Title
Stress patterns and acoustic correlates of stress in Balti Tibetan

Journal
Himalayan Linguistics, 15(2)

Author
Caplow, Nancy

Publication Date
2016

DOI
10.5070/H915232601

A free refereed web journal and archive devoted to the study of the languages of the Himalayas

Himalayan Linguistics

Stress patterns and acoustic correlates of stress in Balti Tibetan

Nancy J. Caplow

Oklahoma State University

ABSTRACT

In Balti Tibetan, spoken in Baltistan, northern Pakistan, disyllabic non-verbs (nouns, adjectives, and numerals) are stressed on the second syllable (σ2). Fundamental frequency is a robust correlate of this σ2 stress pattern; vowel duration is a weak and inconsistent cue for stress, while intensity does not play a role. Verbs, in contrast, are stressed on the first syllable (σ1); F0, intensity, and vowel duration all contribute to conveying syllable prominence. These findings differ from previous descriptions of Balti in distinguishing stress patterns by lexical category. Further, this is the first work to provide an acoustic characterization of the correlates of stress in Tibetan. As one of the most phonologically conservative varieties of Tibetan, Balti can be considered to preserve the prosodic and acoustic characteristics of Proto-Tibetan. This study thus offers crucial information towards reconstructions of Proto-Tibetan and Proto-Tibeto-Burman, and towards development of hypotheses about Tibetan tonogenesis.

KEYWORDS

Tibetan, Balti, stress

This is a contribution from *Himalayan Linguistics, Vol. 15(2): 1–49*.
ISSN 1544-7502

Himalayan Linguistics, Vol. 15(2). © Himalayan Linguistics 2016
ISSN 1544-7502

Stress patterns and acoustic correlates of stress in Balti Tibetan[1]

Nancy J. Caplow
Oklahoma State University

1 Introduction

Geographically, Balti Tibetan is the western-most language of the entire Tibeto-Burman language family. It is spoken by a reported 347,000 people – roughly 327,000 in Baltistan in northeast Pakistan, and the remainder in Jammu and Kashmir in northwest India, and elsewhere (Lewis et al. 2016). Based on geographic and linguistic criteria, Bielmeier et al. (2008) recognize at least 141 distinct varieties of Tibetan. They classify these into seven main dialect groups, and identify Balti as one of the "Western Archaic dialects". Tournadre (2005, 2008, 2014) has identified more than 200 dialects of the language. He suggests (2014) that these fall into nearly 61 more-or-less mutually unintelligible dialect groups,[2] some consisting of only one variety and others with considerable internal dialectal diversity. These 61 Tibetic languages can in turn be grouped into eight major "geolinguistic continua", or sections. Balti falls within the "Northwestern section".

Phonologically, Tibetan has been traditionally divided into two broad categories (Jaeschke 1871; Róna-Tas 1966). The "Archaic" dialects – spoken at the western and eastern extremes of the Tibetan language area – are non-tonal and preserve the onset and coda consonants and consonant clusters of Old Tibetan. The "Innovative" dialects – spoken in the central and southern regions – make contrastive use of tone, and have simplified onsets and codas. Balti is one of the most conservative of the modern spoken Archaic varieties of Tibetan, retaining features of the language which may even pre-date those preserved in the writing system developed in the 8th century.[3]

---

[1] The Balti data presented here was contributed through elicitation sessions I conducted with Raza Tassawar and Mohammed (Baltistan), Senge Hasnan Sering (Seattle), and Hassan Ibrahim, Fatima Ibrahim, Hamida Ali, and Fatima Ilyas (Rockville, MD). I am grateful for their patience, insights, and careful explanations. Nicolas Tournadre kindly shared recordings he made in Skardu with Senge Hasnan Sering and Muhammad Abbas Kazmi. The collection of Balti data in Rockville, Maryland in 2013 was supported by a travel grant from the English Department at Oklahoma State University. My thanks also to Micol Martinelli who created some of the textgrids used in the analysis, and to two anonymous reviewers for their comments.

[2] Each dialect group is equivalent to what is typically identified politically as a language. For instance, while speakers of the Quebec and Parisian dialects of the French language (dialect group) can understand one another, from the perspective of a linguist, they differ sufficiently in their lexicon and morphosyntax to warrant distinct grammatical descriptions. (Tournadre p.c., June 2016).

[3] For example, note the [t] preserved in [bu.'striŋ] *bu.sring* 'woman, wife', and the initial dorsal fricative in [ʁzor.'ba] *zor.ba* 'sickle' (which also appears in Rebkong Amdo as [xso.'ra]).

Written Tibetan: WT forms were drawn from a number of sources (which are not always in complete accord), including: Bielmeier et al. (2002a, b), Goldstein and Narkyid (1984), Goldstein (2001), Jaeschke (1958 [1881]), Norberg-Hodge and Paldan (1991), and Sprigg (2002). I follow the Extended Wylie Transliteration Scheme developed by the University

Given its geographic location and its phonological conservatism, documentation of the lexical prosody and the acoustic characteristics of Balti offers information which may be crucial in reconstructing Proto-Tibetan and Proto-Tibeto-Burman, and in developing hypotheses about Tibetan tonogenesis.

Previous descriptions of the prosodic characteristics of Balti are summarized in section 2 below. The methodology of the present study is described in section 3. Perceived stress patterns based on disyllabic words produced by nine Balti speakers are described in section 4; measurements of the acoustic correlates of stress from two of these speakers are presented in section 5. Section 6 provides a summary of the findings of this study, and a note on their theoretical significance.

2 Previous Descriptions of Stress in Balti

Previous studies of Balti include works by Grierson (1909),[4] Read (1934), Sprigg (1966, 2002), Bielmeier (1985, 1988), and Backstrom (1994) among others. My discussion here primarily addresses the more detailed analyses by Sprigg.

Grierson (1909: 35) remarks that Balti is non-tonal, but makes no mention of stress. Read (1934: 3) suggests that all syllables of a word receive "equal emphasis"; from his example, however, I think he may have been referring to vowel quality, rather than stress. Backstrom (1994: 5) notes that Balti, like other Western Tibetan varieties, is non-tonal, and simply reports the descriptions of lexical prosody offered by Sprigg (1966) and Bielmeier (1985); stress is not marked in his phonetic transcriptions.

Sprigg initially (1966) analysed pitch patterns observed in Balti as representing tone, rather than stress (though he later reversed his view, as discussed below). First, he observed a "majority" pitch pattern on disyllabic nouns:

- σ1 L, level; σ2 H, level or falling
- [_ ¯] or [_ \ ]

In addressing the question of whether this constituted tone or stress, Sprigg also considered "breath force" (by which he likely meant intensity). He assumed that, in a stress system, breath force should consistently align with prominent pitch. Finding that this was not the case, he concluded that lexical prosody in Balti should be characterized as tone. From here, Sprigg then considered a "minority" pitch pattern observed on borrowed words and exceptions (total n=84):

- σ1 H pitch, level or falling; σ2 level, H or L
- [¯ ¯] or [\ _]

As he was already leaning toward a tonal interpretation, Sprigg considered this to be in opposition to the majority pitch pattern, and construed it as supporting the view that tone was contrastive.

---

of Virginia's Tibetan and Himalayan Library (www.thlib.org), based on Wylie (1957), except that I use a period "." rather than a blank space to represent boundaries between syllables; this leaves the blank space available to mark word boundaries. A question mark indicates spellings I am not certain of; "n/a" indicates I could not identify an appropriate WT form.

Phonetic transcription: Elicited examples were transcribed according to the 2015 IPA charts, except that I use [y] rather than [j] for the palatal approximant, and [ny] rather than [ɲ] for the palatal nasal. These modifications are made to be consistent with the WT transliteration.

[4] Grierson was the editor of the Linguistic Survey. In the "Introductory Note" he states that it was his assistant, Sten Konow, who prepared the work, based on materials analysed in part by "Professor Conrady of Leipzig", referring to the German linguist and sinologist August Conrady.

There are several weaknesses in Sprigg's initial analysis. First, it is not necessary, in a stress system, for F0 (pitch) to align with intensity ("breath force"). Cross-linguistically, relative syllable prominence can be conveyed by F0, intensity, vowel duration, and/or vowel quality. Not all languages draw upon these acoustic resources to express stress in the same way, and even within a given language, not all speakers utilize these resources in the same way (as I demonstrate in section 5). Second, the argument that a pitch pattern observed on a limited number of borrowed words and exceptions constitutes a contrastive tone is rather weak (which Sprigg himself acknowledged, p. 200). Third, and most significantly, as Sprigg did not have acoustic evidence available to him, he was not led to consider the intrinsic effects of vowel quality on intensity, and this must surely have played a role in his perception of "breath force".

As reported by Lehiste (1970: 120), intensity is intrinsically higher on low vowels, and lower on high vowels. This is illustrated in Figure 1 below for examples from my own recordings: in [smɪn.ˈma] *smin.ma* 'eyebrows', where [σ1.σ2] exhibits a [high.low] vowel sequence, intensity is higher on the low vowel in σ2 (increasing from 64 dB to 68 dB); in [ndaʁ.ˈʑu] *mda'.gzhu* 'arrow and bow pair', with a [low.high] vowel sequence, intensity is higher on the low vowel in σ1 (decreasing from 70 dB to 64 dB).

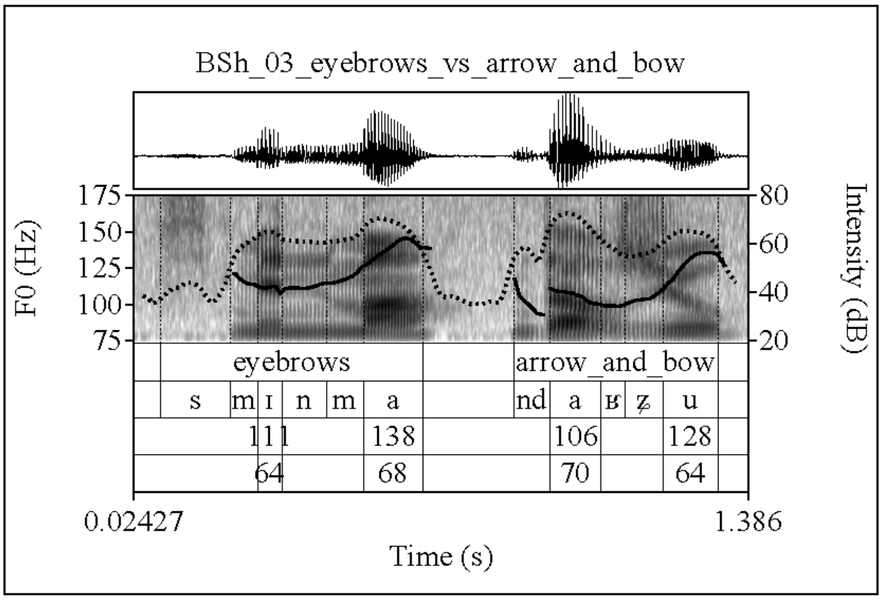


Figure 1. Intensity vs. vowel height (Solid line: F0; dotted line: intensity. Speaker BSh_03)

When vowel height is the same in both syllables, there is no intrinsic contrast in intensity. In Figure 2 below, intensity is about the same in the [low.low] vowels in [gar.ˈba] *mgar.ba* 'blacksmith' (66 dB *vs.* 65 dB) and the [mid.mid] vowels in [lo.ˈskor̥] *lo.skor* '12-year cycle of years' (68 dB *vs.* 69 dB).

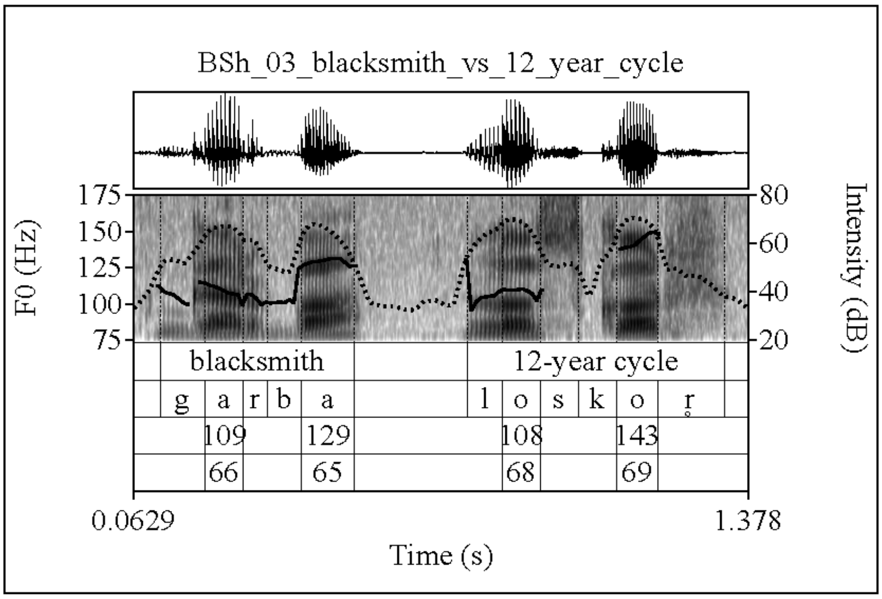


Figure 2. Controlled vowel height (Solid line: F0; dotted line: intensity. Speaker BSh_03)

While not all tokens exhibit such a tidy correlation between vowel height and intrinsic intensity, the point is that this factor was not taken into consideration in Sprigg's description of "breath force". Without controlling for contrasts in vowel quality, a potential correspondence between intensity and stress must have been obscured in some words and exaggerated in others, contributing to Sprigg's rejection of stress as an explanation for the prosodic patterns he perceived.[5]

Intrinsic intensity also comes into play in the later work of Sprigg (2002: 4-5) and Bielmeier (1985, 1988: 47), who both claim that stress falls on the final syllable of a word regardless of lexical category (and regardless of word length). This is indeed consistent with the σ2 stress I describe here for disyllabic non-verbs. To account for the σ1 stress observed on disyllabic verbs, they suggest that the *-pa*/*-ma* which occurs on citation forms / infinitives / verbal nouns (e.g., *rdung.ma* 'hitting; to hit') belongs to a limited set of non-stress-bearing morphemes which cause stress to shift to the preceding syllable (e.g., Sprigg 2002: 4-5).

However, I find this analysis unconvincing, as discussed in detail in Caplow 2009 (48-51, 99-106). First, the majority of Sprigg's twenty-one 'non-stress-bearing suffixes' are clitics – oblique case markers, clause connectors, quantifiers, and participial endings – which I agree do not bear stress.

Next, having identified these clitics as non-stress-bearing, we can imagine that Sprigg was readily inclined to interpret suffixes in the same way – which I do not agree with. This is indeed his analysis of *-po* and *-mo* on a subset of nouns and adjectives. However, all but one of Sprigg's illustrative examples here have a [σ1.σ2] sequence of [low.high] or [low.mid] vowels, as in [dak.mo] *dag.mo* 'housekeeper, person in charge of all food, etc.' and [lyaks.mo] *legs.po* 'good, clean, well, nice'. This contrast in vowel height would yield an intrinsically higher intensity and longer duration on σ1 – and the resultant perceived prominence of σ1 has nothing to do with the phonological status of

[5] The intrinsic correlation between vowel height and vowel duration may also have influenced Sprigg's perception of relative syllable prominence. Lehiste (1970: ff 18) summarizes several studies demonstrating that low vowels tend to be longer than high vowels.

the σ2 suffix. Thus what Sprigg and Bielmeier identify as a shift of stress leftward from a non-stress-bearing suffix can be explained in terms of a contrast in vowel height.

Finally, having proposed the existence of a set of non-stress-bearing suffixes – the evidence for which is debatable, as I have demonstrated above – Sprigg analogously included in this set the *-pa*/*-ma* suffix which is added to verb stems to produce citation forms / infinitives / verbal nouns. This is one way to account for the σ1 stress which is so consistently observed on disyllabic verbs.

Here, though, I offer an alternative analysis: I consider the σ1 stress on verbs to be contrastive, and specific to the lexical category. Note that this is not fundamentally different than the tacitly-accepted, purely descriptive view that the *-pa* in the noun [tʰaq.'pa] *thag.pa* 'rope' is stressed, while the *-pa* in the verb ['tʰaq.pa] *'thag.pa* 'to grind' is unstressed. However, it has the advantage of providing a unified account for stress in verbs of various structures. In addition to infinitives ending in *-pa*, σ1 is also stressed in infinitives ending in *-bya,* such as ['sna.bya] *rna.bya* 'to listen'. These forms, too, were provided in response to verbs elicited with English or Urdu infinitives, and were clearly translated back into English infinitives during discussion. [6] Stress also falls on σ1 in expressions such as ['lut.taŋ] *lud.btang* '[to] spread manure' and ['zan.ɸtɕo] *zan.bco* '[to] cook food'. In the former, which I call a "noun + verbalizer construction" (following Bartee 2000), the verb (here 'to send') is semantically light, can be used with a wide variety of nouns to express different actions, and draws its meaning from the specific collocation. In the latter, a "noun-complement + verb structure", the verb is semantically richer and used more narrowly.[7] Imperatives are also stressed on σ1, as in ['zɡo.tɕukʰ] *sgo.bcug* 'close the door! (door+close).[8] In these structures, the morphemes in σ2 are richer in phonetic and semantic content than Sprigg's "non-stress-bearing suffixes", and are not counted among them. Thus σ1 stress cannot readily be explained as a shift leftward from σ2. Vowel quality does not play a role either; the data (below) shows that σ1 stress in verbs is robustly conveyed by both F0 and intensity, eclipsing contrasts in vowel height.

In summary, then: First, in the cases where Sprigg and Bielmeier see non-stress-bearing suffixes on nouns and adjectives, I see a contrast in vowel height, whose intrinsic acoustic properties account for the stress pattern they described. Second, where Sprigg and Bielmeier see a non-stress-bearing suffix on verbs, I see a different stress pattern altogether, by taking into consideration verbs of a different morphological composition. These claims are supported by the description and acoustic characterization of stress which follows in sections 4 and 5 below.

A contrast between σ1 stress on verbs and σ2 stress on non-verbs has been reported for several other varieties of Tibetan. As described by Zemp (2013: 51) this is the case in Purik, a Western Tibetan dialect spoken in Ladakh. He provides several minimal pairs which contrast only in terms of stress (similar to those I offer here in section 4.2.1). For Zhongu Tibetan, spoken in the Aba

[6] *-bya* may have originated from the verb 'to do' ['bi.(y)a] *bya* (perhaps from *bya.ba* > *bya:* > *bya),* grammaticized as an infinitive / verbal noun / citation form suffix. Zemp (2013) describes both *-pa* and *-bya* infinitives for Purik, and Norberg-Hodge and Paldan (1991) present infinitive verbs with [tʃes] *byas* for Ladakhi.

[7] These examples were elicited with bare stems (rather than infinitives), but it was clear from discussion that they were not imperatives. It is possible that speaker responses are clipped forms of trisyllabic verbs (such as ['xlu.ˌtã.ma] *glu.btang.ma* 'to sing' [song+send]), presented as a bare stem without the final *-pa / -ma* (as discussed in Caplow 2016*).*

[8] Imperative forms were elicited with imperatives; their form and meaning were confirmed through discussion and by translation back into English.

(Ngaba) Tibetan Autonomous Prefecture in Sichuan Province,[9] Sun (2003) distinguishes between a σ2 stress pattern in disyllabic nouns, and a σ1 stress pattern in disyllabic verb complexes. Neither of these descriptions include a detailed phonetic characterization.

3 Methodology

The methods used in data collection and data analysis are described in sections 3.1 and 3.2 below.

3.1 Data Collection

This study is based on disyllabic words elicited from nine speakers of Balti (six males and three females, ranging in age from 30 to 60) between 2002 and 2013. Analysis of the acoustic correlates of stress is based on data from the two speakers from whom I obtained the most extensive list of lexical items and the best quality audio recordings.

3.1.1 Word Lists

Disyllabic words are highly favored in Tibetan, and are the focus of the present study. As noted by Denwood (1999: 87-88) monosyllabic words generally denote common referents, such as people, body parts, nature and the elements, agriculture, and household objects. Trisyllabic and quadrisyllabic words also occur, but are most often composed of monosyllabic and disyllabic elements joined together, with clipping as necessary.

The lexical items discussed here are drawn from several elicitation lists: (a) my own list of ~500 words; (b) Nicolas Tournadre's 650-item list of 'pan-dialectal' vocabulary items; (c) lists developed by Jackson T.-S. Sun to investigate correspondences between WT spelling and pronunciation across dialects; and (d) portions of a list which was developed by Roland Bielmeier for use by the researchers involved in compilation of the Comparative Dictionary of Tibetan Dialects (CDTD; Bielmeier et al. 2002a, b). Finally, of course, the consultants I worked with contributed other words to illustrate points of interest.

My description here focuses on the disyllabic words included in these sources. Most of these lists included more non-verbs (nouns, adjectives, and numerals) than verbs.

3.1.2 Speakers

The two speakers whose recordings I analyzed acoustically were BM_01 and BSh_03. Speaker BM_01 was recorded in the village of Machulo, in Baltistan, northern Pakistan, in August 2003. He was 50 years old at the time, had little formal education, and was not literate. He was also fluent in Urdu, and had contact with languages other than Balti through his work with foreigners as a trekking guide. I recorded Speaker BSh_03 in Seattle, WA in September 2006, during the 39th International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL). He was born in the Shigar area of Baltistan; his mother tongue is the variety of Balti spoken there. When he was a boy, his father found work in Islamabad, so he lived and attended school there, but spent several months each year back in his native village. After completing a university degree, he returned to

[9] More specifically: in Hóngtŭ and Hóngzhá Townships, and in the villages of Píng-ān, Gūnà, and Luóbŭchōng in Xiăoxìng Township in the Zhongu (Tibetan: *zho.ngu.khog*; Chinese: Rèwùgōu) valley of Sōngpān County (Sun 2003: 770).

Baltistan for several years to work in the promotion of Balti language and culture. He is fluent and literate in Urdu and English. At the time I recorded him, he was in his 30s and had already been living in the U.S. for several years. He had recently begun to study the Tibetan writing system.[10]

In addition to acoustic analysis, my study of Balti is based on qualitative, perceptual observations of stress patterns in other elicitation sessions and recordings, working with seven other speakers. My very first opportunity to hear and transcribe Balti was at an elicitation session conducted by Jackson T.-S. Sun during the 8th Himalayan Languages Symposium (HLS) held in Berne, Switzerland in 2002. The speaker was a well-educated scholar from Skardu, a middle-aged male, who also spoke Urdu and English. In 2003 I recorded a male speaker in Baltistan, in his native town of Skardu; he was in his 30s, well-educated, and fluent in English as well as Urdu. In 2013 I recorded a shorter, targeted list of words from members of a family currently residing in Rockville, MD, including a couple in their 50s and two of their daughters-in-law, both in their 30s. The older husband and wife, their three sons, and one of their daughters-in-law were all born in the village of Surmo, in eastern Baltistan. When they were young, the sons moved to Karachi to study, and married Balti-speaking women.[11] The extended family moved to the U.S. around 1999. They all speak English and Urdu in addition to Balti; the six small grandchildren speak English and Urdu, but have not been learning Balti.

My analysis is also based on stress patterns perceived in two recordings which Nicolas Tournadre made in Skardu in 2003 of his 650-item word list. In one, the speaker was the same BSh_03 mentioned above. The other speaker was the same scholar recorded in Berne, Switzerland.

3.1.3 Recording Procedure

Most recordings were made using a solid state recorder (or, in earlier years, a mini-disk recorder) with a head-worn microphone.[12] In a few situations recording equipment was not available, and I simply transcribed the forms I heard, attending carefully to stress.

During sessions I conducted myself, I elicited words both in isolation and in a sentence frame which placed the target word in utterance-medial position, such as [ˈbal.ti ˈskat.diŋ X za.ret] "In [the] Balti language, we X say". On several occasions I attended (and usually recorded) elicitation sessions conducted by other scholars in which target words were produced only in isolation.

3.1.4 Acoustic Measurements

The Praat phonetics software (Boersma 2001, Boersma and Weenink 2016) was used to measure three potential acoustic correlates of stress: fundamental frequency, intensity, and vowel duration. As a matter of standardization, textgrid interval boundaries were drawn at the nadir of the trough of a completed waveform, closest to the point where waveform, F2, and spectrogram cues were in alignment. Vowel onset and offset were identified based on changes in the complexity of the waveform, as well as formant and spectrogram displays.

---

[10] He did not learn the Tibetan script during his early formal education, as – until quite recently – Balti was written using only the Urdu script, which is not ideally suited to the task.

[11] There is a large Balti-speaking community in Karachi. According to my consultants, the Balti population there exceeds 150,000 people.

[12] The solid-state recorders used were a Marantz PMD-660II and an Edirol R-09; the minidisk recorder was a Sharp MD-722. For most recordings I used a Shure SM10A monaural head-worn cardioid condenser microphone with a Shure A96F line-matching transformer.

These cues do not provide a consistent means of defining the end of a vowel in utterance-final position (e.g., open syllables in isolation forms). Vowels in this phonetic environment display considerable variation in the fading out of waveforms, voice bars, and formant structure, and may exhibit aspiration or creaky phonation. In these cases, vowel offset was identified as the point at which intensity had declined to 10 dB less than the vowel's peak intensity (about $^7/_8$ of the peak intensity).[13]

Vowel duration was measured from the onset to the end, as defined above. For fundamental frequency and intensity, a Praat script was used to measure the average value over the middle 50% of the nuclear vowel in each syllable, which controlled for the influence of preceding or following consonants.

Acoustic measurements were collected from recordings of speakers BSh_03 (from Shigar) and BM_01 (from Machulo). With speaker BSh_03, I recorded approximately 150 disyllabic and monosyllabic words, each produced twice in isolation and twice in the sentence frame. Only the second token of each was selected for analysis. With speaker BM_01, I recorded a lengthy elicitation session conducted by Nicolas Tournadre of his 650-item lexicon. Target words were produced only in isolation. Some words were produced two or more times, while some were produced just once; again, only one token was analyzed acoustically. This speaker was older, less educated, and less experienced, and words were sometimes mumbled, drawled, or called out enthusiastically; consequently, acoustic features exhibit a broader range in values, and patterns are a bit more diffuse. Despite these differences, Balti's patterns of σ2 stress for nouns, adjectives, and numerals, and σ1 stress for verbs are clearly manifested.

Not all of the disyllabic words recorded from the two speakers were suitable for acoustic analysis, given their intrinsic phonetic composition. Segmental boundaries cannot be accurately determined in words with glides or approximants (e.g., nouns like [mɨ.'yul] *mi.yul* 'world' and verbs like ['xku.wa] *rku.ba* 'to steal'). Since the span of the nuclear vowels could not be defined, there was no way to collect meaningful measurements of the vowel's duration, average F0 over the medial 50% of the vowel, etc. It was also not possible to collect measurements from words with very short or highly fricated vowels, or with sonorant syllabic nuclei. Though dozens of words thus had to be excluded from acoustic analysis, all of them were still considered in evaluating perceptual stress patterns.

3.2 Data Analysis

Since all of the words considered in this study are disyllabic, F0, intensity, and vowel duration can only be greater on σ1, greater on σ2, or about the same on both syllables. Paired-sample t-tests are used to calculate the difference in corresponding measurements from σ1 and σ2 of each individual word, and then determine whether the mean difference for the entire sample of words defines a statistically significant pattern. A two-way analysis of variance (ANOVA) is used when appropriate to take into consideration the intrinsic effects of syllable structure and vowel height. Graphical representation is also used to directly compare values across syllables.

Statistical significance is a necessary but not sufficient condition to establish that an acoustic feature is being utilized by speakers to convey syllable prominence. A contrast across syllables can be

[13] This heuristic was suggested to me by Ian Maddieson (p.c., 2004). In practice, this solution turned out to select an endpoint which was generally consistent with the signature of the waveform, spectrogram, formant, and/or F0 and intensity traces. When tested on words with a closed final syllable, it usually identified a vowel endpoint that coincided with the actual nucleus-coda boundary, and thus seems to be a reliable approach.

statistically significant but not perceptually significant; this is the case for intensity in all adjectives produced by speaker BM_01 (see section 5.3.2). In this analysis, I control for and rule out incidental factors which might influence the parameters of interest. Both F0 and intensity are intrinsically affected by vowel height: F0 is high on high vowels and low on low vowels, while intensity is high on low vowels and low on high vowels. Vowel duration is affected by syllable closure and vowel height (as well as compensatory lengthening, and position in the utterance).

Threshold values used in evaluating the perceptual significance of contrasts in F0, intensity, and vowel duration across syllables are shown in Table 1 below. For F0, Lehiste (1970: 62-67) and Laver (1994: 451) cite a number of experimental studies which identify the just-noticeable difference (the difference limen) as ± 1 Hz (for frequencies within the range of a typical male voice – i.e., 80 Hz to 160 Hz). Lehiste (1970: 79-80) also refers to experimental studies in which listeners were able to discriminate tones in synthesized Thai words when the difference in fundamental frequency of the tones was on the order of ± 5 Hz (again, for frequency ranges typical of the male voice). Extending this finding from Thai to Tibetan, I consider a pitch difference across syllables which is statistically significant and is ~ 5 Hz to be perceptually significant. However, I regard it as perhaps a "weak" and "unreliable" cue for stress, since a mean pitch difference of 5 Hz entails that the sample includes a number of words with a pitch difference < 5 Hz, which is not perceptually significant. When a statistically significant mean pitch difference across syllables is ~ 10 Hz, I consider it a clear correlate of stress; a difference of ~ 20 Hz is a "strong" or "robust" correlate; a difference of ~ 40 Hz is a "dramatic" correlate. Based on similar considerations, I established discriminable differences and weak and meaningful correlates of stress for intensity (drawing on Lehiste p. 121 and Laver p. 502) and vowel duration (drawing on Lehiste pp. 10-13, 33-34). Table 1 below summarizes the levels of perceptual significance which are relevant to evaluation of the acoustic parameters considered in this study.

Acoustic parameter	Discriminable difference	Weak difference	Meaningful difference
F0	1 Hz	5 Hz	10 Hz
Intensity	1 dB	2 dB	5 dB
V Duration	10-30 msec	20 msec	30 msec

Table 1. Perceptual significance levels

Finally, acoustic correlates are evaluated in terms of the context in which they occur. A contrast across syllables might be neither statistically significant nor perceptually significant, but still be contextually significant. This is the case for vowel duration in verbs produced by speaker BM_01 (see section 5.4.2). Here, several aspects of the phonetic environment favor longer vowels in σ2, yet there is a tendency for longer vowels to occur in σ1.

4 Results: Observed Stress Patterns

Perceived stress patterns based on the hundreds of tokens elicited from all nine speakers are clear and consistent. Disyllabic non-verbs (nouns, adjectives, and numerals) are stressed on the second syllable (σ2), and disyllabic verbs are stressed on the first syllable (σ1). Details are presented below; additional examples can be found in Caplow 2009 and Caplow 2016.

4.1 Stress Patterns on Disyllabic Non-Verbs

The σ2 stress perceived on nouns, adjectives, and numerals pertains regardless of morphological composition, as illustrated below. Exceptions are presented in section 4.1.6.

4.1.1 Stem + Formative Suffix

Examples of disyllabic nouns and adjectives ending in a variant of *-pa*, *-po*, *-ma*, or *-mo* are provided in Table 2 below. This pleonastic "formative" suffix[14] contributes a second syllable to the word and sometimes conveys gender. It is clipped in the formation of compounds, but is generally present in citation forms, and when a noun or adjective stands as in independent word in an utterance. Stress falls on this σ2 suffix even though it is an open, light syllable and contributes no meaning or function.

Lex cat	*Gloss*	*Transcription*	*WT*
Nouns	smoke	tut.ˈpa	*dud.pa*
	bellows	zbuk.ˈpa	*sbug.pa*
	dust	tʰal.ˈba	*thal.ba*
	lungs	χlo.ˈa	*glo.ba*
	donkey	ɓɔ̃ŋ.ˈbu	*bong.bu*
	wool felt	pʰɯŋ.ˈma	*phying.'ma*
	mare	rɡud.ˈmo	*rgod.mo*
	goddess	ɬa.ˈŋo	*lha.mo*
Adjs	new	sar.ˈpa	*gsar.pa*
	wet	qχɕer.ˈfa	*gsher.ba*
	clear	xsal.ˈpo	*gsal.po*
	empty	stoχ.ˈma	*stong.ma*
	sweet	ŋar.ˈmo	*mngar.mo*
	easy	ᵊvdo.ˈŋo	*bde.mo*
	thin	stra.ˈŋo	*sra.mo*

Table 2. σ2 stress in stem + formative nouns and adjectives

4.1.2 Compounds

Compounding is a highly productive means of forming disyllabic words, which are preferred in Tibetan. Compounds may be composed of two one-syllable words, a one-syllable word and a two-syllable word, two two-syllable words, etc., with clipping as necessary. Here, too, stress falls on σ2, even though both morphemes are semantically rich, and again regardless of syllable structure and weight.

[14] Denwood (1999: 89) suggests these second-syllable elements originated as particles.

Lex cat	*Gloss*	*Transcription*	*WT*
Nouns	yoke (neck+wood)	ŋya.'ɕiŋ	*gnya.shing*
	local dialect (place+language)	yul.'skat	*yul.skad*
	laborer (work+person)	las.'mi	*las.mi*
	gun (fire+arrow)	mɛ.'ða	*me.mda'*
	horse saddle (horse+saddle)	ɬtaz.'ga	*rta.sga*
Adjectives	little (small+measurement)	tɕʰun.'tse	*chung.tshad*
	blue (blue+color)	sŋo.'raŋ	*sngo.rang*[15]
Numerals	thirteen (ten+three)	tɕʊχ.'sum	*bcu.gsum*
	thirty (three+ten)	xsʊm.'tɕu	*gsum.bcu*
	fifteen (ten+five)	tɕo.'ʀa	*bco.lnga*
	fifty (five+ten)	ᵊʀaf.'tɕu	*lnga.bcu*

Table 3. σ2 stress in compound nouns, adjectives, and numerals

The compound numerals provide robust evidence of the correlation between syllable position and stress. Comparing 'thirteen' and 'thirty', for example, it is clear that the same morphological, semantic, and phonetic material is unstressed or stressed depending on its position in the word. Pitch and intensity traces for these numerals are compared in Figure 3 below. In these two words, pitch increases more than 20 Hz from σ1 to σ2, while there is little difference in intensity.

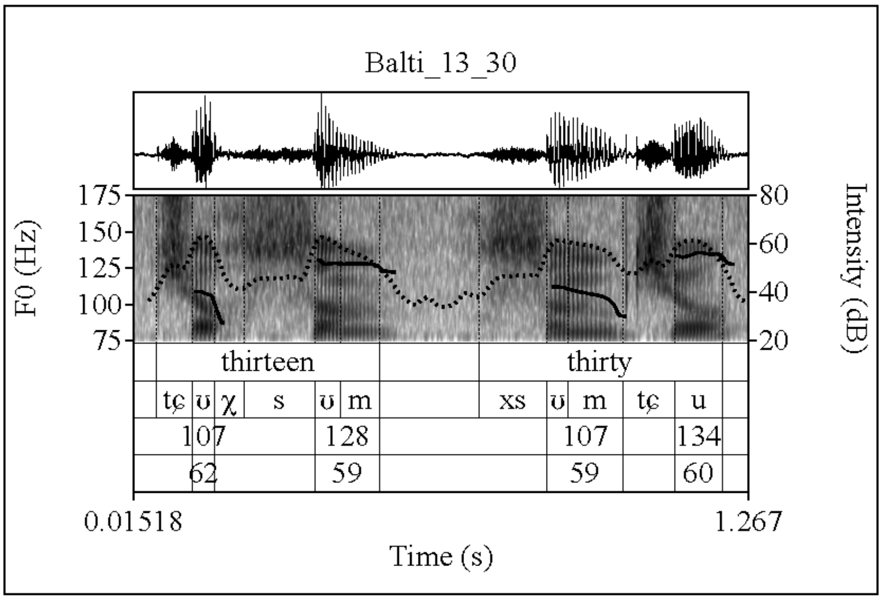


Figure 3. F0 and intensity curves for compound numerals. (Solid line: F0; dotted line: intensity. Speaker BSh_03)

Compound nouns also provide evidence of the association of stress with syllable position. In Table 4 below, one- and two-syllable words are variously combined to form compounds. Stress always falls on σ2: as highlighted in bold, the same morpheme is stressed or unstressed according to its position. For instance, in *a* 'arrow' occurs in σ1 and is unstressed, but in *b* it occurs in σ2 and is stressed.

[15] *rang* 'color' < Urdu < Persian (etymology suggested by Roland Bielmeier p.c.)

		1st element	2nd element	Compound
a.	Gloss	arrow	bow	bow-and-arrow
	Transcription	nda	ʁʐu	**nda**.'ʁʐu
	WT	*mda'*	*gzhu*	*mda'.gzhu*
	Lex cat	N	N	N
b.	Gloss	fire	arrow	gun, rifle
	Transcription	mɛ	nda	mɛ.'**ða**
	WT	*me*	*mda'*	*me.mda'*
	Lex cat	N	N	N
c.	Gloss	arrow	to shoot, to throw	archery
	Transcription	nda	pʰãɰ̃.ma	**da**.'pʰaŋ
	WT	*mda'*	*'phang.ma*	*mda'.'phang*
	Lex cat	*N*	*V*	*N*
d.	Gloss	year	new	new year
	Transcription	lo	**sar**.'pʰa	**lo**.'**sar**
	WT	*lo*	*gsar.pa*	*lo.gsar*
	Lex cat	N	A	N
e.	Gloss	horse	year	horse year[16]
	Transcription	ʂta	lo	ʂta.'**lo**
	WT	*rta*	*lo*	*rta.lo*
	Lex cat	N	N	N
f.	Gloss	year	head	first of the year
	Transcription	lo	gɔ	**lo**.'**mgo**
	WT	*lo*	*mgo*	*lo.mgo*
	Lex cat	N	N	N
g.	Gloss	head	brain	brain, mind
	Transcription	gɔ	**χlʌt**.'pa	**ŋgo**.'**ʁlʌt**
	WT	*mgo*	*klad.pa / glad.pa*	*mgo.glad*
	Lex cat	N	N	N
h.	Gloss	black	bear	black bear
	Transcription	naq.'po	**drɛn**.'mɔ̃	naʁ.'**dɛn**
	WT	*nag.po*	*dren.mo*	*nag.dren*
	Lex cat	A	N	N

Table 4. Compound nouns

4.1.3 Reduplicated Forms

Reduplicated nouns and adjectives again provide clear evidence of the association of stress with syllable position. The two syllables are identical in their phonetic and semantic content, yet stress always falls on σ2, as illustrated by the examples in Table 5.

[16] 'Horse year' is a designation in the Tibetan calendar, which is not used in Baltistan. My consultant was familiar with it nonetheless, and readily provided this form.

Lex cat	*Gloss*	*Transcription*	*WT*
Nouns	buttocks	pon.'pon	*pon.pon*
	paper	ɕoq.'ɕoq	*shog.shog*
	mother.hon	zi.'zi[17]	*n/a*
	elder brother	ka.'ka[18]	*n/a*
Adj	same, alike	tɕik.'tɕik	*gcig.gcig*
	very hot	tsʰik.'tsʰik	*'tshig.'tshig*
	hard, tough	ʈak.'ʈak	*tag.tag ?*
	crazy, foolish	tsʰu.'tsʰu	*'tshub.'tshub ?*

Table 5. σ2 stress in reduplicated nouns and adjectives

4.1.4 Nominalized Forms

As discussed in section 4.2 below, disyllabic verbs are stressed on σ1. However, when a nominalizer is added to the root, stress shifts to σ2. This is illustrated by comparing the verb ['tsoŋ.ma] *btsong.ma* 'to sell' to the agent- and place-nominalized forms [tsʰoŋ.'pa] *tshong.pa* 'merchant' and [tsʰoŋ.'sa] *tshong.sa* 'market'. Similarly, the verb ['ɕɛs.pa] *shes.pa* 'to know' is stressed on σ1, but the agent-nominalized form [ɕɛs.'kˣan] *shes.mkhan* 'expert' is stressed on σ2. This same stress pattern is observed in the nominalized forms [kʰyer.'kʰan] *'khyer.mkhan* 'trader' (from the verb 'to carry'), [tsʰɛm.'kʰan] *tshem.mkhan* 'tailor' (from the verb 'to sew'), and [ɬtsaŋ.'kʰan] *lsang.mkhan* 'beggar, mendicant'.

4.1.5 Borrowed Words

Non-verbs which enter the Balti lexicon through borrowing generally adopt σ2 stress. Words borrowed from English are shown in Table 6 below.[19] In *a,* English stress is retained on σ1 in 'penny' and on σ2 in 'hotel'. The disyllabic words in *b* all show a shift of English σ1 stress to Balti σ2 stress. The example in *c* shows an English trisyllabic word with initial stress reduced to a disyllabic word with σ2 stress. Finally, in *d,* an epenthesized vowel creates a preferred two-syllable word from a dispreferred one-syllable word: in [gɪ.'læs], stress falls on the original (σ2) vowel; in [ɸi.'lim] it falls on the epenthesized (σ2) vowel.

[17] Sprigg (2002: 182) tentatively suggests this word may have been borrrowed from Urdu, originating in Arabic.
[18] Sprigg (2002: 80) indicates that this is borrowed from Urdu, orginating in Persian.
[19] It is possible that these words entered the Balti lexicon indirectly – borrowed first into Urdu or into Pakistani English, for instance, and then into Balti – and may preserve the stress pattern of some such intermediate language. However, all of them were provided by speakers who had direct exposure to English, either through work (as a trekking porter) or through education.

	English		Balti
a.	coin	'pʰɛ.ni	'pɛ.nɛ
	hotel	ho.'tʰɛl	ho.'ʈɨl
b.	doctor	'dɑk.tər	dɑk.'tər
	thermos	'θr̩.mɨs	tθʌr.'mʊɕ
	camera	'kʰæm.rə	kɪm.'ɾa
	border	'bor.dər	bɑr.'dər
c.	officer	'ɐ.fɪ.sr̩	ʌɸ.'sɨr
d.	glass	'glæs	gɪ.'læs
	film	'fɪlm	ɸi.'lim

Table 6. Stress in nouns borrowed from English

Words borrowed from Urdu also exhibit assimilation in terms of stress, as illustrated by examples in Table 7.[20] For instance, stress falls on σ1 in the Urdu adjective meaning 'cheap, inexpensive', but shifts to σ2 when borrowed into Balti. In most varieties of Tibetan, the native word for 'cheap, inexpensive' is quite different, corresponding to the Written Tibetan spelling *khe.po*.

Gloss	*Urdu*	*Balti*	*Tibetan*
cheap, inexpensive	'sʌs.tə	sas.'ta	*khe.po*
forest, jungle	'dʑʌŋ.gl̩	dʑʌŋ.'gɐl	*shing.nags*
car, automobile	'ga:.ɽi	ga.'ɾi	n/a
saw (for cutting)	'a:.ɽa	a.'ra	*sog.le*
frog	'bʌ.tʌx [duck]	ɓa.'tek[21]	*sbal.pa*

Table 7. Stress in adjectives and nouns borrowed from Urdu

4.1.6 Categorical and Lexical Exceptions

The overwhelming majority of the hundreds of non-verbs elicited from the nine Balti speakers exhibited the σ2 stress pattern described here. Nonetheless, there were a number of categorical exceptions (kinship terms, deictics, and quantifiers) as well as idiosyncratic lexical exceptions.

Some disyllabic kinship terms are stressed on σ1, such as ['ʔa.ta] *a.ta* 'father', ['ʔã.ŋõ] *amo* 'mother', ['nɔ̃.ŋõ] *no.mo* 'sister of a girl', and ['nɛ.nɛ] *nene* 'maternal aunt'. Other kinship terms have σ2 stress, such as [ʔa.'po] *a.po* 'paternal grandfather', [ʔa.'pi] *a.pi* 'maternal grandmother', [strɪŋ.'mo] *sring.mo* 'sister of a boy', and [bu.'strɪŋ] *bu.sring* 'wife, woman'.

[20] Some of the examples here come from Sprigg 2002; Roland Bielmeier p.c., 2008; and Sadaf Munshi p.c., 2013.

[21] The word [ɓa.'tek] 'frog' provided by one Balti speaker apparently corresponds to the Urdu word meaning 'duck'. According to Sprigg 2002 (27, 207), [ba.'ʈek] can indeed mean 'duck' in Balti (and the word for 'frog' is given as [la.'ɖa.ŋo].) Regardless of the meaning, or of any potential confusion during elicitation with this speaker, stress clearly shifted from Urdu σ1 to Balti σ2 as the word was borrowed.

A few deictics also exhibit σ1 stress, including ['ye.ka] *'i.ka* 'here' from one Balti speaker, and ['di.ba] *di.ba* 'here' and ['a.ba] *a.ba* 'there' from another. Quantifiers with σ1 stress were ['ᵐbaχ.tɕi] *bags.tse* 'some' and ['tɕʰøn.tɕi] *chung.tse* 'a little'.[22]

I also encountered a number of idiosyncratic lexical exceptions, including ['dʌt.pa] *dad.pa* 'faith', ['pʰɘ.tɘ] 'wooden bowl', ['ⁿda:.man] *lda.man* 'small drum', ['nor.baŋ] *nor.bang* 'wealth (in livestock)', ['ʁzuk.po] *gzhug.po?* 'buttocks', ['tsʊn.mo] *btsun.mo* 'queen', and ['ᵐbaχs.pa] *pags.pa* 'animal hide'.

4.2 Stress Patterns on Disyllabic Verbs

In contrast to nouns, adjectives, and numerals, verbs are never stressed on σ2. Regardless of their morphological composition, they are stressed on σ1. The most explicit evidence of this contrast is illustrated by minimal pairs, discussed in section 4.2.1 below; σ1 stress for other verb forms is presented in section 4.2.2.

4.2.1 Non-verb vs. verb Minimal Pairs

In their descriptions of Balti, both Sprigg (1966: 188-189) and Bielmeier (1988a: 48) provide a handful of noun-verb minimal pairs which contrast only in terms of stress placement. The verbs considered are forms ending in some variant of *-pa* or *-ma*, which Sprigg and Bielmeier refer to as verbal nouns.[23]

I recorded some of these same minimal and near-minimal pairs, along with twenty or so additional examples, and confirmed the perceived stress pattern with acoustic measurements. For instance, the waveform, spectrogram, and pitch and intensity traces for the minimal pair [tʰaq.'pa] *thag.pa* 'rope' vs. ['tʰaq.pa] *'thag.pa* 'to grind' are shown in Figure 4 below. As shown, σ2 stress on the noun is conveyed primarily by pitch: the difference in average F0 across syllables is +24 Hz (110 Hz in σ1, 134 Hz in σ2), while average intensity is about the same (69 dB in σ1, 68 dB in σ2). In the verb, both pitch and intensity are used to convey σ1 stress: the difference in average pitch across syllables is -48 Hz (138 Hz in σ1, 90 Hz in σ2), and the difference in average intensity is -11 dB (71 dB in σ1, 60 dB in σ2). Vowel duration also contributes to the contrast in stress: in 'rope', the vowel durations in σ1 and σ2 are 71 msec and 132 msec, respectively (Δdur = +61 msec); in 'to grind', they are 102 msec and 86 msec (Δdur = -16 msec). As discussed below, this use of acoustic resources as correlates of stress is typical for this speaker.

[22] According to Muhammad Abbas Kazmi (p.c., Nicolas Tournadre), ['ᵐbaχ.tɕi] is used with humans [and perhaps other animate nouns?], while ['tɕʰøn.tɕi] is used with inanimate non-count nouns; [kʰa.'ik] *kha.cig* 'a few' is used with inanimate count nouns.

[23] Other scholars have called them infinitives or citation forms. My consultants usually translated the verbs as 'to V', and occasionally as 'V-ing'.

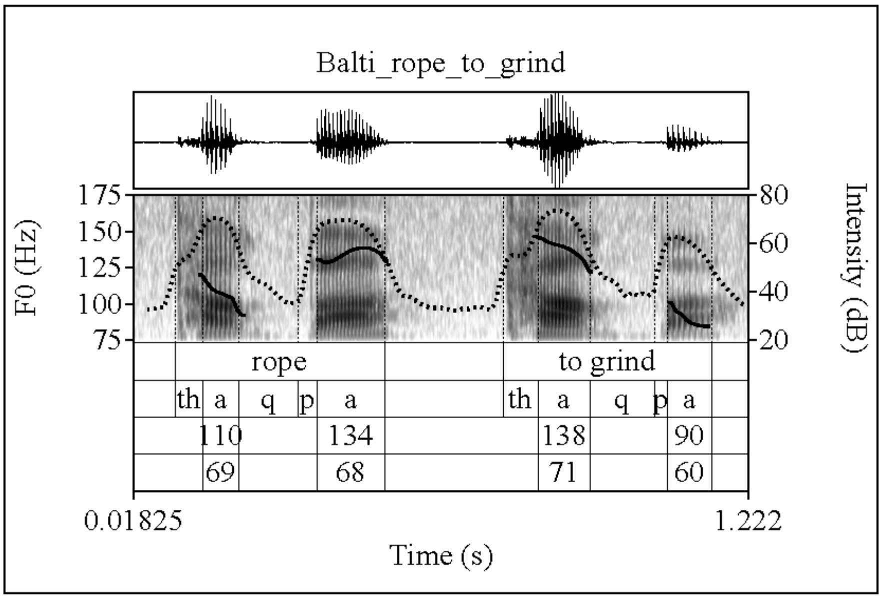


Figure 4. F0 and intensity curves for a noun-verb minimal pair. (Solid line: F0; dotted line: intensity. Speaker BSh_03)

Other non-verb *vs*. verb minimal pairs and near-minimal pairs are shown in Table 8 below; a more extensive list of examples can be found in Caplow 2016. In most cases, there is no obvious semantic relationship between the paired words.

Gloss	Lex cat	Transcription	WT
brain	N	xlʌt.ˈpa	*klad.pa*
to be tired	V	ˈxlʌt.pa	*glad.pa*
bone	N	ᵊrus.ˈpa	*rus.pa*
to dry, evaporate	V	ˈᵊrʌs.pa	*ras.pa*
star	N	skar. ˈma	*skar.ma*
to measure, limit	V	ˈskar.βa	*skar.ba*
horse's polo accessory	N	χmɪt.ˈpaχ	*??*
to devour, swallow	V	ˈχmɪt.pa	*mid.pa / dmid.pa*
high	A	tʰøn.ˈmo	*mthon.mo*
to arrive	V	ˈtʰøn.ma	*'thon.ma*
rare	A	ʂkøn.ˈmo	*dkon.mo*
to clothe (trans.)	V	ˈskøn.ma	*skon.ma*

Table 8. Balti non-verb vs. verb minimal pairs and near-minimal pairs

4.2.2 Stress on other verb forms

In addition to those which form minimal pairs, my data includes scores of other disyllabic verbs ending in some variant of *-pa* or *-ma;* these are likewise stressed on σ1, as illustrated in Table 9 below. Stress also falls on σ1 – and never on σ2 – in infinitives ending in *-bya,* and in noun + verbalizer constructions, noun-complement + verb structures, and imperatives.

Structure	*Gloss*	*Transcription*	*WT*
Infinitive w/ *-pa*	to forget	ˈβʐˠɛt.pa	*brjed.pa*
	to put	ˈyaq.pa	*'jog.pa*
	to leak, to drip down	ˈᵊʁzar.ba	*gzar.ba*
	to be able; can	ˈyɐn.ma	*nyan.ma*
	to lose, to squander	ˈstor.ʋa	*stor.ba*
	to wash	ˈkʰr̥o.wa	*'khru.ba*
Infinitive w/ *-bya*	to work (work+do)	ˈlaz.bya	*las.bya*
	to listen (ear+do)	ˈsna:.bya	*rna.ba.bya*
	to smell (smell+do)	ˈtri.bya	*dri.bya*
N + Verbalizer	[to] spread manure (manure+send)	ˈlut.taŋ	*lud.btang*
N-complement + V structure	[to] call, shout (language+say)	ˈskʰat.zɛr̥	*skad.zer*
	[to] cook (food+prepare)	ˈzan.ɸtɕo	*zan.bco*
	[to] boil water (water+boil)	ˈtɕʰu.skʊl	*chu.skol*
Imperative	close the door (door+close)	ˈzgo.tɕukʰ	*sgo.bcug*

Table 9. σ1 stress on other verbs

5 Results: Acoustic Correlates of Stress

Table 10 below summarizes the contributions made by F0, intensity, and vowel duration to the stress patterns perceived in Balti. F0 is the most consistent and most robust correlate of stress for words in all lexical categories – both the non-verbs, stressed on σ2, and the verbs, stressed on σ1. A listener would be able to identify the σ2 stress pattern on BSh_03 nouns, adjectives, and numerals from the F0 contrast alone. Intensity is also a robust correlate of stress for verbs. It is a limited and weak correlate of stress for nouns produced by speaker BM_01, but otherwise is not important to the perception of stress in Balti non-verbs. Vowel duration is also a correlate of σ1 stress in verbs. In non-verbs produced by speaker BSh_03, duration shows only a weak, general correlation with σ2 stress.

	Speaker	Setting	F0	Intensity	Vowel duration
Nouns (σ2)	BSh_03	isol	robust	no	weak
		frame	robust	no	weak
	BM_01	isol	strong	weak	no
Adjectives (σ2)	BSh_03	isol	robust	no – SS	weak – SS
		frame	robust	no – SS	weak – SS
	BM_01	isol	strong	no	no

	Speaker	Setting	F0	Intensity	Vowel duration
Numerals (σ2)	BSh_03	isol	robust	no – SS	weak – SS
Verbs (σ1)	BSh_03	isol	robust	robust	yes
		frame	robust	robust	no
	BM_01	citation	robust	robust	yes
		N+Vblzr	robust	yes	robust

SS Small Sample

Table 10. Acoustic correlates of stress for speakers BSh_03 and BM_01

In sum, fundamental frequency is a robust cue for stress in Balti for all lexical categories. Intensity is a robust correlate of stress for verbs, but not for non-verbs. Vowel duration is a robust correlate of stress for verbs; it also seems to be manipulated in accord with the σ2 stress perceived on non-verbs, but the effect is limited, given the incidental effects of vowel quality and syllable structure.

These acoustic features are considered in greater detail in the subsections below. The interaction of F0 and intensity is discussed in section 5.1 below. A more detailed analysis of the individual parameters follows, taking into consideration the intrinsic effects of syllable structure and vowel height.

5.1 Overview of F0 and Intensity

Considering F0 and intensity together shows that the categories of verbs and non-verbs occupy distinct acoustic spaces. In the graphs below, the difference in F0 across syllables is plotted on the x-axis, and the difference in intensity across syllables is plotted on the y-axis. Since the value measured in σ1 is subtracted from that measured in σ2, a negative difference means that F0 (or intensity) is higher on σ1 and a positive difference means that F0 (or intensity) is higher on σ2.

For speaker BSh_03, F0 and intensity differences are plotted in Figure 5 and Figure 6 for isolation and frame forms, respectively. The verbs fall in the lower-left quadrant: both F0 and intensity are markedly higher in σ1, the stressed syllable. The non-verbs plot on the right side of the graph: F0 is higher on σ2, but points straddle the horizontal 'ΔIntensity=0' line, meaning that intensity is sometimes higher on one syllable, and sometimes higher on the other. In both plots, one noun falls among the verbs; this lexical exception ['ŋgøn.mo] 'inner shirt' (an undergarment) was pronounced with strong σ1 stress and is excluded from further analysis.

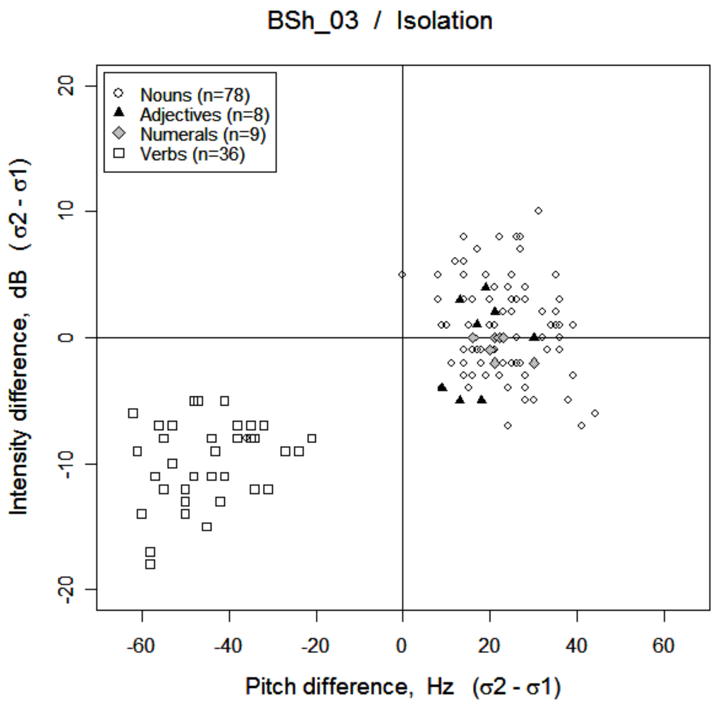


Figure 5. BSh_03 / Isolation: F0 difference *vs.* intensity difference

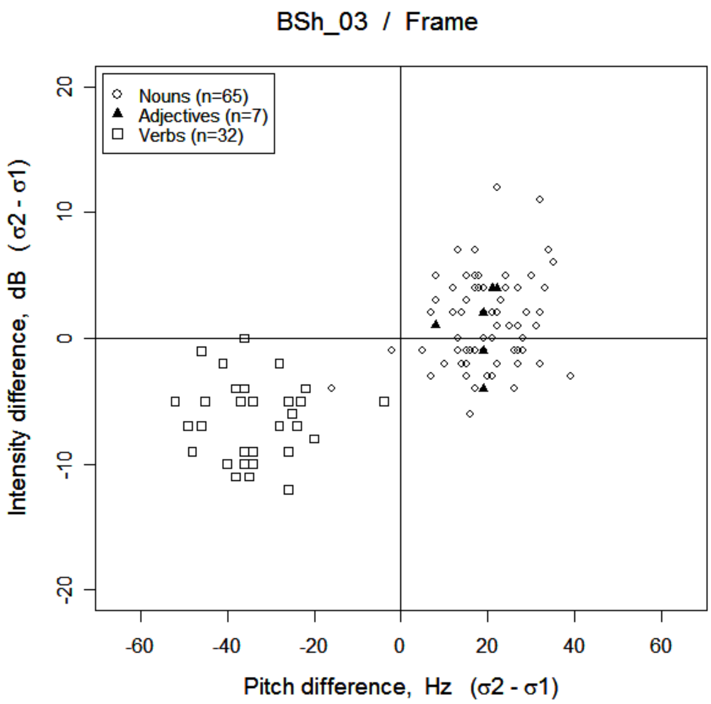


Figure 6. BSh_03 / Frame: F0 difference *vs.* intensity difference

Figure 7 below shows F0 and intensity differences across syllables for speaker BM_01. Again, nouns and adjectives pattern together and are distinct from verbs. The lone exception is the point at coordinates (-88,-8), which represents the noun ['ʁzuk.po] *gzhug.po?* 'buttocks', pronounced with a strong stress on the first syllable, just like a verb. This anomalous form was excluded from statistical analyses.

Speaker BM_01 appears to make less of a distinction between verbs and non-verbs than speaker BSh_03. This may reflect the more casual variability observed throughout the recording session with this speaker.

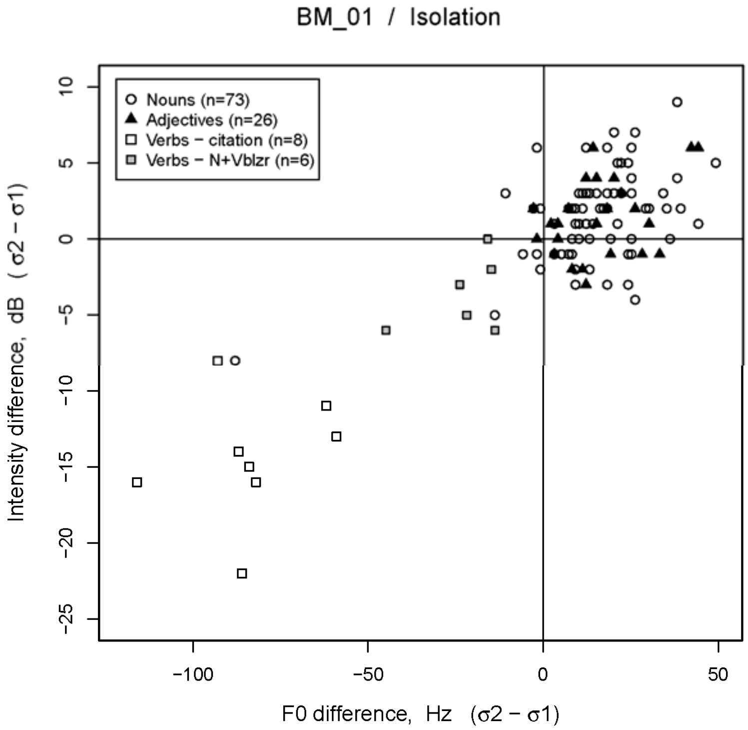


Figure 7. BM_01 / Isolation: F0 difference *vs.* intensity difference

5.2 Fundamental Frequency (F0)

For both speakers BSh_03 and BM_01, fundamental frequency is a robust correlate of stress for all lexical categories: F0 is consistently and significantly higher on σ2 in non-verbs, and on σ1 in verbs. Syllable structure influences average F0 measurements for speaker BM_01, as discussed below.

In the scatter plots below, a diagonal dashed line corresponds to the theoretical points where $F0_{\sigma 1} = F0_{\sigma 2}$. Points falling to the left of the dashed line represent words with a higher F0 on σ1, and points falling to the right of the dashed line represent words with a higher F0 on σ2. A dotted line has been added for reference, corresponding to an F0 difference across syllables of 10 Hz.

5.2.1 F0 for Speaker BSh_03

Figure 8 below shows fundamental frequency measurements on σ1 and σ2 for words produced in isolation. For non-verbs (nouns, adjectives, and numerals), points fall to the right of the dashed line – regardless of morphological composition – demonstrating that F0 is higher on σ2 for nearly every token. (The lone exception is the compound noun [ŋgøn.'tɕɛs] *gon.chas* 'clothing', which falls exactly on the dashed line.) Most tokens also fall to the right of the dotted reference line, indicating that the F0 increase across syllables is at least 10 Hz. For verbs, all points fall to the left of the dashed line (and well to the left of the dotted line representing a difference of 10 Hz), indicating that F0 is higher on σ1 for every token. (The lone Noun+Verbalizer form is [ˈzaːn.za] *zan.za* 'to eat food'.)

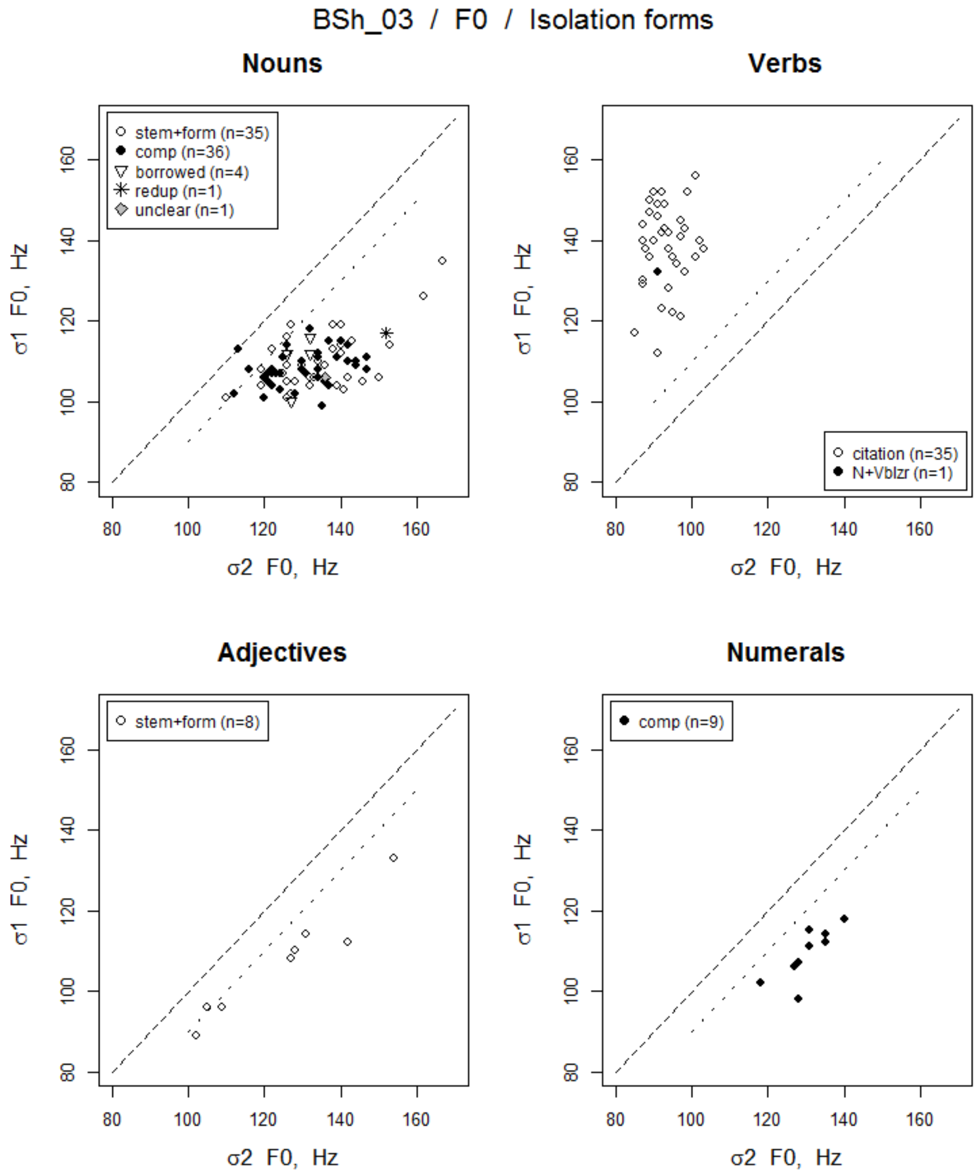


Figure 8. BSh_03 / Isolation: F0 contrasts. (Dotted reference lines at +10 Hz) [24]

Disyllabic words produced within a sentence frame show the same robust contrast in F0 across syllables, as illustrated in Figure 9. (There is no plot for numerals; they were not recorded in the sentence frame.) The lone noun with higher F0 in σ1 is [xpe.ra] *dpe.sgra* 'conversation, talk' (ΔF0 = -2 Hz); in this word, indeed, it is hard to hear a stress difference across syllables. Again, the lone Noun+Verbalizer form is [ˈzaːn.za] *zan.za* 'to eat food'.

[24] stem+form = stem + formative suffix, comp = compound, redup = reduplicated, N+Vblzr = noun plus verbalizer structure.

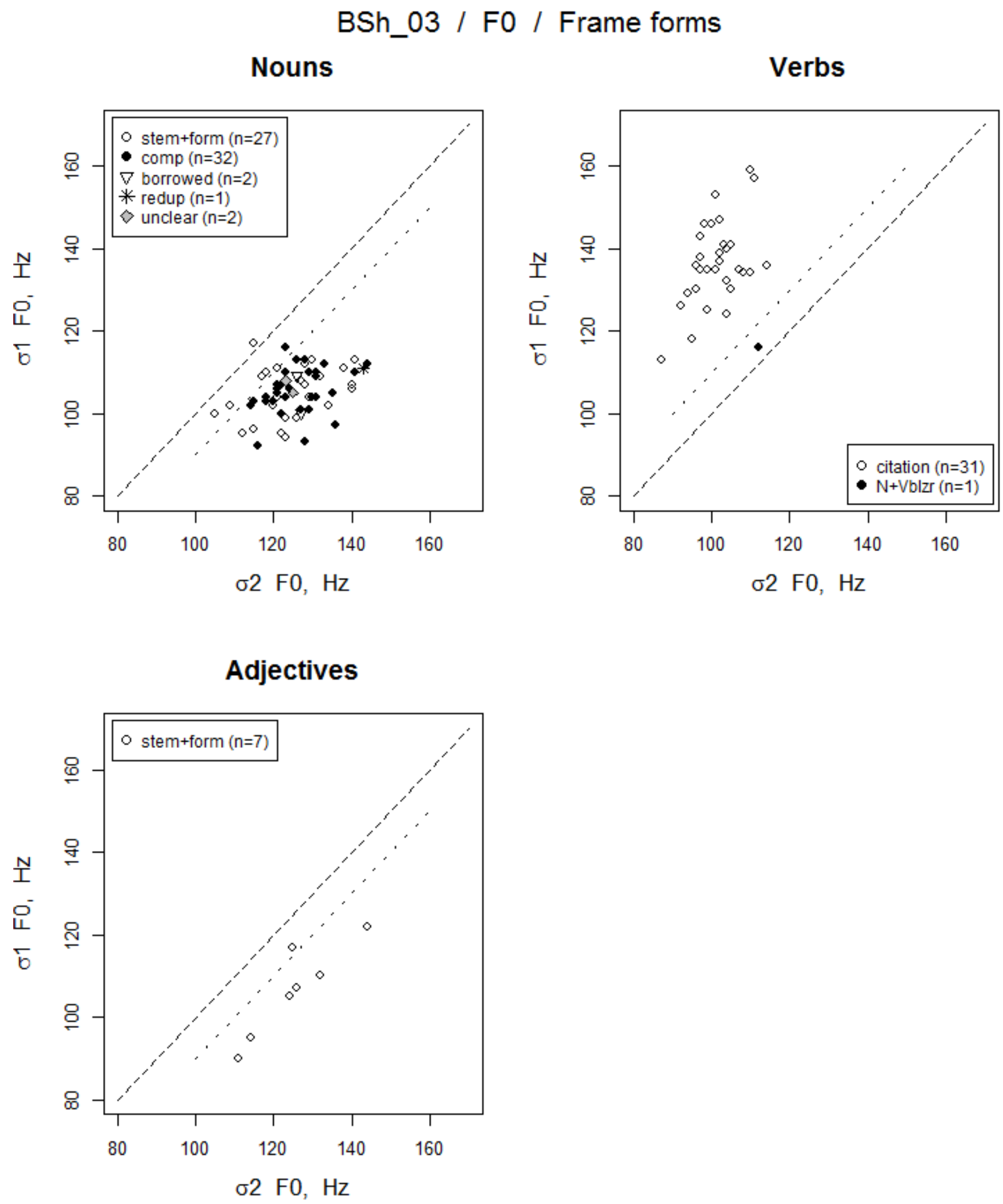


Figure 9. BSh_03 / Frame: F0 contrasts (Dotted reference lines at +10 Hz)

Paired-sample t-tests confirm the statistical significance of the F0 contrast across syllables. As shown in Table 11 below, $p << 0.05$[25] for all lexical categories, for both isolation and frame forms. For the non-verbs, F0 in σ2 is typically 20 Hz higher than in σ1. For verb citation forms / infinitives / verbal

[25] That is, there is a less than 5% probability that the mean pitch difference obtained could occur if there were no difference across syllables – i.e., if the pitch measurements on σ1 and σ2 simply represented the random variation which occurs within a single, normal population.

nouns, the contrast is even more dramatic: F0 in σ1 averages 40 Hz higher than in σ2. In all cases, even the lower 95% confidence limit[26] represents a perceptually significant contrast in F0.

Lex Cat	t	DF	p-value	Mean diff (Hz)	95% conf. limits	
					lower	upper
Nouns - I	22.0007	76	< 2.2e-16	23	21	25
Nouns - F	19.738	63	< 2.2e-16	20	18	22
Adjectives - I	7.7709	7	0.0001097	18	12	23
Adjectives - F	10.1103	6	5.441e-05	18	14	23
Numerals - I	15.3106	8	3.287e-07	21	18	24
Verbs (cit) – I *	-24.3064	34	< 2.2e-16	-45	-49	-41
Verbs (cit) – F *	-22.3284	30	< 2.2e-16	-35	-38	-32

* The lone BSh_03 N+Verbalizer form was excluded.

Table 11. Speaker BSh_03 / F0: Results of paired-sample t-tests (two-tailed) (I = isolation forms; F = frame forms)

5.2.2 F0 for Speaker BM_01

Speaker BM_01 showed greater variability in F0 contrasts, and measurements were influenced by differences in syllable structure, as discussed below. Despite these factors, F0 again functions as a strong perceptually and statistically significant correlate of σ2 stress in nouns and adjectives, and σ1 stress in verbs. Disyllabic numerals were not elicited from speaker BM_01.

As shown in Figure 10 below, almost all nouns – regardless of morphological composition – plot to the right of the dashed line, indicating that F0 is greater on σ2. Again, most also plot well to the right of the dotted reference line representing an F0 contrast of 10 Hz. For some tokens, however, the difference is less than 10 Hz, or is even negative, with F0 slightly higher on σ1 (e.g., in [ɕuk.'pa] *shug.pa* 'juniper'; ΔF0 = -3 Hz [27]).

[26] The 95% confidence limits indicate that, if one repeatedly recorded sample populations of disyllabic words from speaker BSh_03 and calculated the mean difference in pitch across syllables for each sample, there is a 95% probability that these mean pitch differences would fall between 21 and 25 Hz for nouns produced in isolation, between 12 and 23 Hz for adjectives, etc.

[27] Intensity and vowel duration are both greater on σ2 in this word, which is perceptually much more prominent than σ1.

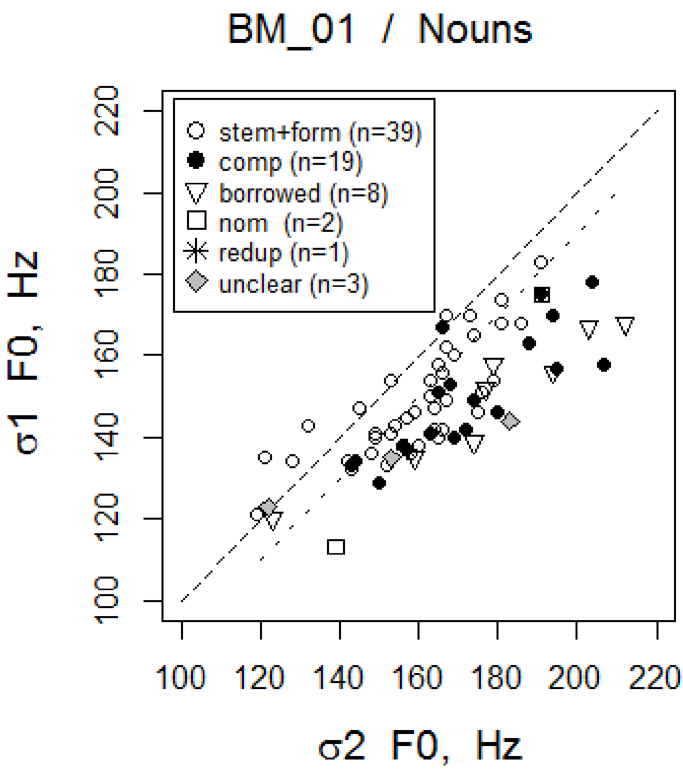


Figure 10. BM_01 / Nouns: F0 contrasts (Dotted reference lines at +10 Hz)

One factor which influences this distribution is whether the second syllable of the word is open or closed. Pitch traces for the nouns [ŋgo.'ʁlat] *mgo.glad* 'brain, mind' and [broq.'pa] *'brog.pa* 'nomad, pastoralist' illustrate this effect in Figure 11 below. The pitch traces are similar in the first syllable of the two words, and average F0 values measured over the middle 50% of the nucleus are essentially the same (152 Hz and 154 Hz). In [ŋgo.'ʁlat] σ2 is closed. Here, pitch over the vowel remains high and the utterance-final decline does not begin until the coda consonant is reached; the average F0 over the middle 50% of the nucleus is 173 Hz. In [broq.'pa] σ2 is open, and pitch begins to decline quite early. Even though the peak pitch is similar to that in [ŋgo.'ʁlat], the average is lower: 163 Hz. The increase in the measured F0 values across syllables is thus 21 Hz for [ŋgo.'ʁlat] 'brain', and 9 Hz for [broq.'pa] 'nomad'. (The latter is represented in Figure 10 by a point which falls fairly close to the dashed line.)

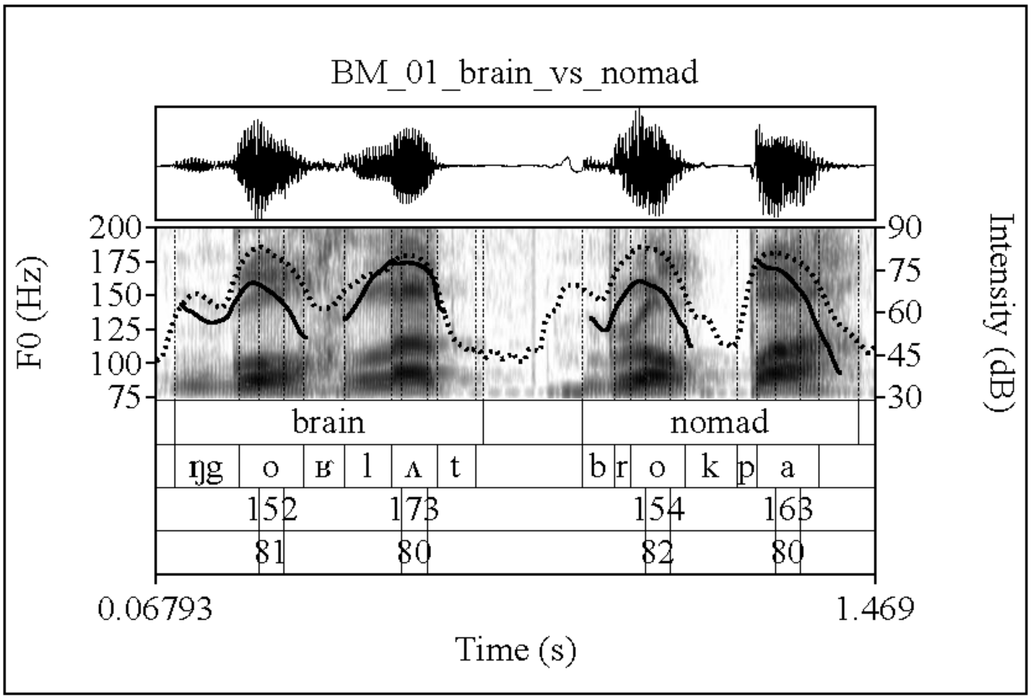


Figure 11. F0 vs. syllable structure: [ŋgo.ˈʁlat] *mgo.glad* 'brain, mind' vs. [broq.ˈpa] *'brog.pa* 'nomad' (Solid line: F0; dotted line: intensity. Speaker BM_01)

The effect of this depression of F0 in open syllables is illustrated for nouns in Figure 12 below. On the right, when σ2 is closed, nearly all points fall well to the right of the dotted +10 Hz reference line. On the left, when σ2 is open, the increase in F0 across syllables is less consistent and less pronounced, and in some words F0 is higher on σ1.

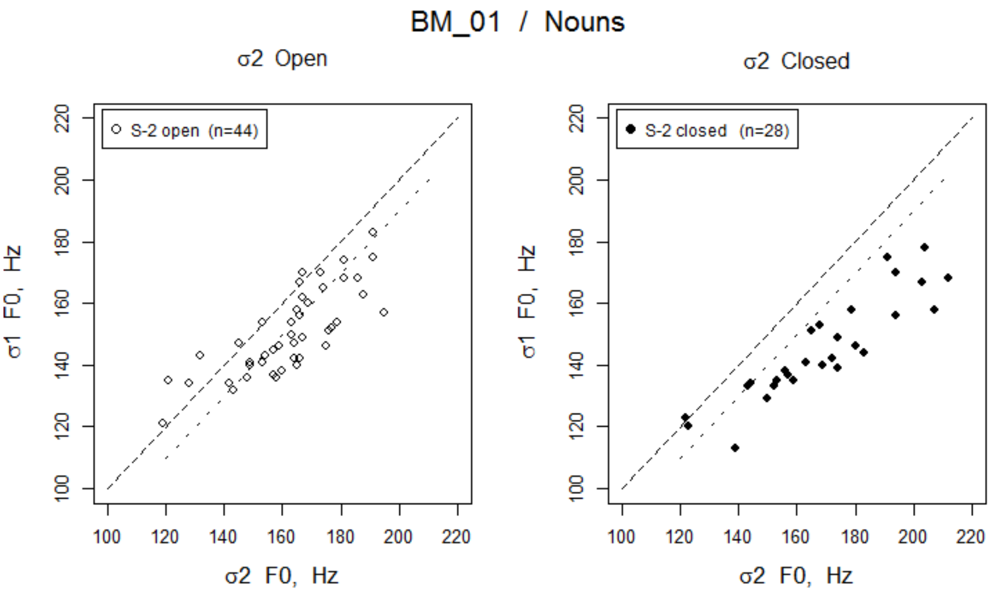


Figure 12. BM_01 / Nouns: F0 *vs.* σ2 closure (Dotted reference lines at +10 Hz.)

The differences here are a reflection of the technique of measuring F0 over the middle 50% span of the vowel. This factor likely does not affect perception of relative syllable prominence; listeners may make a subconscious adjustment for this effect, they may focus on peak F0 as a cue for stress, or they may focus on the fall in F0 over the syllable nucleus. (And, of course, F0 is not the only acoustic cue that listeners rely on as a correlate of stress.)

Most adjectives also have a higher F0 on σ2, regardless of morphological composition, as shown in the left panel in Figure 13 below. For many tokens the contrast is >10 Hz. Some elements of the distribution can again be accounted for in terms of syllable structure. There are only five adjectives in which σ2 is closed (e.g., [sŋo.ˈraŋ] *sngo.rang* 'blue', [tɕʰaʈ.ˈtɕʰaʈ] *chad.chad* 'short, few') but – as illustrated in in the right panel below – the difference in F0 across syllables for all of them is > 10 Hz.

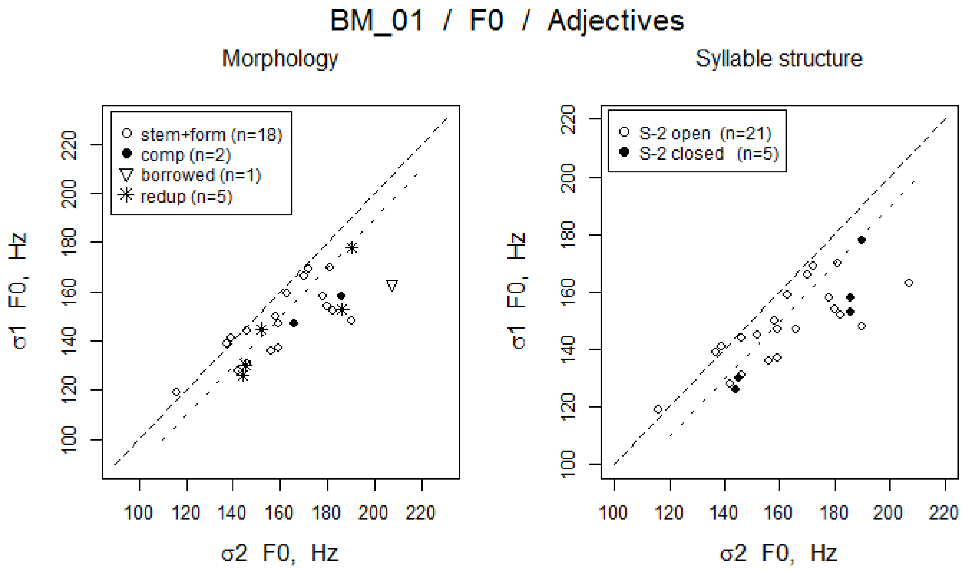


Figure 13. BM_01 Adjectives: F0 (Dotted reference lines at +10 Hz.)

For verbs, plotted in Figure 14 below, all points fall to the left of the dashed line, indicating that F0 is higher on σ1 for every token; they also fall to left the of the dotted line indicating that the F0 difference is greater than 10 Hz. The contrast in F0 across syllables varies with morphological composition, and a closer examination reveals that individual tokens may also be influenced by contrasts in syllable structure, vowel height, onset voicing, etc. Given the small sample size, the effect of these factors cannot be distinguished. However, they are all moot: the contrast in F0 is consistent and sometimes dramatic, in one case even exceeding 100 Hz.

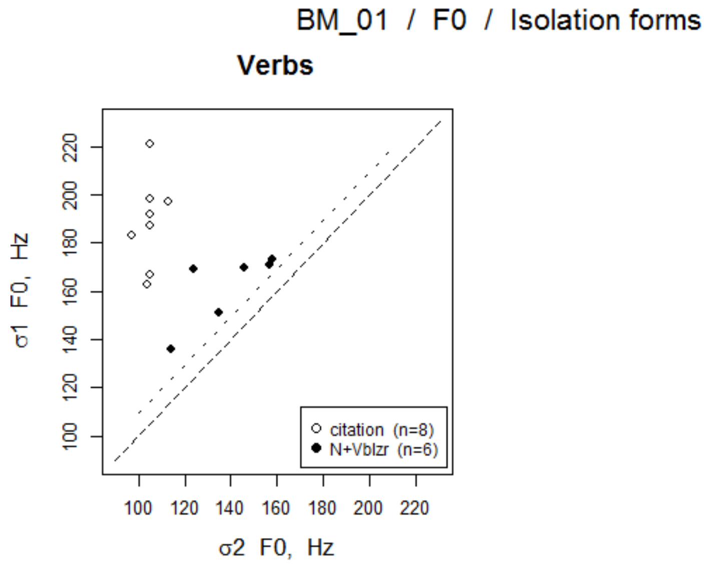


Figure 14. BM_01 Verbs: F0 contrasts (Dotted reference lines at +10 Hz.)

Results of paired-sample t-tests for nouns, adjectives, and verbs produced by speaker BM_01 are shown in Table 12 below, taking into consideration syllable structure and morphological composition as appropriate. In all scenarios, the contrast in F0 across syllables is not only statistically significant ($p < 0.05$), but is also perceptually significant. The lowest lower 95% confidence limits are 8 Hz, for nouns and adjectives in which σ2 is open; even this is a clear cue for stress. For all other subsets, an F0 contrast of at least 10 Hz can be predicted; for verbs the difference is likely to exceed -10 Hz.

Lex Cat	t	DF	p-value	Mean diff (Hz)	95% conf. limits lower	upper
Nouns - all	11.0504	71	< 2.2e-16	16	13	19
Nouns - σ2 open	7.1379	43	8.1e-9	12	8	15
Nouns - σ2 closed	10.711	27	3.191e-11	24	19	28
Adjectives - all	6.132	25	2.071e-06	15	10	21
Adjs - σ2 open	4.8079	20	0.0001070	14	8	20
Adjs - σ2 closed	5.31	4	0.006046	21	10	32
Verbs - all	-6.2009	13	3.214e-05	-78	-37	-58
Verbs - citation	-13.252	7	3.258e-06	-84	-98	-69
Verbs - N+Vblzr	-4.7633	5	0.005045	-23	-35	-10

Table 12. Speaker BM_01 / F0 / Results of paired-sample t-tests (two-tailed)

5.3 Intensity

Both speakers use intensity as a consistent and robust correlate of the σ1 stress perceived on verbs. However, they draw upon this acoustic resource differently for nouns and adjectives. For nouns produced by speaker BSh_03, intensity is definitively not a correlate of stress. For speaker BM_01, intensity is a limited correlate of stress, higher on σ2 except in words where it is obscured by the intrinsic intensity of a lower vowel in σ1. For adjectives and numerals produced by speaker BSh_03, the sample size is small, but does not indicate that intensity is used as a correlate of σ2 stress. For adjectives produced by speaker BM_01, the contrast in intensity across syllables is neither statistically nor perceptually significant.

5.3.1 Intensity for Speaker BSh_03

As demonstrated below, for speaker BSh_03, intensity is a consistent and robust correlate of stress for σ1 stress in verbs. It is definitively not a correlate of stress for σ2 stress in nouns. The sample size for adjectives and numerals is small, but provides no suggestions that intensity is used as a correlate of σ2 stress.

If intensity served as an acoustic cue for stress in BSh_03 nouns, one would expect to see a consistently and significantly higher intensity on σ2. However, this is not the case for either isolation or frame forms. As shown in Figure 15 below, points fall to either side of the dashed line, indicating that intensity is sometimes higher on σ1 and sometimes higher on σ2. Nearly all words show an intensity contrast across syllables between -5 dB and +5 dB. Morphological structure does not govern the distribution.

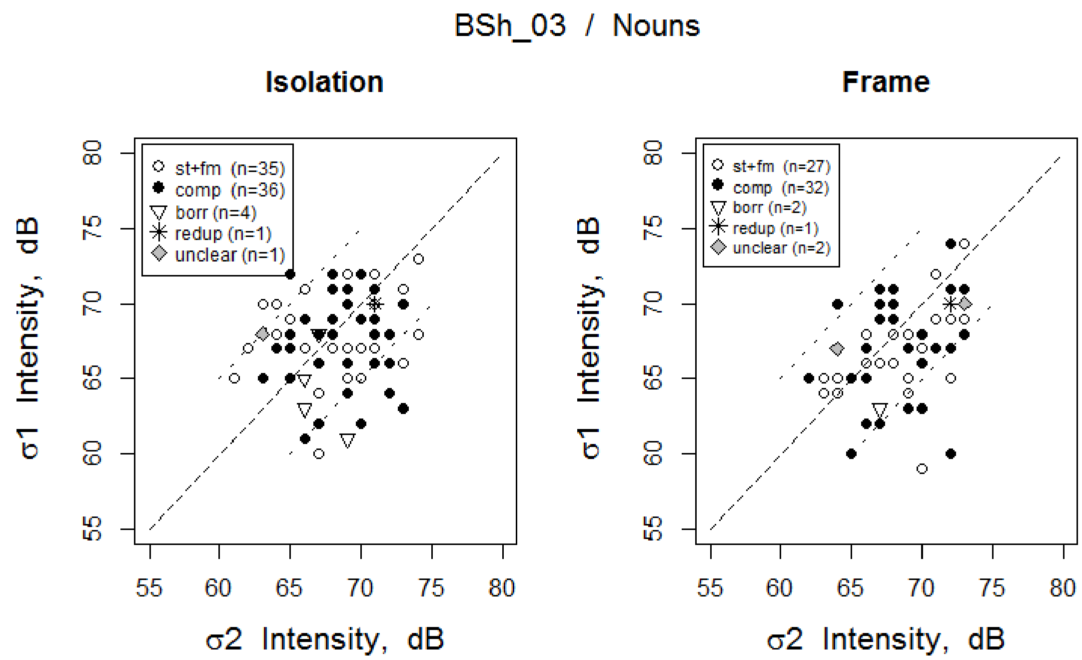


Figure 15. BSh_03 / Nouns: Intensity (Dotted reference lines at ±5 dB.)

Except for a few tokens, this distribution is entirely attributable to the intrinsic variation of intensity with vowel height. For forms produced in isolation, the plot on the left in Figure 16 below shows that intensity is greater on whichever syllable has the lower vowel. Nouns in the [lower.higher] group – such as [ndaʁ.'ʑu] *mda'.gzhu* 'arrow and bow' and [ɓaχ.'mo] *bag.mo* 'bride' – have a higher intensity on σ1 and fall to the left of the dashed line. Nouns in the [higher.lower] group – such as [mɛ.'ða] *me.mda'* 'gun, rifle' and [tut.'pa] *dud.ba* 'smoke' – have a higher intensity on σ2, and fall to the right of the dashed line. The plot on the right below shows that nouns with vowels of the same height in both syllables – such as [spal.'ba] *dpral.ba* 'forehead' and [bu.'striŋ] *bu.sring* 'woman' – are about evenly distributed to either side of the dashed line. This is a random distribution; there is no evidence that the speaker is manipulating intensity to convey stress.

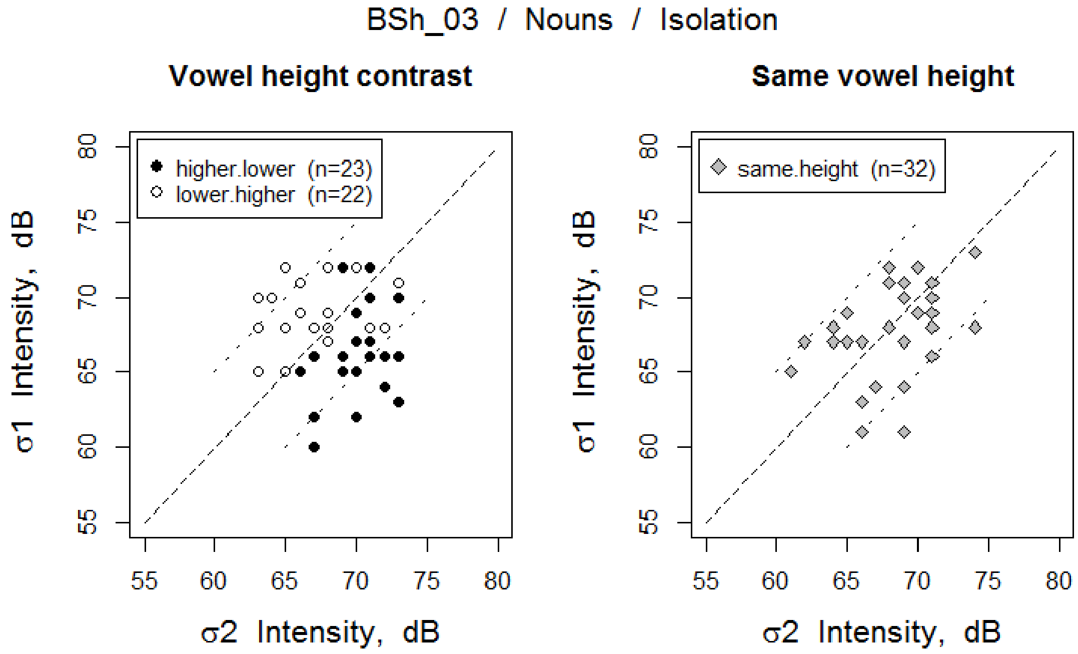


Figure 16. BSh_03 / Nouns / Isolation: Intensity *vs.* vowel height contrast. (Dotted reference lines at ±5 dB.)

Similar patterns are exhibited by nouns produced in the sentence frame "In the Balti language, we *X* say" / ['bal.ti 'skat.diŋ *X* za.ret], as illustrated in Figure 17 below. When there is a contrast in vowel height across syllables, intensity tends to be higher on whichever syllable has the lower vowel; nouns in the control group are more or less evenly distributed to either side of the dashed line.

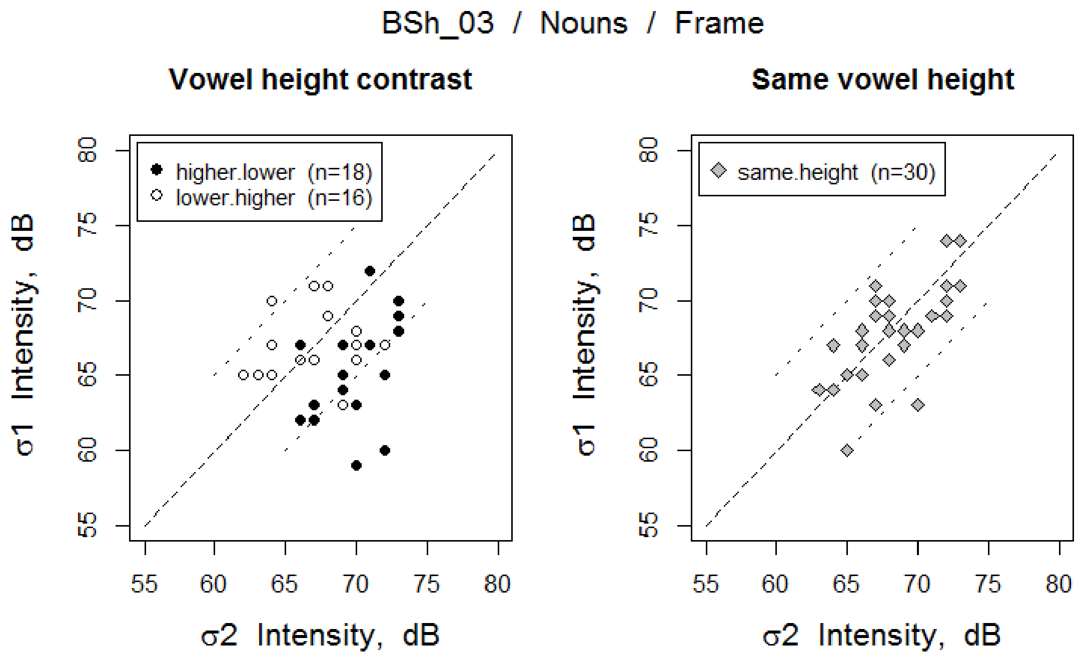


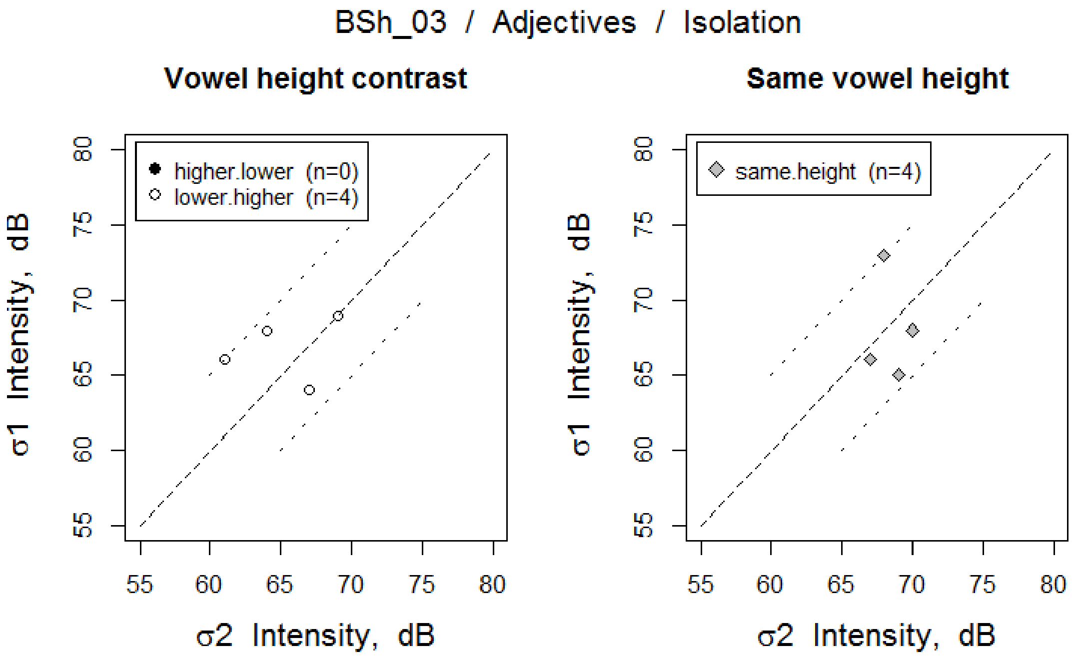
Figure 17. BSh_03 / Nouns / Frame: Intensity *vs.* vowel height contrast. (Dotted reference lines at ±5 dB.)

Paired-sample t-tests, summarized in Table 13 below, confirm that intensity differences across syllables do not define a significant contrast: $p > 0.05$, the 95% confidence interval includes zero, and the mean difference is <1, which is below the difference limen for intensity. In conclusion, intensity is definitively not a correlate of stress for BSh_03 nouns, either in isolation or in the sentence frame.

Vowels	t	DF	p-value	Mean diff (dB)	95 % conf. limits	
					lower	upper
same height - I	0.2088	31	0.836	0.12	-1.1	1.3
same height - F	0.5801	29	0.5664	0.27	-0.7	1.2

Table 13. BSh_03 / Nouns: Intensity, controlled V height: paired-sample t-tests (two-tailed)

Only a small number of adjectives recorded from Speaker BSh_03 could be reliably segmented for acoustic analysis: eight adjectives in isolation and seven in the sentence frame.[28] The scatter plots below show that, even when vowel height is controlled, intensity is sometimes higher on σ1 and sometimes higher on σ2. Among the isolation forms, the word [maː̃.'mo] *mang.mo* 'many' falls exactly on the "ΔInt=0" line.



[28] For the forms in the sentence frame, [naq.'po] *nag.po* 'black' and [maː̃.'mo] *mang.mo* 'many' overlap at (σ2, σ1) coordinates (67, 63), so it appears that there are only three data points with contrasting vowel height.

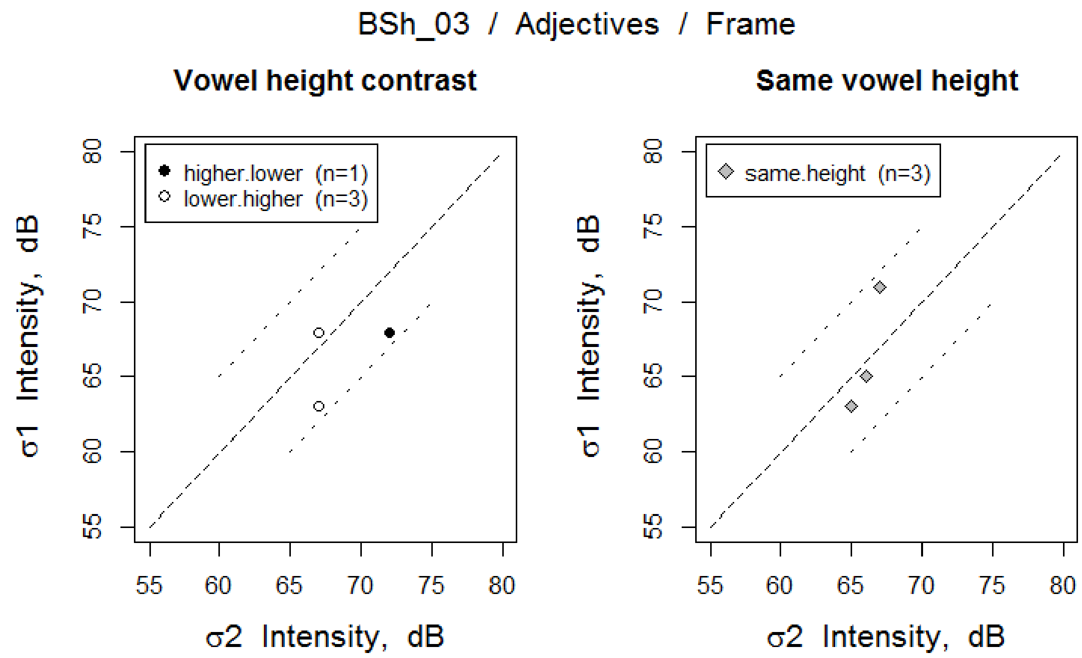


Figure 18. BSh_03 / Adjectives: Intensity *vs.* vowel height contrast. (Dotted reference lines at ±5 dB.)

For numerals (produced only in isolation), intensity is never higher on σ2, as illustrated in Figure 19 below.

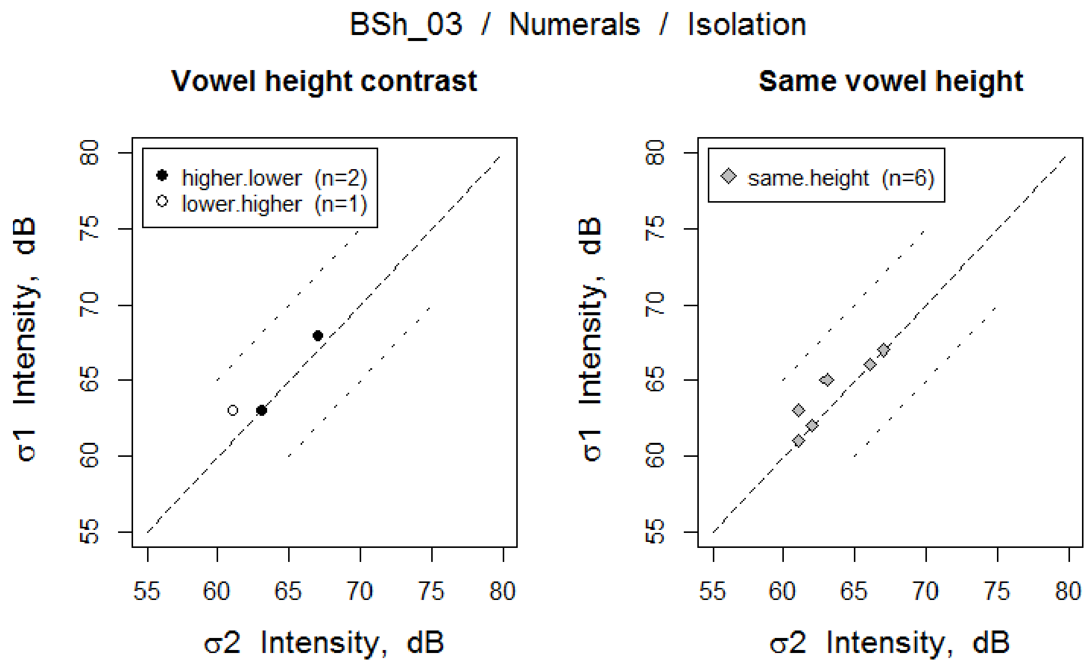


Figure 19. BSh_03 / Numerals: Intensity *vs.* vowel height contrast. (Dotted reference lines at ±5 dB.)

Given the small sample size and the scattered distribution for adjectives and numerals, t-tests do not provide a helpful assessment of intensity contrasts. Instead, the ranges and mean values of the difference in intensity from σ1 to σ2 are shown in Table 14 below. Most of the mean values are smaller than the discriminable difference of 1 dB. There is no evidence to indicate that speaker BSh_03 utilizes intensity as an acoustic correlate of stress in adjectives or numerals.

Lexical category	Setting	Vowels	n	Intensity (dB) minimum	maximum	mean
Adjectives	Isolation	all	8	-5	4	0.5
		same height	4	-5	4	0.5
	Frame	all	7	-4	4	1.4
		same height	3	-4	2	-0.3
Numerals	Isolation	all	9	-2	0	-0.8
		same height	6	-2	0	-0.7

Table 14. BSh_03 / Adjectives and Numerals / Intensity

Intensity values for verbs produced by speaker BSh_03 are plotted in Figure 20 below. Except for the frame form ['smɪn.ma] *smin.ma* 'to be ripe' which falls exactly on the dashed line, all other verbs have a higher intensity on σ1. In most cases the contrast across syllables is >5 dB, and is generally more pronounced in forms produced in isolation than in the sentence frame. For example, the intensity differences in the isolation and frame forms of the verb ['mɪn.ma] *min.ma* 'to give' are -11 dB and -7 dB, respectively. These values are particularly striking given that the [high.low] contrast in vowel height actually favors a higher intensity on σ2. The lone N+Vblzr form ['zaːn.za] *zan.za* 'to eat food' plots among the citation forms in terms of intensity.

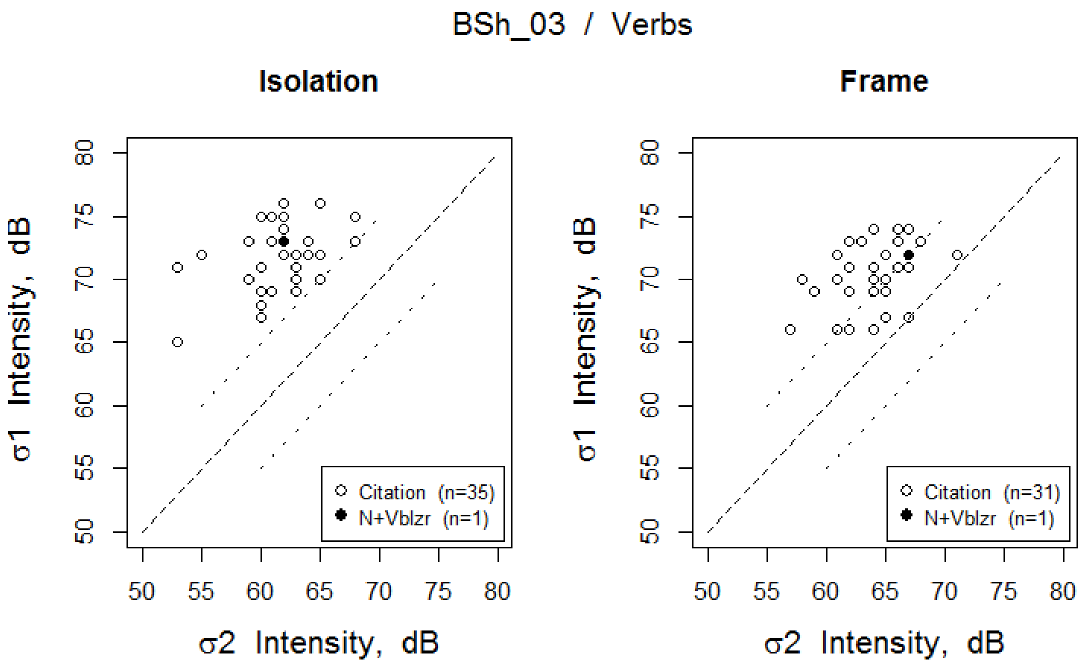


Figure 20. BSh_03 / Verbs: Intensity. (Dotted reference lines at ±5 dB.)

Paired-sample t-tests, summarized below, confirm the statistical significance of intensity differences across syllables, with p-values << 0.05. Even the smallest probable contrast of -5.5 dB – represented by the lower 95% confidence limit for the frame forms – would be highly perceptible.

Setting	t	DF	p-value	Mean diff (dB)	95 % conf. limits	
					lower	upper
isolation	-17.486	34	< 2.2e-16	-10	-11	-8.8
frame	-11.7417	30	9.593e-13	-6.7	-7.8	-5.5

* The lone BSh_03 N+Verbalizer form was excluded.

Table 15. BSh_03 / Verbs / Citation / Intensity: results of paired-sample t-tests (two-tailed)

It is thus clear that intensity is a robust correlate of the σ1 stress on citation forms of verbs produced by speaker BSh_03, both in isolation and in the sentence frame. While we know that intensity varies as a function of vowel height, for BSh_03 verbs this variation in no case yields a higher value on σ2. That is, the correlation between intensity and stress is exceptionless, and any intrinsic variation is moot.

5.3.2 *Intensity for Speaker BM_01*

For speaker BM_01, intensity is again a consistent and robust correlate of stress for σ1 stress in verbs. It is a limited correlate of stress for nouns, higher on σ2 except when the intrinsic intensity of a lower vowel in σ1 obscures the contrast. For adjectives, the contrast in intensity across syllables is generally consistent with σ2 stress, but it is neither statistically nor perceptually significant.

As shown in Figure 21 below, about two-thirds of nouns and adjectives have a higher intensity on σ2, regardless of morphological composition.

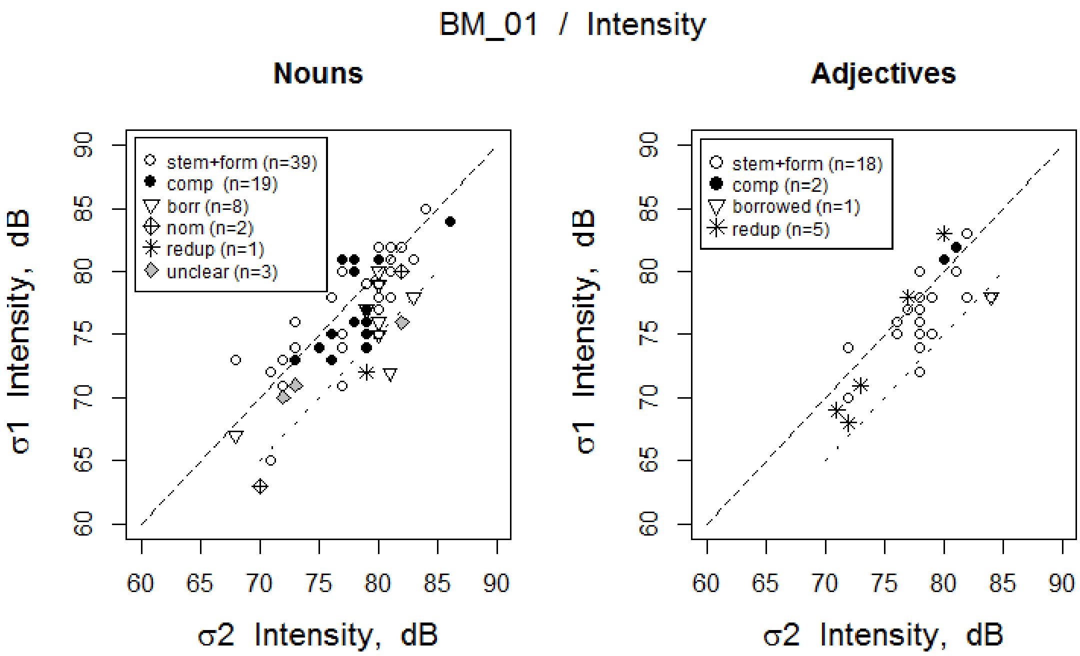


Figure 21. BM_01 / Nouns and adjectives: Intensity (Dotted reference lines at +5 dB.)

In contrast to speaker BSh_03, intrinisic intensity associated with vowel height does not dominate the pattern here. For nouns, when vowel height is controlled, as shown on the right below, intensity is generally higher on σ2, sometimes by more than 5 dB. But this is also likely to be the case when vowel height differs across syllables, as shown on the left below, and this pattern appears to be random. A closer examination of the data did not reveal a correlation with any other factors, such as syllable structure. Adjectives show much the same pattern.

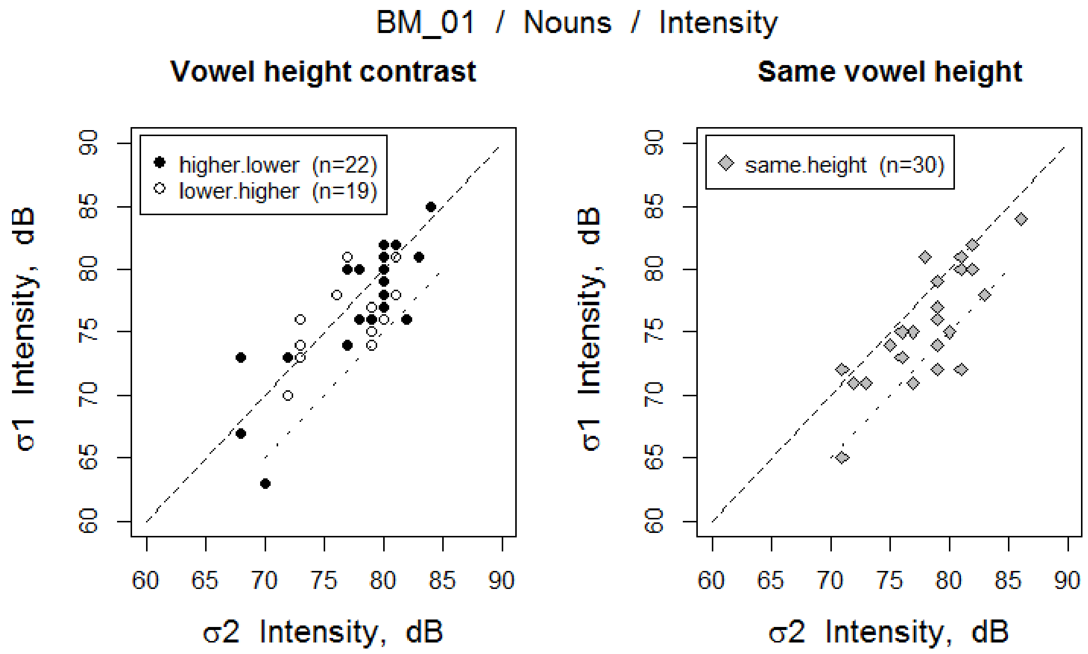


Figure 22. BM_01 / Nouns and adjectives: Intensity *vs.* vowel height.[29] (Dotted reference lines at +5 dB.)

The intensity contrast across syllables is nonetheless statistically significant for nouns, as shown in Table 16 below. This contrast is perceptually significant when vowel height is controlled, but when the full set of nouns is considered, the contrast is barely at the difference limen, with a lower 95% confidence limit of 1 dB. Thus intensity can be considered a weak cue for stress; it will be higher on σ2 except when a low vowel in σ1 obscures this.

When all adjectives are considered, the difference in intensity is statistically significant but not perceptually significant; when vowel height is controlled, the difference is neither statistically nor perceptually significant. Thus even though intensity is higher on σ2 in many words, the contrast is not strong or consistent enough to be identified as a correlate of stress.

[29] Words with diphthongs were excluded in considering vowel height contrasts. For nouns n=71; [tʰaũŋ.ˈboŋ] *thang.bong* 'wild donkey' was omitted. For adjectives n=24; [leaʀ.ˈmo] *legs.mo* 'good' and [mãõ.ˈmo] *mang.mo* 'many' were omitted.

Lexical category	Vowels	t	DF	p-value	Mean diff (dB)	95 % conf. limits lower	upper
Nouns	all	5.1692	71	2.074e-06	1.7	1.0	2.3
	same height	5.7111	29	3.528e-06	2.6	1.7	3.6
Adjs	all	2.8881	25	0.007888	1.5	0.4	2.5
	same height	1.9387	14	0.07297	1.3	-0.14	2.8

Table 16. BM_01 / Nouns and adjectives / Intensity: Paired-sample t-tests (two-tailed)

For verbs, on the other hand, intensity is a strong and reliable correlate of stress, regardless of morphological composition – and, again, regardless of contrasts in vowel height, which in some verbs actually favors a higher intensity on σ2. As shown in Figure 23 and Table 17 below, intensity is almost always higher on σ1; the lone point which falls on the dashed line is ['zan.ɸtɕøs] *zan.bcos* 'to make food'. The contrast is particularly robust for citation forms, averaging -14 dB.

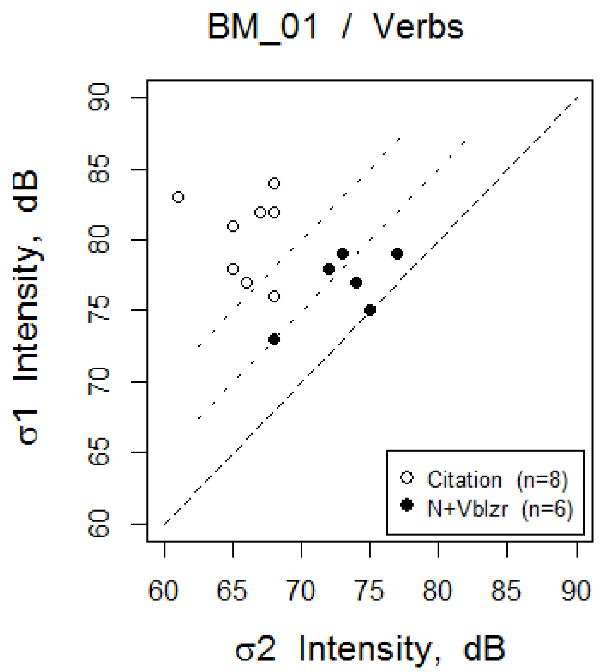


Figure 23. BM_01 / Verbs: Intensity. (Dotted references lines at -5 dB and -10 dB.)

Subset	t	DF	p-value	Mean diff (dB)	95 % conf. limits lower	upper
all	-5.6792	13	7.551e-05	-9.7	-14	-6.1
citation	-9.9081	7	2.274e-05	-14	-18	-11
N+Vblzr	-3.7081	5	0.01388	-3.7	-6.2	-1.1

Table 17. BM_01 / Verbs / Intensity: Paired-sample t-tests (two-tailed)

5.4 Vowel Duration

For speaker BSh_03, vowel duration shows a general, weak correlation with σ2 stress in non-verbs. It is a correlate of σ1 stress in verbs produced in isolation, but not in those produced in the sentence frame. For speaker BM_01 vowel duration is an acoustic correlate of σ1 stress in verbs, but not of σ2 stress in non-verbs.

Several incidental factors which affect vowel duration were taken into consideration in this analysis: (a) vowels may be lengthened in the final syllable of an utterance; (b) vowels are longer in open syllables than in closed syllables; (c) low vowels are longer than high vowels (Lehiste 1970: 18ff). A contrast in any of these features across syllables may influence relative vowel duration. Tibetan is certainly susceptible to such effects: in many words, the second syllable is open and has the low vowel [a] as its nucleus, while the first syllable may vary widely in both respects. Furthermore, forms produced in isolation may be subject to utterance-final lengthening.

5.4.1 Vowel duration for Speaker BSh_03

For nouns produced by speaker BSh_03, vowel duration shows a general, weak correlation with the perceived σ2 stress, but the contrast is statistically and perceptually significant only when it is exaggerated by favorable contrasts in syllable structure or vowel height. The speaker does not robustly manipulate vowel duration to lend prominence to σ2; on the other hand, we do not observe a contradictory pattern of longer vowels on σ1. For adjectives and numerals, the sample size is too small to reveal a convincing pattern. Vowel duration is again generally consistent with the σ2 prominence conveyed by F0.

For verbs produced in isolation, vowels are longer in σ1, except when the duration contrast is obscured by the intrinsic effects of vowel height contrasts. Duration is not a correlate of stress for verbs produced in the sentence frame.

In nouns, as illustrated in Figure 24 below, vowels are generally longer in σ2 – sometimes by more than 30 msec, as indicated by the dotted reference line. Utterance-final lengthening may have a minor effect: the distribution is similar in both settings, but among the isolation forms, more tokens fall further from the dashed line. For some nouns, the duration contrast across syllables is likely below the discriminable difference; in a handful, the vowel is longer in σ1, the unstressed syllable.

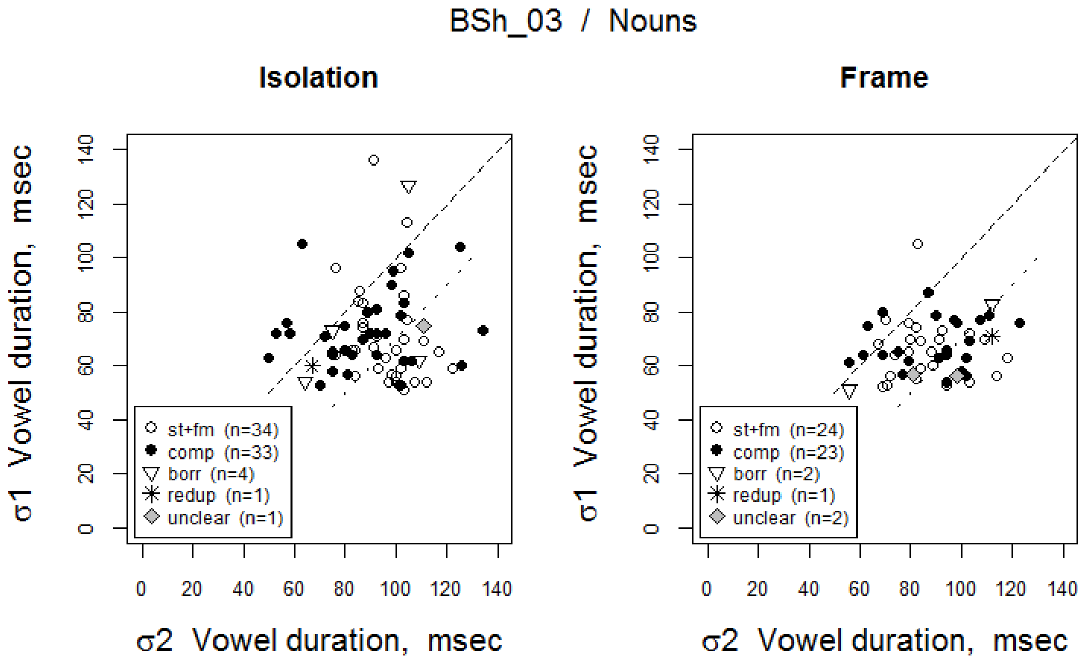


Figure 24. BSh_03 / Nouns: Vowel duration data (words with diphthongs and compensatory lengthening excluded) (Dotted reference lines at +30 msec.)

A two-way analysis of variance (ANOVA) was used to determine the effect of incidental factors. As summarized in Table 18, syllable structure is a main effect for vowel duration in isolation forms (F=5.18, p=0.003). For frame forms, vowel height is significant at the level of 0.10, though not at the of 0.05 level (F=2.92, p=0.065). These factors do not interact (F=0.70, p=0.65 for isolation forms; F= 0.85, p=0.52 for frame forms).

		DF	Sum squares	Mean squares	F value	Pr (>F)
Isolation	Syllable structure	3	7407	2469.1	5.185	0.00295 **
	Vowel height	2	1398	698.8	1.468	0.23851
	Interaction	6	2011	335.2	0.704	0.64755
	Residuals	61	29048	476.2		
Frame	Syllable structure	3	937	312.4	1.013	0.3966
	Vowel height	2	1801	900.5	2.921	0.0652
	Interaction	5	1309	1309	0.849	0.5234
	Residuals	41	12641	308.3		

Table 18. BSh_03 / Nouns / Duration: Vowel duration vs. syllable structure and vowel height; Two-way ANOVA.

For isolation forms, the effect of syllable structure on the difference in vowel duration across syllables is illustrated by the box-and-whisker plot in Figure 25. (The median is shown by the dark band at the waist; the mean is marked with an asterisk; the notches represent the 95% confidence interval.) Nouns with a [closed.open] pattern show the greatest contrast in vowel duration; most of these have the "stem+formative" composition in Figure 24 above, where σ2 is a variant of -*pa* or -*ma*. In Figure 25, the overlap of the notches which represent 95% confidence intervals shows that other syllable structures pattern alike; mean and median values are positive but small. Overall, vowels are longer in σ2, but this is most pronounced in the [closed.open] structure.

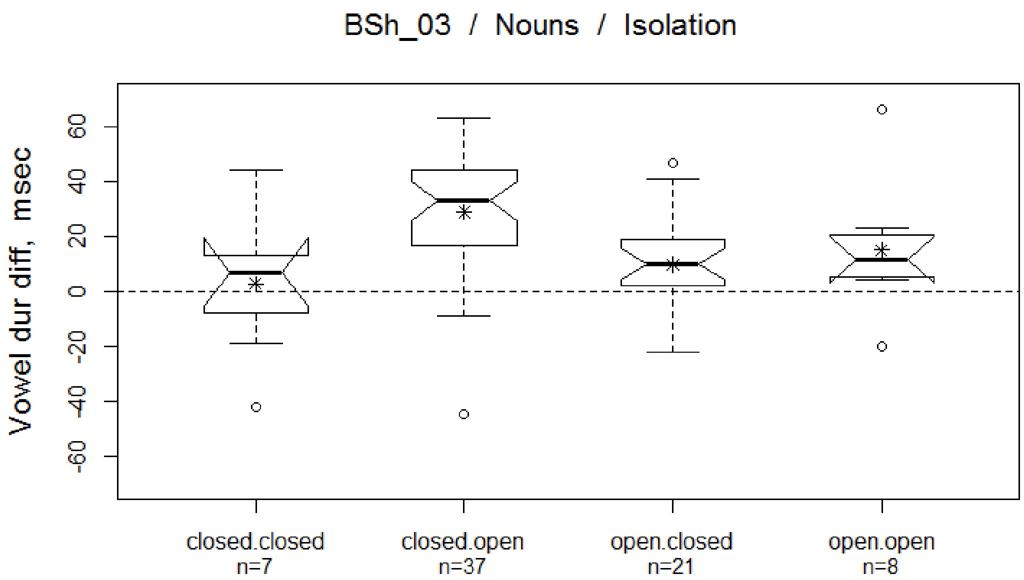


Figure 25. BSh_03 Nouns, isolation: Vowel duration difference vs. syllable closure (words with diphthongs and compensatory lengthening excluded)

Post-hoc paired-sample t-tests confirm that the contrast in vowel duration across syllables is both statistically and perceptually significant only when emphasized by contrasts in syllable structure: that is, with a [closed.open] pattern. (The contrast is statistically significant for nouns with an [open.closed] syllable structure, but the lower 95% confidence limit of 1.6 msec is below the difference limen.) These results demonstrate that, while vowel duration generally corresponds to σ2 stress, the only circumstance in which it is perceptually and statistically significant is when it is exaggerated by the contrast in syllable structure. Thus Speaker BSh_03 does not employ this acoustic resource as a cue for his listener.

Syllable type	t	DF	p-value	Mean diff (msec)	95 % conf. limits	
					lower	upper
closed.closed -I	0.26407	6	0.8006	2.7	-22	28
closed.open -I	**7.9237**	**36**	**2.099e-09**	**29**	**21**	**36**
open.closed -I	2.5126	20	0.02067	9.7	1.6	18
open.open -I	1.7475	7	0.124	15	-5.3	35
all –Isolation	7.0135	72	1.044e-09	19	14	24
all - Frame	8.502	51	2.406e-11	21	16	26

Table 19. BSh_03 / Nouns / Isolation: Duration vs. syllable structure: paired-sample t-tests (two-tailed)

The effect of vowel height contrasts on duration in frame forms is illustrated in Figure 26. For the group of nouns with a [higher.lower] height contrast, duration is slightly longer in σ2 than it is in other groups.

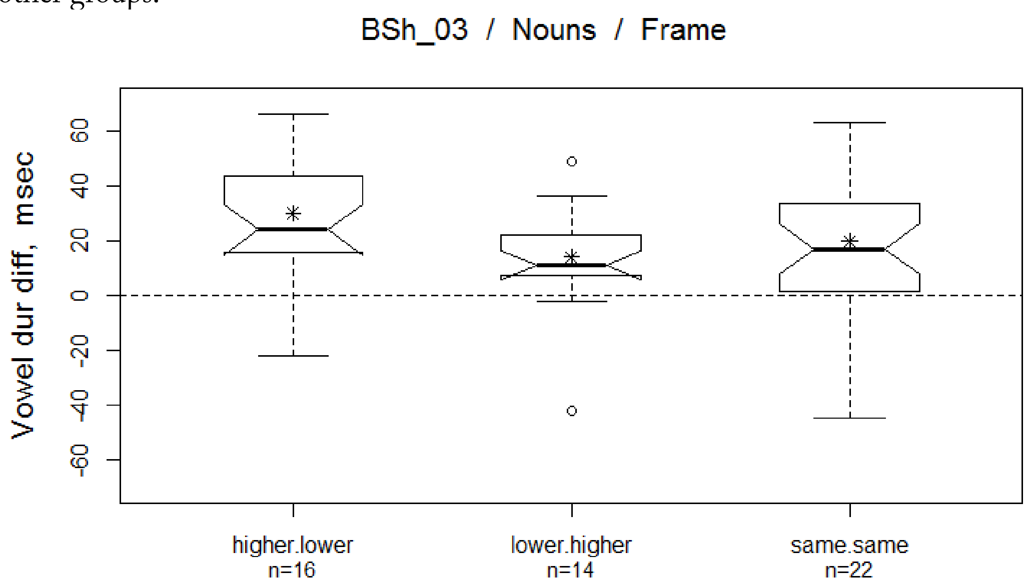


Figure 26. BSh_03 Nouns / frame: Vowel duration difference vs. vowel height contrast (words with diphthongs and compensatory lengthening excluded)

Post-hoc paired-sample t-tests confirm that, for the frame forms, the contrast in vowel duration is statistically and perceptually significant only when emphasized by contrasts in vowel height – i.e., in nouns with a [higher.lower] height pattern. The mean difference for this subset is 30 msec, with a lower 95% confidence limit of 21 msec.

Vowel height	t	DF	p-value	Mean diff (msec)	95 % conf. limits	
					lower	upper
higher.lower - F	**7.3159**	**15**	**2.543e-06**	**30**	**21**	**39**
lower.higher -F	3.7542	13	0.002408	14	6.0	22
same.height - F	4.6922	21	0.0001244	20	11	28
all –Frame	8.502	51	2.406e-11	21	16	26
all - Isolation	7.0135	72	1.044e-09	19	14	24

Table 20. BSh_03 / Nouns: Duration vs. syllable structure: paired-sample t-tests (two-tailed)

For the frame forms, then, vowel duration shows a weak correspondence with the σ2 stress perceived on nouns, but the only circumstance in which it is perceptually and statistically significant is when it is exaggerated by the contrast in vowel height. Thus speaker BSh_03 cannot be said to employ this acoustic resource as a cue for his listener.

For adjectives and numerals, there are too few measurements to evaluate statistically, and there is no evidence to indicate or counter-indicate a correlation between vowel duration and perceived σ2 stress.

Adjectives are plotted in the left panel in Figure 27 below.[30] Tokens fall to either side of the dashed line, with a longer vowel sometimes in σ1 and sometimes in σ2. The three isolation and frame forms which plot left of or close to the dashed line have an [-r] coda in σ1 ([mar.'pʰo] *dmar.po* 'red', [kar.'pʰo] *dkar.po* 'white', and [sar.'pʰa] *gsar.pa* 'new'), which may have a phonological lengthening effect on the preceding vowel.[31] All of the other adjectives have a [closed.open] structure, which may contribute intrinsically to a longer vowel in σ2.

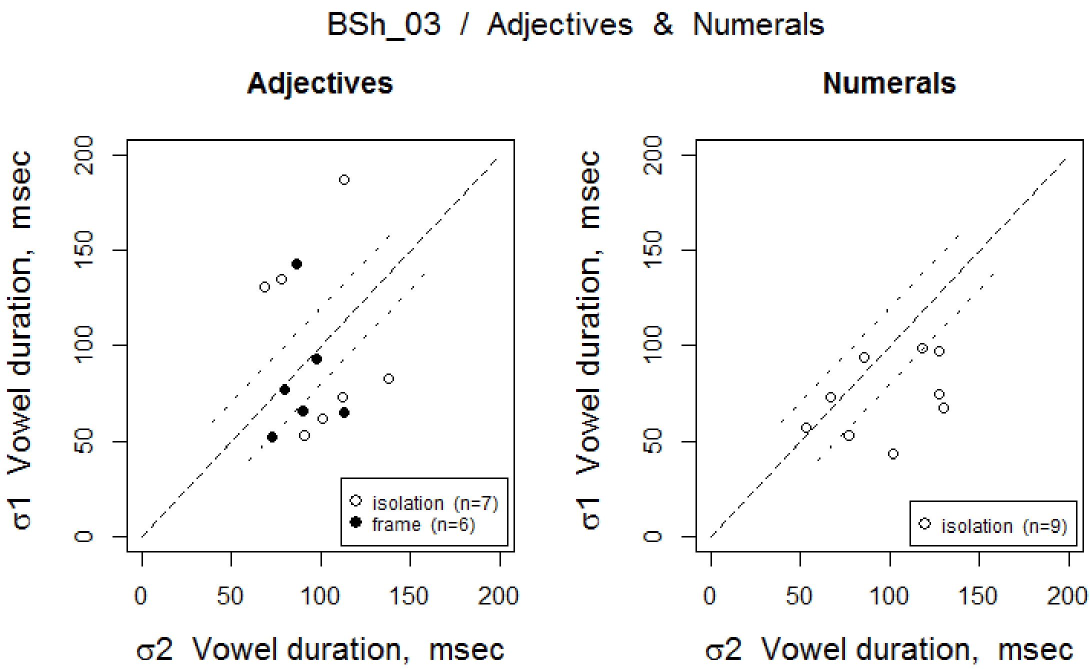


Figure 27. BSh_03 / Adjectives & Numerals/ Vowel duration.

For numerals (produced only in isolation) vowel duration is generally greater in σ2. The few points which fall to the left of the dashed line – [tɕʊχ.'sum] 'thirteen', [tɕu.'ruk] 'sixteen', and [tɕop.'gyɛt] 'eighteen' – are not distinguished from the other tokens in terms of contrasts in syllable structure or vowel height.

[30] [maː̃.'mo] *mang.po* 'many', the one token with compensatory lengthening in σ1, has been omitted. This is the only adjective with an [open.open] syllable structure.

[31] They also all have an aspirated onset in σ2. When segment boundaries are identified in the textgrid, the period of aspiration is assigned to the onset consonant rather than to the nuclear vowel. This may result in a relatively short vowel, compared to words in which the onset consonant is unaspirated.

In verbs, vowel duration contrasts are dominated by the intrinsic effects of vowel height.[32] For isolation forms, the vowel in σ2 is often longer when its vowel is lower (e.g., ['tʰik.pa] *thig.pa* 'to leak': Δduration = +27 msec), as shown in the left panel of Figure 28 below. When vowel height is the same in both syllables, there is a tendency for the vowel in σ1 to be longer (e.g., ['gʊn.ma] *gon.ma* 'to wear': Δduration = -13 msec), as shown in the right panel.[33] This contrast is statistically significant, but the upper 95% confidence limit is only -5 msec (see Table 21 below).[34] We can conclude that vowel duration shows a weak tendency to correspond with the perceived σ1 stress, but is dominated by the effects of vowel height.

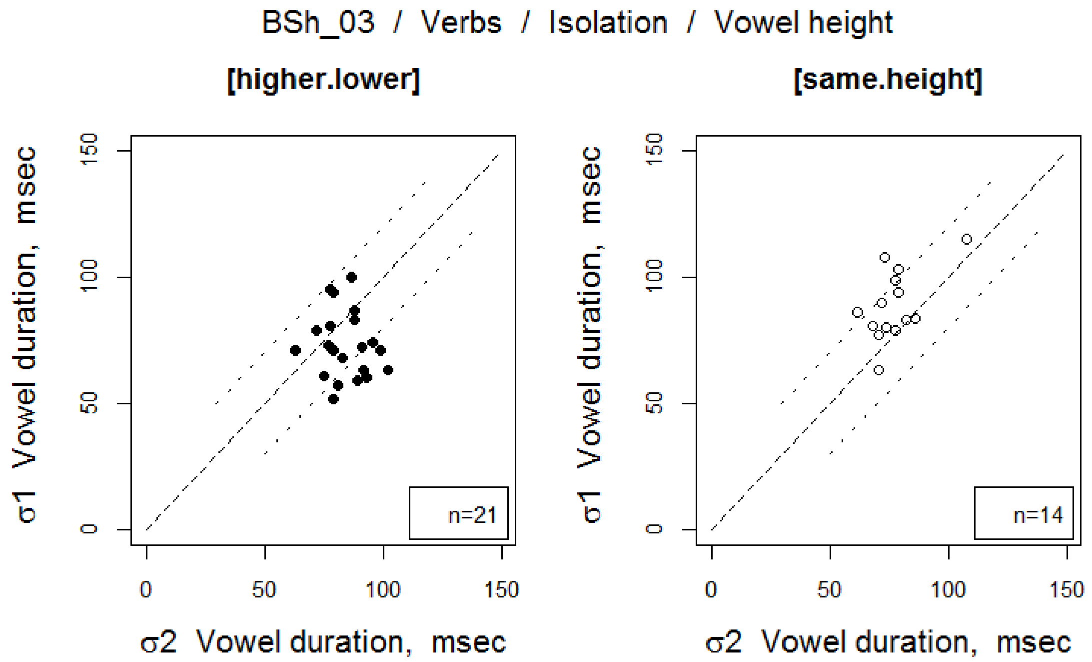


Figure 28. BSh_03 / Verbs / Isolation: Vowel duration *vs.* vowel height contrasts. (Dotted reference line at Δduration = ±20 msec).

In the sentence frame, as shown in Figure 29, when contrasts in vowel height are controlled, vowel duration is randomly distributed about the Δduration = 0 line.

---

[32] Excluded from analysis were words with compensatory lengthening in σ1 (frame forms such as ['straː̆.ma] *srang.ma* 'to straighten' and ['laː̆.ma] *lang.ma* 'to get up, to rise') as well as the lone N+Vblzr form ['zaːn.za] *zan.za* 'to eat food' which has a phonetically long vowel in both isolation and frame forms.

[33] There are no verbs with a [lower.higher] pattern, since citation forms / infinitives end in *-pa* or *-ma*. Syllable structure does not influence vowel duration: with one exception, isolation and frame forms all have a [closed.open] pattern, which may contribute to lengthening of σ2. (Only ['tri.bya] *dri.bya* 'to smell' has an [open.open] pattern.)

[34] Even these small differences in vowel duration may be of interest, as they occur despite factors favoring a longer vowel in σ2: the [closed.open] syllable structure, and the utterance-final position. It is possible that a listener subconsciously taking this phonetic context into consideration might perceive that the speaker does, in fact, manipulate vowel duration to lend prominence to σ1. From this one might conclude that, for verbs produced in isolation, vowel duration is a limited correlate of stress, apparent only in verbs with controlled height. Otherwise the speaker's use of vowel duration is obscured by contrasts in vowel height.

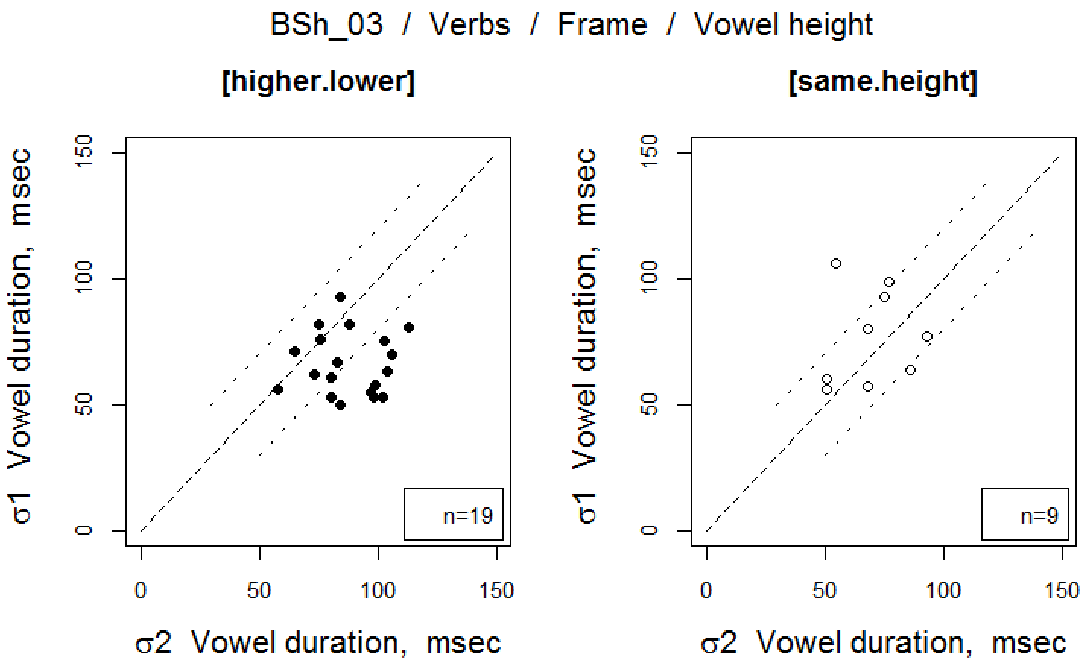


Figure 29. BSh_03 / Verbs / Frame: Vowel duration *vs.* vowel height contrasts. (Dotted reference line at Δduration = ±20 msec).

A one-way ANOVA confirms that vowel height is a main effect (for isolation forms F=18.15, p=0.00016; for frame forms F=12.38, p=0.00162). As anticipated based on the plots above, paired-sample t-tests show that the contrast in vowel duration across syllables is never both statistically significant and perceptually significant for any subset. The mean differences are all <15 msec.

Setting	Vowel height	t	DF	p-value	Mean diff (msec)	95 % conf. limits	
						lower	upper
Isolation	higher.lower	2.9639	20	0.007674	11	3.3	19
	same ht	-3.5824	13	0.003342	-12	-18	-4.6
Frame	higher.lower	0.61696	18	0.545	5.7	-14	25
	same ht	0.40785	8	0.6941	5.3	-25	35

Table 21. BSh_03 / Verbs: Duration vs. vowel height: paired-sample t-tests (two-tailed)

5.4.2 *Vowel Duration for Speaker BM_01*

For speaker BM_01, all words were recorded in isolation only. Vowel duration does not serve as an acoustic correlate of the σ2 stress perceived in nouns. After contrasts in syllable structure are taken into consideration, vowel duration is shown to vary randomly, and the mean difference across syllables is negligible. Vowel duration is not a correlate of stress for adjectives, either.

In contrast, vowel duration does function as a correlate of the σ1 stress perceived in verbs. For those with a N+Vblzr structure, the correlation is dramatic and robust. For citation forms, the statistical significance is weakened by a contrast in vowel height in several tokens. But it is noteworthy that duration is generally longer in σ1 despite a phonetic context which favors longer vowels in σ2.

As shown in Figure 30 below, vowel duration was greater in σ2 for many nouns (especially those with a stem+formative structure, ending in a variant of *-pa* or *-ma*) and greater in σ1 for others (especially compounds).

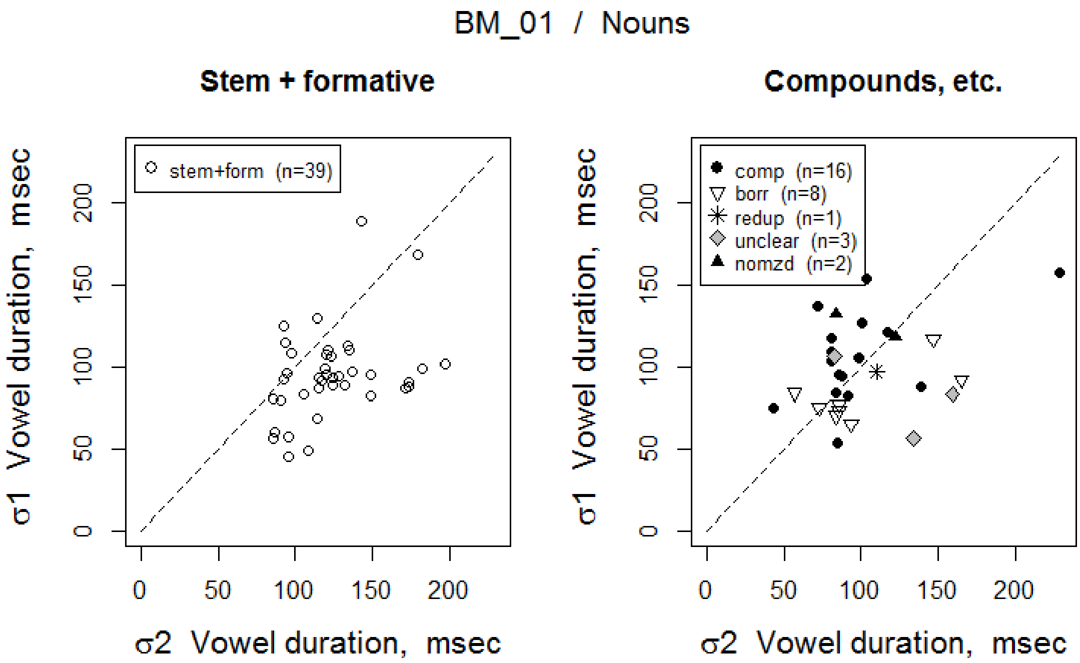


Figure 30. BM_01 / Nouns: Vowel duration

Results of a two-way ANOVA, summarized below, demonstrate that syllable structure is a main effect for vowel duration in isolation forms (F=6.33, p<<0.05), but vowel height is not (F=1.63, p=0.20).

		DF	Sum squares	Mean squares	F value	Pr (>F)
Isolation	Syllable structure	3	20032	6677	6.327	0.000885 ***
	Vowel height	2	3436	1718	1.628	0.205298
	Interaction	6	8847	1474	1.397	0.231635
	Residuals	57	60153	1055		

Table 22. BM_01 / Nouns / Duration: Vowel duration vs. syllable structure and vowel height; Two-way ANOVA.

As shown in the plot below, when σ2 is open, its vowel is longer, regardless of the structure of σ1; when σ2 is closed, vowel duration is more or less evenly distributed across the dashed "duration difference = 0" line – a random distribution.

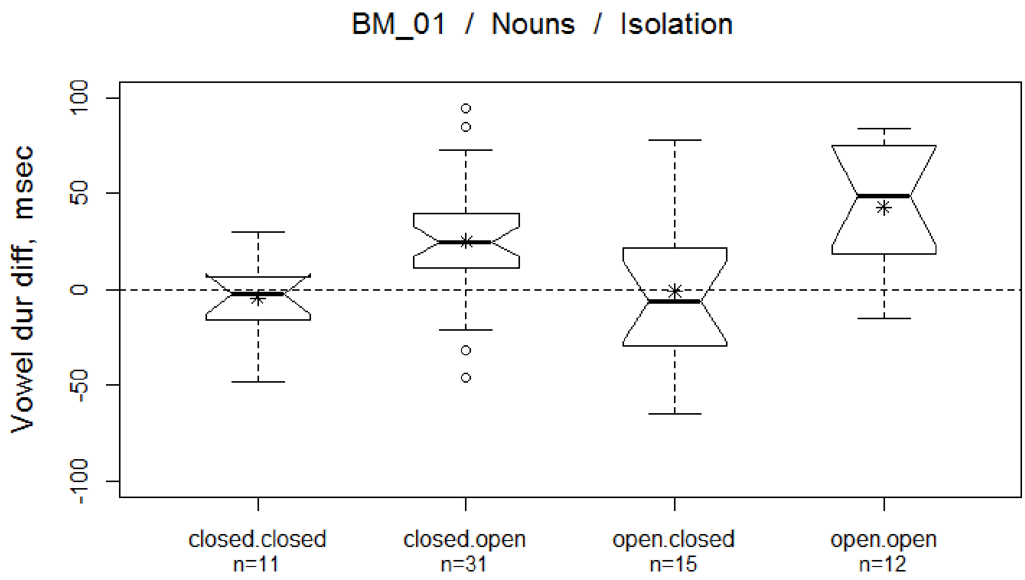


Figure 31. BM_01 Nouns, isolation: Vowel duration difference vs. syllable closure (words with diphthongs and compensatory lengthening excluded)

Conflating these four categories into two, post-hoc paired-sample t-tests comparing measurements across syllables confirm that the contrast in vowel duration is statistically and perceptually significant only in those with σ2 open. Without this enhancing effect, there is no meaningful difference in vowel height. Thus there is no evidence that speaker BM_01 controls vowel duration to convey stress in nouns.

Syllable type	t	DF	p-value	Mean diff (msec)	95 % conf. limits	
					lower	upper
σ2 open	**6.0623**	**42**	**3.232e-07**	**30**	**20**	**40**
σ2 closed	-0.34492	25	0.733	-2.3	-16	12
all	4.0547	68	0.0001316	18	9	27

Table 23. BSh_03 / Nouns: Duration vs. syllable structure: paired-sample t-tests (two-tailed)

Vowel duration is not utilized as a correlate of stress for adjectives, either. As shown in Figure 32 below, vowels are sometimes longer in σ1 and sometimes longer in σ2.

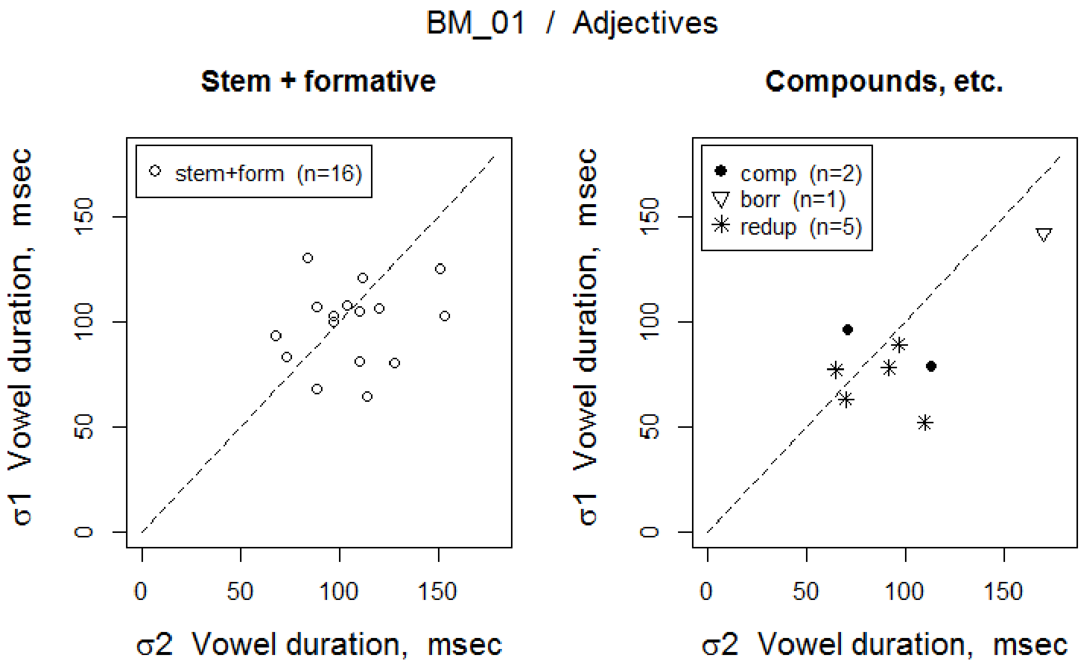


Figure 32. BM_01 / Adjectives: Vowel duration

A two-way ANOVA, summarized in Table 24 confirms that this distribution is not affected by either syllable structure or vowel height (p>0.05). A paired-sample t-test (Table 25) shows that the difference in mean duration between σ1 and σ2 is neither statistically significant (p>0.05) nor perceptually significant. The mean duration difference is 10 msec, which is barely discriminable, and the 95% confidence interval spans zero.

		DF	Sum squares	Mean squares	F value	Pr (>F)
Isolation	Syllable structure	3	2223	740.9	1.055	0.394
	Vowel height	2	2638	1319.1	1.878	0.183
	Interaction	1	87	86.9	0.124	0.729
	Residuals	17	11943	702.5		

Table 24. BM_01 / Adjectives / Duration: Vowel duration vs. syllable structure and vowel height; Two-way ANOVA.

	t	DF	p-value	Mean diff (msec)	95 % conf. limits	
					lower	upper
Adjectives	1.7626	23	0.09125	9.9	-1.7	21

Table 25. BM_01 / Adjectives: Paired-sample t-tests (two-tailed). (Words with compensatory lengthening omitted.)

For verbs produced by speaker BM_01, there is generally a strong correlation between vowel duration and the perceived σ1 stress. As shown in Figure 33, verbs fall to the left of the dashed line (with only two exceptions[35]); many also fall to the left of the -20 msec and -40 msec reference lines.

With all of the variation in syllable closure type and vowel height, it is not possible to identify a control group for either morphological type of verb. It would certainly be useful to evaluate a larger sample of verbs in future, but even with this limited sample, it is clear that vowels are indeed mostly longer in σ1, the stressed syllable.

For the N+Vblzr form ['skʰat.zɛr̥] *skad.zer* 'to call, to shout', the difference in duration across syllables is only -8 msec, which is below Lehiste's minimal difference limen of 10 msec. For all of the other N+Vblzr forms, the difference ranges from -46 msec to -146 msec.

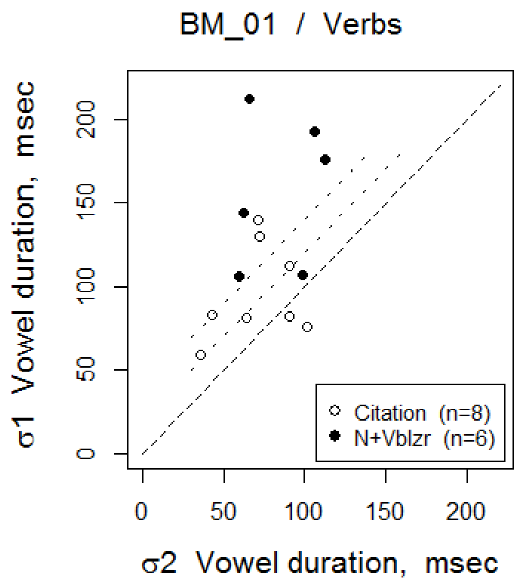


Figure 33. BM_01 / Verbs: Vowel duration (Dotted reference lines at Δdur = -20 msec and -40 msec).

For the N+Vblzr forms, a paired-sample t-test shows that this correlation is statistically significant ($p < 0.05$, as shown in Table 26 below). For the citation forms, the contrast in duration across syllables is *not* statistically significant ($p > 0.05$); however, this pattern is still of note, given the phonetic context. There are three factors that favor a longer vowel in σ2: (a) utterance-final position; (b) a [closed.open] syllable structure; and (c) a [higher.lower] vowel height contrast. Despite these, vowels are nonetheless generally longer in σ1. Thus even a small contrast indicates that the speaker is manipulating vowel duration to override these intrinsic effects, and to instead correspond with σ1 stress.

[35] These are ['dʑik.pʰa] *'jigs.pa* 'to fear' and ['mɪn.ma] *min.ma* 'to give'. The intrinsic effects of vowel quality likely play a role here; vowels are shorter in high vowels than in low vowels, and these are the only two verbs with [i] or [ɪ] in σ1, contrasting with [a] in σ2.

Speaker	Subset	t	DF	p-value	Mean diff (msec)	95 % conf. limits	
						lower	upper
BM_01	N+Vblzr	-3.8206	5	0.01236	-72	-120	-24
	citation	-2.1448	7	0.06914	-24	-51	2.5

Table 26. BM_01 / Verbs / Vowel duration: Results of paired-sample t-tests (two-tailed)

6 Summary and Implications

In Balti Tibetan, disyllabic nouns, adjectives, and numerals (non-verbs) are stressed on the second syllable, σ2. For the two speakers whose recordings I analyzed acoustically, this σ2 stress is conveyed primarily by fundamental frequency (F0). Vowel duration plays a weaker and inconsistent role as an acoustic correlates of stress, while intensity is not a factor. Verbs, in contrast, are stressed on the first syllable, σ1. Syllable prominence is conveyed by all three acoustic parameters considered: F0, intensity, and vowel duration. These findings differ from previous descriptions of Balti in distinguishing stress patterns by lexical category. Further, this is the first work to provide an acoustic characterization of the correlates of stress in Tibetan.

As Balti is one of the most phonologically archaic varieties of Tibetan, its prosodic and acoustic characteristics can be regarded as conservative reflexes of Proto-Tibetan. This study thus contributes vital information towards the prosodic reconstruction of Proto-Tibetan and Proto-Tibeto-Burman. Hypotheses about Tibetan tonogenesis must be refined to take into consideration the transphonologization of stress to tone.

References

Backstrom, Peter Charles. 1994. *A phonological reconstruction of proto-western Tibetan*. MA thesis, The University of Texas at Arlington.

Bartee, Ellen. 2000. "A note on verbalizers in Lhasa Tibetan". In: Bartee, Ellen; and Dromo, Nyima, *A beginning textbook of Lhasa Tibetan,* 193-196. Beijing: National Press for Tibetan Studies.

Bielmeier, Roland.1985. *Das Märchen Vom Prinzen Čobzan. Eine Tibetische Erzählung Aus Baltistan. Text, Übersetzung, Grammatik Und Westtibetisch Vergleichendes Glossar.* Sankt Augustin: VGH Wissenschsaftsverlag.

Bielmeier, Roland. 1988. "On tone in Tibetan". In: Uebach, Helga; and Panglung, Jampa L. (eds.), *Tibetan studies: Proceedings of the 4th Seminar of the International Association for Tibetan Studies*. Munich: Kommission für Zentralasiatische Studien, Bayerische Akademie der Wissenschaften.

Bielmeier, Roland; Häsler, Katrin; Haller, Felix; Huber, Brigitte; and Volkart, Marianne. 2002a. *The Comparative dictionary of Tibetan dialects (Third Preprint). Vol 1.1: Verbs*. University of Berne: Tibetan Dialects Project.

Bielmeier, Roland; Häsler, Katrin; Haller, Felix; Huber, Brigitte; and Volkart, Marianne. 2002b. *The comparative dictionary of Tibetan dialects (Fourth Preprint). Vol 2.1: Nouns*. University of Berne: Tibetan Dialects Project.

Bielmeier, Roland; Haller, Felix; Hässler, Katrin; Huber, Brigitte; and Volkart, Marianne. 2008. "A short guide to the *Comparative dictionary of Tibetan dialects* (CDTD)". Unpublished manuscript. University of Bern.

Boersma, Paul. 2001. "Praat, a system for doing phonetics by computer". *Glot International* 5.9-10: 341-345.

Boersma, Paul; and Weenink, David. 2016. "Praat: doing phonetics by computer". [Computer program]. Retrieved 2016 from http://www.praat.org/.

Caplow, Nancy J. 2009. *The role of stress in Tibetan tonogenesis: A study in historical comparative acoustics*. PhD dissertation, University of California Santa Barbara.

Caplow, Nancy J. 2016. "Reconstructing stress in Proto-Tibetan: Evidence from Balti and Rebkong Amdo". *Linguistics of the Tibeto-Burman Area* 39.2: 180-221.

Denwood, Philip. 1999. *Tibetan*. Amsterdam, Philadelphia: John Benjamins.

Goldstein, Melvyn C. 2001. *The new Tibetan-English dictionary of modern Tibetan*. University of California Press.

Goldstein, Melvyn C. and Narkyid, Ngawang Thondup. 1984. *English-Tibetan dictionary of Modern Tibetan*. Berkeley: University of California Press.

Grierson, George A. 1909. *Linguistic Survey of India, Vol. III, Part I. Tibeto-Burman family: Tibetan dialects, the Himalayan dialects, and the North Assam group*. Delhi: Motilal Banardsidas.

Jaeschke, Heinrich A. 1871. *Handwörterbuch der tibetischen Sprache*. Gnadau: Universitätsbuchhandlung.

Jaeschke, Heinrich August. 1958 [1881]. *A Tibetan-English dictionary with special reference to the prevailing dialects; to which is added a Tibetan-English vocabulary*. London: Kegan Paul.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge MA: MIT Press.

Lewis, M. Paul; Simons, Gary F.; and Fennig, Charles D. (eds.). 2016. *Ethnologue: Languages of the world, nineteenth edition*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Norberg-Hodge, Helena; and Paldan, Gyelong Thupstan. 1991. *Ladakhi-English English-Ladakhi dictionary*. Leh, Ladakh: The Ladakh Ecological Development Group and The Ladakh Project.

Read, A.F.C. 1934. *Balti grammar*. London: The Royal Asiatic Society.

Róna-Tas, Andras. 1966. *Tibeto-Mongolica: The Tibetan loanwords of Monguor and the development of the Archaic Tibetan dialects*. The Hague: Mouton [Indo-Iranian Monographs VII].

Sprigg, R.K. 1966. "Lepcha and Balti Tibetan: Tonal or non-tonal languages?". *Asia Major* 12: 185-201.

Sprigg, R.K. 2002. *Balti-English English-Balti dictionary*. New York: RoutledgeCurzon.

Sun, Jackson T.-S. 2003. "Phonological profile of Zhongu: A new Tibetan dialect of northern Sichuan". *Language and Linguistics* 4.4: 769-836.

Tibetan and Himalayan Library. http://www.thlib.org/reference/transliteration/#!essay=/thl/ewts

Tournadre, Nicolas. 2005. "L'aire linguistique tibétaine et ses divers dialectes". Lalies 25 : 7-56.

Tournadre, Nicolas. 2008. "Arguments against the concept 'Conjunct' / 'Disjunct' in Tibetan". In: Huber, Brigitte; Volkart, Marianne; and Widmer, Paul (eds.), *Chomolangma, Demawend und Kasbek: Festschrift für Roland Bielmeier zu Seinem 65. Geburtstag. Band I: Chomolangma*, 281–308. Halle (Saale): International Institute for Tibetan and Buddhist Studies (IITBS) GmbH.

Tournadre, Nicolas. 2013. "The Tibetic languages and their classification". In: Owen-Smith, Thomas and Hill, Nathan (eds.), *Trans-Himalayan linguistics: Historical and descriptive linguistics of the Himalayan area*, 105-130. Berlin: Mouton de Gruyter.

Wylie, Turrell V. 1957. "A standard system of Tibetan transcription". *Harvard Journal of Asiatic Studies* 22: 261–267.

Zemp, Marius. 2013. *A historical grammar of the Tibetan dialect spoken in Kargil (Purik)*. PhD dissertation, University of Berne.

Nancy J. Caplow
nancy.caplow@okstate.edu