Acoustic correlates of word stress and focus marking in Udmurt

Lena Borise & Ekaterina Georgieva

Abstract

In this paper, we investigate the prosodic realization of stress and focus in Udmurt (Uralic, Permic). According to the literature, Udmurt has fixed final stress, but also has several sets of morphosyntactic exceptions with initial stress. We report the results of two production studies. The first one targets nominals with final stress, and the second one investigates the stress properties of minimal pairs consisting of (i) indicative verbs (PRS.3SG; final stress) and (ii) imperative verbs (IMP.2SG/PL; initial stress). To control for the information-structural contexts, the test words are presented in contexts that elicit narrow focus either on the test word ("F" condition) or on another constituent ("non-F" condition). The results show that all four acoustic parameters surveyed in the paper – duration, intensity, $f_{0,}$ and F1 – participate in stress marking in Udmurt. The results for focus marking vary by study and demonstrate that all cues except for intensity may be involved in focus marking. At the same time, we find wide interspeaker variation with respect to the acoustic cues discussed to mark stress and/or focus. Finally, we outline a preliminary Autosegmental-Metrical interpretation of our f_0 results; a full account of Udmurt intonation awaits further research.

Keywords: word stress, focus, duration, vowel quality, Udmurt, Uralic

***

1 Introduction

Udmurt, a Uralic language spoken in Russia, is commonly described as having fixed stress: word stress regularly targets the final syllable of a word, in all word classes (Lytkin & Tepliashina 1962; Winkler 2001, 2011).[1] This illustrated in (1).

(1) a. kɨrˈd͡ʑan
song
'a song'
b. veˈra
say.PRS.3SG
's/he says'

At the same time, there are several classes of morphosyntactically conditioned exceptions to the stress-finality in Udmurt. These include, e.g., imperative verbs, which regularly have initial stress (Lytkin & Tepliashina 1962; Csúcs 1990; Winkler 2001; 2011), as shown in (2). Indicative and imperative verbs frequently form minimal pairs that only differ in stress placement, as illustrated by (1b) and (2).

(2) ˈvera
say.IMP.2SG
'say!'

[1] Stress placement in other Uralic languages varies: e.g., in Hungarian, stress is fixed on the initial syllable (Siptár & Törkenczy 2000), and in Hill Mari, on the penult (Saarinen 2022); in Komi-Permyak, stress placement depends on morphological factors (Lytkin 1962); in Meadow Mari, stress targets the last full/non-reduced vowel in the word (Alhoniemi 2010). See Pajusalu (2022) for an overview.

In this paper, we report on two studies aimed at investigating the acoustic correlates of stress in Udmurt: vowel duration, intensity, fundamental frequency (f_0) and vowel quality (the properties of formants F1 and F2). The first study targets the realization of nominals (mainly nouns and adjectives), as illustrated by (1a). The second study compares the stress properties of minimal pairs, as illustrated by (1b) and (2). Following the methodological suggestions in Roettger & Gordon (2017), we also control for the information-structural context and elicit the experimental items both when focused (F) and non-focused (non-F).

This paper is structured in the following way. Section 2 introduces the relevant aspects of Udmurt phonology (2.1) and the stress system (2.2), summarizes the existing work on the phonetics of stress in Udmurt (2.3), and provides a brief overview of the existing work on the acoustic marking of stress and focus in a variety of languages (2.4). Section 3 provided the information on the methods: the stimuli (3.1), experimental procedure and participants (3.2), data processing (3.3), and analysis (3.4). Section 4 reports on the results of two studies and is organized by acoustic measures: duration (4.1), intensity (4.2.), fundamental frequency (f_0) (4.3), and vowel height (F1) (4.4). Section 5 contains the discussion of the results, providing a summary of the main findings (5.1), information about interspeaker variation (5.2) and a preliminary Autosegmental-Metrical interpretation of the f_0 findings (5.3). Section 6 concludes.

2 Previous work

2.1 Relevant Udmurt phonology

Udmurt has a seven-vowel system. The positions of the vowels in the vowel space (based on Vakhrushev & Denisov 1992: 26–27; Winkler 2011: 18) and transliterations are provided in Table 1. Since Udmurt uses the Cyrillic script, the Cyrillic orthographic symbols and their equivalents in the Finno-Ugric transcription (FUT), also called Uralic Phonetic Alphabet (UPA) (see Setälä 1901), which is standardly used for the transliteration of Udmurt in the field of Finno-Ugristics, are provided in brackets.

Table 1. The Udmurt vowels

	front	central	back
high	i (и/й)	ɨ (i̮; ы)	u (у)
mid	e (е/э)	ɘ (e̮; ö)	o (о)
low		a (а)	

2.2 Final and non-final stress in Udmurt

Stress in Udmurt is commonly described as targeting the final syllable of a word. Final stress is not conditioned by morphological structure: with the addition of inflectional suffixes to the stem, stress shifts to the rightmost one (Winkler 2001: 10; Vakhrushev & Denisov 1992: 64). Final stress is also regularly found in borrowings: e.g., /kɲiˈga/ < Russ. /ˈkɲiga/ 'book' – though this may not hold in cases of intensive bilingualism (Winkler 2011: 11) and may depend on the type of borrowing (Vakhrushev & Denisov 1992: 66). Even if a Russian borrowing retains its stress pattern, the inflected forms adopt the Udmurt pattern: the rightmost inflectional affix carries stress (Winkler 2011: 22).

At the same time, there are some morphosyntactically conditioned exceptions to the stress-finality: e.g., imperative verbs regularly have initial stress, as was shown in (2) above. Stress does not create important phonological contrasts, though: it only differentiates the members of minimal pairs formed by 3SG indicative verbs and 2SG or 2PL imperative verbs, depending on conjugation class (Tarakanov 1959: 175).

Udmurt verbs form two conjugation classes: Conjugation I and II, also called ɨ-verbs and a-verbs, respectively, based on the final syllable of the stem (which is visible in the infinitival form: e.g., /budɨ-nɨ/

‘grow-INF’ and /vala-nɨ/ ‘understand-INF’). In the ɨ-verbs, minimal pairs are formed by present-tense 3SG indicatives and 2PL imperatives; in the a-verbs, minimal pairs are formed by present-tense 3SG indicatives and 2SG imperatives. This is shown in (3), with the minimal pairs boldfaced.[2]

(3)	a.	**buˈd-e**	ˈbud(ɨ)	ˈ**bud-e**
		grow-PRS.3SG	grow.IMP.2SG	grow-IMP.2PL
	b.	**vaˈla**	ˈ**vala**	ˈvala-le
		understand.PRS.3SG	understand.IMP.2SG	understand-IMP.2PL

Similarly, negated verbs, which are preceded by a negative auxiliary in Udmurt (see Edygarova 2015), are also stressed on the initial syllable: /əm ˈt͡ɕaʃete/ ‘we didn’t make noise’ (Lytkin & Tepliashina 1962: 47; Winkler 2011: 22). Outside of the realm of verbs, initial stress is found in reduplicated adjectives and monosyllabic onomatopoetic words: /ˈgord-gord/ ‘very red’ (lit. ‘red-red’) (Lytkin & Tepliashina 1962: 47; Winkler 2011: 22).

Additionally, stress placement in certain words and/or word classes is described as varying between the initial and final syllables. To the best of our knowledge, these impressionistic claims have not been investigated instrumentally. It is unclear what conditions the variability – it has been described as dependent on “utterance type” (Lytkin & Tepliashina 1962: 48; Csúcs 1990: 29) or “emotional context of an utterance and/or logical emphasis” (Alatyrev 1983). These cases include, e.g., pronouns formed with /vaɲ-/ ‘all, every-’, /koc-/ ‘every-/any-’, /kud-/ ‘which’, /so-/ ‘that’, /ta-/ ‘this’, /ma-/ ‘what’, /no-/ ‘no-’, /og-/ ‘approximately’ (Lytkin & Tepliashina 1962: 48); a few illustrative examples are provided in (4).

(4) ˈvaɲmɨ ~ vaɲˈmɨ ‘we all’; ˈkockin ~ kocˈkin ‘everybody/anybody’, ˈnokin ~ noˈkin ‘nobody’

Some other instances of reported variable stress placement include:

- certain adverbials (/ˈt͡ɕaʎak/ ~ /t͡ɕaˈʎak/ ‘quickly’; /ˈjalan/ ~ /jaˈlan/ ‘always’) (Lytkin & Tepliashina 1962: 48),
- wh-words (Vakhrushev & Denisov 1992: 66),
- adjectives derived with the suffix /-pɨr/: /ˈgordpɨr/ ~ /gordˈpɨr/ ‘reddish’ (Winkler 2011: 23),
- prohibitive verbs, in which stress may target either the negative particle or the first syllable of the lexical verb: /ˈen vera/ ~ /en ˈvera/ ‘don’t say!’ (Lytkin & Tepliashina 1962: 47; Vakhrushev & Denisov 1992: 66; Winkler 2011: 23).

Since the phonological nature of this reported variability is unclear at present, we do not discuss these cases further. The interested reader is directed to the sources cited for more information.

Individual Udmurt dialects present further classes of exceptions to strict stress-finality. In some varieties spoken in Northern Udmurtia, like the Middle Cheptsa dialect (Karpova 2005) and Beserman Udmurt (Tepliashina 1970), as well as in the Kukmor dialect, which belongs to the Southern Peripheral dialects (Kelmakov 1998: 74–75), verbs with plural agreement markers can be stressed on the penult or the ultima: /tuˈbomʌ/ ~ /tuboˈmʌ/ ‘we will climb’.[3] A similar pattern of non-final stress is found in agreeing

[2] To the best of our knowledge, there is no consensus in the descriptive literature whether indicative forms like /buˈde/ are bimorphemic (stem+Tense/Agr) or monomorphemic. We have chosen to represent these verbs as bimorphemic (and this applies to all mid- and high+mid items in our second study, see Appendix 2); nothing hinges on this choice.

[3] The vowel /ʌ/ (ə̂, ъ) is used only in some of the Northern, Southern and Southern-Peripheral dialects and corresponds to /ɨ/ in the standard language (Lytkin & Tepliashina 1962: 40–41; Kelmakov 1998: 47).

converbs, inflecting postpositions, inflecting pronouns and possessed nouns, which take on the same agreement markers: /turnaˈkudʌ/ ‘when you are/were mowing’, /bərˈśamʌ/ ‘behind/after us’, /vicˈnamʌ/ ‘the five of us’, /bakt͡ɕaˈjamʌ/ ‘in our garden’ (Georgieva 2017). Our elicitation materials did not include any items that could be subject to variable stress placement.

2.3 Previous studies on Udmurt stress

Some early instrumental studies investigating the nature of Udmurt stress are available, but their conclusions are quite limited in scope. Lytkin & Tepliashina (1962) conclude, based on a handful of experimental tokens produced by one native speaker, that stressed (i.e., final) syllables are about 1.5 times longer than unstressed ones – though this is only the case in words uttered in isolation; in running speech the difference between the duration of stressed and unstressed vowels is reported as much smaller (Lytkin & Tepliashina 1962: 22−24, 49). The authors also note that greater intensity and f_0 may be used as a secondary means of marking stress. In contrast, Baitchura (1973) observes, based on data from four native speakers (number of experimental tokens not reported), that initial syllables are marked by greater intensity and f_0, while final ones are 1.5-2 longer than the initial ones. Baitchura (1973) interprets these findings as evidence for initial stress.

Finally, Vakhrushev & Denisov (1992), building on Denisov (1980), use di- and trisyllabic words as well as minimal pairs of 3SG indicative and 2SG/PL imperative verbs; the stimuli were tested with two native speakers. Their results show that the duration of stressed syllable is 1.6 times greater than that of the unstressed syllable in disyllables and 1.7 times greater than that of unstressed syllables in trisyllables. The authors also compare the duration results for stimuli uttered in isolation with those used in connected speech and conclude that the duration of the stressed (final) syllables is greatest in words uttered in isolation and phrase finally. This points to a strong effect of final lengthening, which the authors themselves acknowledge (Vakhrushev & Denisov 1992: 74). In minimal pairs, the stressed syllables, both non-final or final, are shown to have greater duration than their unstressed counterparts: i.e., the stressed initial syllables of imperatives have greater duration than the unstressed initial syllables of indicatives, and the stressed final syllables of indicatives have greater duration than the unstressed final syllables of imperatives. With respect to intensity, Vakhrushev & Denisov (1992: 77) conclude that it does not consistently cue stress, though, generally, stressed syllables in the minimal pairs have greater intensity than their unstressed counterparts. With respect to f_0 contours, Vakhrushev & Denisov (1992: 79) show that, in words uttered in isolation, the mean f_0 of the second syllable is lower than that of the first syllable (which is also attributable to declarative intonation). In minimal pairs, the f_0 results are more variable, but seem to point to a tendency for stressed syllables, both initial and final, to be associated with lower f_0 values. Overall, while the quantitative results in Vakhrushev & Denisov (1992) are not reported in detail, some of the general trends are clear; the study has served as an inspiration for our work.

2.4 Acoustic marking of stress and focus in other languages

The acoustic cues that have been mentioned in the existing studies of Udmurt stress, summarized in the previous section, are some of the cues canonically associated with the expression of stress. A non-exhaustive list of these cues, which are commonly discussed in the literature on the topic, includes duration of the stressed vowel/syllable, intensity (either overall or frequency-sensitive, also known as spectral tilt), formant frequency, as well as higher or lower f_0 values on the stressed vowel/syllable. A detailed overview of the relative importance of these cues in a number of languages, based on a meticulous survey of the existing studies, is provided in Gordon & Roettger (2017). The overview shows that stressed vowels/syllables (or, occasionally, syllable codas/onsets) tend to have greater duration and/or greater intensity as compared to unstressed counterparts. With respect to formant frequency/vowel quality, stressed vowels are often more peripheral/lower in the vowel space than unstressed ones. Finally, higher or lower f_0 values on the stressed vowel/syllable, in languages without lexical tone-based distinctions, are typically due to alignment with intonational f_0 targets (high or low). The realization of focus/emphasis commonly relies on the cues that come from the same set; the cues for stress and focus in a given language may overlap, fully or partially (Vogel, Athanasopoulou & Pincus 2016).

The unstressed counterparts that the properties of the stressed syllables/vowels are compared to may come from the same lexical item (i.e., precede or follow the stressed one in the same word – being in a so-called syntagmatic relationship). For instance, in the English noun /ˈpɝmɪt/ the properties of the stressed first syllable/vowel may be compared to those of the unstressed second one. Alternatively or additionally, in languages that allow for variable stress placement, a stressed syllable/vowel may be compared to an unstressed counterpart in the same position in a different word (a so-called paradigmatic relationship). For example, the realization of the stressed first syllable/vowel in the noun /ˈpɝmɪt/ may be compared that in the unstressed initial syllable of the verb /pɚˈmɪt/.

With this background in mind, our hypotheses are the following. First, we expect stress in Udmurt to be cued by one of more of duration, intensity, vowel quality, and f_0, with the relevant comparisons being syntagmatic or paradigmatic in nature. Second, we expect for focus to be expressed by one or more of the cues from the same set.

3 Methods

3.1 Stimuli

Our investigation consisted of two production studies. The first one targeted Udmurt nominals, and the second one investigated the stress properties of minimal pairs formed by indicative and imperative verbs. The test words were collected from a dictionary (Kirillova 2008) and checked with a native speaker who did not participate in the study. The choice of the test words, as described below, was determined by syllable structure and vowel height properties. Since the test words were selected from a dictionary of standard Udmurt, none of the stimuli were dialectal in a strict sense – i.e., used only in a particular dialect. Standard Udmurt sometimes codifies more than one lexeme/variant (often coming from different dialects) as standard, and speakers of Udmurt are usually familiar with the different lexemes in these cases, especially if, like our participants, they have studied standard Udmurt. Nevertheless, the speakers were instructed to skip any test words if they felt that they could not pronounce them in a natural way.

The materials for the first study comprised 109 Udmurt nouns, adjectives, and postpositions (which correspond to a nominal base inflected with a case suffix). All stimuli consisted of CV syllables and were controlled for syllable count (di- and trisyllabic) and vowel height (low, mid, and high; all vowels in given word are of the same height). Both voiced and voiceless onsets were allowed, in order not to restrict the size of the dataset (for the purposes of f_0 analysis, the first 20ms of the vowel were discarded; Xu (2013)). Because morphological structure is not mentioned in previous works as relevant for the purposes of stress assignment, both mono- and polymorphemic test words were used (excluding any morphology that may influence stress assignment, as described in Section 2.2). The breakdown of the dataset by syllable count and vowel height is provided in Table 2. The smaller number of 'low' stimuli is due to the fact that there is only one low vowel in Udmurt, /a/, as compared to three mid vowels (/e/, /ɘ/, and /o/) and three high vowels (/i/, /ɨ/, and /u/), which limits the number of possible stimuli with low vowels. The full list of stimuli used in the first study is provided in Appendix 1.

Table 2. Dataset used in the first study

	low vowels	**mid vowels**	**high vowels**
disyllabic	14	28	30
trisyllabic	7	12	18

All stimuli were embedded in carrier sentences as direct quotes; stress was not marked in any of the words. To control for phrasal prosodic environment, two sets of carrier phrases were constructed, following the set of recommendations in Roettger & Gordon (2017). In the first set, the test word was under narrow (contrastive) focus (henceforth referred to as “F”). This was ensured by explicitly contrasting the test word with another (following) word, of the same syllable count and structure, which was part of the carrier sentence. This is illustrated in (5) for the test word /baka/ ‘frog’, contrasted with the word /daga/ ‘horseshoe’. Only the first of the two words (boldfaced) was analyzed.

(5)	Mon	“**ba**ˈ**ka**”	kɨˈlez	veˈraj,	a	“daˈga”	kɨˈlez	əj.
	I	frog	word.ACC	say.PST.1SG	but	horseshoe	word.ACC	NEG.PST.1SG

‘I said the word “**frog**”, but not the word “horseshoe”.’

In the second set of carrier sentences, the test word was explicitly out of focus (henceforth referred to as “non-F”). This was ensured by placing contrastive focus on another constituent (an adverb), which was part of the carrier phrase. Two subtypes of this kind of carrier phrase were used, containing different pairs of adverbs, to make the test sentences less repetitive. They are provided in (6).

(6)	a.	Mon	“**ba**ˈ**ka**”	kɨˈlez	ʃɨp	veˈraj,	zol	əj.
		I	frog	word.ACC	quiet(ly)	say.PST.1SG	loud(ly)	NEG.PST.1SG

‘I said the word “**frog**” quietly, not loudly.’

b.	Mon	“**ba**ˈ**ka**”	kɨˈlez	kaˈʎːen	veˈraj,	d͡ʒog	əj.
	I	frog	word.ACC	slow(ly)	say.PST.1SG	quick(ly)	NEG.PST.1SG

‘I said the word “**frog**” slowly, not quickly.’

109 test words by two phrasal conditions (F and non-F) yielded 218 test sentences. The test sentences were presented to participants in a randomized order. Three different randomized orders were created in order to control for effects of newness/familiarity of stimuli. Each participant was assigned to one of the randomizations.

The second study consisted of 43 minimal pairs formed by indicative and imperative verbs – i.e., 86 verb forms in total. Like the nominals in the first study, the verbs in the second study were di- and trisyllabic, consisted of CV syllables and were controlled for vowel height (low, mid, high). For morphological reasons, though, the final syllables, corresponding to PRS.3SG/IMP.2 markers, could only contain mid or low vowels, as was illustrated in (3) in Section 2.2. This means that in the ‘high vowels’ category (i.e., the test words in which all vowels were supposed to be high), only the root vowel(s) were high, and the final syllable contained a mid vowel, /e/. Accordingly, we label this type of stimuli ‘high+mid’; for analytical purposes, the final mid vowels of the verbs in the ‘high+mid’ category were grouped together with the other ‘mid’ vowels. The breakdown of the dataset by syllable count and vowel height is provided in Table 3. The full list of stimuli used in the second study is provided in Appendix 2.

Table 3. Dataset used in the second study; the numbers refer to minimal pairs

	low vowels	**mid vowels**	**high(+mid) vowels**
disyllabic	9	9	9
trisyllabic	5	5	6

Like in the first study, the phrasal prosodic context in which the test words appeared was controlled with the help of carrier sentences. In the focused (F) context, a test verb was explicitly contrasted with another verb of the same type (i.e., indicative or imperative) and same syllabic structure, as shown in (7). In the non-focused (non-F) context, with two subtypes, an explicit contrast was established between other elements on the carrier sentence (adverbs), as illustrated in (8). Similarly to the first study, the test words were used as direct quotes within the carrier sentences and stress was not marked on any of the words. Given that the second study is about minimal pairs, we had to indicate whether the speakers should produce an indicative or an imperative verb; this was done in the following way. If the verb was meant to be used as an imperative, it was accompanied by an exclamation mark within the direct quote; the indicative verbs were left unmarked. The participants were informed that the exclamation marks mark imperative verbs but are not meant to elicit exclamative intonation.

(7) Mon “**vaˈla**”/ “ˈ**vala!**” kɨˈlez veˈraj, a “gaˈʒa”/ “ˈgaʒa!”
I understand.PRS.3SG/IMP.2SG word.ACC say.PST.1SG but respect.PRS.3SG/IMP.2SG
kɨˈlez ɘj.
word.ACC NEG.PST.1SG
‘I said the word “**understands**”/ “**understand!**”, but not the word “respects”/ “respect!”.

(8) a. Mon “**vaˈla**”/ “ˈ**vala!**” kɨˈlez ʃɨp veˈraj, zol
I understand.PRS.3SG/IMP.2SG word.ACC quiet(ly) say.PST.1SG loud(ly)
ɘj.
NEG.PST.1SG
‘I said the word “**understands**”/ “**understand!**” quietly, not loudly.’

b. Mon “**vaˈla**”/ “ˈ**vala!**” kɨˈlez kaˈʎːen veˈraj, d͡ʒog
I understand.PRS.3SG/IMP.2SG word.ACC slow(ly) say.PST.1SG quick(ly)
ɘj.
NEG.PST.1SG
‘I said the word “**understands**”/ “**understand!**” slowly, not quickly.’

The 43 minimal pairs, equaling 86 verbs, multiplied by two phrasal contexts, produced 172 test sentences. Like in the first study, the stimuli were randomized; three different randomizations were used. Each participant was assigned to one randomization.

3.2 Procedure and participants

During the recording sessions for both studies, the test sentences in standard Udmurt orthography were presented to the participants on a computer screen, one sentence at a time. The participants were instructed to familiarize themselves with the sentence and then pronounce it using natural intonation. Each test sentence was uttered once by a participant. If the participant was not happy with the way they pronounced the sentence, they were allowed to re-do it; in such cases, all but the final responses were discarded. Before proceeding to the test sentences in each study, the participants were required to complete a short training phase, consisting of four simple Udmurt sentences of various structure that they were asked to pronounce, in order to get accustomed to the experimental setting.

The studies were conducted in June 2020 in Budapest, Hungary. The recordings were carried out in a quiet room, using a Zoom H4n recorder and a close-range head-worn Shure SM10A microphone. Six native speakers of Udmurt took part in the first study, five of the same six native speakers, except Sp3, also took part in the second study. The speakers received a small remuneration for their participation in the experiment (a gift card). The speakers were all female (Sp1-Sp6); age range: 22−39, mean age: 29.5

years. All were studying/working in Budapest, Hungary, at the time of the recording (the duration of residency in Hungary ranged from 1 month to 8 years and 10 months, mean: 4.625 years). All participants were Udmurt-dominant Udmurt-Russian bilingual speakers.

Four of the speakers were born and raised in central Udmurtia (Sp1, Sp2, Sp5, Sp6), one in northern Udmurtia (Sp4), and one in central-southern Udmurtia as well as in Izhevsk, the capital of Udmurtia (Sp3). All speakers have lived in Izhevsk as adolescents, before relocating to Hungary, and have studied the standard variety of Udmurt in school and/or at the university. Given their background, we assumed that the participants' speech would show both features characteristic of the respective (sub)dialects, as well as those of standard Udmurt; as far as we can tell, this is indeed the case. This is in line with recent sociolinguistic studies: Edygarova (2014) describes the colloquial language spoken among Udmurts from different dialect groups, primarily in an informal urban setting, as a so-called 'cross-local vernacular variety of Udmurt', which is a mix of local dialects with the standard variety and Russian code-switching. Furthermore, Edygarova (2014) argues that standard Udmurt is not a native language for Udmurt speakers, but rather an acquired literary style, primarily mastered through explicit linguistic training. Because of this complex sociolinguistic context, we instructed the speakers to pronounce the test sentences in the way that is most natural for them, as our intention was to study their native varieties as spoken by young urban-based speakers who also frequently use the cross-local Udmurt vernacular.

It is important to note that the (cross)dialectal background of our consultants does not differ from the standard language with respect to the vowel inventory. According to Kelmakov (1998: 47, 60–61), the dialects spoken in Udmurtia (Northern, Central, Southern), as well as the standard language, have a seven-vowel system (as in Table 1) – as opposed to vowel systems that include up to ten vowels, which are characteristic of Udmurt-speaking communities outside the Udmurtia proper. The main point of variation among some of the dialects spoken within Udmurtia proper is the use of /ʌ/ instead of /ɨ/ (see also fn. 3). As far as we can tell, this does not apply to the speech of our participants, which is confirmed by the formant distribution plots in Figure 9 and Figure 11. Accordingly, we expect no pronounced qualitative differences between the varieties spoken by our participants.

We also carefully controlled for the differences related to stress between Udmurt dialects, making sure not to include any test material where stress location may vary (see Section 2). Potential differences in phrasal prosody among Udmurt dialects have not been studied; it is a question for further research to determine whether the interspeaker variation that we notice in our data (Section 5.2) is to be explained as a dialectal or idiolectal in nature. Due to the limited number of speakers in our study, we refrain from making any claims to this effect.

The recording sessions for the first study lasted between 16 and 47 minutes per participant, and between 12 and 25 minutes for the second study; there was a 30-minute break between the two studies. In total, 1,308 test sentences were recorded during the first study (218 test sentences * six participants), and 860 for the second study (172 test sentences * five participants).

3.3 Data processing

The audio files were manually annotated in Praat (2021) by trained research assistants, based on the segmentation criteria in Machač & Skarnitzl (2009), and checked by the authors. Disfluent responses (due to pauses, errors, false starts, throat clearing, etc.) were eliminated: 42 in the first study, and 14 in the second study.

While listening to the recordings, we identified a potential for a prosodic ambiguity in the carrier sentences in which the test word carries narrow focus – i.e., those like (5) and (7). Because negation in the second part of these sentences is expressed with a negative auxiliary, the sentences can be understood either as contrasting the test words in the two parts of the sentence or contrasting the verb in the first part with the negative auxiliary in the second part. That is, in examples like (5) and (7), either the two test words carry narrow (contrastive) focus (I said the word **“frog”**, and not the word **“horseshoe”**.), or the two test words are interpreted as contrastive topics, and the verbs are narrowly (contrastively) focused (As for the word “frog”, I **said** it, but the word “horseshoe”, I **didn’t**).[4] Because there is no way to construct sentences of this type in Udmurt other than with a negative auxiliary in the second part of the carrier sentence, the ambiguity is unavoidable. Accordingly, we eliminated the responses in which the verbs carried the main accent and were contrasted with each other. In total, we eliminated 279 “verb-focus” responses in the first study and 91 “verb-focus” responses in the second study. Because the “verb-focus” confound only applied to the “F” condition, the number of “F” responses ended up being lower than the number of “non-F” ones, in both studies. The “verb-focus” reading was especially favored by some of the participants: speakers Sp3 and Sp6 produced all of their “F” responses with focus on the verb, which lead to the elimination of these responses.

Additionally, a native speaker of Udmurt who did not take part in the study listened to the recordings of the second study and eliminated the responses that were not produced on target (e.g., an indicative verb erroneously produced instead of an imperative one and vice versa). The responses eliminated for this reason totalled 22. The final counts of responses for both studies, broken down by focus type, syllable count, vowel height, and, in the second study, verb type, are provided in Table 4 and Table 5, respectively.

Table 4. Final counts of responses in the first study

Focus type	**Syllable count**	**Vowel height**	**N**
Focused	**disyll**	high	112
		mid	99
		low	52
	trisyll	high	67
		mid	37
		low	27
Non-focused	**disyll**	high	168
		mid	147
		low	77
	trisyll	high	98
		mid	63
		low	40
Total analyzed			**987**

Table 5. Final counts of responses in the second study

Verb type	**Focus type**	**Syllable count**	**Vowel height**	**N**
Indicative	**Focused**	**disyll**	**high**	32
			mid	35
			low	35
		trisyll	**high**	24
			mid	21
			low	19
	Non-focused	**disyll**	**high**	42
			mid	41
			low	45
		trisyll	**high**	28
			mid	25
			low	25
Imperative	**Focused**	**disyll**	**high**	30
			mid	33
			low	35
		trisyll	**high**	23
			mid	18
			low	20

[4] We thank Erika Asztalos and Balázs Surányi for pointing out this confound.

	Non-focused	**disyll**	**high**	41
			mid	43
			low	42
		trisyll	**high**	28
			mid	24
			low	24
			Total analyzed	**733**

3.4 Analysis

3.4.1 Measurements

The ProsodyPro Praat script (Xu 2013) was used to collect the acoustic parameters of the annotated segments (vowel duration, intensity, average f_0 per vowel, f_0 at 10 fixed points per vowel, and F1 and F2 values). The visualization of f_0 data was done with Matplotlib in Python (Hunter 2007).

In order to ensure comparability between the data from two studies, as well as between di- and trisyllabic test words, only the acoustic parameters on the initial and final syllables were analyzed (i.e., middle syllables of trisyllables were discarded). This does not necessarily mean that no stress cues are realized on the middle syllables of disyllables. A preliminary exploration of the middle-syllable data points to some potentially relevant tendencies. In indicatives, the middle (i.e., pre-tonic) syllable may exhibit a degree of stress-related lengthening. In imperatives, there is wide variation in f_0 on the second (i.e., post-tonic) syllable; this is not surprising, given that the f_0 contour may be meaningful on the pre- and/or post-tonic syllables as well as the stressed syllable. As far as we can tell, though, any stress-related effects on the unstressed middle syllable in trisyllables are supplementary to the cues expressed on the stressed syllables themselves. For reasons of space, we leave a dedicated discussion of stress cues realized of syllables other than the stressed ones outside the scope of the current paper.

3.4.2 Statistical analysis

The statistical analysis was carried out in R (R Core Team 2020), using packages lme4 (Bates et al. 2015), lmerTest (Kuznetsova, Brockhoff & Christensen 2017), and emmeans (Lenth 2022). Each of the acoustic measures (duration, intensity, f_0, and F1 and F2 values) were analyzed using linear mixed effects models, using the lmer( )function, with the acoustic measure as the dependent variable. Following the guidelines in Gries (2021), for each acoustic parameter, the most complex model was fit first, with the following fixed effects and their interactions: in the first study, FOCUS TYPE (with levels F and NON-F), VOWEL HEIGHT (with levels HIGH, MID, and LOW), and SYLLABLE NO. (with values INITIAL and FINAL); in the second study, FOCUS TYPE (with levels F and NON-F), VOWEL HEIGHT (with levels HIGH, MID, and LOW), SYLLABLE NO. (with values INITIAL and FINAL), and VERB TYPE (with levels INDICATIVE and IMPERATIVE). The starting models also included random effects of WORD, nested in NO. OF SYLLABLES, and SPEAKER; random slopes for SYLLABLE NO. and FOCUS TYPE were also included. The starting models for the first and second studies are provided in (9a) and (9b), respectively.

(9) a. dependent variable ~ FOCUS TYPE* VOWEL HEIGHT* SYLLABLE NO. +
(1 + FOCUS TYPE + SYLLABLE NO. | SPEAKER) +
(1 + FOCUS TYPE + SYLLABLE NO. | NO. OF SYLLABLES/ WORD)

b. dependent variable ~ FOCUS TYPE* VOWEL HEIGHT* SYLLABLE NO.* VERB TYPE +
(1 + FOCUS TYPE + SYLLABLE NO. | SPEAKER) +
(1 + FOCUS TYPE + SYLLABLE NO. | NO. OF SYLLABLES/ WORD)

The starting models did not converge for any of the acoustic measures. Next, the random effect structure was simplified, with the effects that accounted for the least amount variance dropped first and the resulting models compared via the function anova(). After that, the fixed effect structure was simplified, via the function drop1(). Eventually, for each acoustic measure, the most complex model that converged without numerical problems and with all predictors being significant was selected and evaluated (reported in the individual subsections in Section 4). *P*-values were obtained with the lmerTest package. If the interactions between the fixed effects proved significant for a particular acoustic measure, further pairwise comparisons were carried out using the package emmeans().

4 Results

For the ease of comparison of individual acoustic measures across the two studies, this section is organized by acoustic measures, rather than by studies.

4.1 Duration

4.1.1 First study (nominals)

Table 6 provides the mean vowel duration values in the test words of the first study. As these results show, final (stressed) syllables typically have greater duration than initial syllables, in non-F and especially in F contexts, for all vowel heights and in all syllable counts.

Table 6. Mean vowel duration in the first study

Focus type	Syllable count	Vowel height	N	Initial syllable		Final syllable	
				Duration (ms)	SD (ms)	Duration (ms)	SD (ms)
Focused	**disyll**	**high**	112	72.67	22.66	93.89	53.61
		mid	99	85.84	19.82	107.19	42.31
		low	52	95.07	16.99	116.73	41.70
	trisyll	**high**	67	60.43	19.55	97.65	51.72
		mid	37	76.51	15.81	107.86	44.98
		low	27	75.69	10.71	106.85	39.10
Non-focused	**disyll**	**high**	168	63.44	20.58	78.55	44.37
		mid	147	77.09	18.65	95.87	40.07
		low	77	81.00	16.94	101.98	35.52
	trisyll	**high**	98	52.54	18.46	82.29	47.33
		mid	63	70.96	20.47	94.29	40.01
		low	40	72.11	19.56	96.67	38.87

For the statistical analysis, the duration values were log-transformed (the raw data had a long right tail, corresponding to the outliers in Figure 1). A model that fit the data best included VOWEL HEIGHT and SYLLABLE NO. as fixed effects, SPEAKER and WORD as random effects, and random slopes for SYLLABLE NO. for both SPEAKER and WORD (more complex slopes for random effects led to the non-convergence of the model); the model is summarized in (10). The interaction of fixed effects did not improve the model fit, suggesting that the effect of vowel height does not vary by syllable position. The remaining potential fixed effect, FOCUS TYPE, did not significantly affect the duration values, which means that vowel duration does not vary significantly depending on focus context. Notably, among the random effects, NO. OF SYLLABLES turned out to be redundant in the presence of WORD. A likelihood ratio test showed that the model is highly significant ($\chi^2(3) = 82.26$, p<0.001), with the conditional R^2 of 0.788.

(10) log(Duration) ~ VOWEL HEIGHT + SYLLABLE NO. +
(1 + SYLLABLE NO. | SPEAKER) +
(1 + SYLLABLE NO. | WORD)

The duration data, organized according to the fixed effects that proved to be significant (VOWEL HEIGHT and SYLLABLE NO.), is visualized in Figure 1.

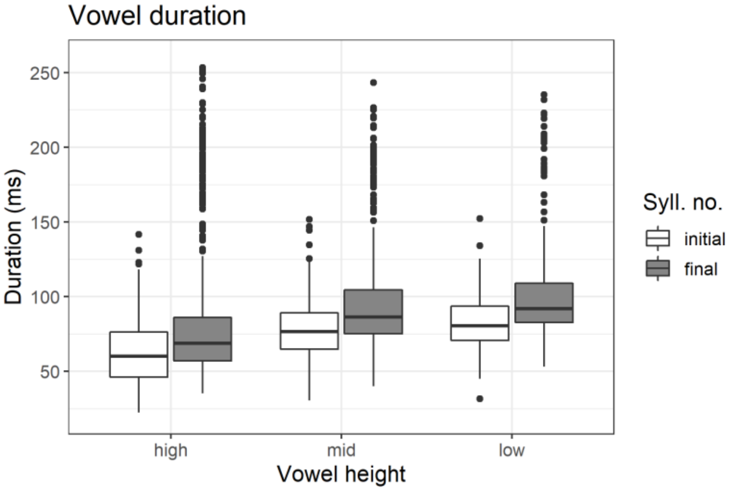


Figure 1. Vowel duration in the first study, broken down by the significant fixed effects (VOWEL HEIGHT and SYLLABLE NO.)

The output of the model is provided in Table 7. The results show that syllable number (which also corresponds to stress in the first study) has a significant effect on vowel duration, and so does vowel height. The lack of the significant effect of interaction between the fixed effects suggests that syllable number has a comparable effect on duration in all vowel heights.

Table 7. Model output for duration in the first study

		Estimate	SE	df	t-value	p-value
[1]	Intercept (final, high)	6.208	0.191	5.132	32.452	<0.001***
[2]	**initial**, high	-0.339	0.140	5.362	-2.227	<0.05*
[3]	final, **low**	0.396	0.043	107.702	9.144	<0.001***
[4]	final, **mid**	0.307	0.036	109.463	8.636	<0.001***

4.1.2 Second study (verbs)

The mean vowel duration values for initial and final syllables in verbs, broken down by verb type, focus type, syllable count, and vowel height are provided in Table 8. The greyed-out cells indicate that there were no final high vowels attested. The mid vowels that were used in the final syllables in the “high-vowel” contexts instead are pooled with the other final mid vowels. As the results show, both initial and final syllables, when stressed, are typically longer than their unstressed counterparts. The duration of vowels in focused verbs is greater than that in their non-focused counterparts – especially in indicatives.

Table 8. Mean vowel duration in the second study

Verb type	Focus type	Syllable count	Vowel height	Initial syllable			Final syllable		
				n	Duration (ms)	SD (ms)	n	Duration (ms)	SD (ms)
Indicative	**Focused**	**disyll**	**high**	32	74.65	14.06			
			mid	35	82.48	21.95	67	123.00	53.78
			low	35	94.50	18.50	35	124.28	53.16
		trisyll	**high**	24	70.86	17.25			
			mid	21	87.91	20.89	45	117.97	54.52
			low	19	82.81	13.14	19	104.62	45.04
	Non-focused	**disyll**	**high**	42	65.24	16.71			
			mid	41	81.28	17.43	83	118.54	52.77
			low	45	91.14	22.25	45	112.93	49.42
		trisyll	**high**	28	60.10	15.48			
			mid	25	80.55	13.21	53	106.91	46.59
			low	25	78.15	15.50	25	107.18	43.02
Imperative	**Focused**	**disyll**	**high**	30	104.76	35.95			
			mid	33	114.96	38.98	63	116.72	45.73
			low	35	126.13	35.45	35	102.04	41.44
		trisyll	**high**	23	87.35	29.38			
			mid	18	111.44	27.63	41	104.22	41.45
			low	20	108.55	36.78	20	95.96	33.15
	Non-focused	**disyll**	**high**	41	99.19	30.16			
			mid	43	121.38	31.06	84	113.48	44.37
			Low	42	128.18	38.92	42	98.21	31.39
		trisyll	**high**	28	82.73	25.49			
			mid	24	104.50	26.18	52	97.76	41.52
			low	24	116.16	26.21	24	87.28	28.78

Like in the first study, the duration values were log-transformed for the statistical analysis. A model that fit the data best was more complex than in the first study: it included all four fixed effects and two interactions: SYLLABLE NO.*VOWEL HEIGHT + VERB TYPE*FOCUS TYPE, suggesting that (i) all fixed effects (or the interactions included) have a significant effect on duration, (ii) the effect of syllable number on duration varies by vowel height, and (iii) the effect of focus type on duration varies by verb type. The random effects, again, included SPEAKER and WORD, but a random slope for SYLLABLE NO. could only be used with the random effect SPEAKER; the model is summarized in (11). A likelihood ratio test showed that the model is highly significant ($\chi^2(5) = 96.839$, $p<0.001$), with the conditional R^2 of 0.610.

(11) log(Duration) ~ VOWEL HEIGHT * SYLLABLE NO. + FOCUS TYPE * VERB TYPE +
(1 + SYLLABLE NO. | SPEAKER) +
(1 | WORD)

The duration data, broken down according to three significant fixed effects (VOWEL HEIGHT, SYLLABLE NO., and FOCUS TYPE), and presented separately based on VERB TYPE, is provided in panels (a) and (b) of Figure 2. As the panel (b) of Figure 2 shows, initial stress in low-vowel imperatives leads to initial vowels being longer than final ones, reversing the pattern shown for low-vowel indicatives in panel (a) – in a syntagmatic fashion described in Section 2.4. For mid-vowels, though, initial stress in imperatives leads to greater duration of initial vowels that makes them longer than their unstressed counterparts in indicatives, though not longer than final mid vowels in imperatives; this exemplifies paradigmatic signaling of stress.

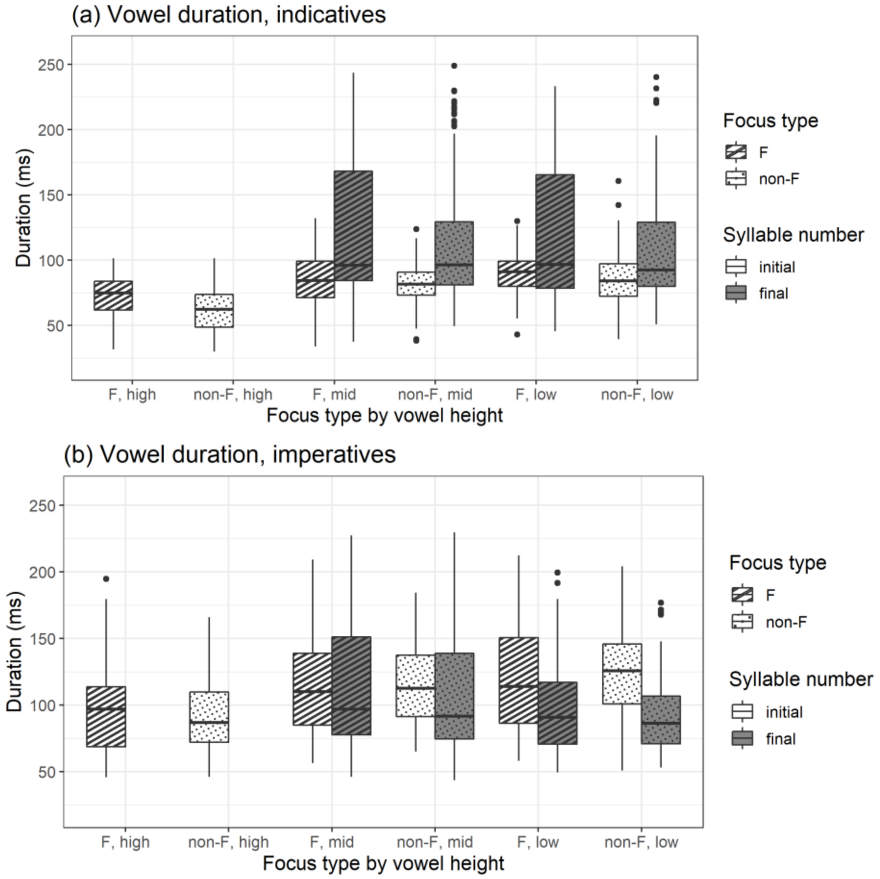


Figure 2. Vowel duration in the second study, shown for indicatives (a) and imperatives (b), and organized by VOWEL HEIGHT (individual bars), SYLLABLE NO. (white vs. grey), and FOCUS TYPE (stripes vs. dots). Note that high vowels are only available in non-final syllables.

The output of the model is provided in Table 9.[5] It shows that syllable number alone does not significantly affect duration, while vowel quality, verb type, and focus type do (the latter to a lesser degree, though the effect is still significant). Additionally, there are significant interaction effects of syllable number and vowel height, and verb type and focus type.

Table 9. Model output for duration in the first study

[5] In linear mixed-effect models in the second study, initial high vowels (in imperatives) acted as the intercept – as opposed to final high vowels in the first study. This is because final high vowels are not attested in the second study, and, as such, would not make a for a meaningful intercept.

		Estimate	SE	df	t-value	p-value
[1]	Intercept (initial, high, imperative, non-F)	6.380	0.115	5.746	55.266	<0.001***
[2]	**final**, high, imperative, non-F	0.098	0.155	4.204	0.633	0.559
[3]	initial, **low**, imperative, non-F	0.408	0.046	83.132	8.797	<0.001***
[4]	initial, **mid**, imperative, non-F	0.351	0.040	610.526	8.773	<0.001***
[5]	initial, high, **indicative**, non-F	-0.264	0.054	1409.670	-4.926	<0.001***
[6]	initial, high, imperative, **F**	-0.054	0.025	1412.925	-2.191	<0.05*
[7]	**final**, **low**, imperative, non-F	-0.152	0.044	1432.021	-3.451	<0.001***
[8]	final, high, **indicative**, **F**	0.076	0.034	1410.513	2.244	<0.05*

Interactions of fixed effects indicate that their effect on duration is non-uniform. To start with the effect of syllable number, row [2] in Table 9 shows that there is no significant durational difference between initial and final high vowels (where the values for the latter are estimated by the model). A pairwise comparison with the emmeans() function shows that this is also the case for low vowels (Estimate = -0.0541, SE = 0.155, df = 4.24, t = -0.349, p = 0.744) and mid vowels (Estimate = -0.0980, SE = 0.155, df = 4.20, t = 0.633, p = 0.560). Next, row [5] in Table 9 shows that there is a significant difference between vowel durations in imperative and indicative verbs in the non-F condition. A pairwise comparison shows that is the case in the F condition, too (Estimate = 0.105, SE = 0.0278, df = 1414, t = 3.774, p <0.001***). Finally, row [6] in Table 9 shows that, within imperative verbs, the effect of focus on duration is significant. A pairwise comparison shows that this is not the case in indicative verbs, though (Estimate = -0.0237, SE = 0.0268, df = 1414, t = -0.882, p = 0370). In other words, focus is marked by duration in imperatives but not in indicatives.

To sum up, the first study has shown that vowel height and syllable number, but not focus type, affect vowel duration, which means that vowel duration is a cue for stress but not focus type. The results of the second study are more complex, demonstrating that vowel height, syllable number, verb type and focus type all affect duration. A more in-depth look shows that duration is a cue for stress, but cues focus marking in imperatives only. We also observe considerable inter-speaker variation with respect to using duration as a cue for stress and/or focus; more on this in Section 5.

4.2 Intensity

4.2.1 First study (nominals)

The mean vowel intensity values obtained in the first study are summarized in Table 10. As these results show, final (stressed) and initial (unstressed) syllables have comparable intensity values. High vowels consistently have higher intensity in final syllables, in both focus types and syllable counts. The same cannot be said about mid or low vowels though: they consistently have lower intensity values in the final syllables.

Table 10. Mean vowel intensity in the first study

Focus type	Syllable count	Vowel height	n	Initial syllable		Final syllable	
				Intensity (dB)	**SD (dB)**	**Intensity (dB)**	**SD (dB)**
Focused	**disyll**	**high**	112	57.09	3.10	58.56	2.66
		mid	99	59.78	4.60	59.13	3.60
		low	52	60.92	3.02	60.53	3.02
	trisyll	**high**	67	57.33	3.08	59.09	2.46
		mid	37	60.15	3.33	58.08	4.60
		low	27	61.50	3.05	60.44	3.18
Non-focused	**disyll**	**high**	168	56.31	3.30	57.74	2.94
		mid	147	59.12	3.44	58.99	3.35
		low	77	60.55	3.25	60.76	3.23
	trisyll	**high**	98	56.57	3.36	58.52	2.78
		mid	63	60.13	3.23	58.53	3.37
		low	40	59.82	3.40	59.12	3.69

A mixed-effects model that provided the best fit included VOWEL HEIGHT and SYLLABLE NO. as interacting fixed effects, SPEAKER and WORD as random effects, and by-SPEAKER and by-WORD random slopes for SYLLABLE NO. The interaction of fixed effects improved the model fit significantly, suggesting that the effect of vowel height on intensity varies by syllable position. The model is summarized in (12). FOCUS TYPE did not significantly affect the intensity values, which suggests that intensity does not mark focus. According to a likelihood ratio test, the model is highly significant ($\chi^2(5) = 98.61$, $p<0.001$), with the conditional R^2 of 0.592.

(12) Intensity ~ VOWEL HEIGHT * SYLLABLE NO. +
(1 + SYLLABLE NO. | SPEAKER) +
(1 + SYLLABLE NO. | WORD)

The intensity results, broken down by VOWEL HEIGHT and SYLLABLE NO. (the only significant fixed effects) is visualized in Figure 3. As it demonstrates, despite the significant results, stress does not systematically correspond to higher or lower intensity values in different vowel heights.

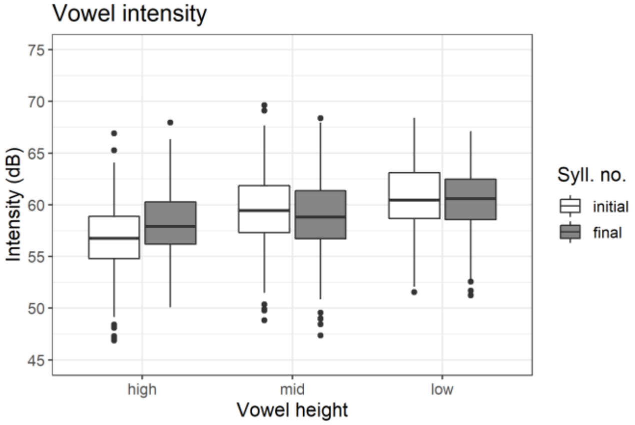


Figure 3. Vowel intensity in the first study, organized by VOWEL HEIGHT and SYLLABLE NO. (the only significant fixed effects).

The output of the model is provided in Table 11. According to it, syllable number (which corresponds to stress in the first study) has a significant effect on vowel intensity. So does vowel height, and the interaction of syllable number and vowel height suggests that the effect of syllable number/stress on intensity varies by vowel height. Row [2] in Table 11 shows this effect for high vowels. An emmeans() calculation of the missing pairwise comparisons showed that this effect also holds for mid vowels (Estimate = -0.810, SE = 0.311, df = 25.6, t = -2.606, p < 0.05*) but not for low vowels (Estimate = -0.351, SE = 0.381, df = 49.1, t = -0.923, p = 0.36).

Table 11. Model output for intensity in the first study

		Estimate	SE	df	t-value	p-value
[1]	Intercept (final, high)	57.943	0.842	6.440	68.788	p<0.001***
[2]	**initial**, high	-1.597	0.278	19.939	-5.739	p<0.001***
[3]	final, **low**	2.012	0.309	105.607	6.509	p<0.001***
[4]	final, **mid**	0.517	0.255	109.185	2.026	p<0.05*
[5]	**initial, low**	1.948	0.387	105.390	5.031	p<0.001***
[6]	**initial, mid**	2.407	0.320	109.834	7.512	p<0.001***

4.2.2 Second study (verbs)

Mean vowel intensity values for initial and final syllables in verbs, broken down by verb type, focus type, syllable count, and vowel height are provided in Table 12. As before, the greyed-out cells indicate the cells for which no vowels were attested. The mid vowels that were used in the final syllables in the "high-vowel" contexts are pooled with the other final mid vowels. Similarly to the picture for low and mid vowels in the first study, the intensity values in the final syllables are typically lower than in the initial syllables, across contexts. The difference between intensity values in the initial and final syllables is more pronounced in the imperatives. This is consistent with the overall tendency for intensity to fall throughout a word/prosodic constituent. In indicatives, this tendency is mitigated somewhat by the fact that final stress brings up the intensity values on the final vowel, leading to more levelled intensity values between the two syllables. In contrast, in imperatives, this tendency is more pronounced, because initial stress gives an extra intensity boost to the initial vowel. This picture is also consistent with paradigmatic cuing of stress.

Table 12. Mean vowel intensity in the first study

Verb type	**Focus type**	**Syllable count**	**Vowel height**	**Initial syllable**			**Final syllable**		
				n	**Intensity (dB)**	**SD (dB)**	**n**	**Intensity (dB)**	**SD (dB)**
Indicative	**Focused**	**disyll**	**high**	32	55.09	2.79			
			mid	35	57.83	3.10	67	57.48	3.44
			low	35	59.27	4.31	35	59.02	4.35
		trisyll	**high**	24	56.36	2.69			
			mid	21	58.73	3.52	45	57.81	3.83
			low	19	59.00	4.24	19	57.71	4.43
	Non-focused	**disyll**	**high**	42	56.02	2.45			
			mid	41	57.55	2.50	83	57.80	3.08
			low	45	59.43	3.90	45	58.39	3.69
		trisyll	**high**	28	55.41	1.91			

			mid	25	57.49	3.38	53	56.89	3.25
			low	25	58.80	3.75	25	57.04	3.91
Imperative	**Focused**	**disyll**	**high**	30	57.65	2.99			
			mid	33	59.64	3.05	63	58.93	2.68
			low	35	61.57	4.18	35	59.66	3.94
		trisyll	**high**	23	57.60	2.17			
			mid	18	59.79	4.58	41	56.82	3.53
			low	20	61.66	4.19	20	57.98	3.88
	Non-focused	**disyll**	**high**	41	57.47	2.76			
			mid	43	59.86	2.98	84	57.81	3.87
			low	42	60.97	3.57	42	57.76	3.87
		trisyll	**high**	28	58.17	2.15			
			mid	24	60.22	3.69	52	56.22	3.79
			low	24	61.11	3.87	24	55.92	4.99

A mixed-effects model that fit the data best included VOWEL HEIGHT, SYLLABLE NO. and VERB TYPE as fixed effects, SPEAKER and WORD as random effects, and random slopes for SYLLABLE NO. in both random effects. Possible interactions of fixed effects did not improve the model fit. The model is summarized in (13). Like in the first study, FOCUS TYPE did not significantly affect the intensity values, which suggests that intensity does not mark focus. According to a likelihood ratio test, the model is highly significant ($\chi^2(4) = 120.7$, $p<0.001$), with the conditional R^2 of 0.670.

(13) Intensity ~ VOWEL HEIGHT + SYLLABLE NO. + VERB TYPE
(1 + SYLLABLE NO. | SPEAKER) +
(1 + SYLLABLE NO. | WORD)

The intensity results, broken down by the significant fixed effects (VOWEL HEIGHT and SYLLABLE NO.) and visualized separately for the two VERB TYPES (indicatives and imperatives) are presented in the two panels of Figure 4.

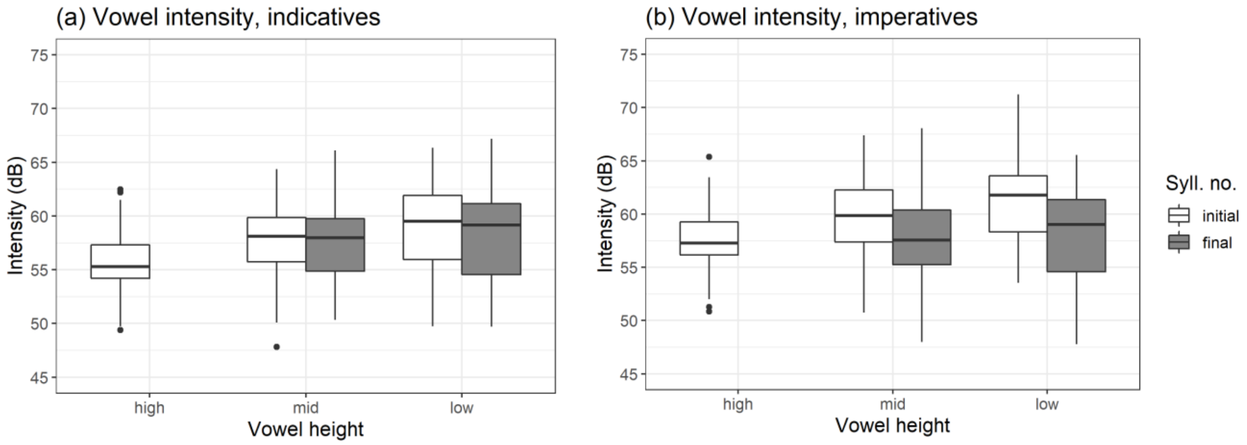


Figure 4. Vowel intensity in the second study, shown separately for indicatives (a) and imperatives (b), and organized according to the significant fixed effects VOWEL HEIGHT and SYLLABLE NO. High vowels are only attested in non-final syllables.

The output of the model is provided in Table 13. As it demonstrates, syllable number has an effect on intensity (smaller than the other factors, but still significant), and so do vowel height and verb type (both highly significant). Lack of interactions between the fixed effects suggest that they affect intensity in a uniform way.

Table 13. Model output for intensity in the second study

		Estimate	SE	df	t-value	p-value
[1]	Intercept (initial, high, imperative)	57.188	1.131	4.459	50.572	p<0.001***
[2]	**final**, high, imperative	-1.783	0.608	5.078	-2.931	p<0.05*
[3]	initial, **low**, imperative	3.303	0.367	42.372	9.011	p<0.001***
[4]	initial, **mid**, imperative	2.442	0.344	44.957	7.107	p<0.001***
[5]	initial, high, **indicative**	-1.002	0.118	1374.342	-8.527	p<0.001***

To sum up the intensity results, we have seen that intensity consistently marks stress in both studies but does not mark focus.

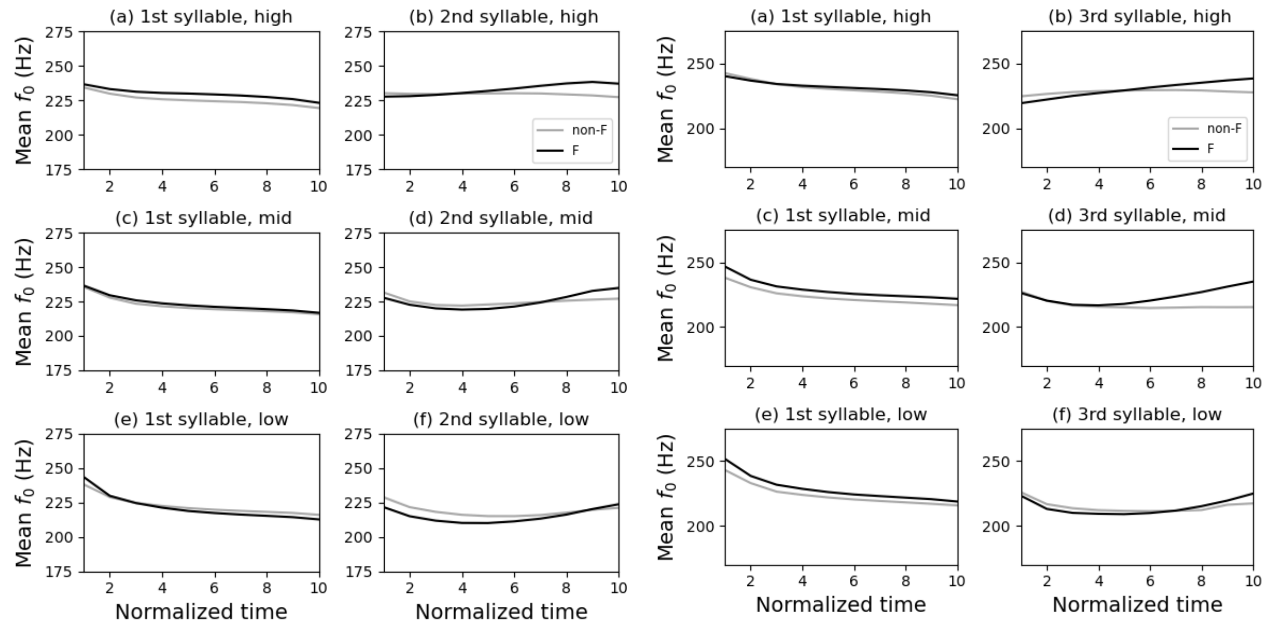
4.3 Fundamental frequency (f_0)

4.3.1 First study (nominals)

The mean f_0 values per vowel, collected from the test words in the first study, are summarized in Table 14. Additionally, as an illustration, Figure 5 demonstrates average f_0 contours per vowel (the figure is divided by syllable number rather than by focus type for an easier comparison of the effect of focus). As these results show, average f_0 values are similar across contexts, with a slight fall from the initial to the final syllable being common. As Figure 5 shows, focus is often marked by a slight rise on toward the end of the vowel – but it is too subtle to be reflected in the mean values.

Table 14. Mean f_0 per syllable in the first study

Focus type	**Syllable count**	**Vowel height**	**n**	**Initial syllable**		**Final syllable**	
				f_0 (Hz)	**SD (Hz)**	**f_0 (Hz)**	**SD (Hz)**
Focused	**disyll**	**high**	112	229.57	24.16	232.86	18.52
		mid	99	223.36	26.91	225.02	18.60
		low	52	221.48	24.49	215.41	17.95
	trisyll	**high**	67	231.98	24.34	229.82	19.62
		mid	37	228.92	22.50	223.56	27.76
		low	27	228.49	21.37	214.57	30.13
Non-focused	**disyll**	**high**	168	225.40	29.15	229.46	25.16
		mid	147	221.74	26.17	225.09	20.65
		low	77	222.62	24.74	218.99	24.76
	trisyll	**high**	98	230.90	25.75	228.08	22.17
		mid	63	223.54	26.44	217.07	26.19
		low	40	223.87	25.71	214.85	28.56



(a) Averaged f_0 contour per vowel, disyllables

(b) Averaged f_0 contour per vowel, trisyllables

Figure 5. f_0 contours per vowel in the first study

A mixed-effects model that provided the best fit for the data included VOWEL HEIGHT and SYLLABLE NO. as interacting fixed effects, SPEAKER and WORD as random effects, and by-SPEAKER and by-WORD random slopes for SYLLABLE NO. The interaction of fixed effects significantly improved the model fit. The model is summarized in (14). FOCUS TYPE did not significantly affect the mean f_0 values (though, as Figure 5 demonstrates, the effect of focus may be reflected in the final rise, which is not captured by the model). A likelihood ratio test shows that the model is highly significant ($\chi^2(5) = 60.296$, $p<0.001$), with the conditional R^2 of 0.588.

(14) Mean (f_0) ~ VOWEL HEIGHT * SYLLABLE NO. +
(1 + SYLLABLE NO. | SPEAKER) +
(1 + SYLLABLE NO. | WORD)

The mean f_0 values per vowel, organized according to the significant fixed effects (VOWEL HEIGHT and SYLLABLE NO.), are shown in Figure 6.

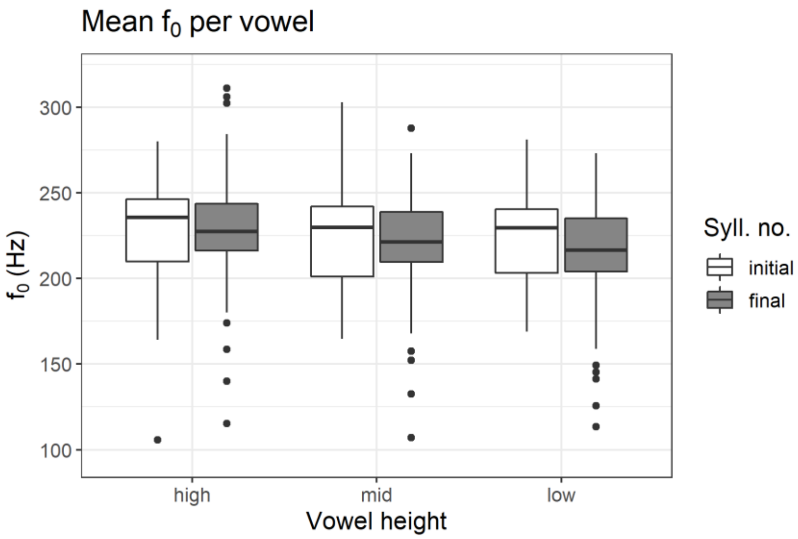


Figure 6. Mean f_0 per vowel in the first study, broken down by VOWEL HEIGHT and SYLLABLE NO. (the significant fixed effects).

The output of the model is provided in Table 15. It shows that syllable number by itself does not affect f_0, while vowel height does. Additionally, there is a significant effect of the interaction between syllable number and vowel height, though only for initial low vowels (row [5]) and not for initial mid vowels (row [6]). The interaction between the fixed effects also allows for looking into whether there is a difference in f_0 values between initial and final syllables for vowel heights other than "high" (row [2]). An emmeans() calculation of the missing pairwise comparisons shows that there is no significant difference in f_0 values between initial and final syllables either for mid vowels (Estimate = -0.184, SE = 8.70, df = 7.58, t = -0.021, p = 0.984) or low vowels (Estimate = -6.824, SE = 8.81, df = 8.02, t = -0.774, p = 0.461), consistently with the high vowels.

Table 15. Model output for f_0 in the first study

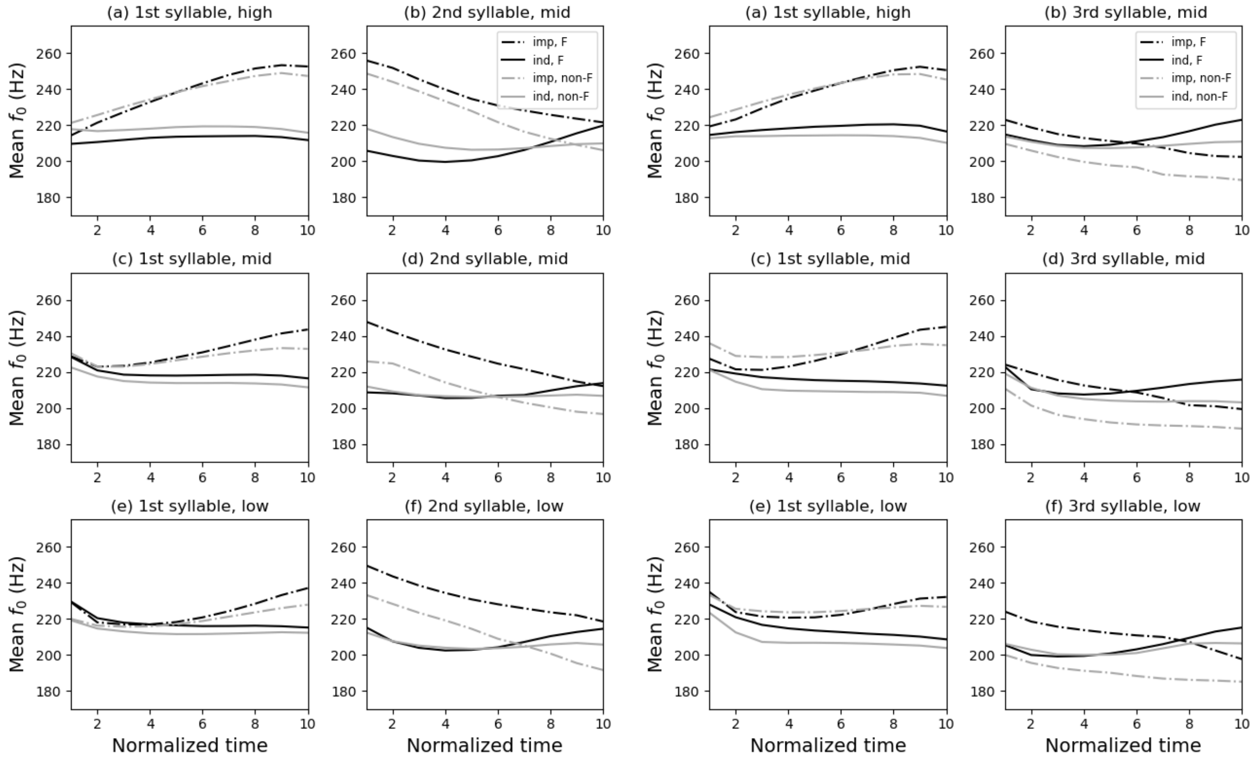
		Estimate	SE	df	t-value	p-value
[1]	Intercept (final, high)	229.199	5.524	6.286	41.491	p<0.001***
[2]	**initial**, high	-1.475	7.939	6.271	-0.186	p=0.858
[3]	final, **low**	-13.286	1.758	104.867	-7.556	p<0.001***
[4]	final, **mid**	-6.727	1.458	110.156	-4.613	p<0.001***
[5]	**initial, low**	8.299	2.438	104.216	3.405	p<0.001***
[6]	**initial, mid**	1.659	2.023	109.653	0.820	p=414

4.3.2 Second study (verbs)

The mean f_0 results per vowel that were obtained in the second study are provided in Table 16. Figure 7 additionally presents the averaged f_0 contours over the vowels in all experimental contexts. As the results show, in both the focused and non-focused condition, imperative verbs typically have higher f_0 values than their indicative counterparts. Interestingly, this holds for both initial syllables and final syllables. Within each verb type, focused verbs also typically have higher overall f_0 values than their unfocused counterparts. The drop in f_0 between the initial and final syllables is steeper in the imperatives than in the indicatives.

Table 16. Mean f_0 values per vowel in verbs

Verb type	Focus type	Syllable count	Vowel height	Initial syllable			Final syllable		
				n	f_0 (Hz)	SD (Hz)	n	f_0 (Hz)	SD (Hz)
Indicative	Focused	disyll	high	32	212.50	35.49			
			mid	35	219.30	35.36	67	207.45	18.90
			low	35	217.98	40.20	35	208.02	19.52
		trisyll	high	24	218.15	31.51			
			mid	21	215.92	28.94	45	212.96	22.64
			low	19	214.75	32.53	19	205.07	20.18
	Non-focused	disyll	high	42	217.96	39.85			
			mid	41	214.85	31.79	83	208.54	24.16
			low	45	213.02	34.14	45	205.82	25.86
		trisyll	high	28	213.43	33.00			
			mid	25	210.71	23.89	53	208.02	25.12
			low	25	208.35	31.90	25	203.27	24.38
Imperative	Focused	disyll	high	30	238.17	24.53			
			mid	33	231.61	23.93	63	231.61	35.16
			low	35	224.23	20.87	35	231.43	35.77
		trisyll	high	23	238.93	14.42			
			mid	18	231.00	23.86	41	210.34	39.54
			low	20	225.97	20.52	20	211.15	34.63
	Non-focused	disyll	high	41	237.86	22.70			
			mid	43	228.37	21.74	84	217.59	32.99
			low	42	220.14	21.82	42	211.99	37.63
		trisyll	high	28	239.39	21.03			
			mid	24	231.83	21.75	52	196.10	33.13
			low	24	225.96	16.27	24	190.13	29.27


(a) Averaged f_0 contour per vowel, disyllables

(b) Averaged f_0 contour per vowel, trisyllables

Figure 7. f_0 contours per vowel in the second study

A mixed-effects model that fit the data best included VOWEL HEIGHT, VERB TYPE, and FOCUS TYPE as fixed effects, and SPEAKER and WORD as random effects. No random slopes were included (adding them to the model led to non-convergence). Interestingly, SYLLABLE NO. did not turn out to have a significant effect on f_0, in contrast with the other acoustic measures discussed so far. Including interactions of the fixed effects did not improve the model fit. The model is summarized in (15). A likelihood ratio test showed that the model is highly significant ($\chi^2(4) = 138.24$, $p<0.001$), with the conditional R^2 of 0.410.

(15) Mean (f_0) ~ FOCUS TYPE + VOWEL HEIGHT + VERB TYPE
(1 | SPEAKER) +
(1 | WORD)

The distribution of the mean f_0 values, organized by the significant fixed effects VOWEL HEIGHT and FOCUS TYPE, and shown separately for the two VERB TYPES, is illustrated in the two panels of Figure 8. Note that because SYLLABLE NO. was not a significant factor, initial and final vowel are lumped together in Figure 8.

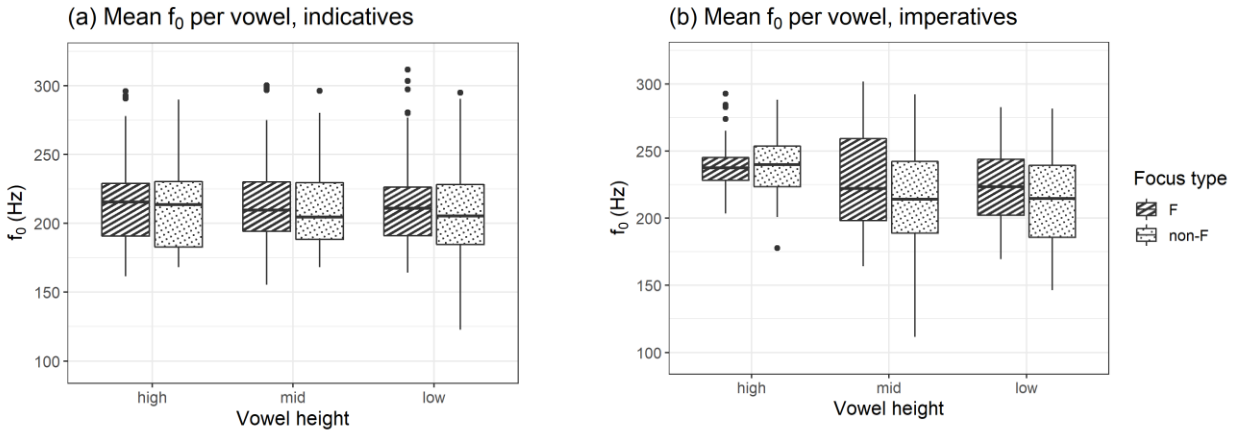


Figure 8. f_0 values in the second study, shown for indicatives (a) and imperatives (b), organized according to VOWEL HEIGHT and FOCUS TYPE (the significant fixed effects). Note the absence of SYLLABLE NO. as a fixed effect.

The output of the model is provided in Table 17. As it demonstrates, each of vowel height, verb type and focus type has a highly significant effect on f_0. Lack of interactions between the fixed effects suggests that they affect f_0 in a uniform way.

Table 17. Model output for f_0 in the second study

		Estimate	SE	df	t-value	p-value
[1]	Intercept (high, imperative, non-F)	226.088	8.635	4.72	26.184	p<0.001***
[2]	**low**, imperative, non-F	-12.859	2.089	88.045	-6.157	p<0.001***
[3]	**mid**, imperative, non-F	-11.502	1.861	475.173	-6.181	p<0.001***
[4]	high, **indicative**, non-F	-11.763	1.266	1428.234	-9.291	p<0.001***
[5]	high, imperative, **F**	4.385	1.223	1425.121	3.584	p<0.001***

Let us sum up the f_0 results. The first study shows that, in a set of data with uniformly final stress, f_0 is used to cue syllable number/stress but not focus. The second study shows that, in a dataset that contains stress-based minimal pairs, f_0 is used to cue both the verb type – imperative (i.e., with initial stress) versus indicative (i.e., with final stress) – and presence versus absence of focus. Interestingly, f_0 is not used to differentiate syllable position (initial versus final), which suggests that a given verb type exhibits characteristic f_0 values that differentiate it from verbs of the opposite type on both the final and initial syllables. As was the case with duration, we also observe wide inter-speaker variation in the f_0 contour utilized; more on this in Section 5.

4.4 Vowel height (F1) [6]

4.4.1 First study (nominals)

The mean F1 values per vowel in the first study are summarized in Table 18. Additionally, as an illustration, Figure 9 demonstrates both F1 and F2 parameters of the vowels. As these results show, the F1

[6] We refrain from making conclusions about the properties of the other formant that defines vowel quality, F2. This is due to the fact that the "horizontal" grouping of vowels employed here, organized by vowel height, does not provide a good context for comparing changes that have to do with the "vertical", frontness-backness dimension. An alternative solution would be to consider F2 (as well as F1) values of individual vowels, as opposed to vowels grouped by height. In this case, though, the only significant fixed effect turns out to be vowel identity, and not stress or focus, which is not informative for the current study. In view of this, we are only discussing properties of F1, for which the "horizontal" vowels grouping that we are using makes sense.

values are typically higher in the final syllables than in the initial syllables and tend to be lower in the non-F condition as compared to the F condition.

Table 18. Mean F1 values per vowel in the first study

Focus type	Syllable count	Vowel height	n	Initial syllable		Final syllable	
				f_0 (Hz)	**SD (Hz)**	**f_0 (Hz)**	**SD (Hz)**
Focused	**disyll**	**high**	112	438.86	64.20	451.35	51.34
		mid	99	590.89	76.23	612.16	85.04
		low	52	793.57	73.90	844.06	52.50
	trisyll	**high**	67	447.67	60.04	427.39	46.99
		mid	37	605.64	82.75	593.56	85.59
		low	27	779.89	49.51	840.45	51.76
Non-focused	**disyll**	**high**	168	437.13	68.41	452.42	49.93
		mid	147	568.54	71.87	611.48	84.04
		low	77	757.70	88.74	807.09	70.98
	trisyll	**high**	98	432.36	55.23	435.08	45.28
		mid	63	586.60	84.24	600.41	77.92
		low	40	749.06	60.84	774.88	69.78

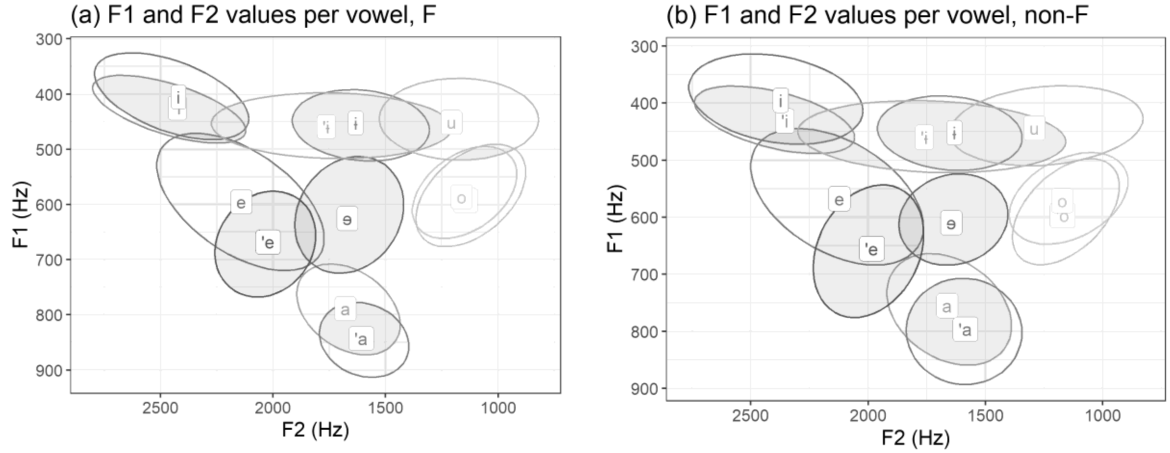


Figure 9. F1 and F2 per vowel in the first study, shown separately for the focused (a) and non-focused (b) conditions, and broken down by stress. The stressed vowels are identified with the IPA sign for primary stress.

A mixed-effects model that fit the data best included SYLLABLE NO., VOWEL HEIGHT and FOCUS TYPE as fixed effects, with VOWEL HEIGHT and FOCUS TYPE interacting, SPEAKER and WORD as random effects, and random slopes for SYLLABLE NO. in both random effects. The interaction of the two out of three fixed effects significantly improved the model fit. The model is summarized in (16). A likelihood ratio test shows that the model is highly significant ($\chi^2(6) = 357.96$, $p<0.001$), with the conditional R^2 of 0.906.

(16) Mean (F1) ~ FOCUS TYPE * VOWEL HEIGHT + SYLLABLE NO. +
(1 + SYLLABLE NO. | SPEAKER) +
(1 + SYLLABLE NO. | WORD)

The F1 results are, organized according to the significant fixed factors (FOCUS TYPE, SYLLABLE NO., and VOWEL HEIGHT), are visualized in Figure 10. They demonstrate the same overall tendencies as the results in Table 18: both stress and focus are associated with higher F1 values.

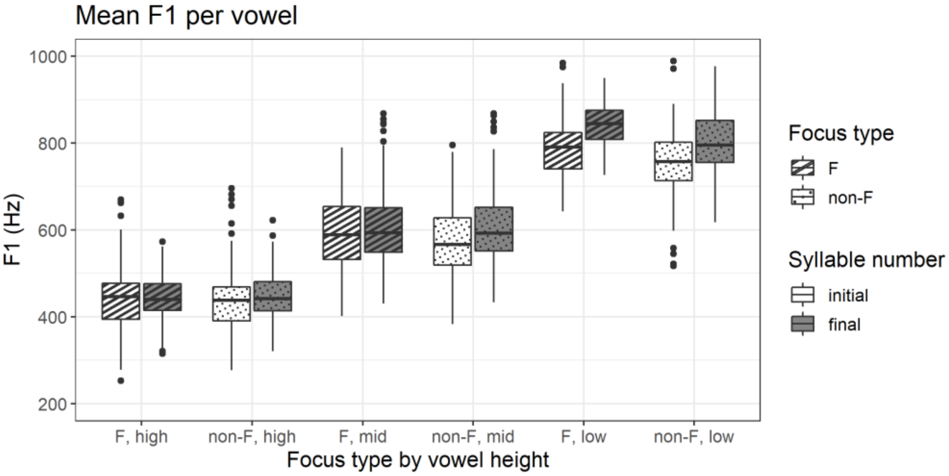


Figure 10. F1 per vowel in the first study, broken down by the significant fixed factors (FOCUS TYPE, SYLLABLE NO., and VOWEL HEIGHT).

The output of the model is provided in Table 19. It shows that the F1 values are affected by syllable number and focus type (in addition to vowel quality, which is directly tied to differences in F1). The interaction between vowel height and focus type also allows for looking into whether the significant difference in F1 between the two focus contexts holds for vowels of all heights. Row [3] in Table 19 shows that it does for high vowels, and an emmeans() calculation of the missing pairwise comparisons shows that the same is true for low vowels (Estimate = -27.07, SE = 4.89, df = 1756, t = -5.532, p <0.001***) but not for mid vowels (Estimate = 2.54, SE = 3.77, df = 1761, t = 0.674, p = 0.5).

Table 19. Model output for f_0 in the second study

		Estimate	SE	df	t-value	p-value
[1]	Intercept (final, high, non-F)	453.513	13.511	8.166	33.567	p<0.001***
[2]	**initial**, high, non-F	-21.962	8.558	12.271	-2.566	p<0.05*
[3]	final, high, **F**	-12.365	3.327	1751.999	-3.717	p<0.001***
[4]	final, **low**, non-F	337.263	8.620	124.217	39.125	p<0.001***
[5]	final, **mid**, non-F	150.331	7.077	125.853	21.242	p<0.001***
[6]	final, **low**, **F**	39.434	5.733	1749.579	6.878	p<0.001***
[7]	final, **mid**, **F**	9.824	4.817	1752.277	2.039	p<0.05*

4.4.2 Second study (verbs)

Table 20 shows the mean F1 values for the vowels in the second study. Like with the F1 results in the first study, in Figure 11 we are also providing the F1 by F2 distribution for the vowels in the second study, divided by verb type and focus type. Similarly to the first study, there is a tendency for stressed syllables (initial in imperatives, final in indicatives) to have higher F1 values, and for F1 values in the F condition to be higher than in the non-F condition.

Table 20. Mean F1 in verbs

Verb type	**Focus type**	**Syllable count**	**Vowel height**	**Initial syllable**			**Final syllable**		
				n	**F1 (Hz)**	**SD (Hz)**	**n**	**F1 (Hz)**	**SD (Hz)**
Indicative	**Focused**	**disyll**	**high**	32	411.64	63.21			
			mid	35	617.62	88.49	67	616.83	94.39
			low	35	790.08	49.69	35	852.63	69.49
		trisyll	**high**	24	430.78	58.73			
			mid	21	602.41	75.25	45	685.11	85.63
			low	19	792.82	38.26	19	822.79	64.88
	Non-focused	**disyll**	**high**	42	412.51	64.53			
			mid	41	607.19	104.27	83	606.12	97.00
			low	45	767.45	78.70	45	817.29	92.64
		trisyll	**high**	28	419.53	59.10			
			mid	25	556.99	89.49	53	651.08	97.80
			low	25	784.86	58.47	25	801.95	95.49
Imperative	**Focused**	**disyll**	**high**	30	407.36	60.07			
			mid	33	635.71	97.70	63	660.68	101.87
			low	35	880.17	70.51	35	817.74	81.72
		trisyll	**high**	23	435.87	53.40			
			mid	18	632.11	101.46	41	690.91	113.46
			low	20	790.79	99.15	20	870.68	45.64
	Non-focused	**disyll**	**high**	41	411.36	72.11			
			mid	43	610.88	92.76	84	620.74	110.20
			low	42	854.12	74.90	42	798.01	96.93
		trisyll	**high**	28	434.76	66.38			
			mid	24	609.41	98.40	52	665.96	121.95
			low	24	866.61	59.67	24	766.90	102.67

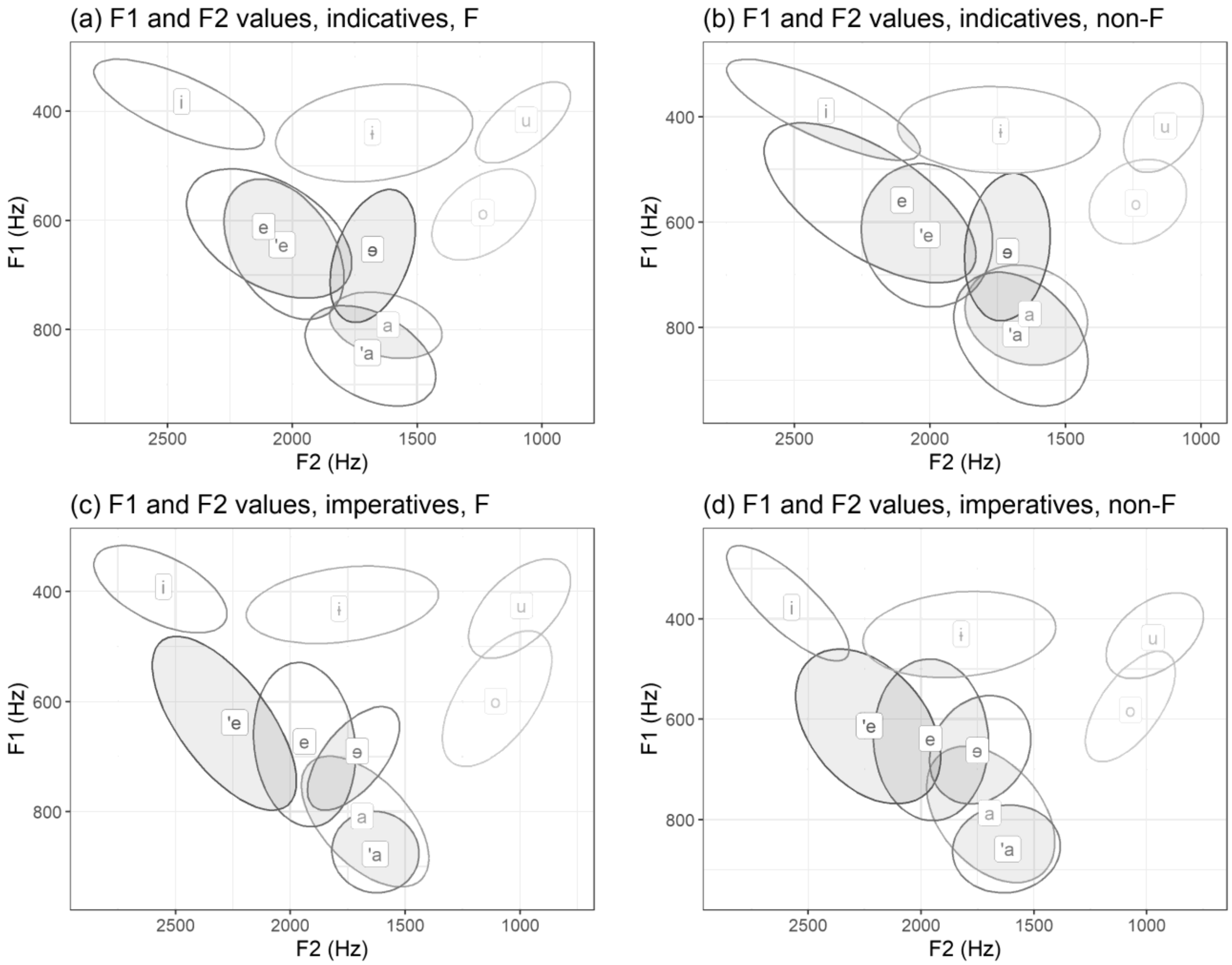


Figure 11. F1 and F2 values per vowel in the second study, shown separately for the focused (a, c) and non-focused (b, d) conditions, as well as indicatives (a, b) and imperatives (c, d), and broken down by stress. The stressed vowels are identified with the IPA sign for primary stress.

A mixed-effects model that provided the best fit for the data consisted of VOWEL HEIGHT and VERB TYPE as fixed effects, and SPEAKER and WORD as random effects. Including random slopes led to the non-convergence of the model. Including an interaction of the fixed effects did not improve the model fit. Interestingly, neither FOCUS TYPE nor SYLLABLE NO. turned out to be significant factors. The model is summarized in (17). A likelihood ratio test showed that the model is highly significant ($\chi^2(3) = 1286.1$, $p<0.001$), with the conditional R^2 of 0.878.

(17) Mean (F1) ~ VOWEL HEIGHT + VERB TYPE
(1 | SPEAKER) +
(1 | WORD)

Figure 12 visualizes the distribution of the F1 data, shown separately for the two verb types, and organized according to the only remaining significant fixed factor, VOWEL HEIGHT.

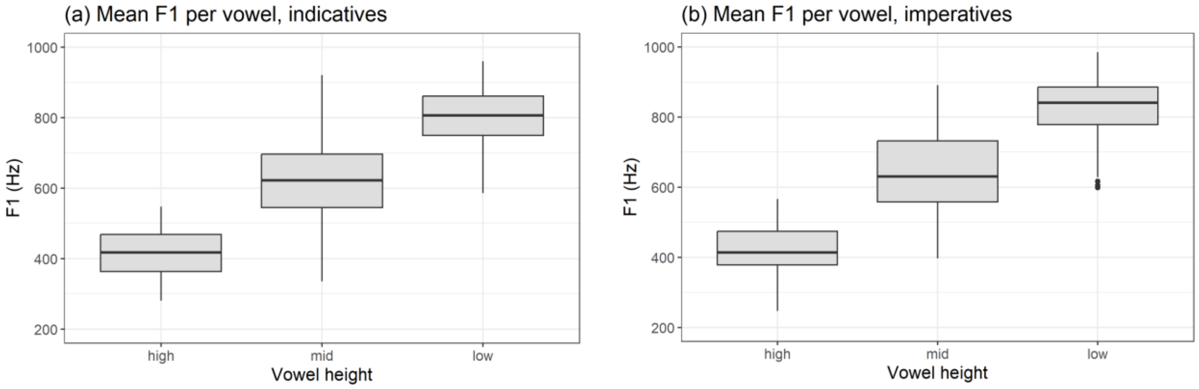


Figure 12. F1 values per vowel in the second study, shown separately for indicatives (a) and imperatives (b), and broken down by the only significant fixed factor (other than VERB TYPE) – VOWEL HEIGHT. Note that neither FOCUS TYPE nor SYLLABLE NO. are significant in this model.

The output of the model is provided in Table 21. As it shows, there is a highly significant effect of vowel height, which is expected, given the intrinsic connection between F1 and vowel height. It also shows that there is a systematic difference between the two verb types, but not the two syllable positions. This probably has to do with the morphological reasons, though: the set of stressed vowels in indicatives (/a, e/) corresponds to the unstressed vowels in imperatives, and vice versa. Because not all vowels are represented in each set, a difference between verbs on the whole but not individual syllable positions is detected. Lack of interaction between the fixed effects does not allow for looking into whether this is true of all vowel heights.

Table 21. Model output for F1 in the second study

		Estimate	SE	df	t-value	p-value
[1]	Intercept (high, imperative)	418.29	33.457	4.337	12.502	p<0.001***
[2]	high, **indicative**	-20.743	3.088	1419.227	-6.718	p<0.001***
[3]	**low**, imperative	402.259	10.293	53.785	39.083	p<0.001***
[4]	**mid**, imperative	220.115	5.141	1419.67	42.812	p<0.001***

To sum up the F1 results, in the first study, F1 was shown to systematically differ for vowels of different height (which is expected), but also syllable number and focus type. In the second study, F1 was not involved in focus marking, and instead only differed for vowels of different height and for different verb types.

5 Discussion

5.1 General

To recap, the goal of the two studies reported here was to investigate the acoustic expression of stress and focus in Udmurt, using a predetermined inventory of acoustic cues (duration, intensity, f_0, F1), in the context of fixed and contrastive stress. Our results show that different acoustic cues may be involved in marking both stress and focus. The most systematic behavior among the cues that we surveyed is exhibited by intensity: it was shown to consistently mark stress, in both studies, but was not involved in marking focus. The behavior of duration is more complex: in the first study, it marked stress but not focus; in the second study, it differentiated both verb types and syllable numbers, as well as focus types. Similarly to duration, f_0 in the first study cued stress but not focus, while in the second study it was shown to be a significant predictor for both focus and verb type. Finally, F1 in the first study cued both stress and focus, but only different verb types in the second study. The results of both studies are summarized in Table 22. Overall, our results show that all four acoustic cues investigated systematically participate in stress marking, while focus is expressed by fewer cues, which also differ from study to study.

Table 22. Summary of the results

	First study		**Second study**	
	Stress position	Focus type	Stress position/ verb type	Focus type
Duration	✓		✓	✓
Intensity	✓		✓	
f_0	✓		✓	✓
F1	✓	✓	✓	

While the studies aimed at investigating stress cues that also control for the focus structure of the utterance, as recommended by Roettger & Gordon (2017), are still relatively few, our results can be compared to some of those obtained for other languages. Suomi et al. (2001) show that, in Finnish, (contrastive) focus is marked both by f_0 and duration, while the position of stress is not marked by f_0 (the role of duration as a cue for stress is not discussed in detail). Similar results, with duration cuing stress and f_0 being reserved for intonational prominence, were obtained for Georgian (Borise 2023). Finally, in a study targeting four languages (Hungarian, Turkish, Greek, and Spanish), Vogel et al. (2016) highlight the cross-linguistic variability in stress- and focus-marking. Among other results, they show that, in Spanish and Greek, f_0 is the main cue for word stress, while duration and intensity, respectively, acted as important cues for focus – in contrast with the results for Finnish and Georgian. For Hungarian, a language with contrastive vowel length, they show that duration is not reliably used to cue stress or focus – both are expressed mainly with f_0. Further work within this methodology should help uncover more reliable cross-linguistic and language-specific tendencies.

5.2 Interspeaker variation

As noted in the sections on duration and f_0, we found that individual participants used the acoustic cues that we investigated differently, and also, in some cases, used them differently between the two studies. The small sample size does not allow us to identify these differences as merely idiolectal or representative of a particular variety of Udmurt, but we hope that highlighting them here can be instructive for future work on the prosodic phonology of Udmurt. For example, Table 23 shows that Sp1 uses a much greater increase in duration to mark stress than all other speakers, and does so consistently between the two studies, whereas, e.g., Sp4 does not use duration in the first study, and Sp5 does not in the second study.

Table 23. Duration (ms) results by speaker

		First study		**Second study**			
Speaker	Syllable number	F	Non-F	Indicative, F	Indicative, Non-F	Imperative, F	Imperative, Non-F
Sp1	initial	87.69	86.35	85.95	87.21	138.93	132.76
	final	172.24	177.45	191.72	193.21	167.13	164.94
Sp2	initial	71.43	78.77	75.70	75.42	85.77	87.69
	final	77.88	68.25	90.36	94.69	92.05	94.37
Sp3	initial		55.62				
	final		75.51				
Sp4	initial	80.86	79.70	88.90	85.79	133.11	139.15
	final	77.49	81.19	130.03	125.45	93.92	104.43

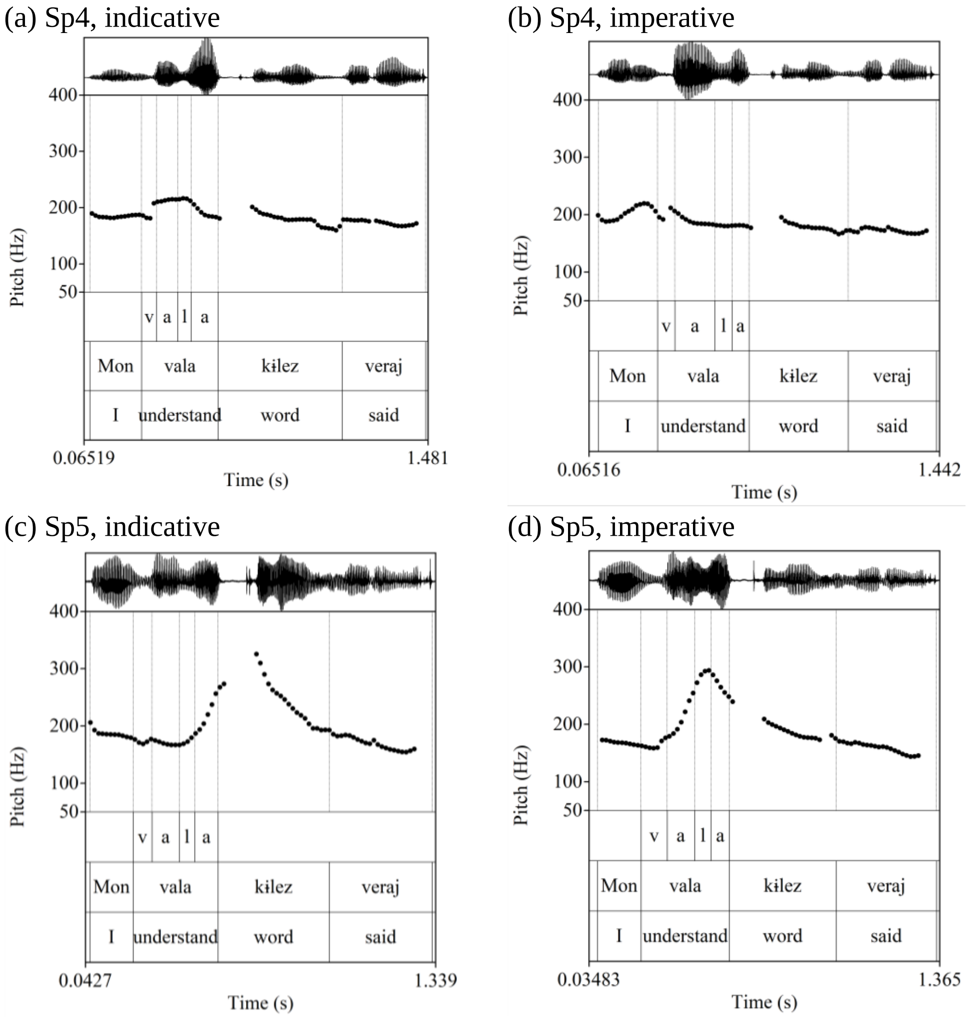
Sp5	initial	70.47	69.82	80.69	74.37	82.77	85.05
	final	81.17	77.57	77.85	81.53	76.10	81.59
Sp6	initial		58.53		63.26		109.33
	final		64.51		77.17		72.44

Similarly, Table 24 shows that there is also considerable variation in the use of f_0. For instance, Sp2 and Sp3 do not vary f_0 between the stressed and unstressed syllables in the first study and the indicatives in the second study (Sp2), and Sp1 and Sp4 do not use f_0 to make any contrasts (indicatives vs. imperatives, F vs. non-F, on corresponding syllables) in either study, with all differences being below 10Hz. Table 24 also presents evidence for qualitative differences in the use of f_0 between the speakers. It shows that in the first study, Sp1 and Sp4 used falling f_0 contours, while Sp5 and Sp6 used rising ones. In the second study, we see that Sp2 uses a rising contour on the imperatives while not varying f_0 on the indicatives, Sp5 continues to use a rising contour in most contexts in the second study, and Sp6 uses a rising contour with the indicatives and a falling one with the imperatives (i.e., aligns the stressed syllable with higher f_0).

Table 24. f_0 (Hz) results by speaker

		First study		**Second study**			
Speaker	Syllable number	F	Non-F	Indicative, F	Indicative, Non-F	Imperative, F	Imperative, Non-F
Sp1	initial	243.89	239.44	257.87	258.60	256.54	254.48
	final	209.91	215.33	212.66	213.22	207.07	207.04
Sp2	initial	242.88	254.04	228.57	230.91	234.54	239.66
	final	240.99	252.88	229.89	231.02	269.17	250.74
Sp3	initial		239.77				
	final		235.99				
Sp4	initial	229.96	225.07	210.60	207.52	212.59	213.74
	final	214.34	213.01	187.94	186.32	183.86	181.85
Sp5	initial	192.63	195.65	174.44	177.69	221.95	209.54
	final	236.21	213.34	203.30	186.21	236.17	202.28
Sp6	initial		197.70		195.56		233.69
	final		211.73		216.62		197.63

Some of the attested dimensions of variation are illustrated in Figure 13. Panel (a), an indicative verb produced by Sp4, demonstrates the increased duration of the stressed (final) vowel, and a falling f_0 contour; in panel (b), the effect of duration is even more apparent on the stressed (initial) vowel of an imperative verb, produced by the same speaker. The stressed syllable is also aligned with a low tone; there may be a leading high tone on the preceding pronoun. The overall magnitude of f_0 movement is quite small. Panels (c) and (d) provide the realizations of an indicative and imperative, respectively, by Sp5. Here, the stressed vowels are aligned with a high tonal target, with the rise on the stressed vowel and the peak reached on the following syllable. There is little evidence for greater duration of the stressed vowel. The magnitude of f_0 movement is also much larger, with the rise covering more than 100Hz. The f_0 scale is kept constant in panels (a-d) to allow for an easier cross-speaker comparison.


(a) Sp4, indicative

(b) Sp4, imperative

(c) Sp5, indicative

(d) Sp5, imperative

Figure 13. A sample of individual realizations from the second study

The availability of this variation with respect to acoustic cues used to mark stress and focus raises non-trivial questions about the processing and perception of stress and the nature of phonetic-phonology interface. It aligns with the available neurolinguistic evidence suggesting that speakers expect varying individual acoustic cues to be utilized in marking stress in a single language (Honbolygó & Csépe 2011). It provides support to the view that phonetic evidence may not provide straightforward one-dimensional physical corroboration for phonological concepts like stress (Keating 1996).

5.3 Autosegmental-Metrical interpretation of the f_0 contours

No Autosegmental-Metrical account of intonation in Udmurt has been developed so far, which means that we can only offer a tentative interpretation of the attested f_0 contours associated with stress in terms of individual tonal targets. Due to the scarcity of evidence, we refrain from addressing other issues of Udmurt intonational phonology at this time (e.g., boundary tones, phrasing patterns, etc.).

As Figure 7 shows for f_0 movements in the second study, the initial stressed syllable in imperatives is associated with a rise in f_0 that is mostly confined to the stressed syllable, with the peak reached towards its end, and a gradual fall throughout the rest of the word. This is likely due to the availability of the H* pitch accent in Udmurt. Final stress, as shown in both Figure 5 for the first study and Figure 7 for the second study, is associated with a drop or drop and rise in $f_{0.}$ This suggests a pitch accent with an L component, like L* or L*+H. As panels (a) and (b) in Figure 13 show, there may also be evidence for a high leading tone accompanying the low pitch accent, H+L*. In panels (c) and (d) of Figure 13 show, the pitch accent may also be realized as a steep rise on the stressed vowel, with the peak reached on the post-tonic syllable, preceded by a stretch of lower f_0 values. It may be analyzable as H*, or, again, as L*+H. Whether the emerging inventory of H*, L*, H+L* and H+L* pitch accents is substantiated for Udmurt should be explored in future research.

6 Conclusion

Our results show that all four acoustic parameters surveyed in the paper – duration, intensity, $f_{0,}$ and F1 – participate in stress marking in Udmurt. The results for focus marking vary by study and demonstrate that all cues except for intensity may be involved in focus marking. As expected, we also found that vowel height leads to significant differences in all acoustic cues, but, somewhat surprisingly, we found that number of syllables was not a significant effect in the presence of the random effect of word. The wide interspeaker variation demonstrates that the averaged results may present a somewhat simplified picture, while individual speakers may rely more heavily on a subset of the acoustic cues discussed to mark stress and/or focus. Finally, we offer a tentative Autosegmental-Metrical interpretation of our f_0 results; a full account of Udmurt intonation awaits further research.

References

Alatyrev, V. I. 1983. Kratkij grammatičeskij očerk udmurtskogo jazyka [Studies in the grammar of Udmurt]. In *Udmurtsko-russkij slovar'*, 561–591. Moscow: Russkij jazyk.

Alhoniemi, Alho. 2010. *Marin kielioppi [Mari grammar]*. Helsinki: Suomalais-Ugrilainen Seura.

Baitchura, Uzbek. 1973. A Few Remarks about Accentuation in Some Fenno-Ugric Languages. *Ural-Altaische Jahrbücher* 45. 80–87.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/doi:10.18637/jss.v067.i01.

Boersma, Paul & David Weenink. 2021. Praat: doing phonetics by computer [Computer program]. http://www.praat.org/. (25 March, 2021).

Borise, Lena. 2023. Disentangling word stress and phrasal prosody: a view from Georgian. *Phonological Data and Analysis* 5(1). 1–37.

Csúcs, Sándor. 1990. *Chrestomathia Votiacica*. Budapest: Tankönyvkiadó.

Denisov, V. N. 1980. *Foneticheskaja xarakteristika udarenija v sovremennom udmurtskom jazyke*.

Edygarova, Svetlana. 2014. The varieties of the modern Udmurt language. *Finnisch-Ugrische Forschungen* 62. 376–398.

Edygarova, Svetlana. 2015. Negation in Udmurt. In Matti Miestamo, Anne Tamm & Beáta Wagner-Nagy (eds.), *Negation in Uralic Languages*, 265–291. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.108.10edy.

Georgieva, Ekaterina. 2017. Person agreement on converbs in Udmurt. In F. Kiefer, J. Blevins & H. Bartos (eds.), *Perspectives on Morphological Organization: Data and Analyses* (Empirical Approaches to Linguistic Theory), vol. 10, 86–122. Leiden: Brill.

Gordon, Matthew & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1). https://doi.org/10.1515/lingvan-2017-0007.

Gries, Stefan Thomas. 2021. *Statistics for linguistics with R: a practical introduction* (Mouton Textbook). 3rd revised edition. Berlin: de Gruyter Mouton.

Honbolygó, Ferenc & Valéria Csépe. 2011. Processing of stress related acoustic cues as indexed by ERPs. In *Twelfth Annual Conference of the International Speech Communication Association*.

Hunter, John D. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 9(3). 90–95.

Karpova, L. L. 2005. *Srednečepeckij dialekt udmurckogo jazyka [The Middle Cheptsa dialect of Udmurt]*. Izhevsk: The Udmurt Institute of History, Language and Litreature of the Ural Branch of the Russian Academy of Sciences.

Keating, Patricia A. 1996. The phonology-phonetics interface. *UCLA Working Papers in Phonetics*. UNIVERSITY OF CALIFORNIA 45–60.

Kelmakov, V. K. 1998. *Kratkij kurs udmurckoj dialektologii. Vvedenie. Fonetika. Morfologija. Dialektnye teksty. Bibliografija [A short course in Udmurt dialectology; introduction, phonetics, morphology, dialectal texts, bibliograpy]*. Izhevsk: Udmurt University Publishing.

Kirillova, Lyudmila E. (ed.). 2008. *Udmurtsko-russkij slovar*. Izhevsk: Udmurtskiy Institut istorii, yazyka i literatury Ural'skogo otdeleniya Rossiyskoy akademii nauk.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H.B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13). https://doi.org/doi: 10.18637/jss.v082.i13.

Lenth, Russel V. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans.

Lytkin, V. I. & T. I. Tepliashina. 1962. Fonetika. In P. N. Perevoshchikov (ed.), *Grammatika sovremennogo Udmurtskogo jazyka [A grammar of contemporary Udmurt]*, vol. I, 7–58. Izhevsk: Udmurtskoe knižnoe izdatelstvo.

Lytkin, Vasiliy I. (ed.). 1962. *Komi-permyackiy yazyk: Vvedenie, fonetika, leksika, morfologiya*. Kudymkar: Komi permyackoje knizhoje izdatelstvo.

Machač, Pavel & Radek Skarnitzl. 2009. *Principles of phonetic segmentation*. Praha: Epocha.

Pajusalu, Karl. 2022. Prosody. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*, 868–878. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0043.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Roettger, Timo & Matthew Gordon. 2017. Methodological issues in the study of word stress correlates. *Linguistics Vanguard* 3(1). https://doi.org/10.1515/lingvan-2017-0006.

Saarinen, Sirkka. 2022. Mari. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (eds.), *The Oxford Guide to the Uralic Languages*, 432–470. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0024.

Setälä, Eemil Nestor. 1901. Über Transskription der Finnisch-ugrischen Sprachen. *Finnisch-Ugrische Forschungen* 1. 15–52.

Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

Suomi, Kari, Juhani Toivanen & Riikka Ylitalo. 2001. On distinguishing stress and accent in Finnish. *Working papers/Lund University, Department of Linguistics and Phonetics* 49. 152–155.

Tarakanov, I. V. 1959. Ob udarenii v udmurtskom jazyke [On stress in Udmurt]. *Izvestija Akademii Nauk Estonskoj SSR. Serija obščestvennyx nauk.* 2. 170–177.

Tepliashina, T. I. 1970. *Jazyk besermian [The language of Besermans]*. Moscow: Nauka.

Vakhrushev, V. M. & V. N. Denisov. 1992. *Sovremennyj udmurtskij jazyk. Fonetika. Grafika i orfografija. Orfoépija [Contemporary Udmurt. Phonetics. Orthography. Pronunciation]*. Izhevsk: Udmurtija.

Vogel, Irene, Angeliki Athanasopoulou & Nadya Pincus. 2016. Prominence, contrast, and the functional load hypothesis: An acoustic investigation. In Jeffrey Heinz, Rob Goedemans & Harry G. van der Hulst (eds.), *Dimensions of phonological stress*, 123–167. Cambridge: Cambridge University Press.

Winkler, Eberhard. 2001. *Udmurt*. München-Newcastle: Lincom Europa.

Winkler, Eberhard. 2011. *Udmurtische Grammatik*. Wiesbaden: Harrasowitz.

Xu, Yi. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France.* 7–10.

Appendix 1

All items used in the first study are provided below, accompanied by English glosses. In morphologically segmentable items, morphemes are marked with a hyphen. The notation “(-)” indicates that a word is not fully transparent morphologically. This is the case for postpositions containing the illative case /-e/, and nouns formed with non-productive nominalizers /-ʎi/ and /-ri/; in these instances, only a lexical translation is provided. Stress is indicated here in order for presentational purposes, but it was not marked in the experimental materials (see Section 3.1).

low		mid		high	
disyllabic	**trisyllabic**	**disyllabic**	**trisyllabic**	**disyllabic**	**trisyllabic**
baˈka ‘frog’	kaʎaˈga ‘swede’	beˈr(-)e ‘after’	deʎeˈt-o ‘grace-ADJZ’	buˈsɨ ‘field’	bubɨ(-)ˈʎi ‘butterfly’
daˈga ‘horseshoe’	paʎaˈka ‘quail’	boˈko ‘monster’	zor-oˈno ‘rain-PTCP.FUT’	vɨˈʒɨ ‘root’	gudɨ(-)ˈri ‘thunder’
kaˈna ‘cupboard’	tamaˈʃa ‘funny’	veˈme ‘communal work’	kenoˈs-o ‘barn-ADJZ’	giˈʒɨ ‘nail’	d͡ʑugɨ(-)ˈri ‘(woven) skein’
maˈza ‘peace, free rein’	t͡ɕaraˈka ‘ball’	voˈʒo ‘Udmurt mythical creature’	keneˈm-o ‘hemp-ADJZ’	guˈbi ‘mushroom’	d͡ʑuzɨ(-)ˈri ‘icicle’
ɕaˈla ‘hazel grouse’	ʃabaˈla ‘mouldboard’	goˈn-o ‘fur-ADJZ’	keseˈg-o ‘piece-ADJZ’	gɨˈʒɨ ‘grain of sand’	kibɨ(-)ˈʎi ‘bug’
ɕaˈna ‘except (for)’	ʃakaˈʎa ‘trifle’	goˈp-o ‘hole-ADJZ’	ket͡ɕeˈto ‘double’	duˈrɨ ‘ladle’	kiʑi(-)ˈʎi ‘star’
taˈba ‘pan’	ʃaraˈka ‘wooden toy’	zoˈr-o ‘rain-ADJZ’	nomeˈr-o ‘number-ADJZ’	d͡ʒiˈbɨ ‘gusset’	kisɨ(-)ˈri ‘wrinkle’
taˈza ‘healthy’		keˈm(-)e ‘approximately’	poleˈso ‘multiple’	d͡ʒuˈbɨ ‘tab, eyelet’	kuʑɨ(-)ˈʎi ‘ant’
taˈka ‘ram’		keˈn-o ‘daughter.in.law-ADJZ’	pɘroˈs-o ‘talent-ADJZ’	d͡ʒɨˈbɨ ‘slot, notch’	niʑi(-)ˈʎi ‘earthworm’
caˈpa ‘sole’		koˈlo ‘ford’	sereˈg-o ‘corner-ADJZ’	d͡ʑiˈt͡ɕɨ ‘fox’	nugɨ(-)ˈʎi ‘home-made vermicelli’
caˈca ‘father’		koˈto ‘wet’	toleˈʑ-o ‘moon-ADJZ’	kuˈdɨ ‘basket’	pukɨ(-)ˈʎi ‘a fat person’
t͡ɕaˈt͡ɕa ‘toy’		koˈt͡ʃo ‘magpie’	ʃebeˈʎo ‘rolled out’	kɨˈdɨ ‘awn’	pɨgɨ(-)ˈʎi ‘lambswool’
ʃaˈra ‘aloud’		kɘˈt-o ‘belly-ADJZ’		lɨˈmɨ ‘snow’	sukɨ(-)ˈri ‘loaf’

ʃaˈt͡ɕa ‘rod’		ʎoˈgo ‘hill-ADJZ’		ʎuˈgɨ ‘burdock’	supɨ(-)ˈʎi ‘chatterbox’
		meˈɲ-o ‘mole-ADJZ’		muˈʎɨ ‘berry’	tɨgɨ(-)ˈʎi ‘round, circle’
		moˈko ‘monster’		muˈmɨ ‘mother, female’	t͡ɕutɨ(-)ˈri ‘ornate, curled’
		moˈso ‘fractional’		nuˈnɨ ‘baby’	ʃɨgɨ(-)ˈri ‘tray, trough’
		neˈne ‘mother’		piˈt͡ɕi ‘small’	ʃɨmɨ(-)ˈri ‘frill, gather’
		peˈʎ-o ‘ear-ADJZ’		puˈʒɨ ‘pattern’	
		peˈɲ-o ‘ash-ADJZ’		puˈnɨ ‘dog’	
		tɘˈl-o ‘wind-ADJZ’		puˈɲɨ ‘spoon’	
		tɘˈro ‘head(man), elder’		suˈzɨ ‘grus’	
		teˈʎ-o ‘forest-ADJZ’		tuˈzi ‘pretty, fashionable’	
		t͡ɕoˈte ‘(into) account’		tuˈri ‘crane’	
		t͡ʃoˈʒ(-)e ‘during’		tɨˈpɨ ‘oak’	
		ʃeˈp-o ‘ear/spike-ADJZ’		t͡ɕiˈɲɨ ‘finger’	
		ʃoˈko ‘braggart’		t͡ɕiˈpɨ ‘chick’	
		ʃɘˈt͡ɕe ‘hog, pig’		t͡ɕuˈɲɨ ‘foal’	
				ʃuˈʑi ‘crazy’	
				ʃuˈkɨ ‘foam’	

Appendix 2

All items used in the second study are provided below, accompanied by English glosses. In morphologically segmentable items, morphemes are marked with a hyphen. The relevant suffixes are verbalizers /-(j)a/ and /-om/, the frequentative marker /-l/, as well as the present tense third person singular marker /-e/ and the second person plural imperative marker /-e/ (both found in Conjugation I verbs). The notation "(-)" indicates that the combination of the stem and the verbalizer is not fully transparent morphologically; in these cases, only the translation of the verb is given. We list both the indicative and the imperative verbs that form minimal pairs, together with their glosses. Stress is marked in the table for presentational purposes; it was not indicated in the experimental materials (see Section 3.1).

low		**mid**		**high(+mid)**	
disyllabic	**trisyllabic**	**disyllabic**	**trisyllabic**	**disyllabic**	**trisyllabic**
vaˈl(-)a 'understand[PRS.3SG]' ˈval(-)a 'understand[IMP.2SG]'	daga-ˈja 'horseshoe-VBZ[PRS.3SG]' ˈdaga-ja 'horseshoe-VBZ[IMP.2SG]'	voˈʑ-e 'hold-PRS.3SG' ˈvoʑ-e 'hold-IMP.2PL'	ber-oˈm-e 'late-VBZ-PRS.3SG' ˈber-om-e 'late-VBZ-IMP.2PL'	biˈɲ-e 'wind-PRS.3SG' ˈbiɲ-e 'wind-IMP.2PL'	budɨ-ˈl-e 'grow-FREQ-PRS.3SG' ˈbudɨ-l-e 'grow-FREQ-IMP.2PL'
gaˈʒ(-)a 'respect[PRS.3SG]' ˈgaʒ(-)a 'respect[IMP.2SG]'	kabaˈn-a 'hayrick-VBZ[PRS.3SG]' ˈkaban-a 'hayrick-VBZ[IMP.2SG]'	koˈʒ-e 'turn-PRS.3SG' ˈkoʒ-e 'turn-IMP.2PL'	d͡ʒog-oˈm-e 'fast-VBZ-PRS.3SG' ˈd͡ʒog-om-e 'fast-VBZ-IMP.2PL'	buˈd-e 'grow-PRS.3SG' ˈbud-e 'grow-IMP.2PL'	bɨʑɨ-ˈl-e 'run-FREQ-PRS.3SG' ˈbɨʑɨ-l-e 'run-FREQ-IMP.2PL'
ʒaˈʎ-a 'pity-VBZ[PRS.3SG]' ˈʒaʎ-a 'pity-VBZ[IMP.2SG]'	ɕala-ˈja 'hazel.grouse-VBZ[PRS.3SG]' ˈɕala-ja 'hazel.grouse-VBZ[IMP.2SG]'	ləˈd-e 'flog-PRS.3SG' ˈləd-e 'flog-IMP.2PL'	zək-oˈm-e 'thick-VBZ-PRS.3SG' ˈzək-om-e 'thick-VBZ-IMP.2PL'	bɨˈʑ-e 'run-PRS.3SG' ˈbɨʑ-e 'run-IMP.2PL'	gudɨ-ˈl-e 'dig-FREQ-PRS.3SG' ˈgudɨ-l-e 'dig-FREQ-IMP.2PL'
paˈʑ(-)a 'sprinkle[PRS.3SG]' ˈpaʑ(-)a 'sprinkle[IMP.2SG]'	tamaˈk-a 'tobacco-VBZ[PRS.3SG]' ˈtamak-a 'tobacco-VBZ[IMP.2SG]'	meˈd-e 'intend-PRS.3SG' ˈmed-e 'intend-IMP.2PL'	keɲeˈʃ-e 'advise-PRS.3SG' ˈkeɲeʃ-e 'advise-IMP.2PL'	vɨˈd-e 'lie-PRS.3SG' ˈvɨd-e 'lie-IMP.2PL'	vɨdɨ-ˈl-e 'lie-FREQ-PRS.3SG' ˈvɨdɨ-l-e 'lie-FREQ-IMP.2PL'
paˈl-a 'peel-VBZ[PRS.3SG]' ˈpal-a 'peel-VBZ[IMP.2SG]'	ʃara-ˈja 'aloud-VBZ[PRS.3SG]' ˈʃara-ja 'aloud-VBZ[IMP.2SG]'	nəˈd-e 'soil-PRS.3SG' ˈnəd-e 'soil-IMP.2PL'	kereˈt-e 'quarrel-PRS.3SG' ˈkeret-e 'quarrel-IMP.2PL'	guˈd-e 'dig-PRS.3SG' ˈgud-e 'dig-IMP.2PL'	dɨgɨ-ˈl-e 'hit-FREQ-PRS.3SG' ˈdɨgɨ-l-e 'hit-FREQ-IMP.2PL'
paˈɕ-a 'hole-VBZ[PRS.3SG]' ˈpaɕ-a 'hole-VBZ[IMP.2SG]'		poˈt-e 'go.out-PRS.3SG' ˈpot-e 'go.out-IMP.2PL'		gɨˈr-e 'plough-PRS.3SG' ˈgɨr-e 'plough-IMP.2PL'	zibɨ-ˈl-e 'put.pressure-FREQ-PRS.3SG' ˈzibɨ-l-e 'put.pressure-FREQ-IMP.2PL'

ɕaˈl-a ‘saliva-VBZ[PRS.3SG]’ ˈɕal-a ‘saliva-VBZ[IMP.2SG]’		toˈd-e ‘know-PRS.3SG’ ˈtod-e ‘know-IMP.2PL’		duˈm-e ‘tie-PRS.3SG’ ˈdum-e ‘tie-IMP.2PL’	
taˈl(-)a ‘take.away[PRS.3SG]’ ˈtal(-)a ‘take.away[IMP.2SG]’		t͡ʃoˈg-e ‘chop.off-PRS.3SG’ ˈt͡ʃog-e ‘chop.off-IMP.2PL’		dɨˈg-e ‘hit-PRS.3SG’ ˈdɨg-e ‘hit-IMP.2PL’	
t͡ɕaˈg-a ‘kindling-VBZ[PRS.3SG]’ ˈt͡ɕag-a ‘kindling-VBZ[IMP.2SG]’		ʃɘˈd-e ‘feel-PRS.3SG’ ˈʃɘd-e ‘feel-IMP.2PL’		ziˈb-e ‘put.pressure-PRS.3SG’ ˈzib-e ‘put.pressure-IMP.2PL’	