Acoustic correlates of stress in Udmurt: inter-speaker variation and implications for stress processing

In this paper, we investigate the prosodic realization of stress in Udmurt (Uralic, Permic), in the contexts of minimal pairs consisting of (i) indicative verbs (PRS.3SG; final stress) and (ii) imperative verbs (IMP.2SG/PL; initial stress). Averaged results show that vowel duration and alignment of the stressed vowel with \( f_0 \) targets (high \( f_0 \) for imperatives, high/low \( f_0 \) for indicatives) are used to cue stress. At the same time, individual speakers may preferentially rely on one of these acoustic parameters to cue stress, disregarding the other. This serves as evidence that widely different acoustic cues may be used in marking a single phonological category, stress (cf. also Llistterri et al. 2003) which, in turn, has interesting implications for the phonetics-phonology interface and neural processing of stress.

**Background.** There is no consensus on the stress properties of Proto-Uralic (Szinnyei 1922; Itkonen 1955; Collinder 1960; Steinitz 1964). Contemporary Uralic languages include those with fixed stress (initial, penultimate, or final), morphologically/phonologically-driven stress, or no lexical stress (Lytkin 1964; 1970). Udmurt has fixed final stress (Yemelyanov 1927; Lytkin & Tepliashina 1962), which is manifested in indicative verbs. In contrast, some word classes, e.g., imperative verbs, are stressed on the initial syllable.

According to the only detailed instrumental investigation of Udmurt stress (Denisov 1980), di- and trisyllables with final stress are marked by greater duration of the final vowel and lower \( f_0 \) values, as compared to the vowel in the penultimate syllable. Notably, the test words in the Denisov (1980) study were uttered in isolation, which alone may explain the results. In minimal pairs formed by indicative and imperative verbs, the stressed vowels (either initial or final) had greater duration than their unstressed counterparts within the minimal pair; the \( f_0 \) results were not consistent. No statistical analysis was offered.

**Materials and methods.** The current experiment also targeted string-identical minimal pairs formed by indicative and imperative verbs (di- and trisyllables; total n=172), aiming to replicate and build upon Denisov (1980). The test words were controlled for syllable shape (CV), vowel height ([+low] vs. [−low]), and information structure (backgrounded vs. focused). All items were selected from Kirillova’s (2008) dictionary and checked by an Udmurt speaker who did not participate in the experiment. All items were embedded in carrier phrases: (a) focused: \( I \) [Foc ___] word said, but ___ word didn’t; (b) backgrounded: \( I \) ___ word [Foc quietly/slowly] said, but loudly/quickly didn’t. Six native speakers of Udmurt (5F, 1M, age range 20-40) participated in the experiment. The recordings were made in a quiet room with a head-worn microphone. The sound files were manually annotated in Praat (Boersma & Weenink 2019); vowel duration and \( f_0 \) measurements at 10 points per vowel were collected (Xu 2013). Statistical analysis was carried out using the \texttt{lmer} function in R (R Core Team 2017).

**Results.** According to the averaged results, the stressed vowel in both verb types has greater duration than the other potential stress target. A between-word comparison showed that stressed initial vowels in imperatives are significantly longer than their unstressed counterparts in indicatives; similarly, most stressed final vowels in indicatives are longer than their unstressed counterparts in imperatives. The averaged values are boldfaced in Table 1. Only values for [−low] vowels are shown, for reasons of space; the results for [+low] are parallel.

With respect to \( f_0 \), averaged results show that the stressed initial syllable in imperatives is associated with high \( f_0 \) values and a rising-falling \( f_0 \) contour, while final stressed syllable in indicatives may be associated with a rise or a fall in \( f_0 \), depending on context and/or speaker (contrastive interpretation favors a rise in \( f_0 \). This is shown for [−low] disyllables in Fig. 1.

At the same time, individual speakers differed with respect to the cues that they used to mark stress: three speakers relied mainly on duration, two on \( f_0 \), and one on both cues. For example, speaker 5 rather consistently used vowel duration but not \( f_0 \); this is shown with the values in italics in Table 1 (duration) and in Fig. 2 (\( f_0 \)). In contrast, speaker 6 utilized changes in \( f_0 \) to
differentiate verb types (as well as focus types), but not vowel duration, as illustrated with the values highlighted grey in Table 1 (duration) and in Fig. 3 ($f_0$). To the best of our knowledge, the differences between speakers are not attributable to sociolinguistic, dialectal, age- or gender-related differences.

<table>
<thead>
<tr>
<th>σ count, σ position</th>
<th>Data</th>
<th>Focused</th>
<th>Backgrounded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
</tr>
<tr>
<td>Disyllabic, initial</td>
<td>mean</td>
<td>68.09 (16.85)</td>
<td>116.78 (27.94)***</td>
</tr>
<tr>
<td>sp 5</td>
<td>86.51 (15.53)</td>
<td>128.49 (27.55)***</td>
<td></td>
</tr>
<tr>
<td>sp 6</td>
<td>72.28 (22.98)</td>
<td>81.31 (14.59)</td>
<td></td>
</tr>
<tr>
<td>Disyllabic, final</td>
<td>mean</td>
<td>86.1 (16.93)</td>
<td>74.894 (15.7)***</td>
</tr>
<tr>
<td>sp 5</td>
<td>132.06 (37.9)</td>
<td>107.79 (37.59)</td>
<td></td>
</tr>
<tr>
<td>sp 6</td>
<td>81.9 (20.3)</td>
<td>77.21 (17.47)</td>
<td></td>
</tr>
<tr>
<td>Trisyllabic, initial</td>
<td>mean</td>
<td>57.98 (14.76)</td>
<td>94.98 (17.39)***</td>
</tr>
<tr>
<td>sp 5</td>
<td>84.67 (20.44)</td>
<td>109.2 (28.17)***</td>
<td></td>
</tr>
<tr>
<td>sp 6</td>
<td>89.05 (19.79)</td>
<td>85.04 (27.17)</td>
<td></td>
</tr>
<tr>
<td>Trisyllabic, final</td>
<td>mean</td>
<td>70.82 (11.48)</td>
<td>72.05 (10.89)*</td>
</tr>
<tr>
<td>sp 5</td>
<td>123.9 (49.58)</td>
<td>90.84 (31.39)</td>
<td></td>
</tr>
<tr>
<td>sp 6</td>
<td>74.27 (21.24)</td>
<td>76.86 (16.54)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Durations of vowels in initial and final syllables (ms) in verbs with [−low] vowels; values in brackets = SD. Asterisks mark the values in the imperatives that are significantly different from their counterparts in the indicatives, based on a linear mixed-effects model.

**Significance.** The inter-speaker variation with respect to acoustic cues used to mark stress in Udmurt raises non-trivial questions about neural processing of stress and the nature of phonetic-phonology interface. First, it aligns with the neurolinguistic evidence suggesting that speakers expect varying individual acoustic cues to be utilized in marking stress in a single language (Honbolygő & Csepe 2011). Second, it provides support to the view that phonetic evidence may not provide straightforward one-dimensional physical corroboration for phonological concepts like stress (Keating 1996).

**Future work and broader picture.** We are currently analyzing intensity values and spectral properties of vowels in the same dataset, as well as a dataset of non-minimal-pair test words.

The results above align with our earlier results that showed that initial stress in negated verbs in Udmurt (another class of exceptions to stress-finality) is also cued by vowel duration.

![Figure 1](image1.png)

**Figure 1.** Average $f_0$ contours on initial and final vowels in disyllabic verbs, all speakers.

![Figure 2](image2.png)

**Figure 2.** Average $f_0$ contours on initial and final vowels, speaker 5.

![Figure 3](image3.png)

**Figure 3.** Average $f_0$ contours on initial and final vowels, speaker 6.
References:


