LANGUAGE SPECIFICITY VS SPEAKER VARIABILITY OF ANTICIPATORY LABIAL COARTICULATION IN GERMAN AND ENGLISH

Justin J. H. Lo[1], Christopher Carignan[1], Marianne Pouplier[2], Roy Alderton[2], Francesco Rodriquez[2], Bronwen G. Evans[1], Eva Reinisch[3]

[1]University College London, UK, [2]Ludwig-Maximilians-Universität München, Germany, [3]Austrian Academy of Sciences, Austria

{justin.lo|c.carignan|bronwen.evans}@ucl.ac.uk, {pouplier|f.rodriquez}@phonetik.uni-muenchen.de, roy.alderton@city.ac.uk, eva.reinisch@oeaw.ac.at

ABSTRACT

This paper presents a study of the sources of cross-linguistic variation in anticipatory labial coarticulation. We collected audiovisual data from speakers of German and American English, two languages distinguished by the presence of phonological rounding contrast for vowels, to examine (i) how patterns of coarticulation differ in these languages and (ii) to what extent such patterns are speaker-specific.

Our results indicate that anticipatory labial coarticulation is temporally extensive in both German and English. While our data demonstrate that in both languages there is considerable variability between speakers in their individual degree of coarticulation, there is nonetheless an overarching influence of language on the patterns of coarticulation. In particular, anticipatory coarticulation is less extensive and more constrained in German than in English. These findings contribute to our understanding of the roles of phonological contrast and speaker variability in the temporal organization of anticipatory coarticulation.

Keywords: Lip rounding; coarticulation; individual variation; language specificity; speech production.

1. INTRODUCTION

In speech, the production of individual sound units is influenced by, and in turn influences, their surrounding contexts, as gestures required to articulate successive sounds overlap and interact with each other. The anticipatory initiation of articulatory movement, such as lip protrusion before rounded vowels, provides important articulatory and acoustic cues that are useful for perception in advance of the target sound itself [16]. Anticipatory lip movement in relation to rounding, the focus of this study, is particularly known to be temporally far-reaching, easily extending up to hundreds of milliseconds [10].

The temporal dynamics of anticipatory coarticulation are known to exhibit much cross-linguistic variation [2, 4, 20], but the source of such variation remains debated. Language-specific sound systems and contrasts have been argued to be influential factors governing the degree of coarticulation [12], in that the maintenance of phonological contrasts is expected to restrict the freedom employed by coarticulatory gestures [5, 17]. Yet, evidence for such a claim from various types of coarticulatory phenomena has thus far been mixed (e.g. [13, 18]). In particular, speakers of Swedish, in which rounding is phonologically contrastive, have been found to initiate anticipatory lip protrusion earlier, not later, than speakers of American English, which has no rounding contrast [11]. Even within individual languages, there is substantial speaker variability in the extent of coarticulation (e.g. [7, 19]), casting further doubt on the role of phonological contrast in the temporal organization of coarticulation.

In a cross-linguistic comparison of American English and Canadian French, Noiray et al. [14] examined the initiation of rounding movement in progressively longer [i]-C_n-[u] sequences and showed that the onset of lip constriction extended in proportion to the possible window of coarticulation at a speaker-specific rate. With no evidence of systematic patterns by language, they concluded that the implementation of anticipatory lip rounding is chiefly speaker-dependent.

In this study, we pursue the issue of the sources of cross-linguistic variation in coarticulation. Specifically, we investigate anticipatory labial coarticulation in German, which has phonological rounding contrast, and North American English, which does not, by tracking the dynamic movement of the lips. We examine how the temporal extent of coarticulation differs between German and English, and address the question of whether patterns of coarticulation are predominantly driven by language-level variation or speaker variability.

2. METHODOLOGY

2.1. Materials

Stimuli consisted of a list of (near-)minimal pairs, wherein the target vowel differed by rounding (e.g. German *Sehne–Söhne* /zenə/-/zønə/). Each target word was embedded in a carrier sentence (German: *Aber Elsa legt gern __ beiseite*; English: *But Tessa had said __ pleasantly*), designed to be maximally permissible for anticipatory labial coarticulation. This environment, demarcated by the bilabial anchor /b/ (in *Aber* and *But*) and the onset of the target vowel, formed our region of interest (ROI). Table 1 lists the vowel pairs tested, but the German pairs involving /ɛ/ were excluded from the present analysis so that pairs were matched by phonological vowel height across languages. Three repetitions of the list were elicited in randomized blocks, yielding 144 utterances in German and 174 utterances in English per speaker.

Language	Vowel pairs
German	i–u, i–y, e–o, e–ø ɪ–ʊ, ɪ–ʏ, (ɛ–ɔ)*, (ɛ–œ)*
English	i–u, e–o, ɪ–ʊ

Table 1: Vowel pairs in German and English. Pairs marked * were excluded from analysis.

2.2. Data collection

We collected recordings from 30 native speakers each of German and North American English, using an adapted version of the "blue lip" technique developed by Lallouache [9] to track lip movement. Two video cameras were placed at right angles to capture frontal and side views of the speaker's blue-painted lips at 50 frames per second. Speakers were shown the stimuli on a computer monitor and prompted for production one at a time using *SpeechRecorder* [6], which simultaneously collected audio recordings. Utterances where prosodic breaks were identified immediately before the target word, or where the target word involved mispronunciation or false starts, were excluded from further analysis.

2.3. Measuring lip rounding & coarticulation

In line with previous research [14], we used the area of constriction between the lips (aperture) as a reliable indication of lip rounding. Within each utterance, we processed each frame from the front camera in Julia [1], first applying thresholding in HSV (hue, saturation, value) to isolate the blue lip and reference components, then fitting an ellipse to the area bounded by the inner edges of the upper and lower lips. We took the area of the ellipse to be the aperture and scaled it from pixels to mm^2. At this stage, data from one German and two English speakers were excluded due to technical difficulties with lip tracking. We smoothed the time-varying aperture signals by means of Hampel and Savitzky-Golay filters, and removed any trials with remaining erroneous measurements. A total of 4,074 trials in German and 4,726 trials in English remained.

To determine the onset of coarticulation, we analysed lip aperture in the rounded condition in relation to that in the unrounded condition and considered their point of divergence in time. We calculated the time-varying distance between the aperture of each utterance in the rounded condition and the mean aperture of all utterances in the unrounded condition by the same speaker in the corresponding word pair, within a fixed window of up to 500 ms before the acoustic onset of the target vowel. We then fitted sigmoid curves to each of the resulting difference curves (using the *sicegar* package [3] in R [15]), taking the point where the tangent line with the maximum slope intersected with the minimum asymptote as the onset of coarticulation. The example item in Fig. 1 illustrates this derivation. We excluded poor fits whose RMS error was amongst the highest 5% of each language, leaving 4,139 items in total. For the purposes of reporting and statistical analysis, we used the absolute value of the timing (and henceforth refer to onset of coarticulation as such), such that a higher positive value corresponds to an earlier onset of coarticulation.

2.4. Statistical analysis

To evaluate the effect of language on coarticulation, we fitted a linear mixed-effects model to the onsets of coarticulation obtained above (using the *lmerTest* package [8]). The main predictor of interest, language, was included as a sum-coded fixed effect. As controls, we also included phonological vowel height (sum-coded), ROI duration (mean-centred, as proxy for speech rate) and their interactions with language as fixed effects. To account for

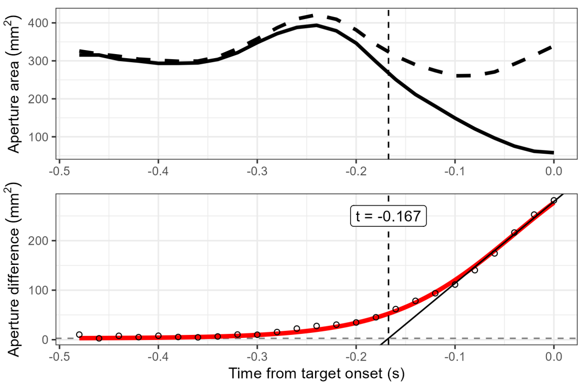

Figure 1: (Top) Aperture preceding /ø/ (solid) vs mean aperture preceding /e/ (dashed) in German *Löhne–Lehne*. (Bottom) Difference curve and its sigmoidal fit, with derived onset of coarticulation (dashed vertical line).

variability across items and speakers, we included random intercepts by word pair and by speaker, as well as random slopes for ROI duration by speaker. In addition, to assess speaker-level effects, we calculated the mean and standard deviation (SD) of onset of coarticulation for each speaker to examine speaker variability in closer detail.

3. RESULTS

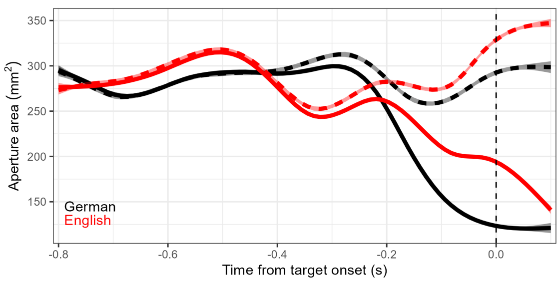

Figure 2: Overall aperture (mm^2) from 800 ms before to 100 ms after target onset (rounded: solid, unrounded: dashed; German: black, English: red; ribbons show ±2SD).

The overall trajectories of aperture in German and English are shown in Fig. 2. In German, aperture in both rounded and unrounded conditions progressed in tandem until shortly after 400 ms before the onset of the target vowel, when the aperture in the rounded condition grew increasingly smaller than that in the unrounded condition. Aperture in English followed trajectories that displayed broadly similar trends, with divergence between the two conditions emerging after 400 ms before target vowel onset. The distance between the conditions continued to rise, as aperture preceding unrounded vowels increased while that preceding rounded vowels fell.

Notwithstanding the similarities in the overall trajectories for German and English, our measure of onset of coarticulation, calculated on the level of individual utterances, revealed a significant effect of language ($p = .0002$). Onset of coarticulation was 53 ms earlier in English than in German, although, as illustrated in Fig. 3, divergence between rounded and unrounded conditions consistently commenced over 200 ms before target vowel onset in both languages. Vowel height was found to be significant as a main effect, but not in interaction with language: coarticulation was more extensive before /i/–/u/ or /y/ than before /ɪ/–/ʊ/ or /ʏ/ ($p = .0081$), and in turn more extensive than before /e/–/o/ or /ø/ ($p = .0024$). There was a significant interaction between language and ROI duration ($p = .0183$), such that earlier onset of coarticulation was found for longer ROIs, to a greater extent in German than in English (Fig. 4).

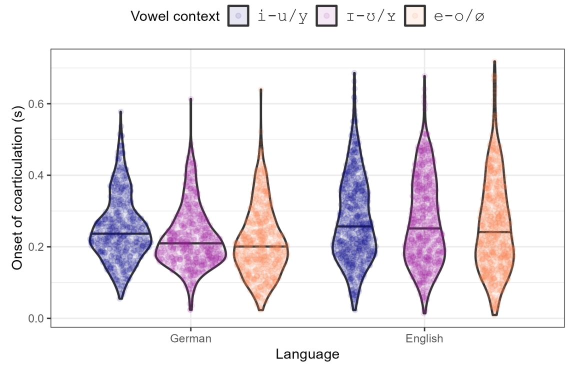

Figure 3: Distributions of onset of coarticulation (in s), measured relative to target vowel onset, in German and English by vowel context.

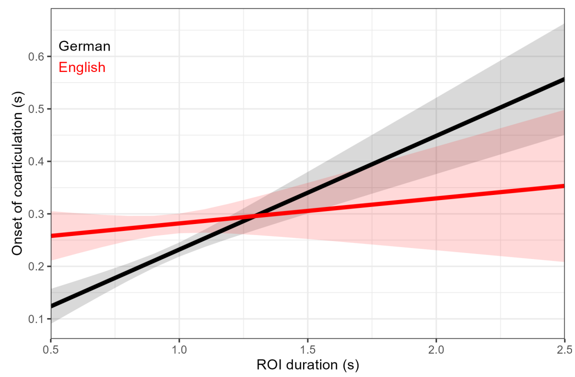

Figure 4: Model-predicted effect of ROI duration on onset of coarticulation (in s) in German (black) and English (red). Ribbons show ±2SD.

In addition to systematic cross-linguistic differences, we found substantial between-speaker variability in both languages, as is evident in Fig.

5. Mean onset of coarticulation for individual speakers, which shows considerable overlap between the two languages, spanned a similarly wide range in German (152–331 ms) and English (182–352 ms). It is also clear from Fig. 5 that onset of coarticulation among English speakers generally had a higher SD than among German speakers, suggesting that English, whose speakers reported higher means on the whole, also exhibited greater within-speaker variability.

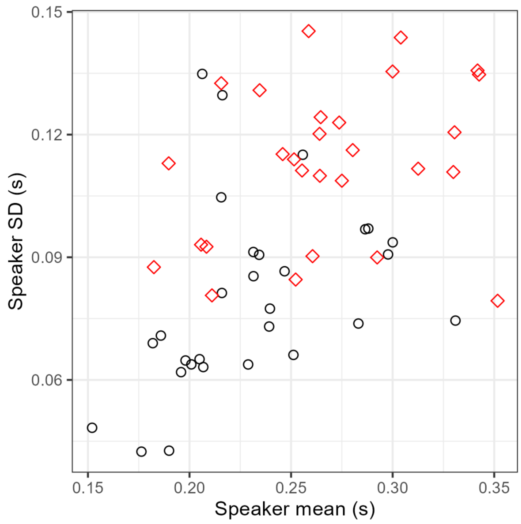

Figure 5: Scatterplot showing mean and SD of onset of coarticulation per speaker (German: black circles; English: red diamonds).

4. DISCUSSION

In this study, we compared the temporal extent of anticipatory labial coarticulation in German and English. Using audiovisual data to capture the dynamics of lip aperture in speech production, we have shown that anticipatory labial coarticulation is extensive in both languages. Coarticulatory information, in the form of differential lip aperture before rounded and unrounded vowels, regularly emerged over 200 ms before the onset of the target vowel in both languages, closely matching the temporal extent of anticipatory lip rounding reported in the literature [16].

Our findings here shed light on two opposing views on the driving force behind cross-linguistic variation in coarticulation. The evidence here lends support to the argument that language-specific phonological contrasts (in this case, rounding contrast in German) can act as a constraining factor on the degree of coarticulation (cf. [12]). Specifically, contrary to previous findings for Swedish [11], we showed that speakers of German, which similarly has phonological rounding contrast, initiated anticipatory coarticulation later, not earlier, than English speakers across all vowel contexts examined. Moreover, on an individual level, English speakers were comparatively less restricted in how much their extent of coarticulation varied from utterance to utterance.

The results here are also consistent with the position that implementation of labial coarticulation is highly speaker-specific [14]. Onset of coarticulation showed an impressive range of inter-speaker variation in both languages, ranging from an average of under 200 ms preceding the onset of the rounded vowel for speakers with the latest onset to over 330 ms for speakers with the earliest onset. Unlike in English and French [14], however, in the present comparison of English and German, speaker idiosyncrasy alone does not override the role of language. The significant influence of language is evidenced not only in the overall degree of coarticulation, but also in its differential response to speech rate variation (cf. [17]). It may be the case that a relatively small difference between languages is more effectively captured in our much larger set of speakers. Another possibility is that, while German and French both contrast rounding, the systemic influence on coarticulation is not a uniform one. We plan to address these issues with a parallel corpus of French speakers.

In the current study, we have focused on the temporal dynamics of lip aperture, yet anticipatory labial coarticulation may well vary by language in other dimensions. In future analyses, we plan to incorporate measures such as lip protrusion and spread to further investigate the question of how coarticulatory information is conveyed. Speakers of different languages may also follow distinct trajectories of coarticulation. While not our present focus, it is worth noting the dynamic differences within the target vowel itself in Fig. 2: in both rounded and unrounded vowels, lip aperture in German was maintained after reaching target onset, whereas in English sharp divergence between the two conditions continued well into the target vowel. More in-depth examination is thus warranted to better understand the time course of lip movement, as well as to fully explore the scope of coarticulatory variability across individuals and languages.

ACKNOWLEDGEMENTS

This work is jointly funded by the British Arts and Humanities Research Council and the German Research Foundation (AH/V002082/1).

5. REFERENCES

[1] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.
[2] S. E. Boyce, "Coarticulatory organization for lip rounding in Turkish and English," *J. Acoust. Soc. Am.*, vol. 88, no. 6, pp. 2584–2595, 1990.
[3] M. U. Caglar, A. I. Teufel, and C. O. Wilke, "Sicegar: R package for sigmoidal and double-sigmoidal curve fitting," *PeerJ*, vol. 6, p. e4251, 2018.
[4] H. Clumeck, "Patterns of soft palate movements in six languages," *J. Phon.*, vol. 4, no. 4, pp. 337–351, 1976.
[5] V. Delvaux, D. Demolin, B. Harmegnies, and A. Soquet, "The aerodynamics of nasalization in *French*," *J. Phon.*, vol. 36, no. 4, pp. 578–606, 2008.
[6] C. Draxler and K. Jänsch, "SpeechRecorder - a universal platform independent multi-channel audio recording software," in *Proc. LREC '04*, Lisbon, 2004, pp. 559–562.
[7] M. Grosvald, "Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation," *J. Phon.*, vol. 37, no. 2, pp. 173–188, 2009.
[8] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *J. Stat. Softw.*, vol. 82, no. 13, pp. 1–26, 2017.
[9] M. T. Lallouache, "Un poste visage-parole couleur: Acquisition et traitement automatique des contours des lèvres," Ph.D. dissertation, Institut National Polytechnique de Grenoble, 1991.
[10] J. Lubker, "Temporal aspects of speech production: Anticipatory labial coarticulation," *Phonetica*, vol. 38, no. 1–3, pp. 51–65, 1981.
[11] J. Lubker and T. Gay, "Anticipatory labial coarticulation: Experimental, biological, and linguistic variables," *J. Acoust. Soc. Am.*, vol. 71, no. 2, pp. 437–448, 1982.
[12] S. Y. Manuel, "The role of contrast in limiting vowel-to-vowel coarticulation in different languages," *J. Acoust. Soc. Am.*, vol. 88, no. 3, pp. 1286–1298, 1990.
[13] P. P. K. Mok, "Does vowel inventory density affect vowel-to-vowel coarticulation?" *Lang. Speech*, vol. 56, no. 2, pp. 191–209, 2013.
[14] A. Noiray, M.-A. Cathiard, L. Ménard, and C. Abry, "Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in french and english speakers," *J. Acoust. Soc. Am.*, vol. 129, no. 1, pp. 340–349, 2011.
[15] R Core Team, *R: A language and environment for statistical computing*, 2019. [Online]. Available: https://www.R-project.org/
[16] M. Redford, J. Kallay, and J. Potratz, "Temporal extent of anticipatory lip rounding in child and adult speech," *J. Acoust. Soc. Am.*, vol. 146, no. 4, p. 2925, 2019.
[17] M.-J. Solé, "Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization," *Lang. Speech*, vol. 38, no. 1, pp. 1–23, 1995.
[18] T. Ushijima and H. Hirose, "Electromyographic study of the velum during speech," *J. Phon.*, vol. 2, no. 4, pp. 315–326, 1974.
[19] G. Zellou, "Individual differences in the production of nasal coarticulation and perceptual compensation," *J. Phon.*, vol. 61, pp. 13–29, 2017.
[20] S. E. G. Öhman, "Coarticulation in VCV utterances: Spectrographic measurements," *J. Acoust. Soc. Am.*, vol. 39, no. 1, pp. 151–168, 1966.