THE ACOUSTICS OF VOWEL HARMONY IN SAKHA

May Pik Yu Chan, Jianjing Kuang

University of Pennsylvania
pikyu@sas.upenn.edu, kuangj@sas.upenn.edu

ABSTRACT

Sakha (Yakut) is a Turkic language that is typologically rare to have both backness and roundness vowel harmony. This study aims to investigate the phonetic realizations of vowel harmony conditioned by vowel features (e.g., backness and roundness), monophthong or diphthong, as well as the distance of the target syllable from the trigger syllable. One native speaker of Sakha produced inflected or conjugated lexical items that were up to five syllables in length in a carrier phrase. Results show that F2 is the most important acoustic cue involved in both backness harmony and roundness harmony. And there is a salient effect of syllable position for vowel harmony: target vowels are gradually centralized when further away from the trigger; and this position-conditioned gradient vowel harmony effect is further modulated by vowel quality features: gradient backness harmony is observed for both monophthongs and diphthongs, but gradient roundness harmony is only observed for diphthongs.

Keywords: Yakut, Sakha, Vowel Harmony, Turkic language

1. INTRODUCTION

Sakha (or Yakut) is a North Siberian Turkic language spoken in Sakha/Yakutia. The eight monophthongs (/i, e, y, œ, ɯ, ɑ, u, o/) in Sakha differ in [+/-high], [+/-back] and [+/-round] dimensions respectively, with /i, ɯ, e, ɑ/ being unrounded counterparts to /y, u, œ, o/, and /i, ɯ, y, u/ being high counterparts to /e, ɑ, œ, o/. There are also four diphthongs (/ie, ɯɑ, yœ, uo/), which contrast in backness and roundess. As an important part of the morphosyntactic process, vowel harmony operates from left to right within an agglutinated word. Vowels consistently harmonize in backness with the word's initial vowel. Rounding harmony applies for all [-high, +round] triggers, but [+high, +round] vowels and diphthongs may only spread their roundness feature to [+high] vowels. There has been limited work looking at the phonetic realization of Sakha vowel harmony, except for [1] showing that Yakut-Russian bilinguals have robust vowel harmony adaptations in Russian word production.

One point of interest in studying vowel harmony is the comparison between alternating and non-alternating vowels. While vowel assimilation is found to be categorical and complete for some cases [2, 3, 4], gradient assimilatory effects between vowels are also fairly common [5, 6, 7]. For example, some other Turkic languages (e.g. Kazakh, Uyghur) have demonstrated centralization in vowel backness harmony patterns [5, 6]. This study focuses on describing gradient vowel harmony in Sakha as Sakha offers two extra dimensions of interest:

First, Sakha has both backness and roundness harmony. Unlike backness harmony, which is common across Turkic languages, systematic rounding harmony is much rarer typologically and often occurs with extremely limited contexts (see [5] for a detailed review). Subsequently, prior work on Turkic vowel harmony has focused on backness harmony only [5]. In comparison, Sakha's rounding harmony rules are less restrictive. In Sakha, low round vowels (/o/ and /œ/) may spread their roundness feature to both high and low vowels, but high round vowels (/y/ and /u/) may only spread roundness to following high vowels. The presence of both backness and relatively less restrictive roundness harmony allows us to explore whether gradient harmony effects are feature specific. The relatively freer rounding harmony rule also means that the entire vowel inventory could be examined, unlike languages like Uzbek, Kazakh, and Uyghur, where non-high round vowels may only occur in initial syllables [5].

Second, diphthongs (/ie, ɯɑ, yœ, uo/) also participate in harmony in Sakha. Specifically, diphthongs as triggers pattern like high vowels in Sakha. Words like /duorɑt/ 'to make an echo' exemplifies this, since the second syllable in the word is unrounded. This is consistent with height restrictions of rounding vowel harmony in Sakha. Furthermore, verbal suffixes in Sakha also contain diphthongs that are underlying only specified in height but not backness and roundness (e.g. /bɑr/ 'to go'; /bɑrɯɑm/ 'go-1SG.FUT'). This allows us to test whether diphthongs follow the same vowel harmony mechanisms.

The goal of this study is to offer a better understanding of the phonetic realizations of Sakha vowel harmony, as instrumental studies have not been done. We aim to capture the differences between alternating and non-alternating vowels in Sakha vowel harmony addressing the following: (i) whether there is gradient realization of harmony conditioned by syllable position in Sakha (ii) whether backness and roundness harmony follow the same harmony pattern and (iii) whether monophthongs and diphthongs pattern similarly.

2. METHODS

2.1. Procedure

One native speaker of Sakha in her mid-twenties participated in the study. Apart from Sakha, she is also fluent in English and Russian, and has beginner-level Italian and Japanese.

The speaker was given lists of lexical items containing nouns and verbs. The speaker learned to associate English translations of sentences with corresponding affixes. For instance, when given the English translation ‘their leathers’, she would produce [tirii-lere] ‘leather-3pl.poss’ instead of the full phrase [kiniler tiriilere] ‘their leathers’. Similarly for verbs, when given the translation ‘I will smile right now’, she would produce only the conjugated target word [yœr-yœm] ‘smile-1sg.fut’, and when given the translation ‘I will smile tomorrow’, she would produce [yœr-yœʁym] ‘smile-1sg.fut’. The speaker was asked to produce each target word in a carrier phrase [biligin _ dien tɯlɯ et] ‘Say the word _ now’, such that the target word was in a focused sentence medial position, avoiding phrase final prosody. This also avoids effects that may favor an interpolation-based account for harmony. The word domain edge would not align with the utterance edge, which may or may not be featurally specified in backness based on default articulatory settings. The carrier phrase and the root of the target word is presented in Cyrillic, and the full set of target word forms were presented in sets as English translations on the same slide as the target root word and carrier phrase. Each word was repeated three times.

Recording sessions were conducted online via Zoom, with the speaker recording the audio locally to her computer to avoid issues with data bandwidth filters. The data were recorded to the default microphone of a MacBook Air 2017 at a sampling rate of 32kHz and 32-bit float. Recordings were made over ten sessions; each lasted between 10 to 20 minutes.

2.2. Stimuli

The participant was presented with a set of target words containing all vowel quality contrasts in Sakha. This includes all eight monophthongs (/i, ɯ, e, ɑ, y, u, œ, o/) and four diphthongs (/ie, ɯɑ, yœ, uo/). Although Sakha does exhibit vowel length contrasts, words containing a short vowel as the initial syllable were chosen if possible. The comparative morpheme (-tAAʁA) also includes a long vowel. Two monosyllabic roots and two disyllabic roots per vowel were included for nouns, apart from the vowel /œ/, which had an extra disyllabic root token. There were 49 noun roots in total. For verbs, one monosyllabic and one disyllabic root were included per vowel, resulting in 24 verbal roots. This adds up to 73 roots in total. To the extent possible, monosyllabic roots and the first syllable of disyllabic roots ended in either a sibilant, liquid, or rhotic, with exceptions for coda-less lexical items, or those ending with a stop or /h/.

For nouns, seven forms were elicited: (i) singular, (ii) plural, (iii) first singular possessive, (iv) third plural possessive, (v) plural with a comparative, (vi) accusative, (vii) plural with an accusative marker. There is backness harmony that targets all non-initial vowels. This results in the following four alternations in backness: i∼ɯ, y∼u, e∼œ, ɑ∼o. With respect to roundness, since rounding harmony is parasitic on height, there are relatively fewer high-rounded vowels compared to unrounded vowels in the elicited forms.

		/kus/ ‘duck’
a. SG	-Ø	kus
b. PL	-LAr	kustɑr
c. SG-1POSS	-(I)m	kuhum
d. PL-3POSS	-LAr-A	kustɑrɑ
e. PL-COMPAR	-LAr-TAAʁAr	kustɑrdɑɑʁɑr
f. ACC	-(n)I	kuhu
g. PL-ACC	-LAr-I	kustɑrɯ

Table 1: Examples of elicited nominal forms

Nine forms of the verbs were elicited: (i) the root in isolation (often corresponding to the imperative form), (ii) two forms of the first singular future, corresponding to an immediate or distant future, (iii) two forms of the third singular future, also corresponding to an immediate or distant future, (iv) first plural future, (v) third plural future, (vi) third singular progressive, (vii) third singular past.

As diphthongs consist of a high vowel followed by a low vowel of the same backness and roundness features, and they are also subject to backness harmony therefore, there are two alternating pairs in terms of backness (ie∼ɯɑ, yœ∼uo). We elicited fewer rounded than unrounded tokens for the same reason of parasitic rounding harmony.

		/bɑr/ 'to go'
a. root	-Ø	bɑr
b. 1SG.FUT	-IAm	barɯɑm
c. 1SG.FUT	-IAʁIm	barɯɑʁɯm
d. 3SG.FUT	-IA	barɯɑ
e. 3SG.FUT	-IAʁA	barɯɑʁɑ
f. 1PL.FUT	-IAxpIt	barɯɑxpɯt
g. 3PL.FUT	-IAxtArA	barɯɑxtɑrɑ
h. 3SG.PROG	-Ar	barɑr
i. 3SG.PST	-PItA	barbɯtɑ

Table 2: Examples of elicited verbal forms

Some complications to the word list include that when disyllabic verbs end in a coda /s/, the preceding vowel is deleted. For example, [umus] 'to dive' has a first singular future form of [umsuom] or [umsuoʁum] and not *[umusuom] nor *[umusuoʁum]. Two verbs have word final /s/. One verb [ɯarɯj] 'to become sick' has irregular conjugations and was excluded from the analysis. For monosyllabic roots, words are maximally four syllables in length, for disyllabic roots, words are maximally five syllables in length. This is the case for both nouns and verbs.

2.3. Segmentation and Measurement

Sound files were segmented manually in Praat [8]. The beginning and end of each vowel were segmented to the onset and offset of high-frequency energy. In cases where the vowel is adjacent to a sonorous consonant like laterals, rhotics and glides, the point of amplitude change was used to separate the vowel and the consonant. In cases where the token has noticeable background noise, the entire token (at the word level, not vowel level) was removed. After removing unwanted tokens, a total of 4948 vowel tokens entered the analysis. Among these tokens 3977 were monophthongs (short and long vowels included), 971 were diphthongs.

The first three formants as well as vowel duration were measured. Measurements were taken at 25, 50, and 75% of the vowel's entire duration using a Praat script. The "To Formant (burg)" function was used. The data were analyzed in R. For monophthongs, formants measured at 50% of the vowel duration were used for analyses. For diphthongs, formants measured at 25% and 75% of the vowel duration were analyzed. Since only one speaker's production was analyzed, no normalization was done.

3. RESULTS AND DISCUSSIONS

We first begin by exploring the overall vowel space in Sakha. The left facet of figure 1 shows the F1 and F2 of all monophthongs (short and long vowels included) measured at 50% of the vowel's total duration. It could be observed that /ɑ/ is located at the central portion of the vowel plot, suggesting that the [back] property of /ɑ/ is more phonological than phonetic. The facet on the right shows a corresponding F2-F3 plot. Generally, vowels differing in roundness do not differ much from the F3 dimension. Instead, vowel pairs contrasting in roundness differ mostly on the F2 dimension. Therefore, subsequent analysis will focus on F2.

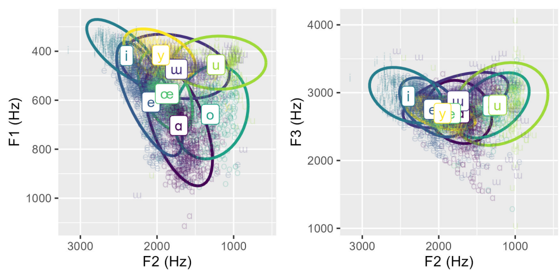

Figure 1: F1-F2 and F2-F3 vowel space plot.

We now turn to the effect of syllable position on vowel realization in vowel harmony. A linear mixed effects regression model (lme4 package [9]) was fitted to examine how vowel harmony is conditioned by syllable position, as well as the vowel features (backness and roundness) of the trigger and target vowels. Statistical significance is determined by Satterthwaite approximation via the lmerTest package [10]. Specifically, we predict F2 based on the fixed effects of initial vowel backness (which would be identical to target vowel backness), target vowel height, syllable position counting from the trigger, initial vowel roundness (which may differ from target vowel roundness), and whether the target vowel matches in roundness with the trigger (henceforth "roundness match"). The following two-way interactions are included: initial vowel backness and vowel height, initial vowel backness and syllable position, vowel height, and syllable position, initial vowel roundness and syllable position, and roundness match and syllable position. Three-way interaction effects of initial vowel backness, vowel height, and syllable position are also included in the model. Random intercepts of the preceding consonant, the following consonant and the target vowel are included, and by-target vowel random slopes for syllable number are also included. Results show significant main effects of initial vowel backness (Reference = Front vowels; β = -1084.95, SE = 153.65, df = 3.110, p = 0.005), of initial vowel roundness (Reference = Unround vowels; β = -495.24, SE = 108.86, df = 3.134, p = 0.018), and of roundness match (Reference = Different; β = -369.56, SE = 112.12, df = 3.52, p = 0.036), confirming that both backness and roundness harmony modulates F2. Crucially, there is a significant interaction effect of initial vowel backness and syllable position (β = 183.24, SE = 37.85, df = 3.3, p = 0.013), suggesting that the later the syllable, the higher the F2 for back vowels. In other words, back vowels are fronted/centralized depending on syllable position. This effect can be clearly seen in Figure 2, which summarizes the F2 of vowels by position and backness harmony pairing. Both front and back vowels shift towards neutralization. However, there is no significant interaction between initial vowel roundness and syllable position, suggesting that the gradient neutralization effect is specific to backness for monophthongs.

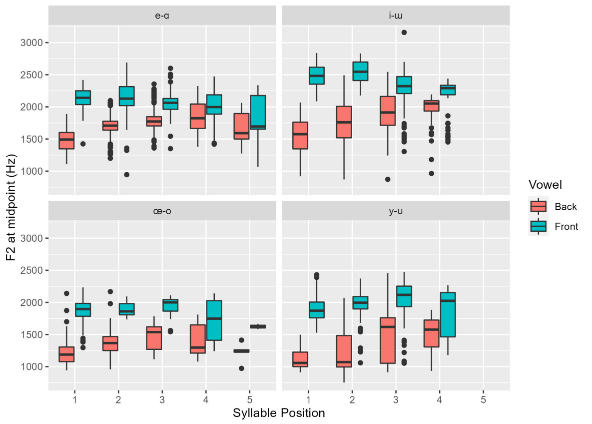

Figure 2: F2 at 50% by backness pairs (monophthongs).

Moving onto diphthongs, figure 3 visualizes F2 by syllable position and backness harmonic pairings in diphthongs. Facets correspond to F2 measurements taken at 25% and 75% of the vowel's total duration. Effects of F2 centralization appear to be generally present as well.

Two linear regression models are run to evaluate the effects of syllable position, initial vowel backness, and vowel roundness on F2 measured

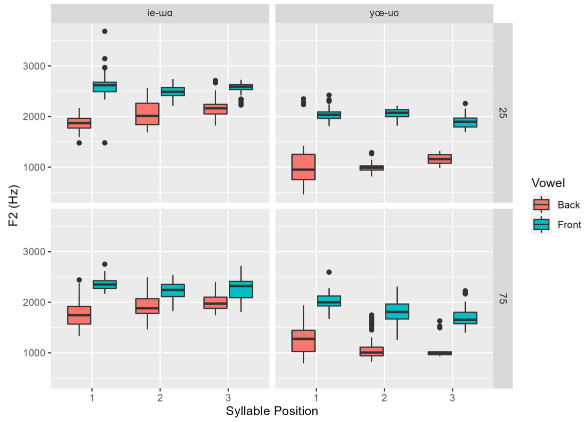

Figure 3: F2 at 25 and 75% of the vowel's duration by backness pairs (diphthongs).

at either 25% or 75% of the vowel's duration. For both models, the main effects of initial vowel backness, syllable position, and target vowel roundness are included. Two-way interaction effects between syllable position and initial vowel backness, and between syllable number and target vowel roundness are also included. Random intercepts by target-vowel are incorporated into the model. Due to a considerably lesser number of diphthongs in the dataset, other main and random effects are not included to ensure model convergence. Reference levels are identical to the monophthong models. Results find significant interactions of initial vowel backness and syllable position (25%: β = 140.36, SE = 16.10, df = 964.15, p < 2e-16; 75%: β = 111.254, SE = 17.74, df = 964.30, p = 5.37e-10) and initial vowel roundness and syllable number (25%: -63.12, SE = 16.33, df = 964.00, p = 0.0001; 75%: β = -162.1, SE = 17.99, df = 964.00, p < 2e-16) for both models. Therefore, for diphthongs, gradient centralization is conditioned by both backness and roundness.

4. CONCLUSIONS

This study aims to provide a better understanding of the phonetic realizations of Sakha vowel harmony. We show that F2 is the most important acoustic cue involved in both backness harmony and roundness harmony. And there is a salient effect of syllable position for vowel harmony: target vowels are gradually centralized when further away from the trigger; and this vowel harmony is further modulated by vowel quality features: gradient backness harmony is observed for both monophthongs and diphthongs, but gradient roundness harmony is only observed for diphthongs.

5. REFERENCES

[1] L. Vasilyeva, "Production and perception of vowel harmony: Phonological predictors of ratings and on-line adaptations of russian vowels in Yakut (Sakha)," PhD dissertation, University of Alberta, 2017.

[2] S. Hess, "Assimilatory effects in a vowel harmony system: an acoustic analysis of advanced tongue root in Akan," *Journal of Phonetics*, vol. 20, no. 4, pp. 475–492, 1992.

[3] B. Gick, D. Pulleyblank, F. Campbell, and N. Mutaka, "Low vowels and transparency in Kinande vowel harmony," *Phonology*, vol. 23, no. 1, pp. 1–20, 2006.

[4] S. G. Guion, M. W. Post, and D. L. Payne, "Phonetic correlates of tongue root vowel contrasts in Maa," *Journal of Phonetics*, vol. 32, no. 4, pp. 517–542, 2004.

[5] A. McCollum, "Gradience and locality in phonology: Case studies from Turkic vowel harmony," PhD dissertation, University of California, San Diego, 2019.

[6] ——, "Gradient morphophonology: Evidence from Uyghur vowel harmony," in *Proceedings of the Annual Meetings on Phonology*, vol. 6, 2019.

[7] ——, "Vowel harmony and positional variation in Kyrgyz," *Laboratory Phonology*, vol. 11, no. 1, 2020.

[8] P. Boersma, "Praat: doing phonetics by computer," http://www.praat.org/, 2006.

[9] 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.

[10] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.