An ultrasound study of frequency and coarticulation

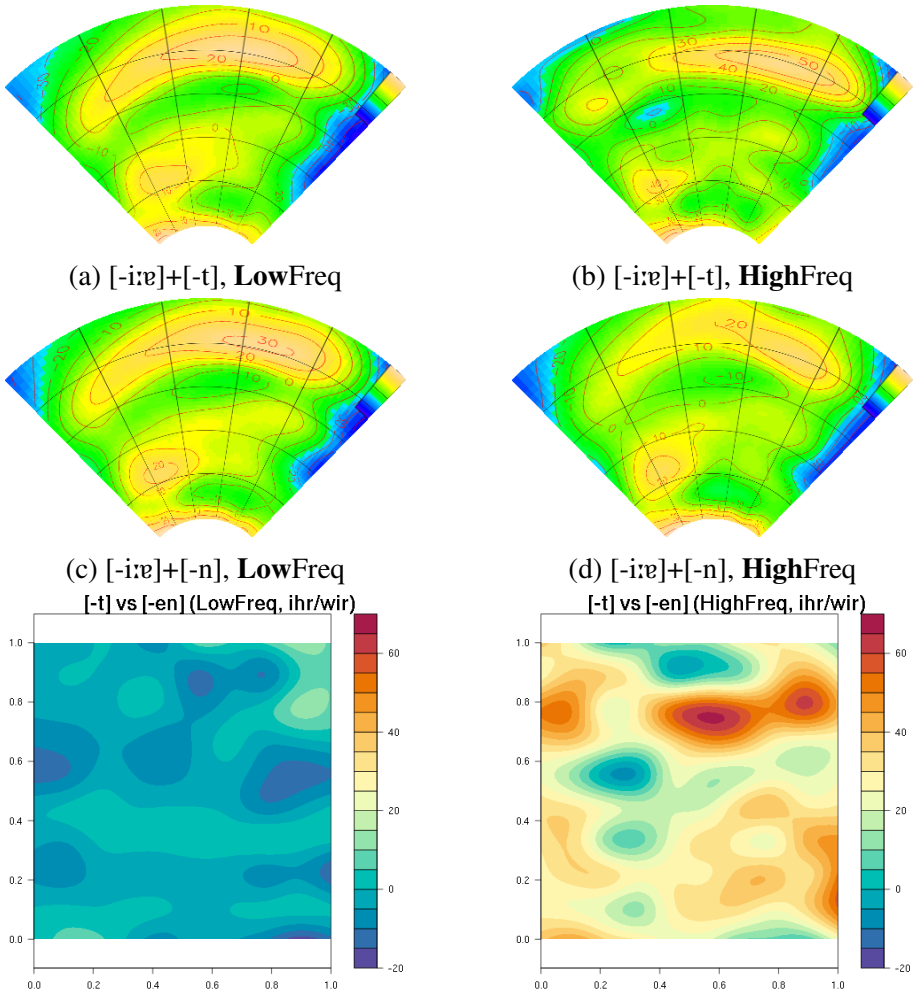
Motoki Saito, Fabian Tomaschek, Harald Baayen
University of Tübingen

In the speech production model proposed by Levelt, Roelofs, and Meyer (1999), word frequency is mainly involved in the selection of target words. Fine details of phonetic realization are supposed to occur after the retrieval of syllabic gestural scores. As a consequence, this model predicts that word frequency cannot be a co-determinant of how articulation is executed. Contrary to this prediction, a number of studies have suggested that tongue movements are actually co-determined by word frequency (Tomaschek, Arnold, Bröker, & Baayen, 2018; Tomaschek, Tucker, Fasiolo, & Baayen, 2018). Tomaschek, Tucker, et al. (2018) observed co-articulation of stem vowels of German inflected verbs with the following suffixes even in the presence of intervening stem-final segments such as [l] in *sie **m**a**l**t* ([zi: m**a:**l**t**], "they paint"). Since co-articulation can be caused not only by following segments but also by preceeding segments (Öhman, 1966), the question remains of whether a frequency effect shaping the extent of co-articulation can be found even after controlling for overlay co-articulatory effects from preceding pronouns, such as [i] in *s**ie** m**a**lt* ([z**i:** m**a:**lt], "they paint"). The present study examines the word frequency effect on co-articulatory patterns in stem vowels of German inflected verbs, taking preceding segments that belong to pronouns preceding and separated from inflected verbs into account.

Generalized Additive Mixed-effects Models (GAMMs) were fitted to ultrasound images with brightness of pixels as a dependent variable. Main predictors were x- and y-coordinates of target pixels, word frequency, preceding pronouns, segments between pronouns and stem vowels, and segments between stem vowels and inflectional suffixes. The target stem vowel was [aː]. The target suffixes were [-t], e.g. *sie malt* ([zi: maːlt], "she paints"), and [-(ə)n], e.g. *sie malen* ([zi: maːl(ə)n], "they paint"). For computational efficiency, five separated GAMMs were fitted, each of which corresponds to five points in time, i.e. 0.00, 0.25, 0.50, 0.75, 1.00, where 0.00 is the beginning and 1.00 is the ending of stem vowels.

The effects of word frequency, preceding, pronouns, and following suffixes were all highly significant at each of these quantiles of time ($p < 0.000$). Figures 1a-1d show fitted ultrasound images during the stem vowel, where likely positions of the tongue is indicated by brighter colors and preceding pronouns are *ihr* or *wir*. Figures 1e and 1f represent differences between these fitted ultrasound images in low and high frequency conditions each. Differences are more pronounced in the high frequency condition, where colors range from blue to red, compared to the low frequency condition, which is covered almost entirely by blue. In addition, the blue region in the middle top in Figure 1f together with the red region immediately below itself suggests that the tongue body should be lower for [-t], compared to [-n]. Also, the red blob in the top left corner in Figure 1f indicates that the shadow in the left side of the images, which is introduced by the hyoid bone located in the root of the tongue, grows bigger when the suffix is [-n]. It is correlated with the fact that the middle part of the tongue is raised in this condition (Figure 1d).

These findings provide further evidence that how exactly vowels of inflected words are realized is determined not only by the surrounding segments, but also by how often these segments co-occur in words and word sequences, contrary to what the production model of Levelt et al. (1999) predicts. Our results indicate that higher frequency of use can be understood as affording improved motor control for articulation, resulting in better optimized articulation.



(a) [-iːɐ]+[-t], **Low**Freq

(b) [-iːɐ]+[-t], **High**Freq

(c) [-iːɐ]+[-n], **Low**Freq

(d) [-iːɐ]+[-n], **High**Freq

(e) [-t]-[-n], **Low**Freq

(f) [-t]-[-n], **High**Freq

Figure 1: Fitted ultrasound images (1a, 1b, 1c, 1d) and differences of these images in each frequency condition. The range of differences (z-axis values) is much bigger (ranging from deep blue to deep red) for high frequency, compared to low frequency (almost all in blue).

References

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Öhman, S. E. G. (1966). Coarticulation in VCV Utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America*, *39*, 151–168. doi: 10.1121/1.1909864

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

Tomaschek, F., Tucker, B. V., Fasiolo, M., & Baayen, R. H. (2018). Practice makes perfect: the consequences of lexical proficiency for articulation. *Linguistics Vanguard*, *4*.