COARTICULATION AND CONTRAST IN A VOWEL HARMONY SYSTEM: COARTICULATORY PROPENSITY IN KHALKHA MONGOLIAN V-C-V SEQUENCES

Auromita Mitra[1], Meghavarshini Krishnaswamy[2], Indranil Dutta[3]

New York University[1], University of Arizona[2], Jadavpur University[3]
auromita.mitra@nyu.edu, mkrishnaswamy@arizona.edu, indranildutta.lnl@jadavpuruniversity.in

ABSTRACT

Vowel harmony has been understood to emerge when listeners fail to perceptually compensate for acoustic variation due to coarticulation. Assuming such an account, what explains the maintenance of non-harmonic domains in the grammar? Towards understanding this, we examine coarticulation within a synchronic system with well-established patterns of harmony and non-harmony. In Khalkha Mongolian, vowels in non-compound words share the features [ATR] and [round], harmony operating in the carryover (left-to-right) direction. The high-front vowel /i/ does not participate in harmony, giving "non-harmonic" VCV sequences. We quantify coarticulatory variation by comparing dependencies in first- and second-formant frequencies (F1&F2) of vowels in harmonic vs non-harmonic VCV sequences. Unlike the former, the latter show greater coarticulation in the anticipatory (right-to-left) direction—opposite to that of vowel harmony. /i/, which is transparent to harmony, demonstrates high coarticulatory resistance [1]. We argue that in systems where vowel harmony is well-established, synchronic patterns of coarticulatory propensity serve to limit feature-sharing in non-harmonic domains.

Keywords: Khalkha Mongolian, vowel harmony, coarticulation, vowels, formants

1. COARTICULATION AND VOWEL HARMONY

Vowel harmony is a grammatical feature of languages whereby sequences of vowels (contiguous or non-contiguous) within a certain domain, e.g. the word, obligatorily have the same value for certain phonological features. This is a productive rule, and leads to predictable variation in vowels depending on phonological context. This could be left-to-right: *carryover*, where features of a preceding vowel affect those of the following vowels, or right-to-left: *anticipatory* where the converse happens. Independent of this grammatical process, an essential feature of speech is that sounds are not produced in discrete sequences. Rather, articulatory gestures overlap during connected speech, and this kind of *coarticulation* results in acoustic variation. Coarticulation could be carryover when the gestures of an earlier target perturb the latter in a mechanico-inertial fashion, or anticipatory, where articulatory planning for a subsequent vowel affects the production of an earlier vowel. To extract stable phonological categories, from such a variable signal, listeners must perceptually compensate for the acoustic variation resulting from coarticulation [2, 3]. Given the parallels between this kind of physiological-psychoacoustic process and grammatical harmony processes, vowel harmony has been understood to diachronically emerge when listeners fail to perceptually compensate for acoustic variation due to coarticulation, thereby perceiving the coarticulated variation as part of the phonological feature of the category [4, 3]. Under this hypothesis, the greater persistence of carryover coarticulation would favor carryover harmony, leading to left-to-right harmony systems, while the persistence of anticipatory coarticulation would result in to anticipatory, right-to-left harmony systems. Of note here is that coarticulation is usually understood to be a non-voluntary by-product of articulation, and therefore arguably continues to exist even after harmony has developed. Assuming the account of causation, once a psycho-acoustic process such as the perceptual persistence of e.g. carryover coarticulation has been set in motion, what prevents the system from coarticulating further, assimilating more features, and eventually resulting in a radical system of vowel copying in the grammar? While such systems exist (e.g. Telugu; [5]), other harmony systems appear to be in a synchronic state of equilibrium, where harmony operates over certain features and segments, but not others. Specifically, most harmony systems also contain some form of *non-harmony*– sequences or segments that do not participate in harmony, either by completely ignoring it (transparent segments), or blocking it (opaque segments). How do such sequences persist in synchronic systems, given the continued existence of coarticulation in the direction of harmony, and the system's proclivity for its perceptual non-compensation? That is, what checks does the language system put in place to prevent the spillover of the very processes that initiated harmony, into its grammatically non-harmonic domains? Towards understanding this, this project examines how coarticulation functions within a system with well-established and robust vowel harmony. We measure the acoustic consequences of coarticulation in harmonic and non-harmonic sequences, and compare the patterns of coarticulatory propensity in these sequences.

2. VOWEL SYSTEM OF KHALKHA MONGOLIAN

Khalkha Mongolian has seven monophthongal vowel categories, classified as non-pharyngeal (+ATR) and pharyngeal (-ATR) [6]:

	[+ATR]	[-ATR]	neutral
high	u	ʊ	i
non-high	e, o	a, ɔ	

Table 1: Monopthongs in Khalkha Mongolian, classified by harmony class

The inventory has four diphthongs that are contrastive: ai, ɔi, ui, ʊi. For the purposes of vowel harmony, these pattern identically to their corresponding monophthongal counterparts (a, ɔ, u, ʊ). There is also a vowel length distinction that does not play a role in vowel harmony, and is not discussed it here.

Vowel harmony

Vowels in non-compound words must have identical value for the [ATR] feature. A subset of vowels (non-high: e, o, a, ɔ) also show rounding harmony, whereby any instance of e, o, a, ɔ in a non-initial position must match the [round] feature of the initial vowel. We focus on ATR harmony. Vowel harmony proceeds left-to-right (carryover harmony): ATR is contrastive only in word-initial position, and vowels in non-initial positions must match the value of the initial vowel. This means that the language has two types of words: [+ATR] and [-ATR]. /i/ does not show phonological alternation; it can occur in both [+ATR] and [-ATR] environments. Thus, although the phonetic realization has been shown to vary slightly across contexts, causing researchers to analyze it as an allophone [6], the category is phonologically unaffected by harmony. Moreover, when /i/ intervenes between the initial and subsequent vowels, harmony proceeds across it – it is 'transparent' to harmony. In our study, we focus exclusively on disyllablic words. Thus, words in which the second vowel is /i/ do not demonstrate harmony, resulting in 'non-harmonic' sequences.

2.1. Questions and Hypotheses

[ATR] (advanced tongue root) is a phonological feature, whose acoustic correlates include the first formant frequency (F1) [7, 8]. Our aim is to examine whether patterns of coarticulation differ between words that involve grammatical ATR harmony (harmonic sequences), and those that don't (non-harmonic sequences). Since we are using acoustic data, we want to identify acoustic consequences of coarticulation in the acoustic dimensions of interest (here, formant frequencies), and quantify those. One approach is to think about coarticulation as constraining the random variability of individual articulatory gestures– if the articulation of sound B is physiologically affected by that of sound A, then we expect that each time a speaker produces the sequence AB, some of the variability in the articulation of B will be explained by the presence of A (compared to, e.g., CB). Assuming (simplistically) that articulatory variability has acoustic consequences, we operationalizie V-to-V coarticulation as constraints on variability in vowel acoustics. We assume that if the articulation of vowel 1 affects that of vowel 2, then some of the variability in the acoustics of 2 will be explained by the identity of 1. We will treat the degree of dependency as a proxy for the extent of coarticulatory effect. Within a V1CV2 sequence, a greater dependency of V1 on V2 suggests greater anticipatory coarticulation, while the opposite suggests greater carryover coarticulation. This gives us the *directionality* of coarticulation within a sequence. The aim of the current study is to examine how coarticulation behaves in synchronic data in a system where vowel harmony (and its domains) have already been grammaticalized. Therefore, we ask whether the extent and directionality of coarticulation differ between harmonic and non-harmonic VCV sequences.

Since vowel harmony in Khalkha Mongolian is in the left-to-right carryover direction, we expect an overall privileging of carryover coarticulation in harmonic domains of the language. This would follow from previous accounts of the diachronic development of harmony. However, this does not necessarily translate to an assumption about a greater extent of coarticulation per se. The accounts in, e.g. Ohala (1994) [4] are perceptual. Carryover coarticulation could be privileged by means of lesser perceptual *compensation*, even if the extent of coarticulation were equal in both directions.

Our main interest in this study is in the non-harmonic sequences. We expect that if the system actively preserves non-harmony in certain domains, then such domains will differ in coarticulatory patterns, compared to domains in which harmony has developed (diachronically) as a result of coarticulation. Exactly what devices a language might employ to preserve grammatical domains is not something we have a priori assumptions about. However, we hypothesize that carryover coarticulation in non-harmonic sequences would be restricted in some way. This could manifest as either: (i) extent: coarticulation is primarily in the carryover direction, but the magnitude is smaller than in harmonic sequences; (ii) directionality: coarticulation has a different pattern across harmonic and non-harmonic words.

3. METHODOLOGY

3.1. Participants and materials

Participants were 14 first-language speakers of Khalkha Mongolian, recruited at the EFL University campus, Hyderabad, where they were enrolled in a short-term English language course. All the participants were from Ulaanbataar, between 25 and 40 years of age. We recorded speakers individually in a sound-attenuated booth, reading a set of words in frame sentences, and a passage, all presented on a computer screen in the Cyrillic script that is used commonly to read and write Khalkha Mongolian. Only the former was analyzed for this study. Each participant read 4 blocked repetitions of 59 target words, in the frame [*pi X gesen*] "I said X" (critical items: $59 \times 14 \times 4 = 3304$). Targets were disyllabic non-compound words of the form (C)V1CV2(C), taken from a book on the description of the sound system of Mongolian [6]. We included words from examples throughout the text, creating a set with all allowed sequences of vowels in initial (v1) and non-initial (V2) positions. Words in which /i/ occupied the V2 position were classified as "non-harmonic", and all others as "harmonic". We removed one item from an initial set of 60 targets, because it was found that it is pronounced as monosyllabic, in spite of having two vowel characters in the orthography.

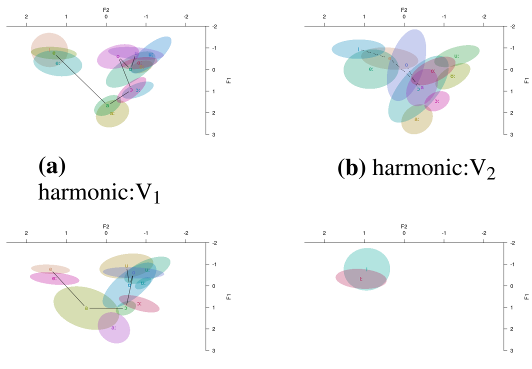
3.2. Acoustic analysis

A subset of recordings (including reading passages and frame sentences) was used to train an acoustic model with the Montreal Forced Aligner [9] to segment and annotate the data. F1 and F2 data were extracted with a customized Praat script [10]. 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).

4. RESULTS

4.1. Vowel space diffusion

Plotting F1 and F2 values of vowels in the initial (V1) and non-initial (V2) position (figure 1) shows the following patterns: for harmonic sequences, the vowel space for V2 is more diffused than V1, suggesting greater variability due to coarticulation. Since the non-harmonic sequences contain only a subset of the vowel space (which are in complementary distribution to V1), the V2 vowel space is not informative. However, the vowel space for V1 shows that the low front vowel /a/ and the high back vowel /u/ are more diffused in the initial position of a non-harmonic sequence, compared to a harmonic sequence. This suggests differences in patterns of variability between harmonic and non-harmonic subsets. This is probed in the statistical analyses.



(a) harmonic:V_1 **(b)** harmonic:V_2 **(c)** non-harmonic:V_1 **(d)** non-harmonic:V_2

Figure 1: Steady-state vowel spaces for harmonic and non-harmonic vowel sequences in Khalkha Mongolian

4.2. Statistical analyses

For any vowel, we expect F1 and F2 values to be predicted by the phonological identity of the vowel. To quantify coarticulation, we used mixed effects models [11] in R (version 4.1.1) [12] to probe how well the formant frequency of a segment (dependent variable) is predicted by the identity of the *other* vowel in the word (independent variable), giving a measure of coarticulatory variability. Moreover, we examine whether this tendency differs between harmonic and non-harmonic words by comparing effect sizes of the independent variable in each model using the *effectsize* [13] package in R. For each model, we started with the maximal random effects structure motivated by the study design: by-item and by-speaker random slopes. This was iteratively simplified until the model converged [14, 15]. For all mixed models, the alpha criterion was set at $|t| > 2$. Table-1 summarizes the final model and output for F1. Here t5 refers to the formant measure at the steady state of the vowel. The results indicate robust coarticulation in both directions, with larger effects in the carryover (left-to-right) direction.

5. DISCUSSION AND CONCLUSION

The statistical analyses demonstrate an asymmetry in coarticulatory propensity between harmonic and non-harmonic sequences in Khalkha Mongolian. Specifically, while harmonic sequences show greater coarticulation in the carryover direction, parallel to that of harmony, non-harmonic sequences show the opposite pattern: coarticulation is greater in the anticipatory direction, which is opposite to that of harmony. We suggest that greater anticipatory coarticulation in these sequences counteracts the overall tendency of the system to privilege carryover coarticulation perceptually, and thus serves to check the gradual development of harmony in grammatically non-harmonic domains. Note that in Khalkha Mongolian, the only vowel that occurs in the non-initial position in non-harmonic sequences if /i/. The lack of significant carryover coarticulation in these sequences suggests that the segment has high coarticulatory resistance compared to the other vowels. This phonetic 'special-ness' tallies with research that has ascribed a special status to the segment in the phonological inventory of Mongolian, where it functions as a 'default' vowel in diphthongs and epenthetic processes [6]. These facts are compatible with two possible threads of interpretation: (i) the language system maintains non-harmony by developing higher coarticulatory resistance in the segments that occur in non-harmonic domains; (ii) non-harmonic domains develop diachronically around segments that have higher coarticulatory resistance for independent reasons. More cross-linguistic investigations in a variety of harmony systems, particularly those with non-harmonic domains involving multiple segments, will allow us to tease these apart. The findings of the present study suggest that in vowel harmony systems, articulatory and psychoacoustic processes might themselves function as checks to maintain synchronic equilibrium and limit harmony to certain domains.

Harmony type	Direction	Model fixed effects	ChiSq	Df	p	effect size (η^2)
harmonic	anticipatory	F1V1t5 ~ V1+**V2**	17.174	9	0.04606 *	0.322
	carryover	F1V2t5 ~ V2+**V1**	34.131	11	0.003443 ***	0.536
non-harmonic	anticipatory	F1V1t5 ~ V1+**V2**	100.87	1	< 2.2e-16 ***	0.133
	carryover	F1V2t5 ~ V2+**V1**	133.41	10	< 2.2e-16 ***	0.174

Table 2: Model outputs for coarticulation in F1, compared to a null model (bold)

Harmony type	Direction	Model fixed effects	ChiSq	Df	p	effect size (η^2)
harmonic	anticipatory	F2V1t5 ~ V1+**V2**	9.3863	9	0.4024	0.191
	carryover	F2V2t5 ~ V2+**V1**	22.79	11	0.01892 *	0.404
non-harmonic	anticipatory	F2V1t5 ~ V1+**V2**	110.57	1	< 2.2e-16 ***	0.146
	carryover	F2V2t5 ~ V2+**V1**	74.809	10	5.182e-12 ***	0.101

Table 3: Model outputs for coarticulation in F2, compared to a null model (bold)

6. REFERENCES

[1] D. Recasens, J. Fontdevila, and M. D. Pallarès, "Velarization degree and coarticulatory resistance for /i/ in Catalan and German," *Journal of Phonetics*, vol. 23, no. 1–2, pp. 37 – 52, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S009544709580031X

[2] P. S. Beddor and R. A. Krakow, "Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation," *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2868–2887, 11 1999. [Online]. Available: https://doi.org/10.1121/1.428111

[3] P. S. Beddor, J. D. Harnsberger, and S. Lindemann, "Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates," *Journal of Phonetics*, vol. 30, no. 4, pp. 591 – 627, 2002.

[4] J. J. Ohala, "Towards a universal, phonetically-based, theory of vowel harmony," in *Third International Conference on Spoken Language Processing*, 1994.

[5] I. Dutta, S. Irfan, and K. Harsha, "Using ANNs for vowel identification from V-to-V coarticulation in non-harmonic VCV sequences," in *15th Conference on Laboratory Phonology, Ithaca, NY*, 2016. [Online]. Available: http://labphon.org/labphon15/long_abstracts/LabPhon15_Revised_abstract_279.pdf

[6] J.-O. Svantesson, A. Tsendina, A. Karlsson, and V. Franzén, *The phonology of Mongolian*. OUP Oxford, 2005.

[7] C. G. A. Starwalt, "The acoustic correlates of atr harmony in seven-and nine-vowel african languages: A phonetic inquiry into phonological structure," Ph.D. dissertation, The University of Texas at Arlington, 2008.

[8] P. Olejarczuk, M. A. Otero, and M. M. Baese-Berk, "Acoustic correlates of anticipatory and progressive [atr] harmony processes in ethiopian komo," *Journal of Phonetics*, vol. 74, pp. 18–41, 2019.

[9] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced aligner: Trainable text-speech alignment using kaldi." in *Interspeech*, vol. 2017, 2017, pp. 498–502.

[10] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2009. [Online]. Available: http://www.praat.org

[11] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[12] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[13] M. S. Ben-Shachar, D. Lüdecke, and D. Makowski, "effectsize: Estimation of effect size indices and standardized parameters," *Journal of Open Source Software*, vol. 5, no. 56, p. 2815, 2020.

[14] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of memory and language*, vol. 68, no. 3, pp. 255–278, 2013.

[15] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, "Parsimonious mixed models," *arXiv preprint arXiv:1506.04967*, 2015.