## Abstract

Characterizing prosodic prominence relations in African tone languages is notoriously difficult, as typical acoustic cues to prominence (changes in F0, increases in intensity, etc.) can be difficult to distinguish from those which mark tonal contrasts. The task of establishing prominence is further complicated by the fact that tone, an important cue to syllable prominence and prosodic boundaries cross-linguistically, plays many roles in African languages: tones often signal lexical contrasts, can themselves be morphemes, and can also interact in important ways with prosody. The present study builds on phonological generalizations about tonal patterns in Medʉmba, a Grassfields Bantu language, and uses the speech cycling paradigm (Cummins & Port 1998) to investigate relative timing of syllables varying in phrase-level prominence. Specifically, we investigate timing asymmetries between syllables hypothesized to occur at the edge of a phonological phrase, which carry a high phrase accent, and those in phrase-medial position, which do not. Results indicate significant differences in the temporal alignment of accented vs. non-accented syllables, with accented syllables occurring significantly closer to positions established as prominence-attracting in previous speech cycling research. We show that these findings cannot be attributed to differences in tone alone. Findings demonstrate the importance of relative temporal alignment as a correlate of prosodic prominence. Findings also point to increased duration as a phonetic property which distinguishes between syllables bearing phrasal prominence from those which do not.

## 1. Introduction

The notion of relative syllable prominence has played an important role in theoretical work in phonology on Bantu tone languages dating back several decades. In particular, a number of studies have posited a role for *accent* in explaining the distribution of tone and other suprasegmental phenomena in Bantu languages (Clements & Goldsmith 1984; Goldsmith 1987; Hyman 1978). Importantly, different researchers have used the term 'accent' in fundamentally different ways. While some have followed Bolinger (1958), for example, in using the term to refer to phonetically-measurable prominence at the word or phrase level (Hyman 1978; Patin 2016), others have used the term in reference to a structural property of syllables or words (e.g., Bickmore, 1995; Clements & Ford 1979; Ford 1975, 1976; Goldsmith 1987; McHugh 1990, 1999).[1] In the latter case, accent is treated as an abstract phenomenon (for some, such as Goldsmith 1987, associated with metrically prominent positions; for others, such as Clements and Ford 1979, determined lexically) and something which may exist independently of any particular acoustic correlate. For the sake of clarity, this notion of accent will be referred to henceforth as 'abstract accent'. For now, however, our focus will be on describing the behavior of a particular phonetic pattern which on its surface is more consistent with Bolinger's definition (though we will also discuss a possible role for abstract accent in Section 2). Specifically, the term 'accent' here will be conceptualized according to the notion of *phrase accent* (Pierrehumbert 1980; Pierrehumbert & Beckman 1988; Grice et al. 2000), or an edge-marking tone occurring at the right edge of a prosodic unit which is smaller than the intonational phrase. Note that this specific term is chosen due to the fact that the phrase accent is an intonational phenomenon whose distribution has been shown to be sensitive both to prosodic boundaries and to metrical prominence (Grice et al. 2000); we will provide evidence that the phrase accent in Medʉmba is sensitive to both of these properties. The phrase accent of interest in this paper will also be variably referred to descriptively as a 'high preverbal phrase accent'. The term 'tone', when used on its own, will refer to the use of F0 for signaling lexical or morphological distinctions, independent of accent. The term 'stress' will be used to refer to phonetically-measurable prominence at the word level. As will be discussed, there are a number of cases in which phonetically-based and abstract accent accounts seem to converge on a similar analysis; for this reason, the term ‘prominence’ will be recruited as a cover-all term for both types of phenomena. Finally, the term ‘rhythm’ will be used to refer to temporally-predictable occurrences of prominent syllables, as in the context of the speech cycling experiment to be described in Section 3.

Evidence for syllable prominence in Bantu tone languages has emerged from a number of sources, including patterns of vowel lengthening and tone distribution. Overwhelmingly, prominence in Niger-Congo languages has been argued to be located either on the initial or penultimate syllable within a word or stem (see Downing 2010 for overview). For example, several Eastern and Southern Bantu languages display increased vowel duration on the penultimate syllable of stems (see Hyman 2014 for overview), which has been interpreted by several researchers (e.g. Cole 1955; Doke 1938; Downing 2010; Hyman 1978) as evidence of stress. On a structural level, abstract prominence on the penultimate syllable of the stem has also been argued by many researchers to be implicated in patterns of tone shift, whereby a tone which is lexically specified shifts its location to the penultimate position of the stem (Bickmore, 1995; Cassimjee & Kisseberth, 1992, 1998; Clements & Goldsmith 1984; Downing, 1990; Goldsmith, 1991; Hyman, 1989; Kisseberth, 1992; McHugh 1990, 1999; de Lacy, 2002; Philippson 1998; Sietsema, 1989; Volk 2011; see also Downing 2010).[2]

Many researchers have also argued for the role of prosodic structure above the word in determining the location of syllable prominence in Bantu languages. For example, penultimate vowel lengthening as described above has been argued for some languages to be strongest in those cases where a stem or word occurs in final position of a phonological phrase (Downing &

Pompino-Marschall, 2013; Kanerva 1990; Kraal 2005; Truckenbrodt 1996, 1999; Zeller 2015). Other work examining tonal distributions has posited an interaction between lexical tones, phrase accent, and prosodic structure. For example, Patin (2016, 2018) (see also Philippson 2005 and Cassimjee & Kisseberth 1998) shows for Shingazidja, a Bantu language of Comoros, that high phrase accents are inserted on penultimate stressed syllables in phonological phrases where lexical high tones are lacking. He furthermore shows that the right edge of a phonological phrase attracts lexical high tones, and that tones cannot cross over a phonological phrase boundary. This proposal is in line with work by Truckenbrodt (1996) which, based on data from a number of Bantu languages, posits that phonological phrases favor edgemost phrasal prominence.

From a phonetic standpoint, while it is fairly well-accepted that patterns such as vowel lengthening reflect prominence in many Bantu languages, some scholars remain skeptical that tonal patterns should be distinguished from f0-related accentual patterns in these languages. This skepticism stems, in part, from the perception that tone languages, unlike stress-based languages such as those in the Indo-European family, lack phonetic evidence for accent which is distinguishable from that attributed to tonal distinctions alone.[3] Odden (1999) points out the fact that, not only are tone and stress/phrase accent both cued primarily through fundamental frequency (F0), but changes in amplitude, another cue for stress in many Indo-European languages, may also be highly correlated with a syllable's lexical tone across tone languages (Abramson, 1962). Odden concludes that, where evidence lacks to verify the existence of syllable prominence in Bantu languages, a more conservative analysis of tone patterns is one in which prominence is not assumed. Unfortunately, the relative dearth of phonetic studies on Bantu languages (and African languages as a whole) has meant that decisions about the presence or absence of phonetic evidence for prominence are based, in many cases, on impressionistic judgments, typically from non-native speakers of the languages of study. While impressions may serve well enough in cases where cues to stress or phrasal prominence are very conspicuous (as may be the case for penultimate lengthening in some languages), it is quite possible that subtler phonetic differences may exist where non-native listeners may have difficulty perceiving them. It is for just this reason that in-depth phonetic investigation has proven essential in disentangling tone-related and stress/accent-related cues from one another in research on tone languages. For example, careful phonetic analysis of Mandarin Chinese, a language with four lexical tones, has yielded evidence for stress feet for which head-dependent relations are cued through a combination of F0 range and duration (Moore, 1993), as well as tonotactic effects (Yip, 1980, Duanmu, 1990). Presence of contrastive/corrective focus have also been shown to influence both duration and f0 of focused constituents in Mandarin (Chen 2006; Grief 2010). In Thai, a language which also has five tones, Potisuk et al. (1994, 1996) find evidence for word stress such that unstressed syllables have shorter duration, more reduced F0 range, and more reduced vowels. Remijsen (2002) finds that Ma'ya, an Austronesian language spoken in Papua New Guinea, shows evidence of both stress and lexical tone, with stress cued primarily through duration, vowel quality, and spectral balance, and tone cued primarily through F0. Remijsen and Van Heuven (2005) find that Curaçao Papiamentu, a language which has been analyzed as having two tones (Rivera-Castillo & Pickering, 2004), cues stress through a combination of duration, intensity, and vowel quality.

In addition to the acoustic measures described above, another promising approach to the cross-linguistic study of syllable prominence involves examining relative timing of syllables in speech. For example, in their work using the speech cycling paradigm, Cummins and Port (1998) and Tajima (1998) find that in repeated utterances of English phrases such as *Buy the day* *for Guy* and *Buy the 'Today' for Guy*, stressed syllables within each phrase preferentially align with certain intervals of the repetition cycle, namely, lower order fractions of the cycle (such as the halfway point). Thus, alignment of the third syllable in the first sentence (i.e., *day*, a stressed syllable) is more similar to that of the fourth syllable in the second sentence (i.e., the stressed syllable in *Todáy*) than to the third syllable in that sentence (i.e., the unstressed syllable in *Todáy*).

Interestingly, the speech cycling paradigm has been useful in highlighting relative syllable prominence both for languages like English which are believed to be *head-prominent*, or which mark prominence on metrical heads, and those which are *edge-prominent*, or which mark prominence at the edges of prosodic units (Jun 2005; 2014). For example, similar to English, Zawaydeh et al. (2002) found that stressed syllables in Jordanian Arabic, a head-prominent language, were aligned to lower order phase positions in speech cycling. In Korean, an edge-prominent language, Chung and Arvaniti (2013) found that syllables in the initial position of an Accentual Phrase were drawn to lower order phase positions in the task.[4] Finally, for Japanese, a head/edge prominence language (where prominence is marked both on metrical heads and edges of prosodic units), Tajima (1998) found that both foot-initial syllables and pitch accented syllables were attracted to lower order phase positions in speech cycling.

Given that speech timing is not a variable which is prone to the confounds associated with acoustic correlates of tone, stress, and phrase accent, it makes for an especially appealing domain in which to investigate relative syllable prominence in Bantu. Furthermore, examining speech in the context of the speech cycling task provides us with a window into the intuitions of speakers about how their language can be rhythmically organized. This study therefore attempts to investigate the timing behavior of syllables hypothesized to bear phrase-level prominence in

Medʉmba, a Grassfields Bantu language of Cameroon. Before describing the details of the study itself, Section 2 describes the distribution of a particular high ‘tone’ which seems to support its status as a phrase accent—in particular, a high preverbal phrase accent—rather than a tone, proper. We hypothesize that this phrase accent serves to mark syllables which bear phrase-level prominence; as a result, these syllables should show specific prominence-related timing patterns in speech cycling. We then present results from a duration analysis which provide further corroboration for the phrase accent as a marker of phrase-level prominence.

## 2. Distribution of the High Preverbal Phrase Accent in Medʉmba

This section presents a phonological analysis of a high tone in Medʉmba that systematically occurs at the end of a preverbal prosodic phrase. We will refer to this tone throughout the paper as the *high preverbal phrase accent* and investigate the role of this tone with respect to prominence in applying the speech cycling paradigm. An example of this tone is provided in (1), where the H tone is posited to occur on the phrase-final negation marker [kə́] in (1a), but not the phrase-medial negation marker [kə̀] in (1b); words to be compared are marked in boldface.[5,6]

(1) a. High/Accented

[ Mvɛ́l ù **kə́** ] [ ↓zwí ] [ ↓ʃún àm **]**

Brother 2.SG.POSS NEG hurt friend 1.SG.POSS

‘Your brother won’t hurt my friend’

b. Low/Unaccented

[Nʧó ] [↓lù **kə̀** kəʔ ] [ ↓mvɛ́l àm ]

Thief RPST. NEG. cut brother 1.SG.POSS

‘The thief didn’t cut my brother.’

Such distributional patterns lead us to analyze the H tone on [kə́] in examples like (1a) as a phrase accent in the sense of Grice et al. (2000). In the paragraphs that follow, relevant background on the Medʉmba language and the linguistic factors which constrain the location of the high preverbal phrase accent will be described in order to give context for interpreting the experimental approach and results to come in Sections 3 and 4.

To begin with, Medʉmba is spoken predominantly in the West Region of Cameroon by approximately 200,000 speakers (Ethnologue, 2017). Words in the language are largely monosyllabic, and there is a two-way tonal contrast including high and low tones (Voorhoeve, 1971). While most morphemes are specified for an underlying tone or tonal melody, some morphemes are toneless. The language also features downstep, a process by which the second of two high tone words in a sequence will have its pitch lowered relative to the first word. This process is quite common across Bantu languages, and, as in other languages, appears to be prosodically conditioned.[7] Specifically, downstep occurs at the boundary of a prosodic unit of a size consistent with an intermediate phrase (Beckman & Pierrehumbert 1986) or, from a syntactically-based perspective, with the phonological phrase[8]. The phonological phrase has been shown to correspond roughly in size to a syntactic XP (Nespor & Vogel, 1986; Selkirk, 1986, 1995; Truckenbrodt, 1996, 1999). In example (2a), downstep is indicated with downward arrows between the subject noun phrase and verb, and between the verb and its object (see also the associated pitch track in Figure 1). All three words and the pronominal clitic are underlyingly high toned. In (2b), the syntactic phrasing for this sentence is provided, and in (2c), its prosodic phrasing is provided (including phonological phrase and intonational phrase level bracketing). As shown in (2c), the object noun, embedded within the larger verb phrase in (2b), is parsed as its own separate phonological phrase in (2c).

| | | | | |
|---|---|---|---|---|
| (2) | a. | ʃún=ʧɑ́m | ↓jə́n | ↓mɛ́n |
| | | friend =1POSS.CLII9 | see | child |
| | | 'My friend has seen the child.' | | |

b. [ [ ʃún=ʧám ]$_{NP}$ [ ↓jə́n [ ↓mɛ́n]$_{NP}$ ]$_{VP}$ ]$_{TP}$

c. [ ( ʃún=ʧám )$_{PPh}$ ( ↓jə́n )$_{PPh}$ ( ↓mɛ́n )$_{PPh}$ ]$_{IPh}$[10]

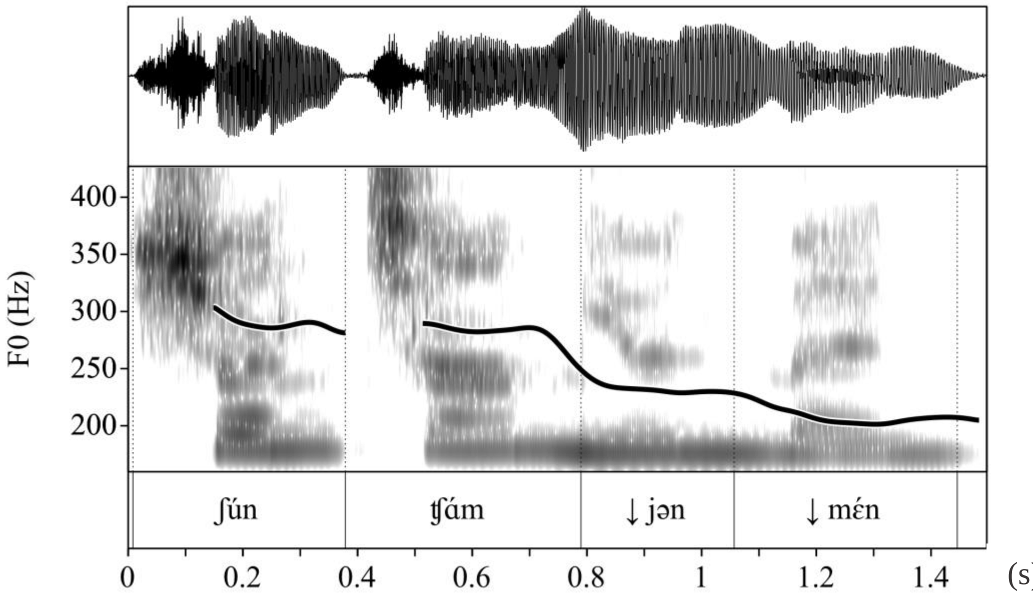


Figure 1: Pitch track for example (2a) shows downstep occurring on successive phonological phrases

Another aspect of Medʉmba phonology which will be important to understand moving forward is the presence of a high ‘tone’—what we are referring to as the high preverbal phrase accent--which occurs within the preverbal domain. As we will show, the patterning of the high preverbal phrase accent indicate that its distribution is constrained both at the level of the phonological phrase and the intonational phrase. Specifically, we will propose that it occurs at the right edge of the first phonological phrase within an intonational phrase.

In (3), we see that the high preverbal phrase accent occurs with the rightmost element in the subject noun phrase. In (3a), a low-toned noun *mvə̀n* ‘chief’ occurs in subject position, and the high preverbal phrase accent associates with its right edge, resulting in a low-high rising tone. In (3b), we see the same tone dock to the low toned pronoun *ɑ̀m* (note that the tone of this pronoun differs based on the class of its head noun), again resulting in a low-high rising tone. The same pronoun does not receive a rising contour in final position of any of the sentences in (3), indicating the absence of any phrase accent. Finally, if, as in (3c), a toneless pronoun (in this case, the first person singular *mə*) occurs in subject position, the high preverbal phrase accent will dock to it, resulting in a simple high tone on the subject. Pitch tracks for (3a-c) showing different positioning of the high peak associated with the preverbal phrase accent are shown in Figure 2.

It should be noted that the high preverbal phrase accent does not appear to be a lexical high tone, as it does not consistently surface with any one specific morpheme or class of morphemes. Furthermore, its effects are present in a variety of different tenses and modalities, and it does not appear to be straightforwardly associated with any particular morphosyntactic function. Rather, the tone appears to behave as a phrase accent, introduced by the prosodic structure as a marker of a phonological phrase edge.

(3) High preverbal phrase accent docks to the right edge of subject

| | | | | | |
|---|---|---|---|---|---|
| a. | mvə̀n | Mvə̌n | ↓jə́n | ↓mɛ́n=ɑ̀m | |
| | chief | chief.**H** | see | child=1POSS.CLIV | |
| | | ‘The chief has seen my child.’ | | | |
| b. | ɑ̀m | Mvə̀n=ɑ̌m | ↓jə́n | ↓mɛ́n=ɑ̀m | |
| | 1POSS.CLIV | chief 1POSS.CLIV.**H** | see | child=1POSS.CLIV | |

‘My chief has seen my child.’

c. mə Mə́ ↓jə́n ↓mɛ́n=àm

1SG 1SG.**H** see child=1POSS.CLIV

‘I have seen my child’

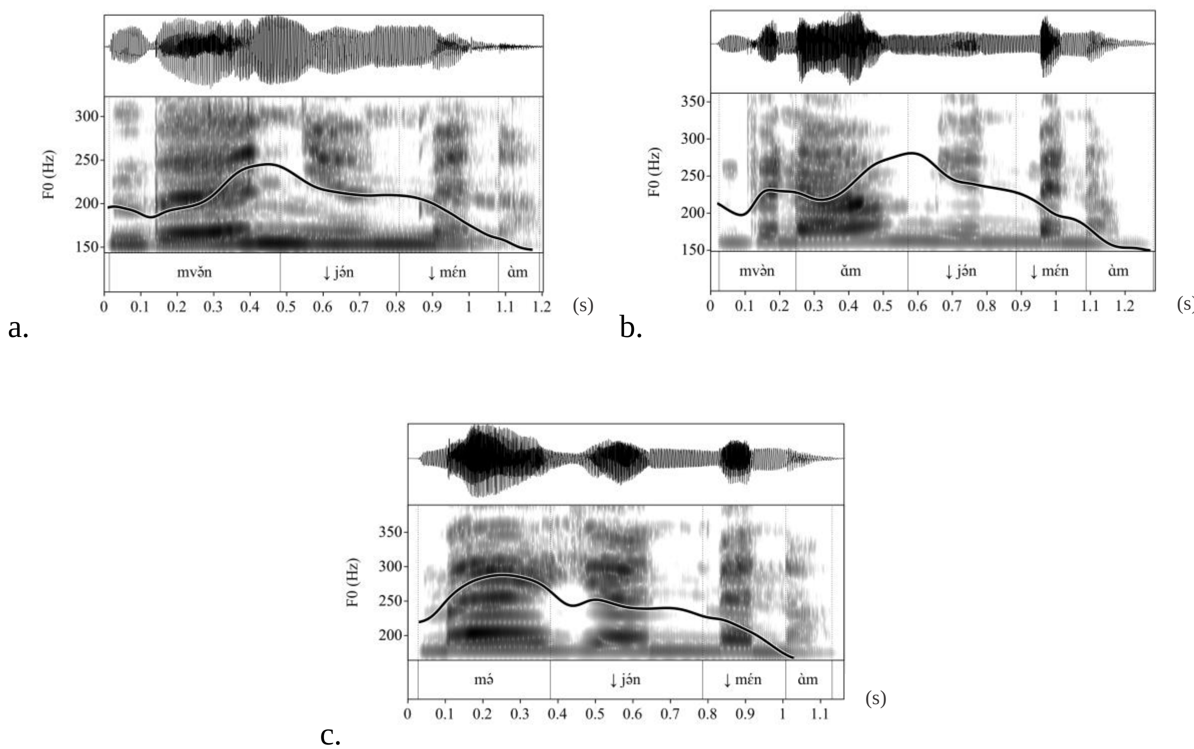


Figure 2: Pitch tracks show variation in position of the high peak associated with preverbal phrase accent for sentences in examples 3a-c

Further inspection of the distribution of the high preverbal phrase accent shows that it is not just the subject noun phrase which can host it. In certain tenses, the negation marker, which we analyze as being underlyingly toneless, can bear the high preverbal phrase accent. In (4a), in the negated future tense, we see that the noun phrase surfaces with all low tones, and the high preverbal phrase accent is assigned to the negation marker. A similar pattern is found in the negated unmarked recent past tense. In (4b), which is the negated version of (3b), the tone also associates to the negation marker, though the marker also bears a final glottal stop.

(4) High preverbal phrase accent docks to negation markers

a. Mvə̀n= ɑ̀m kə́ ↓jə́n ↓mɛ́n=ɑ̀m

chief=1POSS.CLIV NEG.FUT.**H** see child=1POSS.CLIV

'My chief will not see my child.'

b. Mvə̀n=ɑ̀m kə́ʔ ↓jə́n ↓mɛ́n=ɑ̀m

chief=1POSS.CLIV NEG.PST.**H**[11] see child=1POSS.CLIV

'My chief has not seen my child.'

One final distributional generalization is that if a tense marker, such as the low tone recent past tense marker *fə̀*, intervenes between the subject noun phrase and the negation marker, the high preverbal phrase accent will not be able to dock to the negation marker, and instead attaches at the right edge of the subject. In this case, the negation marker receives a default low tone (5).

(5) High preverbal phrase accent docks to a possesive pronoun

Mvə̀n=ɑ̌m ↓fə̀ kə̀ jə́n ↓mɛ́n

chief=1POSS.CLIV.**H** RPST NEG.RPST see child

'My chief did not see the child.'

It therefore seems that the negation marker, when no tense marker is present, can form a prosodic constituent with the subject such that it becomes a candidate for hosting the high preverbal phrase accent. When a tense marker is present, only the subject is contained within this constituent.[12] A possible analysis of these patterns would be to say that the negation marker, when it occurs in the future tense (4a) or in the recent past/perfect (4b), is parsed with the subject as part of the same phonological phrase (6).

(6) Prosodic phrasing for negated unmarked past and future tenses

a. ( Mvə̀n=ɑ̀m  kə́ )$_{\text{PPh}}$  (↓jə́n )$_{\text{PPh}}$  (↓mɛ́n )$_{\text{PPh}}$

b. ( Mvə̀n=ɑ̀m  kə́ʔ )$_{\text{PPh}}$  (↓jə́n )$_{\text{PPh}}$  (↓mɛ́n )$_{\text{PPh}}$

This contrasts with the proposed phrasing in the recent past, where the subject is thought to form its own phonological phrase, and the tense and negation markers phrase with the following verb (7a,b).

(7) Prosodic phrasing for negated recent past tense with tense marker *fə̀*

a. ( Mvə̌n )$_{\text{PPh}}$  (↓fə̀  kə̀  jə́n )$_{\text{PPh}}$  ( ↓mɛ́n )$_{\text{PPh}}$

b. ( Mvə̀n=ɑ̌m )$_{\text{PPh}}$  (↓fə̀  kə̀  jə́n )$_{\text{PPh}}$  (↓mɛ́n)$_{\text{PPh}}$

Note that the phrasing in (6) also helps account for the fact that a high tone subject occurring in the negated future and recent past/perfect cannot trigger downstep on the negation marker in either case (8a,b)[13]. Since downstep is thought to occur only across a phonological phrase boundary, it follows that the high tone of the subject and the high tone of the negation markers should not be differentiated by a downstep, as they occur within the same phonological phrase.

(8) No downstep between subject and negation marker

a. Mɛ́n (*↓) kə́ ↓jə́n ↓mʙʉ́

child NEG.FUT.**H** see dog

'The child will not see the dog.'

b. Mɛ́n (*↓) kə́ʔ ↓jə́n ↓mʙʉ́

child NEG.PST.**H** see dog

'The child has not seen the dog.'

Thus far, we have demonstrated that the high preverbal phrase accent tends to occur at the right edge of the first phonological phrase within a sentence/clause. The importance of a clause boundary in determining the presence of a preverbal phrase accent is further illustrated by the fact that embedded clauses, like main clauses, introduce their own preverbal phrase accents (9).

(9) Both main clauses and embedded clauses introduce a preverbal phrase accent

Mvə̌n ↓ʧúp mbə̀ nʧwɛ̌t ↓ɣúʔ

chief.**H** say COMP chefferie.**H** large

'The chief said that the chefferie is large.'

In this case, we assume that each clause aligns with its own intonational phrase unit, yielding the prosodic parse in (10) for the sentence in (9). Here, each intonational phrase corresponds with its own high preverbal phrase accent.

(10) Prosodic phrasing for embedded clauses

[ ( Mvə̌n )$_{PPh}$ ( ↓ʧúp mbə̀ )$_{PPh}$ ]$_{IPh}$ [ ( nʧwɛ̌t )$_{PPh}$ ( ↓ɣúʔ )$_{PPh}$ ]$_{IPh}$

Finally, focus and topicalization can result in the introduction of an additional high preverbal phrase accent within a clause. An example involving subject focus is given in (11). Here, focusing the subject results in one high preverbal phrase accent being associated with the right edge of the focused subject, and another with the right edge of the low-toned tense marker that follows it. We can make sense of this pattern if we assume that the high preverbal phrase accent is obligatory at the level of the intonational phrase, and that the subject is in fact parsed into an intonational phrase on its own. The proposed prosodic phrasing for (11) is provided in (12). That sentence-initial focused and topicalized elements should form their own intonational phrases is familiar from work by Selkirk (2011) on the Bantu language Xitsonga, spoken in South Africa and Mozambique. Selkirk argues that such elements lie outside of the 'standard clause' of the sentence (which she defines as the complement of Comp$_0$), and that the introduction of an additional intonational phrase boundary reflects the high ranking of a constraint STRONG START which promotes material early in the sentence to have equal or greater prosodic strength to what follows it. Similarly, Downing and Mtenje (2011) show that focused elements in Chichewa, a Narrow Bantu language spoken in southeast Africa, introduce additional prosodic boundaries. They show that these boundaries tend to align with *v*P- and CP-level syntactic phase boundaries, a possibility which may also hold for Medʉmba. Keupdjio (2015) analyzes various types of focus in Medʉmba as involving movement of the focused constituent across a *v*P and/or CP-level syntactic phase boundary.[14][15]

(11) Subject focus results in an additional high preverbal phrase accent

| | | | |
|---|---|---|---|
| Mvə̌n | fə̌ | njə́n | ↓mɛ́n |
| chief.H | RPST.H | see | child |

‘The CHIEF saw the child.’

(12) Prosodic phrasing for a sentence with subject focus

[ ( Mvə̌n )$_{PPh}$ ]$_{IPh}$ [ ( fə̌ )$_{PPh}$ ( njə́n )$_{PPh}$ ( ↓mɛ́n )$_{PPh}$ ]$_{IPh}$[16]

From all of this we can conclude that the high preverbal phrase accent is constrained in its distribution based on the location of both phonological phrase boundaries and intonational phrase boundaries. Specifically, the distributional generalization about the high preverbal phrase accent is that it occurs at the right edge of the first phonological phrase of an intonational phrase. Given that there does not seem to be a corresponding high preverbal phrase accent assigned to other phonological phrases within the intonational phrase, it appears the phrase accent is *culminative* at the level of the intonational phrase (meaning that only one of these phrase accents can occur per intonational phrase), but that it is *attracted* to the right edge of a phonological phrase, a position which has been argued to bear metrical prominence in several Bantu languages (see Section 1). It is interesting to note that similar patterns have been attested in some Indo-European languages. For example, Frota (2003) documents a ‘phrasal peak’ in European Portuguese which is realized as a high tone associated to the stressed syllable of the first prosodic word of the intonational phrase. Myrberg (2010) and Roll et al. (2009) also document cases of ‘initiality accent’ in Swedish which is argued to align with the first stressed syllable within an intonational phrase. In both of these cases, the prosodic event of interest is thought to be introduced by the left boundary of an intonational phrase, but attracted to a prominent syllable occurring *near* (but not necessarily aligned with) the left edge of that constituent. Recall that the phenomenon of the phrase accent, as described by Grice et al. (2000), has precisely these characteristics: it is at its core a ‘boundary tone’ (meaning it is introduced by the boundary of a prosodic constituent), but its surface position will be determined primarily through its attraction to a metrically-prominent syllable (see also Barnes et al. 2006). Indeed, in the case of Swedish, Myrberg argues that the initiality accent in Swedish arises in order to satisfy a Phrase Edge Prominence constraint (Selkirk 1995) requiring phonological phrase edges to bear greater metrical prominence than non-edges.[17] This brings up the question for Medʉmba as to whether the preverbal phrase accent constitutes a similar kind of prominence-sensitive tone, rather than simply marking a phrase boundary. If this is the case, then we would expect that syllables marked with the phrase accent should show temporal properties which are consistent with their status as metrically-strong syllables. In the sections that follow, we will explore precisely this possibility. Specifically, we investigate whether or not these syllables display attraction to certain relative temporal positions in repeated speech, and whether they show increased duration relative to syllables which are hypothesized not to bear phrasal prominence. Section 3 presents an overview for the methodology of the experiments conducted, including an introduction to speech cycling, the experimental paradigm used to investigate these questions.

## 3. Method

### 3.1 The Speech Cycling Paradigm

The method used in this work, the *speech cycling paradigm*, was developed by Cummins (1997), Cummins and Port (1998), and Tajima (1998) to examine speech timing as a measure of rhythmicity in language. The paradigm, which has subjects repeat short sentences in time to a metronome at progressively faster speeds (=shorter metronome periods), has yielded striking results across languages, particularly with respect to relative syllable prominence. Specifically, it has been found that prominent syllables across a typologically diverse range of languages, including English (Cummins, 1997; Cummins & Port, 1998; Tajima, 1998; Tajima & Port, 2003; Tilsen, 2009), Japanese (Tajima, 1998, Tajima & Port, 2003), and Jordanian Arabic (Zawaydeh et al., 2002), are preferentially aligned with certain intervals of the repetition cycle, corresponding to lower order fractions of the cycle, such as the halfway, third, and two-thirds fractions of the cycle. These intervals are known as Simple Harmonic Phases, or SHPs (Cummins & Port, 1998). Previous studies have varied in the levels of prosodic prominence which have been considered. In studies on English, the focus has typically been on the final stressed syllable of the sentence, which is generally also the syllable which carries nuclear pitch accent. For example, in Figure 3, drawn from work by Cummins and Port (1998), the beginning of the vowel in the word *duck* in the (neutral focused) sentence *Dig for a duck* is found to preferentially align with the halfway, third, and two-thirds fraction of the repetition cycle, depending on how rapidly the subject is speaking. For example, while a pattern with *duck* occurring around 1/3 of the way through the repetition cycle is commonly found in slower metronome speeds, as speech rate increases, it becomes more comfortable for speakers to switch the position of duck to the halfway point in the cycle. It has also been found that syllables occurring at SHP positions display relatively less variability in alignment as compared with syllables occurring farther from these positions (Cummins & Port, 1998; Tilsen, 2009), a pattern which we also explore in the present work.

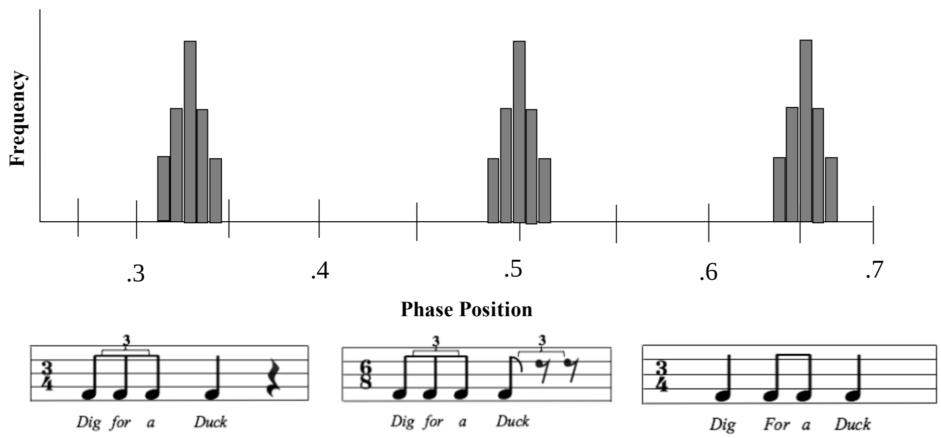


Figure 3: Typical alignment patterns for the word *duck* in speech cycling for the sentence *Dig for a duck*, with corresponding musical rhythmic notation. Histograms in the top portion of the figure represent clustering of syllable repetitions around phase positions of .333, .500, and .667 of the cycle.

The results from the speech cycling paradigm have been explained in terms of a dynamical model of prosodic structure in which distinct levels within the prosodic hierarchy are implemented as planning oscillators which are coupled together in time. Coupling between oscillators at the level of phrase repetitions and prosodic units such as the metrical foot occur at certain frequency-locking ratios such that the phrase is ultimately subdivided into binary or ternary metrical feet. In Figure 4, for example, we see side-by-side comparison of a traditional representation of the prosodic hierarchy (left) based on Selkirk (1986) (though missing the prosodic word level, which could be included in the case of a language, such as English, where a prosodic word can span multiple feet), and the same hierarchy represented as coupled oscillators with 2:1 syllable-to-foot ratio (yielding bisyllabic feet) and 2:1 foot-to-phrase frequency-locking pattern (right). Coupling and frequency-locking thus provide a mechanism by which to capture the well-documented constraint of prosodic Strict Layering found across languages (Selkirk 1981, 1984, 1986; Nespor & Vogel 1986).

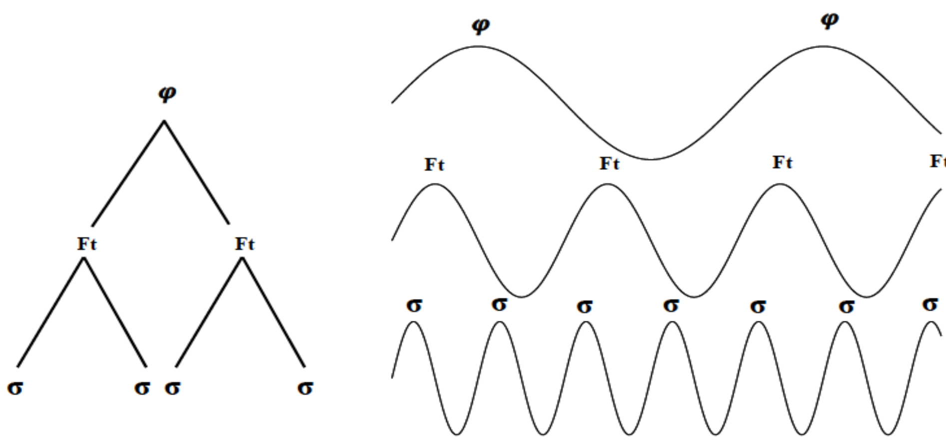


Figure 4: The prosodic hierarchy (left) modeled as coupled oscillators (right) with 2:1 syllable-to-foot and foot-to-phrase frequency locking

In essentially all existing work on speech cycling, phrase- and word-level prominence have not been teased apart, although it is predicted that syllables at both levels of prominence would be preferred over non-prominent syllables to occur at SHP positions in the repetition cycle (see Port, 2003 and Tilsen, 2009 for overviews). It is also the case that the majority of work in speech cycling has focused on syllables occurring late in a sentence. An exception to this generalization is work by Tajima (1998) which provides analysis of alignment patterns of sentences in English and Japanese looking at syllables in all positions. As one might expect, prominent syllables occurring earlier in the sentence (particularly where the sentence is more than four syllables long) tend to occur earlier than the stated SHP positions of 1/2, 1/3, and 2/3 phase. Tajima finds that stressed syllables occurring earlier in sentences of English (such as the initial syllable in *Bombay* in the sentence *Great Bombay Demons*) most frequently occur at 1/6 or 1/4 of the way through the repetition cycle. Unstressed and unaccented syllables (such as the second syllable of *Betty* in the sentence *Betty forgot the bag* or the second syllable of *giving* in the sentence *Giving the girl a doughnut*), on the other hand, tend to occur at higher order fractions such as 1/8 or 1/12 of the way through the cycle. Tajima found similar differences in the alignment of accented versus unaccented syllable in Japanese. This will be important given that the present study is also concerned with temporal patterning of syllables occurring relatively early in the sentence.

Finally, given the number of findings indicating increased duration for syllables bearing prominence across languages, we will also investigate the relative vowel durations of syllables hypothesized to occur with and without a phrase accent in the present study. As has been found previously, we expect that phrase-accented syllables will be realized with longer duration than unaccented syllables.

### 3.2 Stimuli

The present study consisted of two sub-experiments. In the first experiment, three pairs or 'sets' of six-syllable utterances were elicited (Table 1). Multiple items were included in each set to ensure the generalizability of findings across different sentences with different segmental makeups. The target word for all sentences was the third syllable, which in all cases was a negation marker. In the High/Accented condition, the negation marker is thought to occur in final position of a phonological phrase and carry a high phrase accent (refer to Section 2). In the Low/Unaccented condition, the negation marker carries a low tone, and is thought to occur in medial position of a phonological phrase. All efforts were made to control for phonetic context so as to make the paired sentences as similar as possible apart from the target manipulation.

Note that the final [l] of the noun in each subject for the High/Accented condition in Table 1 arises through a process of lenition, whereby the final consonant in each of those words, underlyingly a /d/, undergoes lateralization where it occurs between two vowels. Perceptually, the resulting [l] sounds quite similar to the one in onset position of the tense marker for each sentence in the Low/Unaccented condition.

| SET | CONDITION | IPA TRANSCRIPTION | GLOSS | TRANSLATION |
|---|---|---|---|---|
| A | High/ Accented | [ Nʃúl ù **kə́ ] [** ↓kwág ] [ ↓bán àm ] | [ Ant 2.SG.POSS NEG ] [ touch ] [ couscous 1.SG.POSS ] | 'Your ant won't touch my couscous' |
| | Low/ Unaccented | [ Nʧó ] [ ↓lù **kə̀** kə́ʔ ] [ ↓mvέl àm ] | [ Thief ] [ RPST NEG cut ] [ brother 1.SG.POSS ] | 'The thief didn't cut my brother' |
| B | High/ Accented | [ Wʉ́l ù **kə́** ] [ ↓kúʔ ] [ ↓mbál àm ] | [ Body 2.SG.POSS NEG] [ climb ] [ hill 1.SG.POSS ] | 'Your body won't climb my hill' |
| | Low/ Unaccented | [ Mʙʉ́ ] [ ↓lù **kə̀** béb ] [ ↓mέn àm ] | [ Dog ] [RPST NEG wait ] [ child 1.SG.POSS ] | 'The dog didn't wait for my child.' |
| C | High/ Accented | [ Mvέl ù **kə́** ] [ ↓zwí ] [ ↓ʃún àm **]** | [ Brother 2.SG.POSS NEG ] [ hurt ] [ friend 1.SG.POSS ] | 'Your brother won't hurt my friend.' |
| | Low/ Unaccented | [ Mʙwə́ ] [ ↓lù **kə̀** vwáʔ ] [ ↓ʧwεn àm ] | [ Goat ] [ RPST NEG throw ] [ wood 1.SG.POSS ] | 'The goat didn't throw my wood' |

Table 1: Stimuli for Experiment 1 (phonological phrase boundaries indicated with brackets)

Given that the relative pitch of the target syllables in Experiment 1 differ, an additional experiment was run to ensure that any differences in timing observed in Experiment 1 could not be attributable to differences in pitch alone. Relative pitch is known to affect perception and production of syllable duration and timing (Franich 2018; Genzel & Kügler 2018; Manyah 2006; Yu 2010), and therefore constitutes a potential confound to Experiment 1. Therefore, Experiment 2 tested the difference in temporal patterning between lexical high and low tones, independent of the high phrase accent. The second experiment also involved three sets of sentences, this time with five syllables (though note that the final word in the High condition of Set A was variably produced with a final schwa, creating a sixth syllable) (Table 2). The final syllable in all cases was the target syllable, and in all cases this was the initial (or only) syllable of a noun. The target words appear in the final position of an intonational phrase. In the High condition, the syllable bears a high tone which is lexically specified, and in the Low condition, the syllable bears a low tone which is lexically specified. The target word was also the second of two in a 'noun of noun' associative construction, in which the second noun, if high, can be a target of downstep. As mentioned in Section 2, downstep is a diagnostic for the presence of a phonological phrase boundary, and thus the final noun of the sentence forms a phonological phrase unto itself.

| SET | CONDITION | IPA TRANSCRIPTION | GLOSS | TRANSLATION |
|---|---|---|---|---|
| A | High | [ Mɛ́n ǎm ] [ ↓zwí ] [↓ʃûn ] [ ↓**sáŋ**(ə) ] | [ Child 1.sg.poss ] [ kill ] [ friend.of ] [ bird ] | 'My child killed the friend of the bird.' |
| | Low | [ Bán ǎm ] [ ↓zwí ] [ ↓ʃûn] [ **mvəǹ ]** | [ Couscous 1.sg.poss ] [ kill ] [ friend of ] [ chief ] | |

| | | | | |
|---|---|---|---|---|
| | | | | 'My couscous killed the friend of the chief.' |
| B | High<br><br>Low | [ Mbál ǎm] [ ↓bút] [ ↓mɛ̂n ] [↓**nʧó ]**<br><br>[ Mvɛ́l ǎm ] [ ↓bút ] [ ↓mɛ̂n ] [ **ngò ]** | [ Hill 1.sg.poss ] [ fatigue ] [ child.of ] [ chief ]<br><br>[ Chief 1.sg.poss ] [ fatigue ] [ child.of ] [ country ] | 'My hill has fatigued the child of the thief.'<br><br>'My brother has fatigued the child of the country.' |
| C | High<br><br>Low | [ Nʃún ǎm ] [ ↓lʉ́m ] [ ↓mvɛ́t ] [↓**mɛ́n ]**<br><br>[ Nʧwɛ́n ǎm ] [ ↓lʉ́m ] [ ↓mvɛ́t ] [ **mɛ̀n ]** | [ Friend 1.sg.poss ] [ hurt ] [ brother.of ] [ child ]<br><br>[ Wood 1.sg.poss ] [ hurt ] [ brother.of ] [ someone ] | 'My friend has hurt the brother of your child'<br><br>'My wood has hurt someone's brother |

Table 2: Stimuli for Experiment 2 (phonological phrase boundaries indicated with brackets)

## 3.3 Participants and Procedure

There were ten participants in the study, 6 female and 4 male. Ages of participants ranged from 20 to 60 years old (mean age of 32). All were native speakers of Medʉmba who had spent the majority of their lives living in Banganté, the city in southwestern Cameroon where Medʉmba is primarily spoken, or the villages surrounding it.

Participants were seated at a table in a quiet hotel room in front of a Macbook Pro 13" laptop. All were fitted with a Shure SM35 head-mounted condenser microphone which rested just over their ears and attached around the back of the head. Over this they wore a pair of Sony MDR 7506 studio headphones through which the metronome beats were played. The metronome sound consisted of a synthetic drumbeat created in version 2.1.2 Audacity® recording and editing software, an open-source program for sound editing. Each of the six target sentences were elicited at 15 different speech rates, from slowest (Speed 1) to fastest (Speed 15). This resulted in a maximum of 90 total trials per participant (though see below for repetition success rates for each subject), each of which consisted of 8 repetitions of the target sentence (720 total utterances per subject). The slowest speech rate corresponded to a 1770 ms metronome period, and, following Tajima (1998), the period was reduced by 3% for each subsequent speed, such that the fastest speed corresponded to a 641 ms metronome period. For each trial, participants heard a total of twelve clicks of the metronome. Similar to the procedure described in Tajima (1998), participants were asked to listen to the first four beats so as to acclimate to each new speed and then begin repeating on the fifth beat, saying the sentence once per beat. Sentences were presented in a random order (though the participant saw the same sentence in a row for all 15 metronome speeds).

The target sentence was displayed on the computer screen in Powerpoint in white font against a black background in both French and Medʉmba, in the native Medʉmba orthography. In case participants were not able to read the target sentences, the experimenter read the sentence in both French and Medʉmba and asked the participant to repeat both forms back to verify they had understood what the target sentence was. Participants underwent several practice trials using a separate set of sentences prior to starting the experiment. Once the participant felt comfortable with the task, the experimenter advanced to the experimental trials. During the course of the experiment, subjects were given periodic breaks to rest and drink water.

Not all speech rates were possible for all participants. Once a participant made repetition errors in successive utterances of a particular sentence, they were advanced to the next sentence. None of the participants were able to produce the sentences at the highest speech rates, and some subjects were only able to produce the target sentences fluently up to about Speed 9 or 10. A summary table of the speeds each participant was able to reach, organized by Set, is provided in Table 3.

| | **s1** | **s2** | **s3** | **s4** | **s5** | **s6** | **s7** | **s8** | **s9** | **S10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Set A** | 10 | 9 | 10 | 10 | 11 | 9 | 13 | 12 | 9 | 12 |
| **Set B** | 9 | 10 | 10 | 11 | 7 | 8 | 11 | 11 | 10 | 14 |
| **Set C** | 8 | 13 | 9 | 10 | 10 | 9 | 10 | 10 | 9 | 13 |

Table 3: Number of metronome speeds attainable for sentence repetitions by Subject (s) and Set

### 3.4 Data Processing

Data were annotated semi-automatically using the beat extractor method developed by Cummins (1997) and Scott (1993) and implemented in Praat using the BeatExtractor script written by Barbosa (2003). The script is designed to insert boundaries at each perceptual center (p-center), or the instantaneous 'beat' where listeners perceive a syllable to occur (Morton et al., 1976); this point typically lies close to the vowel onset in Medʉmba (Franich, 2018). The script works by applying a second-order Butterworth filter to the speech signal, after which the signal is rectified and low-pass filtered. 'Beats' are then inserted, in the form of TextGrid boundaries, at points corresponding to the local maxima of the first derivative of each amplitude envelope. Annotations were subsequently hand-corrected, and spurious boundaries removed. Timing of each syllable (referred to henceforth as the syllable's *relative phase*) was calculated in terms of the phase repetition cycle, or the time of the interval spanning successive repetitions of the target sentence, measured from the p-center of the first syllable in each repetition. This measurement is demonstrated in Figure 5: the interval of *b*, which extends from the first p-center of the repetition cycle to the p-center of the third syllable, is divided by the interval of *a*, the duration of the entire repetition cycle (p-center of one repetition of Syllable 1 to the p-center of the following repetition of Syllable 1). This gives the relative phase measure for Syllable 3; similar measures are taken for other syllables.

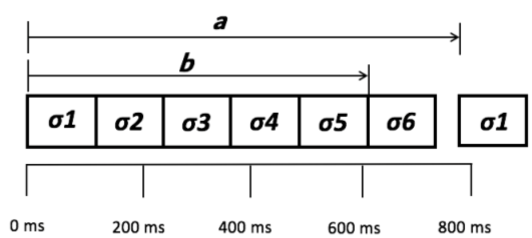


Figure 5: Relative phase for Syllable 6 in a six-syllable utterance is interval *b* divided by interval *a*

After data processing, individual subjects' alignment patterns were compared visually. As can be seen in Figure 6 (where alignment patterns are shown for Syllable 3 across all utterances), nine out of the ten subjects showed very similar overall alignment patterns, while one subject (s5) adopted a quite different alignment strategy. Different alignment strategies were not unexpected, as individuals' distinct baseline speaking rates were predicted to determine their most comfortable alignment strategy. Subject s5 was believed to have an overall slower baseline speaking rate than the other subjects, leading to the adoption of a different alignment strategy likely with different SHP positions.[19] This subject's data was therefore excluded from the analysis. Additionally, one subject (s4) displayed difficulty falling in rhythm with the metronome on their first trial (which was a sentence in the Low/Unaccented condition from Set A); data for this trial for that subject was thus removed. Finally, datapoints corresponding to alignment values exceeding 2 standard deviations from the mean for a given syllable and metronome speed were removed as they reflected disfluencies in repetitions where subjects stumbled over a word; this resulted in the removal of less than 4% of the remaining data.

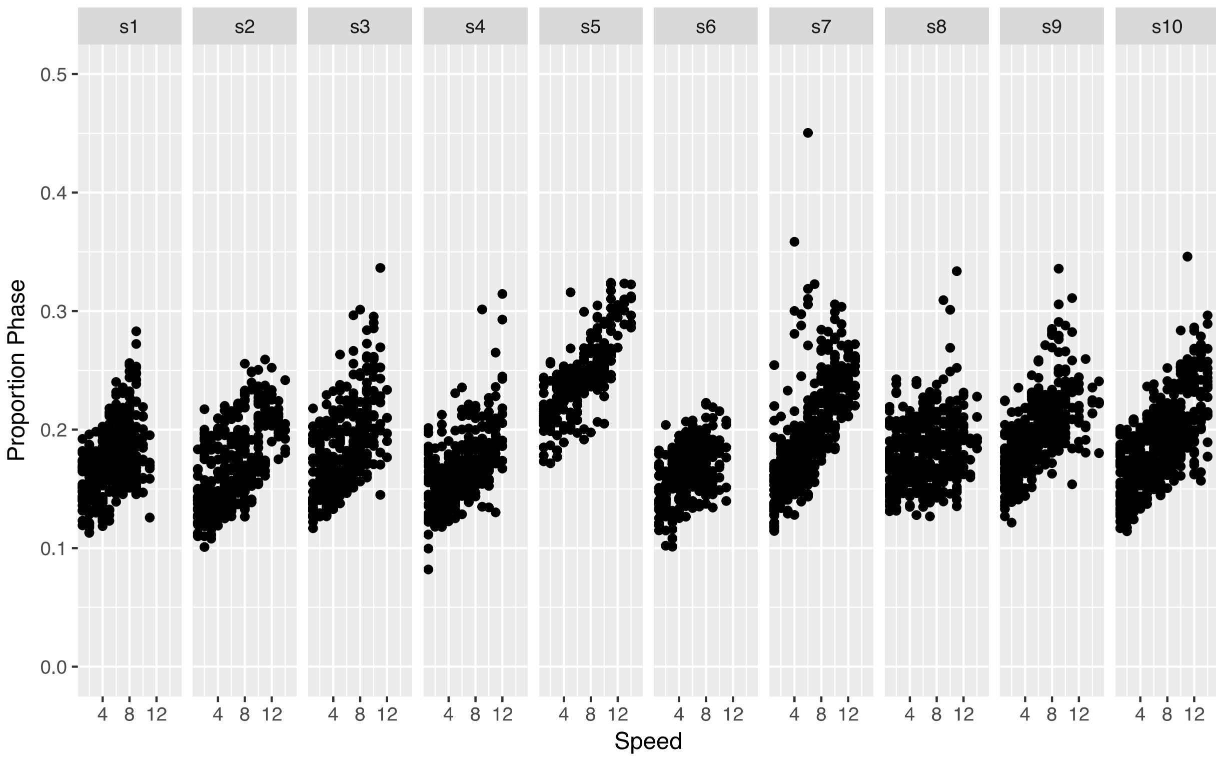


Figure 6: Alignment of Syllable 3, Experiment 1, across subjects. Subject s5 can be shown to display a very different pattern of alignment from the other subjects.

## 3.5 Statistical Modeling

Syllable alignment was compared statistically for each syllable of interest using linear mixed effects modeling implemented with the *lme4* statistical package (Bates et al. 2015) for *R* software

(R Core Team 2018). Each model included fixed effects terms for Condition (coded as a factor with 2 levels; High/Accented and Low/Unaccented for Experiment 1, and High and Low for Experiment 2) and Metronome Speed (coded as a numeric variable and mean-centered). In addition, by-subject and by-set random slopes were included for each of the two variables.

### 3.6 Predictions

For Experiment 1, it is predicted that relative alignment of the target syllable—Syllable 3—should be significantly different across conditions, such that the target syllable in the High/Accented condition should occur significantly closer to simple harmonic phase positions than the target syllable in the Low/Unaccented condition. It may also be the case that syllables surrounding the target syllable (Syllables 2 and 4) will show some differences in alignment, as well. In Experiment 2, it is predicted that nouns occurring in Syllable 5 position should show no difference in alignment based on tone. For both experiments, it is also predicted, based on previous speech cycling research, that target phase position should show an effect of metronome speed, such that increases in metronome speed (= increases in speech rate) should lead to an increase in relative phase. For our duration comparison, we predict that the vowels of target syllables in the High/Accented condition will have longer duration than those in the Low/Unaccented condition for Experiment 1, and that High and Low tone syllables in Experiment 2 should show no difference in duration (in line with findings from Franich 2018).

## 4. Results

### 4.1 Experiment 1

For Experiment 1, alignment patterns were evaluated for the target syllable (Syllable 3) as well as the two surrounding syllables (Syllables 2 and 4) across the two sentence conditions. A full summary of findings from mixed effects modeling is found in Table 4. As expected, there was a significant effect of CONDITION on alignment of Syllable 3, the target syllable ($p < 0.01$). Findings indicated that Syllable 3 occurred significantly closer to the closest SHP position of .167 (1/6 of the way through the repetition cycle) in the High/Accented condition than in the Low/Unaccented condition (Figure 7)—while the mean phase position of the target syllable in the High/Accented condition was .172, the target syllable in the Low/Unaccented condition was generally timed later, with a mean phase position of .188. There was also a significant difference in alignment for Syllable 2, the syllable preceding the target syllable ($p < 0.001$); similarly to the target syllable, timing was slightly earlier in the High/Accented condition than in the Low/Unaccented condition. No significant difference in alignment was found across conditions for Syllable 4 ($p = .72$).

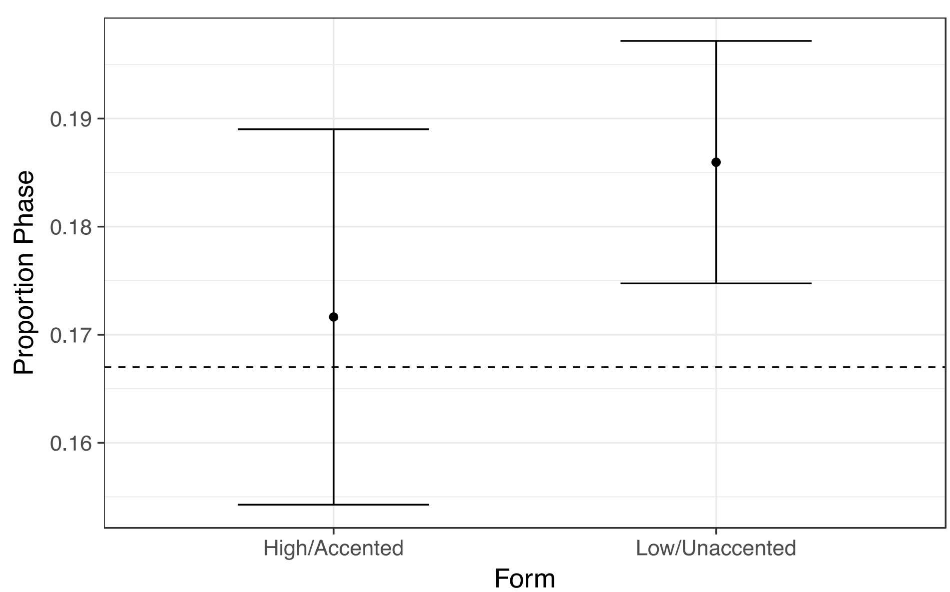

Figure 7: Alignment for Syllable 3 (target) across High/Accented and Low/Unaccented conditions, Experiment 1. Horizontal dotted lines represent closest SHP position of .167. Error bars represent 95% confidence interval.[20]

| **Syllable 3 (Target syllable, Expt. 1)** | | | | | |
|---|---|---|---|---|---|
| | Estimate | Standard error | *df* | t-value | *p*-value |
| (intercept) | 0.180 | 0.006 | 8.114 | 31.945 | < 0.001 *** |
| Condition | -0.008 | 0.002 | 8.043 | -3.574 | < 0.01 ** |
| Met. Speed | 0.018 | 0.003 | 5.426 | 6.132 | < 0.01 ** |
| **Syllable 2 (Preceding syllable, Expt. 1)** | | | | | |
| | Estimate | Standard error | *df* | *t*-value | *p*-value |
| (intercept) | 0.083 | 0.005 | 5.968 | 17.997 | < 0.001 *** |
| Condition | -0.013 | 0.001 | 8.017 | -9.160 | < 0.001*** |
| Met. Speed | 0.006 | 0.002 | 4.351 | 3.353 | < 0.05 * |
| **Syllable 4 (Following syllable, Expt. 1)** | | | | | |
| | Estimate | Standard error | df | *t*-value | *p*-value |
| (intercept) | 0303 | 0.011 | 3.161 | 27.397 | < 0.001 *** |
| Condition | 0.001 | 0.003 | 7.969 | 0.372 | = 0.720 |
| Met. Speed | 0.034 | 0.005 | 5.404 | 7.269 | < 0.001 *** |
| **Syllable 5 (Target syllable, Expt. 2)** | | | | | |
| | Estimate | Standard error | *df* | *t*-value | *p*-value |
| (intercept) | 0.539 | 0.010 | 9.629 | 50.516 | < 0.0001 *** |
| Condition | 0.001 | 0.005 | 7.958 | 0.233 | = 0.822 |

| | | | | | |
|---|---|---|---|---|---|
| Met. Speed | 0.061 | 0.009 | 4.671 | 6.732 | < 0.01 ** |

Table 4: Fixed effects results, mixed effects models for syllable alignment by Condition and Metronome Speed.

For all syllables examined, a significant effect of METRONOME SPEED was found: syllable phase position increased as speech rate increased ($p < .01$).

## 4.2 Experiment 2

Experiment 2 examined differences in alignment of nouns bearing high versus low tones. Statistical analysis revealed no significant effect of TONE on syllable alignment. The mean phase position for the target syllable in the High tone condition was .543, while in the Low tone condition it was .528. As has been found in previous work in speech cycling, there was overall greater variability in alignment of syllables occurring later in the utterance (Tajima 1998; Zawaydeh et al. 2002): The overall alignment variability for Syllable 5 in Experiment 2 was higher (*IQR* = .111 in the High condition; *IQR* = .109 in the Low condition) than the values found for Syllables 2, 3, and 4 for Experiment 1 (*IQR* = .032, .044, and .067, respectively). While the closest SHP position to the target syllable for both conditions in Experiment 2 was .500, in both conditions the target syllable can be seen to be produced slightly after that position (Figure 8).

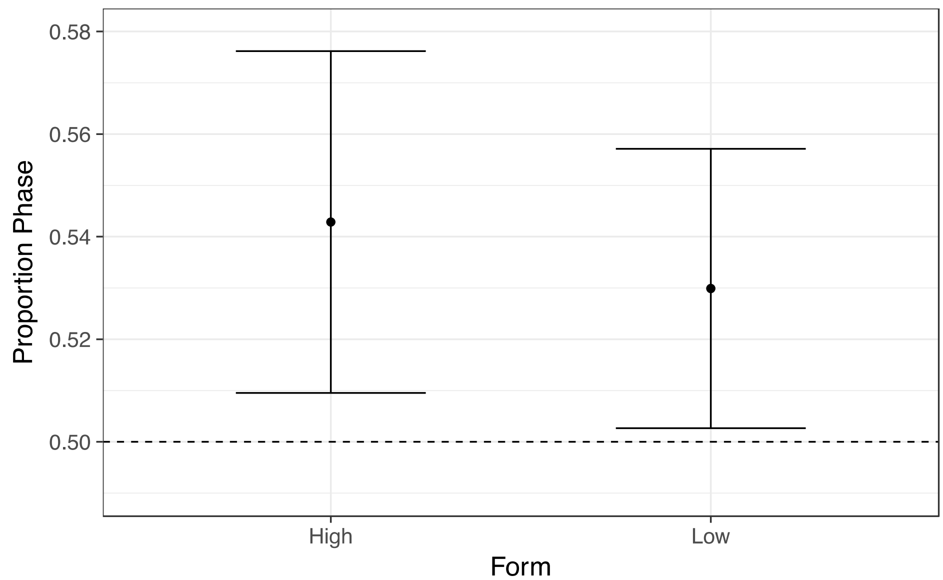


Figure 8: Alignment for Syllable 5 (target) across High tone and Low tone conditions, Experiment 2. Horizontal dotted lines represent closest SHP position of .500. Error bars represent 95% confidence interval.

As was found for all syllables in Experiment 1, there was a significant effect of METRONOME SPEED on phase position of the target syllable, with phase position increasing as metronome speed increased ($p < .01$).

## 4.3 Vowel Duration

Finally, vowel duration was compared for all 10 subjects for Syllable 3 as produced at Speeds 4-6 in Experiment 1. In order to extract vowel duration measures, recordings were force-aligned using FAVE (Rosenfelder et al. 2014) and then visually checked for accuracy; vowel durations were then extracted via an in-house script with Praat (Boersma & Weenink 2018). Six repetitions from each of the three sentences per sentence condition were measured for each subject, for a total of 540 tokens per condition. Speeds 4-6 were chosen as these were the rates that were universally most comfortable across speakers. Since absolute duration measures are expected to be more sensitive to changing speech rate than relative timing measures are (Port 2003; Tajima 1998), it was especially important to have comparable data across subjects for duration analysis. A linear mixed effects model including CONDITION as a single, sum-coded fixed effect was run. The model also included by-subject and by-set random slopes for CONDITION. A summary of results is presented in Table 5. Results indicated that Syllable 3 in the High/Accented condition was produced with significantly longer vowel duration (mean = 61 ms) than Syllable 3 in the Low/Unaccented condition (mean = 49 ms) ($p < 0.05$) (Figure 9). No significant difference in vowel duration was found for the target syllables in Experiment 2 ($p = .43$).

| **Target syllable (Syllable 3), Expt. 1** | | | | | |
|---|---|---|---|---|---|
| | Estimate | Standard error | *df* | *t*-value | *p*-value |
| (intercept) | 60.162 | 4.624 | 6.614 | 13.010 | < 0.001 *** |
| Condition | 10.657 | 3.333 | 7.236 | 3.197 | < 0.05 * |
| **Target syllable (Syllable 5), Expt. 2** | | | | | |
| | | | | | |
| (intercept) | .048 | .005 | 9.025 | .000 | < 0.001 *** |
| Condition | .005 | .006 | .947 | .431 | = .431 |

Table 5: Fixed effects results, mixed effects models for duration by Condition.

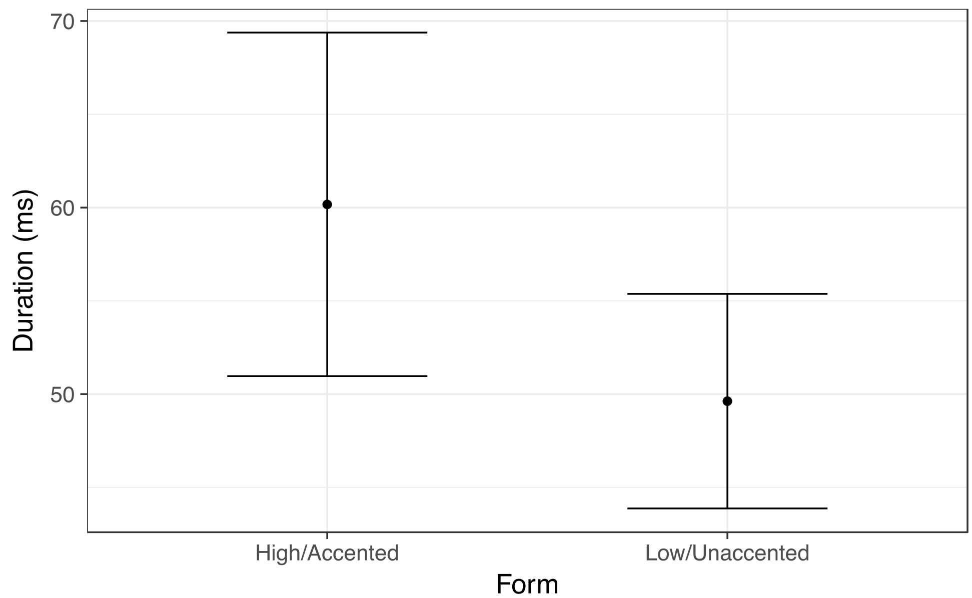


Figure 9: Comparison of vowel duration for Syllable 3 (target) across High/Accented and Low conditions, Experiment 1. Error bars represent 95% confidence interval.

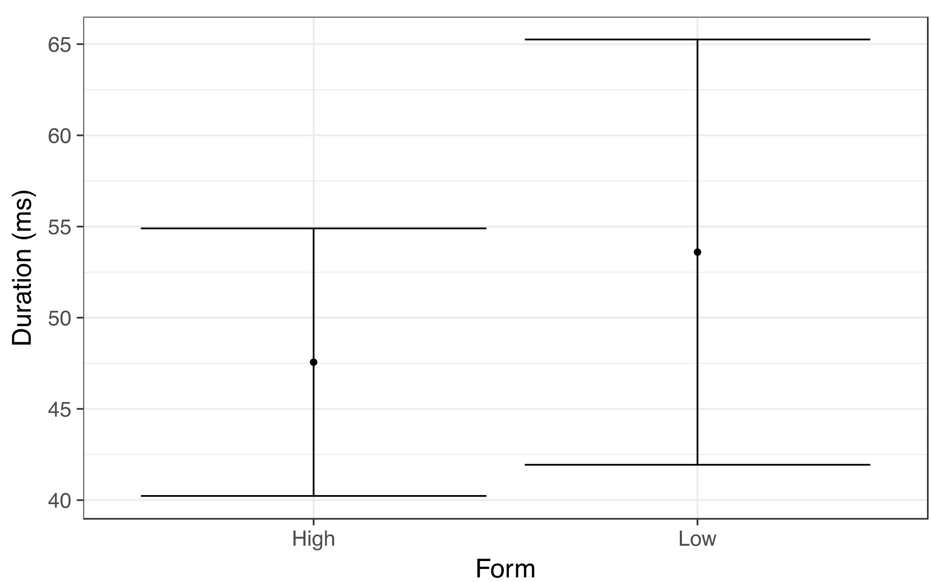


Figure 10: Comparison of vowel duration for Syllable 5 (target) across High/Low conditions, Experiment 2. Error bars represent 95% confidence interval.

As can be seen in Figure 10, vowel duration for low tones was numerically higher than that for high tones in Experiment 2. While average vowel duration between low tone syllables in Experiment 2 and low/unaccented syllables in Experiment 1 differed on average by only a few ms (53 ms in the former and 49 ms in the latter; a difference of 4 ms), there was a considerable difference in duration between high tone syllables in Experiment 2 and high/accented syllables in Experiment 1, which were 48 ms and 61 ms, respectively, a difference of 13 ms. This constitutes around a 25% increase in vowel duration between the plain high tone and high/accented conditions.

## 5. Discussion

Our experimental findings lend support to the analysis in Section 2 which posits that high ‘tones’ occurring preverbally on morphemes such as the negation marker in the unmarked future tense are better analyzed as high preverbal phrase accents. Specifically, our prediction for Experiment 1 that syllable alignment should vary between the high/accented negation marker in the unmarked future tense and the low/unaccented negation marker in the recent past tense was borne out: syllables carrying the high phrase accent were found to occur significantly closer to harmonic phase positions than non-accented (low) syllables. Target syllables in the High/Accented condition also showed slightly less temporal variability than the corresponding syllables in the Low/Unaccented condition. The finding from Experiment 2 that alignment patterns are similar between lexical high and low toned syllables occurring at phrase edges further supports the conclusion that the findings from Experiment 1 cannot be attributed to tone/F0 differences alone: high tones are not, generally speaking, more prominent than low tones.

Interestingly, the phrase-final words compared in Experiment 2 occurred farther from an SHP position than might have been expected if we were to assume that all right edges of phrases bear some sort of abstract prominence (the halfway or .5 phase position was the closest SHP position for both conditions). This would suggest that phonological phrase-final position, in general, is not a position of greater prominence, and that the right edge of the clause-initial phonological phrase is special, in this regard. We note, however, that the complexity of the onsets for the words used in Experiment 2 was greater than that of the plosive-initial words used in Experiment 1, perhaps leading to less precise estimation of the perceptual centers for those words. Onset duration/complexity is known to affect p-center location (Morton et al. 1976), and though words were matched well for duration and complexity across tonal conditions within Experiment 2, they were not matched well across Experiment 2 and Experiment 1. Specifically, p-centers tend to occur earlier in words with longer/more complex onsets, so the estimations given may have skewed too late in the syllable (meaning actual p-centers for target words may have occurred closer to the .5 phase position for the two conditions).

For the vowel duration analysis, accented syllables were found to display longer duration than nonaccented syllables in Experiment 1, consistent with cross-linguistic findings on syllable prominence in other tonal (and non-tonal) languages. Similar to results from Franich (2018), no significant difference in duration was found between high and low tone words in Experiment 2, where phrase accent was not a factor (in contrast with findings from other African languages, where tone has been found to have an effect on segmental duration, e.g. Manyah 2006). High/accented syllables in Experiment 1 also showed considerably longer duration than the ‘plain’ high tone syllables in Experiment 2. Given the difference in sentence position and segmental makeup across experiments, we did not do a direct statistical comparison across these conditions, but numerical comparisons suggest that accented syllables carry greater duration than nonaccented syllables regardless of tone. It remains to be seen whether or not this difference is perceptually salient to Medʉmba speakers, and what role it may play in speech processing. Kembler, Choi, Yu and Cutler (this volume) show that phrase-level acoustic prominence is associated with better recall of target words in typologically-diverse languages such as English and Korean, suggesting that acoustic-prosodic prominence serves an important role in processing. A study similar to this could be useful in further understanding the prosodic status of high/accented syllables in Medʉmba.

Recall that low tone syllables bearing the Medʉmba high phrase accent are realized as a low-high rising tone. It is interesting to note, given the duration differences noted above, that rising tones are also associated with longer duration cross-linguistically due to the fact that they take longer to articulate than level or falling tones (Ohala & Ewan 1973; Sundberg 1973; Xu & Sun 2002). In a number of tone languages, rising contours also appear to be relegated to phrase-final position (Coupe 2007; Goldsmith 1988; Michaud & Vaissière 2015; see also Ou & Guo, this volume), perhaps owing to the facilitative effects of prosodic boundary lengthening on the production of rising tones. We note that this natural connection between f0 rise, duration, and phrase boundaries may contribute to the salience of rising f0 as a marker of perceptual prominence across languages, as has been documented in two other papers in this volume, both tonal and non-tonal (Ou & Guo, this volume, for Taiwanese Southern Min and Destruel & Féry, this volume, for French).

An additional finding which bears discussion is the difference in alignment for Syllable 2 in Experiment 1, the syllable which precedes the target syllable. Similar to the target syllable, Syllable 2 was found to occur significantly earlier in the High/Accented condition. This finding is consistent with previous speech cycling research: Tajima (1998) finds, for example, that a syllable's prominence not only affects the timing of that specific syllable, but the alignment of surrounding syllables, as well. In particular, the duration of syllables preceding the target syllable may be reduced in order for the prominent syllable to occur close to an SHP position. One question which might arise, however, is whether differences in alignment of the target syllable in Experiment 1 may come about as a result of differences in alignment of the preceding syllable, as opposed to the other way around. In other words, perhaps the relevant distinction in alignment is between the syllables occurring in Syllable 2 position, rather than those occurring in Syllable 3 position. After all, despite the fact that Syllable 2 comprises a function word in both conditions, the precise morphosyntactic status of the second syllable differs across conditions: in the High/Accented condition, the syllable is a pronoun and forms part of a possessive noun phrase, whereas in the Low/Unaccented condition, it is a tense marker. Furthermore, and perhaps more importantly, the pronoun in the High/Accented condition is an enclitic (cliticizing to the preceding noun; see Danis et al. 2011), likely forming a binary foot with the noun. This could have an effect on its timing properties. However, given that the observed temporal alignment patterns not only confirm a difference in timing between the target syllables across conditions, but a very close alignment of the target syllable in the High/Accented condition with the 1/6 SHP position, we conclude that the findings still point to an accentual difference on Syllable 3 as the key variable driving our results.

One somewhat unexpected finding was that there was no difference in alignment of Syllable 4 (the verb which followed the target syllable) across sentence conditions. We might have expected, given the prosodic parses given in Table 1, that the verb in the High/Accented condition, being at a phrase edge, should have shown a different pattern of alignment from the verb in the Low/Unaccented condition, which does not occur at a phrase edge. A possible reason for this finding, assuming a recursive prosodic phrase structure (Itô & Mester 2012), is that the verb is enclosed in a 'minimal' prosodic phrase by itself, and then parsed with the preceding tense and negation markers in a 'maximal' prosodic phrase, as indicated in (13) for the sentence in Set A:

(13) ( Nʧó )$_{PPh}$ ( ↓lù **kə̀** ( kə́ʔ )$_{PPh}$)$_{PPh}$ ( ↓mvɛ́l ə̀m )$_{PPh}$

Such an analysis would be in line with work by Truckenbrodt (1996, 1999) who proposes a recursive phonological phrase structure for the Tanzanian Bantu language Kimatuumbi.

Given the convergence of evidence, including the attraction of the high 'tone' in Experiment 1 to the phrase edge, the alignment patterns for syllables bearing this tone in speech cycling, and their increased duration, our findings are consistent with an analysis of the Medʉmba high phrase-final tone as a high phrase accent. As mentioned previously, the association between the high preverbal phrase accent and a prosodic boundary would suggest that it could be treated as a simple boundary tone with no prosodic prominence tied to it. However, given that the high preverbal phrase accent shows a striking resemblance to prosodic phenomena such as the 'initial accent' in Swedish, including culminativity at the level of the intonational phrase paired with attraction to the right edge of the phonological phrase (a position which is argued to be associated with metrical prominence in other Bantu languages), it seems likely that the high preverbal phrase accent associates with phrasal prominence, similar to findings from Portuguese and Swedish (Frota 2003; Myberg 2010; Roll et al. 2009), and in line with the definition of *phrase accent* as applied to multiple languages by Grice et al. (2000).

Even if metrical prominence is not at play in the case of Medʉmba, however, our results still have important implications for our understanding of rhythmic timing in the language. As discussed previously, others have argued that prosodic domain edges, such as the initial position of the Accentual Phrase in Korean, serve the same role as metrical heads in languages which lack metrical prominence relations (Chung & Arvaniti 2013). In other words, the same instincts which drive English speakers to align stressed syllables to certain positions in the speech cycling task also drive Medʉmba and Korean speakers to link phrase edges to these positions; otherwise, we would not have expected any principled timing differences across syllables to arise in speech cycling for the latter two languages. Thus, whether the key element driving the speech cycling results for Korean and Medʉmba speakers is metrical prominence or the prosodic domain edge, it is clear that not all syllables are treated as equally rhythmically prominent in these languages.

## 6. Conclusion

To conclude, we have shown that syllables bearing high phrase accents in Medʉmba are more likely to be attracted to positions of rhythmic prominence (Simple Harmonic Phases) in speech cycling than syllables not bearing the phrase accent. These results have important implications both for empirical phonetic work on Bantu tone languages, and for linguistic theory. On the empirical front, our results suggest that Bantu languages—and even those with relatively more elaborate tone systems such as Medʉmba—should be investigated for both tonal and intonational/accentual properties. Our results illustrate how instrumental techniques can (and should) be applied in situations where phonetic distinctions are subtle and otherwise hard to perceive. In terms of theory, these results provide further support for the importance of prosodic domain edges in regulating rhythmic timing across languages, perhaps in particular where a language lacks metrical heads which would otherwise serve this purpose. Going forward, it will be important to understand how prominence is manifested acoustically and temporally across other Bantu languages in order to confirm the cross-linguistic generalizability of the present results.

References:

Abramson, A.S. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. *International Journal of American Linguistics, 28*(2).

Audacity Team (2018). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.3.0 retrieved December 20th 2018 from https://audacityteam.org/

Barbosa, P.A. (2003). Beat extractor Praat script. Retrieved from https://uk.groups.yahoo.com/neo/groups/praat-users/conversations/topics/921.

Barnes, J., Shattuck-Hufnagel, S., Brugos, A., and Veilleux, N. (2006). The domain of realization of the L- phrase tone in American English. *Proceedings of Speech Prosody, Dresden, Germany*.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2017). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3,* 15–70.

Bickmore, L.S. (1995). Stress in Lamba. *Phonology, 12*, 307–341.

Boersma, P. and Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

Bolinger, D. L. (1958), Intonation and grammar. *Language Learning, 8*, 31–37. doi:10.1111/j.1467-1770.1958.tb01214.x.

Bruce, G. (1998). *Allmän och svensk prosodi* (Praktisk Lingvistik 16). Lund: Lund University. [Republished 2012 by Studentlitteratur, Lund]

Cassimjee F. and Kisserberth, C.W. (1992). The tonology of depressor consonants: Evidence from Mijikenda and Nguni. *Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society: Special Session on The Typology of Tone Languages*, 26–40.

Cassimjee, F. and Kisseberth, C.W. (1998). Optimality Domains Theory and Bantu tonology: a case study from Isixhosa and Shingazidja (33–132). In L.M. Hyman and C.W. Kisseberth (Eds.), *Theoretical Aspects of Bantu Tone*. Stanford, California: CSLI, 33–132.

Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics*, *34*, 176–201.

Chung, Y., and Arvaniti, A. (2013). Speech rhythm in Korean: Experiments in speech cycling. *Proceedings of Meetings on Acoustics*, *19*, 1–9.

Clements, G.N. and Ford (1979). Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry*, *10*(2), 179–210.

Clements, G.N. and Goldsmith, J. (eds.). (1984). *Autosegmental Studies in Bantu Tone*. Dordrecht: Foris, 1–17.

Cole, D.T. (1955). *An Introduction to Tswana Grammar*. London: Longmans & Green.

Coupe, Alexander R. 2007. *A grammar of Mongsen Ao*. Berlin & New York: Mouton de Gruyter.

Cummins, F. (1997). *Rhythmic Coordination in English Speech: An Experimental Study*. Indiana University dissertation.

Cummins, F. and Port, R.F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.

Danis, N., Barnes, J., and O'Connor, C. (2011). Downstep and contour formation in Medumba: a prosodic account. Selected Proceedings of the 42nd Annual Conference on African Linguistics, ed. Michael R. Marlo et al., 23–31. Somerville, MA: Cascadilla Proceedings Project.

Dayal, V. (2012). The syntax of scope and quantification. *The Cambridge Handbook of Generative Syntax*, 827–859. Cambridge, UK: Cambridge University Press.

Doke, C.M. (1938). *Textbook of Lamba Grammar*. Johannesburg: Witwatersrand U.P.

Downing, L.J. (1990). Local and metrical tone shift in Nguni. *Studies in African Linguistics, 21*, 261–317.

Downing, L.J. (2010). Accent in African languages. In Harry G. van der Hulst et al. (eds.), *A Survey of Word Accentual Patterns in the Languages of the World*. Berlin, Mouton de

Gruyter: 381–427.

Downing, L.J. and Mtenje, A. (2011). Prosodic phrasing of Chichewa relative clauses. *Journal of African Languages and Linguistics*, 32, 65–112.

Downing, L.J. and Pompino-Marschall, B. (2013). The focus prosody of Chichewa and the Stress-Focus constraint: a response to Samek-Lodovici (2005). *Natural Language & Linguistic Theory, 31*(3), 647–681.

Duanmu, S. 1990. *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. PhD Dissertation, MIT.

Ford, K.C. (1975). Tone and Intonation in Kikamba, Ms., University of Nairobi.

Ford, K.C. (1976). Tone in Kikamba and the Central Kenya Bantu Languages. Studies in *African Linguistics, 7*(3), 261–293.

Franich, K. (2018). Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics, 67*, 21–33.

Frota, S. (2003). The phonological status of initial peaks in European Portuguese. *Catalan Journal of Linguistics, 2*, 133-152.

Genzel, S. and Kügler, F. (2018). Production and perception of question prosody in Akan. *Journal of the International Phonetic Association*, page 1 of 32 doi:10.1017/S0025100318000191

Goldsmith, J. (1987). Tone and accent and getting the two together. In *Proceedings of Berkeley Linguistic Society, 13,* 88–104.

Goldsmith, J. (1988). Prosodic trends in the Bantu languages. In N. Smith and H. van der Hulst & Smith (eds.) *Autosegmental Studies in Pitch Accent*. Dordrecht: Foris Publications , 81-93.

Goldsmith, J. (1991). Tone and accent in Llogori. In F. Katamba (Ed.), *Lacustrine Bantu Phonology* (7–30). Cologne: University of Cologne.

Grice, M., Ladd, D.R., and Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology 17*, 143–185.

Grief, M. (2010). Contrastive focus in Mandarin Chinese. *Proceedings of Speech Prosody.*

Hyman, L.M. (1978). Tone and/or accent. In Donna Jo Napoli, ed. Elements of Tone, Stress, and Intonation. Washington, DC: Georgetown University Press, 1–20

Hyman, L.M. (1989). Accent in Bantu: An appraisal. *Studies in the Linguistic Sciences, 19*, 115–134.

Hyman, L.M. (2006). Word-prosodic typology. *Phonology, 23*, 225–257.

Hyman, L.M. (2010). Do all languages have word accent? Or: What's so great about being universal? *UC Berkeley Phonology Lab Annual Report*, 297–315.

Hyman, L.M. (2014). Do all languages have word accent? In H. van der Hulst, ed. *Word stress: Theoretical and typological issues.* Cambridge, UK: Cambridge University Press. 56–82.

Itô, J. & Mester, A. (2012). Recursive prosodic phrasing in Japanese. In T. Borowsky, S. Kawahara, T. Shinya & M. Sugahara (eds.), *Prosody Matters: Essays in Honor of Elisabeth Selkirk*, 280–303. London: Equinox Press.

Jun, S.-A. (ed.). (2005). Prosodic Typology: The Phonology of Intonation and Phrasing . Oxford: Oxford University Press.

Jun, S.-A. (ed.). (2014). Prosodic Typology II: The Phonology of Intonation and Phrasing . Oxford: Oxford University Press.

Kanerva, J.M. (1990). *Focus and Phrasing in Chichewa Phonology*. New York: Garland.

Kember, H., Choi, J., Yu, J. and Cutler, A. (This volume). Levels of processing of linguistic prominence.

Keupdjio, H. (2015). Asymmetries on wh-movement in Bamileke medumba. *Oral Presentation at the Conference of the Canadian Linguistics Association.* University of Ottawa, May, 2015.

Kisseberth, C.W. (1992). Metrical structure in Zigula tonology (227–259). In D.F. Gowlett (Ed.), *African Linguistic Contributions*. Pretoria: Via Afrika Limited.

Kisseberth, C.W. (2010). Phrasing and relative clauses in Chimwiini. *ZAS Papers in Linguistics 53*, 109–144.

Kisseberth, C.W. & Abasheikh, M.I. (2011). Chimwiini Phonological Phrasing Revisited. In Nicole Dehé et al. (eds.), Lingua 121, Special issue 'New insights into the Prosody-Syntax interface: Focus, phrasing, language evolution': 1987–2013.

Ko, E-S. (2013). A metrical theory of Korean word prosody. *The Linguistic Review*, 30. 10.1515/tlr-2013-0004.

Kraal, P. (2005). *A Grammar of Makonde (Chinnima, Tanzania)*, PhD Dissertation. Leiden University.

de Lacy, P. (2002). The interaction of tone and stress in optimality theory. *Phonology, 19,* 1–32.

Ladefoged, P. (2005). *Vowels and consonants* (2nd ed.). Hoboken, NJ: Blackwell Publishing.

Manyah, K.A. (2006). Relation between tone and vowel quality in Twi. In *Proceedings of the Second International Symposium on Tonal Aspects of Languages*, 37–40. La Rochelle.

McHugh, B. (1990). The phrasal cycle in Kivunjo Chaga tonology. In Sharon Inkelas & Draga Zec (eds), *The Phonology-Syntax Connection. Chicago*, The University of Chicago Press: 217–242.

McHugh, B. (1999). Cyclicity in the Phrasal Phonology of Kivunjo Chaga. *LINCOM Studies in African Linguistics 03*. Munich, LINCOM EUROPA.

Michaud, A. and Vaissière, J. (2015). Tone and intonation: introductory notes and practical recom- mendations. KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik, Institut für Skandinavistik, Frisistik und Allgemeine Sprachwissenschaft (ISFAS) Abt. Allgemeine Sprachwissenschaft Christian - Albrechts - Universität zu Kiel, *Special Issue: Theoretical and Empirical Foundations of Experimental Phonetics*, 3, pp.43-80.

Moore, C.B. (1993). Phonetic observations on tone and stress in Mandarin Chinese, *Journal of the Acoustical Society of America*, *93*, 2297.

Morton, J., Marcus, S. and Frankish, C. (1976). Perceptual centers (p-centers). *Psychological Review*, *83*, 405–408.

Myrberg, S. (2010). *The Intonational Phonology of Stockholm Swedish* (ACTA Universitatis Stockholmiensis 53: Stockholm Studies in Scandinavian Philology New Series). Stockholm: Department of Scandinavian Languages, Stockholm University.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Dodrecht, Holland and Riverton, NJ: Foris.

Nganmou, A. (1991). *Modalités Verbales*. University of Yaounde dissertation.

Ohala, J.J. and Ewan, W.G. (1973). Speed of pitch change. *Journal of the Acoustical Society of America*, *53*, 345.

Odden, D. (1999). Typological issues in tone and stress in Bantu (187–215). In S. Kaji (Ed.), *Proceedings of the Symposium on Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics*. Tokyo: ILCAA.

Ou, S-c. and Guo, Z-c. (This volume). The language-specific use of F0 rise in segmentation of an artificial language: Evidence from listeners of Taiwanese Southern Min.

Patin, C. (2016). Tone and intonation in Shingazidja. In Laura J. Downing & Annie Rialland (eds.), *Intonation in African Tone Languages*. Berlin, Mouton de Gruyter: 285–320.

Patin, C. (2018). From tones to accents. *Proceedings of the Sixth International Symposium on Tonal Aspects of Language*, 17–21.

Philippson, G. (1991). Tons et accent dans les langues bantu d'Afrique orientale: étude typologique et diachornique. Thèse d'Etat, Université Paris V René Decartes, Paris.

Philippson, G. (1998). Tone reduction vs. metrical attraction in the evolution of Eastern Bantu tone systems. In *Theoretical Aspects of Bantu Tone*, ed. by L.M. Hyman & C.W. Kisseberth, 315–329, Stanford, CSLI.

Philippson, G. (2005). Pitch accent in Comorian and Proto-Sabaki tones. In Bosten, K. and J. Maniacky (eds.), *Studies in African Comparative Linguistics with Special Focus on Bantu and Mande: Essays in Honour of Yvonne Bastin and Claire Grégoire*, 199-220. Tervuren: Musée Royal de l'Afrique Centrale (MRAC).

Pierrehumbert, J. (1980) The phonology and phonetics of English intonation. PhD Dissertation, MIT. Distributed 1988, Indiana University Linguistics Club.

Pierrehumbert, J. and Beckman, M.E. (1988). *Japanese Tone Structure*. Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Port, R.F. (2003). Meter and speech. *Journal of Phonetics, 31*, 599–611.

Potisuk, S., Gandour, J. and Harper, M. (1994). F0 correlates of stress in Thai. *Linguistics of the Tibeto-Burman Area 17*(2), 1–27.

Potisuk, S., Gandour, J. and Harper, M. (1996). Acoustic correlates of stress in Thai. *Phonetica 53*, 200–220.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Remijsen, B. (2002). Lexically contrastive stress accent and lexical tone in Ma'ya. In C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7*, 585–614. Berlin and New York: Mouton de Gruyer.

Remijsen, B., and van Heuven, V.J. (2005). Stress, tone, and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology, 22*, 205–235.

Riviera-Castillo, Y. & Pickering, L. (2004). Phonetic correlates of stress and tone in a mixed system. *Journal of Pidgin and Creole Languages 19*(2), 261–284.

Roll, M., Horne, M. & Lindgren, M. (2009). Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses: An ERP study. *Journal of Neurolinguistics 22*, 55–73.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., and Yuan, J. (2014). FAVE (Forced Alignment and Vowel Extraction). Program Suite v1.2.2 10.5281/zenodo.22281

Rouder, J. N., & Morey, R. D. (2005). Relational and arelational confidence intervals: A comment on Fidler, Thomason, Cumming, Finch, and Leeman (2004). *Psychological Science*, 16, 77-79.

Scott, S.K. (1997). *P-centers in Speech*. University College London dissertation.

Selkirk, E.O. (1981). Epenthesis and degenerate syllables in Cairene Arabic. In H. Borer and J. Aoun eds., *Theoretical Issues in the Grammar of the Semitic Languages.* Cambridge, MA: MIT Press, 209-232.

Selkirk, E.O. (1984) *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Selkirk, E.O. (1986). On derived domains in sentence phonology. *Phonology, 3,* 371–405.

Selkirk, E.O. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory*. London: Blackwell.

Selkirk, E.O. (2011). The syntax-phonology interface. In J.Goldsmith, J. Riggle, and A. Yu (Eds.), *The Handbook of Phonological Theory*.Wiley, John and Sons, Incorporated.

Sietsema, B.M. (1989). Metrical Dependencies in Tone Assignment. MIT dissertation.

Simons, G.F., and Fennig, C.D. (2017). *Ethnologue: Languages of the World* (20th ed.). Dallas, TX: SIL International.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, 7, 71–79.

Tajima, K. (1998). *Speech rhythm in English and Japanese: Experiments in Speech Cycling*. Indiana University dissertation.

Tajima, K. and Port, R.F. (2003). Speech rhythm in English and Japanese. *Papers in Laboratory Phonology VI: Phonetic Interpretation.* Cambridge, UK: Cambridge University Press. 135–149.

Tchana, D. (2003) *Swăndə Medʉmba*. Bangangte, Cameroon: Comité de Langue pour l'Etude et la Production des Œuvres Bamiléké-Medumba.

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science, 33*(5), 839–879.

Truckenbrodt, H. (1996). *Phonological Phrases: Their Relation to Syntax, Focus and Prominence*. MIT dissertation.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, 30, 219–255.

Tyrchan, C. (2019). *Negation in Medʉmba*. Unpublished Bachelor's thesis, Universität Potsdam.

Volk, E. (2011). *Mijikenda tonology*. Ph.D. dissertation, Tel Aviv University.

Voorhoeve, J. (1968). Noun classes in Bamileke. *Lingua*, 21, 584–593.

Voorhoeve, J. (1971). Tonology of the Bamileke noun. *Journal of African Languages, 2*, 44–53.

Voorhoeve, J. (1976). *Contes Bamileke.* Tervuren: Musee royal de l'Afrique centrale.

Xu, Y. and Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America*, *111*, 149–174.

Yip, M. (1980). *The tonal phonology of Chinese*. MIT dissertation.

Yu, A.C.L. (2010). Tonal effects on perceived vowel duration. In *Laboratory Phonology 10,* 151–168. Berlin: Mouton de Gruyter.

Zawaydeh, B.A., K. Tajima, and M. Kitahara. (2002). Discovering Arabic rhythm through a speech cycling task. In *Perspectives in Arabic Linguistics, XIII-XIV*, Arabic Linguistic Society.

Zeller, J. (2015). Argument prominence and agreement: Explaining an unexpected object asymmetry in Zulu. *Lingua,* 156, 17–39.

[1] Still others have used 'accent' as a synonym for tone in languages where tone is culminative within a word, but where the tone is not necessarily assumed to associate with prominence (Kisseberth 2010; Kisseberth & Abasheikh 2011; Philippson 1991).
[2] Though see Downing (2010) and Hyman (2010, 2014) for discussions of cases of mismatch between penultimate lengthening and tone shift which complicate analyses involving penultimate prominence.
[3] This discussion has been further complicated by a fundamental disagreement about what the proper definition of a tone language should be. Here, we will follow Hyman (2006) in defining a tone language as one in which 'an indication of pitch enters into the lexical realization of at least some morphemes' (Hyman, 2006: 229).
[4] Though note that Ko (2013) provides evidence from acoustics and rhythmic intuition that Korean does have stress/metrical structure (and thus head prominence), which can often be manifested in the first syllable of an Accentual Phrase.

5 The abbreviations in glosses can be interpreted as follows: 1=first person; 2=second person; CL=clitic; COMP=complementizer; NEG=negation; POSS=possessive; PST=unmarked past; RPST=recent past; SG=singular; PL=plural

6 For ease of interpretation, Medʉmba translations are given in IPA format rather than in the native Medʉmba orthography

7 See arguments from Voorhoeve (1971) that some instances of downstep—particularly within the domain of possessive constructions—may be conditioned by floating tones. Note also that this analysis deviates from Danis et al.'s (2011) which posits the prosodic word as the relevant domain across which downstep can apply. We opt to treat the phonological phrase as the key domain due to the fact that there exist many places in the language where downstep fails to apply between word-size units (see the sentences in (8), for example).

8 A benefit of the hierarchical model of prosody is that it allows for explicit characterization of distinct phonological processes associated with different units within the hierarchy. While the facts described in this paper could also be described in terms of a grid-based framework, we opt for a hierarchical model in order to more clearly distinguish between processes applying at different levels of phrasing.

9 Glosses for noun classes reflect Voorhoeve's (1968) reclassification of the Bamileke noun classes from Proto-Bantu, which includes classes I-V. Therefore, CLII = noun class II, and CLIV = noun class IV.

10 One could also posit a recursive structure [ ( ʃún=ʧɑ́m )$_{PPh}$ **(** ↓jə́n ( ↓mɛ́n )$_{PPh}$ **)$_{PPh}$** ]$_{IPh}$ to account for the data (though see discussion of example (5) and footnote 16).

11 This negation marker has been described in other work as bearing a falling tone (i.e. *kə̂ʔ*) (Tyrchan 2019), and indeed, it is typically realized with a falling f0 pattern. However, native speakers appear to vary in whether they identify the word as bearing a high tone or a high-low falling contour, leading us to posit that the latter may be an innovation in some dialects which has arisen due to the lowering effect of the glottal stop (c.f. Ladefoged 2005). We leave further discussion of this variation to future work.

12 The availability of both a verb phrase constituent reading (e.g., My chief did not <u>see the child</u>, where negation only applies to the verb phrase) and a clausal negation reading (e.g., It's not the case that my chief saw the child') for the low toned negation marker in (5), but only a clausal negation reading for (4a,b), suggests that the negation marker may reside in a different syntactic position across these examples. See Dayal (2012) for an overview of cross-linguistic correlations between syntactic position and scope-taking behavior.

13 A reviewer points out another possibility, that the high preverbal phrase accent, being an *intoneme*, (rather than a *toneme*) might not be able to be targeted for downstep the way that a lexical high tone can. The implication of this analysis would be that the high preverbal phrase accent can participate as a trigger, but not a target, of downstep.

14 Note that Keupdjio (2015) analyzes the high portion of the rising contour on the tense marker in examples like (11) as an agreement marker that results from movement of a focused constituent. This account does not consider other instances of what we refer to as the high preverbal phrase accent, or the relationship between the tone of the subject and that of the negation marker.

15 Extraction of the subject to a higher clause could explain why we don't see downstep between the subject and tense marker in this construction, as downstep appears to be a clause-internal phenomenon.

16 Note that this structure incorrectly predicts that the verb should be downstepped with respect to the preceding tense marker, which occupies its own phonological phrase. Therefore, a more appropriate phrasing might involve a recursive structure as follows, where the tense marker and following verb form a maximal phonological phrase, and the verb occupies its own minimal phonological phrase: ( mvə̌n )$_{PPh}$ ( fə̌ ( njə́n )$_{PPh}$ )$_{PPh}$ ( ↓mɛ́n )$_{PPh}$. Downstep, then, could be said to target only maximal phonological phrases.

17 Others, including Bruce (1998) have argued that this prosodic element in Swedish is prominence-lending, though his analysis is considerably different from Myrberg's.

18 Franich (2018) shows that nasals in the onsets of words like the ones used for target words in Experiment 2 are not syllabic.

19 Given the proximity of subject 5's third syllable to the .25 phase position in Figure 7, it seems likely this speaker was aligning this syllable around the 1/4 SHP position, rather than the 1/6 position (and thus subdividing entire cycle in 4 beats, rather than 6).

20 Confidence intervals were calculated using the following SummarySE function in R (based on the method used in Rouder & Morey 2005): https://www.rdocumentation.org/packages/Rmisc/versions/1.5/topics/summarySE