

# Evidence for Metrical Prominence Asymmetries in Medumba

Kathryn Franich  
*University of Chicago*

## 1 Introduction

Central to the study metrical structure is that of *prominence*: whether treated in terms of hierarchically-organized headed prosodic constituents (Lieberman & Prince, 1977; Selkirk, 1978, 1986; Nespor & Vogel, 1986) or as marks on a metrical grid (Prince, 1983), syllables in a word or phrase are characterized in terms of their prominence relative to other surrounding syllables. Historically, a focus within metrical phonology on languages which have stress has enabled a straightforward characterization of metrical prominence: metrically prominent syllables are those which bear stress. More recently, however, a variety of patterns in a number of typologically different languages have led phonologists to suggest that metrical relations are active in shaping phonology independent of, or even in the absence of, stress. A prime example of this is in African tone languages: while these languages typically lack stress, they often feature phonological patterns at the tonal level, the segmental level, or both, which can be neatly characterized in terms of metrical constituents such as the foot (Akinlabi & Urua, 2003; Downing, 2004, 2010; Green, 2010, 2015; Leben, 1997, 2003; Pearce, 2006, 2007; Rose & Jenks, 2011; Weidman & Rose, 2006).

There have been various proposals for how these types of metrical structures should be viewed in relation to more traditional stress feet. Some researchers have suggested that tonal feet (as well as other types of ‘featural constituents’) may best be treated as a separate entity from stress feet (or any other constituent operating on the segmental tier), perhaps based on a different set of prosodic structures which are specific to the tonal tier (Leben, 2002, 2003; Bennett, 2012). Under such an account, tonal foot constituents might not be considered ‘metrical’ at all as far as reflecting head-nonhead prominence asymmetries (Bennett, 2012). Others have argued that stress, rather than being essential to metrical foot structure, is simply one possible phonetic manifestation of headedness which a language can (but need not) display (van der Hulst, 1999; van der Hulst & Goedemans, 2009; van der Hulst, 2012; Gussenhoven, 2004). Such treatments in principle require no division between feet which interact with stress and those which interact with tone. Indeed, given headedness itself is conceptualized as an entirely abstract phenomenon under this account, it may be possible for some languages to have metrical heads which are not detectable through any measurable means.

This paper aims to address two questions. First, is it possible to use a single type of foot structure to explain patterns at both the segmental and tonal levels in African languages? In regards to this initial question, I draw on data from Medumba, a Grassfields Bantu language spoken in Cameroon, to show that segments and tones display parallel distributional constraints which suggest that a unified foot structure is possible for both elements. A second question is: is this type of foot structure comparable to that found in languages with stress such as English? More specifically, if a single type of metrical foot structure is applicable in both cases, how can the notion of head-based prominence be properly characterized? Given the range of patterns used to evidence headedness across languages without stress, it is of great importance to establish a useful diagnostic for head-based prominence which can be applied across languages. To address this question, and working from a coupled oscillators model of the prosodic hierarchy, I present results from a series of *speech cycling* experiments (Cummins & Port, 1998) in which subjects repeat sentences in time to a metronome set to different rates. I show that metrically prominent syllables (i.e. foot heads) in Medumba, just as in English,

---

\* Thank you to John Goldsmith, Jeff Good, Alan Yu, and audience members of AMP 2017 for valuable comments on this research. All errors are my own. This work was supported by National Science Foundation Linguistics Program Grant No. BCS-1423865 (co-PIs: Kathryn Franich and Alan C.L. Yu). The National Science Foundation does not necessarily endorse the ideas and claims in this paper.

are drawn to certain lower-order phases of the repetition cycle. Thus, I claim, a useful general characterization of head-based metrical prominence is in terms of its temporal and coordinative properties in speech.

## 2 Evidence for metrical feet in African languages

A number of proposals have emerged in recent years suggesting that segmental and tonal patterns in African languages can be modeled in terms of binary metrical feet, despite the fact that these languages apparently lack word level stress. Akinlabi & Urua (2003), for example, show that the structure of verb roots in Ibibio, a Cross-River language spoken in Nigeria, is subject to size constraints which suggest roots are preferentially parsed into bisyllabic trochees. In Kera, an Eastern Chadic language, patterns of vowel harmony and tone spreading are analyzed by Pearce (2006, 2007) in terms of quantity-sensitive iambic feet. Rose & Jenks (2011) analyze patterns of tone distribution and spreading in Moro, a Khordofanian language of Sudan, through the use of trochaic feet. Downing (1990) argues for two tone spreading rules in Nguni, including a foot-based metrical spreading rule. Downing (2004, 2010) surveys segment phonotactics in a number of African languages across different families, arguing that distributional asymmetries of consonant and vowel contrasts across positions within a root (e.g. first vowel of the root versus second vowel of the root) cannot easily be captured without reference to metrical foot structure. For example, stem-initial syllables in Gokana, a Delta-Cross language, can display one of fifteen different contrastive consonants in the language, but stem-final syllables can only bear three out of these fifteen consonants (Hyman, 1990). Downing proposes that such generalizations can be captured if stem-initial syllables are treated as initial in metrical feet. Dimmendaal (2012) makes a similar argument for metrical structure in a number of Nilotic languages, citing evidence from segmental and morphophonological processes, as well as tonal processes.

Another language, Bambara, has been studied extensively with regard to foot structure. A Mande language spoken in Mali, Bambara has been the subject of several studies from the perspective of both tonal and segmental processes. Leben (1997, 2003), in some of the earliest work on foot structure in Bambara, sought to explain the distribution and association of tonal melodies to syllables. In this work, Leben proposed that feet are parsed at least on the segmental tier (and possibly also on the tonal tier) into maximally binary foot structures. Weidman & Rose (2006), looking at a different dialect of Bambara spoken in Bamako, propose an Optimality Theoretic analysis for accounting for so-called ‘edge-in’ patterns of tone assignment. Such patterns, addressed in previous work by Rialland & Badjimié (1989), pose a challenge for traditional left-to-right or right-to-left word level tone association rules as they involve (on quadrisyllabic words) a sequence of two highs followed by two lows, or vice versa. Weidman & Rose (2006) showed that, by appealing to constraints on foot structure, both tone spreading and high tone liaison patterns can be explained. While certain details of the two foot-based accounts of Bambara differ, Weidman and Rose’s proposal, like Leben’s, treats feet as maximally bisyllabic trochees which parse a word exhaustively. A third line of work by Green & Diakite (2008) and Green (2010) provides segmental evidence for trochaic feet in different varieties of Bambara. Green (2010) argues, for example, that processes of segmental reduction are constrained to occur the domain of a trochaic foot. In the most comprehensive overview of Bambara foot structure to date, Green (2015) draws direct parallels between segmental and tonal processes in Bambara, demonstrating that both can be explained with reference to a single trochaic foot structure, and eschewing the need for a separate level of foot-based organization on the tonal tier, as had been proposed by Leben (1997, 2003).

**2.1 Headedness in non-stress feet** As described above, there are a wide variety of phonological patterns which have been used to motivate non-stress foot structure across African languages. Crucially, each of these patterns operates in only a subset of the languages investigated, meaning that phonological diagnostics for foot structure across languages can be quite distinct. Furthermore, given the lack of stress in these languages, there appears to be no single property which can be used to index a particular syllable as the head of a foot. Rather, headedness may be assessed based on a syllable’s ability to attract tone (e.g. Moro), or bear a particular set of consonant or vowel contrasts (e.g. Gokana), or through some combination of these factors (e.g. Kera and Bambara). In the absence of a unified diagnostic for foot headedness across languages, the door is left open to the possibility that foot-based phonological patterns reflect distinct types of structures. For units which are diagnosed purely in terms of tone, it is possible, as suggested by Bennett (2012), that traditional notions of metrical prominence play no role, and that a separate type of organizational structure is needed. Even where tonal processes seem to parallel phonotactics at the segmental level, claims

of metrical structure may still be unwarranted. For example, it is well-known that foot-initial syllables are subject to processes of consonant fortition and other domain-initial strengthening processes, even when those syllables are not themselves metrical heads (Cho & Jun, 2000; Byrd, 2000; Cho & Keating, 2009). Such processes have been argued to stem from the privileged status of domain-initial positions, which may be crucial for word segmentation, regardless of metrical prominence (Beckman, 1998; Smith, 2002). Thus, in order to adequately assess whether feet in many African languages contain metrical prominence asymmetries, an additional diagnostic which is independent of acoustic patterns would be useful. As will be described in §4, one such diagnostic could involve examining temporal patterns of syllables hypothesized to bear different levels of prominence as they are produced in repeated phrases. Before discussing the specifics of this approach, I outline some of the segmental and tonal evidence for foot structure in the Medumba language.

### 3 Segmental and tonal evidence for foot structure in Medumba

Similar to Gokana and other languages surveyed by Downing (2004, 2010), Medumba realizes more consonantal and vocalic contrasts in stem-initial syllables than in any other position. The majority of non-compound stems are either (N)CV or (N)CVC monosyllabic or (N)CVCV disyllabic. In non-initial stem positions, only 7 of the 33 consonants and 1 of the 8 vowels are allowed (Figure 1). While stops are always realized as voiced stem-internally between vowels (and, in the case of velar and coronal stops, lenited to [ɣ] and [l], respectively), stops devoice in final position in (N)CVC forms. Only schwa is allowed in non-initial syllables. Medumba displays few segmental affixes, but those that do exist exclusively contain the vowel [ə].

Consonants				
Initial Stem (33)	Stem-medial (7)	Stem-final (7)	Prefixes (1)	Suffixes (1)
mb, mb̥, nt, nd, nts, ndz, ntʃ, nɕ ŋk, ŋg, b, tʰ, t, d kʰ, k, kʷ m, n, ɲ, ɲ, β, f, v s, z, ʒ, ts, dz, tʃ, ɕ j, l	b, ʔ, l, ɣ, m, n, ɲ	p, t, k, ʔ, m, n, ɲ	n	d
Vowels				
Initial Stem (8)	Stem-medial	Stem-final (1)	Prefixes (1)	Suffixes (1)
a, a, ə, e, o, u, u, i	N/A	ə	ə	ə

**Figure 1:** Consonant and vowel distributions by stem position and affix type

The distributional asymmetries of consonants have been shown by Danis (2011) to stem from at least in part from a word-internal lenition process, as demonstrated in examples like (1). One class of nouns which includes words such as *mbóɣó* ‘fire’ are realized disyllabically in isolation or in phrase-final position (1a), but monosyllabically phrase-initially or -internally (1b-1d).

- (1) a. *mbóɣó*  
‘fire’  
b. *mbóɣ àm*  
‘my fire’  
c. *mbók ʔʒúmó* (\**mbóɣ ʔʒúmó*)  
The fire is hot’  
d. *mbók ʔÀriâne* (\**mbóɣ ʔÀriâne*)  
‘Ariane’s fire’

As seen in (1a), the velar stop /k/ is voiced and spirantized word-internally to [ɣ] within a stem and preceding a possessive pronoun, but does not undergo these processes (collectively referred to henceforth as ‘lenition’) word-finally, even before another vowel-initial word (1d). Danis proposed the prosodic structures in (2) for (1a-1d).

- (2) a.  $\omega(\text{mbó}\gamma\acute{o})_{\omega}$   
b.  $\omega(\text{mbó}\gamma \grave{a}m)_{\omega}$



- (8) fríbán̩  
‘fry pan’

A similar set of constraints on contour tone distribution is found where the English source word would carry a diphthong, as in words such as *table* ( [ <sup>h</sup>tʰej̩b̩l̩ ] ) and *mango* ( [ <sup>h</sup>mæŋgəw̩ ]<sup>2</sup>). Diphthongs—and particularly [ej] and [əw]—are realized with a falling contour. As seen in (9a), diphthongs occurring on an initial, primary stressed syllable in the source word are realized with a falling tone contour in Medumba. However, as seen in (9b), diphthongs occurring on the second syllable of the source word (and bearing non-primary stress) require the addition of an extra syllable in the loan word, such that the contour is instead distributed as level tones over two syllables.

- (9) a. t̩bli  
‘table’
- b. m̩ŋg̩l̩l̩  
‘mango’

Since falling contour tones appear to be perfectly fine before a pause as shown in (6a-c), the generalization seems to be that such contours are dispreferred specifically on non-initial stem syllables. Or, to unify this tonal finding with those discussed for segmental distributions, contour tones can only occur in head position of a trochaic foot (10).

- (10) a.  $_{FT}(\hat{\sigma} \sigma)_{FT}$
- b. \*  $_{FT}(\underline{\sigma} \hat{\sigma})_{FT}$

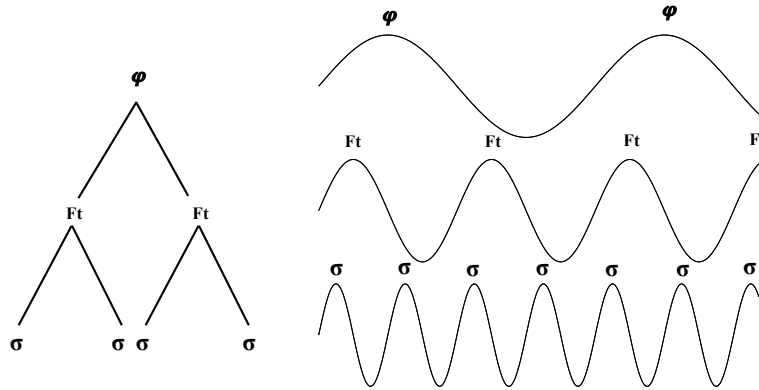
In sum, results from both segmental and tonal processes in Medumba converge on an analysis involving trochaic feet. In the next section, we test to see whether or not heads of feet in Medumba display similar properties as have been reported for stressed syllables in languages such as English in terms of their temporal alignment in repeated phrases.

#### 4 Headedness and temporal alignment: Speech cycling study

**4.1 Prosody as coupled oscillators and the speech cycling paradigm** In traditional linguistic models of prosodic structure (Selkirk, 1984; Nespor & Vogel, 1986), prosodic units are modeled within a tree-based hierarchy such that elements at each level of the hierarchy are parsed exhaustively into elements at the next highest level, resulting in a nested structure of prosodic units. For example, on the left side of Figure 2, a prosodic hierarchy is given in which syllables ( $\sigma$ ) are organized into metrical feet (Ft), which in turn are organized into prosodic phrases ( $\phi$ ).<sup>3</sup> This type of static prosodic structure can be implemented dynamically in terms of a system of multifrequency oscillators which are coordinated together or *coupled* in time (Cummins & Port, 1998; Tajima, 1998; Port, 2003; Tilsen, 2009). Such a system mimics the nested structures of the traditional prosodic hierarchy, but in addition, is predicted to display certain behaviors. First off, certain modes of coordination are predicted to be mathematically more stable than others (Haken et al., 1985; Cummins & Port, 1998; Tajima, 1998). For example, oscillators operating at different relative frequencies are predicted to be preferentially coordinated in a limited set of *frequency locking* patterns: each foot, for instance, is predicted to exhibit 1:2 or 1:3 foot to syllable frequency locking patterns, resulting in either binary or ternary feet (a 1:2 binary pattern is shown in Figure 2). Additionally, each phrase is expected to preferentially contain either two or three feet, for a 1:2 or 1:3 phrase to foot frequency locking pattern.

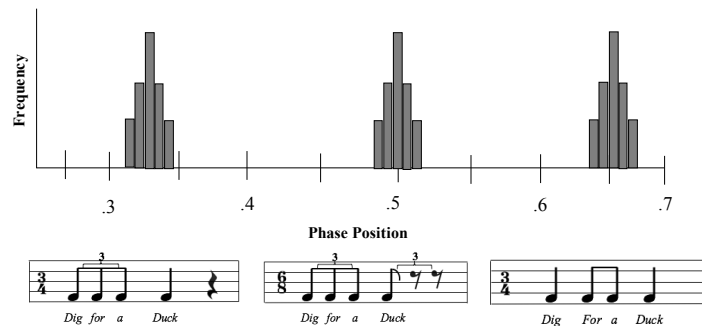
<sup>2</sup> Though it is hard to know exactly when many English words were borrowed into Medumba, it is most likely the case that these words were borrowed from British English.

<sup>3</sup> Note that the hierarchy shown excludes a prosodic word level ( $\omega$ ), which in Medumba is typically equivalent in size to the foot. One exception is where adjunction takes place between a foot and an additional syllable from a prefix or suffix.



**Figure 2:** Traditional model of hierarchical prosodic structure (left) compared with a system of multifrequency coupled oscillators (right)

A preference for lower-order 1:2 and 1:3 ratios of phrase-to-foot frequency locking can be observed in individuals' behavior on the *speech cycling* task, designed by Cummins & Port (1998) and Tajima (1998). In the task, participants repeat short phrases or sentences (e.g. 'dig for a duck') in time to a metronome set to progressively shorter periods to encourage faster and faster speech rates. Participants of different language backgrounds are routinely found to exhibit the *harmonic timing effect*, or a preference for aligning the final stressed or accented syllable of the phrase with either the  $\frac{1}{2}$ ,  $\frac{1}{3}$ , or  $\frac{2}{3}$  fractions (or *simple harmonic fractions*) of the repetition cycle, which is measured as the distance between successive repetitions of the initial syllable of the phrase. Such alignment patterns are given in musical notation in Figure 3, adapted from Cummins & Port (1998). In the first and final patterns (both in  $\frac{3}{4}$  time), the feel of the pulse is in three, as in a waltz beat, with the final stressed word *Duck* landing either one-third of the way through the repetition cycle or two-thirds of the way through. For the second pattern, shown in  $\frac{6}{8}$ , the feel is in two instead of three, with *Duck* falling on the halfway point in the repetition cycle.



**Figure 3:** Common speech cycling patterns in musical notation (adapted from Cummins & Port 1998)

Port (2003), drawing on the dynamic attending theory of Jones (1976) and Jones et al. (2002), proposes that the especially constrained temporal behavior of heads of feet in speech cycling is due to their status as perceptually prominent motor events which are timed to occur at periods of high neuronal excitation. These syllables subserve entrainment to auditory rhythmic stimuli such as (in our case) a beating metronome. Importantly, the special coordinative role that stressed foot heads in English appear to play in speech cycling is also exhibited by stressed foot heads in Jordanian Arabic (Zawaydeh et al., 2002), accented heads of bimoraic feet in Japanese (Tajima, 1998; Tajima & Port, 2003), and heads of accentual phrases in Korean,

a language which is argued to lack foot structure (Chung & Arvaniti, 2013). An open question concerns whether metrically prominent syllables lacking stress or accent might also occur in harmonic positions in speech cycling. In the following sections, we test for this possibility by examining the behavior of foot-initial syllables in Medumba within the paradigm.

**4.1.1 Stimuli** The study included two sub-experiments, each utilizing a different set of four-syllable sentences. The focus in both experiments was on the final two syllables ( $\sigma_3$  and  $\sigma_4$ ). Experiment 1 compared two Forms, STEM-INITIAL/STEM-FINAL (SISF), in which the penultimate syllable is a verb stem-initial syllable and the final syllable is verb stem-final, and STEM-INITIAL/STEM-INITIAL (SISI), in which the penultimate syllable is a verb stem-initial syllable and the final syllable is a noun stem-initial syllable (Table 1). Experiment 2 compared an additional two Forms: PREFIX/STEM-INITIAL (PreSI), where the penultimate syllable was an infinitival prefix and the final syllable was the initial syllable of the verb stem, and STEM-INITIAL/ENCLITIC (SIENC), in which the penultimate syllable was the initial syllable of a noun stem and the final syllable was a pronoun enclitic on the noun (both Forms constitute a single prosodic word) (Table 2). Assuming that stem-initial syllables in both experiments constitute foot heads, these syllables should be attracted to simple harmonic phase positions of .333, .500 and .667 with greater frequency than non-initial/non-stem syllables.

Experiment 1

Form	IPA Transcription	Gloss	Translation
SISF	mén tʃák <b>mbíbó</b>	child TNS wait	‘The child waited.’
SISI	mén tʃák <b>mbí ba</b>	child TNS lose bicycle	‘The child lost the bicycle.’
SISF	bún tʃák <b>njúmó</b>	children TNS go.out	‘The children went out.’
SISI	bún tʃák <b>ntúm mén</b>	children TNS accuse child	‘The children accused the child.’
SISF	sáj tʃák <b>mfúló</b>	bird TNS fly	‘The bird flew.’
SISI	sáj tʃák <b>njú lá</b>	bird TNS eat pineapple	‘The bird ate the pineapple.’

Table 1: Stimuli for Experiment 1. Syllables hypothesized prominent are in bold.

Experiment 2

Form	IPA Transcription	Gloss	Translation
PRESI	mén lɛ́n nɔ̀- <b>tsí</b>	child know INF-choose	‘The child knows how to choose.’
SIENC	mén lɛ́n nɔ̀= <b>tsó</b>	child know problem=their	‘The child knows their problems.’
PRESI	sáj lɛ́n nɔ̀- <b>tʃú</b>	bird know INF-enter	‘The bird knows how to enter.’
SIENC	mén lɛ́n <b>tsò=tʃú</b>	child know in.law=your	‘The child knows your in-laws.’
PRESI	sáj lɛ́n nɔ̀- <b>nú</b>	bird know INF-drink	‘The bird knows how to drink.’
SIENC	sáj lɛ́n <b>tɔ̀ŋ=ú</b>	bird know ear=your	‘The bird knows your ear.’

Table 2: Stimuli for Experiment 2. Syllables hypothesized prominent are in bold. Clitic boundaries indicated with ‘=’ and affix boundaries with ‘-’

**4.2 Method** All data were collected in Bangangté, the central city around which Medumba is spoken, and Yaoundé, the capital of Cameroon (roughly 250 km southeast of Bangangté). Participants included 13 Medumba speakers, six male, seven female. Ages of participants ranged from 19 years to 52 years. All were native speakers who had spent the majority of their lives in Bangangté or the villages surrounding it, where Medumba is the primary language of communication and commerce.

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 Audacity.

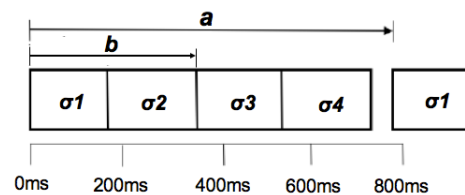
Each of the six target sentences were elicited at 15 different speech rates, from slowest (Speed 1) to fastest (Speed 15): this resulted in a total of 90 trials per participant. The slowest speech rate corresponded to a 1600 ms metronome period, which was established after pilot testing with several subjects indicated they

could not speak slower than this and maintain coordination with the metronome. Following Tajima 1998, the period was reduced by 3% for each subsequent speed, such that the fastest speed corresponded to a 579 ms metronome period. For each trial, subjects heard a total of twelve clicks of the metronome. They were asked to listen to the first four beats so as to acclimate to the new speed and then begin repeating on the fifth beat, saying the sentence once per beat.

The target sentence was displayed on the computer screen in Powerpoint in white font against a black background in both French and Medumba, in the native Medumba orthography. Most subjects could not read Medumba, and many struggled to read the French as well. Thus, for each new sentence, the experimenter read the sentence in both French and Medumba and asked the subject to repeat both forms back to verify they had understood what the target sentence was. Subjects underwent several practice trials prior to starting the experiment. Practice trials involved a separate set of sentences, and were all conducted using the slowest metronome rate. Once the subject 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.

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 Medumba (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. Datapoints corresponding to more than 2 standard deviations from the mean for a given syllable and metronome speed were removed as they most often reflected disfluencies in repetitions where subjects stumbled over a word.

P-center locations were extracted using Praat. Timing of each syllable (referred to henceforth as the syllable’s *relative phase*) was calculated relative to the phase repetition cycle (PRC), 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 4: the interval of  $b$ , which extends from the beginning of the repetition cycle to the start of the second syllable, is divided by the interval of  $a$ , the duration of the entire repetition cycle (beginning of Syllable 1 to the beginning of the following Syllable 1). This gives the External Phase measure for Syllable 2; similar measures are taken for Syllable 3 and Syllable 4.



**Figure 4:** Sample measure of External Phase. External Phase for Syllable 2 is the interval of  $b$  divided by the interval of  $a$ .

To assess clustering patterns in the data, Gaussian mixture modeling was performed on data for  $\sigma_3$  and  $\sigma_4$  in each Form, whereby datapoints were assigned probabilistically to one of several Gaussian distributions based on maximum likelihood estimation. Parametric bootstrapping of likelihood ratio statistics was carried out over 1000 iterations to evaluate the tradeoff between model fit and complexity for models consisting between 1 and 4 component Gaussian distributions. The number of components was justified at the  $p < 0.005$  level. Results of the bootstrap procedure are presented in Table 3; optimal models for each syllable consisted of either 2 or 3 component distributions.

**4.3 Results** Final mixing weights  $\lambda$  (proportion of data attributable to a particular component), component means  $\mu$  and standard deviations  $\sigma$  are provided for each syllable ( $\sigma_3$  and  $\sigma_4$ ) in each Form condition in Table 4. For the sake of space, discussion will be limited to the positions of *dominant components*, or those components which accounted for the most data while having the smallest standard deviation. In all cases, dominant components across Form conditions were also the closest in alignment across

		EXPERIMENT 1				EXPERIMENT 2			
		SISI		SISF		PRESI		SIENC	
		Obs. LRS	p-value	Obs. LRS	p-value	Obs. LRS	p-value	Obs. LRS	p-value
$\sigma_3$	$K = 1$ vs. $K = 2$	217.766	< .005	151.273	< .005	403.304	< .005	291.045	< .005
	$K = 2$ vs. $K = 3$	1.375	= .15	3.6201	< .005	87.711	=.01	9.876	= .03
	$K = 3$ vs. $K = 4$	N/A	N/A	3.052	=.53	N/A	N/A	N/A	N/A
$\sigma_4$	$K = 1$ vs. $K = 2$	113.375	< .005	135.381	< .005	364.268	< .005	377.094	< .005
	$K = 2$ vs. $K = 3$	29.685	< .005	25.260	< .005	129.168	< .005	71.296	< .005
	$K = 3$ vs. $K = 4$	26.749	=.04	17.011	=.63	.176	=.83	17.448	=.14

**Table 3:** Observed log likelihood ratio statistics and  $p$ -values for parametric bootstrap test for number of components  $K$ , Experiment 1 & 2.

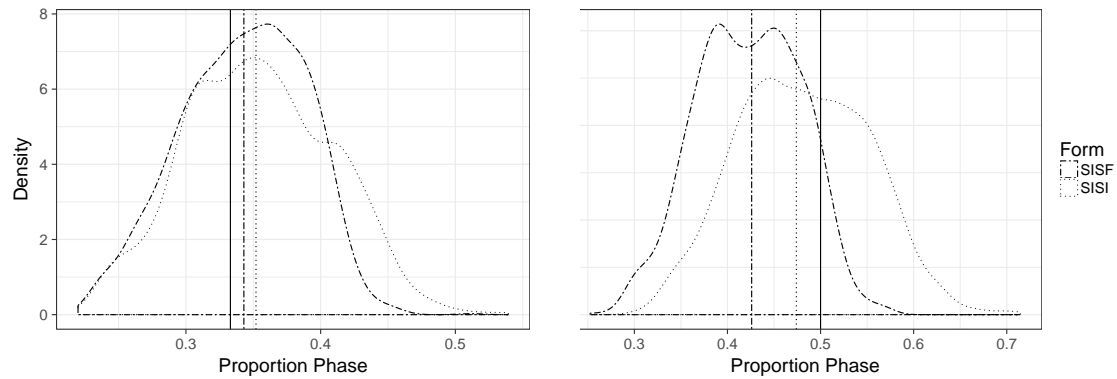
conditions. Therefore, if a significant difference was found in alignment between dominant components across conditions, it is entailed that any other comparison of components across conditions would also have been significant.

		EXPERIMENT 1						EXPERIMENT 2					
		SISI			SISF			PRESI			SIENC		
		$\mu$	$\sigma$	$\lambda$	$\mu$	$\sigma$	$\lambda$	$\mu$	$\sigma$	$\lambda$	$\mu$	$\sigma$	$\lambda$
$\sigma_3$	Component 1	.352	.056	.843	.343	.048	.510	.278	.047	.682	.321	.056	.620
	Component 2	.491	.056	.157	.453	.045	.213	.373	.070	.318	.450	.072	.380
	Component 3	N/A	N/A	N/A	.446	.112	.277	N/A	N/A	N/A	N/A	N/A	N/A
$\sigma_4$	Component 1	.474	.069	.813	.426	.063	.619	.372	.038	.221	.384	.027	.083
	Component 2	.528	.007	.019	.559	.057	.329	.487	.059	.528	.493	.067	.607
	Component 3	.627	.070	.168	.711	.059	.052	.621	.102	.251	.662	.090	.310

**Table 4:** Model parameters for Syllable 3 and Syllable 4: Experiments 1 & 2

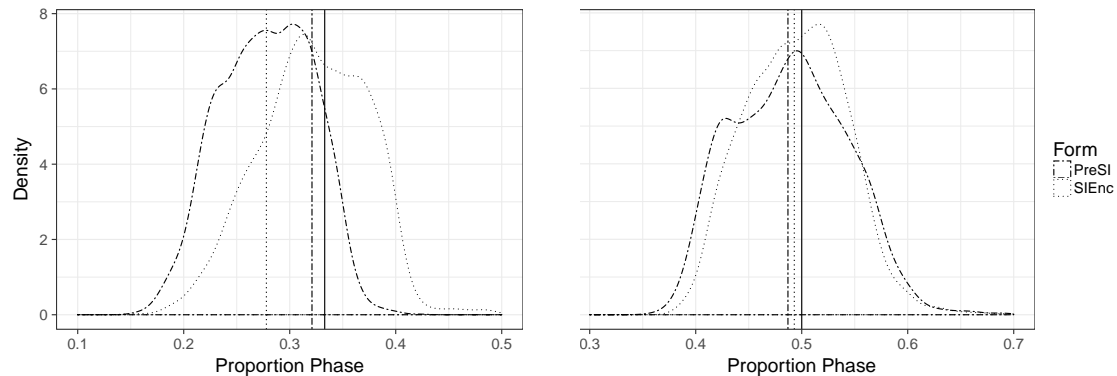
Results for the SISI and SISF conditions are shown in Figure 5. Dotted and dashed lines indicate dominant component distributions and component means (dotted for the SISI condition; double-dashed for the SISF condition). We see that the distributions of Syllable 3 stem-initial syllables in SISI and SISF occur quite close together, and both occur close to the simple harmonic phase value of  $\frac{1}{3}$ , indicated with the solid black line. For Syllable 4, the stem-final syllable in the SISF condition appears to have a much earlier phase position than the same syllable in the SISI condition, and also appears to occur farther from the closest simple harmonic phase position of  $\frac{1}{2}$ , again indicated with the solid black line. A linear mixed effect model was constructed with PROPORTION PHASE as the response variable, and which included fixed effects of FORM and METRONOME SPEED (as well as their interaction), as well as by-subject random slopes for FORM and METRONOME SPEED. The factor FORM was sum-coded and the variable SPEED was treated as continuous and was mean-centered. Results revealed no significant difference in alignment of Syllable 3 between SISI and SISF conditions ( $t = .418$ ,  $df = 10$ ,  $p = .684$ ). For Syllable 4, the stem-final syllable in the SISF condition was found to occur significantly earlier than the stem-initial syllable in the SISI condition ( $t = -13.070$ ,  $df = 12$ ,  $p < .0001$ ). A significant interaction was also found between FORM and SPEED for Syllable 4 ( $t = -5.665$ ,  $df = 3357$ ,  $p < .0001$ ), reflecting the fact that phase position varied more as a function of metronome speed in the SISF condition than in the SISI condition.

Comparing syllable alignment across the PRESI and SIENC conditions, as seen in Figure 6, we see that alignment of the prefix syllable in the PRESI condition (double-dashed line) occurred much earlier than the stem-initial syllable in the SIENC condition (dotted line), which occurred very close to the  $\frac{1}{3}$  simple harmonic phase position (solid black line). Results of mixed effects modeling revealed alignment for the PRESI condition occurred significantly earlier than that of the SIENC condition ( $t = -11.837$ ,  $df = 16$ ,  $p < .0001$ ). There was also a significant interaction between METRONOME SPEED and FORM, indicating greater change in phase position by speed in the PRESI condition than in the SIENC condition ( $t = 3.478$ ,  $df = 6175$ ,  $p < 0.001$ ). Results of mixed effects modeling indicated no significant difference in alignment position



**Figure 5:** Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 1

between the two conditions for Syllable 4 ( $t = 1.286$ ,  $df = 13$ ,  $p = .24$ )



**Figure 6:** Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 2

## 5 Discussion

Patterns of syllable alignment were largely found to support the hypothesis that stem-initial syllables in Medumba—thought to occupy the head positions of metrical feet—are associated with greater prominence than non-initial syllables and than certain types of non-stem syllables (namely, prefixes). Specifically, across both experiments, stem-initial syllables were found to be attracted more strongly to lower-order harmonic phase positions than were stem-final syllables and prefixes. These results bear a striking similarity to those which have been found for stressed syllables in languages such as English and Jordanian Arabic, as well as for accented moras or syllables in languages such as Japanese and Korean. The results thus support the notion that metrical prominence, whether cued through stress, through accent, or through neither of these properties, is observable through temporal alignment patterns of syllables. It also suggests that metrical prominence asymmetries may be much more common in the world’s languages than has been ascertained through the use of acoustic measures alone.

From a theoretical standpoint, the findings from the present work are important for several reasons. First off, the findings lend support to theories which distinguish between the notion of headedness and that of stress: Medumba provides a solid example of a language which does not have phonetic evidence of stress, but nonetheless exhibits metrical prominence asymmetries at the foot level. The findings are also important in that they shed light on another possible notion of headedness as stemming from constraints on speech production and planning. The results show that, while metrical heads are not tied to any particular acoustic

property, neither can they be treated as entirely abstract entities, as they clearly show temporal alignment patterns which are observable and measurable. A unified treatment of metrical heads might thus be one in which headedness is seen as a coordinative property of syllables in speech. A remaining question concerns why so many languages opt to manifest headedness acoustically through stress or accent.

From a practical standpoint, these results are also satisfying in that they point to a way of assessing headedness in cases where acoustic evidence is scarce. The speech cycling task is relatively easy to carry out and requires minimal special equipment to run. It is thus an especially useful tool for field-based settings where a language consultant may not have strong intuitions about where prominent syllables lie in their language. The task may also be used, as it has been here, to confirm patterns of footing or prosodic structure which are evident from phonological patterns.

One remaining question concerns the alignment of the pronominal enclitics examined in Experiment 2. These enclitic syllables, which display phonological dependence on the nouns that precede them (e.g. enclitics trigger lenition on the final consonant of a preceding noun), were hypothesized to be prosodically weak and occur far from harmonic phase positions. Instead, these syllables patterned similarly to stem-initial syllables. While this question is beyond the scope of the current paper, it is interesting to note that pronominal enclitics, similar to stem-initial syllables, also realize nearly the full range of vowel and consonant contrasts in Medumba. Recall that other prosodically weak elements, such as prefixes, only realize schwa. Thus, it seems enclitics display evidence of both phonological strength and weakness and constitute an intriguing puzzle for future work to explore.

## 6 Conclusion

In this paper, I have shown that a single type of metrical foot structure is useful for characterizing both segmental and tonal processes in Medumba, a Grassfields Bantu language. Furthermore, I have shown that, despite the lack of stress in Medumba, heads of trochaic feet in the language display temporal properties consistent with metrical prominence found in stress-based languages such as English. This finding supports the notion that only a single type of metrical foot structure is necessary across languages, and that this structure can influence the phonology of both segments and tones. Findings also support the notion that headedness is best characterized in terms of its coordinative properties in spoken language, rather than as a purely abstract phenomenon or an acoustic property of syllables.

## References

- Akinlabi, A. & E.E. Urua (2003). Foot structure in the Ibibio verb. *Journal of African Languages and Linguistics* 24, 119–160.
- Barbosa, P.A. (2003). Beat extractor praat script. Retrieved from <https://uk.groups.yahoo.com/neo/groups/praat-users/conversations/topics/921>.
- Beckman, J. (1998). *Positional Faithfulness*. Ph.D. thesis, University of Massachusetts, Amherst.
- Bennett, R. (2012). *Foot-conditioned phonotactics and prosodic constituency*. Ph.D. thesis, University of California at Santa Cruz.
- Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, 3–16.
- Cho, T. & S-A Jun (2000). Domain-initial strengthening as enhancement of laryngeal features: aerodynamic evidence from Korean. *UCLA Working Papers in Phonetics* 99, 57–79.
- Cho, T. & P. Keating (2009). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics* 29, 155–190.
- Chung, Y. & A. Arvaniti (2013). Speech rhythm in Korean: Experiments in speech cycling. *Proceedings of Meetings on Acoustics* 19:060216, 1–9.
- Cummins, F. (1997). *Rhythmic Coordination in English Speech: An Experimental Study*. Ph.D. thesis, Indiana University.
- Cummins, F. & R.F. Port (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26:2, 145–171.
- Danis, N. (2011). Downstep and contour formation in Medumba: A prosodic account. *Proceedings of the 42nd Annual Conference on African Linguistics*, 23–31.
- Dimmendaal, G. (2012). Metrical structures: A neglected property of Nilotic (and other African language families). *Studies in Nilotic Linguistics* 5, 1–26.
- Downing, L. (1990). Local and metrical tone shift in Nguni. *Studies in African Linguistics* 21, 261–317.
- Downing, L. (2004). What African languages tell us about accent typology. *ZAS Papers in Linguistics* 37, 101–136.
- Downing, L. (2010). Accent in African languages. Goedemans, Rob W. N., Harry G. van der Hulst & Ellen A. van Zanten (eds.), *A Survey of Word Accentual Patterns in the Languages of the World*, Mouton de Gruyter, Berlin, 381–427.

- 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.
- Green, C.R. (2010). *Prosodic phonology in Bamana (Bambara): Syllable complexity, metrical structure, and tone*. Ph.D. thesis, Indiana University.
- Green, C.R. (2015). The foot domain in Bambara. *Language* 91:1, e1–e26.
- Green, C.R. & B. Diakite (2008). Emergent syllable complexity in colloquial bamana. *Journal of West African Languages* 35, 45–56.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.
- Haken, H., J. A. S. Kelso & H. Bunz (1985). A theoretical model for phase transitions in human hand movements. *Biological Cybernetics* 51, 347–356.
- van der Hulst, H. (1999). Word accent. van der Hulst, H. (ed.), *Word prosodic systems in the languages of Europe*, Mouton de Gruyter, Berlin & New York, 3–116.
- van der Hulst, H. (2012). Deconstructing stress. *Lingua* 122, 1494–1521.
- van der Hulst, H. & R. Goedemans (2009). Stresstyp: A database for word accentual patterns in the world's languages. Everaert, M. & S. Musgrave (eds.), *The use of databases in cross-linguistics research*, Mouton de Gruyter, New York/Berlin, 235–282.
- Hyman, L.M. (1990). Boundary tonology and the prosodic hierarchy. Inkelas, S. & D. Zec (eds.), *The Phonology-Syntax connection*, University of Chicago Press, 109–125.
- Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review* 83, 323–335.
- Jones, M.R., H. Moynihan, N. MacKenzie & J. Puente (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science* 13:4, 313–319.
- Leben, W. (1997). Tonal feet and the adaptation of English borrowings into Hausa. *Studies in African Linguistics* 25, 139–154.
- Leben, W. (2002). Tonal feet and the adaptation of English borrowings into Hausa. *Studies in African Linguistics*, vol. 25, 139–154.
- Leben, W. (2003). Tonal feet as tonal domains. Mugane, John (ed.), *Trends in African Linguistics 5: Linguistic Typology and Representations of African Languages*, Africa World Press, Trenton, 129–138.
- Lieberman, M. & A. Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8, 249–336.
- Morton, J., S. Marcus & C. Frankish (1976). Perceptual centers (p-centers). *Psychological Review* 83, 405–408.
- Nespor, M. & I. Vogel (1986). *Prosodic Phonology*. Foris, Dordrecht, Holland and Riverton, NJ.
- Pearce, M. (2006). The interaction between metrical structure and tone in Kera. *Phonology* 23:2, 259–286.
- Pearce, M. (2007). Iambicity in Kera. Payne, D.L. & J. Pe na (eds.), *Proceedings of the 37th Annual Conference on African Linguistics*, Cascadia Proceedings Project, Somerville, MA, 66–76.
- Port, R.F. (2003). Meter and speech. *Journal of Phonetics* 31, 599–611.
- Prince, A. (1983). Relating to the grid. *Linguistic Inquiry* 14, 19–100.
- Rialland, A. & M. Badjime (1989). Réanalyse des tons du bambara: Des tons du nom à l'organisation générale du système. *Studies in African Linguistics* 20, 1–28.
- Rose, S. & P. Jenks (2011). High tone in Moro: Effects of prosodic categories and morphological domains. *Natural Language and Linguistic Theory* 29, 211–250.
- Scott, S.K. (1993). *P-centers in Speech*. Ph.D. thesis, University College London.
- Selkirk, E.O. (1978). On prosodic structure and its relation to syntactic structure. Fretheim, T. (ed.), *Nordic Prosody II*, TAPIR, Trondheim.
- Selkirk, E.O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. MIT Press, Cambridge, MA.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology* 3, 371–405.
- Smith, J.L. (2002). *Phonological augmentation in prominent positions*. Ph.D. thesis, University of Massachusetts, Amherst.
- Tajima, K. (1998). *Speech rhythm in English and Japanese: Experiments in Speech Cycling*. Ph.D. thesis, Indiana University.
- Tajima, K. & R.F. Port (2003). Speech rhythm in English and Japanese. Local, J., R. Ogden & R. Temple (eds.), *Phonetic interpretation: Papers in Laboratory Phonology VI*, Cambridge University Press, Cambridge, UK, 317–334.
- Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33, 839–879.
- Weidman, S. & S. Rose (2006). A foot-based reanalysis of edge-in tonal phenomena in Bambara. *Paper presented at West Coast Conference on Formal Linguistics 25, University of Washington*.
- Zawaydeh, B.A., K. Tajima & M. Kitahara (2002). Discovering Arabic rhythm through a speech cycling task. *Perspectives in Arabic Linguistics, XIII-XIV, Arabic Linguistic Society*.