How is anticipatory coarticulation of suffixes affected by practice?

–DRAFT–

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

Department of General Linguistics, University of Tübingen, Germany Department of Linguistics, University of Alberta, Canada

Corresponding author:

Fabian Tomaschek

Seminar für Sprachwissenschaft

Eberhard Karls University Tübingen

Wilhelmstrasse 19

Tübingen

e-mail: fabian.tomaschek@uni-tuebingen.de

Version: January 20, 2018

Contents

Abstract

More and more studies indicate differences in fine phonetic detail related to the morphological function of words and segments. In the present study, we investigated to what extent these differences arise due to anticipatory coarticulation and practice with individual word forms. We hypothesized to find stronger anticipatory coarticulation of inflectional exponents during the articulation of the stem vowel in American English verbs such as "cleaned", "cleans", "cleaning" in contrast to the bare stem "clean", the better practiced these words are. However, we observed exactly the oposite effect with better practiced words showing stronger hyper-articulated and less co-articulated vowels. We link these results to findings that morphological function is reflected in fine phonetic detail, challenging traditional models of speech production, which assume a separation of lexical information and the phonetic detail.

Index Terms: morphology, anticipatory coarticulation, naïve discriminative learning, practice

1 Introduction

1.1 Background

According to one of the most influential models of speech production – the theory of lexical access by Levelt, Roelofs, and Meyer (1999) – speech production is a modular, sequential process, during which concepts activate lemmata in a linguistic lexicon. These access abstract building blocks in the form of morphemes. After merging the morphemes, these new units are recoded during a postlexical process, first into a phonological form, then into a syllabic form which drive motor programs for articulation. The theory of Levelt et al. (1999) and its computational formalization (Roelofs, 1997) hold that a word's lexical characteristics are encapsulated from the articulation process. In other words, the phonetic signal should not contain any differences due to syntactic or morphological information.

However, a growing number of studies shows that higher-level lexical properties indeed correlate with fine phonetic details, challenging the assumptions about an encapsulated modular speech production process, contradicting the predictions of Levelt et al. (1999)' theory.

For example, Gahl (2008) reported longer acoustic durations for homophones whose lemma has a lower frequency of occurrence (e.g. 'thyme' vs 'time'). This finding was replicated by Lohmann (2017) for noun-verb homophoneous word pairs. Likewise, Drager (2011) observed that segment durations in the English word 'like' vary depending on its different grammatical functions. (Caselli, K., & Cohen-Goldberg, 2016) report that the acoustic durations of monomorphemic words (e.g. 'pride') correlates positively with the number of inflected phonological neighbors (e.g. 'spied') while it correlates negatively with the number of uninflected phonological neighbors (e.g. 'side'). All these findings put the idea into question that homophonous words have identical articulatory motor programs represented in the form of a syllabary, which are independent from other programs. Similar differences due to higher lexical information have been shown at the syllabic and segmental level, providing further evidence which challenges the idea that articulatory motor commands are driven solely from syllabic units.

Cho (2001) reported that the variability of gestural coordination during consonant cluster articulation is larger when the consonants are located at morpheme boundaries than when they are within a morpheme. Pluymaekers, Ernestus, and Baayen (2005) reported that the acoustic duration of affixes were co-determined by the frequency of occurrence of the carrier word. Just recently, I. Plag, Homann, and Kunter (2017) showed that the acoustic duration of homophonous [s] and [z] in American English depend on their morphological function. A similar finding was reported by Seyfarth, Garellek, Gillingham, Ackerman, and Malouf (2017), who also also showed systematic differences in stem duration between inflected words and in non-inflected words (e.g. 'laps' vs. 'lapse'). Furthermore, these studies question assumptions about how much word specific information is present in the mental lexicon, as findings by Kemps, Wurm, Ernestus, Schreuder, and Baayen (2005a) and Kemps, Ernestus, Schreuder, and Baayen (2005b) have shown that speakers seem to be sensitive to these systematic fine phonetic differences associated with morphological categories – in this case stem durations in singular and plural nouns.

1.2 The present study

In the present study, we further examined the amount of fine phonetic detail related to morphological categories during the articulation of stem vowels in monomorphemic and dimorphemic verbs, i.e. verbs with and without inflectional exponents (e.g. "clean", "cleaned", "cleans", "cleaning") by means of electromagnetic articulography.

Since the upcoming context varies systematically between the morphological categories (i.e. [??] vs. [d] vs. [s] vs. [ɪŋ]), we hypothesize to find differences in the articulations of stem vowels as a result of anticipatory coarticulation of the inflectional exponent.

Coarticulation describes the articulatory and acoustic variability of identical phones due to the variation of preceding or upcoming phones. Anticipatory coarticulation has been shown to occur between directly neighboring segments, across consonants and syllable boundaries, as well as across syllables (Bell-Berti & Harris, 1979, 1982; Goffman, Smith,

Heisler, & Ho, 2008; Hoole, Nguyen-Trong, & Hardcastle, 1993; Katz & Bharadway, 2001; Magen, 1997; Öhman, 1966; Recasens, 1984; Sziga, 1992) and affects the articulatory patterns of the entire phone (Fowler & Brancazio, 2000; Hoole et al., 1993). Furthermore, it is well known that spectral and durational changes in the fine phonetic detail are associated with the frequencies of occurrence and conditional probabilities of word sequences, single words, syllables and even segments (Aylett & Turk, 2004; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Caselli et al., 2016; Gahl, 2008; Priva, 2015; Tremblay & Tucker, 2011; Whalen, 1991; Wright, 1979; Zipf, 1935). Studies having investigated vowels report shorter and more centralized vowels in more frequent words (Aylett & Turk, 2006; Meunier & Espesser, 2011; Priva, 2015). In a information theoretic framework, these reductions have been interpreted to go hand in hand with a reduction of information density in the speech signal. In line with this reasoning, we expected to find stronger reduced vowel articulations associated with a higher frequency of occurrence, independently of the morphological category under investigation.
Regarding word frequency as a decontextualized measure of practice, i.e. the articulatory repetition of a word, effects in phonetics can be compared to effects of repetition in other domains of motor control, e.g. hand movements. Here, practice has been found to be associated with greater kinematic skills, such that movement velocity are increased (Platz, Brown, & Marsden, 1998; Raeder, Fernandez-Fernandez, & Ferrauti, 2015) or sequences of movement gestures overlap to a stronger degree (Sosnik, Hauptmann, Karni, & Flash, 2004). These findings are reflected in speech production, where less reduced, i.e. more peripheral vowel articulations associated with higher frequencies of occurrence have been reported (Benjamin Munson, 2001; Tomaschek, Tucker, & Baayen, under revision 2; Tomaschek, Tucker, Wieling, & Baayen, 2014; Tomaschek, Wieling, Arnold, & Baayen, 2013). From this perspective, we should find stronger anticipatory coarticulation of the inflectional exponent, the better words are practiced.

We regard anticipatory coarticulation to reflect kinematic optimization of local gestures for upcoming gestures (Saltzman & Munhall, 1989). As a consequence, we expect

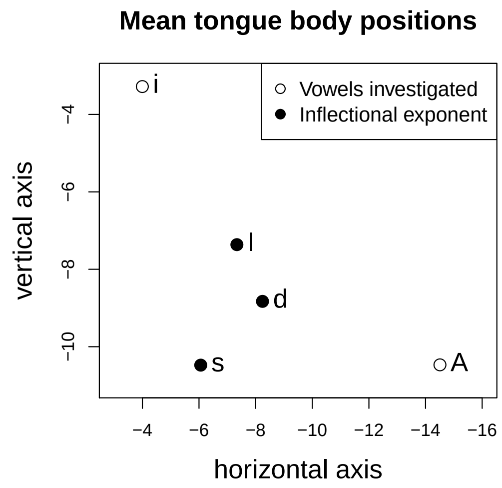


Figure 1. *Mean sensor positions in tongue body for segments of interest pooled across all speakers and words, calculated at the center of the segments.*

anticipatory coarticulation to depend on the spatial constellation between the tongue body position in the vowel under investigation ([ɑː] or [iː]) and in the inflectional exponent. To illustrate this, see Figure 1, which illustrates mean tongue body sensor positions in the vertical/horizontal space at the center of the segments in our data, estimated in two linear mixed-effects models, fitting vertical/horizontal tongue position with segment as a fixed predictor and speaker and word as random intercepts. Empty circles represent the investigated vowels, full circles the inflectional exponents under investigation. With greater practice leading to greater coarticulation (Sosnik et al., 2004), and taking into account the relative mean tongue body position between vowels and the inflectional exponents, we expect greater experience with words to be associated with lower and further back tongue positions in [iː] when verbs have an inflectional exponent in contrast to the bare stem (i.e. "clean"). As for [ɑː], we expect tongue positions to be higher and further front when followed by [d] and [ɪ] in better practiced words in contrast to the bare stem, while preceding [s] should show only stronger fronting in words with greater experience. No changes in the vertical positions should occur because tongue body height is relatively the same in these two sounds.

Table 1
Number of words by vowel and morphological condition.

	present1	past	present2	progressive	Sum	Examples
[ɑː]	20	15	20	20	75	arm, armed, arms, arming
[iː]	37	29	32	35	133	peel, peeled, peels, peeling

2 Methods

2.1 Recording

25 speakers of Canadian and American English (mean age: 29.4, sd: 8.2) were paid to articulate the stimuli. Words were presented in a list. To reduce the effect of repetition, lists were structured according to a latin square design so that one speaker never articulated a verb lemma in two different morphological conditions. The recordings were performed in a sound booth at the Department of Linguistics, University of Alberta, Edmonton. Articulatory movements of the tongue were recorded with an 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 subjects' forehead. Before the 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 to which three sensors in a triangular configuration were attached. Tongue movements were captured by three sensors: one slightly behind the tongue tip, one at the tongue middle and one at the tongue body (distance between each sensor: around 2cm). The present analysis focuses on the tongue tip and the tongue body sensor along both the vertical and the horizontal axis.

2.2 Preprocessing

Tongue movements were corrected for head-movements in an online procedure during recording by the proprietary NDI wave software. The recorded positions of the tongue sensors were centered at the midpoint of the bite plate and rotated in such a way that the back-front direction of the tongue was aligned to the x-axis with more positive values towards the front of the mouth, and more positive z-values towards the top of the oral cavity. No filtering was applied as this would artificially increase the autocorrelation of the data, resulting in anti-conservative p-values. Absolute sensor positions were transformed into distances between the sensor and its maximal vertical/horizontal position. Segment boundaries were determined by automatically aligning the audio signal with phonetic transcriptions by means of a Hidden-Markov-Model-based forced aligner for German (Rapp, 1995). Alignments for the vowel were manually verified and corrected where necessary.

2.3 Word material

We used 406 English words, presented in the infinitive and first person present form (*present1*, *stem*), the third person singular form (*present2*, *stem+s*), the past form (*past*, *stem+d*) and the progressive form (*progressive*, *stem+ing*). We took care that, apart from the progressive form, all other word forms were monosyllabic.

In order to find a sufficiently large number of verbs, our material contained both verb stem structures VC and VCC. We expect sufficient changes in the vowel, even across one consonant, because anticipatory coarticulation has been found to persist across consonant clusters and syllables (Magen, 1997; Sziga, 1992). We controlled for the effects of the intermediate consonant by means of our random effects structure (see Section 2.5). Table 1 shows the number of words for every vowel across the four morphological conditions, in addition to examples for every vowel category under investigation. It is important to note for all words in the [ɑː] category, the vowel was followed by the retroflex [ɹ]. Since (Sziga, 1992) has found anticipatory coarticulation in vowels of consonants separated by another consonant (i.e. in consonant clusters), we expect to find effects of inflectional exponents.

2.4 Operationalization of practice

We decided to use word activations derived from weights in a input-output network calculated by the Naïve Discriminative Learner (package NDL, Version 0.2.17, Baayen, Milin, Đurđević, Hendrix, and Marelli (2011)) in the statistical programming language

R (R version 3.3.3 (2017-03-06), R Core Team (2014)). NDL is based on discriminative learning (Rescorla & Wagner, 1972), an error-driven supervised learning algorithm, according to which speakers/listeners learn to discriminate between lexical classes from sublexical orthographic input (Ramscar, Dye, & McCauley, 2013; Ramscar & Yarlett, 2007). NDL has been shown to predict lexical decision latencies (Baayen et al., 2011; Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017), learning sublexical representations of connected speech (Baayen, Shaoul, Willits, & Ramscar, 2016), and the acquisition of regular and irregular plural forms (Ramscar et al., 2013). Recently, it has successfully been shown to achieve lexical discrimination from acoustic cues automatically derived from spontaneous speech (Arnold, Tomaschek, Sering, Ramscar, & Baayen, 2017) and lexical discrimination from visual cues automatically derived from letter strings (Linke, Bröker, Ramscar, & Baayen, 2017).

We used all verbs from the English Lexicon Project (ELP, Balota et al., 2011) for the estimation of association strength between triphones and wordform. The frequency of occurrence of every cue-to-outcome learning event was based on the HAL frequency provided by the ELP, and the occurrence of the learning events were randomized for learning. Learning rates α and β were set to 0.01 and 0.1, respectively. The maximum possible level of association strength λ was set to 1. As cues we used triphones and word forms as outcomes. Word activations were calculated by summing up the weights between an outcome and all afferent cues in one event.

2.5 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 trajectories were modified by word activation and inflectional exponent. Quantile GAMs (Fasiolo, Goude, Nedellec, & Wood, 2017) integrate quantile regression (Koenker, 2005) with the generalized additive model (GAM, Hastie & Tibshirani, 1990; Wood, 2006, 2011, 2013a, 2013b).

GAMs use spline-based smoothing functions to model nonlinear functional relations

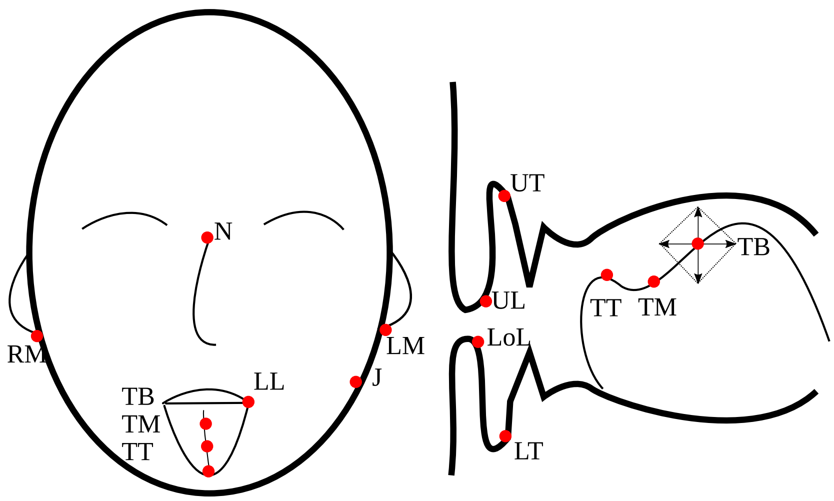


Figure 2. *Illustration of sensor positions. Left: frontal illustration. Right: midsagittal cut through the mouth. The rhombus around the tongue body sensor illustrates the parameterization of its movement area.*

between a response and one or more covariates. This enables 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, Bates, & Kliegl, 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, & 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 (van Rij, Wieling, Baayen, & van Rijn, 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.

Stem vowel duration was normalized to range between 0 and 1. In what follows, we refer to this normalized duration as time. Word activations were ranked, and will henceforth be referred to as activations. Vowel durations were logged and centered. Inflectional exponent, henceforth exponent, was categorized as *present1*, *past*, *present2*, *progressive*. For both vowels and both movement axes, significant modulations in movement patterns across time and activations by exponents were supported by a significant time × activations × exponent interaction, while taking modulations of articulatory trajectories by vowel duration under statistical control by means of a smooth fitting vowel durations and a tensor fitting an interaction between time and vowel duration. The inclusion of the exponent significantly improved model fit in all four models .

Vowels' articulatory trajectories are influenced by the contexts in which these vowels occur. Ignoring the effect of the inflectional exponent, the consonants flanking the vowel are expected to have their own specific effect on how the vowel is articulated (Hoole et al., 1993, , etc.). We therefore included by-verb factor smooths for time in our models (i.e. the lemma). 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 (cf. Baayen et al., 2017). In the QGAMS, an effect of activation has to establish itself independently of the co-articulatory effects of the vowels' context, which we bring under statistical control by including these factor smooths. Since the effect of the inflectional exponents on articulation is investigated with the factor exponent, the combination of the by-verb factor smooths and exponent bring under statistical control all parts of the word forms that potentially co-determine articulation.

Table 2
Summary of partial effects of a linear mixed-effects models, fitting logged vowel duration as a function of morphological condition and number of segments in the base. Absolute t-values larger than 2 are considered to indicate significant partial effects.

Predictor	Vowel	Estimate	Std. Error	t-value
(Intercept)	[iː]	-0.81	0.04	-19.81
Tense: past	[iː]	-0.05	0.01	-4.19
Tense: present	[iː]	-0.05	0.01	-4.40
Tense: progressive	[iː]	-0.20	0.01	-16.91
Num. Segments Base	[iː]	-0.04	0.01	-6.08
Activation	[iː]	-0.00	0.00	-1.01
(Intercept)	[ɑː]	-0.88	0.02	-40.15
Tense: past	[ɑː]	-0.02	0.02	-0.83
Tense: present	[ɑː]	-0.02	0.02	-1.05
Tense: progressive	[ɑː]	-0.12	0.02	-6.79
Num. Segments Base	[ɑː]	-0.07	0.02	-4.21
Activation	[ɑː]	0.00	0.01	0.49

3 Analysis and Results

3.1 Duration effects

Although we kept the effects of vowel duration under statistical control, it is still possible that any effect of practice in the current findings emerged due to significantly shorter vowel durations across activations (cf. Aylett & Turk, 2004; B. Munson & Solomon, 2004; Priva, 2015, for effects of frequency of occurrences or contextual predictability). To rule out this possibility, we first analyzed effects of practice and the inflectional exponent onto vowel duration.

We used a linear mixed-effects regression model (package ”lmer” for R, Version 1.1-15, Bates, Maechler, Bolker, and Walker (2014)). We modeled a functional relation between vowel duration and exponent, activations, and, the number of canonical phones in the stem to to bring compensatory lengthening due to syllable structure under statistical control (cf. Altmann, 1980). We did not use the number of segments in the whole word, since it is collinear with morphological condition. We used speakers, verb, and places of articulation of surrounding consonants as random intercepts. The inclusion was supported by an F-test.

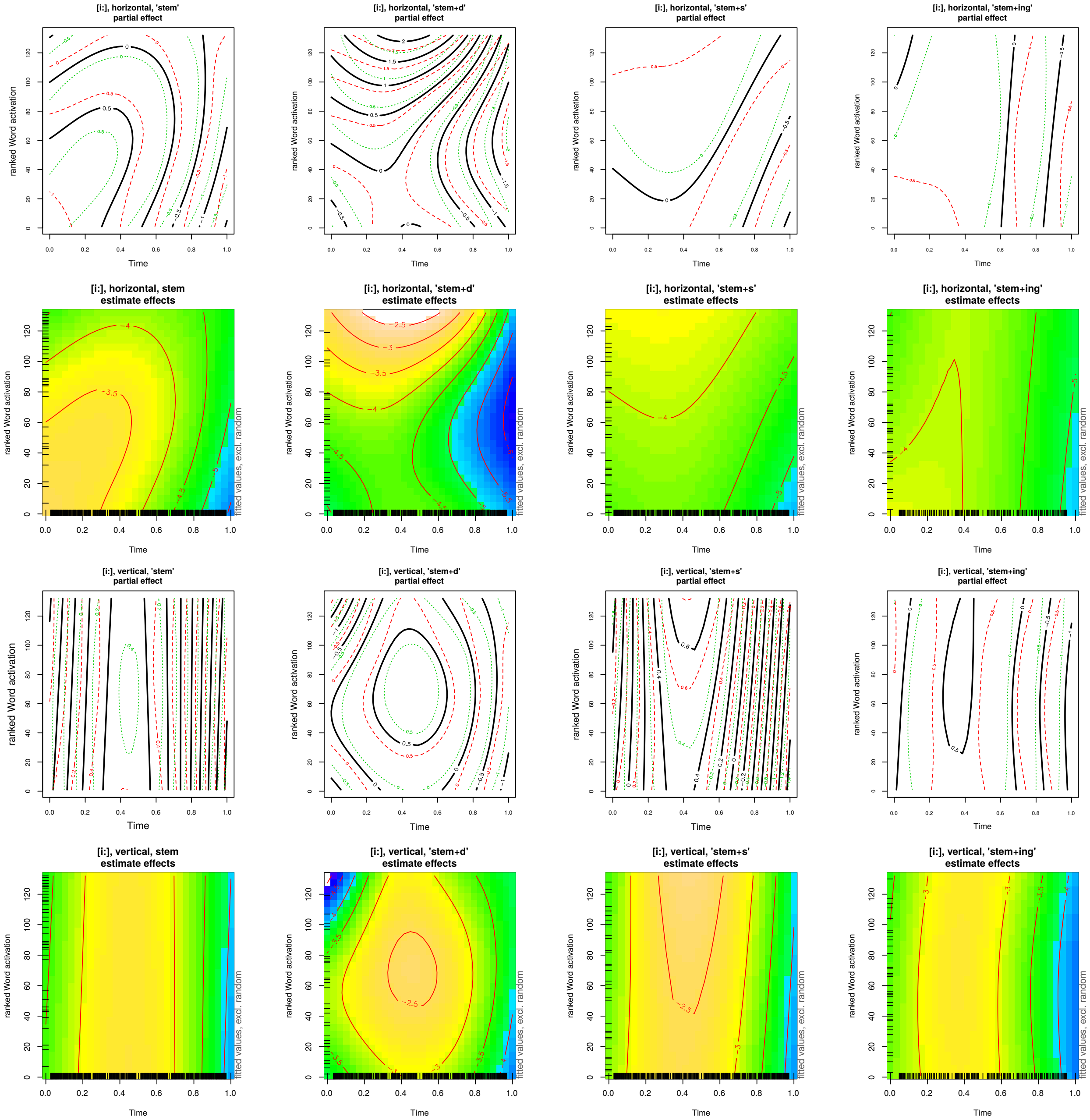
Estimates are summarized in Table 2. While [iː] vowels were significantly shorter in monosyllabic suffixed verbs, no difference between these categories was found for the vowel [ɑː]. In both vowels, vowel durations were significantly shorter in the progressive, i.e. disyllabic, condition. In addition, vowel durations were negatively correlated with the number of segments in the stem. Given that word activations were not significantly correlated with duration in either of the vowels, we regarded all modulations of the articulatory trajectory across activation as a result of activation and not of varying vowel durations.

3.2 Articulatory trajectories

Figures 3 and 4 show the results for horizontal (first & second row) and vertical (third & fourth row) tongue body movements (z-axis) across time (x-axis), modulated by ranked word activation (y-axis) for the [iː] and [ɑː] vowels. Vowel articulation in *stem* verb forms is illustrated in the left most column, *stem+d* in the second column to the left, *stem+s* in the second column to the right, *stem+ing* in the furthest right column. Contour plots in the first and third row illustrate the model's partial effects. Contour lines represent tongue positions. Standard errors colored in green dotted lines indicate an increase of positional values and colored in red dotted lines indicate a decrease of positional values. The second and fourth row illustrate the the estimated tongue movements by means of heat maps, with yellow colors indicating further front / higher positions and green and blue colors indicating further back / lower tongue positions. Ticks at the left border represent the distribution of the data along the activation continuum. In order to trace the vertical/horizontal tongue movement across time, one needs observe how many contour lines an imaginary horizontal line crosses when going from left to right. In the description of the results we will focus on the heat maps, while contour plots serve as a means to evaluate to what extent the change in positional values due to the *Time*Activation* interaction is significant. If word activations had no effect at all, all contour lines would be straight vertical lines. Model comparisons (not presented) testing the present models against models without word activation provided strong support for the relevance of this predictor.

3.2.1 [iː] as stem vowels.

Figure 3. *i TB*



During the production of [iː], we expected that stronger anticipatory movements associated with greater word activations be mirrored by stronger retraction and stronger lowering in inflected verb forms than in the [stem] verb forms. In the following, we will first describe tongue movements in the horizontal, then in the vertical axis. We find a global movement pattern in the horizontal axis to be best described by an inversely u-shaped trajectory across all verb forms (Figure 3, second row): Tongue positions are retracted at the onset, fronted in the center of the vowel and retracted at the offset of the vowel.

In *stem* verb forms, this pattern is modulated by word activations insofar as tongue movements become strongly reduced in words with greater activations: tongue positions at the onset and at the center are more strongly retracted, resulting in tongue movements having smaller movement amplitudes.

This modulation pattern across activation is reversed when verbs containing [iː] have [d] and [d] as inflectional exponents: Higher word activations go hand in hand with with more fronted tongue positions at the onset, in the center, and the offset of the vowel across the entire vowel. Two differences become apparent. First, the size of the effect is smaller in *stem+s* than in *stem+d* verb forms. Second, tongue body positions are more fronted in *stem+d* than in *stem+s* word forms, the greater the practice.

Higher word activations are thus associated with less anticipatory retraction toward the inflectional exponents [d] and [s] in both verb forms. Clearly, these findings are contrary to our predictions to observe stronger anticipatory coarticulation with higher word activations, which should have been reflected by stronger retraction (cf. Figure 1). With higher word activations reflecting more practice with individual word forms, we rather observe anti-coarticulation and a more pronounced articulation of the [iː] vowel in better practiced word forms.

By contrast, higher word activation in the *stem+ing* verb forms is reflected by stronger retraction which goes hand in hand with shallower amplitude of the tongue movements, similar to the *stem* verb forms. This effect could be regarded to indicate anticipatory retraction in [iː] towards the further back located [ɪ] in [ɪŋ].

Turning to the vertical tongue body movements in [iː], the global movement pattern is also characterized by an inverse u-shaped movement trajectory, i.e. one of rising and lowering (Figure 3, bottom row). In the *stem* verb forms, no significant modulations by word activation can be observed.

In the *stem+d* verb forms, the modulations by word activations result in an inversely u-shaped effect across the entire continuum, with a maximum tongue body rising during the vocalic center in words at mid activations. We furthermore observe a shift in tongue positions at the edges, with greater word activations associated with lower tongue positions at the onset, and higher at the offset. We thus observe less coarticulation at the offset of the vowel, the better practiced words become, which is similar to the effects on the horizontal axis. It is possible that the inversely u-shaped pattern across word activation is a result of the extreme fronting in the horizontal axis, with the tongue body being pushed down in more extreme positions due to the curvature of the palate. Note that we observe similar u-shaped effects due to practice in German second person plural verbs containing the vowel [aː] (e.g. 'bahnt', [baːnt]) (Tomaschek et al., under revision 2).

In the *stem+s* verb forms, higher word activations are associated with no effect at the onset of the vowel, but with higher tongue positions at the center and the offset of the vowel. This effect mirrors the anti-coarticulation in better practiced words already attested in the horizontal axis.

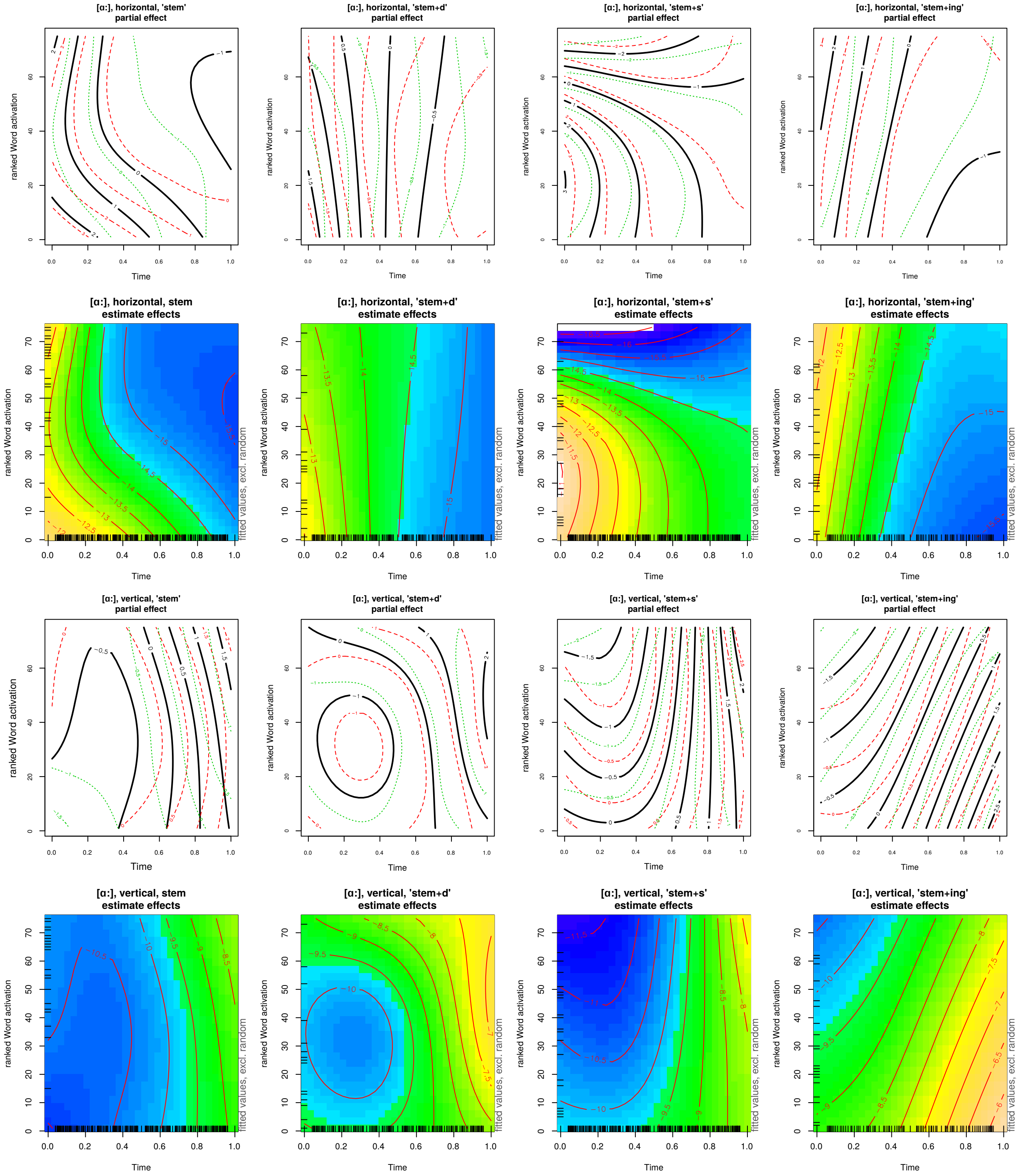
The pattern in the *stem+ing* verb forms are very similar to those in the *stem* verb form. However, we observe a minimal manipulation by word activations which results in a shift of the maximum deflection in the vowel center towards the offset that goes hand in hand with less lowering at the offset. In other words, the tongue body anticipates the lower [ɪ] less.

In summary, we observe less anticipatory coarticulation in better practiced words, especially at the offset of the vowel. This is contrary to our expectations that more practice would lead to stronger coarticulation between the stem vowel and the inflectional exponent (Sosnik et al., 2004). Rather, it seems as though vowels in the inflected stems are hyperarticulated in better practiced words. In the following, we examine whether these effects of hyperarticulation replicate during the articulation of [ɑː].

3.2.2 [ɑː] as stem vowels.

We expected that stronger anticipatory coarticulation in association with greater practice should be reflected by stronger fronting and rising in the *stem+d* and *stem+ing*

Figure 4. *A TB*



word forms than the [stem] verb forms. In the *stem+s* word forms, we further expected to observe effects of practice only in the vertical axis, as there is no difference in the mean vertical tongue body position between [ɑ] and [s] (cf. Figure 1). The global tongue body movement pattern during [ɑː] production in the horizontal axis is characterized by constant retraction across the entire vowel in all verb forms (Figure 4, second row). Horizontal tongue movements in the *stem* verb forms show strong changes across word activations only below word activations of 30, but little to no manipulation above activations of 30. The main cause of the variation in the *stem* verb forms is probably due to the outlier located between rank 10 and 15. Excluding this point from our considerations, we conclude that no changes across word activations can be found for horizontal tongue movements in the *stem* forms.

By contrast, we observe significant changes in the movement trajectory across word activations in the *stem+d* verb forms. Greater activations are associated with less fronted tongue positions towards the onset and further fronted positions towards the offset of the vowel. In other words, horizontal tongue body movements in better practiced words show a shallower retraction movement. A similar effect can be observed in *stem+s* verb forms, but with stronger retraction in words with higher activations. Like for horizontal tongue body movements during [iː], vowel articulations become more distinct with greater practice. In addition, we observe less anticipatory fronting of the inflectional exponents during [ɑː] production.

Similar to articulations of [iː], the modulation pattern across word activations is reversed in the *stem+ing* verb forms, where higher word activations are associated with over all more fronted articulations, resulting in stronger fronting of [ɑː] towards the front [ɪ].

Turning to the vertical tongue body movements during [ɑː] articulation, the global pattern is characterized by a rising movement (Figure 4, bottom row). In the *stem* verb form, this pattern interacts with word activations insofar that tongue positions become higher with increasing word activations. Similar to the effect of word activations along the horizontal axis during [iː] productions, this indicates stronger centralization of the tongue, i.e. stronger reduction, in better practiced words.

The pattern across word activations is similar in *stem+d* verb forms, but with a much stronger effect in words with higher word activations insofar as tongue body positions are high across the entire vowel. The effect is so strong that, while we observe a strong raising pattern during [ɑː] in words with low activations, movements become very shallow in words with high activations. In other words, we observe strong anticipatory raising of [ɑː] towards the higher [d] in better practiced words.

By contrast, the pattern is reversed for the *stem+s* verb forms. Two effects become apparent, the higher word activation becomes: We observe lower tongue body positions in the onset and higher tongue body positions in the offset of the vowel. As a consequence, the amplitude of the movement becomes larger. Recall that we did not predict effect of learning in the vertical axis for the *stem+s* word forms. Rather, we observe larger tongue body movements in better practiced words.

Finally, in the *stem+ing* verb forms, higher word activations are associated with lower tongue positions across the entire vowel. In other words, we observe less anticipatory raising of [ɑː] toward the higher [ɪ], the better practiced a word is.

4 Summary and conclusion

There is a growing number of studies showing that that the durational characteristics of words and segments vary depending on their grammatical and morphological function (Drager, 2011; Kemps et al., 2005a; Lohmann, 2017; I. Plag et al., 2017; Seyfarth et al., 2017). Here, we sought to see whether tongue movements in stem vowels vary due to the morphological function of the verb, they are located in. Concretely, we investigated how inflectional exponents in words such as ’cleaned’, ’cleans’, ’cleaning’ changed the articulatory pattern in contrast to the bare stem ’clean’ (Magen, 1997; Öhman, 1966; Sziga, 1992).

In line with effects of practice observed in kinematic studies of arm movements as well as in articulation, we expected to observe stronger anticipatory coarticulation of inflectional exponents in better practiced words (Sosnik et al., 2004; Tiede, Mooshammer, Goldstein, Shattuck-Hufnagel, & Perkell, 2011) due to stronger overlap between the vocalic gesture and the gesture responsible for the inflectional exponent (e.g. Browman & Goldstein, 1986; Fowler & Saltzman, 1993; Liberman & Mattingly, 1985; Saltzman & Munhall, 1989). However, this pattern was found only in 3 out of the 12 possible environments (2 vowels * 2 axis * 3 inflected verb forms). We mostly observed *anti-coarticulation*, i.e. the vowel was less coarticulated with the upcoming inflectional exponent, the better practiced a word was.

The findings can be summarized in two obervations: vowels in better practiced words were either produced with more extreme tongue movements or at more peripheral tongue positions, i.e. further away from the vowel space center. These findings are at odds with observations that vowels tend to be phonetically reduced in words with a higher frequency of occurrence, which usually is interpreted as a result of lower informativity in studies investigating vowel productions in corpora (Aylett & Turk, 2006; Meunier & Espesser, 2011; Priva, 2015). By contrast, the current findings replicate results observed in experimental studies in which words were articulated in lists (Benjamin Munson, 2001; Tomaschek et al., under revision 2; Tomaschek et al., 2014; Tomaschek et al., 2013). It remains for future research to investigate the source of this difference. However, one potential explanation is that contextual predictability is usually very high and spreads across multiple words in spoken language, increasing the probability of a word and therefore reducing its own informativity. Contextual predictability is not present in list readings, probably allowing the practice effect to emerge. Given the direction of the direction of the effects in the each of the vowel-morphological category combination in the present study, it rather seems as though greater practice results in larger differences between the different morphological categories of the verbs, mirroring the morphological function of the verbs. This challenges predictions by Levelt et al. (1999)'s model of lexical access, according to which phonetic spell out is independent from the morphological information. However, these findings are in line with a growing body of studies showing fine phonetic changes related to higher linguistic structures, especially morphology (Cho, 2001; Ingo Plag, Homann, & Kunter, 2015; Smith, Baker, & Hawkins, 2012). Correlations between morphological structures and the phonetic signal can be interpreted as practice with word forms rather than morphological function themselves, which is why findings in speech production are not a precursor to the assumption that such phonetic detail related to morphology has a lexical representation. However, there is also a growing body of studies showing that listeners are sensitive to these fine spectral and durational cues, attributing them to morphological functions (Balling & R. Baayen, 2008; Blazej & Cohen-Goldberg, 2015; Davis, Marslen-Wilson, & Gaskell, 20002; Kemps et al., 2005b; Kemps et al., 2005a).

Our study therefore shows that speakers acquired greater articulatory practice and this practice becomes part of their mental lexicon, which allows both, hyper-articulation of vowels and simultaneous production of stronger differences between morphological categories. Consequently, like other psychological behavior, speech production is submitted to ongoing fine tuning and mirrors the dynamics of life-long learning (Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014).

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 and Kiki Callihan for their helpful comments and suggestions. We are furthermore thankful to Theresa Schmitt for her help with the experiment.

References

Altmann, G. (1980). Prolegomena to menzerath's law. Glottometrika, 2, 1–10.

Arnold, D., Tomaschek, F., Sering, K., Ramscar, M., & Baayen, R. H. (2017). 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. PLOS ONE.

Aylett, M. & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Language and Speech, 47(1), 31–56.

Aylett, M. & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. Journal of the Acoustical Society of America, 119(5), 3048ff.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. Psychological review, 118(3), 438–481.

Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: A proof of concept with naive discriminative learning. Language, Cognition and Neuroscience, 31(1), 106–128.

Baayen, R. H., Vasishth, S., Bates, D., & Kliegl, R. (2017). The cave of shadows. addressing the human factor with generalized additive mixed models. Journal of Memory and Language, 94, 206–234.

Balling, L. W. & R. Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from danish. Language and Cognitive Processes, 23(7-8), 1159–1190. doi:10.1080/01690960802201010. eprint: http://dx.doi.org/10.1080/01690960802201010

Balota, D., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., . . . Treiman, R. (2011). The english lexicon project. Behavior Research Methods, 39, 445–459.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4. Retrieved from http://CRAN.R-project.org/package=lme4

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational english. Journal of Memory and Language, 60(1), 92–111.

Bell-Berti, F. & Harris, K. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. Journal of the Acoustical Society of America, 65(5), 1268–1270.

Bell-Berti, F. & Harris, K. (1982). Temporal patterns of coarticulation: Lip rounding. Journal of the Acoustical Society of America, 71(2), 449–454.

Blazej, L. J. & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? Journal of experimental psychology. Human perception and performance, 41(1), 50–68.

Browman, C. & Goldstein, L. (1986). Towards an articulatory phonology. Phonology, 3, 219–252.

Caselli, N. K., K., C. M., & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. The Quarterly Journal of Experimental Psychology, 69(3), 432–454. doi:10.1080/17470218.2015.1054847

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

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (20002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance, 28(1), 218–244.

Drager, K. K. (2011). Sociophonetic variation and the lemma. Journal of Phonetics, 39(4), 694–707. Cross-Language Speech Perception and Variations in Linguistics Experience.

Fasiolo, M., Goude, Y., Nedellec, R., & Wood, S. (2017). Fast calibrated additive quantile regression. Manuscript, University of Bristol. Retrieved from https://github.com/mfasiolo/qgam

Fowler, C. & Brancazio, L. (2000). Coarticulation resistance of american english consonants and its effects on transconsonantal vowel-to-vowel coarticulation. Language and Speech, 42(1), 1–41.

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

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

Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The breadth of coarticulatory units in children and adults. Journal of Speech, Language, and Hearing Research, 51(6), 1424–1437.

Hastie, T. & Tibshirani, R. (1990). Generalized Additive Models. London: Chapman & Hall.

Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1993). A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. Language and Speech, 36(2,3), 235–260.

Katz, W. & Bharadway, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: A preliminary report. Clinical linguistics and phonetics, 15(1), 139–143.

Kemps, R. J., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005b). Prosodic cues for morphological complexity: The case of dutch plural nouns. Memory & Cognition, 33(3), 430–446.

Kemps, R. J., Wurm, L. H., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005a). Prosodic cues for morphological complexity in Dutch and English. Language and Cognitive Processes, 20(1/2), 43–73.

Koenker, R. (2005). Quantile regression. Cambridge University Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. The Behavioral and brain sciences, 22(1).

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1–36.

Linke, M., Bröker, F., Ramscar, M., & Baayen, R. H. (2017). Are baboons learning "orthographic" representations? probably not. PLOS ONE, 12(8), 1–14.

Lohmann, A. (2017). Cut(n) and cut(v) are not homophones – lemma frequency affects the duration of noun-verb conversion pairs. Journal of Linguistics, (55).

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in english. Journal of Phonetics, (25), 187–205.

Meunier, C. & Espesser, R. (2011). Vowel reduction in conversational speech in french: The role of lexical factors. Journal of Phonetics, 39(3), 271–278. Speech Reduction.

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. PLOS ONE.

Munson, B. [B.] & Solomon, N. (2004). The effect of phonological neighborhood density on vowel articulation. Journal of speech and hearing research, 47, 1048–1058.

Munson, B. [Benjamin]. (2001). Phonological pattern frequency and speech production in adults and children. Journal of Speech, Language, and Hearing Research, 44(4), 778–792. doi:10.1044/1092-4388(2001/061

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

Plag, I. [I.], Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. Journal of Linguistics, 53(1), 181–216.

Plag, I. [Ingo], Homann, J., & Kunter, G. (2015). Homophony and morphology: The acoustics of word-final S in English. Journal of Linguistics, (doi:10.1017/S0022226715000183). Retrieved from %5Curl%7BS0022226715000183%7D

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

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken dutch. The Journal of the Acoustical Society of America, 118(4), 2561–2569. doi:http://dx.doi.org/10.1121/1.2011150

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

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org

Raeder, C., Fernandez-Fernandez, J., & Ferrauti, A. (2015). Effects of six weeks of medicine ball training on throwing velocity, throwing precision, and isokinetic strength of shoulder rotators in female handball players. J Strength Cond Res. 29(7), 1904–14.

Ramscar, M., Dye, M., & McCauley, S. (2013). Error and expectation in language learning: The curious absence of 'mouses' in adult speech. Language, 89(4), 760–793.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, R. H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. Topics in Cognitive Science, 6(1), 5–42.

Ramscar, M. & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. Cognitive Science, 31(6), 927–960.

Rapp, S. (1995). Automatic phonemic transcription and linguistic annotation from known text with hidden markov models / an aligner for german. In Proceedings of elsnet goes east and imacs workshop. Moscow.

Recasens, D. (1984). Vowel-to-vowel coarticulation in catalan vcv sequences. The Journal of the Acoustical Society of America, 76(6), 1624–1635. doi:http://dx.doi.org/10.1121/1.391609

Rescorla, R. & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. Prokasy (Eds.), Classical conditioning ii: Current research and theory (pp. 64–69). Appleton Century Crofts, New York.

Roelofs, A. (1997). The {weaver} model of word-form encoding in speech production. Cognition, 64(3), 249–284.

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

Schmidtke, D., Matsuki, K., & Kuperman, V. (2017). Surviving blind decomposition: A distributional analysis of the time course of complex word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. Language, Cognition and Neuroscience, 0(0), 1–18.

Smith, R., Baker, R., & Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. Journal of Phonetics, 40(5), 689–705. doi:\url{10.1016/j.wocn.2012.04.002}

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

Sziga, E. C. (1992). Acoustic evidence for gestural overlap in consonant sequences. Haskins Laboratories Status Report on Speech Research, SR-111/112, 43–62.

Tiede, M., Mooshammer, C., Goldstein, L., Shattuck-Hufnagel, S., & Perkell, J. (2011). Motor learning of articulator trajectories in production of novel utterances. In Proceedings of the icphs xvii. Hong Kong.

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

Tomaschek, F., Tucker, B. V., Wieling, M., & Baayen, R. H. (2014). Vowel articulation affected by word frequency. In Proceedings of the 10th issp (pp. 425–428). Cologne.

Tomaschek, F., Wieling, M., Arnold, D., & Baayen, R. H. (2013). Word frequency, vowel length and vowel quality in speech production: An ema study of the importance of experience. In Proceedings of the interspeech. Lyon.

Tremblay, A. & Tucker, B. V. (2011). The effects of n-gram probabilistic measures on the recognition and production of four-word sequences. The Mental Lexicon, 6(2), 302–324.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2015). itsadug: Interpreting time series, autocorrelated data using gamms. R package version 0.8.

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. The Journal of the Acoustical Society of America, 90(4), 2311–2311. doi:http://dx.doi.org/10.1121/1.401072

Wood, S. N. (2006). Generalized additive models: An introduction with r. Boca Raton, Florida, U. S. A: Chapman and Hall/CRC.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B), 73, 3–36.

Wood, S. N. (2013a). A simple test for random effects in regression models. Biometrika, 100, 1005–1010.

Wood, S. N. (2013b). On p-values for smooth components of an extended generalized additive model. Biometrika, 100, 221–228.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. Memory & Cognition, 7(6), 411–419. doi:10.3758/BF03198257

Zipf, G. (1935). The psycho-biology of language. an introduction to dynamic philology. Cambridge: MIT Press.

Appendix: Summary tables

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-4.0814	0.3848	-10.6072	< 0.0001
Tensepast	-0.3289	0.0684	-4.8057	< 0.0001
Tensepresent2	-0.1831	0.0588	-3.1148	0.0018
Tenseprogressive	-0.3010	0.0806	-3.7350	0.0002
B. smooth terms	edf	Ref.df	F-value	p-value
te(Time.normSegment,R.Act):Tensepresent1	7.7745	9.2231	78.8702	< 0.0001
te(Time.normSegment,R.Act):Tensepast	9.1314	10.1966	156.0843	< 0.0001
te(Time.normSegment,R.Act):Tensepresent2	5.0870	5.8386	37.3512	< 0.0001
te(Time.normSegment,R.Act):Tenseprogressive	5.2329	5.9701	24.7807	0.0004
ti(Time.normSegment,cL.SegmentDuration)	2.6378	2.8926	42.0214	< 0.0001
s(cL.SegmentDuration)	1.9309	1.9949	17.7956	0.0001
s(Participant)	23.8835	24.0000	5054.0799	< 0.0001
s(Time.normSegment,Base)	93.5608	107.0000	6973.3834	< 0.0001

Table 3

Summary table for paradigmatic and non-paradigmatic effects in the GAMM fitting ***vertical*** *tongue body movements in [iː].*

[1] "