Asymmetries in V-to-V coarticulation among harmonic and non-harmonic sequences in Khalkha Mongolian

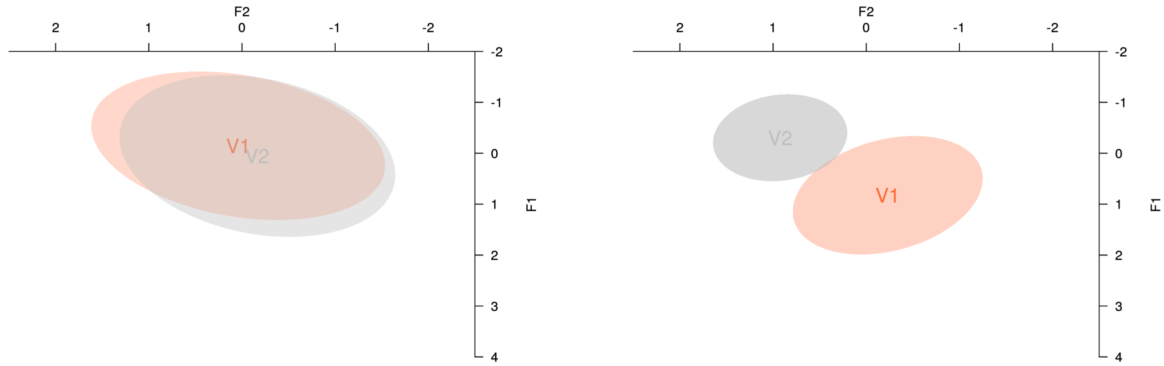
Background: Khalkha Mongolian has a robust vowel harmony system. Vowels in non-compound words share the features [ATR] and [round], with stem-internal vowels triggering harmony in suffixes in the carryover (left-to-right) direction [Svantesson, 2005]. Vowel harmony patterns emerge when listeners fail to perceptually compensate for acoustic variation due to coarticulation [Beddor et al., 2002]. Towards understanding the relationship between these processes, this study quantifies coarticulatory variation in Khalkha Mongolian VCV sequences. We find that unlike harmonic sequences, non-harmonic sequences show greater propensity of coarticulation in the anticipatory (right-to-left) direction - opposite to that of ATR and rounding harmony. We argue that in languages where vowel harmony is well-established, directionality and strength of coarticulatory propensity serves as a means of contrast preservation.

Materials and Methods: We recorded 14 female native speakers reading 4 repetitions of 59 target words, in the frame [pi X gesen] "I said X" (critical items: 59X14X4 = 3304). Targets were disyllabic words of the form $(C)V_1CV_2(C)$ [Svantesson, 2005]. A subset of these recordings was used to train an acoustic model with the Montreal Forced Aligner [McAuliffe et al., 2017] to segment and annotate the data. We analyze Lobanov-normalized F1 and F2 at vowel midpoints. To quantify coarticulatory propensity, we use linear mixed effects models to examine how well the identity of V2 explains variance in the acoustics of V1 (anticipatory) and vice-versa (carryover).

Results and discussion: For a word of the form V_1CV_2, we refer to the model explaining F1 at the midpoint of the initial vowel (V1) as F1V1t50. The initial model had vowel identities of V1 and V2 as fixed effects, and a maximal random effects structure [Barr et al., 2013] with by-subject and by-item random intercepts and random slopes for each fixed effect. We iteratively simplified the random effects structure in accordance with our hypotheses (following Bates et al. [2015]). The final model was of the form F1V1t50 $\sim$ V1 + V2 + (1|word), where V2 is the explanatory variable of interest. We find significant coarticulation in both anticipatory ($\chi^2(11)\_61.95, p = 4.016e-09$) and carryover ($\chi^2(12) = 56.79, p = 8.634e-08$) directions. Adding an interaction term harmony_type*F1V2t50 shows that patterns of anticipatory coarticulation differ across harmonic and non-harmonic sequences ($\chi^2(1) = 13.55, p = 0.00023$). To examine this asymmetry, we fit the model to harmonic and non-harmonic subsets.

In harmonic sequences (2555 observations), coarticulation is robust in both directions (anticipatory: $\chi^2(9) = 19.36, p = 0.02225$, carryover: $\chi^2(11) = 37.51, p = 9.46e-05$). Comparing effect sizes (η^2; Ben-Shachar et al. [2020]) of the explanatory variable across models shows that coarticulation is greater in the carryover direction, i.e., same direction as vowel harmony. In non-harmonic sequences (456 observations), the full model is significant for anticipatory ($\chi^2(2) = 88.89, p = 2.2e-16$) but not carryover ($\chi^2(5) = 1.5, p = 0.913$) coarticulation. η^2 values show a greater effect size in the anticipatory direction, which is opposite to that of vowel harmony.

We find greater diffusion of the acoustic vowel space in harmonic sequences (Fig.1a) compared to non-harmonic sequences (Fig. 1b). Based on these findings, we argue that coarticulation interacts with grammatical knowledge, especially the notion of contrast. Instead of sharing an inimical relationship with the latter, coarticulation serves to preserve contrast among vowels in non-harmonic sequences in Khalkha Mongolian.



(a) Harmonic (b) Non-harmonic

Figure 1: Acoustic vowel space in harmonic and non-harmonic sequences: V1 (red) and V2 (grey)

References

Jan-Olof Svantesson. *The Phonology of Mongolian*. Oxford University Press Inc, New York, 2005.

Patrice Speeter Beddor, James D. Harnsberger, and Stephanie Lindemann. Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4):591 – 627, 2002. ISSN 0095-4470. doi: http://dx.doi.org/10.1006/jpho.2002.0177. URL http://www.sciencedirect.com/science/article/pii/S0095447002901774.

Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. Montreal forced aligner: Trainable text-speech alignment using kaldi. pages 498–502, 08 2017. doi: 10.21437/Interspeech.2017-1386.

Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3): 255–278, 2013.

Douglas Bates, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*, 2015.

Mattan S. Ben-Shachar, Daniel Lüdecke, and Dominique Makowski. effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56):2815, 2020. doi: 10.21105/joss.02815. URL https://doi.org/10.21105/joss.02815.