

A phonological model of Uyghur intonation

Travis Major

University of California, Los Angeles
tjmajor@ucla.edu

Connor Mayer

University of California, Los Angeles
connormayer@ucla.edu

This abstract presents a phonological model of Uyghur (ISO 639-3: uig; Southeastern Turkic) in the autosegmental-metrical framework (AM) (e.g. Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 1996/2008), expanding on Major and Mayer (2018). There has been little research to date on Uyghur intonation, although there has been some on the prosodic systems of related languages like Turkish (e.g., Kamali, 2011; Ipek, 2015).

Uyghur is a stress language that only has edge-marking intonation. It uses a limited set of acoustic features to mark stress compared to other stress languages like English: only duration is a significant predictor of stress location (Yakup, 2013). Stressed syllables in Uyghur cannot be identified by the pitch contour of an utterance. In the AM theory of intonation, tones mark word heads (i.e. stressed syllables) and the edges of prosodic units: if a language has stress, the stressed syllable is expected to be marked by the intonation. Uyghur is thus somewhat unusual from the perspective of the prosodic typology outlined in Jun (2005), 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 (e.g., Lindström & Remijsen, 2005; Kisseberth & Abasheikh, 2011), but none have been modelled in the AM framework, making the model presented here important for the typology of intonation.

Our data were collected from four adult native speakers of Uyghur from Xinjiang. We provided consultants with a randomized list that they first checked for errors and corrected when necessary. They were instructed to read through the list. Each sentence was preceded by a question to supply a context. In most cases the question was *néme boldi?* (what happened?), which elicits a neutral declarative reading. In other cases, we provided a question to elicit focus on a particular item in the sentence.

We first present results of acoustic measurements that confirm that vowel duration is the only acoustic correlate of stress, and show that f_0 is solely used to mark prosodic boundaries. We tested the independence of stress and intonation in Uyghur by eliciting a series of minimal and near-minimal two-syllable stress pairs from (Yakup, 2013) in both sentence-initial and sentence-medial position in a pair of carrier phrases. This design allowed a balanced number of stressed and unstressed vowels in similar contexts with mostly the same vowel quality. This resulted in 16 tokens per speaker, for a total of 64. Fig. 2 contrasts *Acha* “elder sister” and *aCHA* “branching” in sentence-initial position (capital letters indicate stress). Note that vowel duration varies with stress location, but the pitch contour is the same in both utterances.

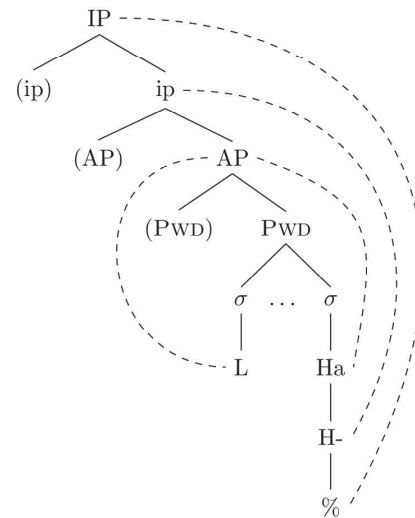


Figure 1: The prosodic hierarchy in Uyghur. Higher prosodic constituents overwrite lower ones (i.e. $\% \gg H- \gg Ha$)

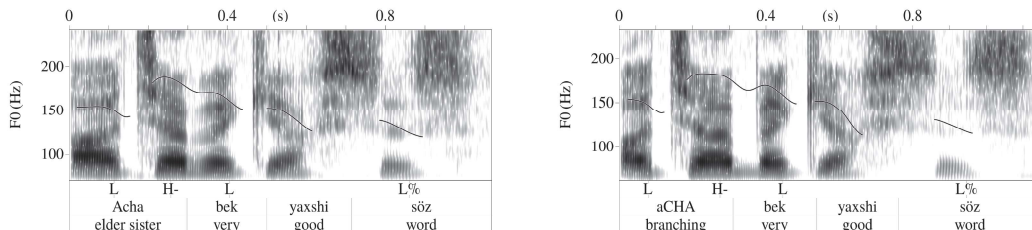


Figure 2: Pitch tracks of word-initial (left) and word-final (right) stress in sentence-initial position.

Linear-mixed effects models showed a significant main effect of stress for duration ($p < 0.01$), with unstressed vowels being significantly shorter, but no effect for pitch. There was also a marginally significant tendency for the final syllable in sentence-medial words to be shorter than the final syllable in sentence-initial words ($p = 0.08$). There was a significant main effect of position in the word on pitch ($p < 0.01$), with second syllables having a higher

pitch, and a significant main effect of position of the word in the sentence ($p < 0.05$), 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.

Based on evidence from the distribution of intonational tonal targets, as well as phonological and syntactic properties, we characterize the Uyghur intonational system by three distinct levels of prosodic constituency: the accentual phrase (AP), the intermediate phrase (ip), and the intonational phrase (IP), which are shown in Fig. 1.

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.

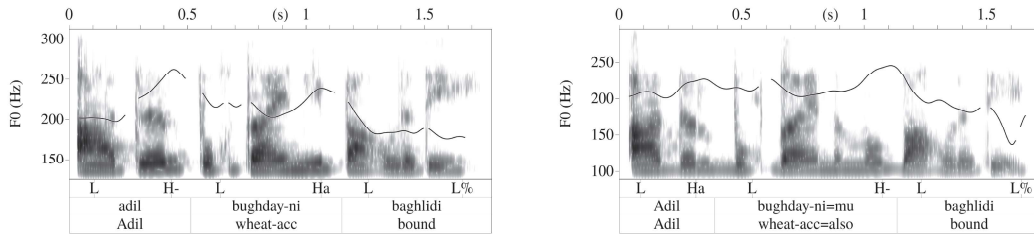


Figure 3: Adil bound the wheat (left); Adil even bound the wheat (right).

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 the left side of Fig. 3 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. 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. An example of the latter is given on the right side of in Fig. 3, 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. The tendency for the final syllable of words in sentence-initial position to be longer than in sentence-medial position described also suggests different boundary strengths (e.g., Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992).

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 yes-no questions which exhibit rising intonation on the final syllable. An example is shown on the left side of Fig. 4.

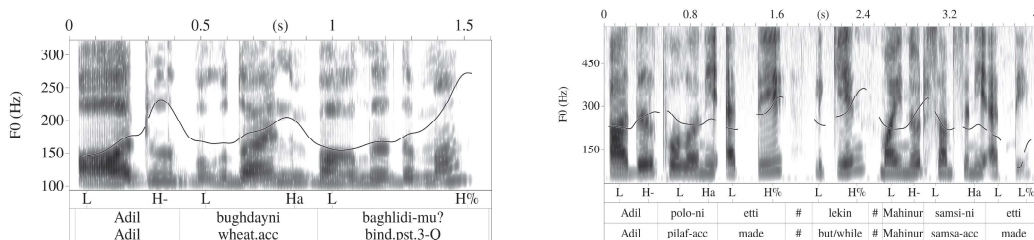


Figure 4: Did Adil bind the wheat? (left); Adil made polu, but Mahinur made samsa (right).

Another environment that licenses the IP-final H% involves clausal coordination/subordination, as in the right side of Fig. 4. 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.

We are in the process of analyzing the data reported above for 10 more speakers recorded in Kazakhstan. In addition to replicating the data reported above, we are currently expanding the empirical coverage to include a variety of different focus and question types.