design. Speakers may articulate words more carefully when speaking to an audience, but in familiar contexts, they may hypo-articulate (Lindblom 1990). On the cline from hyper-articulation to hypo-articulation, words become shorter, vowels more centralized, and segments and syllables deleted (Moon and Lindblom 1989; Lindblom 1990; Junqua 1993; Browman and Goldstein 1986; Browman and Goldstein 1989; Liberman and Mattingly 1985). A second factor is occurrence frequency. High frequency words have shorter acoustic durations when factors such as number of segments are controlled for (Bell et al. 2009; Gahl 2008). Several functional explanations for the negative correlation of frequency have been put forward. For example, Zipf (1949) pointed out that longer words require more articulatory effort (cf Lebedev, Tsui, and Van Gelder 2001, for hand movements), and that general biological constraints to reduce the costs of speaking will 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 (see Cohen Priva 2015, for segments). As frequent words are less informative, the complexity of their acoustic signal is hypothesized to reduce in order to maintain a uniform flow of information. The hypothesis is under scrutiny by Clopper, Turnbull, and Burdin (2018), Cohen Priva and Jaeger (2018), Hall et al. (2018) in the current issue.

A third factor might be 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. For rare words, phonological words become available later in time; for frequent words, they are available earlier.

The terminology to describe shorter variants — articulatory undershoot, hypo-articulation, reduction — reflects the normative status according to the citation form in dictionaries. This negative evaluation does not do justice to the rich communicative values of 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, namely, the increase in skilled execution of articulatory gestures with experience. The three factors discussed above paint a picture of shorter forms to be impoverished and less informative. 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.

2 Kinematic proficiency in hand movements

Before introducing our experiment, we provide an introduction to some relevant results in a related domain: 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 movement accuracy. Experimentally, movement precision is gauged by the magnitude of the error between the executed trajectory and the optimal trajectory, or by the magnitude of the error between the movement's endpoint and its target. According to Fitts' law (Fitts 1954; Bertucco and Cesari 2010),

$$t = a + b \log_2(2d/w). \tag{1}$$

Equation (1) clarifies that decreasing movement time t for a fixed distance d goes hand in hand with an increase in variability w (Langolf, Chaffin, and Foulke 1976). Also, movements which are executed at speeds exceeding the current level of proficiency will be less accurate. When proficiency increases, trajectories become less variable for fixed t and d (Georgopoulos, Kalaska, and Massey 1981; Platz, Brown, and Marsden 1998). Importantly, practice is associated with stronger overlap of two successive gestures and decreasing t for fixed w and d(Sosnik et al. 2004; Raeder, Fernandez-Fernandez, and Ferrauti 2015; Platz, Brown, and Marsden 1998).

3 Kinematic proficiency in articulation

Articulation is a complex motor skill which 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 the articulators. Given the above kinematic principles, we can expect the articulatory record to show that, with more frequent use, articulatory gestures of words become less variable (cf. Tomaschek, Arnold, R. van Rij, et al. under revision), 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 the manner in which 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.

Several 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 in 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 as experience accumulates over the course of one's life, 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, Hendrix, 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 reanalyzed the data from (Tomaschek, Tucker, et al. 2014). Given the literature summarized above, we investigated how kinematic practice, parameterized by a word's occurrence frequency shapes the target of articulation as well as changes 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 native speakers of German (9 female, mean age: 26, sd: 3), undergraduate students at the University of Tübingen, with no known language impairments, took part in the experiment. They were either paid 10 Euro for their participation, or received course credit.

4.2 Stimuli

Twenty-seven German verbs with the vowel [aː] in the stem were used. All verbs were presented in a *sie* ... phrase which is disyllabic in its canonical form (e.g., [ziːtsaːlən]). Nine of these verbs were also presented in a phrase eliciting a monosyllabic verb form ([iːɐtsaːlt]). Verbs were selected to cover a wide range of relative frequencies, extracted from SDEWAC (Faaß and Eckart 2013; Shaoul and Tomaschek 2013), a corpus of written texts collected from the internet. It is conceivable that frequencies of written word forms misrepresent the occurrence frequencies in the spoken language. For the 8 stimuli that also occur in the Karl-Eberhards-Corpus (KEC) of spontaneously spoken German (KEC Arnold et al. 2017), the Spearman rank correlation between the frequencies in the KEC and those in the SDEWAC is 0.9. This indicates that the written frequencies are not so different from spoken frequencies as to invalidate our use of written frequency to study articulation.

Log-transformed relative occurrence frequencies were not a significant predictor of the acoustic durations of the word stimuli ($\beta = -0.0002, s.e. = 0.012, t = -0.017$), and was also not predictive for the acoustic duration of the stem vowel ($\beta = 0.012, s.e. = 0.014, t = 0.884$) in mixed-effects models with random intercepts for subject and word.

4.3 Recording

Recordings took place in a soundproof booth at the Department of Linguistics of the University of Tübingen. Participants uttered words presented randomly on a computer screen. The presentation was divided into three parts. Each part was presented first in a slow and then in a fast speaking condition (inter-stimulus interval: 600/300 ms; presentation-time: 800/450 ms). The tongue's movements were recorded with the Northern Digital WAVE articulograph (sampling rate: 100 Hz). Simultaneously, the audio signal was recorded (Sampling rate: 22.05 kHz) and synchronized with articulatory recordings. Head movements were automatically corrected using a 6DOF reference sensor, attached to participants' forehead. Before tongue sensors were attached, a recording was made to determine the rotation from the local reference to a standardized coordinate system, defined by a bite plate with three sensors in a triangular configuration. Tongue movements were captured by three midsagitally placed 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 1cm). We report the findings for TT and TB, as TM's movement pattern mirrors the one of TT, albeit with a slightly reduced amplitude.

4.4 Preprocessing

Tongue positions were centered at the midpoint of the bite plate and rotated such 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, audio signals were automatically aligned with phonetic transcriptions by means of a Hidden-Markov-Model-based forced aligner for German (Rapp 1995). Vowel alignments were manually corrected, where necessary, on the basis of significant changes in the oscillogram using Praat (Boersma and Weenink 2015). Analyses of movement trajectories were restricted to [aː]'s acoustic boundaries.

4.5 Statistical analysis

We used quantile GAMs (QGams) to investigate how the sensor positions changed over time, and how these articulatory trajectories were modified by frequency of use and inflectional exponent (R package **qgam**, Version 1.1.1, based on the **mgcv** package, Version 1.8-23, for R Version 3.3.3, R Core Team (2014), visualized with **itsadug** (J. van Rij et al. 2015), Version 2.3). QGams (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, enabling the modeling of wiggly curves and wiggly (hyper)surfaces. Wiggly curves were fit with thin plate regression splines. Interactions of covariates with time were modeled with tensor product smooths (Baayen, Vasishth, et al. 2017).

The choice for modeling articulatory trajectories with quantile GAMs was motivated by strong autocorrelations present in the residuals of the Gaussian GAMs initially fit to our data. Time series of 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 implement a distribution-free method for estimating the predicted values of a given quantile of the response distribution, together with confidence intervals, they are a natural and powerful alternative. In our analyses, we investigated the median, but other quantiles can also be of theoretical interest (Schmidtke, Matsuki, and Kuperman 2017).

5 Analysis

Speakers sometimes reduced schwas in [ən], resulting in forms such as [ziːtsaːlən] realized as [ziːtsaːln]. We, therefore, created a three level factor (using treatment dummy coding), inflectional EXPONENT: stem+[t] (N = 286), stem+[n] (N = 197), and stem+[ən] (N = 344). For inclusion in the [ən] group, schwa duration had to exceed 50 ms. The proportion of [ən] to [n] varied between 0.25 and 0.93 across participants and between 0.32 and 0.88 across words. No significant effect of frequency was found in a mixed-effects regression. The reference level of EXPONENT was [ən]. Vowel durations were normalized between 0 and 1, henceforth TIME. Articulatory trajectories are influenced by the contexts in which they occur. As a consequence, for each verb (abstracting away from its inflectional exponents), the consonants flanking [aː] are expected to have their own specific effect on [aː]. We, therefore, included by-stem factor smooths for TIME in our models. These factor smooths are nonlinear equivalents of the combination of by-verb random intercepts and by-verb random slopes for TIME in the linear mixed model (Baayen, Vasishth, et al. 2017). By including factor smooths, we stack the cards against the hypothesis that words' occurrence frequency also co-determines the articulatory trajectories. In our qGAMs, a frequncy effect has to establish itself over and above the co-articulatory effects of [aː]'s context. Since the effect of the inflectional exponents on articulation is probed with the factor EXPONENT, the combination of the by-stem factor smooths and EXPONENT bring all parts of the word forms under statistical control which 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

$$\mathrm{E}[\rho_\tau(y-\eta)|\boldsymbol{x}], \tag{2}$$

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

$$\begin{aligned} \eta_{i,j,t} &= \beta_0 + b_i + \mathrm{fs}(t,j) + \\ &\quad \alpha_{\texttt{exponent}(j)} + \mathrm{te}(t, \texttt{frequency}_j, \texttt{exponent}(j)), \\ &\quad b_i \sim \mathcal{N}(0,\sigma). \end{aligned} \tag{3}$$

For the fast/slow speaking condition, no sensor data were available for the tongue tip sensor for 464/550 measurement points (data loss 6.5%/6.3%), for the tongue mid sensor for 1515/706 data points (17.4%/8.1%), and for the tongue body sensor for 370/338 measurement points (5.2%/3.8%). Separate qGAMs were fit to the remaining data points for each combination of sensor and speaking rate. To take into account variability due to different articulatory complexities as well as temporal variation, we used vowel duration and word duration as covariates in pilot analyses to control for known modulations of the trajectory due to time constraints. While improving the model fit, these covariates did not affect the modulation of articulation by frequency, and are therefore not included in the analyses reported here. Table 1 presents the model summaries, Figure 1 presents the by-word factor smooths for time for the fast condtion, and Figure 2 visualizes the partial effects of the smooths for the time×frequency×exponent interaction for both speaking rates.

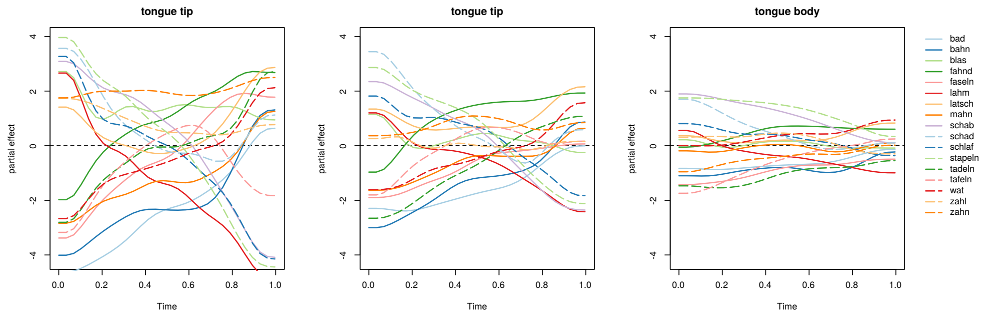

Figure 1: *Partial effects of the by-word factor smooths for time, for tongue tip (left), tongue mid (center) and tongue body sensors (right). Each curve represents a word. Note that the adjustment is largest in tongue tip and smallest in tongue body. Across positions, the same words tend to show roughly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

tongue tip sensor – fast speaking condition				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-6.5294	0.9871	-6.6149	< 0.0001
Exponent=t	0.9679	0.0846	11.4441	< 0.0001
Exponent=n	0.5865	0.0852	6.8808	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time,Frequency):Exponent=en	12.0762	13.1409	454.4291	< 0.0001
te(Time,Frequency):Exponent=t	12.3319	13.2206	432.3513	< 0.0001
te(Time,Frequency):Exponent=n	8.9767	9.7610	392.5233	< 0.0001
s(Participant)	15.9830	16.0000	11197.2793	< 0.0001
s(Time,Word)	109.7432	152.0000	5441.2205	< 0.0001
tongue tip sensor – slow speaking condition				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-6.5626	0.9053	-7.2488	< 0.0001
Exponent=n	-0.3570	0.0821	-4.3463	< 0.0001
Exponent=t	0.8931	0.0740	12.0680	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time,Frequency):Exponent=en	10.3480	10.7772	569.3480	< 0.0001
te(Time,Frequency):Exponent=n	10.2631	10.8002	429.3652	< 0.0001
te(Time,Frequency):Exponent=t	9.8095	10.5061	423.2987	< 0.0001
s(Time,Word)	120.3414	152.0000	4985.3140	< 0.0001
s(Participant)	15.9829	16.0000	10087.2380	< 0.0001
tongue body sensor – fast speaking condition				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-0.3477	1.2850	-0.2706	0.7867
Exponent=t	0.3526	0.0709	4.9764	< 0.0001
Exponent=n	0.4284	0.0737	5.8105	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time,Frequency):Exponent=en	6.1325	6.6809	315.9888	< 0.0001
te(Time,Frequency):Exponent=t	7.5259	7.8432	254.1614	< 0.0001
te(Time,Frequency):Exponent=n	6.7701	7.3977	301.5194	< 0.0001
s(Participant)	15.9941	16.0000	40428.3445	< 0.0001
s(Time,Word)	63.6006	152.0000	5328.7791	< 0.0001
tongue body sensor – slow speaking condition				
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-0.6855	1.4453	-0.4743	0.6353
Exponent=t	-0.2231	0.0697	-3.2026	0.0014
Exponent=n	-0.5463	0.0755	-7.2331	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time,Frequency):Exponent=en	10.0310	10.6392	449.1261	< 0.0001
te(Time,Frequency):Exponent=t	9.5115	10.3115	312.4821	< 0.0001
te(Time,Frequency):Exponent=n	9.5453	10.4671	260.5343	< 0.0001
s(Time,Word)	59.3821	152.0000	1667.3226	< 0.0001
s(Participant)	15.9956	16.0000	58326.4913	< 0.0001

Table 1: *QGAMs for the vertical position of the tongue tip and tongue body sensor in fast (top panels) and slow speaking condition (bottom panels). te: tensor product smooth, s: thin plate regression spline; fs: factor smooth; re: random effect.*

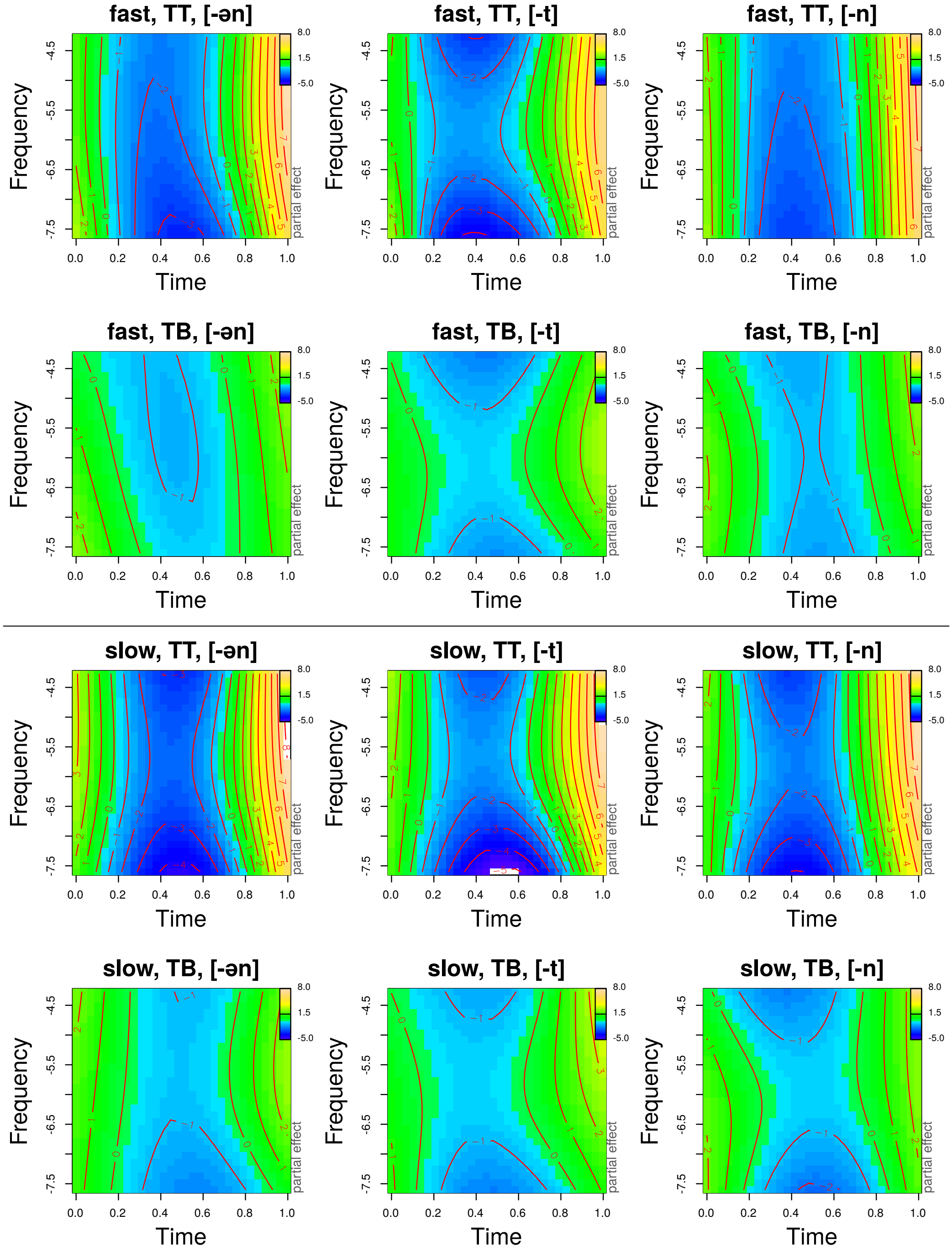

Figure 2: *Partial effect of the interaction of* Time *by log relative* Frequency *on the vertical position of the sensors for the three exponents [-ən], [-t], and [-n]. Deeper shades of blue indicate lower vertical positions, and darker shades of yellow indicate higher vertical positions. Upper panels: tongue tip and tongue body sensor in the fast speaking rate, lower panels: the same sensor positions in the slow speaking rate. TT: tongue tip sensor; TB: tongue body sensor.*

Parametric summaries (labeled A in Table 1) indicate that [aː] in words with [-t] and [-n] exponents were realized higher than in words with the [-ən] exponent. In the slow speaking condition, words with the [-t] and [-n] had lower vertical positions, except for the [-t] for the tongue tip sensor.

Random intercepts (in mm) for participants further modulate the sensor's vertical position on a speaker-by-speaker basis. Standard deviations for 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) in the fast speaking condition. In slow speaking condition: tongue tip: 3.23 (95% confidence interval (2.28,4.57) to tongue mid sensor: 4.27 (95% confidence interval (3.0,6.1), and further for the tongue body sensor: 5.91 (95% confidence interval 4.18, 8.35)

Figure 1 visualizes by-word factor smooths for time, in the fast speaking condition. Each curve represents a word. 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, curves are roughly similar across the three sensor positions. For *stapeln*, for instance, we find a downward trend, and for *bad*, an upward trend. Figure 1 also shows that the movement amplitude decreases from tongue tip to tongue body.

Although the qGAM models identified individual articulatory trajectories for each combination of lemma and exponent, three-way word×frequency×time interactions received solid support in both speaking conditions. The upper two rows of Figure 1 represent the partial effects for the tongue tip and tongue body sensors in the fast speaking condition. The corresponding partial effects for the slow speaking condition are presented in the lower two rows.

The left column in Figure 1 shows the effects for [-ən], the center and right columns present the effects for [-t] and [-n]. Across all panels, deeper shades of blue denote lower vertical positions, and warmer shades of yellow higher positions. Colors are less bright for the tongue body than for the tongue tip sensor; sensors further into the mouth show smaller movement amplitudes.

As expected for [aː], darker shades of blue are found in the center of the time interval (horizontal axis), indicating roughly U-shaped articulatory trajectories (changes from yellow to blue and back to yellow for any line parallel to the x-axis). The exact shape of this U-shaped curve, however, is modulated by word frequency (vertical axis). If frequency would have had no effect at all, all contour lines would be straight vertical lines. Model comparisons pitting the present models against models without frequency and the time×frequency interaction provide strong support for the relevance of frequency as predictor.

[1]

QGAMs reveal three types of modulation of the U-shaped trajectories of the sensors. The first type is found in the fast speaking condition for tongue tip during the articulation of [-ən] and [-n], and in the slow speaking condition for both tongue tip and tongue body when articulating [-ən] (c.f. Figure 1, upper left panel). The lowest point of articulation is reached for the

[1] When the time×frequency×exponent is replaced by time×exponent and frequency×exponent, ML scores increased in all the sensors in the fast (TT: +94.0, TM: +11.6, TB: +244.4), but only in one in the slow speaking condition (TT: -46.4, TM: +141.7, TB: -649.9). We also analyzed tongue movements which decomposed time×frequency×exponent into main effects and the interaction riding on top of these main effects, using the `ti` function instead of the `te` function. The inclusion of the additive `ti` interaction reduced the ML-score for sensors in the fast speaking condition in all the sensors (TT: -123.869, TM: -16.4, TB: -215.0), in the slow speaking condition for tongue tip and tongue body (TT: -34.0, TM: +28.8, TB: -271.4). These models provided less precise fits to the data compared to the models that made use of the tensor product smooths fit with `te`. No test for model comparison is available for qGams, but the magnitude of the reduction in ML scores obtained by including the three-way interaction is sufficient to conclude that a three-way interaction is indeed present.

lowest-frequency words. As frequency is increased, this minimum increases likewise. For everything except for the tongue tip sensor for [-n] in the fast speaking condition, we find that, for medium and high frequency words, the tongue sensor starts to move back up, and reaches a higher offset position, than is the case for the lowest frequency words. In other words, as frequency increases, the amplitude of the U-shaped movement stays roughly the same, but the whole movement is executed at a higher position in the oral cavity. In other words, as articulatory proficiency increases, the upcoming alveolar exponent is anticipated by overall raising of the tongue.

The second type of modulation is found only for the tongue body sensor realizing [-ən] in the fast speaking condition (see the first panel on the second row of Figure 1). Here, we see that the U-shaped trajectory is lowered as frequency is increased, without any indication of earlier anticipatory co-articulation with the upcoming exponent. Articulatory proficiency is visible here in the form of an overall deeper articulation.

For the remaining 7 sensor-exponent-speaking condition combinations, the modulation of the U-shaped trajectory by frequency is intermediate between the preceding two types. As can be seen in the second panel on the first row of Figure 1, the minimum of the U-shaped trajectory first increases with frequency, and then decreases. Medium-frequency words show the earliest onset of anticipatory co-articulation, and sensors reach a higher offset position than is the case for the lowest-frequency words. For the highest-frequency words, the offset position reached tends also to be higher than that reached for the lowest-frequency words, even though a lower minimum is reached than for the medium-frequency words. Compared to the lower-frequency words, higher-frequency words show a mastery of deeper articulation of [aː] in combination with more and relatively earlier anticipatory co-articulation.

This hour-glass pattern characterizes the articulation of [-t] and [-n], but is not present for [-ən]. Since, for the latter exponent, the alveolar target for co-articulation is at a greater distance from the stem vowel, being separated from it by the schwa, the pressure for co-articulation is reduced for words with [-ən]. As a consequence, the two opposing constraints on articulation of the stem vowel, one favoring a low minimun to ensure a clear [aː] (the clarity constraint), and the other favoring co-articulation with the upcoming exponent to ensure smooth articulatory gestures (the smoothness constraint), are optimized differently. Whereas for words with the [-t] and [-n] exponents, they are optimized jointly with the smoothness constraint preceding the clarity constraint as frequency is increased, words with the [-ən] exponent appear to favor the clarity constraint, ignoring the smoothness constraint, thus building a better contrast between the [aː] and the upcoming schwa. It is only for the tongue body sensor in the fast speaking condition that smoothness is favored to the exclusion of vowel clarity.

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 tongue sensors. We controlled statistically for the effects of the consonants surrounding the stem vowel by including by-word factor smooths for time in the analyses. Modeling the median tongue position using quantile regression, we observed significant modulation by frequency of the U-shaped vertical trajectory that characterizes the articulation of the [aː]. These modulations reflect two constraints, one constraint favoring smooth trajectories through anticipatory co-articulation, and one constraint favoring clear articulation of the [aː] by realizing lower minima. The predominant pattern across sensors, exponents, and speech rate suggests that the constraint of clarity dominates for low-frequency words. For medium-frequency words, the smoothness constraint leads to an overal raising of the trajectory. For the higher-frequency words, both constraints are met simultaneously, resulting in low minima and earlier and stronger co-articulation.

This hour-glass pattern of modulation was characteristic of the [-t] and [-n] exponents, but was absent for the [-ən] exponent. We think this is due to the schwa in the exponent lowering the pressure for coarticulation. Furthermore, under time pressure, in the fast speaking condition, the interaction of time by frequency may be almost completely absent. In our data, this is the case for the tongue tip sensor for words with the [-n] exponent. Here, higher frequency words have slightly higher minima, but there is no clear effect of earlier or stronger co-articulation. Apparently, the fast speaking condition induced a loss of clarity without speakers being able to compensate for smoothness.

Considering all sensors together, the lowest minima tend to be present for the lowest-frequency words. This dovetails well with the finding that formants in lower-frequency words are articulated further away from the center of the vowel space (Aylett and Turk 2004; Meunier and Espesser 2011).

Returning to the hour-glass pattern dominating the interaction of time by frequency, we note that the balance of the constraints of clarity and smoothness can also be seen as reflecting the opposing pressures of predictability and discriminability (Blevins, Milin, and Ramscar 2015). The pressure of predictability favors a clear low vocalic center and a strong U-shaped trajectory that can be executed in a similar way across words. The pressure of discriminability, by contrast, is served by co-articulation. Co-articulation causes inflected forms to have somewhat different vowels, and as a consequence, these words are possibly more easily teased apart by the listener (cf. Kemps, Wurm, et al. 2005a; Kemps, Ernestus, et al. 2005b). The optimization for both constraints visible for higher-frequency words is perhaps made possible by the presence of larger, multi-phone planning units for these words (see, e.g., Hickok 2014).

It is tempting to project the three stages of modulation of the U-shaped curve visible in the hour-glass interactions of time by frequency onto the development of kinematic proficiency across the lifespan. Initially, the constraint of clarity would dominate speech production, but with age, speakers would learn to satisfy both the constraint of clarity and the constraint of smoothness. However, the literature comparing child speech with adult speech does not provide clear support for this conjecture, as findings are inconsistent. Barbier et al. (2015) and Zharkova, Hewlett, and Hardcastle (2012) report that adults co-articulate more than children, whereas Sussman et al. (1999), Zharkova, Hewlett, and Hardcastle (2011), and Katz and Bharadway (2001) report exactly the opposite. Furthermore, Noiray, Menard, and Iskarous (2013) failed to find any differences in anticipatory co-articulation between adults and children. As most of these studies are based on non-words, and given the consequences of life-long learning for lexical processing (Ramscar, Hendrix, et al. 2014; Ramscar, Sun, et al. 2017), the demands made on children and adults requested to produce the same nonwords are very different. As a consequence, the question of whether the hour-glass shaped pattern generalizes to changing proficiency over the lifetime has to be left open.

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. (accepted) showed that frequency modulates speed of articulation in relation to the curvature of movement trajectories. Segments with more narrow curves were articulated faster in more practiced words. Furthermore, more frequent words were articulated with smoother trajectories (cf. Sosnik et al. 2004; Tiede et al. 2011), and Tomaschek, Arnold, R. van Rij, et al. (under revision) showed that articulatory movements at more probable word boundaries were produced with greater precision (cf. Goffman et al. 2008).

All these results were obtained for laboratory speech, using a registration technique that requires placement of sensors on the 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; Foulkes et al. 2018, a discussion of potential differences between laboratory and spontaneous speech). Replication studies, ideally based on corpora of spontaneous speech with EMA or ultrasound registration, are essential for consolidating the present body of evidence.

If the present results are pointing in the right direction, they have two important theoretical implications. First, the finding 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 Meyer (1999) , and Goldstein et al. (2009), which assume that the representations driving articulation are assembled of phonemes and morphemes, or of the gestural scores associated with these units. These models cannot straightforwardly accommodate the present evidence that experience at the level of individual words co-determines how articulatory trajectories are actually realized (see also Gahl 2008; Daland and Zuraw 2018).

Second, higher-frequency forms are not necessarily more ‘reduced’. The decrease in acoustic duration that has been observed many times for higher-frequency words (van Bergem 1995; Aylett and Turk 2006; Schulz et al. 2016; Meunier and Espesser 2011) does not logically entail that the constraint of clarity is discarded in favor of the constraint of smoothness. When motor skills improve, complex motions can be executed both faster and with more precision. The argument that shorter acoustic durations go hand in hand with reduction of articulatory movements appears to gain further support from the observation that vowels in higher-frequency words centralize more (Aylett and Turk 2006; Meunier and Espesser 2011) . However, more vowel centralization for higher frequency words is entirely compatible with the present findings, as the minima for high-frequency words tend to be slightly higher than those of low-frequency words while going hand in hand with more tongue raising for co-articulation. Thus, global measures of vowel centralization that do not take into account how centralization varies with time will, also for our data, single out higher-frequency words as more reduced. We note here that evidence present in the articulography record may not be available in the acoustic signal (see Wieling et al. 2016, for the case of different articulatory settings between dialects). We therefore conclude that the consequences of kinematic proficiency for articulation are worthy of further investigation.

Acknowledgements

This research was funded in part by the Alexander von Humboldt professorship awarded to R. H. Baayen (grant 1141527), and in part by a collaborative grant from the Deutsche Forschungsgemeinschaft (BA 3080/3-1). We are indebted to Ryan Callihan, Samantha Tureski, and two anonymous reviewers for their helpful comments on previous versions of this paper.

References

[1] D. Arnold et al. "Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm 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] G. Barbier et al. “Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses.” In: *16th Annual Conference of the International Speech Communication Association (Interspeech 2015).* Proceedings of Interspeech 2015. International Speech Communication Association. Dresden, Germany, Sept. 2015. URL: `https://hal.archives-ouvertes.fr/hal-01200984`.

[7] 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.

[8] 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.

[9] 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.

[10] Paul Boersma and David Weenink. *Praat: doing phonetics by computer [Computer program], Version 5.3.41, retrieved from http://www.praat.org/.* 2015.

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

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

[13] C. G. Clopper, R. Turnbull, and Burdin. “Assessing predictability effets in connected read speech.” In: *Linguistics Vanguard* 4.S2 (2018).

[14] U. Cohen Priva. “Informativity affects consonant duration and deletion rates.” In: *Laboratory Phonology* 6.2 (2015), pp. 243–278.

[15] U. Cohen Priva and F. Jaeger. “The interdependence of frequency, predictability, and informativity.” In: *Linguistics Vanguard* 4.S2 (2018).

[16] R. Daland and K. Zuraw. “Loci and locality of informational effects on phonetic implementation.” In: *Linguistics Vanguard* 4.S2 (2018).

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

[18] 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.

[19] 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.

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

[21] 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.

[22] P. Foulkes et al. “Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory.” In: *Linguistics Vanguard* 4.S2 (2018).

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

[24] S. Gahl and R. H. Baayen. “Twenty-eight years of vowels.” In: *under revision* (2017).

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] K. C. Hall et al. “The Role of Predictability in Shaping Phonological Patterns.” In: *Linguistics Vanguard* 4.S2 (2018).

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

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

[32] 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.

[33] 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.

[34] 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.

[35] W.F Katz and S. Bharadway. “Coarticulation in fricative-vowel syllables produced by children and adults: a preliminary report.” In: *Clinical linguistics and phonetics* 15.1 (2001), pp. 139–143.

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

[37] 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 (2005a), pp. 43–73.

[38] 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.

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

[40] 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.

[41] 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.

[42] 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).

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

[44] 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.

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

[46] 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.

[47] 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.

[48] A. Noiray, L. Menard, and K. Iskarous. “The development of motor synergiers in children: Ultrasound and acoustic measurements.” In: *Journal of the Acoustical Society of America* 133.1 (2013), 444ff.

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

[50] 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.

[51] 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.

[52] 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.

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

[54] M. Ramscar, C. C. Sun, et al. “The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the “Cost” of Learning, Not Cognitive Decline.” In: *Psychological Science* (2017). URL: https://doi.org/10.1177/0956797617706393.

[55] 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.

[56] 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).

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

[58] 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.

[59] 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.

[60] H. M. Sussman et al. “An Acoustic Analysis of the Development of CV Coarticulation – A Case Study.” In: *Journal of Speech, Language, and Hearing Research* 42.5 (1999), pp. 1080–1096. DOI: 10.1044/jslhr.4205.1080. URL: +%20http://dx.doi.org/10.1044/jslhr.4205.1080.

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

[62] F. Tomaschek, D. Arnold, Franziska Bröker, et al. “Lexical frequency co-determines the speed-curvature relation in articulation.” In: *Journal of Phonetics* (accepted).

[63] F. Tomaschek, D. Arnold, R. van Rij, et al. “Proficiency effects on the movement precision during the execution of articulatory gestures.” In: (under revision).

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

[65] R. Turnbull. "Patterns of probabilistic segment deletion/reduction in English and Japanese." In: *Linguistics Vanguard* 4.S2 (2018).

[66] 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.

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

[68] M. Wieling et al. "Investigating dialectal differences using articulography." In: *Journal of Phonetics* (2016).

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

[70] 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.

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

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

[73] N. Zharkova, N. Hewlett, and W. J. Hardcastle. "An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children." In: *Journal of the International Phonetic Association* 42.2 (2012), pp. 193–208.

[74] N. Zharkova, N. Hewlett, and W. J. Hardcastle. "Coarticulation as an indicator of speech motor control development in children: an ultrasound study." In: *Motor Control* 15.1 (2011), pp. 118–140.

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