

MIT Open Access Articles

Phonetic correlates to Khalkha Mongolian vowel contrasts: duration, formants and voice quality

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Kenstowicz, Michael J. 2023. "Phonetic correlates to Khalkha Mongolian vowel contrasts: duration, formants and voice quality."

As Published: <https://doi.org/10.1007/s10831-023-09262-w>

Publisher: Springer Netherlands

Persistent URL: <https://hdl.handle.net/1721.1/152920>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution





Phonetic correlates to Khalkha Mongolian vowel contrasts: duration, formants and voice quality

Michael J. Kenstowicz¹

Received: 1 October 2022 / Accepted: 6 April 2023
© The Author(s) 2023

Abstract The paper reports the results of an analysis of acoustic correlates to phonological contrasts in vowel length, quality and pharyngeal width with data from five native speakers of the Khalkh dialect. The study replicates Svantesson's (1985) discovery of a chain-shift rotation of the round vowels in their presumed evolution from Classical Mongolian. Analysis with VoiceSauce (Shue et al. in: Proceedings of the ICPHS XVII, 1846–1849, 2009) finds that in addition to the first formant, the harmonics-to-noise ratio is a reliable indicator of the ATR versus RTR contrast that has been hypothesized to be the basis of the language's vowel harmony.

Keywords Vowel harmony · Chain shift · ATR/RTR · Harmonics-to-noise ratio (HNR) · Duration ratio

1 Background

The vocalic contrasts and harmony processes found in Khalkha Mongolian have been discussed for over 50 years in the Western (generative) tradition starting with Lightner (1965), Binnick (1969, 1980), Odden (1980), Steriade (1979), Hamp (1980), Rialland & Djamouri (1984), Svantesson (1985), and more recently Svantesson et al. (2005), Vaux (2009), and Ko (2011, 2019), among others. An important milestone in this line of research was Svantesson's (1985) discovery that the presumed front rounded vowels /ü/ and /ö/ of the Classical language are realized as back vowels in several modern varieties including Khalkh. In some Mongolian dialects these derived back vowels merged with the corresponding original back

✉ Michael J. Kenstowicz
kenstow@mit.edu

¹ Department of Linguistics, Massachusetts Institute of Technology, 77 Massachusetts Ave., MA, Cambridge 02139, USA

round vowels /u/ and /o/; but in others, including Khalkh, merger was avoided by lowering /u/ and /o/ in the vowel space. Svantesson made the intriguing proposal that the lowering of the original back vowels reflected a reorganization of the phonological opposition between /ü/ and /ö/ versus /u/ and /o/ from [±back] to a contrast in pharyngeal width. The /ü,ö/ vowels were reassociated with the Advanced Tongue Root [ATR] feature and /u,o/ with Retracted Tongue Root [RTR]. It should be noted that Svantesson (1985) actually uses the feature [pharyngeal] from Wood (1979) implemented by the hyoglossus and pharyngeal constrictors; but he aligns this distinction with the [ATR] feature introduced by Stewart (1967) and associated with pharyngeal width via positioning of the tongue root in Lindau's (1975, 1979) X-ray studies of Akan. Svantesson's hypothesis was supported by the classification of the vowels as 'female' lax (*eme*) versus 'male' tense (*ere*) in the native Mongolian literature. According to Chinggeltei (1979) the tense vowels are produced with the upper part of the throat tensed and the tongue root retracted. The positions in the vowel space implied by the backing and lowering 'rotation' of the Modern Mongolian round vowels have been accepted in most subsequent studies. More recently a controversy has arisen as to whether it is the contrast in tongue-body backness or the tongue-root posture that represents the original contrast: Vaux (2009), Ko (2011, 2019), and Ko et al. (2014) see the Mongolian tongue-root expression of the contrast as part of a more general Northeast Asian sprachbund also observed in Tungusic and Kamchatkan languages, and possibly Korean as well, and treat the oral-cavity front-back tongue-body articulation as an innovation, primarily of the Turkic languages.

While the phonetic correlates underlying the ATR harmony contrasts in (West) African languages have been relatively well studied (Ladefoged 1964, Lindau 1979, Fulop et al. 1998, Hess 1992, Guion et al. 2004, Przewdziecki 2005, Starwalt 2008, among others; see Ko 2019 for a helpful summary), the same is not true for the East Asian languages. The few investigations that do exist are based often on just a single speaker with minimal indication of the words comprising the corpus, the recording methods, and the statistical analyses of the data; Rialland and Djamouri (1984): seven vowel phonemes recorded in isolation (9–12 repetitions) by one male Khalkh speaker; Svantesson (1985): 10 words with three repetitions from two inner Mongolian male speakers and one male Khalkh speaker; Kang and Ko (2011): one male speaker of Western Buriat and one female speaker of Tungusic Ewe. This paper is an attempt to help fill this gap, at least with respect to the Mongolian variety with the largest number of native speakers—the Khalkh dialect of the capital Ulaanbaatar whose population of 1.5 million comprises roughly one half of the entire nation according to a recent (2019) census (Wikipedia.org). Our results broadly but not completely align with earlier findings. We add value to the investigation of the phonetic basis of the Khalkh vowel contrasts by replicating earlier results, especially as they relate to harmony, with a larger number of speakers representing several generations, a larger number of test items, and consistent statistical analyses of the three parameters of contrast defining the Mongolian vocalic inventory: length, height and backness, and the volume/width of the pharyngeal cavity. Our study breaks new ground by investigating a larger number of phonetic correlates to the contrast between the female /e,ö,ü/ and male /a,

o,u/ vowels. Like almost all previous studies, our data are acoustic: duration (for phonological length), formants (for vowel quality), and phonation (for the pharyngeal-laryngeal articulators). The articulatory correlates to the phonological contrasts remain an outstanding gap that needs to be filled in future research.

2 Phonemic vowel system

In the Mongolian phonological literature the vowel phonemes are transliterated from the Cyrillic alphabet with the customary five /i,u,e,o,a/ letters as well as /ɣ/ and /ɵ/ for the rotated front rounded vowels of the Classical language. For the sake of typographic convenience and in conformity with the traditional literature, we represent these elements with the umlaut sign (i.e. as /ü/ and /ö/, respectively) rather than a particular IPA symbol since their phonetic realization is a point of controversy and a research question. This seven-vowel system is crossed for length to yield fourteen vowel phonemes. There is also a diphthong /aj/ which, for our speakers, has coalesced into a long low/mid front vowel /æ:/. We put this vowel aside for the moment and return to it in Sect. 5.2. As seen in Table 1, the Mongolian vowel system (Classical or Modern Khalkh) is distinguished by three basic phonological features: front versus back or ATR versus RTR, high versus nonhigh, and round versus nonround. The /i/ lacks a back/RTR counterpart.

As in most Altaic languages, vowel harmony in Khalkha Mongolian is a root-based progressive assimilation process (there are no prefixes). The /i/ is a ‘neutral’ vowel that does not alternate while the front/ATR /e,ü,ö/ alternate with their back/RTR counterparts /a,u,o/ as a function of the rightmost stem vowel. The neutral /i/ is transparent so that harmony appears to skip over this segment. When it is the sole stem vowel, then suffixes take the /e,ü,ö/ set vowels. In contrast to the Classical language, Khalkha Mongolian has an additional labial-harmony process. It is also progressive and rounds a following nonhigh vowel when the stem contains a round vowel which is also nonhigh (/ö/ or /o/). Svantesson (1985) notes that labial harmony only occurs in Mongolian varieties that have shifted to the tongue-root/pharyngeal based harmony. Table 2 illustrates the vowel-harmony alternations with ablative case forms; the ablative suffix has a lexically determined $n \approx 0$ alternation.

Table 1 Phonological contrasts comprising the Khalkh vowel inventory

front/ATR		back/RTR	
_____	_____	_____	_____
- round	+ round	- round	+ round
+ high i	ü		u
- high e	ö	a	o

Table 2 Ablative case forms illustrating vowel harmony

ar-a:s	'back'
pex-e:s	'ink'
xir-e:s	'dirt'
pug-a:s	'demon'
ot-no:s	'star'
xög-nö:s	'shame'
tüt-ne:s	'devil'
xajr-a:s	'love'

3 Outline of the study

As mentioned above, our study was designed to investigate the acoustic correlates to the phonological contrasts of length, vowel height and backness, and pharyngeal width in the Khalkh dialect. We first describe the methods: our speakers and the word list comprising our corpus; the recording equipment and types of data analysis. We then present our results in three successive sections: length, formant measures, and phonation/voice quality, each followed by discussion of the findings. The paper closes with a brief summary of the overall results.

4 Methods

Our data was provided by five native speakers who were born and raised in the Mongolian capital Ulaanbaatar. They represent three generations: younger (two females c. 25 years of age), middle-aged (one male and one female both c. 50 years of age), and elderly (one female c. 75 years of age). A set of five monosyllabic CVC nouns for each of the fourteen vowel phonemes as well as for the coalesced diphthong /aj/ > /æ:/ was assembled (see Appendix 3). Before recording we went over the word list with each speaker to make sure they were familiar with the words. In a couple of cases, a word was unfamiliar to one of our speakers and so an alternative was substituted. The 75 words comprising the speaker's list were randomized and then successively elicited in four case forms: nominative (identical to stem), accusative *-iig*, ablative *-AAs*, and comitative/possessive *-taj*. The speakers each read the same randomized word list with the minor substitutions noted above. Our data corpus thus comprises 1500 words: 14+1 vowels * 5 word stems * 4 case forms * 5 speakers. The recordings were made in a sound-insulated booth with a Shure SM10A Unidirectional Head-Worn Dynamic Microphone and a USB Pre 2 Preamp at a sampling rate of 44.1 kHz and quantizing resolution of 16 bits. The resulting sound files were analyzed in Praat (Boersma & Weenick 2019) with segmentation and labeling by hand based on auditory and visual inspection of the spectrograms for large changes in the waveform indicative of the transition to and from the nuclear vowel with steady F1. Duration and formant measurements based on vowel midpoint were collected by Praat scripts. The resultant data were stored in

excel files and then analyzed in R (4.1.3) with mixed-effects linear regressions (speaker and word as random factors). At the suggestion of a reviewer, speaker age and gender were entered as fixed effects for the phonation correlates. A t-value of 2.0 was taken as the threshold of statistical significance. Formant measures were normalized with Z-scores based on the mean and standard deviation for each speaker. Presumed tongue-root correlates (H1-H2, H1-A1, H1-A2, HNR05, HNR15, HNR25, NHR35, as well as f0, F1, F2, and F3) were investigated with VoiceSauce (Shue et al. 2009) based on a 20 ms window centered on the vowel's midpoint.

5 Results

5.1 Duration

Our data file permits several research questions to be posed concerning the length of the stem vowel. First, vowel length can be affected by the number of syllables in the word. In particular, the vowel of a monosyllable is expected to be longer than in a disyllable. Indeed, this factor is often phonologized as a so-called “minimality effect”, as in Kansai Japanese where monomoraic words of the standard (Tokyo) dialect like [yu] ‘hot water’ are lengthened to two moras when spoken in isolation [yuu] (Ito 1990). Hence, other things being equal, we may expect the stem vowel in our Mongolian data to be longer in the zero-suffixed nominative case form where the word is monosyllabic compared to the same vowel in the accusative and comitative cases formed with the suffixes *-iig* and *-taj*, respectively, where the word is disyllabic. Another factor influencing the duration of a vowel is syllable structure: vowels are typically longer in open syllables compared to closed syllables (Maddieson 1997). The diachronic sound changes and synchronic phonological processes of Open-syllable Lengthening and Closed-syllable Shortening are reflexes of these phonetic factors. Other things being equal, we may expect the vowel of a CVC stem to be longer in the accusative where the stem-final consonant parses as an onset [CV.Ciig] compared to the comitative where that consonant closes the syllable of the stem [CVC.taj]. Finally, we can ask how the long versus short phonemic contrast might be affected by these three case-form contexts: will the phonemic long-short duration ratio remain stable under these associated shifts in word length and syllable structure?

Figure 1 indicates the log-transformed durations for the stem vowels in three case forms (nominative, accusative, comitative) as a function of the long versus short phonological contrast. In the bare stem nominative context, long vowels are overall more than twice the duration of short vowels: raw score means 350 versus 136 ms (ratio = 2.57). The stem vowels are shortened in the disyllabic accusative *-iig*; but the ratio between the long and short vowel categories remains comparable: raw score means 237 versus 90 ms (ratio = 2.63). The consonant-initial comitative *-taj* closes the stem vowel's syllable leading to shorter stem-vowel durations; but the ratio between the long and short vowel categories remains largely unchanged: raw score means 186 versus 78 ms (ratio = 2.38). Also, the /æ:/ vowel derived from

coalescence of earlier /aj/ has mean durations that are very close to the long-vowel categories: raw score means = 349 ms (nominative), 246 ms (accusative), 177 ms (comitative). A mixed-effects linear regression over the logs of the stem-vowel durations for the three case forms found significant differences between the baseline accusative and the nominative on the one hand (est = 0.15, $t=4.57$) and between the accusative and comitative on the other (est = -0.10 , $t=-2.86$); see Appendix 1 Model 1.1 for details. The same pattern of results obtained when the data was split as a function of the phonemic length of the stem vowel: long vowels (est=0.16, $t=11.42$ for nominative and est = -0.10 , $t=-7.13$ for comitative) and short vowels (est = 0.19, $t=7.33$ for nominative and est = -0.06 , $t=-2.27$ for comitative); see Appendix 1, Models 1.2 and 1.3 for details. Our data are thus consistent with the hypotheses that changing the word's length from one to two syllables shortens the vowel of the stem as does changing the stem vowel's syllable from open to closed. The test also suggested that the former effect is greater in magnitude and reliability compared to the latter. Finally, the ratio between the long versus short vowel contrast was largely stable across these different prosodic contexts.

5.2 Formants

Figure 2 shows the placement of the 15 stem vowels (normalized) in F1*F2 space based on their means in the citation form nominative case. F1 is indicated on the vertical axis and F2 on the horizontal axis.

The following observations can be made concerning this excel plot. First and most importantly the /ü/ and /ö/ counterparts to the Classical Mongolian presumed front round vowels have retracted to the back vowel region with F2 values that are on average lower than /a/. Second, the counterparts to the Classical Mongolian back round vowels have lowered in the vowel space so that /u,u:/ are below /ü,ü:/ and even below /ö,ö:/, respectively. The vowels /o,o:/ are much lower in the vowel space. Our results thus accord well with Svantesson's (1985) original findings and confirm the striking chain-shift rotation of the round vowels that he discovered. It

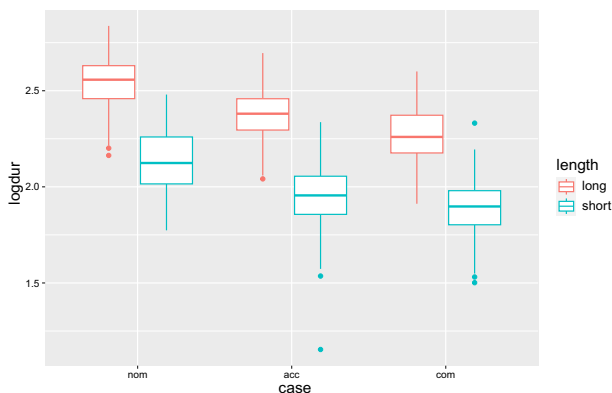


Figure 1 Log-transformed long versus short stem-vowel durations (ms) in three case forms

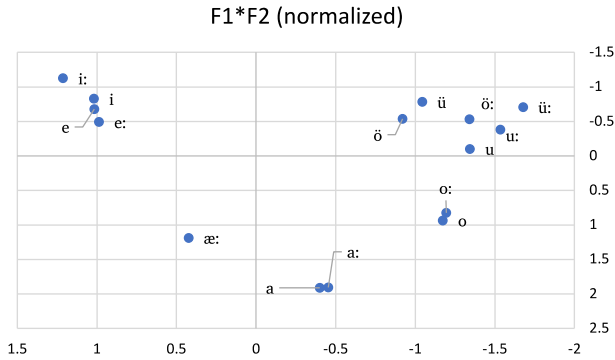


Figure 2 F1(vertical axis) * F2 (horizontal axis) plot of stem vowel means from citation form nominative

should be noted that Rialland & Djamouri (1984) independently reached much the same conclusion. Third, the short vowels /ü/ and /ö/ are located more towards the center of the vowel space compared to their long-vowel counterparts. On the other hand, there is little difference between the short versus long positions of the /i/, /a/ and /o/ vowels. This disparity might be taken to indicate that the backing of /ü/ and /ö/ is a sound change in progress with the short vowels lagging behind their long counterparts. Fourth, the front mid vowel /e/ is very close to /i/, especially when short: no significant differences were found for normalized F1 (est = -0.06, t = -0.99) and normalized F2 (est=0.01, t=0.04); see Appendix 1 Model 2 for details. The long vowel /e:/ is quite a bit higher in the vowel space compared to the back vowels /o,o:/, which sit roughly halfway between the rotated /u,u:/ and /a,a:/. Finally, the coalesced [æ:] from /aj/ is roughly halfway between /e:/ and /a:/ in the normalized F1*F2 space; it occupies a position comparable to the lowered /o/ in vowel height and is well separated from /a:/. Since stem vowel /a/ and /aj/ > [æ:] both choose the same suffixal RTR harmonic alternants (cf. [pag-a:s] ‘team’ abl. and [xæ:r-a:s] ‘love’ abl.), the nonhigh RTR region has introduced a [±back] distinction in addition to the [±round] contrast that separates /a/ from /o/. The plot in Fig. 3 shows the realization of the vowel of the ablative suffix as a function of whether the stem vowel is /a:/ or [æ:] < /aj/. The former belong to a more compact region while the latter are more widely dispersed. A mixed-effects regression test predicting F1 and F2 for the ablative -a:s as a function of stem-vowel backness found a significant difference in normalized F2 (est=0.16, t=3.67) and a marginal difference in normalized F1 (est = -0.22, t=-2.04). See Appendix 1 Model 3 for details. These data suggest that the [±back] feature that distinguishes /a:/ from [æ:] < /aj/ is being spread as an autosegmental dependent of RTR so that when the -aas alternant is derived by tongue-root harmony, it splits into a minor front versus back difference, as suggested in the plot. The data are admittedly quite noisy for the [æ:] < /aj/ case and so this is a point that requires further study.

As a final observation, we note that the Khalkh rotated /ü,ü:/ and /ö,ö:/ vowels have F2 values considerably lower than /a/ (especially when long). In this respect they differ from the front round vowels of Turkish (Korkmaz & Aytug 2018) as well

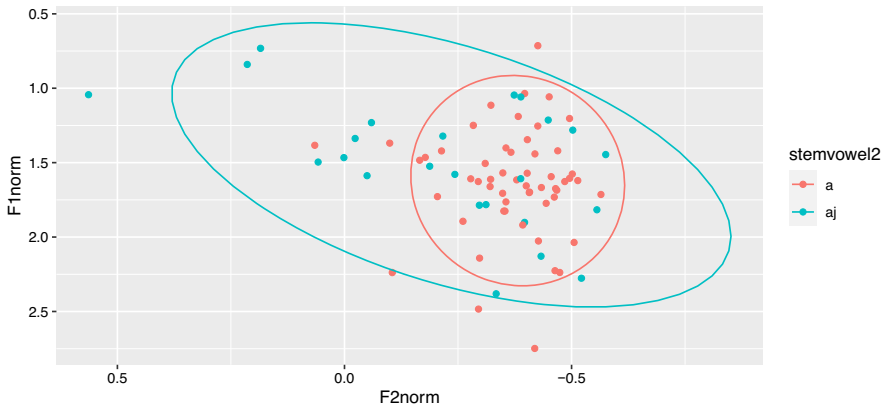


Figure 3 Ablative suffixal vowel as function of /a:/ vs [æ:] < /aj/ stem vowel

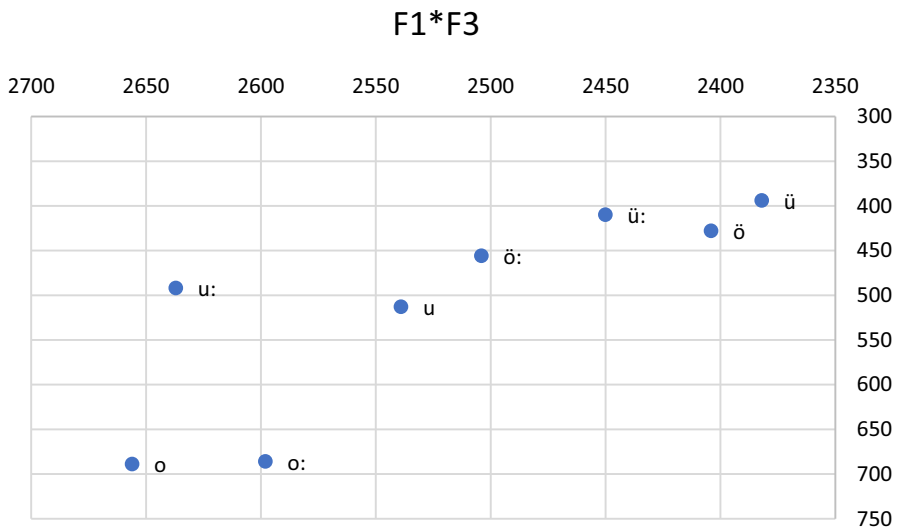


Figure 4 F1 (vertical axis) *F3 (horizontal axis) plot of round stem vowel means from citation form nominative

as French and German (Strange et al. 2007), which have higher values. Figure 4 plots the round vowels in our corpus in F1*F3 space and shows that the third formant does a much better job of isolating the /*ö,ö:,ü,ü:/* set than the second formant does. A mixed-effects linear regression found the difference in normalized F3 as a function of the /*ö,ü/* versus /*o,u/* contrast to be significant (est 0.713, $t = 6.24$). For details see Appendix 1 Model 4. A similar F3 /*i/ > /u/ > /y/; /e/ > /o/ > /ø/* hierarchy appears in French (Law 2008), German (Piroth et al. 2015), and Turkish (Korkmaz and Aytug 2018) suggesting that the lower F3 of the Khalkh /*ü/* and /*ö/* vowels relative to /*u/* and /*o/* may reflect their former front vowel status.

5.3 Voice quality

In order to investigate potential spectral offset and phonation correlates to the ATR-RTR contrast in our Khalkha Mongolian data, we utilized the VoiceSauce (Shue et al. 2009) Matlab program. The program's options were set to their default values: the Straight algorithm (Kawahara et al. 1999) for the harmonics and the Snack Sound Toolkit (Sjölander 2004) for the formant frequencies. The analysis was restricted to the stem vowels in our data. The input sound files for each speaker were accompanied by a Praat textgrid marking off 20 ms intervals centered on the midpoint for each stem vowel in the corpus: hence 75 vowels * 4 case forms * 20 measurements (one per millisecond) * 5 speakers. In a few cases (c. 1% of the data) the program failed to make a measurement and so the resultant file contained 29,664 lines of data. The data were reduced by taking the average of the stem vowel across the 20 measurement points per word and then further reduced by taking the average across the four case forms. This resulted in a file of 377 items across the five speakers. The resultant data was then normalized with Z-scores based on the mean and standard deviation for each speaker. Two of the resultant items were removed because their Z-scores exceeded three standard deviations and were considered outliers. The data was analyzed with mixed-effects linear regression tests in R (4.1.3) with the ATR (baseline) versus RTR specification of the stem vowel as the main predictor variable (coded as VQ). At the suggestion of a reviewer, the gender and age of the speaker were also entered into the models as predictors. Female was the baseline for gender, and age had three levels: old, middle (baseline), and young. Speaker and word were specified as random intercepts and, in full models, by-speaker and by-word random slopes for VQ were called as well. In cases where the full model did not converge, the random slope by speaker (and in a few cases the random slope by word as well) were dropped in order to obtain convergence.

Three groups of dependent variables were regressed on the VQ + gender + age predictor variables. First, the spectral drop-off factors that measure the difference in amplitude (dB) between the first and second harmonics (H1H2c) as well as the difference in amplitude between the fundamental and the harmonics located nearest the first, second, and third formants (H1A1c, H1A2c, H1A3c) were considered. Based on the ATR harmony literature, the Khalkh RTR vowels /a,o,u/ should be associated with less drop-off than the corresponding ATR vowels since ATR vowels are expected to have weaker amplitudes in the higher regions of the spectrum and hence they will diminish the value of H1 less. In other words, given that ATR is the baseline category in our regression models, the change to RTR should be negative (less drop-off) for these variables. The second class of factors we considered involve the harmonics-to-noise ratio (HNR) at various levels of the speech spectrum: HNR05 for 0–500 Hz, HNR15 for 0–1500 Hz, HNR25 for 0–2500 Hz, and HNR35 for 0–3500 Hz. If the RTR vowels involve pharyngeal narrowing we might expect greater turbulence and hence an increase in the noise factor compared to the vowels that advance the tongue root. Again, given that ATR is the baseline in our regression models, increased noise will lower the value of HNR and so we expect that a switch from ATR to RTR be associated with negative coefficients. The final group of factors we considered are the fundamental (sF0) and first three formants. We

already know from Fig. 2 that the RTR vowels are lower in the Khalkh vowel space compared to their ATR counterparts and so again, given that ATR is the baseline, we expect the switch to RTR to result in positive coefficients for sF1. The ATR vowel-harmony literature indicates that F1 is the most widespread and reliable correlate to the ATR-RTR contrast so we might also expect this factor to be associated with greater magnitude and reliability compared to the other predictors. The literature is far from unanimous on the effect of the second formant; but given that the ATR /e,ö,ü/ are located in the more anterior region of the vowel space, we might expect that a switch to RTR should be associated with a negative value in the regression coefficient for sF2 if tongue-root posture is affected by tongue-body position. We saw above that the third formant distinguished between the ATR /ö,ü/ and RTR /o,u/ and so we anticipate that a switch from baseline ATR in VQ should be associated with positive coefficients for sF3. Finally, given that higher vowels are associated with greater inherent f0 (Lehiste 1970), the switch from ATR to RTR should be negative for sF0.

5.3.1 Results

Table 3 summarizes the results of the regressions tests over the VoiceSauced Khalkha Mongolian data. For these tests the predictor variables received default dummy coding. (We ran a separate set of tests in which the predictors were centered with simple coding; the overall pattern of results was largely the same; see Appendix 2). The number of observations was 374 grouped by word (77) and speaker (5). In the second column, 2 means that the model converged with by-speaker and by-word random slopes for VQ (for example, $H1H2cZ \sim VQ + age + gender + (1|speaker) + (1|word) + (1+VQ|speaker) + (1+VQ|word)$); 1 indicates models with by-word random slopes and 0 are models with no random slopes. It

Table 3 Summary of mixed-effects linear regression tests of full data set

Correlate	Slopes	VQ	Estimate	Std. error	df	t-value	Pr(t)	sig
H1H2cZ	2	RTR	-0.75	0.32	4.43	-2.29	0.07	
H1A1cZ	2	RTR	0.77	0.19	4.92	3.95	0.01	*
H1A2cZ	2	RTR	0.79	0.25	6.88	3.13	0.01	*
H1A3cZ	2	RTR	0.17	0.22	7.82	0.76	0.46	
HNR05Z	0	RTR	0.61	0.14	205.12	4.30	2.61e-05	***
HNR15Z	2	RTR	-0.85	0.15	45.85	-5.41	2.15e-06	***
HNR25Z	1	RTR	-0.67	0.16	58.27	-4.07	0.0001	***
HNR35Z	1	RTR	-0.56	0.16	49.93	-3.50	0.0009	***
sF0Z	1	RTR	-0.92	0.14	52.05	-6.39	4.55e-08	***
sF1Z	1	RTR	1.49	0.12	43.08	12.24	1.27e-15	***
sF2Z	0	RTR	-0.37	0.11	357.23	-3.38	0.0007	***
sF3Z	1	RTR	-0.12	0.16	58.54	-0.77	0.44	
sF3Z,rnd	1	RTR	0.71	0.11	44.60	6.24	1.41e-07	***

turned out that in none of these models were there significant effects for gender or age.

Turning now to the results, for the first group of factors testing spectral slope there was either no significant effect (H1H2c, H1A3c) or an effect in the opposite direction from what was expected (H1A1c, H1A2c), where a switch of VQ from baseline ATR to RTR was associated with positive coefficients indicating greater drop-off. This result suggests that the Khalkh contrast between /e,ö,ü/ and /a,o,u/ is modal versus RTR rather than ATR versus modal or binary ATR versus RTR. For the noise factor there is a relatively strong and reliable (***) decrease in HNR (negative coefficients) indicating that the /a,o,u/ vowels are associated with greater noise and thus lower the harmonics-to-noise ratio. This finding held for all but the lowest region of the spectrum where the association went in the opposite direction. This result would be consistent with pharyngeal narrowing (generating turbulence) for the /a,o,u/ set and dovetails with the presence of spectral drop-off for these vowels. As anticipated, there is a very strong and reliable effect of sF1 being positively associated with the RTR vowels. The pattern with sF2 went in the opposite direction from sF1 and is consistent with the more anterior position of /i,e,ö,ü/ in the vowel space. As for sF3, no significant association was found when sF3 was regressed on VQ for the entire vowel set. However, when the same test was run on just the round vowels /ö,ü/ and /o,u/ (205 observations) then there was a significant positive effect for RTR. Finally, sF0 had significant negative coefficients for RTR consistent with the association between vowel height and f0.

A second set of regressions with the same settings (ATR as the VQ baseline and speaker and gender as random factors) was run over a subset of the data comprised of vowels that are harmonic partners for RTR harmony: i.e. a~e; o~ö, and u~ü;

Table 4 Summary of mixed-effects linear regression tests of reduced data set

	Slopes	VQ	Estimate	Std. error	df	t-value	Pr(t)	sig
H1H2cZ	0	RTR	-0.70	0.12	78.5	-5.79	1.33e-07	***
		Age-old	0.25	0.12	241.4	2.08	0.03	*
		Age-young	0.32	0.10	242.43	3.20	0.001	**
H1H2cZ	1	RTR	-0.73	0.12	54.40	-5.76	4.03e-07	***
H1A1cZ	2	RTR	0.91	0.25	5.08	3.63	0.014	*
H1A2cZ	2	RTR	0.66	0.23	8.84	2.84	0.019	*
H1A3cZ	2	RTR	-0.01	0.23	8.44	-0.06	0.94	
		Age:old	0.24	0.10	236.7	2.30	0.02	*
HNR05Z	2	RTR	0.34	0.18	25.9	1.85	0.07	.
HNR15Z	2	RTR	-0.78	0.16	46.04	-4.67	2.62e-05	***
HNR25Z	2	RTR	-0.56	0.20	33.40	-2.75	0.009	**
HNR35Z	2	RTR	0.66	0.23	8.84	2.84	0.01	*
sF0Z	2	RTR	-0.95	0.16	22.09	-5.83	7.12e-06	***
sF1Z	2	RTR	1.32	0.17	12.62	7.65	4.32e-06	***
sF2Z	2	RTR	-0.42	0.15	55.52	-2.81	0.006	**
sF3Z	2	RTR	0.23	0.26	6.16	0.89	0.40	

number of observations: 302, groups: word 60; speaker 5. The results, seen in Table 4, are generally congruent with the larger data set seen in Table 3: positive spectral drop-off for H1A1c and H1A2c, negative coefficients for HNR15 and HNR25, positive coefficients for sF1, and negative coefficients for sF0 and sF2.

The excel plots below indicate how the individual harmonic pairs perform with respect to the dependent variables (raw scores). The error bars are standard errors. For spectral drop-off, the low /a,e/ and mid /o,ö/ vowel pairs pattern together (except for H1A3c) while the high /u,ü/ go in the opposite direction (H1H2c) or find no difference (H1A1c, H1A2c) (Figs. 5, 6, 7, 8).

For the harmonics-to-noise ratios, the /a,o,u/ set vowels contribute more turbulence than /e,ö,ü/ at levels above 500 Hz, lowering the HNR values at all three levels of vowel height (Figs. 9, 10, 11, 12).

For sF0 and sF1 the vowels separate cleanly into the /a,o,u/ versus /e,ö,ü/ sets at each level of vowel height. For sF2 the nonround vowels separate cleanly while for the round vowels the VQ factor is confounded with vowel length. As was seen in Fig. 2, above, the short ATR /ö,ü/ are positioned more toward the center of the vowel space compared to their long vowel counterparts. A regression test over the round vowels found a significant positive effect for sF2 as a function of length:short (est = 0.79, $t = 6.55$); see Appendix 1 Model 5 for details. Lastly, for sF3 the VQ patterns differ as a function of rounding as noted in Table 4 (Figs. 13, 14, 15, 16).

Finally, univariate binomial logistic regression models were run in R with glmer to see how well the individual factors classified the reduced data set (302 observations) into the ATR /e,ö,ü/ versus RTR /a,o,u/ categories. It turned out that the only significant predictor was F1: $z=3.31$, $AIC=71.3$. Among the noise factors, none of the HNR predictors by themselves reached significance. But when combined into a single model, HNR05 and HNR15 covering the lower region of the spectrum, each reached significance: HNR05 ($z=2.62$) and HNR15 ($z=-2.59$). The AIC for this model was 65.2. A model combining the noise factors HNR05 and

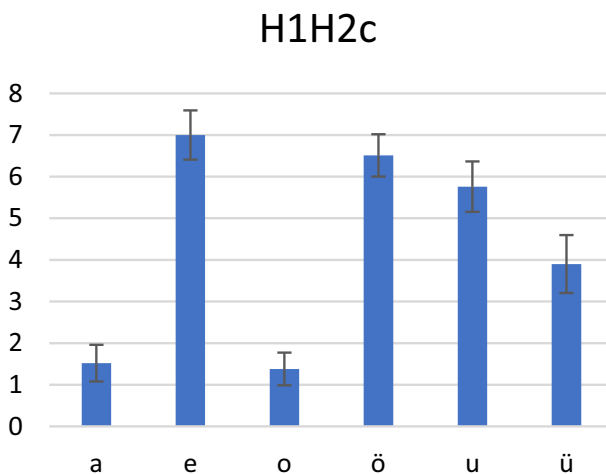


Figure 5 H1H2c for Khalkha Mongolian vowels

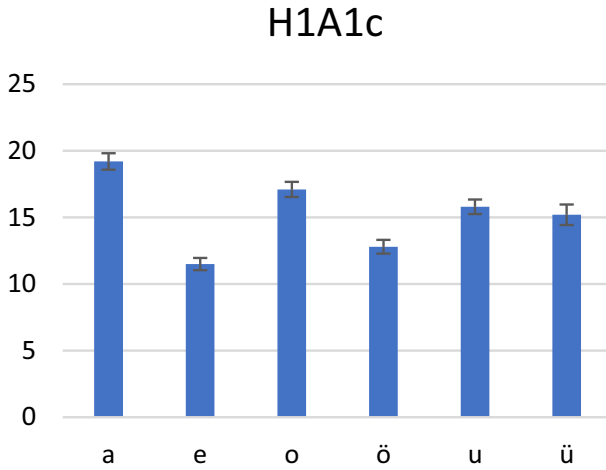


Figure 6 H1A1c for Khalkha Mongolian vowels

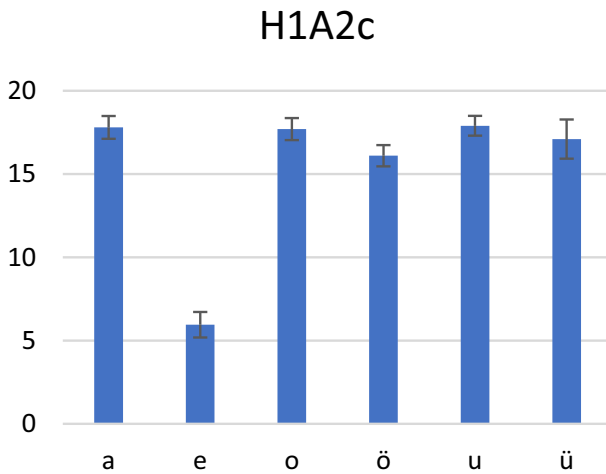


Figure 7 H1A2c for Khalkha Mongolian vowels

HNR15 with F1 had a significantly lower AIC of 59.7 (anova Chisq=15.61, $p=0.0004$) and was the best fitting model that could be constructed.

In order to investigate further the role of the various predictors for the ATR versus RTR distinction in our data, we employed R’s Recursive Partitioning algorithm in the mlr package. The random search parameter was set to 200 iterations with fivefold cross validation. Three different models were investigated. A full model employed both the voice-quality predictors of spectral slope and the harmonics-to-noise ratios as well as the formant-based sF1 and sF2 (plus sF0). Two reduced models employed either the voice-quality predictors alone or the formant-based predictors alone. The models are detailed below in Table 5 along with their

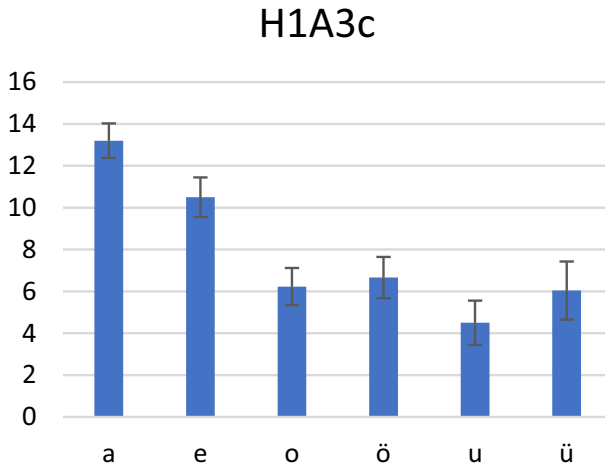


Figure 8 H1A3c for Khalkha Mongolian vowels

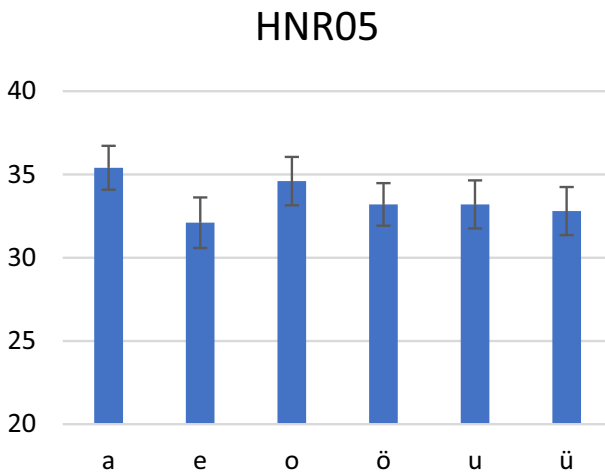


Figure 9 HNR05 for Khalkha Mongolian vowels

mean misclassification error (mmce) scores. The results indicate that the Formants-only model performed at essentially the same rate as the Full model with both correctly classifying roughly 90% of the data. The misclassification rate for the model predicting the RTR versus ATR distinction based on the voice-quality predictors alone was substantially higher with an error rate of over 20%.

In sum, based on the location of a vowel in $F1 * F2$ space c. 90% of the data can be correctly classified with respect to the ATR versus RTR distinction. The voice-quality factors are less reliable in this regard; but they nevertheless succeed in classifying roughly 80% of the data correctly. There is thus considerable collinearity between the vowel quality and voice quality predictors.

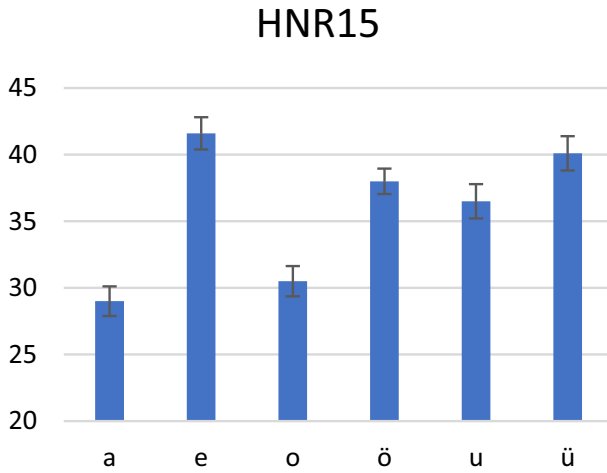


Figure 10 HNR15 for Khalkha Mongolian vowels

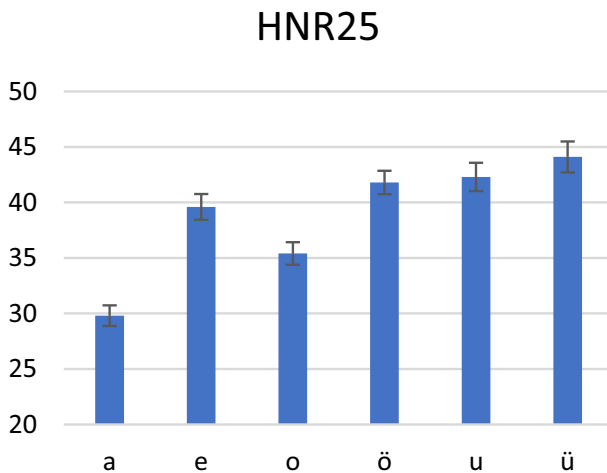


Figure 11 HNR25 for Khalkha Mongolian vowels

5.3.2 Discussion

The most notable result of our investigation into the phonetic parameters underlying the Khalkha Mongolian /a,o,u/ versus /e,ö,ü/ contrast relevant for vowel harmony is that the voice-quality factors of spectral drop-off and acoustic turbulence align with the formant-based F1 and F2 predictors. While not as great in magnitude or reliability, the voice-quality parameters nevertheless do play a significant role. This general result is compatible with and supports the Laryngeal Articulator Model (LAM) of voice quality proposed in Esling (2005) and the phonological potentials model (PPM) of Esling et al. (2019) which it grounds. The LAM is presented as an

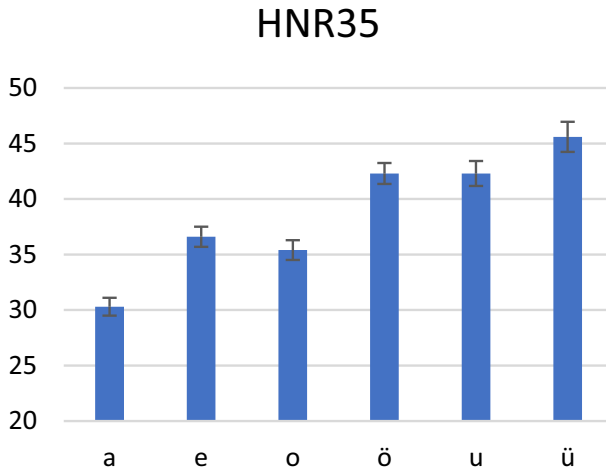


Figure 12 HNR35 for Khalkha Mongolian vowels

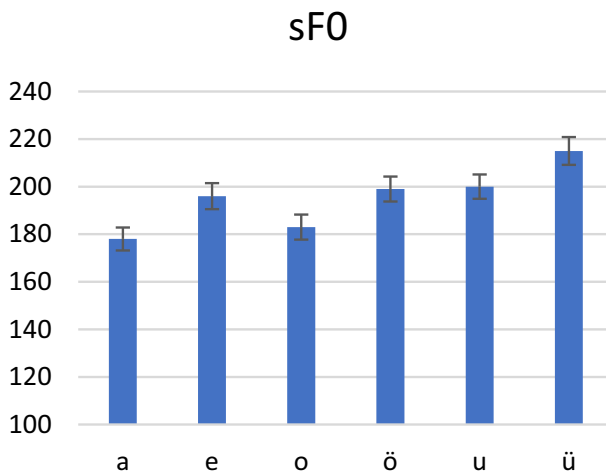


Figure 13 sF0 for Khalkha Mongolian vowels

alternative to the traditional source-filter model of the vocal tract in which the larynx provides the basic sound wave which is then filtered and shaped by articulators located primarily in the oral cavity. LAM recognizes and attempts to model the contribution of articulators in the pharynx and larynx in shaping speech sounds. With regard to the typology of tongue-root harmony, Esling (2005) notes that Lindau's (1979) X-rays for Akan were not able to thoroughly resolve the lower region of the pharyngeal cavity (a point made by Svantesson 1985 as well) though she notes the presence of laryngeal raising for the [-ATR] vowel set raising the question of whether and, if so, how such nonlingual articulations are involved in ATR-harmony. Esling sheds light on this question with photographic

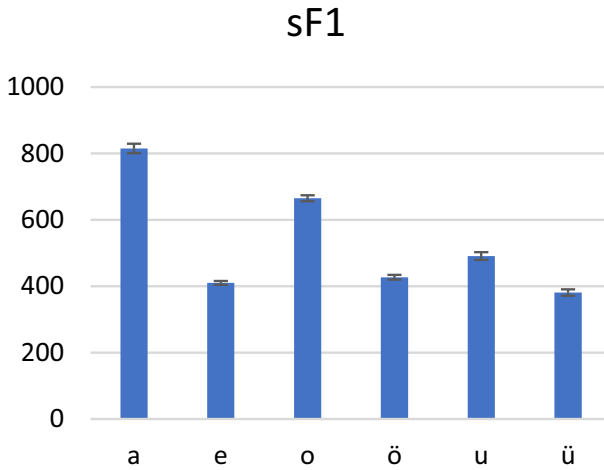


Figure 14 sF1 for Khalkha Mongolian vowels

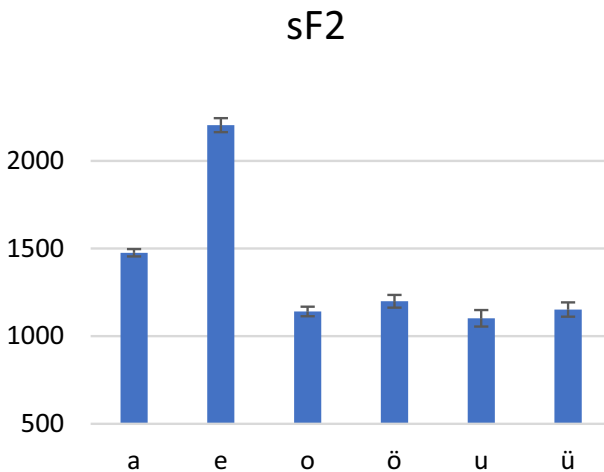


Figure 15 sF2 for Khalkha Mongolian vowels

(laryngoscopic) evidence from the northern Togolese (Gur) language Kabiye. Kabiye has a harmony system that divides its five basic vowels into two sets comparable to many other West African languages such as Akan. Esling finds that the primary gesture distinguishing the Kabiye [-ATR] vowels is an aryepiglottal-epiglottal stricture with concomitant reduction in the size of the pharyngeal cavity. This active articulation is accompanied by tongue retraction, larynx raising, and possible narrowing of the pharyngeal walls; but the primary factor is claimed to be the aryepiglottal-epiglottal stricture. The relevance of this finding for Mongolian harmony is that such a stricture would provide a plausible articulatory basis for the turbulence we have observed in the Khalkh RTR vowels. This, in turn, would make

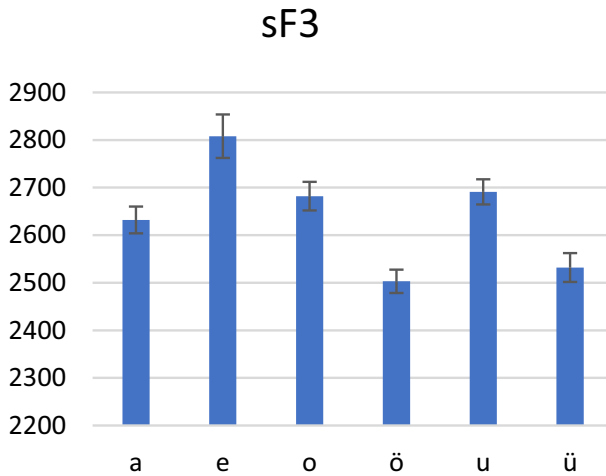


Figure 16 sF3 for Khalkha Mongolian vowels

Table 5 Rpart analysis of reduced data set

Model	Predictors	Tuned mmce	Cross-validated mmce
Full model	H1H2c,H1A1c,H1A2c, HNR05,HNR15,HNR25, sF0,sF1,sF2	0.082	0.117
Voice-quality	H1H2c,H1A1c,H1A2c, HNR05,HNR15,HNR25	0.203	0.211
Formants-only	sF1, sF2, sF0	0.085	0.099

the RTR set the marked or active factor phonologically. As noted by Ko (2019), Clements & Rialland (2008: 53) observe that in many African harmony systems the low vowels are the site of neutralization for ATR contrasts with [ATR] /a/ replaced by /a/ while in the East Asian languages like Mongolian it is the high vowels that are the site of neutral vowels with [RTR] /i/ replaced by /i/ (reversed polarity). We recall that /i/ lacks an RTR counterpart in the phonemic inventory of (1) and is transparent to the spread of tongue-root harmony in Mongolian. The generalization across these two language types would be the tendency for loss of contrast to be found in the regions of the vowel space that are least compatible with the ATR (open pharynx) West African versus RTR (constricted pharynx) East Asian articulations.

Esling et al.'s PPM (2019: 166) depicts the positive and negative synergies at play among the various articulations linking vowel quality, phonatory quality, and tonal quality. Phonological systems may override the positive connections for purposes of contrast; but a natural default state is implicit in the model. Of particular relevance for our Khalkha Mongolian results are the following synergistic (positive)

connections posited in the PPM: lingual (tongue-body and tongue-root) retraction and lowering (leading to increased F1 and decreased F2), epilaryngeal constriction (leading to noise), larynx raising, vocal fold adduction (for voicing), and increased vocal fold tension leading to higher f_0 . The acoustic correlates for the Khalkh /a,o,u/ vowels that we have found are consistent with the first four of these five pharyngeal-laryngeal articulations. The exception is f_0 where we found the /a,o,u/ set to be associated with lower pitch. The PPM also treats tongue-body raising as an anti-synergy for epilaryngeal constriction. This premise aligns with our data for spectral drop-off where H1H2 has the profile /a/ < /e/ and /o/ < /ö/, but /u/ > /ü/ and H1A1 has the profile /a/ > /e/ and /o/ > /ö/, but /u/ \approx /ü/. In both cases it is the high vowel pair that is the odd man out. And as noted above, the high front region of the vowel space lacks an RTR element in Mongolian.

As a final remark, one cannot help but wonder if the throat singing for which Mongolia is justly famous is not recruiting a basic feature of the language's phonological structure for artistic purposes comparable to how Japanese haiku and tanka verse are based on the fundamental role of the mora in shaping that language's prosody.

6 Conclusions

In this article we presented the results of a study of the phonetic correlates of the phonological contrasts of vowel length, vowel quality and voice quality in Khalkha Mongolian. The study is based on a recording of 75 monosyllabic noun stems in four case forms collected from five native speakers. The principal results are as follows. First, we confirmed and replicated Svantesson's (1985) important finding that the presumed front round vowels of Classical Mongolian have retracted to the back vowel region while original [u] and [o] have lowered in the vowel space. Svantesson interpreted this rotation of the round vowels as evidence that the phonological contrast between /ü/ and /ö/ versus /u/ and /o/ changed from an opposition in tongue-body backness to an opposition based on pharyngeal width. Second, we presented the results of an attempt to find voice-quality evidence for the postulated ATR versus RTR contrast. The most reliable correlate found was the Harmonics-to-Noise ratio where the RTR /u/ and /o/ vowels presented consistently lower values compared to /ü/ and /ö/ at most of the spectral levels analyzed. We speculated that this finding supported Esling's (2005) Laryngeal Articulator Model according to which the Khalkh RTR /a,o,u/ set may reflect aryepiglottal-epiglottal stricture. We also examined the positioning of the ablative suffixal vowels in F1*F2 space and found preliminary evidence that the [\pm back] contrast between /a:/ and /æ:/ is transmitted to the suffixal vowel along with assimilation in pharyngeal opening and lip rounding in the two vowel harmony processes for which the Mongolian language is well known. Finally, we found that long stem vowels were approximately two and a half times the duration of short vowels. This ratio was maintained in the face of changes in absolute duration as a function of suffixation and changes in syllable structure. Future research should try to pinpoint the articulatory basis of the tongue-root/pharyngeal harmony contrasts with

glottography and other tools. Also, tests modifying the voice-quality correlates of spectral drop-off and turbulence could be devised to see if these variables cue the ATR versus RTR distinction perceptually.

Acknowledgments Thanks to two anonymous JEAL reviewers for their helpful comments on the form and substance of the paper and to our five Mongolian speakers for sharing their language.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funding 'Open Access funding provided by the MIT Libraries'.

Appendix 1

Statistical models

Multivariate mixed-effects linear-regression models;

Model 1.1: (stem vowel) Log-normalized duration (ms): $\text{logdur} \sim \text{case} + (1|\text{word}) + (1|\text{speaker}) + (1+\text{caselword})$; Number of obs: 1135, groups: word, 235; speaker, 5

	Est	SE	Df	t-value	Pr(> t)	sig
(Intercept)	2.19	0.04	7.27	44.28	4.03e-10	***
Case:com	-0.10	0.03	158.87	-2.86	0.004	**
Case:nom	0.15	0.03	196.70	4.57	8.49e-06	***

Model 1.2 (long stem vowel): Log-normalized duration (ms): $\text{logdur} \sim \text{case} + (1|\text{word}) + (1|\text{speaker}) + (1+\text{caselword})$; Number of obs: 600, groups: word, 132; speaker, 5

	Est	SE	Df	t-value	Pr(> t)	sig
(Intercept)	2.37	0.04	4.45	49.73	2.71e-07	***
Case:com	-0.10	0.01	68.77	-7.13	7.70e-10	***
Case:nom	0.16	0.01	69.32	11.42	< 2e-16	***

Model 1.3 (short stem vowel): Log-normalized duration (ms): $\text{logdur} \sim \text{case} + (1|\text{word}) + (1|\text{speaker}) + (1+\text{case}|\text{word})$; Number of obs: 535, groups: word, 110; speaker, 5

	Est	SE	Df	t-value	Pr(> t)	sig
(Intercept)	1.94	0.04	5.97	44.64	8.96e-09	***
Case:com	-0.06	0.02	61.11	-2.27	0.02	*
Case:nom	0.19	0.02	65.75	7.33	4.12e-10	***

Model 2a for /i,e/: $F1_{\text{norm}} \sim \text{stemvowel} + (1|\text{speaker}) + (1|\text{word})$
 Number of obs: 345, groups: word, 43; speaker, 5

	Est	SE	t-value
(Intercept)	-0.71	0.06	-11.69
Stemvowel:i	-0.06	0.06	-0.98

Model 2b for /i,e/: $F2_{\text{norm}} \sim \text{stemvowel} + (1|\text{speaker}) + (1|\text{word})$
 Number of obs: 345, groups: word, 43; speaker, 5

	Est	SE	t-value
(Intercept)	1.007	0.03	27.26
Stemvowel:i	0.001	0.03	0.04

Model 3a: for ablative -a:s: $F1_{\text{norm}} \sim \text{stemvowel} + (1|\text{speaker}) + (1|\text{word})$
 Number of obs: 79, groups: word, 22; speaker, 5

	Est	SE	t-value
(Intercept)	1.91	0.12	15.85
Stemvowel:æ	-0.22	0.10	-2.04

Model 3b: for ablative -a:s: $F2_{\text{norm}} \sim \text{stemvowel} + (1|\text{speaker}) + (1|\text{word})$
 Number of obs: 79, groups: word, 22; speaker, 5

	Est	SE	t-value
(Intercept)	-0.41	0.04	-8.99
Stemvowel:æ	0.16	0.04	3.67

Model 4: $F3_{\text{norm}} \sim VQ + (1|\text{word}) + (1|\text{speaker}) + (1+VQ|\text{word})$
 Number of obs: 205, groups: word, 40; speaker, 5

	Est	SE	Df	t-value	Pr(> t)	sig
(Intercept)	-0.71	0.13	5.42	-5.16	0.002	**
VQ:RTR	0.71	0.11	44.60	6.24	1.41e-07	***

Model 5: for round vowels: sF2Z ~ VQ * length + gender + age +(1|word) + (1|speaker) + (1+VQ|word) + (1+VQ|speaker)

Number of obs: 202, groups: word, 41; speaker, 5

	slopes	VQ	Estimate	Std. error	Df	t-value	Pr('t)	sig
sF2Z	2	RTR	0.03	0.12	35.50	0.30	0.76	
		length:short	0.79	0.12	18.38	6.55	3.34e-06	***
		gender:male	0.22	0.10	158.39	2.12	0.03	*

Univariate binomial logistic regression models predicting VQ (baseline ATR)

Number of obs: 368, groups: word, 75; speaker, 5

glmer(VQ ~ predictor + (1|word) + (1|speaker), family = binomial)

Predictor	Estimate	Std. error	z-value	Pr(> z)
H1H2cZ	-0.72	0.73	-0.97	0.32
H1A1cZ	1.06	0.81	1.30	0.19
H1A2cZ	1.06	0.97	1.09	0.27
H1A3cZ	-0.04	0.79	-0.05	0.95
HNR05Z	0.63	0.902	0.70	0.48
HNR15	-0.03	0.08	-0.37	0.70
HNR25Z	-0.25	0.97	-0.26	0.79
HNR35Z	-0.19	0.95	-0.20	0.83
sF0Z	-1.26	0.94	-1.33	0.18
sF1Z	30.60	9.09	3.36	0.0007
sF2Z	-0.49	0.68	-0.71	0.47
sF3Z	0.44	0.67	0.66	0.50

Best fit multivariate binomial logistic regression model

glmer (VQ ~ sF1Z + HNR05Z + HNR15Z + (1|word), family = binomial)

AIC	BIC	logLik	deviance	df.resid
59.7	79.2	-24.8	49.7	363

Number of obs: 368, groups: word, 75

	Est	Std. error	z-value	Pr(> z)	Sig
(intercept)	-11.86	6.28	-1.88	0.05	.

sF1Z	20.52	8.26	2.48	0.01	*
HNR05Z	30.80	13.68	2.25	0.02	*
HNR15Z	-41.12	19.86	-2.07	0.03	*

Appendix 2

Simple coding

Correlate	Slopes	VQ	estimate	Std. error	Df	t-value	Pr(lt)	sig
H1H2cZ	0	RTR	-0.76	0.11	70.31	-6.61	6.16e-09	***
H1A1cZ	0	RTR	0.76	0.09	86.21	7.73	1.77e-11	***
H1A2cZ	1	RTR	0.85	0.14	49.91	5.99	2.27e-07	***
H1A3cZ	1	RTR	0.18	0.14	91.33	1.26	0.20	
HNR05Z	1	RTR	0.37	0.17	61.74.12	2.13	0.03	*
HNR15Z	1	RTR	-0.91	0.16	66.92	-5.70	2.87e-07	***
HNR25Z	1	RTR	-0.71	0.17	66.21	-4.16	9.24e-05	***
HNR35Z	0	RTR	-0.38	0.14	208.06	-2.73	0.006	**
sF0Z	1	RTR	-0.93	0.15	60.50	-6.20	5.48e-08	***
sF1Z	1	RTR	1.32	0.09	212.27	14.26	<2e-16	***
		male	-0.13	0.05	293.88	-2.37	0.01	*
sF2Z	2	RTR	-0.51	0.13	93.99	-3.75	0.0002	***
sF3Z	2	RTR	-0.12	0.25	8.54	-0.47	0.64	
sF3Z,rand	1	RTR	0.70	0.11	48.89	6.27	9.02e-08	***
		male	0.76	0.13	164.75	5.56	1.03e-07	***

Appendix 3

List of recorded words (transliterated)

	Gloss	Nom	Acc	Abl	Com
1	shin	ʃiir	ʃiiriig	ʃiirees	ʃiirtai
2	tail	süülʒ	süülʒiig	süülʒnees	süülʒtai
3	hall	tsaalʒ	tsaalʒiig	tsaalʒnaas	tsaalʒtai
4	soup	ʃölʒ	ʃölʒiig	ʃölʒnööös	ʃölʒtai
5	castle, fort	ts ^h aits	ts ^h aitsiig	ts ^h aitsnaas	ts ^h aitstai
6	blood	ts ^h us	ts ^h usiig	ts ^h usaas	ts ^h ustai
7	lock	ts ^h ootʃ	ts ^h ootʃiig	ts ^h ootʃnoos	ts ^h ootʃtai

Appendix

Gloss	Nom	Acc	Abl	Com
8 color	tsüs	tsüsiig	tsüsees	tsüstai
9 foam	xöös	xöösiig	xöösnöös	xööstai
10 flame	tölz	tölziig	tölznöös	tölztai
11 coin	tsoos	tsoosiig	tsoosnoos	tsoostai
12 tent	ger	geriig	gerees	gertai
13 stove	tsuux	tsuuxiig	tsuuxnaas	tsuuxtai
14 team	pag	pagiig	pagaas	pagtai
15 rope	tees	teesiig	teesnees	teestai
16 river	golz	golziig	golzooos	golztai
17 glass	ʃilz	ʃilziig	ʃilznees	ʃilztai
18 ribbon, braid	tuuts	tuutsiig	tuutsnaas	tuutstai
19 love	xair	xairiig	xairnaas	xairtai
20 ink	pex	pexiig	pexees	pextai
21 fire	galz	galziig	galzaas	galztai
22 fat	ööx	ööxiig	ööxnöös	ööxtai
23 spear	tʃat	tʃatiig	tʃatnaas	tʃattai
24 force	xütʃ ^h	xütʃ ^h iig	xütʃ ^h naas	xütʃ ^h tai
25 dirt	xir	xiriig	xirnees	xirtai
26 brush	ʃüür	ʃüüriig	ʃüürees	ʃüürtai
27 demon	pug	pugiig	pugaas	pugtai
28 paint brush	piir	piiriig	piirnees	piirtai
29 corpse	xüür	xüüriig	xüürnees	xüürtai
30 star	ot	otiig	otnoos	ottai
31 back	ar	ariig	araas	artai
32 food	xoolz	xoolziig	xoolzooos	xoolztai
33 bribe	xeelz	xeelziig	xeelznees	xeelztai
34 oil, fat	tos	tosiig	tosooos	tostai
35 scissors	xaitʃ ^h	xaitʃ ^h iig	xaitʃ ^h naas	xaitʃ ^h tai
36 border	tsaag	tsaagiig	tsaagnaas	tsaagtai
37 brown color	por	poriig	porooos	portai
38 shrub	söög	söögiig	söögnöös	söögtai
39 flute	ts ^h uur	ts ^h uuriig	ts ^h urnaas	ts ^h urtai
40 direction	tiiʃ	tiiʃiig	tiiʃees	tiiʃtai
41 shame	xög	xögiig	xögnöös	xögtai
42 ticket	tiits	tiitsiig	tiittees	tiitstai
43 dice	ʃoo	ʃoog	ʃooos	ʃootai
44 daughter-in-law	per	periig	perees	pertai
45 roof-tile	waar	waariig	waarnaas	waartai
46 gown	teelz	teelziig	teelznees	teelztai
47 year	tʃilz	tʃilziig	tʃilzees	tʃilztai
48 hand	gar	gariig	garaas	gartai
49 joy	xöör	xööriig	xöörnöös	xöörtai
50 monkey	pitʃ ^h	pitʃ ^h iig	pitʃ ^h ees	pitʃ ^h tai

Appendix

Gloss	Nom	Acc	Abl	Com
51 help	tus	tusiig	tusaas	tustai
52 anvil	töʃ	töʃiig	töʃnööᠰ	töʃtai
53 river	golʒ	golʒiig	golʒoos	golʒtai
54 steppe	xeer	xeeriig	xeerees	xeertai
55 dinner	tsoog	tsoogiig	tsoognoos	tsoogtai
56 fiddle	xuur	xuuriig	xuuraas	xuurtai
57 leg	xölʒ	xölʒiig	xölʒnööᠰ	xölʒtai
58 beard	xuuts	xuutsiig	xuutsnaas	xuutstai
59 antelope	tseer	tseeriig	tseernees	tseertai
60 court	ʃüüx	ʃüüxiig	ʃüüxeeᠰ	ʃüüxtai
61 pillow	ter	teriig	ternees	tertai
62 bull	pux	puxiig	puxnaas	puxtai
63 tooth	ʃüt	ʃütiig	ʃütnees	ʃüttai
64 violin	xiilʒ	xiilʒiig	xiilʒees	xiilʒtai
65 border	xilʒ	xilʒiig	xilʒnees	xilʒtai
66 bead	ʃür	ʃüriig	ʃürees	ʃürtai
67 paper	ts ^h aas	ts ^h aasiig	ts ^h aasnaas	ts ^h aastai
68 minister	sait	saitiig	saitaas	saittai
69 dream	tsüüt	tsüütiig	tsüütnees	tsüüttai
70 nose-rope	tör	töriig	törnööᠰ	törtai
71 devil	tüt	tütiig	tütnees	tüttai
72 law	ts ^h aats	ts ^h aatsiig	ts ^h aatsaas	ts ^h aastai
73 apartment	pair	pairiig	pairnaas	pairtai
74 kidney	pöör	pööriig	pöörnööᠰ	pöörtai
75 pattern	teg	tegiig	tegnees	tegtai
76 singer	tuutʃ ^h	tuutʃ ^h iig	tuutʃ ^h naas	tuutʃ ^h tai
77 flag	tug	tugiig	tugnaas	tugtai

References

- Binnick, R. 1969. Non-high neutral vowels in modern Mongolian vowel harmony. In *Papers from the fifth regional meeting of the Chicago Linguistic Society*, ed. R. Binnick et al., 295–301.
- Binnick, R. 1980. The underlying representation of harmonizing vowels: evidence from modern Mongolian. In *Issues in vowel harmony*, ed. R. Vago, 113–134. Benjamins.
- Boersma, P., and D. Weenick. 2019. Praat: doing phonetics by computer. Praat.org
- Chinggeltei, Q. 1979. Zhongguo mengguyang fangyan de huafen. The dialectal division of China's Mongolian. *Minzu yuwen* 1 (13-20): 108–112.
- Clements, G. N., and A. Rialland. 2008. Africa as a phonological area. In *A Linguistic geography of Africa*, ed. B. Heine & D. Nurse, 36–85. Cambridge University Press.
- Esling, J. 2005. There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics* 50: 13–44.

- Esling, J., S. Moisić, A. Benner, and L. Crevier-Buchman. 2019. *Voice quality*. Cambridge University Press.
- Fulop, S., E. Kari, and P. Ladefoged. 1998. An acoustic study of the tongue root contrast in Degema vowels. *Phonetica* 55: 80–98.
- Guion, S., M. Post, and D. Payne. 2004. Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics* 32: 517–542.
- Hamp, E. 1980. Mongolian vocalic features and the problem of harmony. In *Issues in vowel harmony*, ed. R. Vago, 101–111. Benjamins.
- Hess, S. 1992. Assimilatory effects in a vowel harmony system: An acoustic analysis of advanced tongue root in Akan. *Journal of Phonetics* 20: 475–492.
- Ito, J. 1990. Prosodic minimality in Japanese, CLS 26-II: *Papers from the Parasession on the Syllable in Phonetics and Phonology*, 213–239. University of Chicago.
- Kawahara, H., I. Masuda-Katsuse, and A. de Cheveigne. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Communication* 27: 187–207.
- Ko, S. 2011. Vowel contrast and vowel harmony shift in the Mongolic languages. *Language Research* 47: 23–43.
- Ko, S., A. Joseph, and J. Whitman. 2014. Comparative consequences of the tongue root harmony analysis for proto-Tungusic, proto-Mongolic, and proto-Korean. In *Paradigm change: In the Transeurasian languages and beyond*, ed. M. Robbeets and W. Bisang, 141–176. Benjamins.
- Ko, S. 2019. *Tongue root harmony and vowel contrast in northeast Asian languages*. Harrassowitz Verlag.
- Korkmaz, Y., and B. Aytug. 2018. Classification of Turkish vowels based on formant frequencies. Research Gate. <https://www.researchgate.net/publication/331277195>.
- Ladefoged, P. 1964. *A phonetic study of West African languages: An auditory-instrumental survey*. Cambridge University Press.
- Law, F. 2008. Acoustical analysis of Canadian and Parisian French word-final vowel productions in varying phonetic contexts. *The Journal of the Acoustical Society of America* 123 (5): 3735.
- Lehiste, I. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lightner, T. 1965. On the description of vowel and consonant harmony. *Word* 21: 244–250.
- Lindau, M. 1979. The feature expanded. *Journal of Phonetics* 7: 163–176.
- Lindau, M. 1975. Features for vowels. *UCLA Working Papers in Phonetics* 30: 1–28.
- Maddieson, I. 1997. Phonetic universals. In *The handbook of phonetic sciences*, ed. J. Laver and W. Hardcastle, 619–639. Blackwell.
- Odden, D. 1980. The irrelevancy of the relevancy condition: Evidence for the feature specification constraint. *Linguistic Analysis* 6: 261–304.
- Piroth, H., P. Skupinski, and B. Pompino-Marschall. 2015. Production of vowel contrasts in Northern Standard German and Austrian Standard German. In *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by The Scottish Consortium for ICPHS 2015, Glasgow, UK: the University of Glasgow.
- Przedziecki, M. 2005. Vowel harmony and coarticulation in three dialects of Yoruba: Phonetics determining phonology. Ph.D. dissertation, Cornell University.
- Rialland, A., and R. Djamouri. 1984. Harmonie vocalique, consonantique et structures de dépendance dans le mot en mongol Khalkha. *Bulletin de la société linguistique de Paris*, 333–383.
- Shue, Y.-L., P. Keating, C. Vicenik, and K. Yu. 2009. VOICESAUCE: A program for voice analysis. In *Proceedings of the ICPHS XVII*, 1846–1849. <http://www.ee.ucla.edu/~spapl/voicesauce>.
- Sjölander, K. (2004). The snack sound toolkit. <https://www.speech.kth.se/snack/>.
- Starwalt, C. 2008. The acoustic correlates of ATR harmony in seven- and nine-vowel African languages: A phonetic inquiry into phonological structure. Ph.D. dissertation, University of Texas at Arlington.
- Steriade, D. 1979. Vowel harmony in Khalkha Mongolian. *MIT Working Papers in Linguistics* 1: 25–50.
- Stewart, J. 1967. Tongue root position in Akan vowel harmony. *Phonetica* 16: 185–204.
- Strange, W., A. Weber, E. Levy, V. Shafiro, M. Hisagi, and K. Nishi. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America* 122 (2): 1112–1119.
- Svantesson, J.-O. 1985. Vowel harmony shift in Mongolian. *Lingua* 67: 283–327.
- Svantesson, J.-O., A. Tsendina, A. M. Karlsson, and V. Franzén. 2005. *The phonology of Mongolian*. Oxford University Press.

- Vaux, B. 2009. “[atr] and [back] harmony in the Altaic languages.” In *Investigations into Formal Altaic Linguistics: Proceedings of the 3rd Workshop on Altaic Formal Linguistics (WAFLL 3)*, ed. S. Tatevosov, 50–67. Moscow: MAKSS Press.
- Wood, S. 1979. A radiographic analysis of constriction locations for vowels. *Journal of Phonetics* 7: 25–43.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.