Towards a phonological model of Uyghur intonation

Travis Major, Connor Mayer

University of California, Los Angeles, USA

tjmajor@ucla.edu, connormayer@ucla.edu

Abstract

In this paper we present a preliminary intonational model for Uyghur (Turkic: China). We use acoustic measurements to support the claim that Uyghur is a stress language that only uses edge-marking intonation. Although this is not unattested in the literature, to our knowledge this is the first AM model of such a language.

Index Terms: intonation, prosody, Uyghur, autosegmental-metrical, AM, ToBI, stress

1. Introduction

The goal of this paper is to provide an initial sketch of a model of intonational phonology of Uyghur in the AM (autosegmental-metrical) framework (e.g. [1], [2], [3]). The AM theory proposes that the continuous pitch contour of utterances can be broken down into a string of discrete pitch targets that consist only of high (H) or low (L) tones, or complex combinations of the two (e.g. LH or HL). These tones are associated with particular parts of the segmental string in two ways: head-marking tones, or pitch accents, associate with a prominent syllable or mora, while edge-marking or boundary tones associate with the edges of prosodic constituents. Phonetic interpolation determines the pitch contour between tonal targets. The AM model provides a useful set of theoretical assumptions for analyzing the intonational systems of languages, and the analysis of Uyghur in this framework will allow for typological comparisons with the systems of other languages. Because Uyghur intonation has not been formally modeled prior to the present study, our goals are modest. We first describe the relationship between stress and pitch in Uyghur: we find evidence supporting past descriptions of Uyghur as a stress language with only edge-marking intonation. We then propose a model describing both the nominal and verbal domains in (elicited) declarative sentences. We additionally discuss the properties of interrogatives and of focus constructions. We intend for this to be the first step toward a full model of the Uyghur language.

1.1. Uyghur background

Uyghur (ISO 639-3: uig) is a southeastern Turkic language with roughly ten million speakers in the Xinjiang Uyghur Autonomous Region in the People's Republic of China, and neighboring regions such as Kazakhstan and Kyrgyzstan. It is a synthetic, agglutinating language with SOV word order and a rich case marking and agreement system. It is typologically most similar to modern Uzbek [4].

1.2. Past work on Uyghur

There has been little work to date on intonation in Uyghur, although there has been some on the prosodic systems of related languages like Turkish and Chuvash.

The status of lexical stress in Turkish has been heavily debated in the literature. Turkish has been traditionally analyzed as a stress language (e.g. [5], [6], [7], [8], [9], [10], [11], [12]), while others have argued that Turkish is a lexical pitch accent language (e.g. [13], [14], [15]). It has been noted that the nuclear pitch accent in Turkish is realized in a more compressed pitch range than the pre-nuclear pitch accent [14][16]. More recently, [11] shows that the nuclear pitch accented word is marked on its left edge by an H tone in addition to the pitch range compression, while there is an additional H target associated with the right edge of NPs and PPs.

The study of the Turkic language Chuvash in [17] also suggests that it is a stress language, with duration and intensity serving as important cues. No correlative measures of pitch were done, however.

Previous work has claimed that Uyghur is also a stress language [18]. The author examined the acoustic correlates of stress in Uyghur and found that in both single word utterances and continuous speech, only duration served as a significant predictor of stress location. Neither intensity nor f0 were found to be reliable cues for stress, suggesting that Uyghur is a stress language, but one that uses a more limited set of acoustic features to mark stress than other stress languages such as English. [19], on the other hand, argues that Uyghur is a predominantly footless language that features intonational prominence on the right edge of prosodic words (with marked exceptions), but does not address the duration data presented in [18].

This indicates that although Uyghur and Turkish both have stress,[1] their intonational systems do not treat it in the same way. Uyghur stress appears to be realized in duration but not in pitch, which means that the stressed syllable cannot be identified from the pitch contour of an utterance, while in Turkish intonational tones do associate with stressed syllables. In the AM theory of intonation, intonational tones mark word heads (i.e. stressed syllables) and the edges of prosodic units. That is, if a language has stress, the stressed syllable is expected to be marked by the intonation. Therefore Uyghur would be somewhat unusual from the perspective of the prosodic typology outlined in [21], which does not identify any languages that have stress word prosody but only edge marking intonation.

Such languages are not completely unknown in the literature, however: [22] suggests that the language Kuot, a non-Austronesian language of Papua New Guinea, displays similar properties, with strong effects of duration for word stress and f0 for intonational marking, but no interaction between the stress and intonational systems. Similarly, [23] report that Chimwiini intonation is independent of vowel length, which correlates with stress. The description of the Turkic language Chuvash in [17] also has some intriguing suggestions of mismatches between stress and intonation, but there is not sufficient data presented to determine whether pitch accents are present or not.

[1] Stress in Turkic languages historically targeted the root, but eventually got shifted to the final syllable with certain exceptions. This tendency for final stress is robust across Turkic, but each language has developed a unique system [20].

If it is indeed the case that Uyghur is a stress language with only edge marking intonation, formalizing this into an AM model will be a useful step towards expanding our typological inventory of intonational systems.

2. Data collection

Our data were collected from four adult speakers of Uyghur, two male and two female. All four speakers are from Xinjiang: three from the greater Urumqi area, and one from Qashqar. All speakers were educated in Uyghur and raised speaking primarily in Uyghur. They are all currently pursuing post-secondary degrees in the United States.

Sentences were elicited by having the consultants read from a randomized list prepared by the authors. The consultants checked the sentences for acceptability and made any necessary corrections before recording them. Each sentence was preceded by a question to supply a context. In most cases this was simply *néme boldi?* (what happened?), which suggests a neutral declarative reading. In other cases, wh- or contrast questions were used to elicit focus on a particular constituent. Examples throughout the paper are in the neutral context unless otherwise stated. The recordings used in this paper were made in sound booths in the UCLA and University of Kansas departments of linguistics using the software Audacity.

3. Stress and intonation in Uyghur

[18] shows that stress in Uyghur is reflected only by vowel length, not pitch or intensity. This suggests that although Uyghur can be described as a stress language, stress and intonation are independent. In this section, we use acoustic measurements to show that vowel duration, but not f0, is the only acoustic correlate of stress, while f0 is used to mark the boundaries of prosodic constituents (cf. the discussion of "Intermediate Phrases" in Section 4). This is largely a replication of [18], but we introduce one important change: we examine the same words in both word-initial and word-medial positions, while the studies in [18] only looked at words in isolation and in word-medial contexts. We will show later that there is a tendency for different prosodic constituents to be used for words in these positions, and that this difference is reflected in the acoustic measurements, although with some variability across speakers.

3.1. Stimuli

We tested the independence of stress and intonation in Uyghur by eliciting a series of minimal and near-minimal stress pairs from [18] in both sentence-initial and sentence-medial position in the following carrier phrases:

- X bek yaxshi söz – "X is a good word"
- Mahinur X deydu – "Mahinur will say X"

These carrier phrases were chosen to accommodate the various parts of speech of the target words. Although these carrier phrases set off the target words somewhat from the rest of the sentence, this applies equally to all target words, and hence comparison between them is justified. Our target words are shown in Table 1. Stress is indicated by capitalization. This design allowed a balanced number of stressed and unstressed vowels in similar contexts and with mostly the same vowel quality. This resulted in 16 tokens per speaker, for a total of 64.

Figs. 1 and 2 contrast *Acha* "elder sister" and *aCHA* "branching" in sentence-initial position. Compare the relative duration of the two vowels in each word: the first and second vowels in Fig. 1 are 131 ms and 78 ms respectively, while in figure 2 they are are 80 ms and 118 ms.

Table 1: *Near-minimal and minimal target words.*

Word 1	Gloss 1	Word 2	Gloss 2
DAka	gauze	daLA	plain
BAza	base	baHA	price
DAcha	villa	daDA	father
DOra	medicine	doQA	forehead
CHAsa	square	chaTAQ	problem
Acha	elder sister	aCHA	branching
BAla	child	baLA	disaster
Ara	fork	aRA	between

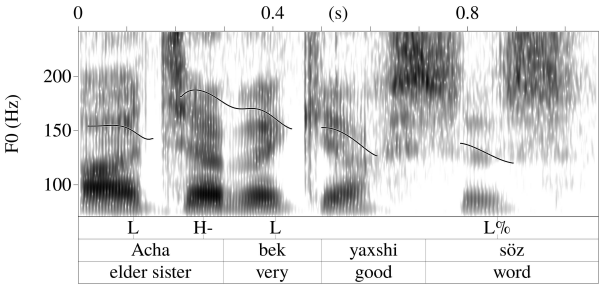

Figure 1: *Pitch track of word-initial stress in sentence-initial position.*

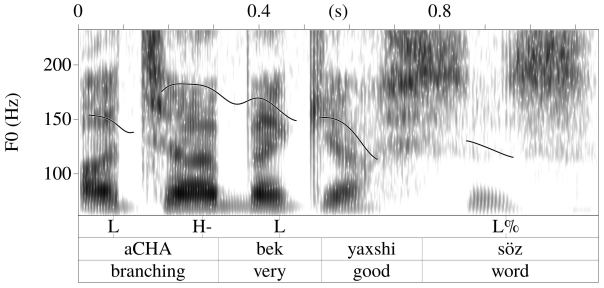

Figure 2: *Pitch track of word-final stress in sentence-initial position.*

3.2. Analysis

The two vowels in each word were segmented using Praat [24], and the average intensity, average f0, and duration were extracted. We ran three linear mixed effects models using the *nlme* package in R [25][26], with duration, intensity, and pitch respectively as the dependent variables. Our independent variables were stress (stressed or unstressed), position in the word (first or second syllable), and position of the word in the sentence (initial or medial). Our random variables were word, vowel, and subject.

3.3. Results

The duration model showed a significant main effect of stress ($\beta = -0.041, t = -3.064, p < 0.01$), with unstressed vowels being significantly shorter. It also showed a marginally significant interaction between position in the word and position of the word in the sentence ($\beta = -0.018, t = 1.76, p = 0.08$): in other words, the final syllable in sentence-medial words tends

to be shorter than the final syllable in sentence-initial words. One subject in particular did not display this effect, and when this subject is removed from the analysis, this interaction becomes significant ($\beta = -0.026, t = -2.34, p < 0.05$). We tentatively take this to indicate a tendency present in most of our speakers, but idiosyncratically absent in one. More will be said about this in Section 4, where we propose a hierarchy of prosodic constituents for Uyghur.

No significant effects were found in the intensity model, though there was a marginally significant effect of position in the sentence ($\beta = -0.92, t = -1.86, p = 0.07$), with intensity lower in word-medial position. The pitch model showed a significant main effect of position in the word ($\beta = 40.631, t = 2.614, p < 0.01$), with second syllables having a higher pitch, and a significant main effect of position of the word in the sentence ($\beta = -11.525, t = -2.342, p < 0.05$), with vowels in sentence-medial words having a lower pitch.

These results show that stress location is a significant predictor of duration, but not pitch. Pitch, rather, is predicted by the position in the word (word-final syllable > word-initial syllable) and the position of the word in a sentence (sentence-initial > sentence-medial), reflecting the edge-marking function of pitch. These results support treating Uyghur as a stress language with only edge-marking intonation, which serves as the basis for the model described below.

4. The Intonational Phonology of Uyghur

Based on evidence from the distribution of intonational tonal targets, as well as phonological and syntactic properties, we argue that the Uyghur intonational system is characterized by three distinct levels of prosodic constituency: the accentual phrase (AP), the intermediate phrase (ip), and the intonational phrase (IP). A schematized representation of these constiuents is shown in Fig. 3.

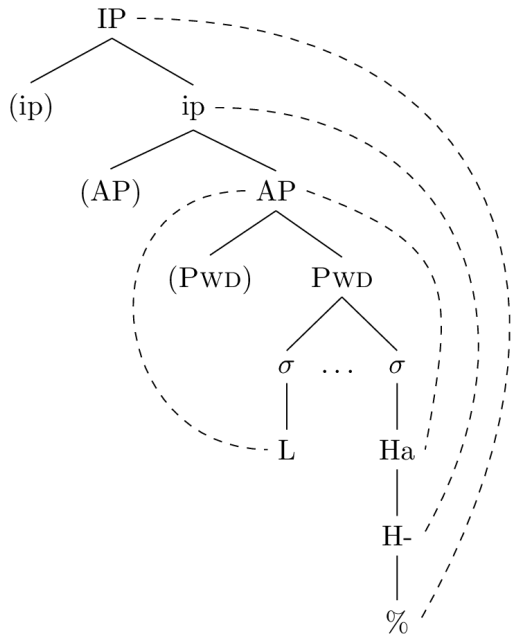

Figure 3: *A schematic representation of the proposed prosodic hierarchy for Uyghur. Prosodic tones associated with higher prosodic constituents overwrite tones associated with lower ones (i.e. % ≫ H- ≫ Ha)*

.

4.1. Accentual Phrase

Section 3 showed that f0 is not correlated with stress, but rather occurred reliably on the final syllable of the constituents we examined. We capture in part with the smallest prosodic constituent that we propose: the Accentual Phrase (AP). The AP is characterized by an L-tone on the left edge and a high tonal target on the right, which we notate as Ha. In many cases an AP consists of a single prosodic word, but multiple words can form a single AP in complex NPs, PPs, relative clauses, and verb phrases. An example is provided for a complex NP containing an adjective in Fig. 4. In this case, both an adjective and a noun form a single AP. The L is associated with the left edge of *bezi*, while the Ha boundary tone is realized on the final syllable of *bughdayni*, i.e. the right edge of the entire complex NP.

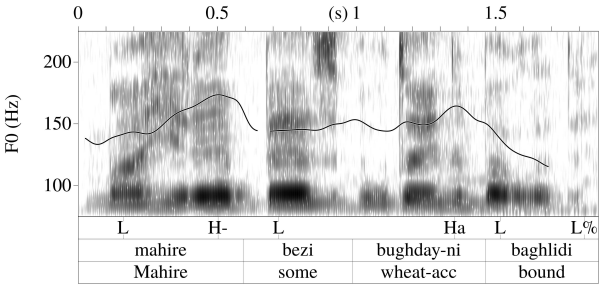

Figure 4: *Mahire bound some wheat.*

Figs. 1 and 2 also show multi-word APs, with all words following the subject falling into a single AP that begins with a L tone (causing the sudden drop from the peak at the end of the subject) and declines further to L% tone, which overwrites the final Ha tone of the AP (cf. Section 4.3).

The remaining evidence for APs comes through comparison with ips, which is presented in the following section.

4.2. Intermediate Phrase

Like the AP, the end of an intermediate phrase (ip) is marked by a high tone, which we label H-. The ip differs from the AP in two ways: the f0 at ip boundaries is higher than AP boundaries, and certain phonological processes occur across AP boundaries, but not ips. For example, hiatus resolution occurs between vowels across an AP boundary but not across an ip boundary, and syllables on an ip boundary are longer than those on an AP boundary.

The subject and object in neutral utterances such as Fig. 5 both have f0 peaks on their final syllables, but the subject-final peak is higher. We account for this difference by suggesting that the subject in this case forms an ip ending in a H- tone, while the object is an AP ending in a Ha.[2] Evidence that this difference in f0 is not simply due to declination over the utterance comes from sentences where the focus falls on an element that is not the subject. In these cases the focused element either forms its own separate ip, or shares an ip with the subject. The latter is the case in Fig. 6, where the final syllable of the subject *Adil* shows a modest peak corresponding to an Ha boundary tone, but the highest f0 is on the focus particle *mu* attached to the object *bughday-ni*, which has an H- boundary tone.

That subjects and objects tend to phrase as ips and APs respectively is also supported by the tendency for the final syllable of words in sentence-initial position to be longer than in

[2]Native speakers tend to judge the juncture between subjects and objects as larger than between objects and verbs in neutral utterances.

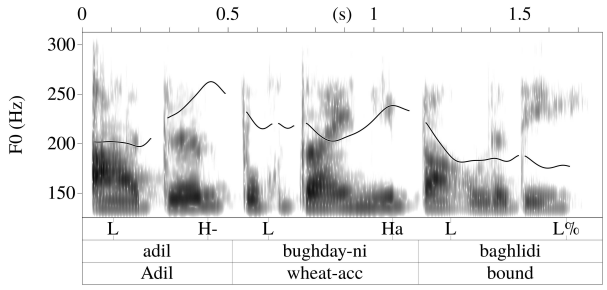

Figure 5: *Adil bound the wheat.*

sentence-medial position (cf. Section 3.3). Research has shown that larger prosodic boundaries are associated with a longer duration on the last syllable of the prosodic unit (e.g. [27], [28], [29]), and the systematic difference in length between the two positions here supports different boundary strengths. This is reflected in Figs. 5, 6, and 7, where the syllables associated with H- tones are longer in each case than those associated with Ha tones (Fig. 5: H-: 311ms, Ha: 182ms; Fig. 6: Ha: 211ms, H-: 230ms; Fig. 7: H-: 179ms, Ha: 118ms).

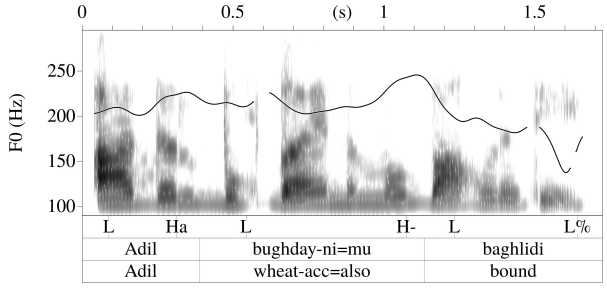

Figure 6: *Adil even bound the wheat.*

Further evidence for a distinction between APs and ips comes from hiatus resolution. More specifically, vowel hiatus does not get resolved by deletion across ip boundaries, but does across AP boundaries. Notice in Fig. 7 that the final [i] in *Adilni* is entirely deleted, while the final [i] in *Ali* and initial vowel in *Adilni* remain unaffected.

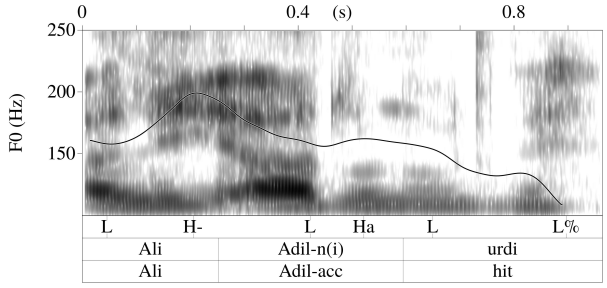

Figure 7: *Ali hit Adil.*

4.3. Intonational Phrase

An IP is marked by a boundary tone on the final syllable, which is also substantially lengthened and followed by an optional pause. We have so far identified two boundary tones: L% and H%. L% marks the end of declaratives and H% marks the end of yes-no questions and sentence-medial continuation rises. Each of the figures provided above shows declarative sentences, which all exhibit the same low target on the final syllable of the sentence. This alternates with the yes-no question in Fig. 8, which exhibits rising intonation on the final syllable.

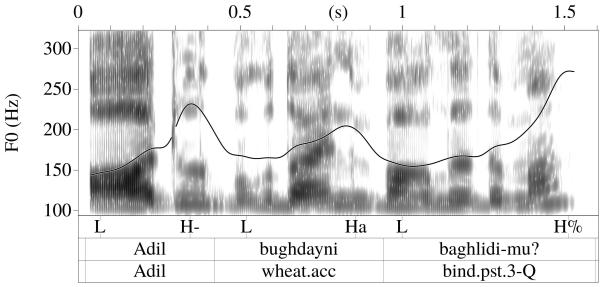

Figure 8: *Did Adil bind the wheat?*

Another environment that licenses the IP-final H% involves clausal coordination/subordination, as in Fig. 9. The two highest peaks in the utterance are the final syllable of the verb of the first clause (*etti*) and the final syllable of the subordinating conjunction (*lekin*). The second clause has the expected L% associated with the right edge of the declarative sentence.

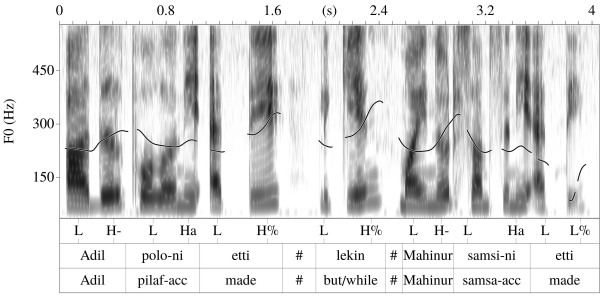

Figure 9: *Adil made polu, but Mahinur made samsa.*

5. Conclusions

In this paper, we have presented a preliminary model of the intonational phonology of Uyghur. We have shown that duration is the most reliable indicator of prominence in Uyghur words, in line with [18]. Furthermore, pitch does not associate with the prominent syllable, i.e. the head, but rather with the edges of phrases. As such, Uyghur appears to be a language with prosodic heads that are ignored by the intonational system. We then motivated the following prosodic constituents in Uyghur: AP, ip, and IP. APs begin with a low tone and end with a high tone. ips also end with a high tone, but this tone is much higher relative to APs. ips can contain multiple APs. Finally, IPs are the largest prosodic constituent in our model, consisting of one or more ips, and indicate sentence type (e.g. declarative vs. interrogative, discourse meaning). This information is marked by a final L% or H% tone at the end of the utterance.

6. Acknowledgements and notes

We would like to thank our consultants Ziba Ablet, Mustafa Aksu, Abduquyum Mamat, and Gülnar Eziz for sharing their language and culture with us. Without their generosity and time, none of this would be possible. We would also like to thank Sun-Ah Jun and the attendees of the UCLA Phonetics Seminar for their invaluable feedback. The authors of this paper are listed in alphabetical order.

7. References

[1] J. Pierrehumbert, "The phonology and phonetics of English intonation," Ph.D. dissertation, MIT, 1980.

[2] M. Beckman and J. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook*, vol. 3, pp. 255–309, 1986.

[3] D. R. Ladd, *Intonational Phonology*, 1st ed. Cambridge: Cambridge University Press, 1996/2008.

[4] T. Engesæth, M. Yakup, and A. Dwyer, *Teklimakandin Salam: hazirqi zaman Uyghur tili qollanmisi / Greetings from the Teklimakan: a handbook of Modern Uyghur*. Lawrence: University of Kansas Scholarworks, 2009/2010.

[5] R. Lees, "The phonology of Modern Standard Turkish," in *Uralic and Altaic Series*. Indiana University Publications, 1961, vol. 6.

[6] E. Kaisse, "Some theoretical consequences of stress rules in Turkish," *Proceedings of the 21st meeting of the Chicago Linguistic Society*, vol. 1, pp. 199–209, 1985.

[7] C. Barker, "Extrametricality, the cycle, and Turkish word stress," *Phonology at Santa Cruz*, vol. 1, pp. 1–34, 1989.

[8] S. Inkelas, "Exceptional stress-attracting suffixes in Turkish: representations versus the grammar," *The prosody-morphology interface*, vol. 134, p. 187, 1999.

[9] S. Inkelas and C. O. Orgun, "Level (non) ordering in recursive morphology: evidence from Turkish," in *Morphology and its relation to phonology and syntax*, P. Farrell, Ed., 1998, pp. 360–392.

[10] B. Kabak and I. Vogel, "The phonological word and stress assignment in Turkish," *Phonology*, vol. 18, no. 3, pp. 315–360, 2001.

[11] C. Ipek and S.-A. Jun, "Towards a model of intonational phonology of Turkish: Neutral intonation," in *Proceedings of Meetings on Acoustics ICA2013*, vol. 19, no. 1. ASA, 2013, p. 060230.

[12] C. Ipek, "The phonology and phonetics of Turkish intonation," Ph.D. dissertation, University of Southern California, 2015.

[13] S. V. Levi, "Acoustic correlates of lexical accent in Turkish," *Journal of the International Phonetic Association*, vol. 35, no. 1, pp. 73–97, 2005.

[14] B. Kamali, "Topics at the PF interface of Turkish," Ph.D. dissertation, Harvard University, 2011.

[15] G. Günes, *Deriving prosodic structures*. LOT, Netherlands Graduate School, 2015.

[16] S. Kan, "Prosodic domains and the syntax-prosody mapping in Turkish," Master's thesis, Boğaziçi University, 2009.

[17] M. Dobrovolsky, "The phonetics of Chuvash stress: Implications for phonology," *In Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 1, pp. 539–542, 1999.

[18] M. Yakup, "Acoustic correlates of lexical stess in native speakers of Uyghur and L2 learners," Ph.D. dissertation, University of Kansas, 2013.

[19] Ö. Özçelik, "Stress or intonational prominence? Word accent in Kazakh and Uyghur," *Turkic Languages*, vol. 19, pp. 163–192, 2015.

[20] K. Menges, *The Turkic Languages and Peoples*, 2nd ed. Harassowitz Verlag, 1995.

[21] S.-A. Jun, "Prosodic typology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 430–458.

[22] E. Lindström and B. Remijsen, "Aspects of the prosody of Kuot, a language where intonation ignores stress," *Linguistics*, vol. 43(4), pp. 839–870, 2005.

[23] C. W. Kisseberth and M. I. Abasheikh, "Chimwiini phonological phrasing revisited," *Lingua*, vol. 121, pp. 1987–2013, 2011.

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer Program]. Version 6.0.37," http://www.praat.org/, accessed: 2018-02-17.

[25] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: https://www.R-project.org/

[26] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team, *nlme: Linear and Nonlinear Mixed Effects Models*, 2017, r package version 3.1-131. [Online]. Available: https://CRAN.R-project.org/package=nlme

[27] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1707–1717, 1992.

[28] D. H. Klatt, "Vowel lengthening is syntactically determined in a connected discourse." *Journal of Phonetics*, vol. 3, no. 3, pp. 129–140, 1975.

[29] I. Lehiste, J. P. Olive, and L. A. Streeter, "Role of duration in disambiguating syntactically ambiguous sentences," *The Journal of the Acoustical Society of America*, vol. 60, no. 5, pp. 1199–1202, 1976.