Relative functional load determines co-articulatory movements of the tongue tip*

Motoki Saito, Fabian Tomaschek, R. Harald Baayen

Eberhard Karls Universität Tübingen

motoki.saito@, fabian.tomaschek@, harald.baayen@uni-tuebingen.de

Abstract

Does morphological structure affect articulation when segmental similarity is strictly controlled? To address this question, we used electromagnetic articulography to study the articulatory trajectories of tongue tip and tongue body during the articulation of German words containing [a(:)] as stem vowels followed by [t] that in roughly half of the words realized an inflectional function. According to a generalized additive model fitted to the articulatory trajectories of tongue body and tongue tip sensors using electromagnetic articulography, a factorial predictor signaling the presence or absence of an inflectional exponent underperformed compared to a quantitative measure derived from a Linear Discriminative Learning model. This quantitative measure is based on the functional load of triphones, i.e., the extent to which a triphone contributes to the meaning of the word. The relative functional load of the stem triphone (centered around [a(:)]) and the triphone pivoted on the [t] emerged as a strong co-determinant of articulation. Importantly, words with a balanced relative functional load (i.e., a value close to zero) revealed optimized smooth co-articulation at tongue body and tongue tip sensors. These results provide evidence for the possibility that differences in the details of articulation straightforwardly reflect differences in meaning as captured by distributional semantics.

Keywords: electromagnetic articulography, Linear discriminative Learning, relative functional load, generalized additive mixed models, optimized articulatory gestures

1. Introduction

The second syllables of the German inflected word *geschafft* ([gə+ʃaf+t] "did/made", past participle) and the German derived word *Fachschaft* ([fax+ʃaft] "(student) association") share the same segments, but the inflected word has a morpheme boundary before the final dental obstruent that is absent in the derived word. This study addresses the question of whether the presence of this inflectional exponent has consequences for how the syllable [ʃaft] is articulated.

According to the WEAVER++ model (Levelt and Wheeldon 1994; Levelt, Roelofs, and Meyer 1999), the answer to this question is a clear no, because articulation is driven by syllables. The syllable [ʃaft] is retrieved from a mental syllabary, and as this syllable is posited to be identical irrespective of whether the final obstruent is a morphological exponent, the model predicts that, apart from prosodic factors such as prominence, articulation should proceed in exactly the same way. However, both acoustic (Plag, Homann, and Kunter 2017; Zimmermann 2016; Seyfarth et al. 2018; Tucker, Sims, and Baayen 2019; Tomaschek et al. 2019) and articulatory studies (Cho 2001; Lee, Kim, and Cho 2019; Strycharczuk and Scobbie 2016; Song et al. 2013) have reported different phonetic realization depending on morphological structure.

In the present study, we addressed the potential consequences for articulation of the presence of the German word-final inflectional exponent (-t) while holding constant the segments in the syllable. We selected words such as *geschafft* and *Fachschaft*, and used electromagnetic articulography to clarify whether systematic differences exist in how the rimes of the final syllables were articulated. In addition to predictors included to control statistically for duration, stress, and frequency of use, we included a predictor for relative functional load, a novel measure that we introduce in the next section.

2. Relative functional load

When we consider words' meanings from the perspective of distributional semantics, the function of the German word final -t is remarkably different for inflected forms such as `geschafft` and uninflected words such as `Fachschaft`. For inflected words, the -t serves to position a word's meaning properly in semantic space with respect to tense, aspect, and number. By contrast, the -t in `Fachschaft` has a general discriminatory function similar to any other segment in the word. In order to quantify this difference in functional load, we made use of Linear Discriminative Learning (LDL, Baayen et al. 2019). This computational model implements direct mapping between high-dimensional form vectors and high-dimensional semantic vectors (word embeddings from distributional semantics).

The present study represents words by numeric vectors indicating which triphones are present in a word's form. For semantic vectors, we explored two methods. The first method simulates semantic vectors while at the same time building in inflectional semantics. For example, the semantic vectors of *painted* and *bought* are similar to each other, because they share the inflectional function of the past tense. Since current implementations of LDL do not provide means to implement semantic similarity between stems, *bought* and *purchased* receive nearly orthogonal semantic vectors. In order to assess the importance of semantic similarity for stems, the present study also made use of word2vec (Mikolov et al. 2013) to create semantic vectors.

Accordingly, two LDL models were trained, one of which was trained with simulated semantic vectors, and one of which was trained with word2vec embeddings. The model architecture is laid out in Figure 1. Triphones are linked with semantic units, the weights on the connections from form to meaning are shown in different colors, one for each of the triphones present in the word `band`. For the triphone `#ba`, the red column vector

*The original title: "Semantic measures determining coarticulatory movements of the tongue tip".

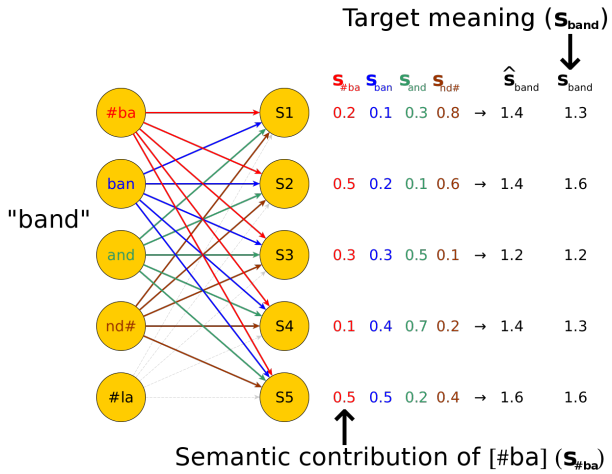

Figure 1: LDL is a two layer network with a form layer with triphones and a semantics layer. The functional load of a triphone is defined as the correlation between the semantic vector of the target word and the contribution that this triphone makes to the semantic vector of the target word. This contribution is the pertinent vector of the network's weight matrix, e.g., for the triphone `#ba` the red vector. $\hat{s}_{\text{band}}$: the predicted semantic vector, s_{band}: the gold standard semantic vector.

$s_{\#\text{ba}}$ represents the contribution that this triphone makes to the semantic vector $\hat{s}_{\text{band}}$. Assuming that $\hat{s}_{\text{band}}$ is a good approximation of s_{band}, the functional load of `#ba` can be defined as the correlation between $s_{\#\text{ba}}$ and s_{band}.

Since the focus of the present study is on the articulation of the rime of words such as `Fachschaft` and `geschafft`, we zoomed in on the functional loads of the triphone with the vowel in its center, and the triphone to its right: [ʃaf] and [aft] for both `Fachschaft` and `geschafft`. We refer to these two functional loads as ϕ_s (which primarily gauges the functional load of the stem) and ϕ_e (which primarily gauges the functional load of the exponent, if present). Across different words, the vectors of values of ϕ_s and ϕ_e ($\vec{\phi}_s$ and $\vec{\phi}_e$) are highly correlated. To avoid adverse effects of collinearity, we calculated the relative functional load, defined as $\phi_s - \phi_e$.

When the stem triphone (ϕ_s) and the transition triphone (ϕ_e) are perfectly balanced with respect to their functional load, the relative functional load is zero. When ϕ_s, which primarily captures the contribution of the stem, has a higher functional load than ϕ_e, which represents the functional load of the exponent, then the relative functional load becomes positive. On the other hand, relative functional load becomes negative when the functional load of the exponent triphone exceeds that of the stem.

We calculated two relative functional loads for each triphone, one based on LDL simulated semantic vectors ϕ_{sim}, and one using `word2vec` embeddings ϕ_{word2vec}. In what follows, we investigate whether relative functional load codetermines the articulation of the final rime of morphologically simple and morphologically complex words.

3. Methods

3.1. Materials

Tongue movement data were extracted from the Karl Eberhards Corpus of spontaneously spoken southern German (KEC) (Arnold and Tomaschek 2016). Tongue movements were recorded with electromagnetic articulography (EMA, NDI WAVE articulograph, sample rate 400 Hz). The present study focuses on vertical and horizontal movements of the tongue tip ("T.T.") and the tongue body ("T.B.") in the midsagittal plane. The words in the present study all contained the sequence [a(ː)(C)t], where [(C)] represents a potential intervening consonant. Words were tagged for whether an inflectional exponent (-t) was present. The total number of word types considered in the analyses was 98, of which 34 have a word final -t exponent. The total number of audio tokens was 8757 , of which 2448 were inflected.

3.2. Analysis

Analyses were carried out with Generalized Additive Mixed-effects Model (GAMM) (Wood 2006), a generalization of multiple regression that enables the analyst to study non-linear relationships between a response variable and one or more (numeric) predictors. The response variable was sensor position, and a four-level factor F was used to obtain smooth functions for position as a function of normalized time for each the four combinations of dimension (horizontal vs vertical) and sensor (T.B., T.T.). A second factor, henceforth Exponent, coded the presence of an inflectional exponent. All factors were set up using treatment dummy coding. Log-transformed frequency of occurrence (based on the *SdeWac* corpus (Faaß and Eckart 2013)), the two relative functional load measures, and the duration of the syllable nucleus were included as covariates. Random intercepts were included for speaker, for intervening segment (including a factor level specifying the absence of such a segment), as well as for the segments preceding the vowel and following the exponent.

4. Results

We first fitted a model to the sensor positions that included `Frequency` and `Exponent` as predictors, leaving out the measures of relative functional load. We then fitted two further models in which `Frequency` and `Exponent` were replaced by one of the two measures of relative functional load (ϕ_{word2vec} or ϕ_{sim}).

Model comparison indicated that the model with ϕ_{sim} (AIC: 1128522, ML: 564386.4) outperformed the models using either ϕ_{word2vec} (AIC: 1228344, ML: 614316.1) or `Frequency` in combination with `Exponent` (AIC: 1228091, ML: 614169.2). Importantly, the model with relative functional load as predictor provided a more accurate fit to the data, while requiring fewer parameters.

Figure 2 shows the estimated movement in the vertical dimension of the tongue tip as a function of normalized time, using the GAMM with ϕ_{sim} as measure of relative functional load. In this contour plot, the X-axis represents normalized time, and the Y-axis represents relative functional load. Colors represent the height of the tongue tip, with warmer colors denoting higher positions and colder colors coding lower positions. As can be seen in Figure 2, the tongue tip is steadily moving upwards during the production of [a(ː)], anticipating the upcoming [t].

When ϕ_{sim} is close to zero, indicating that the functional load of the vowel and exponent triphones are balanced, the vertical trajectory of the tongue tip has very small gradient. For larger positive or negative values of relative functional load, the tongue tip starts out at increasingly lower positions while mov-

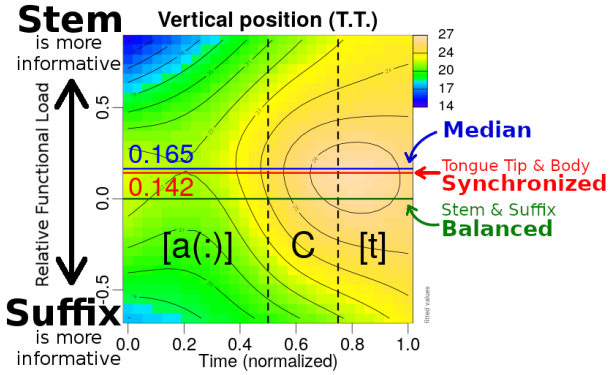

Figure 2: Vertical tongue tip movements along time (x-axis) for different values of relative functional load (y-axis). The vertical dashed lines denote the offset of the vowel for words with a complex nucleus (left) and with only the -t exponent following the vowel (right).

ing upwards with increasingly steep gradients: The number of contour lines crossed for $\phi_{\text{sim}} \approx 0$ is 4, whereas for large negative or positive values, the number of contour lines crossed increases up to 7.

The tongue tip reaches the highest position for relative functional load value slightly bigger than $\phi_{\text{sim}} = 0$, and here the vertical distance traveled reaches its minimum. This suggests that articulation is optimal, in the sense of requiring least effort, at values of ϕ_{sim} that are slightly favoring the functional load of the stem vowel triphone. Interestingly, the median relative functional load is 0.165, which is close to the 'ridge' of highest positions across time (the blue line in Figure 2). This suggests that the tongue tip moves least in the vertical direction for words with the most commonly encountered values of relative functional load.

If indeed the 'ridge' trajectory around $\phi_{\text{sim}} = 0.165$ reflects optimization of articulatory effort, the question arises of whether the tongue body shows similar optimization. Furthermore, does this optimization take place not only in the vertical but also in the horizontal dimension? Interestingly, and unsurprisingly as tongue tip and tongue body are tightly coupled, joint optimization receives support from a further analysis in which we examined the movements of the sensors in the midsaggital plane (Figure 3). Curves are shown relative to their initial position at the onset of the vowel. As a consequence, the trajectories of the two sensors always start at the origin. If two curves have completely parallel trajectories, after being shifted to start at the origin, they will show up with completely overlapping curves. Thus, the larger the differences between the two curves in Figure 3, the less synchronized and parallel the articulatory trajectories of the two sensors are in the midsaggital plane.

Figure 3 presents sensor trajectories for selected values of relative functional load. The red curves represent the tongue body, and the blue-green curves those of the tongue tip. Deeper shades of blue-green and red indicate later points in normalized time. In the leftmost panel, the tongue body sensor starts out at a low, moves back slightly, and then moves up. The tongue tip also starts low, and moves up with very little displacement along the horizontal axis. As relative functional load increases, from left to right in Figure 3, the final horizontal position of the tongue body sensor moves further towards the front. At the same time, it is lowered. For the tongue tip, final vertical positions first decrease and then increase again. A similar pattern is present for the lowest vertical position reached.

Importantly, as relative functional load is increased, the two articulatory trajectories first move closer together, and then move further apart again. Apparently, when relative functional load is close to 0.1, the tongue body and tongue tip trajectories are most similar and most strongly coordinated. For each relative functional load value, we quantified the amount of synchronization of the tongue tip and tongue body sensor using the mean Euclidean distances between tongue tip and body trajectories. The closest, and hence most synchronized, trajectories of the tongue tip and body were found for a relative functional load equal to 0.142 (indicated by the horizontal red line in Figure 2). This is close to the value of relative functional load for which we observe the ridge with maximal vertical tongue positions in Figure 2. In other words, for a relative functional load that slightly favors the stem triphone, we observe minimal displacement of the tongue tip sensor in combination with the tightest synchronization with the tongue body sensor.

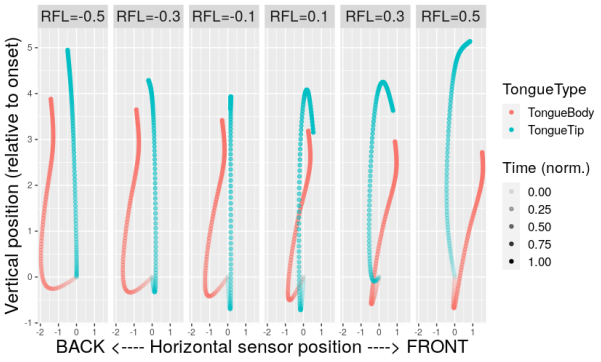

Figure 3: Tongue tip/body positions relative to their initial positions at the onset of [a(:)]. At $t = 0$, both curves start out at the origin, darker shades of color indicate later points in time. The more similar the two curves are, the more strongly coordinated and parallel in the midsaggital plane the articulatory trajectories of the tongue tip and tongue body sensors are.

5. Discussion

In this study, we addressed the question of whether words in which the final [t] realizes an inflectional function are articulated differently from matched controls in which the final [t] does not realize an inflectional function. We documented that a generalized additive mixed model with word frequency and morphological status as predictors did not provide as accurate a fit as a simpler model with relative functional load as covariate. Unsurprisingly, mean relative functional load is smaller for words realizing the inflectional exponent (0.067) compared to the non-inflected controls (0.218, $t(46.333) = 2.423, p = 0.019$). The lower relative functional load for the inflected words indicates that the contribution of the triphone of the exponent to the meaning of the word is somewhat larger for inflected words compared to uninflected controls. It is noteworthy that on average, even for inflected words, the functional load of the triphone of the exponent is slightly above zero. Since inflection typically modifies the syntactic positioning of the baseword without really changing its meaning, the generally somewhat stronger functional load of the stem triphone makes sense.

The finding that, given relative functional load, a factorial predictor for morphological status is no longer required to predict tongue position, suggests that the theoretical construct of a "morpheme boundary" is not useful, just as the theoretical construct of the "morpheme" is highly problematic (Blevins 2016). More generally, theories in psychology that build on morphological units such as stems and exponent, such as the WEAVER++ model (Levelt, Roelofs, and Meyer 1999), are challenged by the finding that low-level sublexical units such as triphones and their contributions to words' semantics are actually driving the fine details of articulation.

The importance of the frequency of occurrence of complex words for speech production is unresolved (contrast, for instance, Levelt, Roelofs, and Meyer (1999) and Janssen, Bi, and Caramazza (2008)). For the present data, frequency was predictive for tongue sensor position in the model from which relative functional load was withheld as predictor. Furthermore, if frequency is added to the model building on relative functional load, then it does help improve model fit. However, the improvement of the Maximum Likelihood score and the AIC score by the addition of frequency as predictor was considerably smaller than the improvement offered by relative functional load. When predictors are compared with respect to their variable importance using Random Forests, we also observe a much smaller importance of frequency (4.226), compared to relative functional load (14.916). The strong effect of relative functional load suggests it may be profitable to revisit chronometric experiments and to clarify whether there too frequency is outperformed by relative functional load.

Our analyses also revealed that tongue tip and tongue body trajectories were synchronized the most tightly, and moving upward with the smallest gradient, for words with a relative functional load close to the median functional load. To understand why this pattern may be present in our data, we note that inflected words have to strike a balance between staying faithful to the meaning of the base word, while at the same time clarifying the syntactic role of the inflected word. When these two functional constraints are in equilibrium, articulatory trajectories are optimized in such a way that movements are as smooth, and possibly as efficient as possible. The further away a word is from this equilibrium, the greater the gradient of the articulatory trajectory becomes, and the less tongue tip and tongue body are synchronized. If this interpretation is on the right track, we are seeing the articulatory consequences of functional markedness.

6. Acknowledgments

This study was funded by the European Research Council (WIDE-#742545 awarded to the third author) and by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 'Spoken Morphology', Project 'The articulation of morphologically complex words', BA3080/3-2).

7. References

Arnold, Denis and Fabian Tomaschek (2016). "The Karl Eberhards Corpus of spontaneously spoken southern German in dialogues — audio and articulatory recordings". In: *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, pp. 9–11.

Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan, and James P. Blevins (2019). "The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning". In: *Complexity*, pp. 1–39. DOI: 10.1155/2019/4895891.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press.

Cho, Taehong (2001). "Effects of Morpheme Boundaries on Intergestural Timing: Evidence from Korean". In: *Phonetica* 58, pp. 129–162.

Faaß, Gertrud and Kerstin Eckart (2013). "SdeWaC – A Corpus of Parsable Sentences from the Web". In: *Language Processing and Knowledge in the Web*. Ed. by Iryna Gurevych, Chris Biemann, and Torsten Zesch. Darmstadt, Germany: Springer, pp. 61–68.

Janssen, N., Y. Bi, and A. Caramazza (2008). "A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds". In: *Language and Cognitive Processes* 23.7-8, pp. 1191–1223.

Lee, Jiyoung, Sahyang Kim, and Taehong Cho (2019). "Effects of morphological structure on intergestural timing in different prosodic-structural contexts in Korean". In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Ed. by Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Levelt, Willem J. M., Ardi Roelofs, and Antje S. Meyer (1999). "A theory of lexical access in speech production". In: *Behavioral and Brain Sciences* 22, pp. 1–75.

Levelt, Willem J. M. and Linda Wheeldon (1994). "Do speakers have access to a mental syllabary?" In: *Cognition* 50, pp. 239–269. DOI: 10.1016/0010-0277(94)90030-2.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). "Efficient estimation of word representations in vector space". In: *Proceedings of the International Conference on Learning Representations (ICLR 2013)*, pp. 1–12. arXiv: 1301.3781.

Plag, Ingo, Julia Homann, and Gero Kunter (2017). "Homophony and morphology: The acoustics of word-final S in English". In: *Journal of Linguistics* 53.1, pp. 181–216. DOI: 10.1017/S0022226715000183.

Seyfarth, Scott, Marc Garellek, Gwendolyn Gillingham, Farrell Ackerman, and Robert Malouf (2018). "Acoustic differences in morphologically-distinct homophones". In: *Language, Cognition and Neuroscience* 33.1, pp. 32–49. DOI: 10.1080/23273798.2017.1359634.

Song, Jae Yung, Katherine Demuth, Stefanie Shattuck-Hufnagel, and Lucie Ménard (2013). "The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound". In: *Journal of Phonetics* 41.3-4, pp. 281–295. DOI: 10.1016/j.wocn.2013.03.004.

Strycharczuk, Patrycja and James M. Scobbie (2016). "Gradual or abrupt? The phonetic path to morphologisation". In: *Journal of Phonetics* 59, pp. 76–91. DOI: 10.1016/j.wocn.2016.09.003.

Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus, and R. Harald Baayen (2019). "Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning". In: *Journal of Linguistics*, pp. 1–39. DOI: 10.1017/S0022226719000203.

Tucker, Benjamin V., Michelle Sims, and R. Harald Baayen (2019). "Opposing forces on acoustic duration". In: *PsyArXiv*, pp. 1–38. DOI: 10.31234/osf.io/jc97w.

Wood, Simon N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida, U.S.A.: CRC Press.

Zimmermann, Julia (2016). "Morphological status and acoustic realization: Findings from New Zealand English". In: *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology (SST-2016)*. December. Canberra: Australasian Speech Science and Technology Association (ASSTA), pp. 201–204.