Metrical Prominence Asymmetries in Medʉmba, a Grassfields Bantu Language

Kathryn Franich
University of Delaware
kfranich@udel.edu

Metrical Prominence Asymmetries in Medʉmba, a Grassfields Bantu Language

Abstract

There has been considerable debate as to whether word-level metrical prominence asymmetries are a universal feature of languages. African tone languages have been at the heart of this debate, as many of these languages do not show clear phonetic evidence of lexical stress. This paper explores metrical prominence asymmetries in Medʉmba, a Grassfields Bantu language, by examining such asymmetries through the lens of speech timing. Forged within a dynamical model of metrical structure, a metronome-based phrase repetition task known as *speech cycling* is used to investigate relative timing of syllables hypothesized to be metrically prominent and metrically weak. Previous research using the task has shown that metrically prominent syllables are attracted to certain relative positions within a repetition cycle. Results of two experiments show that foot heads in Medʉmba also show this behavior, supporting their status as metrically prominent. These results suggest that true metrical prominence asymmetries exist in a broader range of languages than previously been thought, and that relative timing serves as an important unifying property of metrical structure cross-linguistically.[*]

Keywords: metrical structure, prominence, accent, speech timing, Grassfields Bantu, Medʉmba

[*] Many thanks to John Goldsmith, Jeff Good, and Alan Yu for valuable feedback as this work was in development, and to Jennifer Cole and Fred Cummins for very useful discussion. Thanks also to Joe Toscano and Colin Wilson for advice on statistical modeling, and to audiences at AMP 2018, ACAL 2019, the University of Chicago, the University of Delaware, Yale University, Villanova University, Keio University, the International Christian University, and Boston University for comments and discussion. Thanks to Jacob Phillips for his detailed reading of multiple drafts and to Hermann Keupdjio for contributing feedback on the work and judgments about the Bazou dialect of Medʉmba. Thanks to Andries Coetzee, Khalil Iskarous, and two reviewers for very clear and helpful feedback on the manuscript. All errors are of course my own. Finally, many thanks to the many Medʉmba speakers who welcomed me into their community and generously contributed their time to participate in the research. Special thanks to Dr. Ange Bergson Lendja and Dr. Ariane Ngabeu for helping to facilitate the work. 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.

1. INTRODUCTION An ongoing debate in phonology concerns whether word-level metrical PROMINENCE ASYMMETRIES—also sometimes referred to as metrical 'accent'—are a universal feature of human language. One position, expressed by Goedemans and van der Hulst (2009), holds that all languages likely contain such asymmetries, and simply vary in how they are phonetically marked: while some languages utilize stress or pitch accent to convey metrical prominence, other languages use different sets of cues, or perhaps none at all. A contrasting position is expressed by Hyman (2012, 2014) who argues that, in the absence of clear and unambiguous phonetic cues to word stress, there is no satisfactory way to establish the existence of metrically driven prominence in a language. African tone languages are amongst those languages that have played an especially important role in this debate, since many of these languages contain only a subset of (or perhaps none of) the typical phonetic cues to word stress. For example, these languages may contain positional restrictions on segmental contrasts or some patterns of syllable lengthening which are characteristic of stress systems, but lack other hallmarks of a stress system, such as interaction with intonational cues or evidence for cyclic secondary stress (Hyman 2012, 2014). Since the existing cues to 'prominence' could potentially be reduced to morphologically-conditioned phonological processes or word edge effects—neither of which are necessarily 'metrical' in nature—it is argued that there is no need to appeal to metrical prominence in such languages.

This view that metrical prominence asymmetries are necessarily tied to a specific collection phonetic patterns is in large part a product of theoretical perspective. Within generative phonology, and since the development of metrical theory (Halle & Vergnaud 1987, Hayes 1984, Hayes 1995, Liberman 1975, Liberman & Prince 1977, Prince 1983, Selkirk 1984), syllable prominence has been treated as a relational property of syllables: a syllable is prominent by virtue of the fact that it resides in a structural position which is relatively 'stronger' or 'weaker' than that of a neighboring syllable. For example, in treatments of metrical stress in English from Liberman 1975 and Liberman and Prince 1977, the position of syllable stress within words is governed by constituency relations among syllables within metrical trees or grids (Figure 1).

<INSERT FIGURE 1A ABOUT HERE> <INSERT FIGURE 1B ABOUT HERE>

The precise location of 'strong' syllables within the tree or grid is language-specific, and determined based on a set of rules or constraints whose application or ranking will determine, for example, at which phrase edge the most prominent syllable (or foot) occurs, or whether stress is sensitive to syllable weight. Very often, these rules are formulated to refer directly to stress (e.g. the Weight-to-Stress Principle of Prince 1990), the implicit assumption being that the criteria for being a 'strong' syllable in a language like English is the syllable's stress-bearing status. And although stress within metrical phonology was originally treated as independent of any particular phonetic property (akin to the way a musical downbeat is treated independently of any specific musical quality), and rather more closely linked with its global timing patterns (Liberman & Prince 1977: 262), the rise of phonetically-driven accounts of metrical structure (e.g. Gordon 1999, Gordon 2002) has led to a more narrow construal of stress as characterized by increased duration and greater intensity of a syllable, as well as more extreme fundamental frequency. Since constraints within such approaches are designed to make reference to these static phonetic properties (whether directly or indirectly), there is no natural link between languages which utilize 'canonical' stress cues and those which may mark metrical alternations in other ways.

In this paper, I argue for an alternative view of metrical prominence asymmetries, which I will refer to as the DYNAMICAL VIEW, in which metrical prominence is treated as an emergent property of languages, rather than being governed by rules or constraints. The specific prosodic model I will draw on is inspired by the work of Cummins and Port (1998) and Port (2003) who recast the traditional metrical-prosodic hierarchy as a system of coupled oscillators operating at distinct frequencies which are phase-locked and frequency-locked to one another. This model is one of a family of dynamical models of metrical structure which start from the common assumption that phonological behavior can be viewed—similar to fluid flow patterns or particle movement—as resulting from laws, stated in terms of differential equations, which govern the behavior of the linguistic system in time and in response to changing system parameters, such as speech rate (Barbosa 2002, Barbosa 2007, Goldsmith 1994, O'Dell & Nieminen 1999, Prince 1993, Saltzman et al. 2008, Vatikiotis-Bateson 1988, Vatikiotis-Bateson & Kelso 1993).[1] I will show that, viewed from this perspective, we can make predictions about potentially universal behavior of metrically-prominent syllables in terms of the stability of their RELATIVE TIMING in speech production. Specifically, the model predicts that, in repeated speech, syllables residing in metrically-prominent positions should be drawn to specific, lower-order fractions—such as the halfway point—of the repetition cycle; non-prominent syllables, conversely, are predicted to occur farther from these positions, and have the potential to show greater temporal variability. Such behavior, which has already been demonstrated for a number of languages (see §3), is uniquely predicted by the dynamical model, since reference to such dynamic phonetic properties is not possible within traditional rule- or constraint-based models. This behavior is also closely tied to the types of coordinative patterns often observed between metrical prominence and musical beat strength across different languages and musical traditions (Janda & Morgan 1988, Lerdahl & Jackendoff 1983, Nancarrow 2010, O'Keeffe 2010, Pau 2015, Tarlinskaja 1993, Temperley & Temperley 2012). As I will argue in §5, these coordinative properties of metrically prominent syllables constitute an important unifying property of metrical structure across languages.

The empirical focus of this paper is on Medʉmba, a Grassfields Bantu language spoken in Cameroon. Medʉmba is one of many African tone languages which shows evidence consistent with metrical prominence patterns, but which does not meet all of the typical phonetic criteria for a metrically based stress system. In the case of Medʉmba, stem-initial syllables appear to show evidence of greater prominence than non-initial and non-initial syllables. The primary goal of the study is to investigate whether stem-initial syllables in the language show prominence-related patterns of relative timing predicted by the coupled-oscillator model of metrical-prosodic structure. In §2, I provide an introduction to the Medʉmba language and present evidence from positional asymmetries in segment contrasts which provides preliminary motivation for metrical prominence asymmetries. In §3, I introduce the speech cycling paradigm in the context of the coupled oscillator model of metrical-prosodic structure. In §4 and §5, I present results and discussion of two experiments examining temporal alignment of stem-initial, non-initial, and non-stem syllables in the speech cycling task, demonstrating that stem-initial syllables in Medʉmba do, indeed, show parallel behavior to that found for stressed syllables in languages like English. Results from the study also highlight an interesting paradox in the prosodic behavior of enclitic syllables in Medʉmba, suggesting they possess elements of metrical strength and weakness, possibly the result of a prosodic change in progress.

2. BACKGROUND ON THE MEDʉMBA LANGUAGE Medʉmba is one of several Bamileke languages within the Eastern Grassfields subgroup of Grassfields Bantu.[2] While arguably descended from Proto Bantu (Hyman 2003, Voorhoeve 1971), Medʉmba and other Grassfields languages look quite different from many well-studied Eastern and Southern Bantu languages in having more isolating morphology and fewer segmental affixes, instead featuring extensive tonal morphology. Relevant to the present paper is the fact that Grassfields Bantu languages pattern with other Bantu languages from the Northwest regions (those located in Guthrie zones A and B; Guthrie 1948) as well as several non-Bantu languages of West and Central Africa in exhibiting positional prominence effects, such that stem-initial syllables bear greater number of consonantal, vocalic, and tonal contrasts than do non-initial and non-stem syllables (see Hyman et al. 2019 and references therein). In Medʉmba, in stem-initial position (or in monosyllabic stems), 48 consonants and 11 vowels can appear (Table 1).[3] In non-initial position, only 7 consonants and one vowel ([ə]) are found.[4] Medʉmba displays few segmental affixes, but those that do exist exclusively contain the vowel [ə]. Additional positional restrictions on tone can be found in Appendix A.

<INSERT TABLE 1 ABOUT HERE>

As will be described in the sections that follow, while these distributional patterns can be characterized with reference to the stem, evidence suggests that they are more likely driven by metrical prominence asymmetries. First off, native stems are maximally disyllabic, a common restriction on foot structure. Second, the asymmetries in consonantal and vocalic contrasts in Table 1 arise, in part, from a process of lenition which targets medial consonants within this maximally disyllabic domain, a process which is quite common foot-internally across languages. Third, lenition patterns in loanwords suggest that, while stems may be longer than two syllables, they are parsed into maximally disyllabic feet, each bearing a greater number of contrasts at its left edge.

2.1 LENITION PROCESSES IN STEMS The majority of non-compound native stems in Medʉmba are either monosyllabic (N)CV or (N)CVC or disyllabic (N)CVCV. In disyllabic forms, distributional asymmetries of consonants shown in Table 1 derive in part from a lenition process, as demonstrated in examples like (2), which targets consonants occurring in stem-medial position (Danis 2011). Words such as *$^{m}b^{w}$áɣə́* ‘fire’ and *^{m}bálə́* ‘hill’ are realized as disyllabic in isolation or phrase-finally (1a-b), and as monosyllabic phrase-initially or phrase-internally (1c-d). As seen in 1a,c, the velar stop /k/ is realized as [ɣ] word-internally due to spirantization (as well as voicing). A similar pattern is found for the consonant /d/, which lateralizes to [l] in the same environments where /k/ undergoes spirantization (1b,d).

(1) Spirantization of /k/ and lateralization of /d/ (Danis 2011)

a.	^{m}b^{w}ɑ́**ɣ**ə́	‘fire’	/^{m}b^{w}ɑ́**k**ə́/
	sɑ́**ɣ**ə́	‘sauce’	/sɑ́**k**ə́/
b.	^{m}bá**l**ə́	‘hill’	/^{m}bá**d**ə́/
	^{m}vɛ́**l**ə́	‘brother	/^{m}vɛ́**d**ə́/

c. ^{m}b^{w}ɑ́**k** ↓Ánà (*mbɑ́**ɣ** ↓Ánà)
‘Ana’s fire’
sɑ́**k** ↓Ánà (*sɑ́**ɣ** ↓Ánà)
‘Ana’s sauce’

d. ^{m}bá**t** ↓Ánà (*mbá**l** ↓Ánà)
‘Ana’s hill’
^{m}vɛ́**t** ↓Ánà (*mvɛ́**l** ↓Ánà)
‘Ana’s brother’

These processes (collectively referred to henceforth as ‘lenition’) occur word-internally within a stem, but do not occur word-finally, even before another vowel-initial word (1c,d). To explain these patterns, Danis proposed the prosodic word structures in 2 for the forms in 1.

(2) Prosodic word (ω) structures proposed by Danis (2011)

a.	$_{\omega}$(^{m}b^{w}ɑ́ɣə́)$_{\omega}$	$_{\omega}$(^{m}bálə́)$_{\omega}$
b.	$_{\omega}$(^{m}b^{w}ɑ́k) $_{\omega}$(↓Ánà)	$_{\omega}$(^{m}bát) $_{\omega}$(↓Ánà)

As can be seen, lenition, per Danis' analysis, occurs intervocalically within a prosodic word (ω), but not across a prosodic word boundary.[5] This same generalization appears to hold for verbs: compare 3a, in which lenition applies stem-internally and preceding a VC object pronoun, with 3b, in which /k/ and /d/ occur at the right edge of a prosodic word and do not undergo lenition (and /d/ undergoes devoicing).

(3) Lenition in verb stems

a.	ká**ɣ**ə́	$_{\omega}$(ká**ɣ**ə́)$_{\omega}$	tɛ́**l**ə́	$_{\omega}$(tɛ́**l**ə́)$_{\omega}$
	'release'		'meet'	
b.	ká**k** ↓Ánà	$_{\omega}$(ká**k**)$_{\omega}$ $_{\omega}$(↓Ánà)$_{\omega}$	tɛ́**t** ↓Ánà	$_{\omega}$(tɛ́**t**)$_{\omega}$ $_{\omega}$(↓Ánà)$_{\omega}$
	'release Ana'		'meet Ana'	

One additional generalization (not discussed by Danis 2011) is that consonants also do not lenite stem-initially after a CV prefix (4a), despite being contained within the same prosodic word as the preceding and following vowels

(4) No lenition targeting stem-initial consonants within a word

a.	nə̀-**k**áɣə́	(*nə̀-**ɣ**áɣə́)	$_{\omega}$(nə̀-**k**áɣə́)$_{\omega}$
	INF-release		
	'to release'		
b.	nə̀-**t**ɛ́lə́	(*nə̀-**l**ɛ́lə́)	$_{\omega}$(nə̀-**t**ɛ́lə́)$_{\omega}$
	INF-meet		
	'to meet'		

Thus, the emerging generalization appears to be that lenition can target consonants which are stem-internal, but not those that are stem-initial. Due to the fact that native stems in Medʉmba are maximally disyllabic, however, it becomes unclear whether lenition patterns are better characterized with reference to the stem, or perhaps some other type of prosodic domain, such as the foot (5). Indeed, lenition of this type has been argued to be conditioned by metrical feet in other Niger-Congo languages, such as Ibibio (Akinlabi & Urua 2003); see also Downing (2004, 2010) for additional discussion of possible metrically based segmental patterns in African languages.

(5) Possible foot structures for Medʉmba nouns and verbs

a. $_{\text{Ft}}$(^{m}b^{w}ɑ́ɣə́)$_{\text{Ft}}$ $_{\text{Ft}}$(^{m}bálə́)$_{\text{Ft}}$

b. $_{\text{Ft}}$(káɣə́)$_{\text{Ft}}$ $_{\text{Ft}}$(tɛ́lə́)$_{\text{Ft}}$

To observe lenition behavior in stems longer than two syllables, we can look to Medʉmba loanwords. Unfortunately, trisyllabic loanwords in which /d/ or /k/ occur as the onset of the second syllable are scarce, preventing us from examining the precise processes of lateralization and spirantization described above. However, a parallel form of spirantization does occur in loanwords,[6] which turns /t/ and /ʧ/ to [s] when preceding a high front vowel. This process, similarly to those outlined above, does not apply to stem-initial consonants (6a), but also crucially does not target the onset of the third syllable in a trisyllabic word; only the onset of the second syllable can be targeted (6b,c).

(6) Medʉmba Loanword			English Source IPA	English Translation
a.	**t**í	(*sí)	[ti]	‘tea’
	tílò	(*sílò)	[ˈtejlə]	‘tailor’
	ʧíʧà	(*síʧà)	[ˈtiʧə]	‘teacher’
b.	sá**s**ídè	(*sátídè)	[ˈsætəˌdej]	‘Saturday’
	kí**s**ímì	(*kíʧímì)	[ˈkɪʧən]	‘kitchen’
	ʧó**s**ì[7]	(*ʧóʧì)	[ʧəːʧ]	‘church’
c.	kúbá**t**ì	(*kúbásì)	[ˈkʌbəd]	‘cupboard’
	líbá**t**ì	(*líbásì)	[ˈlibəti]	‘liberty’

It therefore appears that our earlier generalization that spirantization targets stem-internal consonants is incorrect: rather, spirantization targets alternating consonants, a type of rhythmic pattern which is also characteristic of metrically driven stress systems. A better generalization appears to be that spirantization is driven by foot-based metrical prominence asymmetries, as had been suggested in 5. The trisyllabic forms are therefore analyzed as in 7, with the initial two syllables parsed as one foot, and the final syllable parsed as its own foot; under this analysis, foot-initial syllables can be analyzed as foot heads, in which initial consonants resist lenition.

(7) Proposed foot structures for loanwords

a. $_{FT}$(sásí)$_{FT}$ $_{FT}$(dè)$_{FT}$
b. $_{FT}$(kísí)$_{FT}$ $_{FT}$(mì)$_{FT}$
c. $_{FT}$(kúbá)$_{FT}$ $_{FT}$(tì)$_{FT}$
d. $_{FT}$(líbá)$_{FT}$ $_{FT}$(tì)$_{FT}$

It should be noted that, while lenition patterns of this type are known to be conditioned by stress in some languages (Harris 1994, Honeybone 2008), it is possible that the observed segmental asymmetries arise simply due to positional effects within the foot. Syllables occurring in the initial position of a foot or word are known to undergo prosodic strengthening effects independently of their metrical prominence properties, meaning that positional prominence does not necessarily signal metrical prominence (Bennett 2012, Bennett 2013, Cho & Keating 2001, Cho et al. 2007, Fougeron & Keating 1997, Hyman 2014). As a result, the experiments presented in §4 will be crucial in providing evidence that these syllables also display rhythmic prominence which cannot be reduced to such positional effects.

2.2 LENITION WITHIN PRONOMINAL CONSTRUCTIONS There is an additional source of data which is of interest in the present work due to its implications for our understanding of foot-based patterns in Medʉmba. As described by Danis (2011), pronominal enclitics appear to trigger lenition processes which parallel those found stem-internally in some dialects of Medʉmba, including the Bangangté and Bangoulap dialects (8).

(8) Lenition in pronominal costructions

a. ^{m}b^{w}á**ɣ**=ám[8] ^{m}b^{w}á**ɣ**=ú

‘my fire’ ‘your fire’

sá**ɣ**=ám sá**ɣ**=ú

‘my sauce’ ‘your sauce’

b. ^{m}bá**l**=ám ^{m}bá**l**=ú

‘my hill’ ‘your hill’

^{m}vɛ́**l**=ám ^{m}vɛ́**l**=ú

‘my brother’ ‘your brother’

As seen in (8), lenition applies to velar and coronal stops preceding the VC pronominal enclitics *ám* and *ú.* This pattern would suggest that enclitics may behave similarly to stem-final syllables in disyllabic stems, forming a disyllabic CVCVC or CVCV foot when combined with monosyllabic stems (9).

(9) Proposed foot structures for pronominal constructions

a. $_{Ft}$(^{m}b^{w}áɣ=ám)$_{Ft}$ $_{Ft}$(^{m}b^{w}áɣ=ú)$_{Ft}$

b. $_{Ft}$(^{m}bál=ám)$_{Ft}$ $_{Ft}$(^{m}bál=ú)$_{Ft}$

An interesting difference between enclitic syllables and stem-final syllables, however, is that enclitics can realize a large range of consonantal contrasts. A sample of Medʉmba possessive and object pronouns is given in Table 2.

<INSERT TABLE 2 ABOUT HERE>

Therefore, it appears that enclitic syllables behave similarly to stem/foot-initial syllables in the range of vocalic contrasts they realize, but similarly to stem-final syllables in triggering lenition.

Another interesting observation about these enclitic syllables is that they do not trigger lenition in all dialects of Medʉmba. Speakers of Medʉmba from Bazou, for example, a chiefdom located south of Bangoulap, show a different pattern, whereby /k/ resists any kind of lenition and /d/ only partially lateralizes (10).

(10) Pronominal forms as Produced in Bazou Dialect of Medʉmba

a. $^{m}b^{w}$á**k**=ám (*$^{m}b^{w}$á**ɣ**=ám) $^{m}b^{w}$á**k**=ú (*$^{m}b^{w}$á**ɣ**=ú)
'my fire' 'your fire'

b. ^{m}bá**dl**=ám (*^{m}bá**l**=ám) ^{m}bá**dl**=ú (*^{m}bá**l**=ú)
'my hill' 'your hill'

This dialectal difference in lenition patterns is only apparent in the pronominal examples: Bazou speakers show similar patterns to Bangangté and Bangoulap speakers when it comes to stem-internal lenition. This difference across dialects could indicate that the lenition pattern in the pronominal forms in Bangangté and Bangoulap dialects reflects a recent change which Bazou speakers have not yet undergone. However these patterns have come about, pronominal enclitics constitute an interesting form to investigate given they seem to show simultaneous evidence of prosodic strength and weakness in Bangangté and Bangoulap dialects.

In §4, we will compare patterns in relative timing across stem-initial and stem-final syllables, affixes, and enclitics in order to evaluate whether asymmetries exist in the timing of these different syllable types in speech production. In particular, we will investigate whether stem-initial syllables in Medʉmba—which are reanalyzed as foot-initial syllables—show the hallmarks of temporal stability and harmonic timing predicted by the coupled oscillator model of metrical-prosodic structure when tested within the speech cycling paradigm. If our metrically-based analysis of lenition patterns is on the right track, it is predicted that stem-initial syllables should show a greater tendency towards temporal stability and harmonic timing than any of the other three syllable types. Given their mixed behavior, however, it is possible that enclitic syllables in Medʉmba will show timing patterns which are more similar to those found for stem-initial syllables than for stem-final syllables or affixes. Before delving into the study, §3 presents some background on the speech cycling paradigm and how findings from the paradigm can be interpreted with respect to the coupled oscillator model of metrical-prosodic structure.

3. METHOD: THE SPEECH CYCLING PARADIGM The speech cycling paradigm was developed by Cummins (1997), Cummins & Port (1998), and Tajima (1998) to examine speech timing as a measure of rhythmicity in language. In the simplest version of the task (see Cummins & Port 1998 for an alternative approach), subjects repeat short sentences in time to a metronome at progressively faster speeds (controlled by incrementally shortening the metronome period). The task has been carried out in a typologically diverse array of languages, including English (Cummins 1997, Cummins & Port 1998, Tajima 1998, Tajima & Port 2003, Tilsen 2009), Japanese (Tajima 1998, Tajima & Port 2003), Jordanian Arabic (Zawaydeh et al. 2002), Korean (Chung & Arvaniti 2013), and Polish (Malisz 2005). Across all of these languages, it has been found that speakers gravitate to a limited set of comfortable coordination strategies in the task, an effect known as the HARMONIC TIMING EFFECT. Specifically, it has been found that speakers prefer to align syllables bearing greater metrical prominence with lower-order fractions of the repetition cycle, such as the halfway point, or 1/3 or 2/3 of the way through the cycle; these positions are known as SIMPLE HARMONIC PHASES, or SHPs (Cummins & Port 1998).

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 2, drawn from work by Cummins & Port (1998), the beginning of the vowel in the word *duck* (or its 'perceptual center'; Morton et al. 1976) 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. While a pattern with *duck* occurring around one-third 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 greater stability (=less variability) in alignment as compared with syllables occurring farther from these positions (Cummins & Port 1998, Tilsen 2009). Tajima (1998) finds that English stressed syllables—whether they are initial or final in a word—are more likely to be drawn to SHP positions than unstressed syllables.

Similar to findings for English, speakers of Jordanian Arabic, which also has metrically conditioned stress, preferentially align stressed syllables to SHP positions (Zawaydeh et al. 2002) regardless of their position within a word. In Japanese, Tajima (1998) finds that heads of moraic trochees are drawn to SHP positions, and that foot heads are more resistant to temporal perturbations than non-heads (Tajima & Port 2003). Interestingly, for Japanese, even foot heads lacking pitch accents were found to be attracted to SHP positions, suggesting a separability of metrical and acoustic prominence. In Korean, Chung and Arvaniti (2013) find that initial syllables in the accentual phrase (AP) are preferentially aligned with SHP positions. While these authors do not posit metrical prominence for AP-initial syllables, subsequent work by Ko (2013), which examines evidence from acoustics, phonological processes, and text-setting in vocative chant, supports the idea that AP-initial syllables are metrically prominent in Korean (regardless of the pitch accent they bear). Results from these studies demonstrate the importance of metrical prominence—rather than word position or acoustic prominence—in determining speech cycling patterns.

<INSERT FIGURE 2 ABOUT HERE>

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 a second order oscillatory system, with planning oscillators which are coupled together in time (Cummins & Port 1998, van Gelder & Port 1995, O'Dell & Nieminen 1999, Port 2003). Coupling between oscillators at the level of phrase repetitions and prosodic units such as the metrical foot occurs at certain frequency-locking ratios such that the phrase is ultimately subdivided into two or three evenly timed feet. In Figure 3 (adapted from Tilsen 2009), we see side-by-side comparison of a traditional representation of the prosodic hierarchy (left) based on Selkirk (1986) and the same hierarchy represented as coupled oscillators with 2:1 syllable-to-foot ratio (yielding disyllabic feet) and 2:1 foot-to-phrase frequency-locking pattern. In a phrase repetition task, due to these frequency-locking patterns, foot-initial syllables are predicted to occur only at certain positions within the phrase-level repetition cycle: namely, at the very beginning, at the halfway point, or at the 1/3 or 2/3 points in the cycle (in the case of a 3:1 foot-to-phrase frequency-locking pattern).

<INSERT FIGURE 3 ABOUT HERE>

From the dynamical perspective, the hierarchy described in formal phonological accounts such as those in Figure 3 is seen to emerge from coordinative constraints on the speech-motor system. Preferred alignment patterns for the task are accounted for by appealing to the notion of ATTRACTORS from dynamical systems theory: certain positions within the repetition cycle (the halfway and 1/3 or 2/3 positions of the cycle) are naturally more comfortable for participants to align prominent syllables to due to (possibly universal) constraints on coordination dynamics (Haken 1996, Kelso 1994, Port 2003). Changes in alignment position during the task can be explained in terms of PHASE TRANSITIONS, which occur when a particular attractor becomes unstable due to changes in the state of the system, such as through increases in speech rate (one of various CONTROL PARAMETERS that can be altered within the system). Under these conditions, participants may move through a period of instability before achieving another stable alignment strategy. The dynamical model of prosodic structure is not only appealing due to its ability to account for observed patterns in speech cycling; recent work in neuroscience supports the idea that speech production and perception rely on entrainment of neural oscillators operating at timescales consistent with that of the syllable, word, and phrase (Ghitza 2012, Poeppel 2003), suggesting the model in Figure 3 is highly plausible from a neurocognitive perspective. Port (2003) proposes that these oscillators produce periodic pulses to which attentionally-salient events, such as stressed syllables, will align. This proposal, too, finds support from studies of atypical entrainment to rhythmic stimuli (Goswami & Leong 2013, Soltész et al. 2013).

Having been designed primarily to explain patterns of speech timing rather than phonological processes, the dynamical prosodic model does deviate somewhat in structure from that of formal models like in Figure 1 (for a comprehensive comparison, see Malisz et al. 2016). In particular, the oscillatory model does not distinguish between trochees and iambs: feet are generally interpreted in the sense of Abercrombie (1967) as including a stressed syllable and everything that follows, up until the next stressed syllable (though see Tilsen 2019 for a proposal on how trochaic vs. iambic patterns might be captured in an oscillatory model through the incorporation of word edge-specific activation patterns during speech-motor planning). Additionally, the level of the prosodic word is omitted in the dynamical model, though the model could, in principle, incorporate additional oscillators with timescales in between that of the foot and phrase. In spite of these differences, the coupled oscillator model captures many of the same generalizations which the traditional prosodic model is designed to handle: for example, coupling and frequency-locking provide a mechanism by which to capture the well-documented constraint of prosodic Strict Layering found across languages (Nespor & Vogel 1986, Selkirk 1981, Selkirk 1984, Selkirk 1986).

In the present study, we will test the hypothesis that heads of metrical feet in Medʉmba (which are argued coincide with stem-initial syllables) will be more likely to occur at simple harmonic phase positions in the speech cycling task. §3.1 describes the experimental stimuli used to test this prediction.

3.1 STIMULI The study included two sub-experiments, each utilizing a different set of four-syllable sentences, similar to those used in previous speech cycling research (Cummins 1997, Cummins & Port 1998, Tajima 1998). The focus in both experiments was on the final two syllables (σ_3 and σ_4). Experiment 1 compared two Prosodic 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 3). Sentences were grouped into sets designed to be as phonetically similar (including in tone) as possible across conditions. Experiment 2 compared an additional two Prosodic 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 4). Again, sentences were designed to be as phonetically similar as possible across conditions.[9]

<TABLE 3 ABOUT HERE>

<TABLE 4 ABOUT HERE>

3.2 STUDY PREDICTIONS Our null hypothesis is that Medʉmba does not display rhythmic differences between stem/foot-initial and non-initial/non-stem syllables. If this is the case, then we predict that any syllable should be equally likely to occur at any of the simple harmonic phase positions described in §3. If, however, rhythmic differences do exist between syllables, we predict that differences in syllable alignment should emerge between stem/foot-initial and non-initial/non-stem syllables occurring IN CORRESPONDING ORDINAL POSITIONS (e.g. σ_3 position across Prosodic Form conditions), such that stem-initial syllables occur with greater frequency at simple harmonic phase positions (SHPs), and stem-final syllables, affixes, and perhaps enclitics occur farther from these positions. The study hypotheses are formally summarized in 11:

(11) **H_0:** Syllables of all types (stem-initial, stem-final, affix, enclitic) will be equally likely to occur at SHP positions (e.g. .33, .5, and .67 proportions of the repetition cycle). Differences in relative timing for syllables in corresponding ordinal positions are not necessarily predicted.

H_1: Stem/foot-initial syllables will occur with greater consistency at SHP positions than stem-final syllables, affixes, and enclitics; as a result, relative timing should differ between syllables of different types occurring in corresponding ordinal positions within the phrase.

Broadly speaking, based on the coupled oscillator model of foot structure, non-initial/non-stem syllables should be parsed with the stem/foot-initial syllables which precede them, while a stem-initial syllable should initiate a foot and be more likely to be repeated close to an SHP position. The precise positioning of syllables across conditions will depend on various factors, however. First off, our predictions for alignment positions of stem/foot-initial syllables will change depending on whether the participant opts to subdivide the repetition cycle in 2 (aligning key foot heads with the .5 phase position) or in 3 (aligning key foot heads with the .33 or .67 phase positions, or both).

Figure 4 shows several possibilities for how syllable alignment might be expected to pattern for Experiment 1 and Experiment 2. Figure 4 pattern (a) shows, for the SISI condition, subdivision in 2 with four equally timed feet (one for each syllable of the sentence), with the final stem-initial/foot-initial syllable occurring at .5 phase position. Pattern (b) for Experiment 1 shows subdivision in 3, so that the final two syllables (both stem-initial) each head their own foot, one at the .33 position, the other at the .67 position (the tense marker in σ_2 position is seen as the least likely syllable to form its own head if only three feet are formed across the four syllables). Looking now at the SISF condition, pattern (c) shows subdivision in 2 with the final stem-initial syllable (σ_3) aligned with the .5 position. Pattern (d) shows subdivision in three with this same syllable aligned with the .33 position. Finally, pattern (e) shows subdivision in 3 with both σ_2 (the tense marker) and σ_3 (the final stem/foot-initial syllable) aligned with SHP positions .33 and .67, respectively. Comparing now between strategies for individual prosodic conditions, we see that, regardless of whether speakers adopt strategy (a) or strategy (b) for the SISI condition, and regardless of whether they adopt strategy (c), (d), or (e) for the SISF condition, σ_4 will occur at a different phase position across conditions, and will occur closer to an SHP position in the SISI condition than in the SISF condition, in line with the alternative hypothesis.

Moving to Experiment 2, patterns (a-e) represent the same alignment possibilities as for the SISI and SISF conditions. If σ_4 (the pronominal enclitic) in the SIEnc condition is found to bear metrical prominence (MP), alignment patterns should look similar to those found in the SISI condition; if it is not, alignment patterns should look similar to those found in the SISF condition (though note that neither of these two conditions will be directly compared with the SIEnc condition). Finally, for the PreSI condition, pattern (f) for Experiment 2 shows a possibility with three equal feet (the second containing both the verb stem-initial syllable in σ_2 position and the infinitival prefix that occurs in σ_3 position), with the final foot initiated at the .5 position (subdivision in 2). Pattern (g) shows these same three feet with the final foot initiated at the .33 condition. Pattern (h) shows three equal feet with both the penultimate and final feet aligned with the .33 and .67 phase positions, respectively. Again, regardless of whether participants adopt strategy (c), (d), or (e) for the SIEnc condition (assuming the enclitic is metrically weak), and regardless of whether they opt for strategies (f), (g), or (h) for the PreSI condition, both σ_3 and σ_4 are predicted to differ in their relative phase positions, with σ_3 occurring at SHP positions in the SIEnc condition, and σ_4 occurring at SHP positions in the PreSI condition, consistent with the alternative hypothesis.

<FIGURE 4 ABOUT HERE>

Note that these alignment patterns are not the only ones that could be posited for these sentences, but, based on prior research using the paradigm, these seem to us to be the most likely patterns to arise. If our predictions are correct, then, it should turn out to be the case for

Experiment 1 that one of two results will be observed: 1) σ_3 and σ_4 will occur later in the SISF condition than in the SISI condition, with σ_3 occurring in a parallel SHP position in the SISF condition to that of σ_4 in the SISI condition (assuming participants opt for patterns (a) and/or (b) with patterns (c) and/or (e)); or 2) σ_3 will be timed similarly across prosodic form conditions and σ_4 will be timed earlier in the SISF condition than in the SISI condition (assuming participants opt for patterns (a) and/or (b) with pattern (d)). What we should NOT expect is for σ_3 to occur earlier in the SISF condition than in the SISI condition, or for σ_4 to occur farther from an SHP position in the SISF condition than in the SISI condition.

In Experiment 2, more possibilities arise due to the uncertain prosodic status of the pronominal enclitic syllable, σ_4, in the SIEnc condition. Results should either show that 1) σ_3 occurs later in the PreSI condition than the SIEnc condition (assuming participants opt for patterns (f) or (h)), with σ_4 occurring in parallel SHP positions in the SIEnc condition to σ_3 in the PreSI condition (assuming participants opt for patterns (a), (b), or (e)); or 2) with σ_4 occurring somewhat earlier in the SIEnc condition than in the PreSI condition, and farther from SHP positions (assuming the enclitic is not prominent and participants opt for patterns (c) or (d)). Another possibility, should participants opt for Pattern (g), is that σ_3 and σ_4 will both occur earlier in the PreSI condition than in the SIEnc condition. What we should NOT expect is for σ_3 to consistently occur closer to SHP positions in the PreSI condition than in the SIEnc condition.

3.3 PARTICIPANTS AND PROCEDURE There were thirteen participants in Experiment 1 (6 female) and 14 for Experiment 2 (7 female). Three male participants were unavailable for Experiment 2 and were replaced with 3 other male participants of the same ages, plus one additional female participant to achieve gender balance. Ages of participants ranged from 19 to 50 years old (mean age of 32). All were native speakers of Medʉmba who had spent the majority of their lives living in or around the town of Bangangté or the village of Bangoulap.

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 potential utterances per subject). The slowest speech rate corresponded to a 1600 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 579 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. After ensuring participants knew what the target sentence was, they were asked to repeat the target sentences only in Medʉmba. 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. Data processing procedures are outlined in Appendix B.

4. ANALYSIS AND RESULTS

4.1 OVERALL ALIGNMENT PATTERNS We first examine how prosodic form condition influenced overall alignment patterns within the task. Alignment patterns were modeled using linear mixed effects models, implemented with the lmer package for *R* statistical software (Bates et al. 2017). Separate models were run for each syllable of interest (σ_3 or σ_4) in each experiment. Models included a fixed effect of PROSODIC CONDITION, which was treated as a factor with two levels (SISF vs. SISI for Experiment 1; PreSI vs. SIEnc for Experiment 2) and sum-coded[10], as well as a fixed effect for METRONOME SPEED, which was coded as a numeric variable. Initially, maximal models were built including by-subject and by-set random slopes for both variables.

These maximal models were found to be singular (i.e. variances of one or more linear combinations of effects were near zero); therefore, following Barr et al. (2013), only those random slope terms whose absence eliminated singularity were removed. This amounted to removing by-set random slopes for CONDITION and METRONOME SPEED from all models. Model *p*-values for fixed effects were derived using Satterthwaite's degrees of freedom method, implemented with the lmerTest package for *R* (Kuznetsova et al. 2017).

EXPERIMENT 1 Recall that σ_4, which was stem-initial in the SISI condition but stem-final in the SISF condition, was predicted to differ in its alignment position across prosodic conditions regardless of the overall subdivision strategy. Indeed, there was a significant effect of CONDITION found for σ_4 (estimate: -.02; $t = -9.00$; $p < 0.001$), with σ_4 occurring earlier in the SISF condition than the SISI condition. No effect of CONDITION was found for σ_3 across prosodic conditions (estimate: .001, $t = 0.64$, $p = 0.53$). Figure 5 shows that the mean alignment position for σ_4 in the SISI condition was right at .5, and slightly earlier, at .47, in the SISF condition. Mean alignment of σ_3 for both the SISI and SISF conditions was somewhat farther from an SHP position (.37 and .38, respectively), but .33 was the closest position in both conditions. This pattern is discussed further in §4.2.

<FIGURE 5 ABOUT HERE>

As has been found in previous work (Cummins & Port 1998, Tajima 1998, Zawaydeh 2002), proportion phase of syllable repetitions did increase with speech rate for both σ_3 (estimate: .01; $t = 12.21$, $p < .001$) and σ_4 (estimate: .02; 14.23; $p < .001$). As has also been found in previous work, variability in syllable alignment was slightly greater for phrase-final syllables (σ_4) than phrase-medial syllables (σ_3), with interquartile ranges (IQRs) at .12 and .10, respectively. This finding is in line with observations that phrase-final syllables tend to occur with phrase-final lengthening, and correspondingly, with greater temporal variability, following the prediction of Weber's Law (Byrd & Saltzman 1998, see also Ivry & Hazeltine 1995). Going by prosodic condition, syllables in the SISF condition had comparably greater variability in alignment than those in the SISI condition: IQR for σ_3 was .11 in the SISF condition and .09 in the SISI condition, while for Syllable 4 there was found to be an IQR of .13 in the SISF condition and .11 in the SISI condition.

EXPERIMENT 2 For Experiment 2, it was unclear whether a difference in alignment would be found for σ_4 across the SIEnc and PreSI conditions, due to the uncertain prosodic status of the pronominal enclitic in the SIEnc condition. A significant effect of CONDITION was found for σ_4, with this syllable timed later in the SIEnc condition (estimate = -.01, t = -4.20, $p < .001$). A significant difference was also found in the alignment of σ_3 across prosodic conditions, with the stem/foot-initial syllable in the SIEnc condition aligned later than the prefix syllable in the PreSI condition (estimate = -.02, t = -6.55, $p < .001$). Mean phase positions for the PreSI and SIEnc conditions were .30 and .35, respectively, for σ_3, and .50 and .53, respectively, for σ_4 (Figure 6).

Once again, proportion phase was found to increase with metronome speed for both σ_3 (estimate = .01, t = 11.69; $p < .001$) and σ_4 (estimate = .02, t = 15.05; $p < .001$). As was found for Experiment 1, IQR for syllable alignment was found to be greater for σ_4 (.14) overall than σ_3 (.11). Going by condition, variability was greater for σ_3 in the SIEnc condition than in the PreSI condition, with IQRs of .09 and .14, respectively. Variability was more comparable for σ_4 across conditions, but IQR was still slightly higher at .15 in the SIEnc condition than in the PreSI condition, where it was found to be .14.

<FIGURE 6 ABOUT HERE>

4.2 PROXIMITY OF SYLLABLE ALIGNMENT TO SHP POSITIONS AND EFFECTS OF METRONOME SPEED Results from Experiment 1 showed patterns which accorded with predictions, in that σ_4 occurred significantly earlier in the SISF condition than in the SISI condition. The lack of difference in alignment of σ_3 also followed from our predictions for alignment strategies (a) and (d) in Figure 4. In Experiment 2, differences in syllable alignment were observed for both σ_3 and σ_4 across prosodic conditions, however, the direction of this difference was different from what was predicted for σ_4. Means for some stem/foot-initial syllables across both experiments occurred farther from SHP positions than predicted. Thus, while results from both experiments suggest that rhythmic differences do exist across syllables in different prosodic positions in Medʉmba, the results do not align neatly with predictions from the coupled oscillator model.

However, as described in §3, there exists the strong possibility that speakers exhibited more than one mode of coordination during the speech cycling task: for example, the same speaker might shift from aligning a prominent syllable to the .33 phase position at slower speech rates to aligning that same syllable to the .50 or .67 position at faster speech rates. In the likely event that multiple phase positions are being targeted for the same syllable in the experiment, then the overall mean alignment value for that syllable will have little meaning in terms of evaluating the predictions of the model. If it turns out that participants are, indeed, using distinct alignment strategies throughout the task, then in order to better understand these alignment strategies, we will need to evaluate the results of the experiments not against a model with a single distribution with a single SHP value as its mean, but to a model with multiple distributions. As can be seen in Figure 7, the distributions for our target syllables all seem to show one major peak, but certain distributions (such as that for SIEnc σ_3) show a clear secondary peak, suggesting the distributions are multi-modal. This multimodality may, then, be associated with distinct alignment strategies at different speech rates, or for individual speakers.

<FIGURE 7 ABOUT HERE>

As has been found in previous speech cycling research, alignment of each syllable of interest in the task was found to move later as metronome speed increased. Taking a closer look at changes in alignment position over different metronome speeds for Set A in each experiment (Figures 8 and 9), the source of apparent multimodality in Figure 7 becomes clear: while some subjects (such as s4) showed a relatively linear change in alignment position as metronome speed increased, many subjects (such as s1, s2, and s10) showed a staircase-like pattern, such that equal changes in metronome speed sometimes led to small, and sometime large, changes in alignment position. To put it another way, as alignment position shifted later and later with increasing metronome speed, we see a pattern whereby speakers ‘stop off’ at certain alignment positions, remaining at roughly the same position over several changes in metronome speed.

To take an example, in the SISI condition, s11 shows a ‘stop off’ for σ_4 (in gray) between speeds 1-4, then shows a larger change in alignment position at around speeds 5 and 6, before stopping off again at the .5 SHP position around speed 7 and remaining at that position for several speed changes. Similarly, in Experiment 2 in the SIEnc condition, we see that s12 shows a plateau around speed 3 at the .33 SHP position for σ_3 (in black), and then another plateau around the .5 SHP position for that syllable. These patterns are consistent with the idea that at least some speakers employed more than one stable alignment strategy in the task. Looking across participants in both experiments, several can be shown to plateau in alignment position around the theoretically-determined SHP positions of .33, .5, and .67. A question now arises: if we examine clustering in the data across conditions within each experiment, are participants equally likely to align stem-final/non-stem syllables with SHP positions as they are to align stem/foot-initial syllables with these positions? This question will be investigated in §4.3.

[FIGURE 8 ABOUT HERE]

[FIGURE 9 ABOUT HERE]

4.3 ASSESSING MULTIMODALITY: GAUSSIAN MIXTURE MODELS Based on the impressionistic patterns of multimodality displayed in Figures 7, 8, and 9, we can now statistically evaluate the likelihood of multimodal patterns in the data. Were the data to represent a single overall alignment strategy, or a simple linear increase in alignment position with speech rate, we assume that alignment for a given syllable within one of the four prosodic form conditions should generally follow a Gaussian distribution, with a mean μ and standard deviation σ. Assuming multiple alignment strategies are present in the data, each associated with its own cluster of data, the data would be better modeled in terms of a *mixture* of Gaussian distributions (Figure 10), each with its own μ and σ, and each accounting for a particular proportion λ of the overall dataset for that syllable.

[FIGURE 10 ABOUT HERE]

In order to evaluate the possibility of multiple modes within the data, the data were fit to a series of Gaussian Mixture Models (GMMs) using the mixtools package for *R* statistical software

(Benaglia et al. 2009). Mathematical details of GMMs are provided in Appendix C. The mixtools package represents mixture modeling as a case of maximum likelihood estimation (MLE) in which observations are treated as incomplete data resulting from non-observed complete data. Starting with some initial parameter values θ, an expectation-maximization (EM) algorithm is applied iteratively 1) to calculate posterior probabilities of a given data point belonging to a particular component distribution, conditional on the data and the current model parameters (referred to as the E-step) and 2) to re-estimated model parameters based on the conditional probabilities obtained (referred to as the M-step). Eventually, convergence is obtained (i.e. the algorithm runs until changes to model parameters are negligible), and data is assigned to a particular component based on the new estimated probabilities. The mixing proportion λ associated with each component distribution will then indicate the proportion of the input data assigned to that component. Practically speaking, this is a similar result to what would be obtained through a clustering procedure such as *k*-means, though the component assignments for a GMM are based on 'soft labels,' meaning that component assignments are probabilistic in nature, rather than deterministic. One benefit of the GMM approach is that it allows us to handle data in which modes within the data may be close together, or even overlapping; the soft labels provide us with an estimate of exactly HOW likely a particular data point is to belong to one component or another in the mixture.

For the present research, each of the GMM parameters will be useful for helping us to understand speaker behavior in the task. Means μ of the component distributions in the GMM models will highlight the phase positions that participants were most likely to gravitate to for each syllable; we expect that stem/foot-initial syllables should occur close to SHP positions such as .33, .5, and .67 at least some of the time. Standard deviations σ of component distributions will show us how consistent participants were in aligning to a particular position: a lower standard deviation associated with a particular component distribution can be interpreted to mean that participants showed relatively lower temporal variability when aligning around that position. Previous work has shown that foot-initial syllables are not only more likely to be attracted to SHP positions in the speech cycling task, but that temporal alignment of foot-initial syllables within a particular alignment mode is overall less variable (Tajima & Port 1998). Finally, the mixing proportion λ of each component distribution can be interpreted as the overall dominance of any one particular alignment strategy for a syllable: a large mixing proportion for a component reflects the large number of data points occurring around that alignment position.

Assessing the number of mixture components An important challenge in implementing a GMM is to figure out what the optimal number of components for the distribution would be (for recent discussion of the challenges associated with this process, see McLachlan & Rathnayake 2014). The harmonic timing model of Cummins & Port (1998) on which the current analysis is based predicts that participants should gravitate to alignment positions of .33, .5, and .67 for prominent syllables, but not necessarily all three of these positions for a given syllable. It was therefore predicted that the ideal number of components necessary to model the data should be 2 or 3. To test this hypothesis, a parametric bootstrap of log-likelihood ratio statistics was implemented to evaluate the optimal number of components for the data in each syllable/prosodic form condition. Details of this procedure are provided in the Supplementary Materials which accompany this paper. Results were found to match our predictions: the bootstrap procedure revealed that distributions for most syllables were optimally modeled with a 3-component model, with an exception for σ_3 in Experiment 2, which was optimally modeled with a 2-component model for both prosodic conditions.

Model fitting Having established optimal component numbers for each syllable of interest, the data for each syllable and prosodic form condition were then fit to two different models. In the first model, means for the components were fixed to two or all three of the predicted SHP positions from Cummins & Port (1998), depending on how many components were deemed optimal for the individual syllable's model. Given that we had no prior expectations as to the mixing proportions or standard deviations for the component distributions for each syllable, initial values for these parameters were randomly chosen.[11] In the second model, initial values for means, mixing proportions, and standard deviations were all randomly chosen. For the means-specified model, where only two means were used (as was the case for those syllables/prosodic forms for which a 2-component distribution was optimal), the two SHP positions were chosen with which the data was most likely to overlap. So, for example, since the overall means for σ_3 in Experiment 1 were .38 and .39, the means were set to .33 and .5 (since there would be few data points occurring close to the .67 SHP position).

Since a more flexible model will almost always provide a better fit to the data than a more restricted model, we anticipate that the means-unspecified model (henceforth MU) will yield a better fit (indicated with a higher log likelihood) than the means-specified model (henceforth MS). However, by comparing the change in fit across the more theoretically restricted MS model with the less restrictive MU model, we can observe how well the theoretically defined means reflecting SHP positions in the more restrictive model account for alignment of the data for each syllable of interest. In other words, for a given syllable in a prosodic form condition, we can interpret smaller differences in log likelihood estimates between models to indicate closer alignment of data for a given syllable to the theoretically determined SHP positions. Based on this, we can predict that the largest differences in log likelihood across the MS and MU models will be for syllables not occurring in stem/foot-initial position.

Models were tested using cross-validation, a technique which ensures generalizability of a model within the dataset. In this case, a 3-fold[12] leave-one-out procedure was performed, in which the data is partitioned into three equal folds. Then, all logically possible pairs of folds are combined to form training samples, while the 'left out' fold for each pairing is treated as the testing set. Models were trained on each of the three training sets and then fit to the corresponding three sets of testing data. In this way, all data points were at some point contained within the training data set as well as within the testing data set. A maximum of 1000 iterations was permitted for convergence to be obtained in the Expectation Maximization procedure, and the criterion for model convergence was set to epsilon = 10^{-8}.

Chi-squared tests were used to evaluate goodness of fit for each of the models trained in the cross-validation procedure to the corresponding testing data. Model parameters and test results can be found in the Supplementary Materials which accompany this paper. For the MU models, with few exceptions, results of the Chi-squared tests were not significant at the $p < .05$ level, indicating that the trained models provided a good fit to the testing data. It was also the case that trained model parameters were quite similar across models within each syllable/prosodic form condition. Ultimately, the model was selected for each condition which provided the best fit to the testing data. This model was then fit to the overall dataset for each condition. Results from those models are presented in section 4.4 below.

4.4 GMM MODEL RESULTS

EXPERIMENT 1 Estimated model parameters for each syllable and prosodic form condition are provided in Table 4. Log likelihood values reflect model fit; higher log likelihood values indicate a better fit to the data. Density curves for best fit models are plotted against the raw data for each syllable of interest in Figure 11. Looking first at the SISI condition, we see that the best-fit model for σ_3 showed a component with a mean at the .34 phase position to which 71% of the data would most likely belong, another smaller component with a mean of .43 to which 21% of the data would most likely belong, and a third component with a mean of .53 to which 8% of the data would most likely belong. In the SISF condition, σ_3 showed a component with a mean around the .34 phase position to which 60% of the data points would most likely belong, a smaller component with a mean around the .45 phase position to which around 21% of the data would most likely belong, and a third component with a mean around .44 and a large standard deviation, to which 8% of the data would most likely belong. Thus, across both conditions, repetitions clustered most heavily around the .33 SHP position, consistent with strategies (a), (b), and (d) for σ_3 from Figure 4.

Turning to σ_4 in the SISI condition, the optimal model had a component with a mean phase position of .41 to which 14% of the data point most likely belonged, another component with a mean at .49 to which 68% of the data most likely belonged, and a third component with a mean of .61, and a relatively larger standard deviation, to which 18% of the data most likely belonged. In the SISF condition, the optimal model had a component with a mean phase position of .41 to which 49% of the data most likely belonged, another component with a mean of .52 to which 27% of the data most likely belonged, and a third component with a mean of .54 to which 24% of the data was most likely assigned. Thus, while σ_4 showed similar mean alignment positions across conditions, this syllable gravitated much more consistently to the .5 SHP position in the SISI condition than in the SISF condition, where the bulk of repetitions occurred around the .41 phase position, far from any of the theoretically-determined SHP positions. The dominant patterns for σ_4 are therefore consistent with strategies (a) and (d) from Figure 4.

[TABLE 4 ABOUT HERE]

[FIGURE 11a ABOUT HERE] [FIGURE 11b ABOUT HERE]

[FIGURE 11c ABOUT HERE] [FIGURE 11d ABOUT HERE]

As predicted, the fit of the MU model was better than that of the MS model. Based on model parameter estimates from Experiment 1, we can conclude that differences in goodness of fit between the MU and MS models for Experiment 1 are driven by the presence of additional modes of alignment present in the data which are not among the theoretically determined SHP positions. For example, participants were found to align their repetitions around the .41 phase position for Syllable 4 in both the SISI and SISF conditions, a position which is far from either the .33 or .5 SHP positions. Similarly, Syllable 3 in both the SISI and SISF conditions showed alignment between the .43 and .45 phase positions for some portion of the task. Overall, model estimates of mixing proportions indicate these alignment positions were relatively marginal compared with the .33 and .5 phase positions, to which participants more frequently aligned their repetitions. The only syllable for which alignment to a position not among the theoretical SHP positions was favored in Experiment 1 was Syllable 4 in the SISF condition, the one syllable which did not represent a stem/foot-initial syllable. Thus, as predicted, Syllable 4 in the SISF condition demonstrated the greatest difference in model fits between the means-unspecified and means-specified models (an average of an 8% drop in log likelihood, compared to averages between 3% and 5% for the other three syllables of interest). From all of this, we can conclude that the dominant patterns in the data are consistent with a scenario in which the SISI condition showed subdivision in 2, with the final syllable attracted to the .5 SHP position, and the SISF condition showed subdivision in 3, with σ_3 attracted to the .33 SHP position, as demonstrated in strategies (a) and (d) in Figure 4 for Experiment 1.

EXPERIMENT 2 Estimated parameters for MU and MS models, along with the log likelihood estimates of each model, are presented by fold for each syllable/prosodic form condition in Table 5 for Experiment 2. Density curves for best fit models are plotted against the raw data for each syllable of interest in Figure 12. The best fit model for σ_3 in the SIEnc condition yielded a component with a mean phase position of .32 to which 76% of the data would most likely be assigned, and another component with a mean position of .46 to which 24% of the data was most likely assigned. In the PreSI condition, the best fit model showed a component with a mean of

.28 to which 74% of the data would most likely be assigned, and another component with a phase position of .37 to which 26% of the data would most likely be assigned. Hence, while the stem/foot-initial syllable in the SIEnc condition occurred regularly very close to the .33 SHP position, the infinitival prefix in the corresponding syllable position occurred regularly at a phase position quite a bit earlier than the .33 position and far from any of the identified SHP positions. While these results are consistent with strategies (a) and (b) for σ_3 from Figure 5 in the SIEnc condition, σ_3 in the PreSI condition did not align with any of the predicted positions in Figure 4.

Turning to σ_4, the best fit model for the SIEnc condition had a component with a mean phase position of .38 to which 9% of the data was most likely assigned, another component with a mean phase position of .49 to which 49% of the data was most likely assigned, and a third component with a considerably larger standard deviation and a mean phase position of .62 to which 42% of the data was most likely assigned. For the PreSI condition, a component was identified with a mean of .38 to which 16% of the data was most likely assigned, another component with a mean of .48 to which 67% of the data was most likely assigned, and a third component with a larger standard deviation with a mean of .67 to which 17% of the data was most likely assigned. Based on these parameter estimates, it seems σ_4 did occur consistently at the .5 SHP position in both the SIEnc and PreSI conditions, though more consistently at that position in the PreSI condition. The third mode of alignment also had a mean position closer to one of the identified SHP positions, .67, in the PreSI condition than in the SIEnc condition. These results are consistent with strategies (a) and (f) for σ_4 in Figure 4.

[TABLE 5 ABOUT HERE]

[FIGURE 12a ABOUT HERE] [FIGURE 12b ABOUT HERE]

[FIGURE 12c ABOUT HERE] [FIGURE 12d ABOUT HERE]

Once again, model parameter estimates for the MU models suggest that participants adopted alignment strategies which fell far from the theoretically-determined SHP positions for at least part of the task. In particular, repetitions of σ_4 in both the SIEnc and PreSI conditions occurred with a phase position of around .38 for some portion of the task. The drop in log likelihood between MU and MS models can therefore be attributed, at least partly, to these differences between theoretical and actual alignment strategies in the task. As with Experiment 1, however, based on the best fit models, these unpredicted alignment positions accounted for a relatively small percentage of the data compared with the theoretically-determined SHP positions. There were also clear differences across syllables in Experiment 2 in terms of how large the drop in log likelihood was between the MU and MS models: in particular, while the MS model for σ_3 in the SIEnc condition showed less than a 1% drop in log likelihood from the MU model, the drop was at 11% for σ_3 in the PreSI condition. The drop in log likelihood was at around 5% for σ_4 in the PreSI condition and 2% for the same syllable in the SIEnc condition.

Alignment patterns in Experiment 2 deviated in some ways from our expectations: while the dominant pattern in the SIEnc condition, where Syllable 3 occurs close to the .33 phase position and Syllable 4 occurs close to the .5 position, is consistent with pattern (a) in Figure 4, alignment positions in the PreSI condition did not match any of the predicted patterns from Figure 4. While σ_4 did consistently occur close to the .5 SHP position in this condition, consistent with strategy (f), σ_3 occurred quite early, consistently surfacing around a phase position of .28. This was considerably earlier than the predicted position given strategy (f) in Figure 5, but still far from any of the theoretically determined SHP positions. This pattern would seem to indicate that speakers, rather than constructing three feet in this condition, created instead a simple two-foot pattern, with one ternary foot headed by σ_1 and with σ_2 and σ_3 as dependents.

4.5 RESULTS SYNTHESIS The goal of the experiments presented in §4 was to test the hypothesis that foot-initial syllables in Medʉmba, which are hypothesized to align with the initial position of stems, would show greater metrical strength in the speech cycling task, observable through their timing at prominence-attracting simple harmonic phase positions in the task. Meanwhile, we predicted that less prominent stem-final syllables and non-stem syllables, such as prefixes, should occur less consistently at these positions. After highlighting overall differences in alignment of foot/stem-initial and non-initial/non-stem syllables, a statistical modeling approach was adopted involving Gaussian Mixtures to evaluate the frequency with which speakers aligned their repetitions of syllables to harmonic phase positions. By examining model fits with and without means restricted to theoretically determined harmonic phase positions, we determined that, while speakers showed a variety of alignment strategies throughout the task, they were more likely to align repetitions of foot/stem-initial syllables to harmonic phase positions and non-initial/non-stem syllables to positions other than harmonic phase positions. These findings are consistent with our hypothesis that stem-initial syllables bear greater metrical prominence than non-stem/non-initial syllables. Pronominal enclitics were found to pattern more like stem/foot-initial syllables in terms of the relative phase positions they gravitated to, suggesting that they, too, bear metrical prominence.

5. DISCUSSION Findings from the present study run counter to suggestions (e.g. from Hyman 2015) that feet in languages like Medʉmba are either absent or lacking in true metrical prominence asymmetries. Heads of feet in Medʉmba, like those in languages like English, show evidence of rhythmic prominence, a hallmark of metrical structure. From a typological perspective, these findings are important, since they confirm that at least some African tone languages have genuine metrical prominence asymmetries at the word level. And while the findings certainly do not, in and of themselves, prove the universality of word-level metrical prominence asymmetries across languages, they are consistent with this possibility.

These findings are also important from an empirical perspective in that they demonstrate that the speech cycling paradigm may be a particularly useful tool for examining metrical prominence asymmetries in languages which lack unambiguous phonetic cues to stress. Though this work has not presented direct evidence of a lack of typical stress cues—such as increased duration—on foot/stem-initial syllables in Medʉmba, as mentioned in §2, the confound of domain edges makes this type of analysis unreliable as a measure of metrical prominence in Medʉmba, as well as in many other languages. The speech cycling results are not prone to this confound. Indeed, given the apparent heterogeneity of phonetic and phonotactic cues to metrical prominence asymmetries across languages (see e.g. Gordon & Roettger 2017), the coordinative property of metrically-prominent syllables highlighted in the present work may be one of the most important unifying factors underlying metrical prominence cross-linguistically.

Having said this, it is interesting to consider why different language groups may have evolved to have such varied correlates to metrical prominence. A proposal by van der Hulst et al. (2017) states that phonetic and phonological correlates of rhythmic prominence at the word level arise through the grammaticalization of 'low-level rhythm' (167) which the authors take to be a universal property of languages. Just what is meant by 'low-level rhythm' is not entirely clear in this context, but one possibility is that the coordinative function of metrical prominence is what initially drives the emergence of rhythmic asymmetries in a language, and that language-specific factors related to, for example, prosody, syntax, and information structure—and perhaps coordinative preferences, themselves—influence how different cues to metrical prominence emerge and evolve. More research will be needed in order to establish exactly how this process might come about. One important step in this direction will be to examine how distinct coordination patterns can contribute to phonetic variation (see Tilsen 2011, for example, for a demonstration of how metrical coordination interacts with articulatory patterns in English).

5.1 PROSODIC STATUS OF PRONOMINAL ENCLITICS Results from the speech cycling study also revealed that participants showed mostly similar alignment strategies for enclitic syllables as they did for stem-initial syllables. In particular, the drop in log-likelihood between models with means specified to harmonic phase positions versus without means specified was small—the smallest for any syllable across the two experiments—indicating the more restrictive model actually fit the data for the pronominal enclitics quite well. This was similar to the findings for stem-initial syllables, suggesting that both syllable types bear relatively greater metrical prominence than stem-final and affix syllables. Given the variability in lenition patterns across dialects for pronominal enclitics, one possibility is that lenition constitutes a change in progress which has started with the Bangoulap dialect and which may or may not spread to other dialects like Bazou. An interesting question concerns whether pronominal enclitics may eventually be reanalyzed as having a different prosodic status, such that they will eventually show timing behavior more similar to that found for metrically weak stem-final syllables. Prosodic weakening of functional elements such as pronouns is of course quite common cross-linguistically (Zwicky 1977, Jeffers & Zwicky 1980), particularly in post-tonic positions (Barnes 2002, Hyman 2008). Of note here is the fact that segment-level evidence of cliticization appears to precede metrical weakening (or 'deaccenting'), while it has been claimed that the typical diachronic trajectory to cliticization involves the opposite sequence of events (Jeffers & Zwicky 1980).

5.2 WORD- VS. PHRASE-LEVEL PROMINENCE Going forward, there are various questions which will be important to investigate further concerning the nature of the speech cycling paradigm and how it operates across languages. First off, discussion has been limited in this paper to variation in metrical prominence at the word level, but the possible influence of phrase-level prominence in determining speech cycling results has not been discussed. As mentioned in §3, the syllables under study in much of the previous work on speech cycling in English have not only borne lexical stress, but also nuclear accent at the phrase level. The present work provides clear evidence that not all syllables within a word in Medʉmba have equal potential for rhythmic prominence; if the opposite were true, we would expect that any syllable in a word would be equally likely to be attracted to the theoretically determined simple harmonic phase positions, which was not the case. The fact that these prominence asymmetries also align with foot-level phonotactic asymmetries provides additional evidence for the role of metrical structure in driving speech cycling patterns. However, the possibility that phrasal prominence may also be playing a role in determining speech cycling patterns has not been ruled out, due to the fact that phrasal prominence may be 'projected' from the foot/word level, as is the case in English (Liberman & Prince 1977). Indeed, recent work by Franich (2019) has shown that syllables in Medʉmba occurring earlier within a sentence (e.g. the second or third syllable within a longer, six-syllable utterance) which bear acoustic evidence of phrase-level accent are also drawn more consistently to SHP positions than syllables not bearing the accent. Future work will need to examine in depth the relative contributions of word- and phrase-level prominence in predicting results in speech cycling cross-linguistically. This is an especially interesting question as it concerns languages, such as Turkish, in which metrical structure has been argued not to play a role in lexical representations of all words, and where rhythmic stress patterns may be conditioned by higher-level prosodic constituents (Kabak & Vogel 2001, Özçelik 2014).

5.3 RELATIVE TIMING AND ITS RELEVANCE TO PHONOLOGY The benefit of the coupled oscillator model of metrical-prosodic structure is that it allows us to characterize beat strength in terms of a potentially universal phonetic property: the harmonic timing effect. The centrality of timing in phonological structure is certainly not unique to the case of metrical structure described here. Models of phonological structure which assume articulatory gestures as phonological primitives have also demonstrated the importance of relative timing as an index of phonological structure (Browman & Goldstein 1988, Browman & Goldstein 1992, Gafos 2002, Shaw et al. 2009, Shaw et al. 2011). Some of these approaches have incorporated the architecture of coupled oscillators and phase-based timing to characterize structural differences in the organization of syllable onsets and codas, for example (Goldstein et al. 2009), as well as to characterize coordination between segments and tones (Gao 2008, Katsika et al. 2014), and between conversation partners' turns (Wilson & Wilson 2005). Variability in the timing of individual speech gestures has been shown to be closely linked to variability of rhythmic timing in the speech cycling task, suggesting a unified system of speech planning which underlies both aspects of timing (Tilsen 2009). As with the results of the present study, the emergence of a small number of stable coordination patterns across these various domains in language production is uniquely captured within the dynamical account of phonological structure, and not so easily explained within generative approaches assuming a separation between static, symbolic phonological units and their phonetic implementations. Having said this, there is still much work to be done in terms of understanding the underlying dynamics which shape speech behaviors in these striking ways. Future work in phonology would benefit greatly from a deeper understanding of the physical laws which govern speech timing.

5.4 LINKS BETWEEN SPEECH CYCLING AND MUSICAL STRUCTURE Temporal patterns being investigated in the present study are constrained to highly regulated speech contexts, in which participants are essentially asked to chant speech, rather than to speak with the typical level of variability found in naturalistic contexts. One might enquire, then, as to whether the findings from this study are truly representative of rhythmic patterns in language as it is typically produced. In fact, many of the introspections that served as the initial basis for metrical theory itself were taken from text-setting in music, poetry, and vocative chant (e.g. Kiparsky 1975, Liberman 1975, Hayes 1983). In the case of English and other languages which have been examined, the speech cycling paradigm appears to tap into the same notion of beat strength which is exploited for creative purposes in music and poetry. By constraining speech to be less susceptible to variability found in conversational speech, the speech cycling paradigm is uniquely fit for examining the harmonic timing effect, while also avoiding the potential for expressive phrasing found in music and poetry, which can complicate the relationship between metrical intuitions and phonetic structure.

Related to this last point, it is interesting to reflect on how closely speech cycling results might align with text-setting patterns in the musical domain across languages and cultures. The speech cycling paradigm is built on a well-supported model of coordination dynamics which correctly predicts that certain types of coordination, despite cultural variation, will be universally observed; these dynamics regulate motor behaviors from walking gait, to multi-joint arm movements, to speech production (Haken et al. 1985, Kelso 1995). This is to say that the speech cycling task, regardless of the cultural environment in which it is carried out, is expected to yield similar overall results with regard to speech coordination. On the other hand, in the musical domain, given the many ways in which musical traditions across the world organize rhythms and map them in creative ways to text, it seems doubtful that syllables found to be metrically strong from a linguistic standpoint in speech cycling will always align predictably to musical beats as straightforwardly as is found, for example, with English folk verse. Looking at the Cameroonian context, for instance, some types of traditional rhythms (frequently referred to by music theorists as POLYRHYTHMS) can invoke the sense of multiple pulses at once, such that it may not be clear which pulse (if any) should be expected to align with rhythmic alternations in speech. These types of rhythms coexist alongside other, more 'square' rhythms (some of which are found in the context of Christian hymns written by European composers) for which text-setting norms may be entirely different. Despite the many cultural variations in rhythm aesthetics, examining text-setting from a rhythmic perspective will no doubt provide additional, rich insight as to the relationship between metrical patterns and beat strength cross-culturally.

Indeed, the utility of exploring links between linguistic structure and musical structure goes beyond the development of a theory of grammar. While language and music have traditionally been viewed as being controlled by separate cognitive systems (Pinker 1997), evidence continues to emerge which suggests a greater level of overlap in processing across the two domains than was previously thought (see Slevc & Okada 2015 for a recent overview). Patterns of rhythmic neural entrainment have been found to be crucial to the ability to decode the speech signal, predict upcoming speech, and coordinate with a conversation partner (Giraud & Poeppel 2012, Lakatos et al. 2019). Links between music and language are also abundant within behavioral research: for example, the ability to manually synchronize to a musical drumbeat has been shown to be a good predictor of speech perception and reading skills among children and adults (independent of factors such as IQ and vocabulary size) (Tierney & Kraus 2013, Woodruff Carr et al. 2014). These findings highlight the potentially deep evolutionary link between language and music, a connection which has been hypothesized to exist since Darwin (1871), but for which far more concrete evidence has emerged as of late. An important role for linguists to play in advancing understanding of possible evolutionary links between language and music is in documenting rhythmic properties of the world's languages from a variety of perspectives.

6. CONCLUSION The evidence provided here for metrical prominence asymmetries in Medʉmba is some of the first concrete evidence of such asymmetries in an African tone language. The results therefore allow us to conclude that metrical prominence asymmetries play a role in a broader range of languages than was previously thought; this finding, while not in itself conclusive, provides fuel for the argument that metrical prominence asymmetries are universal across languages, regardless of their specific phonetic-prosodic profiles. Central to the present approach is the idea that relative timing in speech provides a key unifying correlate of metrical prominence across languages. From the dynamical timing-based perspective presented here, we can make predictions about metrical strength which are extraordinarily detailed: not only do we expect overall timing differences between prominent and non-prominent syllables, but we expect prominent syllables to occur at specific temporal intervals. The speech cycling paradigm could provide an important empirical tool for understanding rhythm and metrical prominence cross-linguistically, particularly for languages where other phonetically observable measures of prominence may be absent or insufficient to motivate metrical structure in the ways it is typically diagnosed.

APPENDIX A: RESTRICTIONS ON THE POSITION OF CONTOUR TONES

Phonemically, Medʉmba has only a binary tonal contrast between high and low tones;[13] falling and rising tones can occur in some contexts, though they are analyzable as sequences of level tones (Clements & Goldsmith 1984, Goldsmith 1976, Leben 1971). For example, untensed and tensed verb stems (the latter bearing a nasal prefix) surface in affirmative, non-focused contexts with one of two tone melodies, H or LH, and can be either monosyllabic or disyllabic. As can be seen in (12), monosyllabic verbs can host LH contours, while the LH melody is distributed as a sequence of two level tones on a disyllabic verb. It is standardly assumed, in order to avoid a stem-internal Obligatory Contour Principle violation (Leben 1971), that the high tone in disyllabic examples such as those in (12a) is a single tone linked to both stem syllables (Hyman & Tadadjeu 1976).

(12) Mono- and disyllabic verbs with H and LH tone melodies

a.	zí	'sleep'	zínə́	'walk'
	ʒʉ́	'eat'	ʒʉ́mə́	'be dry'
b.	bǎ	'be ripe'	bàɣə́	'split'
	sǒ	'press'	sòŋə́	'throw'

Thus, while contour tones are permitted on stem/foot-initial syllables in some contexts, such as (12b), contours never occur on a non-initial syllable. In fact, evidence from loanwords from English suggests that contours are explicitly *disallowed* from occurring on syllables which are not stem/foot-initial. English loanwords are typically incorporated with a high tone assigned to the stressed syllable(s) of the source word (or to the only syllable, if monosyllabic), and a low tone to the unstressed syllable (13a), though HL falling contours are found in certain loanword contexts, such as words with a CVN syllable shape (13b)[14].

(13)	Medʉmba Loanword	English Source IPA	English Translation
a.	tí	[ti]	'tea'
	bá	[bajk]	'bike'

	hámə̀	[ˈhæmə]	'hammer'
	tílò	[ˈtejlə]	'tailor'
b.	pîn	[pɪn]	'pin'
	tâm	[tajm]	'time'
	ⁿgûmnə̀	[ˈgʌvnə]	'governor'
	sîŋlì	[ˈsɪŋglət]	'singlet'

Interestingly, HL contours are only permitted on CVN syllables occurring in stem-initial position: if a disyllabic word with a second syllable of CVN shape is borrowed, an epenthetic vowel is inserted after the nasal (either [i] or [ə], depending on the place of articulation of the preceding consonant), and the HL contour is distributed as separate H and L tones across the final two syllables (14).[15]

(14)	Medʉmba Loanword		English Source IPA	English Translation
	dósínì	(*dósîn)	[ˈdʌzən]	'dozen'
	flébánə̀	(*flébân)	[ˈfɹaj ˌpan]	'fry pan'

The above patterns can be accounted for if we assume that 1) where possible, Medʉmba builds maximally disyllabic feet starting from the left edge of the stem; and 2) syllables bearing contour tones are banned from occurring in the dependent (rightmost) position of a foot; and 3) vowel epenthesis is enacted as a repair strategy to enable the second syllable in the forms in (14) to be parsed as the dependent member of the disyllabic foot. This predicts that stems with odd numbers of syllables will relegate unparsed syllables or degenerate feet to the right edge.

Further evidence of left-to-right foot parsing comes from tone spreading patterns in trisyllabic English loanwords. Loanwords in Medʉmba whose English source words are trisyllabic initial stress reveal an interesting pattern with respect to tone assignment: while high tones are expected on the initial syllable due to stress in the source word, high tones are also found on the second syllable, even where the syllable is unstressed in the source word (15).

(15)	Medʉmba Loanword	English Source IPA	English Translation
	kámárà	[ˈkæməɹə]	‘camera’
	kábí^{n}də̀	[ˈkapɪntə]	‘carpenter’
	ɛ́^{m}básì[16]	[ˈɛmbəsi]	‘embassy’
	bítálì	[ˈbɪtə ˌlif]	‘bitter leaf’

This pattern is reminiscent of one described by Leben (1997, 2002) for languages such as Hausa and Bambara (see also Green 2015, Rialland & Badjimé 1989, and Weidman & Rose 2006). For English loanwords in Hausa, Leben posits that the primary lexical tonal is HL, and that words containing three syllables exhibit foot-based, binary tone spreading between the initial two syllables (the first of which bears stress in the English source). A similar scenario can account for these patterns in Medʉmba (16,17). Recall that stem-internal tone spreading is also common to native words in Medʉmba.

(16) $_{FT}$(kámá)$_{FT}$ $_{FT}$(rà)$_{FT}$
$_{FT}$(kábí)$_{FT}$ $_{FT}$(^{n}də̀)$_{FT}$
$_{FT}$(ɛ́^{m}bá)$_{FT}$ $_{FT}$(sì)$_{FT}$
$_{FT}$(bítá)$_{FT}$ $_{FT}$(lì)$_{FT}$

(17) Association of HL tone melody to trisyllabic loanwords

H L

ka ma ra

Trisyllabic examples such as *sásídè* ‘Saturday,’ *kísímì* ‘kitchen,’ *kúbátì* ‘cupboard,’ and *líbátì* ‘liberty’ demonstrate that lenition and tone spreading converge on the same foot patterns (18) (see §2.1).

(18) $_{FT}$(sásí)$_{FT}$ $_{FT}$(dè)$_{FT}$
$_{FT}$(kísí)$_{FT}$ $_{FT}$(mì)$_{FT}$
$_{FT}$(kúbá)$_{FT}$ $_{FT}$(tì)$_{FT}$
$_{FT}$(líbá)$_{FT}$ $_{FT}$(tì)$_{FT}$

APPENDIX B: DATA PROCESSING

Following Tajima (1998), of the eight total repetitions per metronome speed, the first and last repetitions of each phrase were excluded from analysis since they tended to be more variable in participants’ production. 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).[17] 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. Diagrams indicating script-generated p-center locations are shown for utterances of four different phrases (those constituting Set A for Experiment 1 and 2, respectively) in Figure B1.

[INSERT FIG B1 TOP LEFT ABOUT HERE] [INSERT FIG B1 TOP RIGHT ABOUT HERE]

[INSERT FIG B1 BOTTOM LEFT ABOUT HERE] [INSERT FIG B1 BOTTOM RIGHT

ABOUT HERE]

FIGURE B1. Examples of script-generated p-center locations (dashed lines) inserted for Set A sentences for Experiment 1 (top) and Experiment 2 (bottom)

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 B2: 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.

<INSERT FIG B2 ABOUT HERE>

FIGURE B2. Relative phase for Syllable 3 (σ3) in a four-syllable utterance is interval *b* divided by interval *a*

Not all speech rates were possible for all participants. While most speakers were able to produce the target sentences up until speed 9 or 10, few participants were able to consistently produce the sentences at the highest speech rates. Once a participant made repetition errors in successive utterances of a particular sentence, they were advanced to the next sentence. In order to minimize imbalance in our data between the conditions of interest, within a given experiment, we only compared data for each subject for which there was a match in conditions across metronome speeds. In other words, if a participant was able to complete the task at one speed in one prosodic condition but not the other, we excluded the unmatched data from the higher speeds in the first condition. Summary tables of the speeds each participant was able to reach, organized by Set, are provided for each experiment in Tables S1 and S2 of the Supplementary Materials which accompany this paper.

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 around 10% of the remaining data. Total numbers of data points analyzed by experiment and condition are provided in Table S3 of Supplementary Materials.

APPENDIX C: ARCHITECTURE OF GAUSSIAN MIXTURE MODELS

The mixture model framework assumes a vector of random variables $X_1,\ldots,X_n$ sampled from a finite mixture *m* of arbitrary distributions (components), where $m > 1$. Each distribution X_i has the probability density function in (1), where λ_j stands for the mixing weights, or the probability that a randomly selected observation comes from component *j*.

(19) <INSERT EX19 HERE>

Here, $\theta = (\lambda, \phi) = (\lambda_1,\ldots, \lambda_m, \phi_1,\ldots,\phi_m)$ denotes the parameter and the mixing weights λ_j must be positive and sum to 1. We will assume here that the parameters ϕ_j are drawn from a parametric family $\mathcal{F}$ of Gaussian distributions. The univariate Gaussian probability density function can be written as in (2).

(20) <INSERT EX20 HERE>

Model parameters can then be reduced to (3).

(21) <INSERT EX21 HERE>

REFERENCES

ABERCROMBIE, DAVID. 1967. *Elements of general phonetics*. Chicago: Aldine and Atherton.

AKINLABI, AKIN and ENO E. URUA. 2003. Foot structure in the Ibibio verb. *Journal of African Languages and Linguistic*s 24.119-160.

BARBOSA, PLÍNIO A. 2002. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. *Proceedings of the Speech Prosody*, Aix-en-Provence, France, 163-166.

BARBOSA, PLÍNIO A. 2003. *Beat extractor Praat script*. Retrieved from https://uk.groups.yahoo.com/neo/groups/praat-users/conversations/topics/921

BARBOSA, PLINIO A. 2007. From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Communication* 49.725-742

BARNES, JONATHAN A. 2002. *Positional neutralization: A phonologization approach to typological patterns*. Berkeley, CA: University of California, Berkeley, dissertation.

BARR, DALE J.; ROGER LEVY; CHRISTOPH SCHEEPERS; and HARRY J. TILY. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255-78. DOI: 10.1016/j.jml.2012.11.001.

BATES, DOUGLAS; MARTIN MAECHLER; BEN BOLKER; and STEVEN WALKER. 2017. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1-48.

Benaglia, Tatiana; Didier Chauveau; David R. Hunter; and Derek S. Young. 2009. mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software* 32(6).1-29.

BENNETT, RYAN. 2012. Foot-conditioned Phonotactics and Prosodic Constituency. Santa Cruz, CA: University of California dissertation.

BENNETT, RYAN. 2013. The uniqueness of metrical structure: Rhythmic phonotactics in Huaripano. *Phonology* 30.355-398.

BROWMAN, CATHERINE P. and LOUIS GOLDSTEIN. 1988. Some notes on syllable structure in Articulatory Phonology. *Phonetica* 45.140-155.

BROWMAN, CATHERINE P. and LOUIS GOLDSTEIN. 1992. Articulator Phonology: An overview. *Phonetica* 49.155-180.

BYRD, DANI and ELLIOT SALTZMAN. 1998. Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26.173-199.

CHO, TAEHONG and PATRICIA KEATING. 2001. Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics* 29.155-190.

CHO, TAEHONG, JAMES MCQUEEN, and ETHAN A. COX. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35(2).210-243.

CHUNG, YOUNAH and AMALIA ARVANITI. 2013. Speech rhythm in Korean: Experiments in speech cycling. *Proceedings of Meetings on Acoustics* 19.1-9.

CLEMENTS, G.N. and JOHN GOLDSMITH (eds.). 1984. *Autosegmental Studies in Bantu Tone*. Dordrecht, The Netherlands: Foris.

CUMMINS, FRED. 1997. *Rhythmic Coordination in English speech: An Experimental Study* Bloomington, Indiana: Indiana University dissertation.

CUMMINS, FRED and ROBERT F. PORT. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26.145-171.

DANIS, NIC; JONATHAN A. BARNES; and CATHERINE O'CONNOR. 2011. Downstep and contour formation in Medumba: A prosodic account. *Selected proceedings of the 42nd Annual Conference on African Linguistics* ed. by Michael R. Marlo, Nikki B. Adams, Christopher R. Green, Michelle Morrison, and Tristan M. Purvis, 23-31. Somerville, MA: Cascadilla Proceedings Project.

DARWIN, CHARLES. 1871. *The descent of man, and selection in relation to sex*. London: John Murray.

DOWNING, LAURA J. 2004. What African languages tell us about accent typology. *ZAS Papers in Linguistics* 37.101-136

DOWNING, LAURA J. 2010. Accent in African languages. *A Survey of Word Accentual Patterns in the Languages of the World*, ed. by Harry G. van der Hulst, Rob Goedemans, and Ellen van Zanten, 381-427. Berlin, Germany: Mouton de Gruyter.

FOUGERON, CECILE and PATRICIA KEATING. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101.3728-3739.

FRANICH, KATHRYN. 2017. The interaction of prominence, rhythm, and tone in Medʉmba. Chicago: University of Chicago dissertation.

FRANICH, KATHRYN. 2018. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* 67.21-33.

FRANICH, KATHRYN. 2019. Uncovering tonal and temporal correlates of phrasal prominence in Medʉmba. *Language and Speech*. https://doi.org/10.1177/0023830919887994

GAFOS, ADAMANTIOS. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20.269-337.

GAO, MAN. 2008. Mandarin tones: An Articulatory Phonology account. New haven,
CT: Yale University dissertation.

GHITZA, ODED. 2012. Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Human Neuroscience* 6.340.

GIRAUD, ANNE-LISE and DAVID POEPPEL. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience 18, 15(4).511–517. doi: 10.1038/nn.3063.

GOEDEMANS, ROB and HARRY VAN DER HULST. 2009. StressTyp: A database for word accetual patterns in the world's languages. *The use of databases in cross-linguistic research*, ed. by Martin Everaert and Simon Musgrave, 235-282. Mouton de Gruyter.

GOLDSMITH, JOHN. 1976. *Autosegmental Phonology*. Cambridge, MA: MIT dissertation.

GOLDSMITH, JOHN. 1994. A dynamic computational theory of accent systems. *Perspectives in Phonology*, ed. by Jennifer Cole and Charles Kisseberth, 1-28. Stanford, CA: Center for the Study of Language and Information.

GOLDSTEIN, LOUIS; HOSUNG NAM; ELLIOT SALTZMAN; and IOANA CHITORAN. 2009. Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in Phonetics and Speech Science*, ed. by Gunnar Fant, Hiroya Fujisak and Jiaxuan Shen, 239-249.

GORDON, MATTHEW. 1999. Syllable weight: Phonetics, phonology, and typology. Los Angeles: UCLA dissertation.

GORDON, MATTHEW. 2002. A factorial typology of quantity- insensitive stress. *Natural Language and Linguistic Theory* 20.491-552.

GORDON, MATTHEW and TIMO ROETTGER. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1).

GOSWAMI, USHA and VICTORIA LEONG. 2013. Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology* 4(1).67-92.

GREEN, CHRISTOPHER R. 2015. The foot domain in Bambara. *Language*, 91(1).e1-e26.

GUTHRIE, MALCOLM. 1948. *The Classification of the Bantu Languages*. Oxford University Press.

HAKEN, HERMANN. 1996. *Principles of Brain Functioning* (Vol. 67). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-79570-1

HALLE, MORRIS and JEAN-ROGER VERGNAUD. 1987. *An Essay on Stress*. Cambridge, MA: MIT Press.

HARRIS, JOHN. 1994. *English Sound Structure*. Oxford: Blackwell.

HAYES, BRUCE. 1983. A grid-based theory of English meter. *Linguistic Inquiry* 14.375-393.

HAYES, BRUCE. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15.33-74.

HAYES, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: University of Chicago Press.

HOMBERT, JEAN-MARIE., JOHN H. OHALA, and WILLIAM G. EWAN. 1979. Phonetic explanations for the development of tones. *Language* 55(1).37-58.

HONEYBONE, PATRICK. 2008. Lenition, weakening and consonantal strength: Tracing concepts through the history of phonology. *Lenition and Fortition*, ed. by Joaquim Brandão de Carvalho, Tobias Scheer, and Philippe Ségéral, 9-93. Berlin: Mouton de Gruyter.

HYMAN, LARRY M. 2003. Why describe African languages? *Keynote address at 4th World Congress on African Linguistics / 34th Annual Conference on African Linguistics,* Rutgers University.

HYMAN, LARRY M. 2008. Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46(2).309-350.

HYMAN, LARRY M. 2012. Do all languages have word accent? *UC Berkeley Phonology Lab Annual Report*, 32-54.

HYMAN, LARRY M. 2014. Do all languages have word accent? *Word Stress: Theoretical and Typological Issues*, ed. by Harry van der Hulst, 56-82. Cambridge University Press.

HYMAN, LARRY. 2015. Positional prominence vs. word accent: Is there a difference*? UC Berkeley Phonology Lab Annual Report*, 86-99.

HYMAN, LARRY M., NICHOLAS ROLLE, HANNAH SANDE, EMILY CLEM, PETER S.E. JENKS, FLORIANN LIONNET, JOHN MERRILL, and NICHOLAS BAIER. 2019. Niger-Congo linguistic features and typology. *The Cambridge Handbook of African Linguistics and A History of African Linguistics*, ed. by Ekkehard Wolff, 191-245. Cambridge University Press.

HYMAN, LARRY M. and MAURICE TADADJEU. 1976. Word domains and downstep in Bamileke-Dschang. *Phonology* 2(1).47-83.

IVRY, RICHARD and R. ELIOT HAZELTINE. 1995. Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance* 21(1).3-18.

HAKEN, HERMANN; JAMES A.S. KELSO; and H. BUNZ. 1985. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51.347-356.

ISKAROUS, KHALIL. 2017. The relation between the continuous and the discrete: A note on the first prinnciples of speech dynamics. Journal of Phonetics 64.8-20.

ISAKROUS, KHALIL and LOUIS GOLDSTEIN. 2018. The dynamics of prominence profiles: From local computation to global patterns. *Shaping Phonology*, ed. by Diane Brentari and Jackson Lee, 253-277. University of Chicago Press.

JAMES, GARETH M.; DANIELA WITTEN; TREVOR HASTIE; and ROBERT TIBSHIRANI. 2013. *An Introduction to Statistical Learning: with Applications in R.* New York: Springer.

JANDA, RICHARD and TERRELL MORGAN. 1988. El acentó dislocadó – pues cantadó – castellanó: on explaining stress-shift in song-texts from Spanish (and certain other Romance languages). *Advances in Romance Linguistics*, ed. by David Birdsong and Jean-Pierre Montreuil. Dordrecht: Foris, 151-170.

JEFFERS, ROBERT J. and ARNOLD M. ZWICKY. 1980. The evolution of clitics. *Current Issues in Linguistic Theory* 14, ed. by Rebecca Labrum and Susan C. Shepherd, 221-231. Amsterdam: John Benjamins.

KATSIKA, ARGYRO; JELENA KROVIKAPIĆ; CHRISTINE MOOSHAMMER; MARK TIEDE; and LOUIS GOLDSTEIN. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44.62-82.

KELSO, JAMES A.S. 1994. Elementary Coordination Dynamics. *Interlimb Coordination*, ed. by Stephan Swinnen, Jean Massion, Herbert Heuer, and P. Casaer, 301-318. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-679270-6.50020-6

KELSO, JAMES A.S. 1995. *Dynamic Patterns*. Cambridge, MA: MIT Press.

KEUPDJIO. HERMANN S. 2020. The syntax of A'-dependencies in Bamileke Medumba. Vancouver, BC: University of British Columbia dissertation.

KIPARSKY, PAUL. 1975. Stress, syntax, and meter. *Language* 51(3).576-616.

KO, EON-SUK. 2013. A metrical theory of Korean word prosody. *The Linguistic Review* 30(1).79-115. doi:10.1515/tlr-2013-0004.

KUZNETSOVA, ALEXANDRA; PER B. BROCKHOFF; and RUNE H.B. CHRISTENSEN. 2017. lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). https ://doi.org/10.18637 /jss.v082.i13

LAKATOS, PETER; JOACHIM GROSS; and GREGOR THUT. 2019. A new unifying account of the roles of neuronal entrainment. *Current Biology* 29(18).R890-R905. https://doi.org/10.1016/j.cub.2019.07.075

LEBEN, WILLIAM R. 1971. *Suprasegmental Phonology*. Cambridge, MA: MIT dissertation.

LEBEN, WILLIAM R. 1997. Tonal feet and the adaptation of English borrowings into Hausa. *Studies in African Linguistics* 25(2).139-154.

LEBEN, WILLIAM R. 2002. Tonal feet. *Proceedings of Typology of African Prosodic Systems* ed. by Ulrike Gut and Dafydd Gibbon, 27-40. Bielefeld Occasional Papers in Typology 1.

LERDAHL, FRED and RAY JACKENDOFF. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

LIBERMAN, MARK. 1975. The intonational system of English. Cambridge, MA: MIT dissertation.

LIBERMAN, MARK and ALAN PRINCE. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249-336.

MCLACHLAN, GEOFFREY J. and SUREN RATHNAYAKE. 2014. On the number of components in a Gaussian mixture model: Number of components in a Gaussian mixture model. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 4*(5).341-355. https://doi.org/10.1002/widm.1135

MALISZ, ZOFIA; MICHAEL O'DELL; TOMMI NIEMINEN; and PETRA WAGNER. 2016. Perspectives on speech timing: Coupled oscillator modeling of Polish and Finnish. *Phonetica* 73.229-255.

MARCUS, STEPHEN M. 1981. Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics 30*.247-256.

MORTON, JOHN; STEVE MARCUS; and CLIVE FRANKISH. 1976. Perceptual centers (p-centers). *Psychological Review* 83.405-408.

NANCARROW, CASSY (2010). What's that song about? Interaction of form and meaning in Lardil *Burdal* songs. *Australian Journal of Linguistics*, 30(1), 81-92.

NESPOR, MARINA and IRENE VOGEL. 1986. *Prosodic Phonology*. Dodrecht, The Netherlands: Foris.

O'DELL, MICHAEL and TOMMI NIEMINEN. 1999. Coupled oscillator model of speech rhythm. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, ed. by John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, and Ashlee C. Bailey, 1075-1078.

O'KEEFFE, ISABEL. 2010. Kaddikkaddik ka-wokdjanganj 'Kaddikkaddik Spoke': Language and Music of the Kun-barlang Kaddikkaddik Songs from Western Arnhem Land. *Australian Journal of Linguistics* 30(1).35-51.

OLSON, KENNETH and YOHANN MEYNADIER. 2015. On Medumba bilabial trills and vowels. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, ed. by The Scottish Consortium for ICPhS.

PAU, ANDREW. 2015. "Sous le Rythme De la Chanson": Rhythm, Text, and Diegetic Performance in Nineteenth-Century French Opera. *Music Theory Online*, 21(3).

PINKER, STEVEN. 1997. *How the mind works*. W.W. Norton & Company.

POEPPEL, DAVID. 2003. The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* 41.245-255

PORT, ROBERT F. 2003. Meter and speech. *Journal of Phonetics* 31.599-611.

PRINCE, ALAN. 1983 Relating to the grid. *Linguistic Inquiry* 14.19-100.

PRINCE, ALAN. 1990. Quantitative consequences of rhythmic organization. *Proceedings of the 26th Chicago Linguistic Society: Parasession on the syllable in phonetics and phonology,* 355-98.

PRINCE, ALAN. 1993. In defense of the number *i*: Anatomy of a linear dynamical model of linguistic generalizations. *Technical Report 1.* Piscataway, NJ: Rutgers University Center for Cognitive Science.

RIALLAND, ANNIE and MAMADOU BADJIMÉ. 1989. Réanalyse des tons du Bambara: Des tons du nom à l'organisation Générale du système. *Studies in African Linguistics* 20, 1-27.

SALTZMAN, ELLIOT; HOSUNG NAM; JELENA KRIVOKAPIĆ; and LOUIS GOLDSTEIN. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation.

Proceedings of the 4th International Conference on Speech Prosody, Campinas, 175-84. Online: http://aune.lpl.univaix.fr/~sprosig/sp2008/papers/3inv.pdf.

SCOTT, SOPHIE K. 1993. *P-centers in Speech.* London: University College London dissertation.

SELKIRK, ELISABETH O. 198. Epenthesis and degenerate syllables in Cairene Arabic. *Theoretical Issues in the Grammar of the Semitic Languages*, ed. by Hagit Borer and Yosef Aoun, 209-232. Cambridge, MA: MIT Press.

SELKIRK, ELISABETH O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

SELKIRK, ELISABETH O. 1986. On derived domains in sentence phonology. *Phonology* 3.371-405.

SHAW, JASON A.; ADAMANTIOS I. GAFOS; PHILIP HOOLE; and CHAKIR ZEROUAL. 2009. Syllabification in

Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology* 26.187-215. doi: 10.1017/s0952675709001754

SHAW, JASON A.; ADAMANTIOS I. GAFOS; PHILIP HOOLE; and CHAKIR ZEROUAL. 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology* 28.455-490. doi: 10.1017/s0952675711000224

SLEVC, L. ROBERT, and BROOKE M. OKADA. 2015. Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review* 22.637-652. https://doi.org/10.3758/s13423-014-0712-4

SOLTÉSZ, FRUZSINA; DENES SZŰCS; VICTORIA LEONG; SONIA WHITE; and USHA GOSWAMI. 2013. Differential entrainment of neuroelectric delta oscillations in developmental dyslexia. *PLoS ONE* 8(10).e76608. https://doi.org/10.1371/journal.pone.0076608.

TAJIMA, KEIICHI. 1998. *Speech rhythm in English and Japanese: Experiments in speech cycling*. Bloomington: Indiana University dissertation.

TAJIMA, KEIICHI. and Port, Robert F. 2003. Speech Rhythm in English and Japanese. *Papers in Laboratory Phonology VI: Phonetic Interpretation*, ed. by John Local, Richard Ogden, and Rosalind Temple, 322-339. Cambridge, UK: Cambridge University Press.

TARLINSKAJA, MARINA. 1993. *Strict Stress-Meter in English Poetry Compared with German and Russian*. Calgary: University of Calgary Press.

TEMPERLEY, DAVID. 2000. Meter and Grouping in African Music: A View from Music Theory. *Ethnomusicology 44*(1).65. https://doi.org/10.2307/852655

TEMPERLEY, NICHOLAS and TEMPERLEY, DAVID. 2012. Stress-meter alignment in French vocal music. *Journal of the Acoustical Society of America* 134(1).520-7.

TIERNEY, ADAM T. and NINA KRAUS. 2013. THE ABILITY TO TAP TO A BEAT RELATES TO COGNITIVE, LINGUISTIC, AND PERCEPTUAL SKILLS. BRAIN AND LANGUAGE 124(3).225-231.

TILSEN, SAM. 2009. Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33.839-879.

TILSEN, SAM. 2011. Metrical regularity facilitates speech planning and production. *Laboratory Phonology* 2(1).185-218.

TILSEN, SAM. 2019. Space and time in models of speech rhythm. *Annals of The New York Academy of Sciences*, 1453.47-66 doi:10.1111/nyas.14102.

VAN DER HULST, HARRY; ROB GOEDEMANS; and KEREN RICE. 2017. Areal aspects of Stress. *Handbook of Areal Linguistics*, ed. by Raymond Hickey, 161-203. Cambridge: Cambridge University Press.

VAN GELDER, TIMOTHY and ROBERT PORT. 1995. It's about time: Overview of the dynamical approach to cognition. *Mind as Motion: Explorations in the Dynamics of Cognition*, ed. by Robert Port and Timothy van Gelder, 1-43. Bradford Books/MIT Press.

VATIKIOTIS-BATESON, ERIC. 1988. Linguistic structure and articulatory dynamics. Bloomington: Indiana University dissertation.

VATIKIOTIS-BATESON, ERIC and JAMES A.S. KELSO. 1993. Rhythm type and articulatory dynamics in

English, French and Japanese. *Journal of Phonetics* 21 231-265.

VOORHOEVE, JAN. 1965. The Structure of the Morpheme in Bamileke (Bangangte dialect). *Lingua* 13.319-334.

VOORHOEVE, JAN. 1967. Personal pronouns in Bamileke. *Lingua* 17(4).421-430.

VOORHOEVE, JAN. 1971. Tonology of the Bamileke noun. *Journal of African Languages* 2.44-53.

VOORHOEVE, JAN. 1976. *Contes Bamileke*.Tervuren, Belgium: Museé Royal de L'Afrique Centrale.

WEIDMAN, SCOTT and SHARON ROSE. 2006. A foot-based reanalysis of edge-in tonal phenomena in Bambara. *Proceedings of the 25th West Coast Conference on Formal Linguistics*, ed.

by Donald Baumer, David Montero, and Michael Scanlon, 426-434. Somerville, MA: Cascadilla Press.

WILSON, MARGARET and THOMAS P. WILSON. 2005. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review* 12(6).957-968. https://doi.org/10.3758/BF03206432

WOODRUFF CARR, KALI; TRAVIS WHITE-SCHWOCH; ADAM T. TIERNEY; DANA L. STRAIT; and NINA KRAUS. 2014. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *PNAS* 111(40).14559-14564.

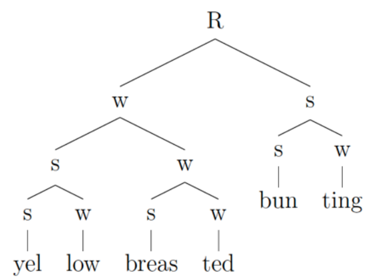
ZAWAYDEH, BUSHRA A.; KEIICHI TAJIMA; and MAFUYU KITAHARA. 2002. Discovering Arabic rhythm through a speech cycling task. *Perspectives in Arabic Linguistics, XIII–XIV*. Arabic Linguistic Society.

ZHANG, JIE. 2002. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. Los Angeles: UCLA dissertation.

[1] For a comprehensive introduction to dynamical systems first principles as they can be applied to phonological theory, see Iskarous 2017. For background on how such principles can be applied specifically within the domain of metrical structure, see Iskarous & Goldstein 2018.
[2] Generalizations in the present paper are based on Medʉmba as it is spoken in the town of Bangangté and in the village of Bangoulap, situated south of Bangangté, in Ndé division, West Region, Cameroon.
[3] This list includes most of the same consonants and vowels described by Voorhoeve (1965, 1976), with some differences. For example, while Voorhoeve argued for a contrast between /k/ and /g/ in the language, I find no evidence that these are distinct phonemes in the Bangangté/Bangoulap dialects. Furthermore, root-internally, Voorhoeve's /mf/ is always produced as /mv/ in the dialects examined here; I therefore transcribe them as such. Aspirated consonants are contrastive in loanwords only. Finally, the vowel inventory is updated in places to reflect more recent acoustic analyses of the Bangangté dialect by Olson & Meynadier (2015). Vowels analyzed as diphthongs by Voorhoeve (1965) are also excluded from the present discussion.
[4] Note that in stem-final position, plosive consonants become devoiced.
[5] Danis also uses this difference in structure to account for the presence of downstep between words in items such as (2b), and the lack of such a downstep in (2a)
[6] This particular process of spirantization is not observed in native Medʉmba words: stem-internally, as well as in cliticized constructions, all coronal stops are realized [l] when they occur intervocalically (e.g. ᵐvɛ́t=í > mvɛlí 'her brother'), and /ʧ/ only ever occurs in stem-initial position. The lack of a separate process turning /t/ to [s] (vs. the observed process changing /d/ to [l]) may be attributable the fact that the /d/ vs. /t/ distinction itself is fairly marginal in Medʉmba native words: the distinction is fully neutralized in all positions except for stem-initially, and even in stem-initial position, very few minimal pairs exist to evidence this distinction.
[7] Illicit syllable structures, such those ending in an affricate (e.g. the English loan 'church'), are remedied through the insertion of a final epenthetic vowel [i].
[8] Note that inalienable possession (such as would apply to nouns like *mbáɣə́* 'fire' and *mbálə́* 'hill') is usually indicated with a pronoun which is pre-posed with respect to the noun it modifies, as opposed to postposed. Though these constructions are somewhat pragmatically odd for speakers, speakers nonetheless have clear intuitions about how the key lenition processes are to apply in such contexts.
[9] Matching the content of consonants and vowels across conditions was considerably more difficult in Experiment 2 than in Experiment 1 since the prefix in the PreSI condition limited the possible onset consonants usable for comparison. As there are relatively few CV noun roots initiated with a nasal, some forms (notably tsə̀=ʧú 'your in-laws' and tɔ̀ŋ=ú 'your ear') were used in the SIEnc condition which deviated from their counterparts in the PreSI condition. These choices were made with care, however. Since the location of perceptual centers (see Appendix B) is most heavily influenced by syllable onset duration (Marcus 1981, see also Franich 2018), *ts* and *t* were chosen as comparison onsets due to the similarity of their durations with that of *n*. Of note, too, is the fact that participants intuited that the final nasal of *tɔŋ* 'ear' was syllabified with the vowel of the pronominal enclitic, giving us a comparison between *nú* and *ŋú* for the final syllable in Set C.
[10] When a factor is sum-coded, the mean of the dependent variable for a given factor level is compared to the overall mean of the dependent variable over all levels.

[11] Mixtools draws initial mixing proportion values randomly from a uniform Dirichlet distribution. Standard deviations are set as the reciprocal of the square root of a vector of random exponential-distribution values whose means were determined according to a binning method.
[12] Generally speaking, the number of folds *k* chosen for cross-validation should be determined so that the train and test groups for the data samples is sufficiently large to be statistically representative of the dataset as a whole, and so as to avoid model overfitting (James et al. 2013). Often, the value is chosen so that the data can be evenly split across folds. Since the dataset contained (a maximum of) 6 repetitions of for each condition per participant, a 3-fold approach was deemed optimal so that each fold could in principle have an equal number of 2 observations per condition per participant.
[13] While researchers largely agree that both high and low tones must be specified in the tonal inventory of Medʉmba (Danis 2011; Franich 2017; Hyman 2003; Voorhoeve 1971, 1976), Keupdjio argues that lexical categories are limited to a low vs. unmarked distinction, in which the unmarked tone surfaces with a default high tone. We maintain that a H vs. L distinction is necessary even for lexical categories given that both high and low tones are both highly phonologically 'active' in the language (c.f. Myers 1998), with both tones undergoing processes such as tone spread and downstep.
[14] The reason for this pattern is likely to do with the commonly-observed lowering effect of voiced consonants on f0 (Hombert et al. 1979); nasals are the only permissible voiced segment to occur in coda position in Medʉmba.
[15] Note that the epenthesis which results on the examples in (14) cannot be attributable to the illicitness of CVCVC as a foot shape, as we do find other loanwords of this shape, e.g. [slípɛ̀t] 'slipper', and [tóvɛ̀t] 'towel'. Native Medʉmba roots of the shape CVCVN can also be found, such as *bə̀lɔ̀ŋ* 'potato'; low tone spreading in such forms seems to indicate they also form a single foot.
[16] Superscript nasals in examples in (15) indicate that the nasal is parsed as part of a 'prenasalized' ^{n}C onset, rather than as a coda-onset sequence.
[17] Note that the word-initial nasal for the verb in both prosodic conditions in Experiment 1 is a prefix. Franich (2018) has shown that these prefixes do not influence p-center location of a CV sequence that they prefix to, suggesting they form their own, separate timing unit. This could indicate the nasal is itself syllabic, though, as noted in Franich (2018), Medʉmba speakers do not have the intuition that they are. Another possibility is that the prefix is 'extra-syllabic' (Goldstein et al. 2009). Given that both prosodic conditions contain the same nasal in the same position relative to the syllable, its presence—whatever its prosodic status—does not pose a problem for the analysis presented here.

(a)



(b)

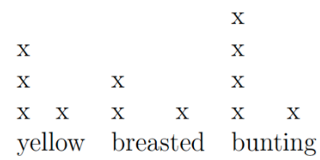


FIGURE 1. Metrical tree (a) and grid (b) structures similar to those proposed by Liberman 1975 and Liberman and Prince 1977. Prominence corresponding to strong (*s*) and weak (*w*) nodes at each level in the hierarchy in (1a) are realized with greater numbers of gridmarks in (1b).

Consonants				
Stem-initial (48)	Stem-medial (7)	Stem-final (7)	Prefix (1)	Suffix (1)
^{m}ʙ, ^{m}b, ^{m}b^{w}, ^{n}t, ^{n}d, ^{ɲ}c, ^{ɲ}c^{w}, ^{ɲ}ɟ, ^{ɲ}ɟ^{w}, ^{ŋ}k, ^{ŋ}g, ^{m}v, ^{n}z, ^{n}z^{w}, ^{n}ts, ^{n}dz, ^{n}ʧ, ^{n}ʤ, ɲ^{w}, ŋ^{w}, ʙ, b, b^{w}, t, d, t^{h}, c, c^{w}, k, k^{h}, k^{w}, m, n, ɲ, ŋ, f, v, s, s^{w}, z, ʒ, ɣ, ts, dz, ʧ, ʤ, j, l	b, ʔ, l, ɣ, m, n, ŋ	p, t, k, ʔ, m, n, ŋ	n	d
Vowels				
Stem-initial (11)	Stem-medial	Stem-final (1)	Prefixes (1)	Suffixes (1)
i, u, ɪ, ʉ, e, o, ɛ, ɔ, ə, a, ɑ	N/A	ə	ə	ə

TABLE 1. Consonant and vowel distributions by stem position and affix type

Possessive Pronouns		Object Pronouns	
ɑm	1.sg	ɑm	1.sg
u	2.sg	u	2.sg
i	3.sg	i	3.sg
jak	1.pl	jak	1.pl exclusive
zin	2.pl	bən	1.pl inclusive
jup	3.pl	zin	2.pl
		jup	3.pl

TABLE 2. Sample of possessive and object pronouns taken from Voorhoeve 1967, Groups III and V (transcribed based on IPA conventions)

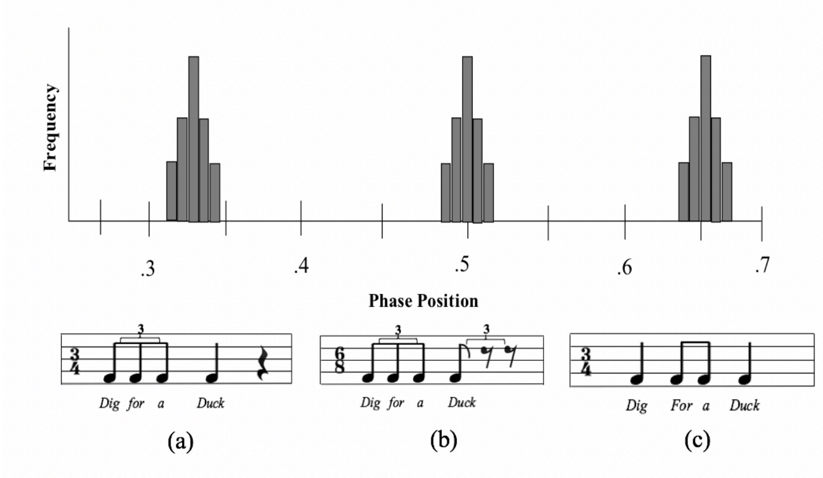


FIGURE 2. Typical alignment patterns for the word *duck* in speech cycling for the sentence *Dig for a duck*, with corresponding musical rhythmic notation. Histograms in the top portion of the figure represent clustering of syllable repetitions around phase positions of 1/3 or .33 (a), 1/2 or .50 (b), and 2/3 or .67 (c) of the repetition cycle.[a]

[a] Cummins & Port (1998) show that, while musical experience has the potential to influence the stability of different alignment strategies within the speech cycling task at different rates, participants with and without musical training show similar patterns of preference for simple harmonic phase angles in the task (in particular, gravitation to the 1/2 and 1/3 phase angles). Another question concerns whether cross-cultural musical preferences might also influence participants' behavior in the task, an issue which is addressed in §5. Given the extraordinary commonness of duple and triple meters such as those shown in Figure 2 across many musical traditions, including many West and Central African musical traditions (Temperley 2000), we have good reason to think that these alignment strategies will be preferred by Medʉmba speakers, similarly to US English speakers.

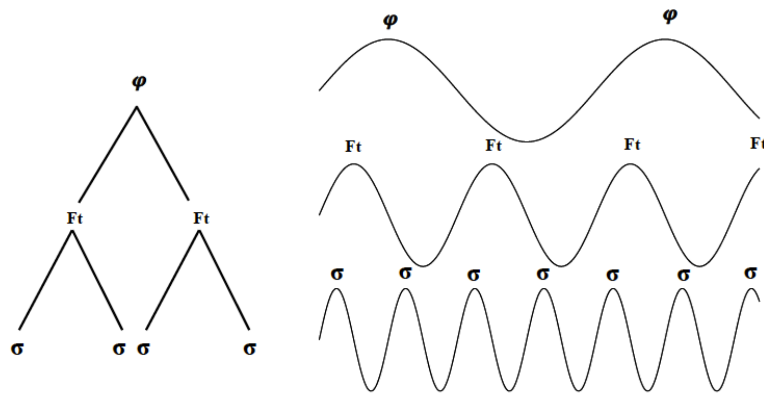


FIGURE 3. The prosodic hierarchy (left) modeled as coupled oscillators (right) with 2:1 syllable-to-foot and foot-to-phrase frequency locking

Set	Form	IPA Transcription	Gloss	Translation
A	SISF	mέn ʧák **mbí**bə́	child TNS wait	'The child waited.'
A	SISI	mέn ʧák **mʙí bá**	child TNS lose bicycle	'The child lost the bicycle.'
B	SISF	bʉ́n ʧák **nʃú**mə́	children TNS go out	'The children went out.'
B	SISI	bʉ́n ʧák **ntúm mέn**	children TNS accuse child	'The children accused the child.'
C	SISF	sáŋ ʧák **mfʉ́**lə́	bird TNS fly	'The bird flew.'
C	SISI	sáŋ ʧák **njʉ́ lá**	bird TNS eat pineapple	'The bird ate the pineapple.'

TABLE 3. Stimuli for Experiment 1. Syllables hypothesized prominent are in bold.

Set	Form	IPA Transcription	Gloss	Translation
A	PreSI	mɛ́n lɛ̆n nə̀-**tsí**	child know INF-choose	'The child knows how to choose.'
A	SIEnc	mɛ́n lɛ̆n **nù**=tsə́	child know problem=their	'The child knows their problems.'
B	PreSI	sáŋ lɛ̆n nə̀-**ʧú**	bird know INF-enter	'The bird knows how to enter.'
B	SIEnc	mɛ́n lɛ̆n **tsə̀**=ʧú	child know in.law=your	'The child knows your in-laws.'
C	PreSI	sáŋ lɛ̆n nə̀-**nú**	bird know INF-drink	'The bird knows how to drink.'
C	SIEnc	sáŋ lɛ̆n **tɔ̀ŋ**=ú	bird know ear=your	'The bird knows your ear.'

TABLE 4. Stimuli for Experiment 2. Syllables hypothesized prominent are in bold. Clitic boundaries indicated with '=' and affix boundaries with '-'

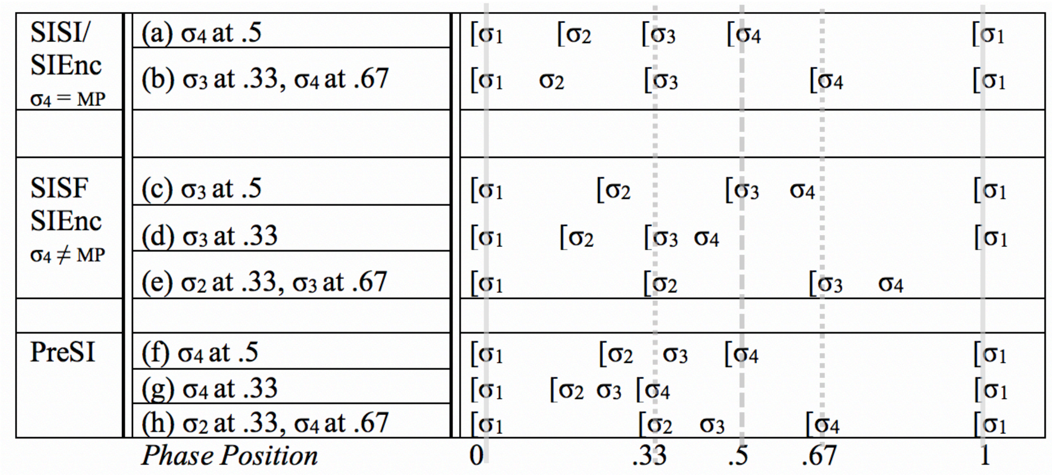


FIGURE 4. Possible syllable alignment strategies based on repetition cycle subdivision strategies in the speech cycling task. Dotted and dashed lines represent SHP positions for subdivision in 2 (dashed) and subdivision in 3 (dotted) strategies. MP stands for 'metrically prominent.'

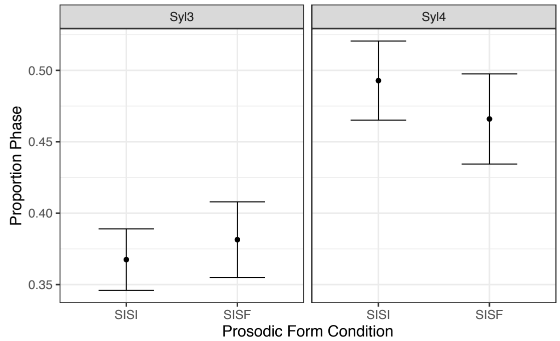


FIGURE 5. Difference in alignment between conditions for Experiment 1, Syllables 3 and 4. Error bars represent 95% confidence intervals.

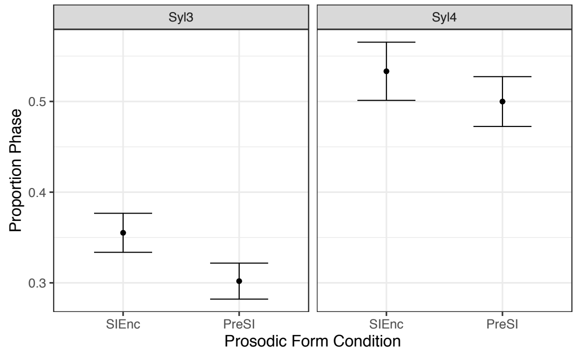


FIGURE 6. Difference in alignment between conditions for Experiment 2, Syllables 3 and 4. Error bars represent 95% confidence intervals.

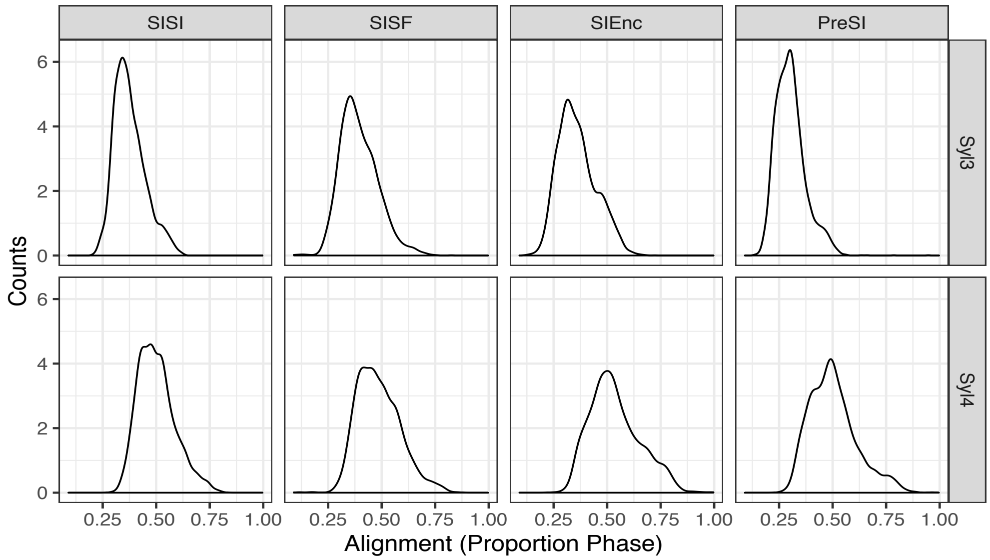


FIGURE 7. Histograms of Phase Alignments for Syllables 3 and 4, Experiments 1 and 2

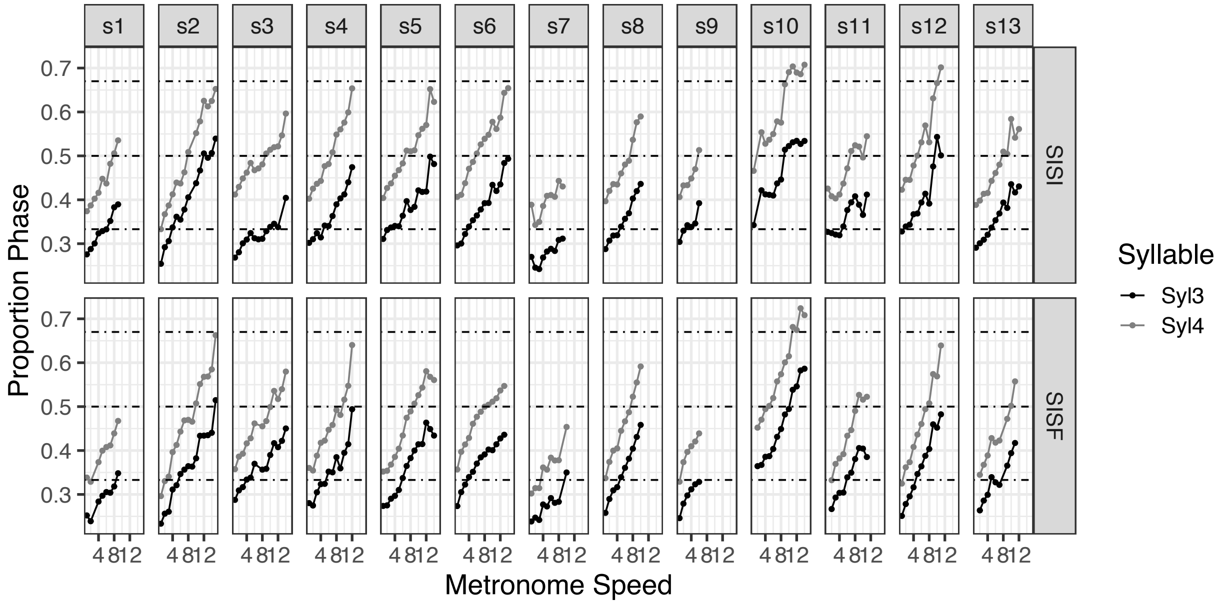


FIGURE 8. Individual variation in syllable alignment across prosodic conditions, Experiment 1

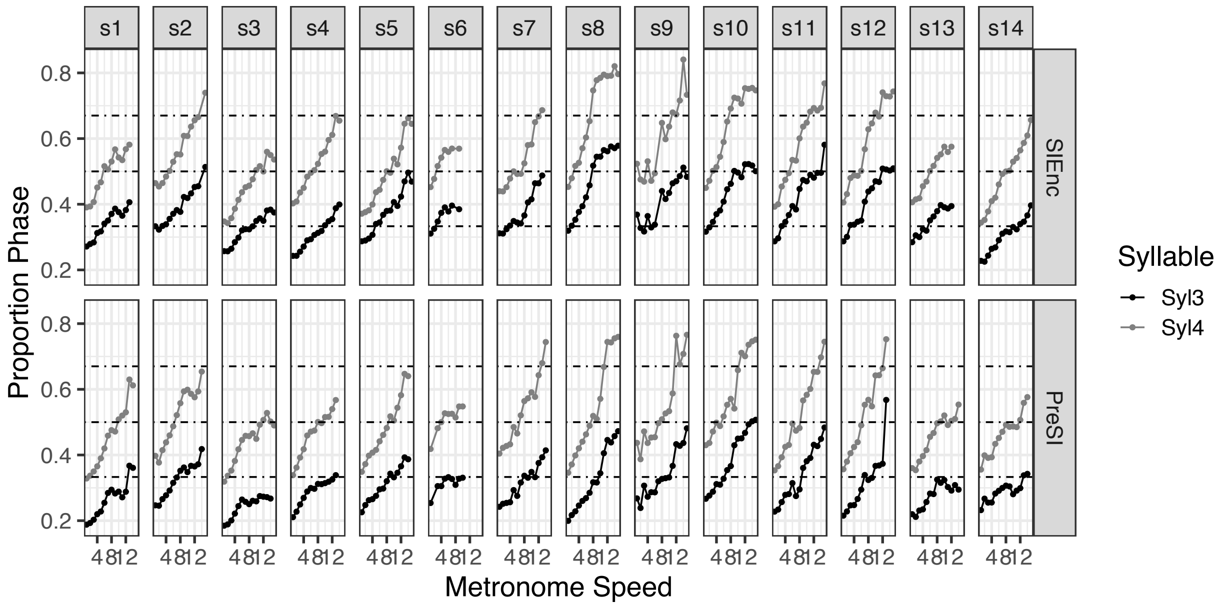


FIGURE 9. Individual variation in syllable alignment across prosodic conditions, Experiment 2

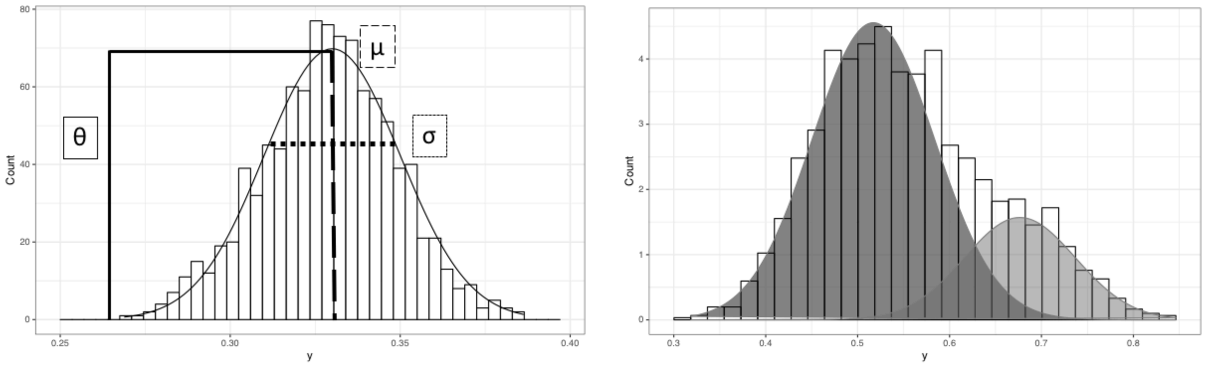
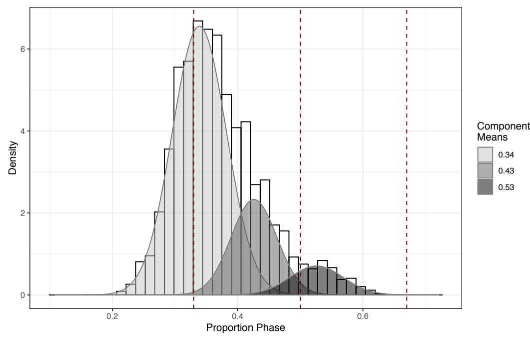


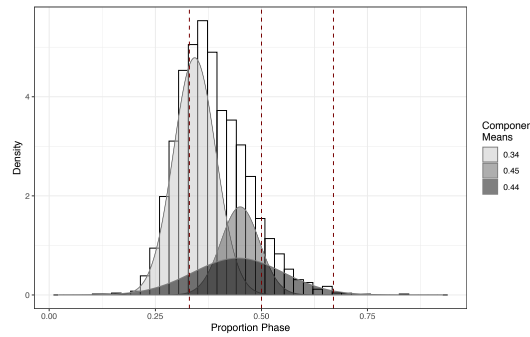
FIGURE 10. Single Gaussian distribution (left) with parameters labeled and a two-component Gaussian mixture (right)

Condition	Model	mean μ			Standard deviation σ			Mixing proportion λ			Log likelihood
Syl. 3, SISI	MU	.34	.43	.53	.04	.04	.04	.71	.21	.08	2942.843
	MS	.33	.50	.67	.05	.06	.02	.82	.17	.01	2801.609
Syl. 3, SISF	MU	.34	.45	.44	.05	.05	.11	.60	.22	.18	2573.634
	MS	.33	.50	.67	.06	.07	.13	.72	.27	.01	2438.445
Syl. 4, SISI	MU	.41	.49	.61	.03	.06	.08	.14	.68	.18	2513.423
	MS	.33	.50	.67	.05	.07	.06	.01	.95	.04	2438.120
Syl. 4, SISF	MU	.41	.52	.54	.05	.05	.11	.49	.27	.24	2280.904
	MS	.33	.50	.67	.04	.09	.10	.08	.91	.01	2108.759

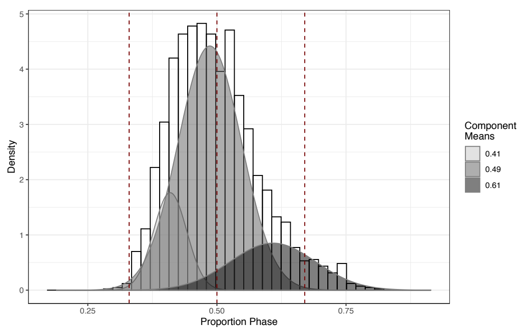
TABLE 4. Final Model Parameters, Experiment 1



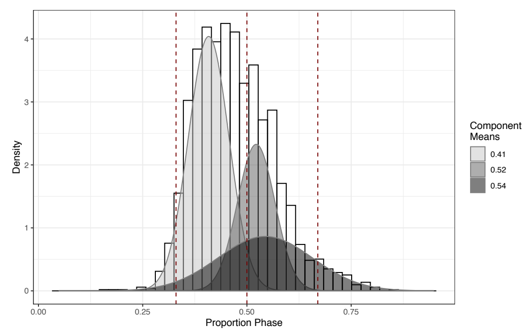
(a) Syllable 3, SISI



(b) Syllable 3, SISF



(c) Syllable 4, SISI

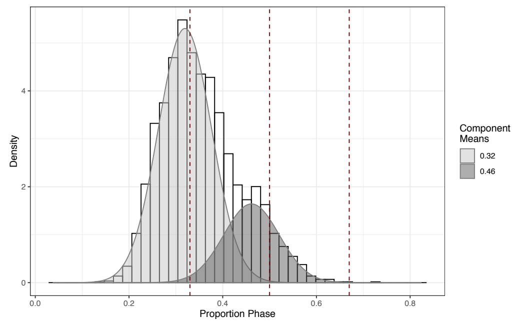


(d) Syllable 4, SISF

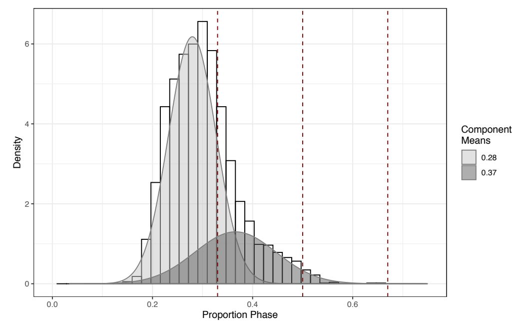
FIGURE 11. density curves for model-theoretic components plotted against histograms of raw data, Experiment 1. Dashed red lines mark locations of SHPs at .33, .5, and .67 phase positions.

Condition	Model	mean μ			Standard deviation σ			Mixing proportion λ			Log likelihood
Syl. 3, SIEnc	MU MS	.32 .33	.46 .50		.06 .05	.06 .08		.76 .54	.24 .46		3274.70 3253.26
Syl. 3, PreSI	MU MS	.28 .33	.37 .50		.05 .07	.08 .30		.74 .99	.26 .01		3814.45 3402.35
Syl. 4, SIEnc	MU MS	.38 .33	.49 .50	.62 .67	.03 .02	.06 .08	.11 .10	.09 .01	.49 .78	.42 .21	2420.52 2368.60
Syl. 4, PreSI	MU MS	.38 .33	.48 .50	.67 .67	.04 .02	.07 .08	.09 .11	.16 .05	.67 .85	.17 .10	2540.92 2417.92

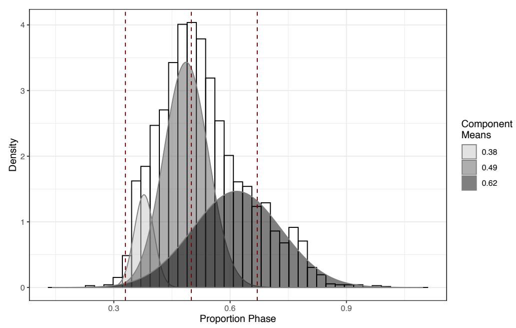
TABLE 5. Final Model Parameters, Experiment 2



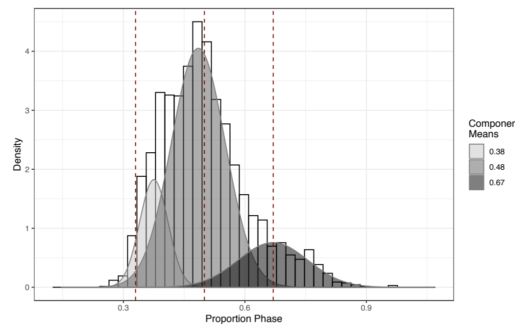
(a) Syllable 3, SIEnc



(b) Syllable 3, PreSI



(c) Syllable 4, SIEnc



(d) Syllable 4, PreSI

FIGURE 12. density curves for model-theoretic components plotted against histograms of raw data, Experiment 2. Dashed red lines mark locations of SHPs at .33, .5, and .67 phase positions.