Phonetic correlates of phonological vowel quantity in Yakut read and spontaneous speech

Lena Vasilyeva,[a] Anja Arnhold, and Juhani Järvikivi
Department of Linguistics, University of Alberta, 2-40 Assiniboia Hall, Edmonton, Alberta T6G 2E7, Canada

(Received 1 October 2015; revised 7 April 2016; accepted 13 April 2016; published online 10 May 2016)

The quantity language Yakut (Sakha) has a binary distinction between short and long vowels. Disyllabic words with short and long vowels in one or both syllables were extracted from spontaneous speech of native Yakut speakers. In addition, a controlled production by a native speaker of disyllabic words with different short and long vowel combinations along with contrastive minimal pairs was recorded in a phonetics laboratory. Acoustic measurements of the vowels' fundamental frequency, duration, and intensity showed a significant consistent lengthening of phonologically long vowels compared to their short counterparts. However, in addition to evident durational differences between long and short quantities, fundamental frequency and intensity also showed effects of quantity. These results allow the interpretation that similarly to other non-tonal quantity languages like Finnish or Estonian, the Yakut vowel quantity opposition is not based exclusively on durational differences. The data furthermore revealed differences in F0 contours between spontaneous and read speech, providing some first indications of utterance-level prosody in Yakut.
 [http://dx.doi.org/10.1121/1.4948448]

[JFL] Pages: 2541–2550

I. INTRODUCTION

This paper investigates the acoustic correlates of phonological quantity in Yakut. Yakut (or Sakha) is a Turkic language spoken by over 400 000 speakers in the Republic of Sakha (Yakutia) of the Russian Federation (Pakendorf, 2007). According to existing accounts, Yakut is a standard quantity language where all vowel phonemes, and many consonants, have short and long counterparts distinguishing lexical meaning (Anderson, 1998; Finch, 1985; Krueger, 2012). It is assumed that the distinction between short and long vowels in Yakut is based solely on a difference in duration (Krueger, 2012, p. 47), although there is currently no phonetic (instrumental) study of the acoustic properties of this distinction. The present study focuses on remedying this situation by inquiring into the binary quantity distinction in Yakut long and short vowels with the purpose of determining its acoustic-phonetic characteristics. To that end, the study uses disyllabic nouns and verbs with long and short vowels as well as monosyllabic and disyllabic minimal pairs extracted from both read and spontaneous speech.

Yakut has four front vowels /i, y, ɛ, œ/ and four back vowels /*ɯ*, u, ɑ, ɒ/ (Anderson, 1998; Krueger, 2012), thus each unrounded Yakut vowel phoneme has a symmetrical rounded counterpart (Sasa, 2009). All vowels appear in word initial, medial, and final positions (Krueger, 2012). Table I represents quantity minimal pairs for each of them. In addition, there are four diphthongs /iɛ, yœ, uɒ, *ɯ*ɑ/ (Finch, 1985; Kaun, 1995; Krueger, 2012). Moreover, Yakut has vowel harmony, specifically, backness and rounding harmony (Anderson, 1998; Finch, 1985; Kaun, 1995; Krueger, 2012; Sasa, 2009). Stress is regularly word final (Anderson, 1998) but is not very prominent and often hardly distinguishable acoustically (Samsonova, 1959). Krueger states that Yakut long vowels are two to three times the length of short vowels (Krueger, 2012, p. 48).

Previous studies on a number of quantity languages, such as Japanese (Kinoshita *et al.*, 2002), Finnish (Järvikivi *et al.*, 2010; Nakai *et al.*, 2009; Ylinen *et al.*, 2005), Estonian (Lehiste *et al.*, 2007; Lippus *et al.*, 2011), Washo (Yu, 2008), Thai (Mixdorff *et al.*, 2002), and Livonian (Lehiste *et al.*, 2007), have shown that in addition to the robust durational cues, also pitch cues may distinguish short and long quantity in these languages (see also, Fox and Lehiste, 1987; Yu, 2010). For example, perception experiments have suggested that speakers of non-tonal languages like Finnish and Japanese use pitch cues to distinguish between short and long phonemes. Järvikivi *et al.* (2007) and Järvikivi *et al.* (2010) showed that participants' decisions to categorize the initial vowel in a word as long or short, e.g., sika "pig" vs siika "whitefish," were significantly influenced by not only duration but also whether the vowel had a level or falling pitch. They further found that, controlling for durational cues, appropriate pitch of auditory word primes facilitated word recognition in cross-modal priming. Vainio *et al.* (2010) moreover demonstrated that Finnish long and short

TABLE I. Minimal pairs of short and long vowels.

[ɑ]/[ɑː]	bɑs "head"	bɑːs "wound"
[ɒ]/[ɒː]	χɒs "room"	χɒːs "empty"
[ɛ]/[ɛː]	ɛhɛ "bear"	ɛhɛː "grandfather"
[œ]/[œːː]	bœrœ "wolf"	bœrœː "wrap!, cloak!, cover up!"
[i]/[iː]	ir "melt"	iːr "go insane"
[y]/[yː]	kyly "ash (accusative)"	kyly: "laughter"
[*ɯ*]/[*ɯ*ː]	k*ɯ*l "horse mane"	k*ɯ*ːl "animal"
[u]/[uː]	kus "duck"	kuːs "hug"

[a]Electronic mail: lvasilye@ualberta.ca

vowels systematically co-varied with differences in pitch contours also in production. Although the primary role of duration in quantity perception is indisputable, listeners used pitch cues when durational information was ambiguous (see also Kinoshita *et al.*, 2002, for Japanese). Pitch cues also play an important role in the perception of quantity in Estonian, another (non-tonal) quantity language. In contrast to Finnish, Estonian has a tertiary quantity distinction. A series of perception studies among Estonian listeners conducted by Lippus *et al.* (2007, 2009, 2011) revealed a significant role of pitch cues in the distinction between quantity 2 (long) and quantity 3 (overlong). Unlike for Finnish, where pitch cues co-vary with duration, in Estonian pitch serves as an independent cue distinguishing between the long and overlong quantity alongside with temporal characteristics, which are primary in distinguishing short and long quantity. Further, in the only phonetic study investigating quantity marking in spontaneous speech that the authors are aware of, Asu *et al.* (2008) observed stable durational and pitch cues to the Estonian quantity distinction in altogether 348 disyllabic words produced by seven speakers.

Taken together, these studies indicate that, unlike standardly assumed, duration is not necessarily the only feature marking phonological quantity. Different languages employ various acoustic correlates and factors related to marking the quantity opposition, including syllable structure, tone, accent, stress, moraic structure, and semantic context, in addition to duration and pitch (Lehiste, 1965; Lippus *et al.*, 2007; Mixdorff *et al.*, 2002; Suomi *et al.*, 2003; Suomi, 2007; Vainio *et al.*, 2010; Yu, 2008). If this tendency is robust enough in the languages discussed above, then there is a high chance that Yakut exhibits the same interplay between pitch and duration in the quantity distinction as well.

The primary research question of the present study is: What are the acoustic correlates of the phonological distinction between short and long vowels in Yakut? As shown in the above overview, recent studies find that long and short vowels are contrasted not only based on duration but also by pitch cues. This article therefore investigates a possible role for F0 (maximum/minimum and slope), duration, and intensity. Experiment 1 analyzes the production of words with short and long vowels in Yakut unscripted spontaneous speech, while experiment 2 assesses these factors by using a carefully designed set of mono- and disyllabic words produced in carrier sentences in a laboratory setting.

II. METHODS

A. Experiment 1

In experiment 1, an acoustic analysis of long and short vowels from spontaneous Yakut speech taken from mostly monologue-based conversations was performed.

1. Participants

The participants were nine female native Yakut speakers from the age of 19 to 77 years old. The speakers were raised in an environment where Yakut is the dominant language and Russian is primarily acquired through formal education and media. All of the participants are Yakut-Russian bilinguals, but Yakut is their mother tongue. The participants reported no speech or hearing impairments.

2. Recordings

The speech data were obtained in fieldwork conducted by the first author and the participants were recorded with an MP3 Dictaphone (sampling frequency: 44.1 kHz). The speakers were asked to talk for about 10–15 min on a chosen topic they felt comfortable with. The only restriction for the recording session was not to mention peoples' names for ethical reasons and to protect their privacy. The author did not interrupt the participants' monologue-based speech during the recording sessions. Paused monologues were repaired by dialogue-based conversations facilitated by the author. The questions that were used were: What season do you like? What do you like to do in your free time? What is your favorite food? and so on. These types of questions were raised only in case a speaker did not know how to proceed with their monologues. The speakers were encouraged to speak as naturally as they could during the recording sessions.

3. Acoustic measurements

The recordings were segmented using the software Praat (Boersma and Weenink, 2014). From eight speakers' sessions the first author extracted 25 disyllabic verbs and nouns with short vowels each. In addition, 25 disyllabic verbs and nouns with initial, final or both syllables containing long vowels were selected for segmentation for each speaker. Hence, each speaker's session yielded a total of 50 segmented words. The aim was to reach a total of 400 words from all the participants, however, two speakers' sessions did not contain a sufficient amount of target words. Therefore, an additional ninth speaker's session was used to extract eighteen words to compensate for the shortfall. The resulting set of 400 words contained mostly nouns and verbs that were either in inflected or citation form. Some monosyllabic words appeared in the disyllabic form due to inflectional suffixation; these were also included in the analysis. Note that although diphthongs are regarded as long vowels, they were not included in the analysis.

In each word token, syllables and vowels were segmented. The following acoustic measurements were obtained: the duration of short and long vowels, maximum and minimum F0 values for each vowel, F0 slope, and intensity (mean intensity within the center of the vowel, maximum intensity within the center of the vowel and mean intensity for the whole vowel). Additional F0 measurements were conducted at ten equidistant time-points within each vowel as the basis for average time-normalized pitch contours. Hereafter, the materials of experiment 1 will be referred to as "interviews."

B. Experiment 2

Experiment 2 investigated whether the binary vowel quantity opposition is based on duration or pitch, or both, in carefully pronounced laboratory speech.

TABLE II. The four word types occurring in the 200-word list.

Word	Number of syllables	Syllable 1	Syllable 2
χɑ:l.t*ɯ*s "necktie"	2	Long	Short
bi.li: "knowledge"	2	Short	Long
ky:s.tɛ:χ "strong"	2	Long	Long
χɑ.rɑχ "eye"	2	Short	Short

1. Participants

One adult female native Yakut speaker, aged 34 (the first author), participated in the study. She is also bilingual in Russian, and regards Yakut as her strongest language. The participant grew up in a dominantly monolingual Yakut-speaking environment. In addition to Russian and Yakut, which were acquired during childhood and adolescence, she also speaks English fluently as her third language.

2. Materials

The speaker read a list of 200 disyllabic target words embedded in a carrier sentence. Four types of words with different sequences of short and long vowels in both syllable positions were included (see Table II). Onset and coda positions were not controlled. There were 50 tokens of each type of word.

The target words were either in their citation or inflected forms. Monosyllabic word stems appeared in disyllabic forms due to declension or conjugation (for instance, the monosyllabic word *tɯ*: "boat" in the nominative case becomes disyllabic *tɯ*:.*nɯ* in the accusative case).

Additionally, a list of 50 minimal pairs, i.e., 100 words in total, was also recorded (see Table III for examples). The list consisted of disyllabic and monosyllabic words containing contrasting long and short vowels. Different categories were included without a restriction to nouns and verbs only.

The word list containing 300 words total was combined and randomized.

3. Procedure and acoustic measurements

The recording session was conducted in a sound-treated booth at the University of Alberta Phonetics Laboratory. A Countryman Isomax Microphone Model E60W5LEV and a Korg MR-2000 recording device were used in the recording session (sampling frequency: 44.1 kHz). The speaker read out each word of the randomized word list in the carrier sentence *Biligin* _____ *diɛn tulɯ ɛt* "Say the word____now" twice. After reading the 300-word list, the speaker had a break for 10 min and read the list again in a different randomized order. There was a very short pause between each item, with the interval ranging from 2 to 3 s. The speaker was asked to read the words as naturally as possible.

TABLE III. Example items from 100-word list of minimal quantity pairs.

Word	Number of syllables	Syllable 1	Syllable 2
ɑ.tɑχ "leg"	2	Short	Short
ɑ.tɑ:χ "spoilt"	2	Short	Long
bɑs "head"	1	Short	N/A
bɑ:s "wound"	1	Long	N/A

Acoustic measurements were the same as in experiment 1. The materials of experiment 2 will be referred to as "lab recordings" hereafter in the paper.

III. RESULTS

All acoustic measures were analyzed using linear mixed-effects modeling as implemented in the package lme4 for the statistical analysis program R (Bates *et al.*, 2014; R Development Core Team, 2014). These models assess the significance of independent variables as predictors (fixed effects), and additionally include random variables to, for example, take variation between participants and items into account (Baayen *et al.*, 2008; Jaeger, 2008). Modeling of the present data tested a number of experimental factors and random variables in order to determine the statistical model that had the best fit to the data. Henceforth, only the best model for each tested variable will be reported. Each ultimate model for a given dependent variable emerged following the same procedure in steps. First, predictors were tested, retaining only variables that significantly improved model fit. Models containing different variables were compared with respect to their log likelihood as a measure of goodness of fit, using the analysis of variance function (Baayen, 2008). After the best model for predictors was found, that model was compared against models with different random effects. Tested predictor variables were vowel quantity (levels: long vs short quantity) and syllable position (levels: first vs second syllable). Tested random effects were word and vowel (modeling variation between different lexical items and vowel qualities, respectively). For the interview data, random by-speaker variation was also considered. In the interpretation of the best models, differences associated with a *t*-value above 2 were considered significant (Baayen *et al.*, 2008).

A. Duration

1. Interviews

Figure 1 shows the mean values and 95% confidence intervals for vowel duration of short and long quantity vowels by syllable position. It demonstrates that there was a strong distinction in duration between the two quantities, with overall shorter durations in short quantity (quantity 1) than in long quantity (quantity 2). As the data were taken from spontaneous speech samples, it showed that speakers might vary the duration for both quantities, and the durations for short (Q1) and long quantity (Q2) could even overlap for outlier values. As far as the syllable number is concerned, Fig. 1 illustrates a slightly longer duration overall in the second syllable position for both quantities.

The best linear mixed-effects model of vowel duration in the interview data included both quantity and syllable number as significant predictors ($p = 2.43 \times 10^{-5}$, $\chi^2 = 17.8$ and $p = 4.607 \times 10^{-5}$, $\chi^2 = 16.6$, respectively). It indicated that vowel duration in the long quantity (quantity 2) was significantly longer than in the short quantity (quantity 1; estimate = 43.323, SE = 8.032, $t = 5.394$). Additionally,

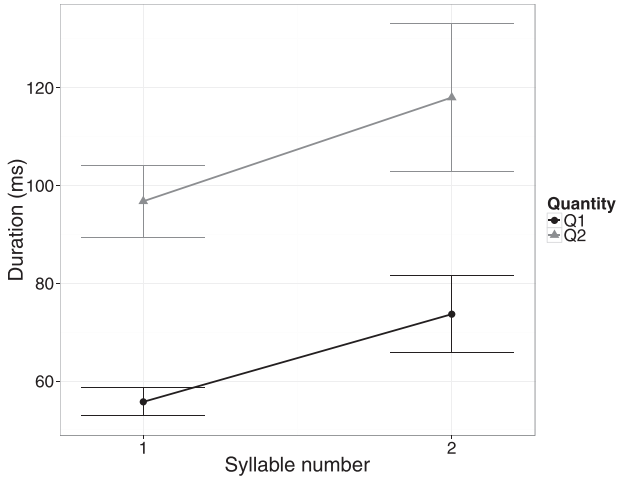

FIG. 1. Vowel duration of short and long quantity by syllable position (interviews).

vowels in the second syllable (syllable 2) were longer than vowels in the first syllable (syllable 1), thus syllable position proved to be a significant predictor of vowel duration (estimate = 15.144, SE = 3.523, $t = 4.299$). A model with an interaction between these predictors was not significantly better ($p = 0.902$, $\chi^2 < 1$).

2. Lab recordings

Figure 2 shows vowel duration in the short and long quantities by syllable position for the read speech sample. Quantity 2 vowels were consistently longer than quantity 1 vowels. This distinction was clearer than in the interview data. Figure 2 also illustrates a tendency for vowels in the second syllable position to be longer in duration compared to the first syllable position, which is likewise more pronounced than for the interview data.

The best statistical model of vowel duration for the lab recordings data had both quantity and syllable position as significant predictors ($p = 6.837 \times 10^{-14}$, $\chi^2 = 56.1$ and $p = 2.2 \times 10^{-16}$, $\chi^2 = 674.7$, respectively). There was no significant interaction between the predictors ($p = 0.304$, $\chi^2 = 1.1$). The model indicated a robust distinction in duration between the short and long quantities (estimate = 127.131, SE = 3.444, $t = 36.92$), showing that quantity 2 was significantly longer in duration than quantity 1. Similar to the results for the interview data, vowels in the second syllable position were consistently longer than vowels in the first syllable position (estimate = 26.403, SE = 3.316, $t = 7.96$).

B. F0 measurements

1. Interviews

Measurements of maximum and minimum F0 were similar for the interview data. For this reason, the maximum F0 only was used as a dependant variable in fitting a statistical model of F0 for the interviews (cf. Fig. 3). The resulting best statistical model of F0 for the interview data included both quantity and syllable number as predictors ($p = 7.615 \times 10^{-5}$, $\chi^2 = 15.7$ and $p = 0.041$, $\chi^2 = 4.1$, respectively), but a model with an interaction between them was not significantly better ($p = 0.259$, $\chi^2 = 0.61$). The model indicated that the F0 maximum was significantly lower in the second syllable as compared to the first syllable (estimate = −8.768, SE = 2.191, $t = -4.00$), as well as being lower in quantity 2 vowels than in quantity 1 vowels (estimate = −5.639, SE = 2.756, $t = -2.05$).

As an illustration, Fig. 4 shows time-normalized mean F0 contours for first and second syllable vowels in the disyllabic words produced by Yakut speakers in spontaneous speech. The numbers on the x axis indicate the ten measurement points for each syllable. Plot symbols and line colors indicate vowel quantity in the first and second syllable. Figure 4 shows relatively flat pitch contours in both syllable positions. There was, however, a slight lowering of F0 in both quantities in the first syllable. By contrast, in the second syllable position, both quantities 1 and 2 followed a relatively flat pattern. Moreover, F0 was lower in quantity 2 than in quantity 1, in line with the significant negative effect. Although the difference between the quantities is more clearly visible in the second syllable, it was consistent in both syllable positions, in agreement with the model not finding an interaction.

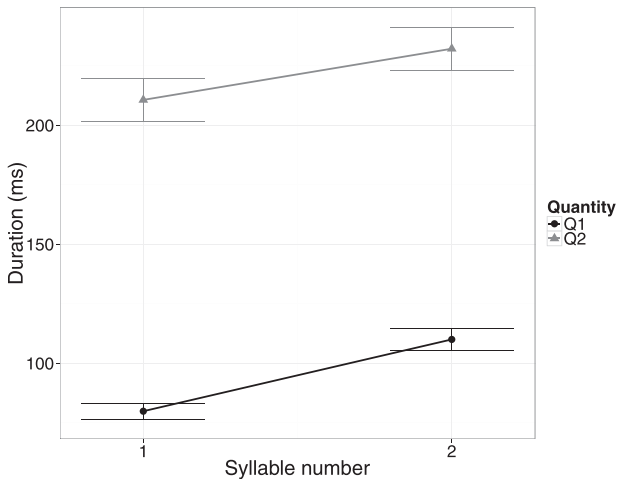

FIG. 2. Vowel duration of short and long quantity by syllable position (lab recordings).

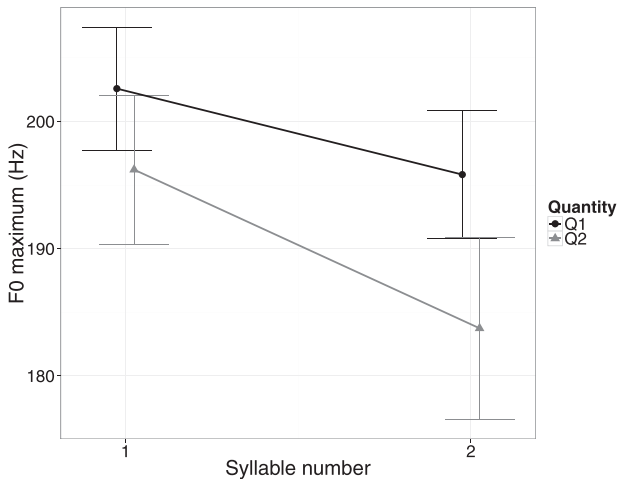

FIG. 3. Maximum F0 values of short and long quantity by syllable position (interviews).

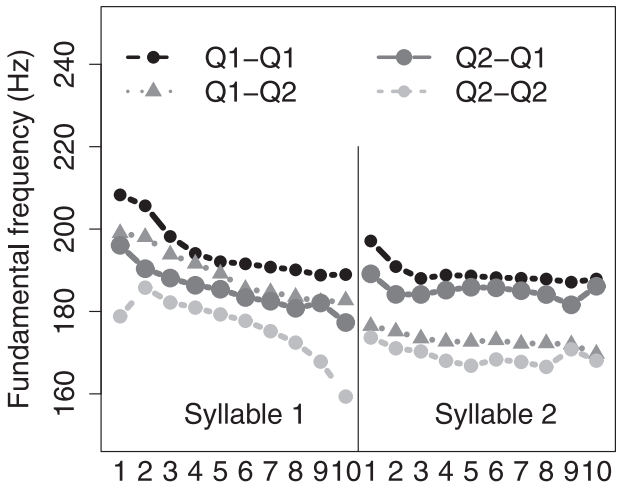

FIG. 4. Average time-normalized pitch contours by syllable position for vowels in short and long quantities (interviews).

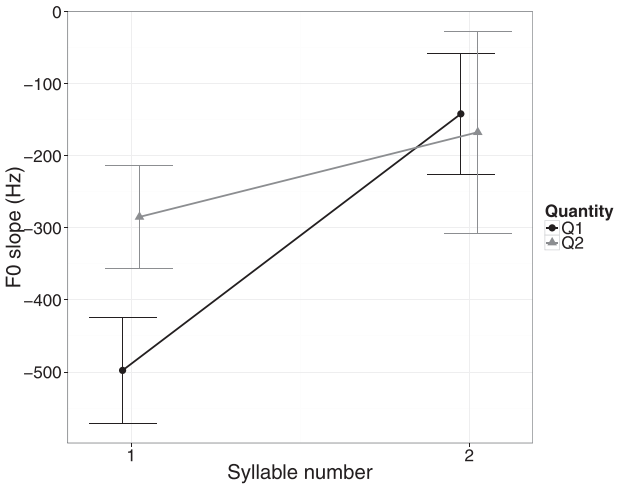

FIG. 5. F0 slope of vowels in short and long quantity by syllable position (interviews).

Since the effect is rather small, however, a relevant question is whether quantity impacted F0 in the same way for all of the speakers. In order to test this, quantity was adjusted for each speaker as a random by-speaker effect in an alternative mixed-effects model. However, this more complex model was not significantly better than the reported model of maximum F0 ($p = 0.955$, $\chi^2 = 0.1$). Thus, there was no indication for a significant variation between the speakers with respect to the effect of quantity on F0.

To assess the variation of the F0 movement visible in Fig. 4 more directly, we analyzed F0 slope. For calculating the F0 slope, minimum F0 was subtracted from maximum F0 and the result was divided by the result of subtracting the time of the F0 minimum from the time of the F0 maximum. The values of the F0 slope make it possible to trace the general direction of F0 contours, as well as the extent of F0 movement within a certain timeframe. Positive values are associated with a rise and negative values with a fall. Steeper F0 contours are shown as larger slope values. The analysis discarded outlier values above 4000 Hz/s and below −4000 Hz/s (0.5% of the data).

Figure 5 illustrates the F0 slope for both quantities by syllable position. As visible in Fig. 4, vowels in the first syllable position displayed slightly falling F0 for both quantities, resulting in negative slope values. Overall, quantity 2 carried a less steep fall than quantity 1 vowels, resulting in less negative slope values. Figure 5 moreover displays a difference between syllable positions, with steeper falls in first syllables. However, this difference was noticeably smaller for quantity 2 vowels than for quantity 1 vowels. At the same time, the two quantities had overlapping confidence intervals and very similar mean values close to 0 in the second syllable.

Accordingly, the best model of F0 slope for the interviews data included an interaction of quantity and syllable number ($p = 0.006$, $\chi^2 = 7.5$). The negative estimate for the intercept (estimate = −503.95, SE = 63.98, $t = -7.877$) reflects the fact that F0 overall had a tendency to fall. The model further suggests that the F0 slope was significantly less negative in quantity 2 (estimate = 237.87, SE = 66.45, $t = 3.580$) and in the second syllable position (estimate = 368.71, SE = 52.22, $t = 7.061$). However, the interaction between the two factors was associated with a negative estimate, indicating that both effects were significantly reduced when combined (estimate = −268.04, SE = 97.17, $t = -2.758$). Thus, the difference between the quantities was neutralized in the second syllable, where F0 slopes where flat for both quantities, as visible in Fig. 4.

2. Lab recordings

In contrast to the interview data, the data from the lab recordings displayed two different patterns for maximum and minimum F0 values (see Figs. 6 and 7). Maximum F0 measurements were very similar across all conditions, although values were slightly lower for the long quantity and in the second syllable position. Minimum F0 measurements were likewise lower in quantity 2, but showed higher values in the second syllable position. Overall, minimum F0 measurements had clearer differences between the conditions.

Since measurements of minimum and maximum F0 values were different from each other, both of them were analyzed. For maximum F0, the linear mixed-effects model

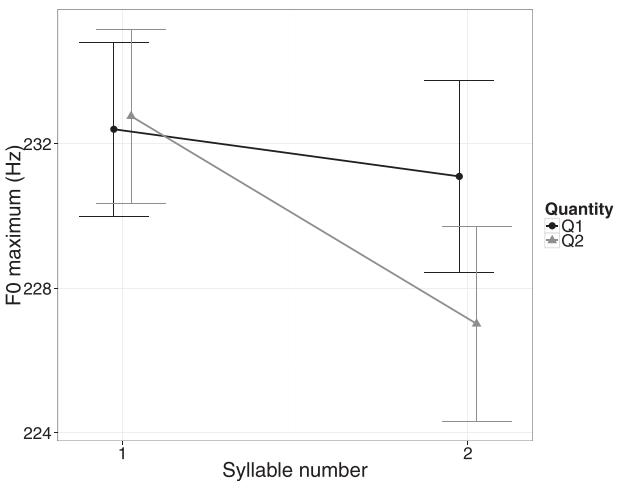

FIG. 6. Maximum F0 values of short and long quantity by syllable position (lab recordings).

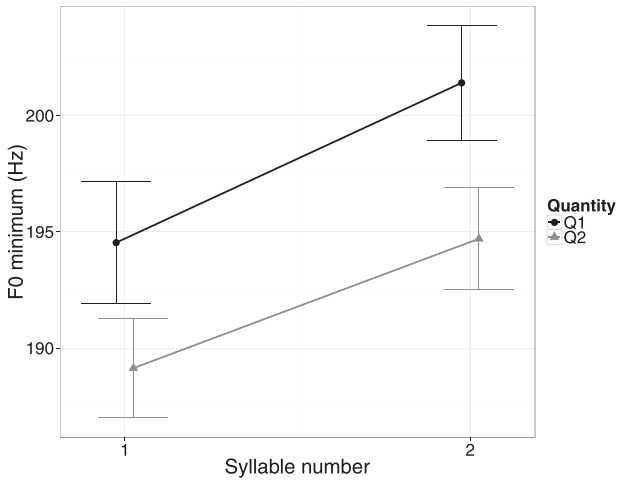

FIG. 7. Minimum F0 values of short and long quantity by syllable position (lab recordings).

analysis showed no significant difference between any of the models. Thus, the best model was the simplest model with no fixed predictors, including only random factors word and vowel. Unlike the best model of maximum F0 for the interviews, where syllable number and quantity were significant predictors, there was no effect of the experimental factors for maximum F0 values obtained from one speaker's production of words in stable carrier sentences. Most of the variance visible in Fig. 6 was thus explained by the variation between different words and vowels.

By contrast, the best model of minimum F0 included syllable number as a significant predictor ($p = 1.529 \times 10^{-13}$, $\chi^2 = 54.5$). Adding quantity only marginally improved the model fit ($p = 0.066$, $\chi^2 = 3.4$), and an interaction between the predictors was not significant ($p = 0.834$, $\chi^2 < 1$). The model indicated that minimum F0 values for the lab recordings were marginally lower in quantity 2, and significantly higher in the second syllable (estimate = 9.659, SE = 1.049, $t = 9.21$).

Figure 8 shows average time-normalized pitch contours. It is structured like Fig. 4, but additionally displays contours for monosyllabic words, which resembled those of second

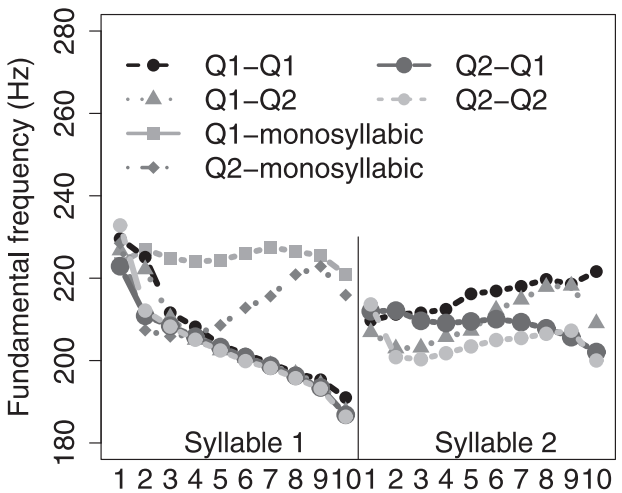

FIG. 8. Average time-normalized pitch contours by syllable position for vowels in short and long quantities (lab recordings).

syllables in disyllabic words. But also for disyllabic words, mean F0 contours differed from those of the interview data. Here, both quantities followed a falling pattern in the first syllable, with virtually congruent time-normalized contours. Moreover, whereas F0 contours for both quantities were relatively flat in the second syllable position in the interview data, once Yakut disyllabic words were produced in a constant carrier phrase, they frequently showed rising F0 on the second syllable, especially in quantity 2. In quantity 1, a much flatter rise appeared when the preceding syllable also contained a quantity 1 vowel, whereas quantity 1 vowels preceded by quantity 2 vowels showed a minimal rise followed by a small fall, similar to quantity 1 vowels in monosyllabic words.

Correspondingly, the best model of the F0 slope for the lab recordings (calculated the same way as for the interview data, excluding 7% of the data as outliers according to the same criterion) included a significant interaction between quantity and syllable number ($p = 0.006$, $\chi^2 = 7.44$).

Like for the interview data, the model included a negative intercept (estimate = −595.61, SE = 42.14, $t = -14.136$), reflecting an overall tendency toward falling F0, while indicating that quantity 2 vowels had less steeply falling slopes (estimate = 367.20, SE = 61.90, $t = 5.932$) than quantity 1 vowels. Slopes in the second syllable were more positive than in the first syllable (estimate = 388.49, SE = 61.68, $t = 6.298$). As for the interview data, the interaction between the factors was associated with a negative estimate (estimate = −259.41, SE = 92.85, $t = -2.794$), suggesting that each effect was weakened when combined with the other one. Figure 8 shows that first syllable vowels in both quantities realized almost identical falls (with the exception of monosyllabic words), but since quantity 1 vowels had shorter durations, the same fall corresponded to a steeper slope, as illustrated in Fig. 9. By contrast, time-normalized contours diverged in second syllables, and the longer rises in quantity 2 spanned a wider F0 range. Note also that whereas the mean F0 contours in Figs. 4 and 8 differ, slope values displayed a similar pattern for both data sets.

Finally, as mean F0 contours of monosyllabic words resembled those in second syllables of disyllabic words, mostly displaying a rise that was larger in quantity 2, we fitted another model with stress as a predictor instead of syllable number. Since stress is word-final in Yakut, this grouped monosyllabic words with second syllables. The resulting model contained more pronounced effects and provided a better fit than the model with syllable number as a predictor ($p = 2.2 \times 10^{-16}$, $\chi^2 = 34.6$). However, it confirmed the same pattern as the previous model, i.e., significant positive main effects of quantity ($p = 4.107 \times 10^{-7}$, $\chi^2 = 25.6$; estimate = 469.17, SE = 67.08, $t = 6.994$) and stress ($p = 7.871 \times 10^{-14}$, $\chi^2 = 55.8$; estimate = 525.79, SE = 59.21, $t = 8.881$), as well as a significant interaction between them ($p = 1.958 \times 10^{-5}$, $\chi^2 = 18.2$; estimate = −398.37, SE = 89.81, $t = -4.436$).

C. Intensity

1. Interviews

In order to investigate the effect of vowel quantity on intensity in Yakut, this section evaluates three measurements of intensity: first, mean intensity within the center of vowels,

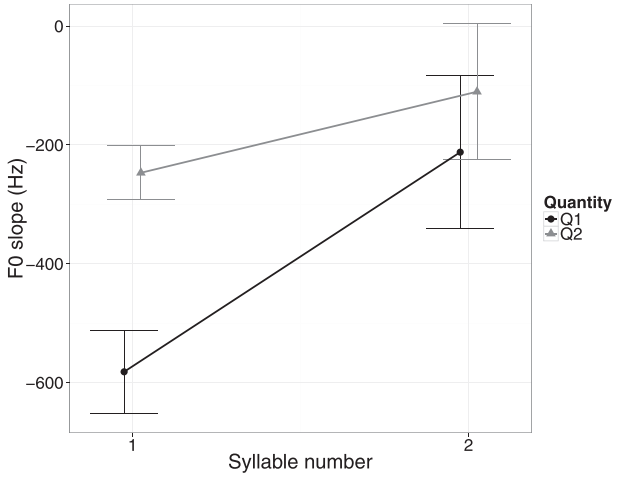

FIG. 9. F0 slope for vowels in short and long quantity by syllable position (lab recordings).

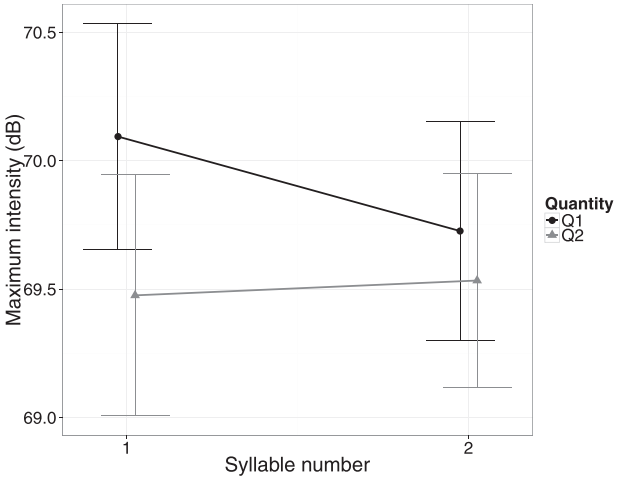

FIG. 11. Maximum intensity for the center of vowels (lab recordings).

i.e., measured during the middle 50% of the vowel duration to exclude any influence of the consonants; second, maximum intensity for the center of vowels; and third, mean intensity for the whole vowel. Interestingly, all the measurements' results looked very similar. Therefore, the analysis concentrated on maximum intensity within the center of the vowel (Fig. 10).

Model comparisons determined that the best model of intensity for the interview data included quantity as its only predictor. Adding syllable number as a predictor did not result in an improved fit ($p = 0.986$, $\chi^2 < 1$), and neither was an interaction between quantity and syllable number found ($p = 0.967$, $\chi^2 < 1$).

The model suggested that measurements of maximum intensity were significantly higher in quantity 2 than in quantity 1 (estimate = 0.4924, SE = 0.2010, $t = 2.45$), a trend that is visible in Fig. 10. In general, the effect of Yakut quantity on intensity was however very small in the data obtained from the interviews, with considerable overlap of measurements and very similar mean values in all conditions. Still, the model provided a better fit to the data than a model with no predictors ($p = 0.026$, $\chi^2 = 5.0$), and was likewise superior to a model with syllable number as its only predictor ($p = 2.2 \times 10^{-16}$, $\chi^2 = 5.0$).

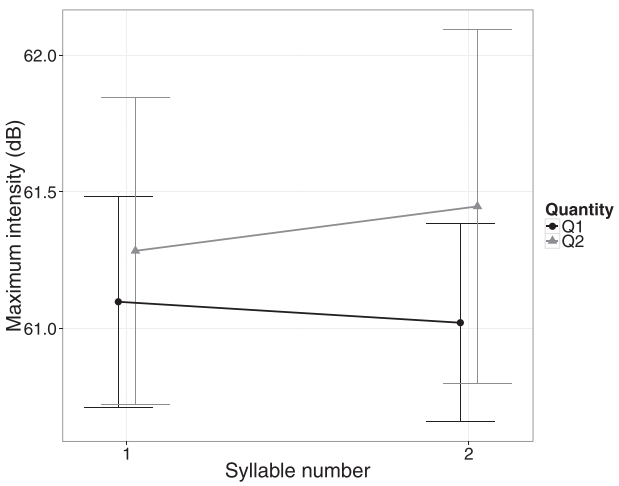

FIG. 10. Maximum intensity for the center of vowels (interviews).

2. Lab recordings

This section evaluates the same three measurements of intensity as analyzed for the interview recordings. Again, results for all three measurements looked quite similar, so statistical analyses concentrated on maximum intensity within the center of the vowel (Fig. 11).

The best linear mixed-effects model included only syllable number as a predictor ($p = 1.162 \times 10^{-5}$, $\chi^2 = 19.2$). Adding the factor quantity or an interaction between both factors did not improve model fit ($p = 0.586$, $\chi^2 = 0.3$ and $p = 0.129$, $\chi^2 = 4.1$, respectively). The model including syllable number alone further outperformed a model that had only quantity as a predictor ($p = 2.2 \times 10^{-16}$, $\chi^2 = 18.6$). Intensity was a little lower in the second syllable position (estimate = −0.9175, SE = 0.2050), and this result was significant ($t = -4.48$). In contrast to the interview recordings, where quantity was a significant predictor of maximum intensity, syllable position was a better predictor in the lab recordings. Notice, however, that as for the interview data, all three measurements of intensity showed a large overlap between all conditions.

IV. DISCUSSION AND CONCLUSION

This study was aimed at revealing phonetic correlates of vowel quantity in Yakut. The binary opposition between long and short vowel quantity was described based on acoustic measurements from two types of data, spontaneous speech ("interviews"), and controlled read production in a phonetics laboratory ("lab recordings"). Since similar data were extracted from spontaneous speech and lab recordings, and both data sets were analyzed the same way, this allowed for an accurate comparison of the acoustic correlates of quantity in the two types of speech. This allowed us to investigate the acoustic correlates that remain stable across speech types and can thus be considered primary cues to quantity, as well as the correlates that vary under the influence of other factors.

The present study thereby differed from existing studies of quantity, which have almost exclusively been based on data recorded in controlled environments. Moreover, the current study investigated Yakut, an understudied language on which no systematic acoustic research has been conducted so far. Interestingly, while the study was aimed at finding correlates of quantity, its results shed some light on other aspects of Yakut prosody, as well.

This section will discuss findings for duration, intensity, and F0 in turn. Previous statements about vowel quantity in Yakut solely mention durational differences as a correlate of quantity (Krueger, 2012). As expected from this assessment, measurements of duration in both interviews and lab recordings showed a robust durational difference between short and long quantity. That is, long vowels were significantly longer than their short counterparts. This distinction was clearer in the lab recordings, where target words were produced in identical frame sentences by a single speaker. Nevertheless, the durational contrast was clearly maintained for the interview data, as well. By contrast, larger differences between the two types of speech appeared for the other two investigated acoustic parameters, as discussed below.

While the data showed a consistent durational marking of quantity, vowel durations were not influenced by quantity alone. In addition, syllable number appeared to be a significant predictor of duration. In both interview data and lab recordings, vowels in both quantities had longer durations in the second syllable position than in the first syllable. This finding can be related to an observation by Krueger (2012), who notes that the second vowel is slightly longer when two successive syllables contain the same vowel, for example, in the word *kykyr* "large." Interestingly, a consistent vowel lengthening in the second syllable was observed in the present study, generalizing Krueger's finding across different vowel phonemes (which were taken into account as random effects) and both quantities. Since consistently longer durations appeared in final syllables, corresponding to the stress position in Yakut (Anderson, 1995, 1998), it is likely that lengthening is a correlate of stress and/or accent, although Samsonova described Yakut stress as not distinctly prominent articulatorily (Samsonova, 1959, p. 21; but note that Krueger, 2012, referred to the second vowel in words like *kykyr* as "accented"). Further studies are needed to determine how stress and/or accent affect duration in Yakut. In addition, a more detailed analysis should control syllable shape, particularly moraic structure, since compensatory lengthening of stressed vowels has been observed, e.g., in Washo (Yu, 2010; also see Mixdorff *et al.*, 2002, on lengthening of codas with preceding short vowels in Thai).

In addition to being consistently marked by duration, quantity also showed effects on the other two acoustic parameters, intensity and F0. Here, differences between read and spontaneous speech were larger than for duration. Both data sets exhibited only small intensity differences and large overlap between conditions. Analyses suggested a significant effect of quantity for the interview data, with higher intensity in long vowels, which afforded more durational space for reaching higher values, but there was no indication of a quantity effect for the lab recordings. Instead, intensity was significantly lower for second syllable vowels than for those in first syllables, indicating a clear intensity downtrend over the course of the word and, potentially, the utterance (on intensity declination and its relation with duration and F0, e.g., see Pierrehumbert, 1979; Streeter, 1978; Strik and Boves, 1995).

Even more noticeably, F0 patterns differed between spontaneous and read speech. In the spontaneous data, pitch contours appeared at an overall lower level in quantity 2 than in quantity 1, reflected by significantly lower maximum F0 values. This effect was absent from the lab data. However, the two data sets not only differed in F0 height, but also in the shape of the F0 contours. Interestingly, analyses of F0 slope revealed the same pattern for both data sets: more positive slope values in quantity 2 and for final syllables, but an interaction associated with a negative estimate. As the normalized F0 contours indicated, the pattern was due to falling F0 on the first syllables of disyllabic words, with the same fall translating to a steeper slope in the shorter quantity 1 vowels, and non-falling F0 on second syllables. The difference between the data sets was thus localized to second, final syllables. Whereas the interview data showed uniformly flat averaged F0 in both quantities, the lab recordings mostly displayed clear rising contours in quantity 2 and a slight rise or rise-fall in quantity 1, depending on the quantity of the preceding vowel.

The differences between F0 contours of spontaneous and read speech are most likely explained by utterance- or phrase-level prosodic characteristics affecting F0 of individual words. Whereas words were extracted from various parts of spoken phrases for the interview data, target words were consistently produced phrase-medially in a carrier sentence for the lab recordings. In particular, it is likely that the speaker interpreted and realized the target words as being in narrow focus (for a definition and semantic account of focus, e.g., see Krifka, 2008). Reading a list of near-identical sentences that only differ in one target word will almost invariably cause the speaker to realize this word as the focus of the sentence, unless she imagines her utterances as answers to questions focusing a part of the frame sentence every time (e.g., *Should I say the word____later? – Say the word____NOW*). Thus, it is likely that the F0 rise visible in the lab recordings constituted a prosodic correlate of focus, especially since it consistently appeared on the stressed syllable.

Since the data did show effects of quantity on F0, the question arises whether F0 is an independent quantity correlate, in contradiction to the statement by Krueger (2012), or whether these effects are simply a by-product of the durational differences. As discussed by Yu (2010), durational difference and F0 are related in many languages; namely, vowels that have rising tones are longer and vowels with falling tones are shorter, while vowels with low tones have longer durations than those with high tones. Both data sets analyzed here roughly conformed to these generalizations: First syllables showed a falling F0 and shorter vowel durations in both quantities for both data sets. In second syllables, where durations were longer, F0 contours rose slightly for most conditions in the lab recordings and were flat on a low level in the interview data. Another consistent pattern in

the relation between F0 and duration is an association of static F0 with shorter duration on the one hand and of dynamic F0 with longer duration on the other hand (Yu, 2010). Also in non-tonal languages like Finnish, this association influences quantity perception (Kinoshita *et al.*, 2002; Järvikivi *et al.*, 2010), and is likewise reflected in production (Vainio *et al.*, 2010). The present finding of steeper falls for quantity 1 on first syllables at first glance seems at odds with this generalization. However, the mean F0 contours suggested that instead of an inherent quantity difference, this finding simply reflects the fact that a similar F0 fall was realized in a shorter time for quantity 1 vowels. Thus, second syllables might be a more likely candidate for an inherent distinction between static and dynamic contours as cues to quantity: In the lab recordings, vowels in the long quantity carried rising F0, whereas F0 was more flat for quantity 1 vowels. This pattern was a mirror image of the study by Vainio *et al.* (2010) where Finnish speakers produced heavy syllables with a falling F0 contour and light syllables with a more static pitch in the word-initial position. As stress is initial in Finnish, but final in Yakut, quantity differences in F0 in both languages appeared in prosodically prominent locations. Noticeably, all other studies finding F0 cues to quantity in Finnish have focused on first syllables, as well. Contra the description by Vainio *et al.* (2010) of different tonal targets, several authors have concluded that F0 is not a primary cue to quantity in Finnish, but a secondary cue underlining duration via the alignment of a uniform tonal contour with the phoneme string: When the initial vowel only carries static high pitch, it has to be short, because a long vowel would have provided time for a fall (Arnhold, 2014; O'Dell, 2003; Suomi, 2005, 2009). In line with this, Finnish native speakers base their quantity judgments on F0 cues only when durational information is ambiguous (Järvikivi *et al.*, 2010). That F0 in Finnish is a secondary cue that co-varies with duration is further exemplified by data showing that native Finnish listeners are unable to use F0 cues to distinguish quantity 2 and quantity 3 in Estonian (Lippus *et al.*, 2009), where F0 functions as the primary cue for this distinction, whereas duration is crucial for the contrast between quantity 1 and quantity 2 (Lippus *et al.*, 2007, 2009, 2011).

As no uniform tonal correlates of quantity appeared across the syllable positions and data types in the present analyses, whereas durational distinctions were stable even in spontaneous speech, it seems unlikely that pitch is an independent cue to quantity in Yakut. Thus, like for Finnish, a uniform postulation of distinct tonal targets for each quantity—one F0 contour expected to appear on all quantity 1 vowels and one expected on all quantity 2 vowels like lexical tones—seems inappropriate. Rather, the present findings suggest that quantity differences in F0 may arise from the interaction of duration with F0 movements associated with word- or utterance-level prosody, although more research is needed to confirm this interpretation. Even if duration is likely the primary acoustic correlate of vowel quantity, native Yakut speakers might well be sensitive to and cued by pitch contours in cases where durational cues are less accessible, like Finnish and Japanese listeners (Järvikivi *et al.*, 2010; Kinoshita *et al.*, 2002)—especially since in these languages F0 could be a fairly systematic co-variant of duration in the quantity distinction. This is even more likely since peoples' perception and word recognition have been shown to be very sensitive to a variety of phonetic detail (e.g., Hawkins, 2003; Smith *et al.*, 2012). Ultimately, perception studies are needed to investigate whether (and how) pitch cues affect Yakut listeners' identification of vowel quantity.

V. CONCLUSION

This study demonstrated that the Yakut vowel quantity distinction is consistently marked through duration, in read speech producing a list of target words in invariable frame sentences as well in spontaneous speech. However, it also found effects of quantity on F0 and, to a lesser extent, on intensity. It is thus possible that F0 functions as a secondary cue or covariant of duration in Yakut, similarly as in Finnish and Japanese. On initial syllables of disyllabic words in read as well as spontaneous speech, both quantities showed nearly identical F0 falls that corresponded to a steeper slope for the shorter quantity. On final, stressed syllables, F0 contours showed larger differences between quantities as well as between the two data sets. In particular, while F0 was flat in spontaneous speech, with a lower level for long vowels, an F0 rise appeared in read speech, which was more extensive in long quantity. This difference between the two types of data was likely connected to the production of target words in invariable frame sentence for the read speech data, suggesting that the F0 rise observed on the stressed syllable is a prosodic correlate of narrow focus. However, since neither overall pitch movements within an utterance nor prosodic focus marking have been investigated for Yakut, more research is needed to confirm this hypothesis.

Anderson, G. D. S. (**1995**). "Diachronic aspects of Russianisms in Siberian Turkic," Annu. Meet. Berkeley Linguist. Soc. **21**(1), 365–376.

Anderson, G. D. S. (**1998**). "Historical aspects of Yakut (Saxa) phonology," Turkic Languages **2–3**, 1–32.

Arnhold, A. (**2014**). "Finnish prosody: Studies in intonation and phrasing," Ph.D. thesis, Goethe University, Frankfurt am Main, Germany, 382 pp.

Asu, E. L., Lippus, P., Teras, P., and Tuisk, T. (**2008**). "The realization of Estonian quantity characteristics in spontaneous speech," in *Nordic Prosody. Proceedings of the Xth Conference, Helsinki 2008*, Helsinki, Finland, edited by M. Vainio, R. Aulanko, and O. Aaltonen (Peter Lang, Frankfurt am Main), pp. 50–56.

Baayen, R. H. (**2008**). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R* (Cambridge University Press, Cambridge, UK), 368 pp.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**(4), 390–412.

Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and S4. R package (version 1.1-6)," http://cran.r-project.org/package=lme4 (Last viewed June 20, 2014).

Boersma, P., and Weenink, D. (**2014**). "Praat: Doing phonetics by computer (version 5.3.75) [computer program]," http://www.praat.org/ (Last viewed July 11, 2014).

Finch, R. (**1985**). "Vowel harmony in Yakut," Sophia Ling. **18**, 1–17.

Fox, R. A., and Lehiste, I. (**1987**). "Discrimination of duration ratios by native English and Estonian listeners," J. Phonetics **154**, 349–363.

Hawkins, S. (**2003**). "Roles and representations of systematic fine phonetic detail in speech understanding," J. Phonetics **31**(3), 373–405.

Jaeger, T. F. (**2008**). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed model," J. Mem. Lang. **59**(4), 434–446.

Järvikivi, J., Aalto, D., Aulanko, R., and Vainio, M. (**2007**). "Perception of vowel length: Tonality cues categorization even in a quantity language,"

in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, edited by J. Trouvain and W. J. Barry, pp. 693–696.
Järvikivi, J., Vainio, M., and Aalto, D. (**2010**). "Real-time correlates of phonological quantity reveal unity of tonal and non-tonal languages," PloS One. **59**, e12603, 1–10.
Kaun, A. R. (**1995**). "The typology of rounding harmony: An optimality theoretic approach," Ph.D. thesis, University of California, Los Angeles, CA, 206 pp.
Kinoshita, K., Behne, D. M., and Arai, T. (**2002**). "Duration and F0 as perceptual cues to Japanese vowel quantity," Perception **1454**, 1–4.
Krifka, M. (**2008**). "Basic notions of information structure," Acta Ling. Hungarica **55**(3-4), 243–276.
Krueger, J. R. (**2012**). *Yakut Manual, Uralic and Altaic Ser*., Vol. 21 (Routledge, New York), 389 pp.
Lehiste, I. (**1965**). "The function of quantity in Finnish and Estonian," Language **41**(3), 447–456.
Lehiste, I., Teras, P., Pajusalu, K., and Tuisk, T. (**2007**). "Quantity in Livonian: Preliminary results," Ling. Uralica **1**, 29–44.
Lippus, P., Pajusalu, K., and Allik, J. (**2007**). "The tonal component in perception of the Estonian quantity," in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, edited by J. Trouvain and W. J. Barry, pp. 1049–1052.
Lippus, P., Pajusalu, K., and Allik, J. (**2009**). "The tonal component of Estonian quantity in native and non-native perception," J. Phonetics **37**(4), 388–396.
Lippus, P., Pajusalu, K., and Allik, J. (**2011**). "The role of pitch cue in the perception of the Estonian long quantity," in *Prosodic Categories: Production, Perception and Comprehension*, edited by S. Frota, G. Elordieta, and P. Prieto (Springer, Dordrecht, the Netherlands), pp. 231–242.
Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., and Charnivivit, P. (**2002**). "Perception of tone and vowel quantity in Thai," in *Proceedings of International Conference on Spoken Language Processing (ICSLP2002)*, pp. 753–756.
Nakai, S., Kunnari, S., Turk, A., Suomi, K., and Ylitalo, R. (**2009**). "Utterance-final lengthening and quantity in Northern Finnish," J. Phonetics **37**(1), 29–45.
O'Dell, M. (**2003**). *Intrinsic Timing and Quantity in Finnish* (Tampere University Press, Tampere, Finland).
Pakendorf, B. (**2007**). "Contact in the prehistory of Sakha (Yakuts): Linguistic and genetic perspectives," Ph.D. thesis, LOT, Utrecht, The Netherlands, 375 pp.
Pierrehumbert, J. (**1979**). "The perception of fundamental frequency declination," J. Acoust. Soc. Am. **66**(2), 363–369.
R Development Core Team (**2014**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria), http://www.R-project.org (Last viewed October 4, 2014).
Sasa, T. (**2009**). "Treatment of vowel harmony in optimality theory," Ph.D. thesis, University of Iowa, Iowa City, IA, 220 pp.
Samsonova, T. P. (**1959**). *Comparative Characteristics of the Sound System of the Russian and Yakut Languages* (Iakutskoe kniz˙noe izdatel'stvo, Yakutsk, Russia), 92 pp. (in Russian).
Smith, R., Baker, R., and Hawkins, S. (**2012**). "Phonetic detail that distinguishes prefixed from pseudo-prefixed words," J. Phonetics **40**(5), 689–705.
Streeter, L. A. (**1978**). "Acoustic determinants of phrase boundary perception," J. Acoust. Soc. Am. **64**(6), 1582–1592.
Strik, H., and Boves, L. (**1995**). "Downtrend in F0 and Psb," J. Phonetics **23**(1), 203–220.
Suomi, K. (**2005**). "Temporal conspiracies for a tonal end: Segmental durations and accentual f0 movement in a quantity language," J. Phonetics **33**(3), 291–309.
Suomi, K. (**2007**). "On the tonal and temporal domains of accent in Finnish," J. Phonetics **35**(1), 40–55.
Suomi, K. (**2009**). "Durational elasticity for accentual purposes in Northern Finnish," J. Phonetics **37**(4), 397–416.
Suomi, K., Toivanen, J., and Ylitalo, R. (**2003**). "Durational and tonal correlates of accent in Finnish," J. Phonetics **31**(1), 113–138.
Vainio, M., Järvikivi, J., Aalto, D., and Suni, A. (**2010**). "Phonetic tone signals phonological quantity and word structure," J. Acoust. Soc. Am. **128**(3), 1313–1321.
Ylinen, S., Shestakova, A., Alku, P., and Huotilainen, M. (**2005**). "The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish," Language Speech **48**(3), 313–338.
Yu, A. C. L. (**2008**). "The phonetics of quantity alternation in Washo," J. Phonetics **36**(3), 508–520.
Yu, A. C. L. (**2010**). "Tonal effects on perceived vowel duration," Lab. Phonol. **10**, 151–168.