How is anticipatory coarticulation of suffixes affected by lexical proficiency?

Abstract

More and more studies find 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 of inflectional exponents and the amount of long-term practice with individual verbs such as American English "clean", "cleaned", "cleans", "cleaning". Kinematic studies of hand movements show that with greater practice, i.e. regular repetition of a sequence of gestures, upcoming gestures are stronger and smoother anticipated. Consequently, we hypothesized to find stronger anticipatory coarticulation of inflectional exponents during the articulation of the stem vowel in verbs for which speakers acquired a greater lexical proficiency, as their articulatory gestures were better practiced. We observed both, stronger anticipatory coarticulation towards the offset of the gesture and less coarticulation concomittant with more hyperarticulation towards the onset of the gesture. 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, proficiency

1 Introduction

1.1 Background

According to one of the most influential models of speech production – the theory of lexical access by Levelt et al. (1999) – speech production is a modular, sequential process, during which concepts activate lemmata in a linguistic lexicon. These access in turn abstract building blocks in the form of morphemes. After merging the morphemes, the resulting new units are recoded during a postlexical process, first into a phonological form, then into a syllabic form which drives motor programs for articulation. This framework assumes that most stages of speech production happen at an abstract, symbolic level. It also assumes that processing is modular, such that symbols at early levels do not co-determine the details of articulatory spell-out, which is why a word's lexical information should not be reflected by articulation processes.

However, a growing number of studies challenges assumptions about an encapsulated modular speech production process by showing that fine phonetic details – articulatory and acoustic – do correlate with higher-level lexical properties, especially morphological information. Cho (2001), for example, 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. Drager (2011) and Podlubny et al. (2015) reported that segment durations in the English word 'like' depend on its different grammatical functions. Lee-Kim et al. (2012) showed that the "darkness" of English [l] depends on the morphological status of its position. Finally, Plag et al. (2017) and Seyfarth et al. (2017) found that the acoustic duration of homophonous [s] and [z] in American English depend on their morphological function (see Tomaschek et al., 2019b, for a replication).

Furthermore, grammatical information interacts with frequency of occurrence, further modulating fine phonetic details of the speech signal. 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. Caselli et al. (2016) showed 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'). Pluymaekers et al. (2005) found that the acoustic duration of affixes were co-determined by the frequency of occurrence of the carrier word.

1.2 The present study

In the present study, we further examined the amount of fine phonetic detail related to morphological categories. We focused on 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.

Our first hypothesis concerns vowel articulation independently of the morphological status of the verb it is located in. It is widely accepted that vowels in more frequent words exhibit shorter durations and more centralized positions in the formant space (Aylett and Turk, 2006; Meunier and Espesser, 2011). In an information theoretic framework (Shannon, 1948), these reductions have been interpreted to go hand in hand with a reduction of information density in the speech signal (see also Cohen Priva (2015) for similar findings in consonants). Given this line of reasoning, we expected to find stronger centralized vowel articulations associated with a higher frequency of occurrence.

Our second hypothesis concerned the inflectional exponent following the stem (i.e. [ø] vs. [d] vs. [s] vs. [ɪŋ]), or in other words the morphological status of the verb. Given that spectral and temporal characteristics of phones vary due to coarticulation with following phone (Öhman, 1966; Bell-Berti and Harris, 1979, 1982; Recasens, 1984; Hoole et al., 1993; Fowler and Brancazio, 2000; Katz and Bharadway, 2001; Goffman et al., 2008) and that coarticulation occurs between directly neighboring phones and also across intervening phones and even syllables (Magen, 1997; Sziga, 1992), we expected articulatory patterns to be modulated by upcoming inflectional exponents. The extent of anticipation should vary with the amount of practice with the words. We will elaborate this in the following paragraphs.

Changes in kinematic skills can be attributed to two kinds of processes: long-term practice across weeks and months, which is typically attested for motor skills required in sports; and short-term practice which is usually the case in kinematic studies, in which single kinematic gestures are practiced within one or several experimental sessions. Kinematic studies of hand movements have repeatedly shown improvements of motor skills due to short-term practice, revealing shorter movements (Raeder et al., 2015; Platz et al., 1998) or gestures that overlap to a stronger degree (Sosnik et al., 2004).

In language, long-term practice with individual verbs can be operationalized by their

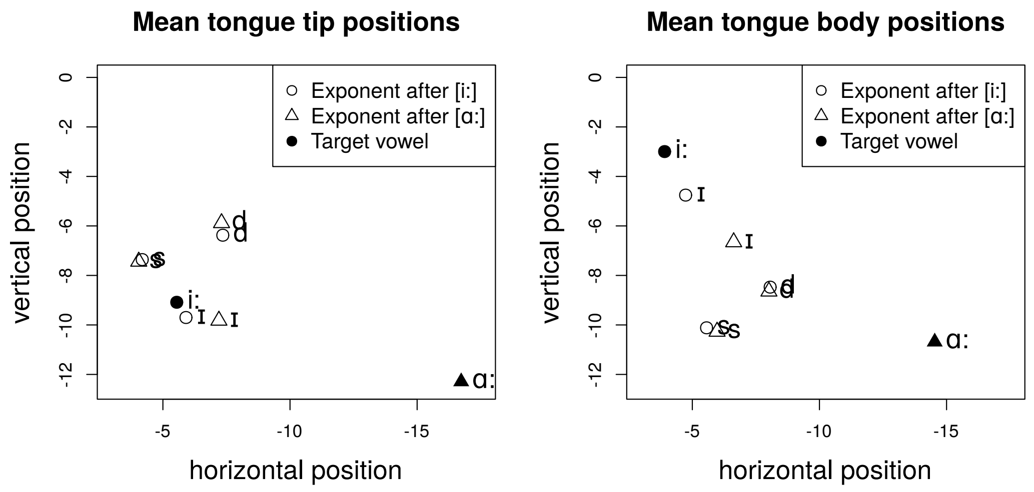

Figure 1: *Mean sensor positions of the tongue tip (left) and tongue body (right) for segments of interest pooled across all speakers and words, calculated at the center of the segments. Segments in words with an [iː] are represented by a circle, segments in words with an [ɑː] are represented by a triangle. Empty circles and triangles represent the exponents, full circles and triangles represent the vowels under investigation.*

frequency of occurrence, and indeed, studies that have investigated its effects on language found shorter durations for more frequent words (Wright, 1979; Whalen, 1991; Bell et al., 2009), which in turn have been attributed to smoother gestural execution (Tiede et al., 2011).

From this perspective, and given the results of Tomaschek et al. (2013, 2014, 2018c), we expected to find stronger anticipatory coarticulation of the inflectional exponents in the stem vowel for better practiced words.

We regard anticipatory coarticulation to reflect kinematic optimization of current gestures for upcoming gestures (Saltzman and Munhall, 1989). As a consequence, we expected anticipatory coarticulation to depend on the spatial constellation of the articulatory target positions of the vowel under investigation ([ɑː] or [iː]) and those of the inflectional exponent. In the following, we parameterized target positions as the mean positions at the center of the pertinent segments.

Figure 1 shows mean horizontal/vertical position of the tongue tip and tongue body sensor at the center of the segments in our data, estimated with linear mixed-effects models, fitting vertical/horizontal tongue position with segment as a fixed factor and random intercepts for speaker and word. Segments in words with an [iː] are represented by a circle, segments in words with an [ɑː] are represented by a triangle. Empty circles and triangles represent the exponents, full circles and triangles represent the vowels under investigation.

We expected to observe anticipatory coarticulation of the [s, d, ɪŋ] exponents during the [iː] and [ɑː] vowel in suffixed verbs in contrast to unsuffixed verbs. Taking into account the relative mean tongue position between the vowels and the inflectional exponents in Figure 1, tongue tip should have higher positions in [iː] when verbs are inflected with [s] and [d], but lower positions when verbs are inflected with [ɪŋ] due to the slightly lower position of [ɪ]. Furthermore, tongue tip positions should be further front, when verbs are inflected with [s] and further back otherwise.

As for the tongue tip articulation during [ɑː], stronger coarticulation of the exponents should be reflected by further front and higher tongue tip positions in inflected verbs. In case of the tongue body, we expect tongue positions to be higher and further front when [ɑː] verbs are inflected with [d] and [ɪŋ] in contrast to the bare stem, while preceding [s] should show only stronger fronting. No changes in the vertical positions should occur because tongue body height is very similar in [ɑː] and [s].

Given the effects of practice on kinematic skills, we expected to observe stronger anticipatory coarticulation in verbs with which speakers have greater experience than in verbs with which speakers have less experience.

2 Methods

Table 1: The table presents the number of words for each vowel and morphological condition and the total number of verbs presented for each vowel.

	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

2.1 Recording

25 speakers of Canadian and American English (mean age: 29.4, sd: 8.2) were paid to articulate the stimuli. Word lists were structured according to a latin square design such 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 dimension.

2.2 Preprocessing

Tongue movements were corrected for head-movements in an online procedure during recording by the 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. Absolute sensor positions were transformed into distances between the sensor and its maximal vertical/horizontal position for each sensor in each speaker. More negative values represent stronger retraction in the horizontal dimension and stronger lowering in the vertical dimension. Segment boundaries were determined by automatically aligning the audio signal with phonetic transcriptions by means of a Hidden-Markov-Model-based forced aligner for English. Alignments for the vowel were manually verified and corrected where necessary.

2.3 Word materials

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

In order to obtain a sufficiently large number of verbs, our materials contained both rime structures, VC and VCC. It is important to note that for all words in the [ɑː] category, the vowel was followed by the retroflex [ɹ]. Nevertheless, we expected coarticulation to occur across both single and doubly filled codas, given the preceding literature (Sziga, 1992; Magen, 1997).

Table 1 shows the number of words for every vowel across the four morphological conditions, in addition to examples for the two stem vowels that we investigated.

2.4 Operationalization of linguistic practice

Ever since Zipf (1935) found that frequency of occurrence is inversely proportional to the average number of phones, the measure has been widely used to assess a speaker's experience with a given word (e.g. Whalen, 1991; Munson and Solomon, 2004; Aylett and Turk, 2006; Gahl, 2008; Meunier and Espesser, 2011; Tomaschek et al., 2018c,a). From a kinematic perspective, word frequency can be regarded to gauge the amount of long-term practice a speaker has with a given word (Tomaschek et al., 2018c,a). From an information theoretic perspective (Shannon, 1948), word frequency can be regarded as a measure of a-priori expectancy of a word independently of its context. Consequently, word frequency measures the occurrence of a word form independently of other word forms. However, it has been shown that the amount of sublexical similarity between words (i.e. at the segment level), which can be gauged by Phonological Neighborhood Density, correlates with the acoustic characteristics of words (Scarborough, 2003; Munson and Solomon, 2004; Gahl and Strand, 2016). Of course it would be possible to combine use these two measures in an analysis. Unfortunately, they are strongly proportional to each other, causing problems of interpretability in the statistical analysis (Tomaschek et al., 2018b).

Word ACTIVATION, a measure derived from weights in a input-output network calculated by the Naïve Discriminative Learner (package NDL, Version 0.2.17, Baayen et al. (2011)) in the statistical programming language R (R version 3.3.3 (2017-03-06), Team (2014)), provides a measure that combines word frequency and phonological neighborhood density.

NDL is based on discriminative learning (Rescorla and Wagner, 1972), an error-driven supervised learning algorithm, according to which speakers/listeners learn to discriminate between lexical classes from sublexical input (Ramscar and Yarlett, 2007; Ramscar et al., 2013). NDL represents a simple two-layer cue-to-outcome network and has been shown to predict lexical decision latencies (Baayen et al., 2011; Milin et al., 2017), acoustic durations of word final [s] depending on the discriminability of its morphological function (Tomaschek et al., 2019a), learning sublexical representations of connected speech (Baayen et al., 2016), and the acquisition of regular and irregular plural forms (Ramscar et al., 2013).

Whereas the mapping of cues onto outcomes is solely driven by the Rescorla-Wagner learning equation (Rescorla and Wagner, 1972), NDL allows to investigate different levels of speech production by varying the cue and the outcome structure. For example, Baayen et al. (2011) trained the network to discriminate lexomes, i.e. word forms that we regard to be pointers to a higher dimensional semantic space (Milin et al., 2017), by means of letter bigrams. Baayen et al. (2016) used diphones to discriminate lexomes from a moving window spanning several words. Recently, Arnold et al. (2017) has successfully achieved lexical discrimination from acoustic cues automatically derived from spontaneous speech. Linke et al. (2017) was able to predict lexical discrimination from visual cues automatically derived from letter strings.

Baayen et al. (2011) used letter bigrams as form cues, as in their study participants had to read out loud target words on a screen. Likewise, participants in the present study had to read out loud a word on a screen (see description of the experiment below). This is why we decided to model practice using a form-to-lexome network which was trained to discriminate lexomic word forms on the basis of triphone cues (following Baayen et al. (2016)). We regard the triphone cues as acoustic targets which speakers try to meet during articulation. Word form outcomes represent pointers to the semantics of the word.

To train the network, we used all verbs from the English Lexicon Project (ELP, Balota et al., 2011). 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.

On the basis of the trained triphone-to-lexome network, we calculated ACTIVATIONs for each of the verb forms used in the present experiment by summing up the weights between a lexome and all its afferent triphone cues (e.g. `#wA wAk Aks ks#` to 'walks'). ACTIVATIONS represent the bottom-up support for that lexome given the cues in the input. While it is positively correlated with word frequency (R = 0.86 for [iː] verbs, R = 0.81 for [ɑː] verbs), this measure differs from word frequency by emerging through the competition of phones for an outcome (e.g. `#wA` for 'walks' and 'walked'). This is why ACTIVATION also captures sublexical relations among words and therefore renders the use of Phonological Neighborhood Density superfluous. Finally, ACTIVATION ignores all semantics connotations that come along with a word and which are potentially co-measured by word frequency. As ACTIVATION emerges

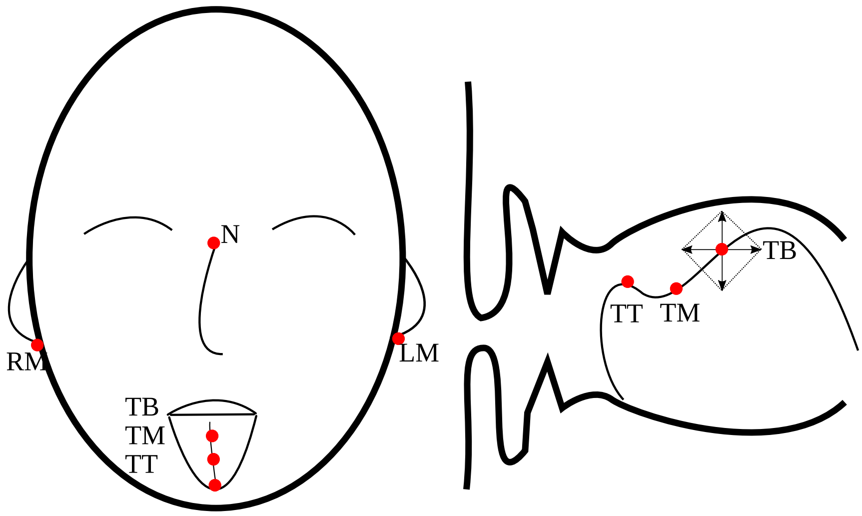

Figure 2: *Illustration of sensor positions. Left: frontal illustration. Right: midsagittal cut through the mouth. The rhombus around the tongue body sensor illustrates the parameterization of its movement area.*

through the diphones and their frequencies, filtered through discriminative learning, we consider it to be a measure of practice with individual word forms and their articulatory gestures independently of other influences[1].

2.5 Statistical analysis

We used quantile GAMs as implemented in the R package **qgam**, to investigate how the positions of the tongue sensors changed over time, and how these articulatory trajectories were modulated by word ACTIVATION and inflectional exponent. Quantile GAMs (Fasiolo et al., 2017) integrate quantile regression (Koenker, 2005) with the generalized additive model (GAM, Hastie and Tibshirani, 1990; Wood, 2006, 2011, 2013a,b).

GAMs use spline-based smoothing functions to model nonlinear functional relations 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 et al., 2017) for an introduction to spline smooths). Quantile GAMs (henceforth QGAMs) implement a distribution-free method for estimating the predicted values of a given quantile of the response distribution, together with confidence intervals. In our

[1]We compared the models with ACTIVATION as a measure of lexical proficiency to model fits using frequency of occurrence. The direction and strenght of the models was roughly similar between ACTIVATION and FREQUENCY. Models, their summaries and plots can be inspected in the Supplementary Material.

analyses, we investigated the median, but other quantiles can also be of theoretical interest (see, e.g., Schmidtke et al., 2017). The **qgam** package builds on the **mgcv** package (version 1.8-5) for R (Version 3.0.2, (Team, 2014)). We used the `itsadug` package (van Rij et al., 2015) (Version 2.2) for visualization.

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

Time in the stem vowel was normalized to range between 0 and 1 with equidistant incriments. In what follows, we refer to this normalized time as TIME. In order to reduce overly strong influences of outliers, word ACTIVATIONs were ranked for each vowel category, and will henceforth be referred to as ACTIVATIONS. Vowel durations were log transformed and centered. Inflectional exponents, henceforth EXPONENT, were categorized as *stem*, *stem+d*, *stem+s*, *stem+ing*, with *stem* as a reference.

We fitted tongue positions using four models (horizontal and vertical dimension for [iː] and [ɑː]), including an interaction between TIME × ACTIVATIONS × EXPONENT interaction using a tensor product smooth. The three-way interaction with EXPONENT was only included when a χ^2-test on the basis of ML-scores indicated an improvement of the model fit. Otherwise a two-way TIME × ACTIVATIONS was included to the model.

To account for know effects of vowel duration on articulatory amplitude (Gay, 1978), we brought potential modulation 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 in all models. We will not further discuss any effects of this control variable onto articulatory trajectories.

Vowels' articulatory trajectories are influenced by the contexts in which these vowels occur. Putting the effects of the inflectional exponent aside, the consonants flanking the vowel are expected to have their own specific effect on how the vowel is articulated (Hoole

Table 2: *Summary of partial effects of linear mixed-effects models, fitting log transformed 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]	-1.51	0.22	-6.90
Tense: past	[iː]	-0.10	0.02	-4.59
Tense: present	[iː]	-0.08	0.02	-4.10
Tense: progressive	[iː]	-0.45	0.02	-22.23
Num. Segments Base	[iː]	-0.07	0.06	-1.09
Activation	[iː]	-0.00	0.00	-0.96
(Intercept)	[ɑː]	-1.47	0.31	-4.81
Tense: past	[ɑː]	-0.06	0.04	-1.33
Tense: present	[ɑː]	-0.03	0.03	-0.86
Tense: progressive	[ɑː]	-0.32	0.04	-8.47
Num. Segments Base	[ɑː]	-0.16	0.08	-2.07
Activation	[ɑː]	0.00	0.00	0.70

et al., 1993). We therefore included random by-verb factor smooths for TIME in our models (i.e. the lemma). These random smooths are the nonlinear equivalent of the combination of random by-verb intercepts and random by-verb slopes for TIME in the linear mixed model (cf. Baayen et al., 2017). See Wieling et al. (2016) and Tomaschek et al. (2018c) for illustrations of random factor smooths. All model summaries can be found in the Appendix. The statistical analysis can be inspected in the Supplementary Material (downloadable from https://osf.io/zpfqk/).

3 Analysis and Results

3.1 Vowel duration

Although we kept the effects of vowel duration on articulatory trajectories under statistical control, it is still possible that any effect of long-term practice in the current findings still emerged due to significantly shorter vowel durations across ACTIVATIONS (cf. Aylett and Turk, 2004; Munson and Solomon, 2004; Cohen Priva, 2015, for effects of frequency of occurrences or contextual predictability). To rule out this possibility, we first analyzed the relation between ACTIVATION and VOWEL DURATION.

We used a linear mixed-effects regression model (package "lmer" for R, Version 1.1-15, Bates et al. (2014)) for each stem vowel, fitting VOWEL DURATION as a function of EXPONENT (with *stem* as a reference), ACTIVATIONS, and the number of canonical phones in the stem. We did not use the number of segments in the whole word, since it is collinear with morphological condition (see Tomaschek et al., 2018b, for a discussion of the problem of collinearity in statistical analysis). We implemented random intercepts for SPEAKERS and VERB.

The models are summarized in Table 2. While [iː] vowels were significantly shorter in monosyllabic suffixed verbs, no significant differences between these categories were 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, when the stem vowel was [ɑː]. Given that verb ACTIVATIONS were not significantly correlated with duration in either of the vowels, we regard all modulations of the articulatory trajectory across ACTIVATION as a result of ACTIVATION and not of varying vowel durations[2].

3.2 Articulatory trajectories

3.2.1 Model description for the [iː] vowel

Before turning our attention to articulatory trajectories in [iː], we discuss the models. For tongue body movements on the horizontal dimension, a three-way TIME × ACTIVATION × EXPONENT interaction provided a better fit to the model than a two-way TIME × ACTIVATION (reduction of ML-score = 33.602).

For tongue body movements on the vertical dimension, a three-way interaction was not supported: its inclusion did not reduce the ML-score, but rather increased it (ΔML: 38.993), indicating that the model with the two-way interaction TIME × ACTIVATION provided a better fit to the data. Summaries of the parametric and the non-linear effects are provided in the appendix.

Modelling tongue tip movements, the three-way TIME × ACTIVATION × EXPONENT provided a significantly better fit than the two-way TIME × ACTIVATION interaction for both, the horizontal dimension (reduction of ML-score = 140.282) and the vertical dimension (reduction of the ML-score =242.469).

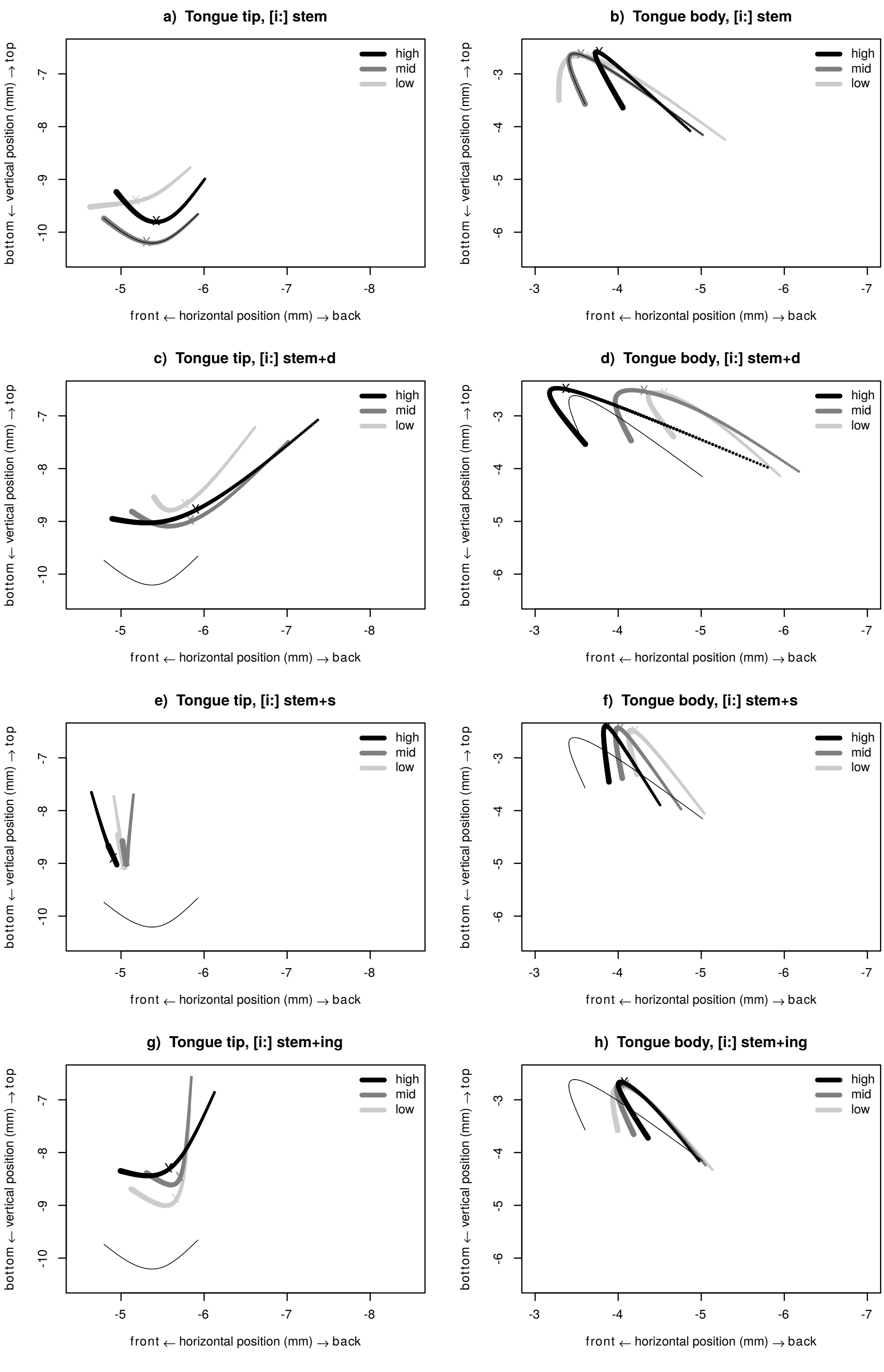

Figure 3: Articulatory trajectories of [iː] in verbs of high (black), mid (dark grey) and low (light grey) ACTIVATION. Left columns illustrate the movement of the tongue tip, the right column illustrates the movement of the tongue body. Axes represent the movement in relation to the most fronted and highest sensor position. The thin solid line represents the articulatory trajectory in *stem* verbs of mid ACTIVATION.

3.2.2 Articulatory trajectories in the [iː] vowel

On the basis of the four models (two dimensions × two sensors) for the [iː] vowel, we calculated a trajectory in the two-dimensional space for the deciles 0.25, 0.50, and 0.75 of

[2]We tested the same models with frequency as a predictor, which also failed to be significantly predictive

ACTIVATIONS (low, mid, high ACTIVATION, respectively) for each of the four EXPONENTS, with the help of the *get_predictions()* function from the **itsadug** package. Contour plots for the tensor product smooths for these models can be found in the Supplementary Material.

The plots in Figure 3 illustrate these trajectories in the midsaggital plane. Black encodes trajectories in verbs of high ACTIVATION, dark grey of mid ACTIVATION and light grey of low ACTIVATION. Time across the vowel is illustrated by means of the trajectory's thickness, with the trajectory tapering off towards the offset. Vowel articulation in *stem* verbs is illustrated in the top row, *stem+d* verbs in the second row, *stem+s* verbs in the third row and *stem+ing* verbs in the bottom row. The left panels present trajectories of the tongue tip sensor, the right panels those of the tongue body sensor.

The *stem* condition

First, we consider tongue movements during the [iː] vowel in *stem* verbs (Figure 3, a & b). We observe an inverse u-shaped movement of the tongue body, which is characterized by a slightly fronting movement as it raises towards the turning point (located at the 'X' which represents the mid time point of the vowel) and a sharp lowering movement as the tongue body gets retracted.

The tongue tip exhibits a constant retraction movement, together with lowering in the first of the trajectory and raising in the second half. Given this u-shaped movement pattern, it it seems to be in phase with the tongue body in the horizontal dimension but anti-phasic in the vertical dimension (Saltzman and Munhall, 1989; Fowler and Saltzman, 1993).

The movement pattern of the tongue body in the *stem* verbs is modulated by ACTIVATION: horizontal movements are more reduced in verbs with greater ACTIVATIONS (Figure 3, b). Tongue positions at specifically the onset and at the center are more retracted. In the vertical dimension, onset positions are slightly lower in verbs of high ACTIVATION than in verbs of low ACTIVATION; the tongue reaches a slightly higher position at the turning point and at the offset.

Movements of the tongue tip in the *stem* verbs (Figure 3, a) differ little in the horizontal dimension. In the vertical dimension, movement trajectories in verbs of high and low ACTIVATION are located at higher positions than in verbs of mid ACTIVATION. Whereas the tongue tip is only retracted and no change in the vertical position can be oserved in verbs of

for vowel duration.

low ACTIVATION, it performs an additional lowering in verbs of high ACTIVATION. It seems as though the anti-phasic coupling between tongue tip and tongue body is more present in verbs of high ACTIVATION.

The *stem+d* condition

We now turn our attention to the *stem+d* verbs with [iː] vowels. Given that the target position of [d] is further back and lower than that of [iː] (Figure 1), we expected the tongue body to be stronger retracted and more lowered in the suffixed verbs than in the *stem* verbs and this effect should be stronger in verbs of higher ACTIVATION. As a reference, we added the trajectory of the *stem* verbs of mid ACTIVATION to the plots for the trajectories in the suffixed verbs. As can be seen in Figure 3 (c & d), no systematic differences between the *stem* and *stem+d* verbs can be observed in the vertical dimension. As for the horizontal dimension, we oberve an effect opposite to our expectation for the first half of the trajectory. The higher the ACTIVATION of the verb, the stronger fronted the trajectory and the turning point become. In addition, the turning point in the trajectory is more fronted than the trajectory in the *stem* condition. In the second half of the trajectory, we find the originally-expected effect of anticipatory coarticulation, with tongue body positions stronger retracted in the *stem+d* than in the *stem* verbs. This retraction emerges for across all ACTIVATIONS. Crucially, although the tongue body is stronger fronted at the turning point in verbs of high ACTIVATION, it is nevertheless fast enough retracted to reach almost the same offset position as found for mid and low ACTIVATION.

Turning to the tongue tip, we expected higher and stronger retracted positions in the suffixed verbs than in the *stem* verbs (Figure 1). Clearly, actual tongue tip trajectories are indeed higher in the *stem+d* than in the *stem* verbs across all ACTIVATIONS, supporting our hypothesis of anticipatory coarticulation in the suffixed verbs (Figure 3, c). Is our hypothesis of stronger anticipatory coarticulation in better practiced verbs supported for the tongue tip trajectory? In the first half of the trajectory, the tongue tip trajectory is lower and stronger fronted in verbs of high ACTIVATION than in verbs of low ACTIVATION, potentially resulting from the coupling between tongue body and tongue tip (Saltzman and Munhall, 1989; Simko and Cummins, 2011). When the tongue body was fronted in verbs of high ACTIVATION, so was the tongue tip. Because the tongue tip is limited by the teeth, it is possible that it is pushed down during the fronting movement. However, the effect changes in the second half of the trajectory during raising, where the tongue tip reaches a more retracted and higher offset position in verbs of high ACTIVATION than in those of low low activation. This means that the upcoming higher [d] is stronger anticipated in better practiced verbs.

In conclusion, greater practice resulted in a more peripheral, i.e. hyperarticulated [iː] pronunciation in the horzontal dimension towards the onset of the vowel concomittant with stronger anticipatory coarticulation of the upcoming [d] towards the offset of the vowel. This effect can be explained by opposing forces onto the movement of the tongue body. On the one hand, greater proficiency allows the tongue to execute articulatory trajectories in a more skilled way, as we already have shown in Tomaschek et al. (2018c) and Tomaschek et al. (2018a). On the other hand, greater proficiency results in stronger anticipation of upcoming targets and gestures, which is in line with findings for hand movements (Sosnik et al., 2004).

The *stem+s* condition

Next, we consider the *stem+s* verbs with [iː] as stem vowels. We expected tongue body positions to be stronger lowered and stronger retracted than in the *stem* condition (cf. Figure 1). Given that the [d] is further back than the [s], the effect of anticipatory coarticulation should turn out to be stronger for [d] as for [s] in the horizontal dimension. Simultaneously, it should be stronger in the vertical dimension, as [s] is lower than [d]

Articulatory trajectories of tongue body in the *stem+s* verbs are characterized by a narrower amplitude in the horizontal dimension than in the *stem+d* verbs across all ACTIVATIONS (Figure 3, e & f). They are furthermore indeed stronger retracted in the *stem+s* than in the *stem* verbs. This is the case at the onset and at the trajectory's turning point.

As for the tongue tip, horizontal positions are more fronted in the *stem+s* verbs than in the *stem* verbs and than in the *stem+d* verbs. Thus, we find effects of anticipatory coarticulation of [d] during [iː] as expected given their position in the articulatory space (Figure 1). Similar to tongue body movements in *stem+d* verbs, tongue body positions were stronger fronted in verbs of high ACTIVATION than in verbs of low ACTIVATION.

For the tongue tip, we expected stronger raised positions in the *stem+s* than in the *stem* position (cf. Figure 1), which is indeed the case (Figure 3, e). No effect of ACTIVATION can be observed in the vertical dimension. However, as can be seen in Figure 1, the distance between the tongue tip sensor and the front of the mouth as well as the alveolar ridge ranges between 6 and 10 mm. Consequently, the tongue tip is constrained to a very limited space.

Any vertical divergence from this position towards the top of the oral cavity might have resulted in a closure of the oral cavity, i.e. a consonant.

Figure 1 lead us to expect stronger fronted tongue tip position in the case of *stem+s* verbs. We indeed find that tongue tip trajectories are more fronted in verbs of high ACTIVATION than in those of mid and low ACTIVATION. This raises the question of why the trajectory in verbs of low ACTIVATION should be more fronted than that of mid ACTIVATION? Given that the tongue body is more fronted in verbs of high than those of low ACTIVATION, it is possible that different strategies to articulate the [s] in lower activated verbs were chosen. Concretely, it is possible that in verbs of low ACTIVATION, the [s] was articulated more with the blade, but in verbs of mid ACTIVATION they were articulated more with the tip of the tongue.

Summarizing the results for the *stem+d* verbs, we regard the findings to indicate two patterns for the articulatory movements of the tongue for [iː] in *stem+s* verbs. On the one hand, more hyperarticulated trajectories in better practiced verbs. On the other hand, stronger anticipatory coarticulation of the upcoming [s].

The *stem+ing* condition

In what follows, we discuss the findings for the *stem+ing* verbs (Figure 3, g & h). Recall that we expected the tongue body to be stronger retracted in these verbs than in the *stem* verbs, to anticipate the more centrally located [ɪ]. This is actually what we find. Furthermore, we expected the tongue body to be lower due to coarticulation. However, we don't observe any changes in the vertical dimension. In addition, no effect of different ACTIVATIONS is present. Noticing the small difference between [iː] and the [ɪ] target in Figure 1, it is actually not surprising that there are no major manipulations of the [iː] trajectory in the context of [ɪ]. There is simply too little room to maneuver during the production of [iː], and a strong anticipation of [ɪ] would risk of elimination the contrast between these two vowels.

For the tongue tip, we expected to observe little to no anticipatory coarticulation in the *stem+ing* verbs, given how near the mean positions of [i] and [ɪ] are. Surprisingly, tongue tip trajectories are actually higher in these verbs than in the *stem* verbs and the raising effect is stronger in verbs of high compared to verbs of low ACTIVATION.

3.2.3 Model description for the [ɑː] vowel

Do the effects of anticipatory coarticulation in tongue tip and hyperarticulation in tongue body that we found for [iː] replicate for the articulation of [ɑː]? The three-way interaction TIME × ACTIVATION × EXPONENT provided a significantly better fit than the two-way interaction TIME × ACTIVATION interaction for both, the horizontal dimension (reduction of ML-score = 53.976) and the vertical dimension (reduction of ML-score = 83.253).

Modelling tongue tip movements, the three-way TIME × ACTIVATION × EXPONENT provided a significantly better fit than the two-way TIME × ACTIVATION interaction for the vertical dimension (reduction of ML-score = 108.546), but not for the horizontal dimension, as the two-way TIME × ACTIVATION interaction yielded a lower ML score (ΔML: 23.799).

3.2.4 Articulatory trajectories in the [ɑː] vowel

From the four models (two dimensions × two sensors) for the [ɑː] vowel, we calculated trajectories in the midsaggital plane for the deciles 0.25, 0.50, and 0.75 of ACTIVATIONS (low, mid, high ACTIVATION, respectively) for each of the four EXPONENTS. Contour plots for the tensor product smooths for these models can be found in the Supplementary Material.

Figure 4 presents the resulting trajectories, again with black encoding trajectories in verbs of high ACTIVATION, dark grey of mid ACTIVATION and light grey of low ACTIVATION.

The *stem* condition

We first discuss the articulation of [ɑː] in the *stem* verbs. The tongue body trajectory is more fronted towards the offset in the verbs of high ACTIVATION than in those of low ACTIVATION (Figure 4, b). Otherwise, we observe little differences in the articulatory trajectory of the tongue body between the verbs of high and low ACTIVATION in either dimension.

In contrast to tongue body and tongue tip movements during [iː], tongue tip and tongue body are in phase during the articulation of [ɑː]. The movement of the tongue tip starts at a back low onset, is subsequently fronted and raises towards the offset (Figure 4, b). We observe that the tongue tip is stronger fronted in the last quarter of the trajectory in verbs of high ACTIVATION than in those of low ACTIVATION. In addition, the offset of the trajectory is slightly higher. This pattern indicates a movement in the direction of the center of the vowel space, i.e. a reduction in verbs of higher ACTIVATION. It mirrors the reduction effect found for [iː] as well as those found in the acoustic domain (Aylett and Turk, 2006; Meunier

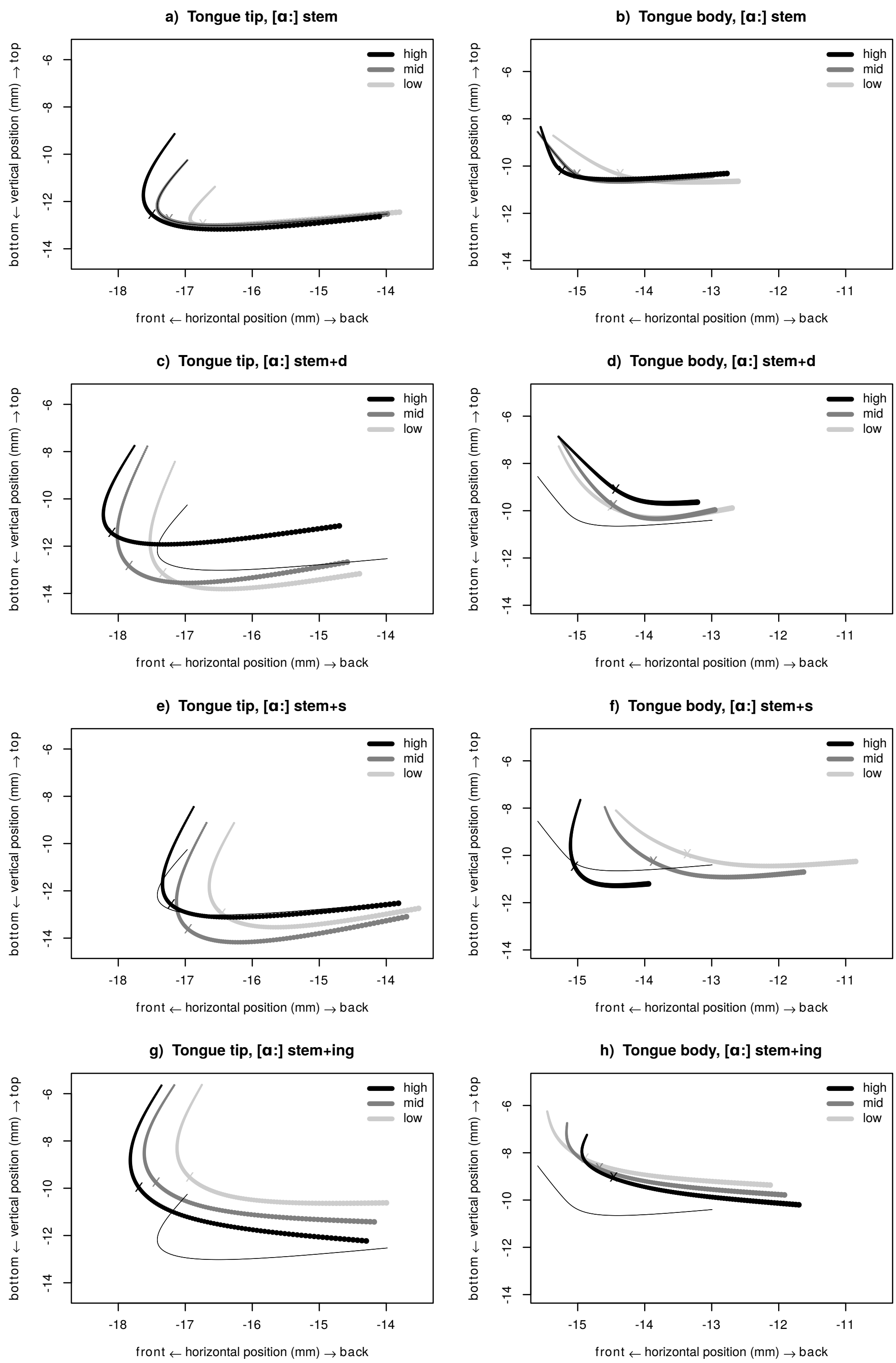

Figure 4: Articulatory trajectories of [ɑː] in verbs of high (black), mid (dark grey) and low (light grey) ACTIVATION. Left columns illustrate the movement of the tongue tip, the right column illustrates the movement of the tongue body. Axes represent the movement in relation to the most fronted and highest sensor position. The thin solid line represents the articulatory trajectory in *stem* verbs of mid ACTIVATION.

and Espesser, 2011).

The *stem+d* condition

We expected that anticipatory coarticulation in [ɑː] would be reflected by a higher and more fronted tongue body position in the *stem+d* than in the *stem* verbs, because the target location for [d] is located higher and more fronted than that for [ɑː] in both sensor positions (cf. Figure 1). We indeed observe the expected effects in the vertical dimension (Figure 4, d). In addition, we find that the trajectory in verbs of high ACTIVATION is higher than that in verbs of mid and low ACTIVATION, indicating stronger anticipatory coarticulation with the [d] with greater practice.

We also expected the tongue tip to be stronger raised and more fronted in the *stem+d* than in the *stem* verbs (thin solid line). Indeed, articulatory trajectories are stronger fronted at the onset of the vowel and and stronger raised towards the offset of the vowel.

Articulatory trajectories in verbs of high ACTIVATION than in verbs of low ACTIVATION are stronger raised at the onset of the vowel and stronger fronted at the offset of the vowel.

The *stem+s* condition

For the *stem+s* verbs, we predicted that tongue body positions should be more fronted compared to the *stem* verbs in order to anticipate the further front [s] (cf. Figure 1). However the opposite is the case (Figure 4, f). Tongue positions are stronger retracted in the *stem+s* verbs than in the *stem* verbs. Articulatory trajectories furthermore change their position relatively to each other, with the trajectories being stronger fronted in verbs of the high ACTIVATION than in verbs of low ACTIVATION, resulting in a smaller movement amplitude.

Since there is only a small difference in height of the tongue body between the target position for [s] and that for [ɑː], we did not expect to observe effects in the vertical dimension. However, we do find some effects in the vertical dimension. In the first half of the trajectory, tongue body positions are lower in verbs of high ACTIVATION than in verbs of low ACTIVATION. For verbs of high ACTIVATION, articulatory trajectories are also lower than in the *stem* condition.

It seems as though the tongue body follows the articulatory path of the tongue tip, for which we predicted higher and more fronted trajectories in the *stem+s* than in the *stem* verbs (Figure 4, e). We find that [ɑː] trajectories differ only minimally at their onset, as well as in comparison to the *stem* verbs. The effect of anticipatory coarticulation and practice emerges only towards the offset. Offsets in *stem+s* verbs are higher than in *stem* verbs and are stronger fronted in in verbs of high ACTIVATION than in verbs of low ACTIVATION, indicating stronger anticipatory coarticulation in better practiced verbs.

The *stem+ing* condition

Finally, we turn our attention to the articulatory trajectories in the *stem+ing* verbs, for which we also predicted higher and more fronted tongue body and tongue tip positions than in the *stem* verbs. Our predictions are partially supported in the vertical dimension (Figure 4, h). However, in the horizontal dimension, the onset of [ɑː] trajectories is significantly further back in the *stem+ing* verbs, as is slightly their offset, than in the *stem* verbs, resulting in larger horizontal movement amplitudes. In addition, inspecting the trajectories across the ACTIVATIONS relatively to each other, we observe minimal, but nevertheless significantly stronger lowering and stronger retraction in verbs of high ACTIVATION than in verbs of low ACTIVATION. Thus, the tongue body became stronger hyperarticulated in better practiced verbs in spite of the need to anticipate the upcoming exponent.

The first half of the tongue tip's trajectory mirrors the effect of the tongue body in the vertical dimension, but is reversed in the horizontal dimension (Figure 4, g). While tongue tip trajectories in [ɑː] are on average higher in the *stem+ing* verbs than in the *stem* verbs, lower, more hyperarticulated trajectories can be observed in verbs of high ACTIVATION. In the second half of the trajectory, however, anticipatory coarticulation seems to be the stronger force, as trajectories have higher offsets than in the *stem* condition and become more fronted in verbs of high ACTIVATION.

4 General discussion

There is a growing number of studies showing that the durational characteristics of words and segments vary depending on their grammatical and morphological function (Kemps et al., 005a; Drager, 2011; Lohmann, 2017; Plag et al., 2017; Seyfarth et al., 2017). Here, we sought to see whether tongue movements in the stem vowels [iː] and [ɑː] recorded by means of electromagnetic articulography vary due to the morphological function realized in the carrier verb. Concretely, we investigated how inflectional exponents in words such as 'cleaned', 'cleans', 'cleaning' changed the vowel's articulatory pattern compared to the bare stem 'clean', due to anticipatory coarticulation (Öhman, 1966; Sziga, 1992; Magen, 1997). In line with effects of practice observed in kinematic studies of arm movements as well as in articulation, we expected that more practice with individual word forms, estimated by their activations in a Naïve Discriminative Learning form-to-lexome network, should be reflected in stronger anticipatory coarticulation of inflectional exponents as a result of stronger overlap between the vocalic gesture and the gesture responsible for the inflectional exponent (e.g. Liberman and Mattingly, 1985; Browman and Goldstein, 1986; Saltzman and Munhall, 1989; Fowler and Saltzman, 1993; Sosnik et al., 2004; Tiede et al., 2011).

Articulatory trajectories in suffixed verbs support this hypothesis. However, the findings have to be broken down into effects of the sensor position under investigation and the part of the articulatory trajectory. We observed anticipatory coarticulation of the upcoming inflectional exponent across the entire articulatory trajectory in the tongue tip position and anticipation is indeed stronger in better practiced verbs than in less practiced verbs. In the tongue body, however, the effect of stronger anticipatory coarticulation is restricted to the second half of the trajectory, i.e. towards the offset of the trajectory. By contrast, in the first half of the trajectory, i.e. towards the onset of the trajectory, exactly an opposite effect can be observed. Articulatory trajectories were produced either with more extreme tongue movements or at more peripheral tongue positions, i.e. further away from the vowel space center, in better practiced verbs than in less practiced verbs. Crucially, modulations of anticipatory coarticulation emerge purely as a result of greater practice rather than changes in vowel durations, which have been attested elsewhere (Aylett and Turk, 2004; Bell et al., 2009), as no significant correlation between activation and vowel duration was attested in the present study.

The finding of more extreme articulation is at odds with observations that vowels tend to be phonetically reduced in words with a higher frequency of occurrence (which we also observed in the bare stem verbs). Such reductional phenomena are usually investigated in corpora by means of monomorphemic words and interpreted as a result of lower informativity in studies investigating vowel productions (Aylett and Turk, 2006; Meunier and Espesser, 2011). By contrast, the current finding replicates results observed in experimental studies in which words were articulated in lists (Tomaschek et al., 2013, 2014, 2018c).

One potential explanation is that word frequency in corpus studies is confounded with contextual predictability which is a good predictor for phonetic reduction (Bell et al., 2009; Jurafsky et al., 2000). Although these studies include conditional probabilities to account for contextual predictability, the measure is usually calculated on the basis of directly neighboring words. It thus misses contextual predictability which emerges from words or even sentences further away than the direct neighbor. Thus, correlations between word frequency and reduction might still be due to the context and not due to the frequency of a word. By contrast, contextual predictability is not relevant for list readings, probably allowing the effect of practice to emerge. It remains for future research to investigate why the different approaches show different effect directions.

The present finding of more centralized [iː] and fronted [ɑː] also indicates that phonetic centralization which has been previously attested as phonetic reduction, might actually be the result of stronger anticipatory coarticulation of upcoming articulatory gestures. Thus, another reason for phonetic reductions might be that speakers become more efficient in their articulatory behavior (as has been shown for hand movements, Sosnik et al., 2004) rather than being simply lazy (Zipf, 1935) or conveying less information (Aylett and Turk, 2004). Furthermore, our study shows that changes in the speech signal originate at different stages of speech production, and that anticipation and centralization can be different at the onset and at the offset of a vowel.

Given the direction of the effects in each of the vowel-morphological category combinations in the present study, it rather seems as though greater practice results in larger differences between the different morphological categories of the verbs, reflecting the morphological function of the verbs by enhancing the vowel. Enhancement of morphological function mirrors the finding of Tomaschek et al. (2019a) who investigated the duration of word final [s] in American English words. They trained a Naïve Discriminative Learning network to discriminate morphological functions of [s] on the basis of contextual word and form cues. They found that [s] durations were enhanced when the activation of the morphological function was high in a given context (see also Kuperman et al. (2007); Cohen (2015) for similar effects and Buz et al. (2016) for phonetic enhancement when a phone has to discriminate minimal pairs and Hall et al. (2018) for how this effect relates to phonetic mergers). Taken together, articulatory gestures are articulated with greater expertise, when practice with articulating a morphological function is high, potentially increasing the discriminability of the phonetic signal.

Crucially, Levelt et al. (1999)'s model of lexical access, predicts that, first, phonetic spell-out is independent from morphological function; second, the morphological function is discriminated only by means of morphemes; and third, no phonetic effect arises in relation with the amount of experience with that morphological function. However, our findings show that a word's morphological function becomes already visible in the stem, i.e. before the uniqueness point after which the morpheme discriminates the morphological function (Kemps et al., 005a; Balling and R. Baayen, 2008). The present results are therefore in line with a growing body of studies showing fine phonetic changes related to higher linguistic structures, especially morphology (Cho, 2001; Smith et al., 2012; Lee-Kim et al., 2012; Plag et al., 2017; Lohmann, 2017; Seyfarth et al., 2017).

The question arises at what level of speech production our results arise. It is possible that the current findings might be attributed to post-lexical processes after a lemma's morphological function has been used to generate a phonological representation of the verb form. The effects attested in this study would thus reflect only greater practice of the articulatory gestures rather than the morphological function. Consequently, such findings would not allow to assume that fine phonetic detail in relation to morphology is lexically valid.

However, Gahl (2008) has shown that the acoustic duration of homophones, i.e. words of different semantics that hypothetically share the same articulatory gestures, is shorter when their lemma frequency is higher. Lohmann (2017) replicated this finding for homophonous noun-verb conversion pairs. Drager (2011) and Podlubny et al. (2015) investigated the articulation of the English word 'like' and showed that the relative durations of its segments varies systematically with its grammatical function. Cohen (2015) found that Russian vowels are articulated in more peripheral positions when they had greater paradigmatic support and Tomaschek et al. (2019a) showed that the duration of word final [s] in American English are enhanced when speakers have greater experience with the morphological function of the carrier word.

All these studies show that the linguistic experience and amount of practice are associated with the lexical rather than post-lexical level affect the phonetic signal. In addition, there is also a growing body of studies showing that listeners are sensitive to these fine spectral and durational cues that emerge due to articulatory processes, attributing them to morphological functions (Davis et al., 0002; Kemps et al., 005a,b; Blazej and Cohen-Goldberg, 2015; Balling and R. Baayen, 2008; Blazej and Cohen-Goldberg, 2015). In other words, these studies show that the mental lexicon contains information about how lexical structure, specifically morphology, affects the speech signal and this information is used during perception, and as our findings show, during speech production.

Our study therefore shows that speakers acquired greater articulatory practice and this practice becomes part of their mental lexicon. Just recently, investigating speech production of the same speakers over a periode of 28 years, Gahl and Baayen (2019) have shown that the positions of vowels in the F1/F2 space shift towards the periphery, the older the speakers are, making them more discriminable. Thus, the changes attested in the present study do not only mirror improved kinematic skills but also greater lexical proficiency. Greater lexical proficiency 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 et al., 2014).

References

Arnold, D., Tomaschek, F., Lopez, F., Sering, K., Ramscar, M., and 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. and 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. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. Journal of the Acoustical Society of America, 119(5):3048–3058.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., and 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., and 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., and 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. and R. Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from danish. Language and Cognitive Processes, 23(7-8):1159–1190.

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

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and 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. and 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. and Harris, K. (1982). Temporal patterns of coarticulation: Lip rounding. Journal of the Acoustical Society of America, 71(2):449–454.

Blazej, L. J. and 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. and Goldstein, L. (1986). Towards an articulatory phonology. Phonology, 3:219–252.

Buz, E., Tanenhaus, M. K., and Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. Journal of memory and language, 89:68–86.

Caselli, N. K., Caselli, M. K., and 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.

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

Cohen, C. (2015). Context and paradigms: two patterns of probabilistic pronunciation variation in russian agreement suffixes. Mental Lexicon, 10(3):313–338.

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

Davis, M. H., Marslen-Wilson, W. D., and 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., and Wood, S. N. (2017). Fast calibrated additive quantile regression.

Fowler, C. and 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. and Saltzman, E. (1993). Coordination and coarticulation in speech production. Language and Speech, 36(2-3):171–195.

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

Gahl, S. and Baayen, R. H. (2019). Twenty-eight years of vowels: Tracking phonetic variation through young to middle age adulthood. Journal of Phonetics, 74:42–54.

Gahl, S. and Strand, J. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. Journal of Memory and Language, 89:162 – 178.

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

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

Hall, K. C., Hume, E., Jaeger, T. F., and Wedel, A. (2018). The role of predictability in shaping phonological patterns. Linguistics Vanguard.

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

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

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2000). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, Frequency and the emergence of linguistic structure. John Benjamins, Amsterdam.

Katz, W. and 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., and 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., and 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.

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, H. (2007). Morphological predictability and acoustic salience of interfixes in dutch compounds. JASA, 121:2261–2271.

Lee-Kim, S.-I., Davidson, L., and Hwang, S. (2012). Morphological effects on the darkness of english intervocalic /l/. Laboratory Phonologya, 4(2):475–511.

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

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

Linke, M., Bröker, F., Ramscar, M., and 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. and 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., and Baayen, R. H. (2017). Discrimination in lexical decision. PLOS ONE.

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

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

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

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

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

Podlubny, R., Geeraert, K., and Tucker, B. (2015). It's all about, like, acoustics. In Proceedings of the ICPHS IIXX, Glasgow.

Raeder, C., Fernandez-Fernandez, J., and 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., and 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., and 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. and 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.

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

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

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

Scarborough, R. (2003). Lexical confusability and degree of coarticulation. Annual Meeting of the Berkeley Linguistics Society, 29(1):367–378.

Schmidtke, D., Matsuki, K., and 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., and Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. Language, Cognition and Neuroscience, 0(0):1–18.

Shannon, C. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27:379–423, 623–656.

Simko, J. and Cummins, F. (2011). Sequencing and optimization within an embodied task dynamic model. Cognitive Science, 35(3):527–562.

Smith, R., Baker, R., and Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. Journal of Phonetics, 40(5):689–705.

Sosnik, R., Hauptmann, B., Karni, A., and 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.

Team, R. C. (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

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

Tomaschek, F., Arnold, D., Broeker, F., and R. Baayen, R. H. (2018a). Lexical frequency co-determines the speed-curvature relation in articulation. Journal of Phonetics, 68:103–116.

Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018b). Strategies for managing collinearity in multivariate linguistic data. Journal of Phonetics, 71:249–267.

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019a). Modeling the duration of word-final s in english with naïve discriminative learning. Journal of Linguistics.

Tomaschek, F., Plag, I., M., E., and Baayen, R. H. (2019b). Modeling the duration of word-final s in english with naive discriminative learning. https://psyarxiv.com/4bmwg.

Tomaschek, F., Tucker, B. V., Baayen, R. H., and Fasiolo, M. (2018c). Practice makes perfect: The consequences of lexical proficiency for articulation. Linguistic Vanguard, 4(s2):1–13.

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

Tomaschek, F., Wieling, M., Arnold, D., and 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.

van Rij, J., Wieling, M., Baayen, R. H., and 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.

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

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

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). On p-values for smooth components of an extended generalized additive model. Biometrika, 100:221–228.

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

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.

Zipf, G. (1935). The Psycho-Biology of Language. An Introduction to Dynamic Philology. MIT Press, Cambridge.

Appendix

Table 3: The table summarizes the parametric effects in the models fitting vertical and horizontal movements of the tongue body and the tongue tip during the pronunciation of [iː]. The summary can be interpreted like standard linear mixed-effect regression tables. Horizontal lines seperate the sensors and dimensions.

Term	Sensor	Dim.	Estimate	SE	z-value	p-value
(Intercept)	TT	vertical	-9.330	0.530	-17.670	< 0.05
Tense=stem+d	TT	vertical	0.960	0.060	15.370	< 0.05
Tense=stem+s	TT	vertical	0.700	0.050	12.970	< 0.05
Tense=stem+ing	TT	vertical	1.050	0.080	13.840	< 0.05
(Intercept)	TT	horizontal	-5.430	0.360	-14.930	< 0.05
Tense=stem+d	TT	horizontal	-0.610	0.060	-10.810	< 0.05
Tense=stem+s	TT	horizontal	0.430	0.050	8.570	< 0.05
Tense=stem+ing	TT	horizontal	-0.240	0.070	-3.380	< 0.05
(Intercept)	TB	vertical	-3.260	0.580	-5.620	< 0.05
Tense=stem+d	TB	vertical	0.100	0.040	2.490	< 0.05
Tense=stem+s	TB	vertical	0.190	0.040	4.720	< 0.05
Tense=stem+ing	TB	vertical	-0.080	0.050	-1.510	0.132
(Intercept)	TB	horizontal	-4.080	0.380	-10.610	< 0.05
Tense=stem+d	TB	horizontal	-0.330	0.070	-4.810	< 0.05
Tense=stem+s	TB	horizontal	-0.180	0.060	-3.110	< 0.05
Tense=stem+ing	TB	horizontal	-0.300	0.080	-3.740	< 0.05

Table 4: The table summarizes the non-linear effects in the models fitting vertical and horizontal movements of the tongue body and the tongue tip during [iː]. Estimated degrees of freedom (edf) of a tensor product smooth larger than 1 indicate a non-linear functional relation between the interacting predictors and the response variable. P-values smaller than 0.05 indicate significance of the effect. Horizontal lines seperate the sensors and dimensions.

Term	Sensor	Dim.	edf	Ref.edf	Chi square	p-value
te(Time,Activations):Tense=stem	TT	vertical	8.580	9.640	174.830	< 0.001
te(Time,Activations):Tense=stem+d	TT	vertical	8.310	9.540	98.750	< 0.001
te(Time,Activations):Tense=stem+s	TT	vertical	7.080	8.200	58.620	< 0.001
te(Time,Activations):Tense=stem+ing	TT	vertical	8.420	9.660	105.800	< 0.001
te(Time,Activations):Tense=stem	TT	horizontal	3.010	3.020	26.270	< 0.001
te(Time,Activations):Tense=stem+d	TT	horizontal	4.390	5.080	51.670	< 0.001
te(Time,Activations):Tense=stem+s	TT	horizontal	7.820	9.130	27.390	0.001
te(Time,Activations):Tense=stem+ing	TT	horizontal	6.270	7.420	20.130	0.006
te(Time,Activations)	TB	vertical	6.870	6.980	223.850	< 0.001
te(Time,Activations):Tense=stem	TB	horizontal	7.770	9.220	78.870	< 0.001
te(Time,Activations):Tense=stem+d	TB	horizontal	9.130	10.200	156.080	< 0.001
te(Time,Activations):Tense=stem+s	TB	horizontal	5.090	5.840	37.350	< 0.001
te(Time,Activations):Tense=stem+ing	TB	horizontal	5.230	5.970	24.780	< 0.001

Table 5: The table summarizes the parametric effects in the models fitting vertical and horizontal movements of the tongue body and the tongue tip during the pronunciation of [ɑː]. The summary can be interpreted like standard linear mixed-effect regression tables. Horizontal lines seperate the sensors and dimensions.

Term	Sensor	Dim.	Estimate	SE	z-value	p-value
(Intercept)	TT	vertical	-12.350	0.790	-15.690	< 0.05
Tense=stem+d	TT	vertical	0.890	0.210	4.300	< 0.05
Tense=stem+s	TT	vertical	0.460	0.170	2.640	< 0.05
Tense=stem+ing	TT	vertical	2.980	0.220	13.560	< 0.05
(Intercept)	TT	horizontal	-16.330	0.660	-24.890	< 0.05
Tense=stem+d	TT	horizontal	-0.410	0.230	-1.840	0.066
Tense=stem+s	TT	horizontal	0.350	0.190	1.790	0.074
Tense=stem+ing	TT	horizontal	0.250	0.240	1.050	0.292
(Intercept)	TB	vertical	-10.080	0.570	-17.540	< 0.05
Tense=stem+d	TB	vertical	0.960	0.170	5.580	< 0.05
Tense=stem+s	TB	vertical	0.130	0.150	0.910	0.36
Tense=stem+ing	TB	vertical	1.470	0.180	8.090	< 0.05
(Intercept)	TB	horizontal	-14.400	0.780	-18.490	< 0.05
Tense=stem+d	TB	horizontal	0.110	0.220	0.530	0.597
Tense=stem+s	TB	horizontal	0.440	0.190	2.340	< 0.05
Tense=stem+ing	TB	horizontal	0.220	0.230	0.980	0.327

Table 6: The table summarizes the non-linear effects in the models fitting vertical and horizontal movements of the tongue body and the tongue tip during [ɑː]. Estimated degrees of freedom (edf) of a tensor product smooth larger than 1 indicate a non-linear functional relation between the interacting predictors and the response variable. P-values smaller than 0.05 indicate significance of the effect. Horizontal lines seperate the sensors and Dimensions.

Term	Sensor	Dim	edf	Ref.edf	Chi square	p-value
te(Time,Activations):Tense=stem	TT	vertical	8.580	9.640	174.830	< 0.001
te(Time,Activations):Tense=stem+d	TT	vertical	8.310	9.540	98.750	< 0.001
te(Time,Activations):Tense=stem+s	TT	vertical	7.080	8.200	58.620	< 0.001
te(Time,Activations):Tense=stem+ing	TT	vertical	8.420	9.660	105.800	< 0.001
te(Time,Activations)	TT	horizontal	8.760	10.020	122.060	< 0.001
te(Time,Activations):Tense=stem	TB	vertical	6.010	7.170	43.460	< 0.001
te(Time,Activations):Tense=stem+d	TB	vertical	7.670	9.080	69.260	< 0.001
te(Time,Activations):Tense=stem+s	TB	vertical	6.740	8.110	83.770	< 0.001
te(Time,Activations):Tense=stem+ing	TB	vertical	4.740	5.450	68.280	< 0.001
te(Time,Activations):Tense=stem	TB	horizontal	7.010	8.360	56.190	< 0.001
te(Time,Activations):Tense=stem+d	TB	horizontal	4.340	4.990	29.650	< 0.001
te(Time,Activations):Tense=stem+s	TB	horizontal	6.970	8.090	120.770	< 0.001
te(Time,Activations):Tense=stem+ing	TB	horizontal	5.350	6.060	58.790	< 0.001