DURATION AND PITCH IN THE VOWEL QUANTITY DISTINCTION OF YAKUT: EVIDENCE FROM SPONTANEOUS SPEECH

Lena Vasilyeva, Anja Arnhold, Juhani Järvikivi

University of Alberta
lvasilye@ualberta.ca, arnhold@ualberta.ca, jarvikiv@ualberta.ca

ABSTRACT

This paper investigates instrumentally for the first time the binary vowel quantity opposition (short vs. long) in Yakut (or Sakha) on the basis of spontaneous production data from nine speakers. Acoustic measurements of vowels in disyllabic words showed a significantly shorter duration of short vowels than their long counterparts. Furthermore, f0 maxima and f0 slope showed effects of both quantity and syllable number. The results suggest that pitch is an additional phonetic correlate of vowel quantity in Yakut, alongside with the robust durational difference between short and long vowels.

Keywords: Vowel quantity, Pitch, f0, Duration, Yakut, Spontaneous speech.

1. INTRODUCTION

Yakut is a Turkic language spoken in the Republic of Sakha (Yakutia) of the Russian Federation. Yakut is a quantity language where all vowels, and many consonants, have phonemically short and long variants (e.g. see minimal pairs like /ɑt/ 'horse' vs. /ɑːt/ 'name' or /kɯrɯs/ 'topsoil' vs. /kɯrɯːs/ 'curse'). Yakut also has regular word-final stress [1]. Krueger [7] claims that the difference between short and long vowels is based on length only, with no change in articulation and long vowels being two to three times longer than short vowels. However, recent perception and production studies have shown that there are additional correlates of quantity in several languages. In particular, these studies highlight the important role of pitch cues in the vowel quantity distinction, even in languages that do not have lexical tone.

Studies of the three-way quantity distinction in Estonian have determined that Estonian listeners rely mainly on durational syllable ratios in identifying quantity 1 (short), but use pitch cues to distinguish between quantity 2 (long) and quantity 3 (overlong) [9]. Speakers of Finnish, which like Yakut has a binary quantity distinction and fixed stress (initial for Finnish, final for Yakut), use pitch for the production and perception of quantity distinctions. In disyllabic words, as recent studies have shown, first syllables frequently carried an f0 fall when they contained a long vowel, whereas f0 was more flat (high) for first syllables with short vowels [12]. In perception, listeners primarily used durational cues, however, when temporal information was ambiguous, listeners resorted to pitch cues to distinguish vowel quantity [5]. This phenomenon is also observed in Japanese, where listeners used pitch cues besides durational information when the latter was not evident enough [6].

These and other studies show that pitch and duration can both correlate with quantity (for an overview and an experiment confirming a general correlation in perception, see [13]). Therefore, we hypothesize that pitch may serve as an additional phonetic correlate of quantity also in Yakut, in contrast to what available accounts assume.

2. MATERIALS

We present an acoustic analysis of long and short vowels in unscripted speech by multiple Yakut speakers. We extracted disyllabic nouns and verbs from mostly monologue-based spontaneous conversations.

2.1. Participants

Nine female native speakers, aged 19 to 77 years, participated in the study. All of the speakers grew up in a predominantly monolingual Yakut-speaking environment. The participants are bilingual in Russian, but Yakut is their strongest language. None of the speakers had any speech or hearing impairment.

2.2. Recordings

The recordings were conducted in fieldwork using an MP3 Dictaphone. The principal investigator asked the speakers to talk for about 10-15 minutes about any topic they liked. The participants were encouraged to have a monologue for as long as they would like. However, in case of a pause during the recording session, the investigator led a dialogue-based conversation. Most importantly, the speakers were asked to speak as naturally as possible.

2.3. Items and acoustic measurements

For eight speakers, 25 disyllabic verbs and nouns with short vowels, and 25 disyllabic verbs and nouns with long vowels in either one or both syllables were selected and segmented. Thus, a total of 50 words were segmented from each speaker. There were not enough words with long vowels in the data from two of the speakers, therefore eighteen additional words were taken from a ninth female speaker in order to reach the target number of 400 words. Syllables and vowels were segmented for each word token using Praat software [4]. As a result, 200 disyllabic words with short vowels in both syllables and 200 disyllabic words with long vowels in either one or both syllables were segmented.

The acoustic analysis involved the following measurements: duration, maximum f0 value and f0 slope for each vowel.

3. RESULTS

We analysed each response variable with linear mixed-effect models implemented in R [2, 8], using the function *anova* for comparisons between models with different variables to obtain the model with the best fit to the data. Tested predictors were vowel quantity and syllable number. Tested random effects were speaker, word and vowel. The final model only retained variables that significantly improved the model fit.

3.1. Duration

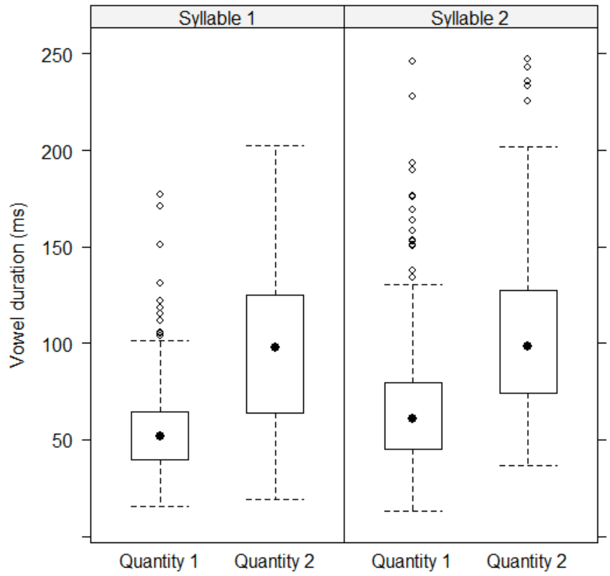
The model of duration with the best fit to the data included quantity and syllable number as significant predictors, as well as speaker, word, and vowel as random factors (see Table 1).

Table 1: Fixed effects of the best statistical model of vowel duration.

	Estimate	Std. Error	t-value
(Intercept)	55.431	5.873	9.438
Quantity 2	43.323	8.032	5.394
Syllable 2	15.144	3.523	4.299

Table 1 shows significantly longer vowel durations in the long quantity (quantity 2) than the short quantity (quantity 1). In addition, vowel duration was significantly longer in the second syllable (syllable 2) than the first syllable (syllable 1). There was no interaction between quantity and syllable number, suggesting that these were independent effects. The boxplot in Figure 1 illustrates the vowel duration of short and long quantity vowels by syllable position.

Figure 1: Vowel duration of short and long quantities by syllable position. Extreme outliers above 250ms are removed for clarity.



As Figure 1 shows, quantity 1 vowels tended to be clearly shorter in duration than the quantity 2 vowels, although there was some overlap. Figure 1 further displays an overall longer duration for both quantities in the second syllable position. However, the effect of position was clearly smaller than that of quantity, as also indicated by the estimates in Table 1.

3.2. Fundamental frequency: f0 maximum

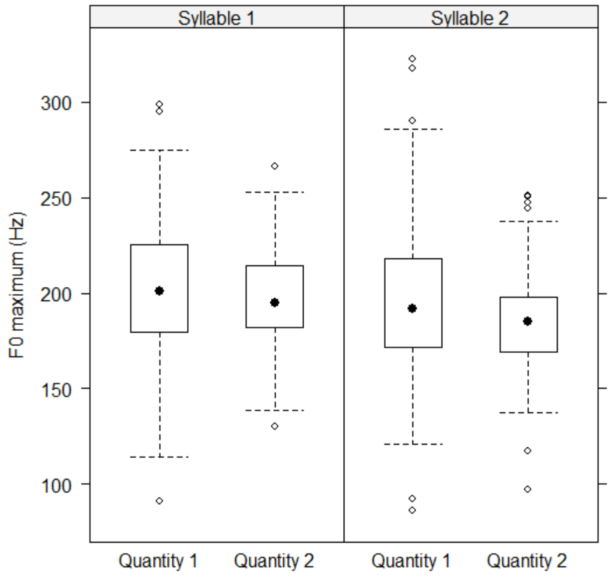
The best statistical model of the f0 maximum appears in Table 2. It included both quantity and syllable number as predictors and speaker and word as random factors.

Table 2: Fixed effects of the best statistical model of maximum f0 values.

	Estimate	Std. Error	t-value
(Intercept)	201.796	5.312	37.99
Quantity 2	-5.639	2.756	-2.05
Syllable 2	-8.768	2.191	-4.00

As illustrated in Figure 2, maximum f0 values were significantly lower in the second syllable than for first syllable vowels. Furthermore, values in quantity 2 were lower than in quantity 1, with this effect just reaching significance. No interaction between quantity and syllable number was found.

Figure 2: Maximum f0 values of short and long quantities by syllable position.



However, the question remains whether quantity impacted all of the Yakut speakers in the same way or whether there was variance from one speaker to another in how they produced long and short quantities. Hence, a model that adjusted the effect of quantity for each speaker as a by-speaker random factor was computed. The comparison of this model with the best model shown in Table 2 showed no evidence that the more complex model would provide a better fit (p = 0.95, χ^2 = 0.09). Thus, there was no indication that quantity affected the speakers differently.

3.2. Fundamental frequency: f0 slope

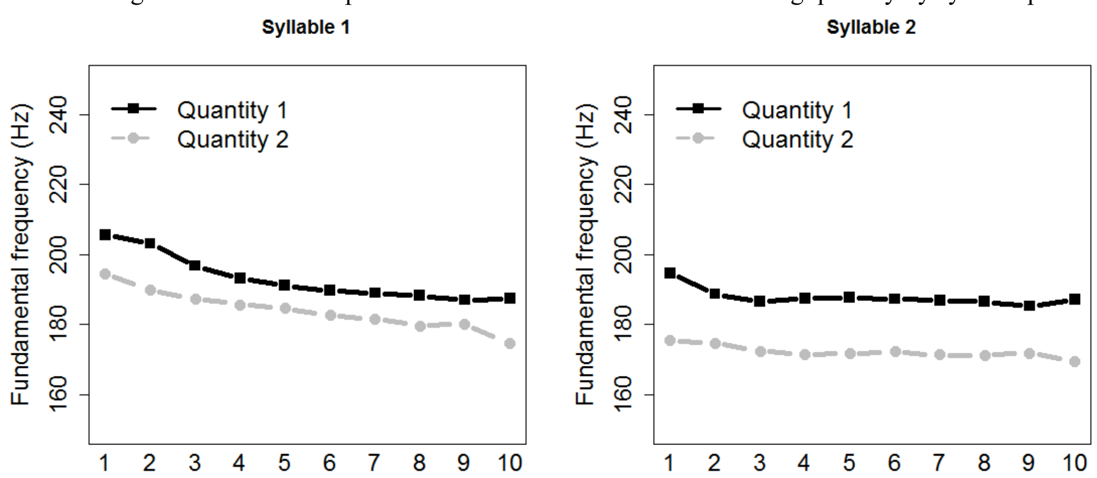
Figure 3 shows average time-normalized pitch contours based on f0 measurements at ten equidistant time-points within each vowel. What is evident is that average f0 was consistently lower for quantity 2 than for quantity 1 in both syllable positions, in line with the significant effect of quantity in Table 2. This seems to be driven mostly by the second syllable (Figure 3), and could be simply due to the pitch fall continuing for longer time in quantity 2. Overall, pitch contours in both syllable positions were relatively flat. However, f0 was falling for both quantities in the first syllable.

To assess possible differences in pitch movement across the vowel, we calculated f0 slope by subtracting minimum f0 from maximum f0 and dividing the result by the result of subtracting the time of the f0 minimum from the time of the f0 maximum. Thus, positive slope values indicate an f0 rise, and negative ones represent a fall, respectively; the more negative or positive the values are, the steeper the f0 movement. We discarded outlier values above 4000Hz/s and below -4000Hz/s (0.5% of the data).

Figure 4 shows the f0 slope for both quantities by syllable position. It shows that most slope values were fairly close to 0, which corresponds to flat f0 contours or no movement. The vowels in the first syllable position carried falling pitch on average (Figure 3), thus their slope values mostly remained in the negative range. By contrast, f0 of second syllable vowels was more flat, yielding slopes closer to 0. Also, the slope of vowels in quantity 2 was higher than that of vowels in quantity 1, meaning that there was a less steep fall of the f0 contours for quantity 2. Interestingly, the difference between the quantities was larger in first syllables than in second syllables, as Figure 4 illustrates.

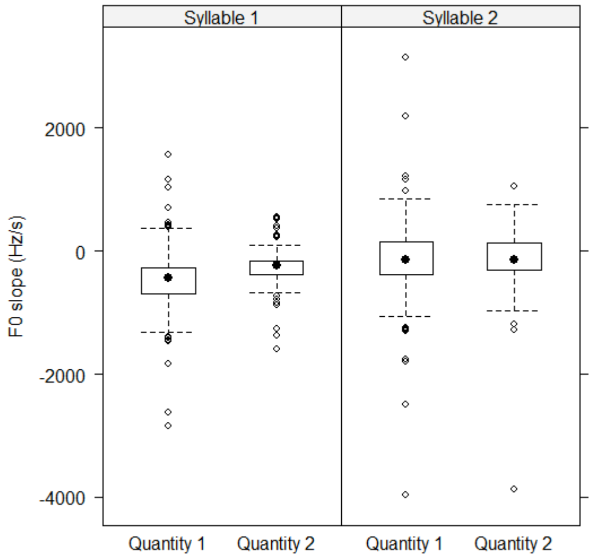
In line with this, the best model of f0 slope included an interaction of quantity and syllable number (Table 3). It also specified speaker as a random factor. The analysis revealed that this model

Figure 3: Average time-normalized pitch contours for vowels in short and long quantity by syllable position.



was significantly better than the model with no interaction between the predictors (p = 0.006**, χ^2 = 7.54).

Figure 4: F0 slope of short and long quantities by syllable position.



The main effects of the model confirmed that the effects of quantity and syllable position were significant. Thus, while f0 movements tended to be falling overall, the fall was less steep for quantity 2 vowels and vowels in second syllables. However, the interaction counteracted these effects, suiting the interpretation that the quantity difference was less pronounced in syllable 2.

Table 3: Fixed effects of the best statistical model of the f0 slope.

	Estimate	Std. Error	t-value
(Intercept)	-503.95	63.98	-7.877
Quantity 2	237.87	66.45	3.580
Syllable 2	368.71	52.22	7.061
Quantity 2 : Syllable 2	-268.04	97.17	-2.758

4. DISCUSSION AND CONCLUSION

The results indicate that it is the robust durational difference between short and long vowels that drives phonological quantity in Yakut. As has been observed for other quantity languages, also Yakut speakers lengthen long vowels compared to the short ones. Most previous studies have looked at laboratory speech and words produced either separately or in a carrier sentence. Our result shows that the durational differences are robust also in spontaneous speech.

Further, we also found that syllable number was a significant predictor of duration. Vowels in both quantities appeared to be a bit longer in the second syllable compared to the first syllable. This result is in line with Krueger's note that when the same vowel appears in two successive syllables, as in *kykyr* 'large', the second vowel, which is accented, is a bit longer [7]. Although the present experiment did not control for the individual vowels, the results show consistent vowel lengthening in the second syllable. Thus, Krueger's finding seems to extend beyond disyllabic words with two identical vowels. The observed tendency for vowels in both quantities to slightly lengthen in the second syllable position in the target disyllabic words is conceivably connected with the regular word-final stress in Yakut mentioned by Anderson [1]. Further studies are needed in order to clarify the relationship between stress, accent and duration in Yakut.

Furthermore, we found that the f0 slope was overall less steeply falling in the second syllable position and for the long quantity. However, there was an additional interaction showing that f0 slope differed between the two quantities only in the first syllable. Therefore, there was no quantity distinction in f0 movement in the stressed second syllable. It stands to reason that quantity signalling can and will co-opt f0 only on syllables where tonal targets are introduced, whether by lexical distinctions (as in Korean [8]), pitch accents (as in Estonian [9]), or phrasal tones (as in Finnish [2]). In this context it is relevant that f0 cues to quantity in Finnish have only ever been investigated in the first (stressed) syllable and not for the second one, which invariably carries low flat f0 [12]. In this sense, Yakut behaves like Finnish, with f0 distinctions between quantities appearing only on syllables that have tonal targets. Further, this also implies that in Finnish f0 is not itself a cue to quantity, despite the fact that people can use f0 to aid word recognition [5], as further attested by the fact that Finnish speakers cannot use f0 to distinguish Estonian quantities 2 and 3 [10].

In conclusion, we have shown that although the Yakut vowel quantity distinction is primarily marked through duration, pitch contours serve as an additional correlate of the binary opposition, as in several other languages.

5. REFERENCES

[1] Anderson, G. D. S. 1995. Historical aspects of Yakut (Saxa) phonology. *Turkic Languages* 2, 1–32.

[2] Arnhold, A. 2014. Finnish prosody: studies in intonation and phrasing. Dissertation, Goethe-University, Frankfurt.

[3] Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

[4] Boersma, P., Weenink, D. 2014. Praat: Doing phonetics by computer. http://www.praat.org

[5] Järvikivi, J., Vainio, M., Aalto, D. 2010. Real-time correlates of phonological quantity reveal unity of tonal and non-tonal languages. *PloS ONE* 5, 1–10.

[6] Kinoshita, K., Behne, D. M., Arai, T. 2002. Duration and F0 as perceptual cues to Japanese vowel quantity. *Proc. of the seventh international conference on spoken language processing*, 757–760.

[7] Krueger, J. R. 2012 [1962]. *Yakut manual, volume 21 of Uralic and Altaic Series.* Reprinted by Routledge: London, New York.

[8] Kwon, K-K. 2003. Prosodic change from tone to vowel length in Korean. In: Fikkert, P., & H. Jacobs (Eds.), *Development in Prosodic Systems*. Berlin: de Gruyter, 67–78.

[9] Lippus, P., Pajusalu, K., Allik, J. 2007. The tonal component in perception of the Estonian quantity. *Proc. 16th ICPhS* Saarbrücken, 1049–1052.

[10] Lippus, P., Pajusalu, K., Allik, J. 2009. The tonal component on Estonian quantity in native and non-native perception. *Journal of Phonetics* 37, 388–396.

[11] R Development Core Team. 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.r-project.org

[12] Vainio, M., Järvikivi, J., Aalto, D., Suni, A. 2010. Phonetic tone signals phonological quantity and word structure. *J. Acoust. Soc. Am*. 128, 1313–1321.

[13] Yu, A. L. 2010. Tonal effects on perceived vowel duration. In: Fougeron, C., Kühnert, B., D'Imperio, M., Vallée, N. (eds), *Laboratory Phonology* 10. Berlin: de Gruyter, 151–168.