METRICAL PROMINENCE ASYMMETRIES IN MEDUMBÁ, A GRASSFIELDS BANTU LANGUAGE

KATHRYN FRANICH

University of Delaware

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 article explores metrical prominence asymmetries in Medumba, 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 the relative timing of syllables hypothesized to be metricaly prominent and metricaly weak. Previous research using the task has shown that metricaly prominent syllables are attracted to certain relative positions within a repetition cycle. Results of two experiments show that foot heads in Medumba also show this behavior, supporting their status as metricaly prominent. These results suggest that true metrical prominence asymmetries exist in a broader range of languages than previously thought, and that relative timing serves as an important unifying property of metrical structure crosslinguistically.*

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

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 metricaly driven prominence in a language. African tone languages are among 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 that 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. Since the existing cues to ‘prominence’ could potentially be reduced to morphologically conditioned phonological processes or word-edge effects—neither of which are neces-

* 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 Medumba. Thanks to Andries Coetzee, Khalil Iskarous, and two referees for very clear and helpful feedback on the manuscript. All errors are of course my own. Finally, many thanks to the many Medumba speakers who welcomed me into their community and generously contributed their time to participate in the research. Special thanks to Ange Bergson Lendja and 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 article.

Printed with the permission of Kathryn Franich. © 2021.
sarily ‘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 of 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, 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 & Prince 1977, the position of syllable stress within words is governed by constituency relations among syllables within metrical trees or grids (Figure 1).

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, 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 that utilize ‘canonical’ stress cues and those that may mark metrical alternations in other ways.

In this article, I argue for an alternative view of metrical prominence asymmetries, which I 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 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 that 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, that govern the behavior of the linguistic system in time and in response to changing system parameters, such as speech rate (Barbosa 2002, 2007, Goldsmith 1994, O’Dell & Nieminen 1999, Prince 1993, Saltzman et al. 2008, Vatikiotis-Bateson 1988, Vatikiotis-Bateson & Kelso 1993). I 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; nonprominent 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 (Dell & Halle 2009, Janda & Morgan 1988, Lerdahl & Jackendoff 1983, Nancarrow 2010, Pau 2015, Tarlinskaja 1993, Temperley & Temperley 2012). As I 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 article is on Medumba, a Grassfields Bantu language spoken in Cameroon. Medumba is one of many African tone languages that shows evidence consistent with metrical prominence patterns, but that does not meet all of the typical phonetic criteria for a metrically based stress system. In the case of Medumba, stem-initial syllables appear to show evidence of greater prominence than noninitial and nonstem 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. I first provide an introduction to the Medumba language and present evidence from positional asymmetries in segment contrasts that provides preliminary motivation for metrical prominence asymmetries (§2). I then introduce the speech cycling paradigm in §3 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, noninitial, and nonstem syllables in the speech cycling task, demonstrating that stem-initial syllables in Medumba do, indeed, show behavior parallel 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 Medumba, suggesting that they possess elements of metrical strength and weakness, possibly the result of a prosodic change in progress.

1 For a comprehensive introduction to the first principles of dynamical systems 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. Background on the Medumbá Language. Medumbá is one of several Bamileke languages within the Eastern Grassfields subgroup of Grassfields Bantu. While arguably descended from Proto-Bantu (Hyman 2003, Voorhoeve 1971), Medumbá 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 article 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 a greater number of consonantal, vocalic, and tonal contrasts than do noninitial and nonstem syllables (see Hyman et al. 2019 and references therein). In Medumbá, forty-eight consonants and eleven vowels can appear in stem-initial position (or in monosyllabic stems); see Table 1. In noninitial position, however, only seven consonants and one vowel ([ǝ]) are found. Medumbá displays few segmental affixes, but those that do exist exclusively contain the vowel [ǝ]. Additional positional restrictions on tone can be found in Appendix A.

As is 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 that targets medial consonants within this maximally disyllabic domain, a process that 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 noncompound native stems in Medumbá 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 oc-

---

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

<table>
<thead>
<tr>
<th>CONSONANTS</th>
<th>STEM-INITIAL (48)</th>
<th>STEM-MEDIAL (7)</th>
<th>STEM-FINAL (7)</th>
<th>PREFIX (1)</th>
<th>SUFFIX (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m̃n, m̃b, m̃bw, ɲ̃t, ɲ̃q, ɲ̃c, ɲ̃cw, ɲ̃f, ɲ̃v, ɲ̃z, ɲ̃z̃w, ɲ̃ts, ɲ̃dz, ɲ̃g, ɲ̃q̃d, ɲ̃p̃w, ɲ̃n, b, b̃w, t, d, t̃, c, c̃w, k, k̃b, k̃w, m, n, p, ɲ, f, v, s, s̃w, z, ɲɈ, ɲɈw, ɲk, ɲɡ, mv, nz, nzw, nts, ndz, nz, ʧ, nʤ, ɲw, ɲw, ɲw, ɲw, b, b, bw, t, d, th, c, cw, k, kh, kw, m, n, ɲ, ɲ, ɲ, f, v, s, s, ʃ, ʒ, ɣ, ʧ, ʤ, j, l, N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOWELS</td>
<td>STEM-INITIAL (11)</td>
<td>STEM-MEDIAL</td>
<td>STEM-FINAL (1)</td>
<td>PREFIX (1)</td>
<td>SUFFIX (1)</td>
</tr>
<tr>
<td>i, u, j, u, c, o, ɛ, ɔ, ǝ, a, a</td>
<td>N/A</td>
<td>ǝ</td>
<td>ǝ</td>
<td>ǝ</td>
<td>ǝ</td>
</tr>
</tbody>
</table>

---

2 Generalizations in the present article are based on Medumbá 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 /ɡ/ in the language, I find no evidence that these are distinct phonemes in the Bangangté/Bangoulap dialects. Furthermore, root-internally, Voorhoeve’s /mv/ is always produced as /mŋ/ 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 and 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.
curring in stem-medial position (Danis 2011, Danis et al. 2011). Words such as "mbwáɣé‘fire’ and mbáɭ ‘hill’ are realized as disyllabic in isolation or phrase-finally (1a–b), and as monosyllabic phrase-initially or phrase-externally (1c–d). As seen in 1a,c, the velar stop /k/ is realized as [ɣ] word-externally 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. mbwáɣé sáɣé ‘fire’ sákó ‘sauce’ /mbwáké/ /mbákó/

b. mbálé mbálé ‘hill’ /mbálé/ /mbálé/

mvéɫó ‘brother’ /mvéɫó/ /mvéɫó/

c. mbwák ɭÁnà (*mbáɣé ɭÁnà) ‘Ana’s fire’ sák ɭÁnà (*sáɣé ɭÁnà) ‘Ana’s sauce’

d. mbáɭ ɭÁnà (*mbáɭ ɭÁnà) ‘Ana’s hill’ mbáɭ ɭÁnà ‘Ana’s brother’

These processes (collectively referred to henceforth as ‘lenition’) occur word-externally 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. ω(mḅwáɣé)ω (mḅwáɣé)ω (mḅwáɣé)ω (mḅwáɣé)ω

b. ω(mḅáɭ)ω (mḅáɭ)ω (mḅáɭ)ω (mḅáɭ)ω

As can be seen, lenition, per Danis’s analysis, occurs intervocally 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-externally 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. káɣé (káɣé)ω (káɣé)ω (káɣé)ω (káɣé)ω ‘release’ téló (téló)ω (téló)ω ‘meet’

b. káɭÁnà (káɭÁnà)ω (káɭÁnà)ω (káɭÁnà)ω (káɭÁnà)ω ‘release Ana’ tét ɭÁnà (tét)ω (tét)ω (tét)ω (tét)ω ‘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. ṇd-káɣé (*ṇd-ɣáɣé) (ṇd-káɣé)ω (ṇd-káɣé)ω INF-release ‘to release’

b. ṇd-téló (*ṇd-léló) (ṇd-téló)ω (ṇd-téló)ω INF-meet ‘to meet’

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.
Thus, the emerging generalization appears to be that lenition can target consonants that are stem-internal, but not those that are stem-initial. Due to the fact that native stems in Medumba 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 Medumba nouns and verbs


To observe lenition behavior in stems longer than two syllables, we can look to Medumba 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, which turns /t/ and /ʧ/ to [s] when preceding a high front vowel, does occur in loanwords.6 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) Medumba loanword          English source IPA       English translation
a. tí (*sí)         [tí]       ‘tea’
    tóló (*siló)     [‘tejlo]    ‘tailor’
    tfójá (*sitjá)   [‘tjíjá]    ‘teacher’
b. sásidé (*sátidé) [‘sætđ,đej] ‘Saturday’
    kísíimi (*kitjími) [‘kítjên] ‘kitchen’
    tfósi7 (*fоjí)   [fоjí]      ‘church’
c. kúbáti (*kúbási) [‘kábåd]    ‘cupboard’
    libáti (*libási) [‘libåti]    ‘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 that is also characteristic of metrically driven stress systems. A better generalization appears to be that spirantization is driven by foot-basedmetrical prominence asymmetries, as was 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


6 This particular process of spirantization is not observed in native Medumba words: stem-internally, as well as in cliticized constructions, all coronal stops are realized as [l] when they occur intervocally (e.g. mvé-ti > mveli ‘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 to the fact that the /d/ vs. /t/ distinction itself is fairly marginal in Medumba 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 as those ending in an affricate (e.g. the English loan ‘church’), are remedied through the insertion of a final epenthetic vowel [i].
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, 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 that cannot be reduced to such positional effects.

2.2. Lenition within pronominal constructions. There is an additional source of data that is of interest in the present work due to its implications for our understanding of foot-based patterns in Medumba. As described by Danis (2011; see also Danis et al. 2011), pronominal enclitics appear to trigger lenition processes that parallel those found stem-internally in some dialects of Medumba, including the Bangangté and Bangoulap dialects (8).

(8) Lenition in pronominal constructions

a. mbwάɣ=ám mbwάɣ=ú
   ‘my fire’ ‘your fire’
   sáɣ=ám sáɣ=ú
   ‘my sauce’ ‘your sauce’

b. mbá=ám mbá=ú
   ‘my hill’ ‘your hill’

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(mbwάɣ=ám)Ft
   Ft(mbwάɣ=ú)Ft
b. Ft(mbá=ám)Ft
   Ft(mbá=ú)Ft

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

<table>
<thead>
<tr>
<th>POSSESSIVE PRONOUNS</th>
<th>OBJECT PRONOUNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>1SG</td>
</tr>
<tr>
<td>am</td>
<td>am</td>
</tr>
<tr>
<td>2SG</td>
<td>2SG</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
</tr>
<tr>
<td>3SG</td>
<td>3SG</td>
</tr>
<tr>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>1PL</td>
<td>1PL EXCLUSIVE</td>
</tr>
<tr>
<td>jak</td>
<td>jak</td>
</tr>
<tr>
<td>2PL</td>
<td>2PL INCLUSIVE</td>
</tr>
<tr>
<td>zin</td>
<td>ban</td>
</tr>
<tr>
<td>3PL</td>
<td>3PL</td>
</tr>
<tr>
<td>jup</td>
<td>jup</td>
</tr>
</tbody>
</table>

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

8 Note that alienable possession (such as would apply to nouns like mbáɣ ‘fire’ and mbál ‘hill’) is usually indicated with a pronoun that is preposed with respect to the noun it modifies, as opposed to postposed. Though the constructions in 8 are somewhat pragmatically odd for speakers, speakers nonetheless have clear intuitions about how the key lenition processes are to apply in such contexts.
Therefore, it appears that enclitic syllables behave similarly to stem/foot-initial syllables in the range of 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 Medumbo. Speakers of Medumbo 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 Medumbo

a. \textit{m\textsuperscript{b}\textsuperscript{\text{"}{\text{d}}k}=\text{"}{\text{m}}} (\textit{\text{"}{mb}\textsuperscript{\text{"}{\text{d}}y}=\text{"}{m}}) ‘my fire’

b. \textit{m\textsuperscript{b\text{"}{\text{d}}l}=\text{"}{\text{m}}} (\textit{\text{"}{mb}\textsuperscript{\text{"}{\text{\text{"}{l}}}=\text{"}{m}}) ‘my hill’

This dialectal difference in lenition patterns is apparent only in the pronominal examples: Bazou speakers show similar patterns to Bangangtè 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 Bangangtè and Bangoulap dialects reflects a recent change that Bazou speakers have not yet undergone. However these patterns have come about, pronominal enclitics constitute an interesting form to investigate, given that they seem to show simultaneous evidence of prosodic strength and weakness in Bangangtè and Bangoulap dialects.

In §4, I 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, I investigate whether stem-initial syllables in Medumbo—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 toward 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 Medumbo will show timing patterns that 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 and 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 one-third or two-thirds 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 that carries nuclear pitch accent. For example, in Figure 2, drawn from work by Cummins and Port (1998), the beginning of the vowel in the word duck (closely corresponding to its ‘perceptual center’; Morton et al. 1976) in the sentence Dig for a duck is found to preferentially align with the halfway, one-third, or two-thirds fraction of the repetition cycle, depending on how rapidly the subject is speaking (subjects are instructed to apply neutral focus to the utterance). While a pattern with duck occurring around one-third of the way through the repetition cycle is commonly found at 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 initial or final in a word—are more likely to be drawn to SHP positions than unstressed syllables.

**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 (a) one-third or 0.33, (b) one-half or 0.50, and (c) two-thirds or 0.67 of the repetition cycle.9

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 nonheads (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 metrical structure is present in various dialects of Korean, and that many AP-initial syllables are in fact metrically prominent. Results from these studies demonstrate the importance of metrical prominence—rather than word position or acoustic prominence—in determining speech cycling patterns.

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 that are coupled together in time (Cummins & Port 1998, O’Dell & Nieminen 1999, Port 2003, van Gelder & Port 1995). 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 2009:849), 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 a 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 one-third or two-thirds points in the cycle (in the case of a 3:1 foot-to-phrase frequency-locking pattern).

From the dynamical perspective, the hierarchy described in formal phonological accounts such as those in Fig. 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 one-third or two-thirds 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. Not only is the dynamical model of prosodic structure appealing due to its ability to account for observed patterns in speech cycling, but recent work in neuroscience also 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 2011, Poeppel 2003), suggesting that the model in Fig. 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 Fig. 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 that 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, 1984, 1986).

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

3.1. Stimuli. The study included two subexperiments, 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, syllable 3 and syllable 4 (σ₃ and σ₄). 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.¹⁰

¹⁰ 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
is the fact that participants intuited that the final nasal of the pronominal enclitic, giving us a comparison between formally summarized in 11.

Table 3. Stimuli for experiment 1. Syllables hypothesized to be prominent are in bold.

<table>
<thead>
<tr>
<th>SET</th>
<th>FORM</th>
<th>IPA TRANSCRIPTION</th>
<th>GLOSS</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>SISF</td>
<td>mén giak mbibó</td>
<td>child TNS wait</td>
<td>‘The child waited.’</td>
</tr>
<tr>
<td>A</td>
<td>SISI</td>
<td>mén giak mui há</td>
<td>child TNS lose bicycle</td>
<td>‘The child lost the bicycle.’</td>
</tr>
<tr>
<td>B</td>
<td>SISF</td>
<td>bán giak nfrámá</td>
<td>children TNS go out</td>
<td>‘The children went out.’</td>
</tr>
<tr>
<td>B</td>
<td>SISI</td>
<td>bán giak ntúm mén</td>
<td>children TNS accuse child</td>
<td>‘The children accused the child.’</td>
</tr>
<tr>
<td>C</td>
<td>SISF</td>
<td>sán giak mfúló</td>
<td>bird TNS fly</td>
<td>‘The bird flew.’</td>
</tr>
<tr>
<td>C</td>
<td>SISI</td>
<td>sán giak njú ló</td>
<td>bird TNS eat pineapple</td>
<td>‘The bird ate the pineapple.’</td>
</tr>
</tbody>
</table>

Table 4. Stimuli for experiment 2. Syllables hypothesized to be prominent are in bold. Clitic boundaries are indicated with '=' and affix boundaries with '-'.

<table>
<thead>
<tr>
<th>SET</th>
<th>FORM</th>
<th>IPA TRANSCRIPTION</th>
<th>GLOSS</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>PreSI</td>
<td>mén lén ná-nsí</td>
<td>child know INF-choose</td>
<td>‘The child knows how to choose.’</td>
</tr>
<tr>
<td>A</td>
<td>SilEnc</td>
<td>mén lén nú-ntsí</td>
<td>child know problem=their</td>
<td>‘The child knows their problems.’</td>
</tr>
<tr>
<td>B</td>
<td>PreSI</td>
<td>sán lén ná-ntsí</td>
<td>bird know INF-enter</td>
<td>‘The bird knows how to enter.’</td>
</tr>
<tr>
<td>B</td>
<td>SilEnc</td>
<td>mén lén tsí-ntsí</td>
<td>child know in.law=your</td>
<td>‘The child knows your in-laws.’</td>
</tr>
<tr>
<td>C</td>
<td>PreSI</td>
<td>sán lén ná-ntsí</td>
<td>bird know INF-drink</td>
<td>‘The bird knows how to drink.’</td>
</tr>
<tr>
<td>C</td>
<td>SilEnc</td>
<td>sán lén tó-ntsí</td>
<td>bird know ear=your</td>
<td>‘The bird knows your ear.’</td>
</tr>
</tbody>
</table>

3.2. Study predictions. Our null hypothesis is that Medumba does not display rhythmic differences between stem/foot-initial and noninitial/nonstem syllables. If this is the case, then we predict that any syllable should be equally likely to occur at any of the SHP 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 noninitial/nonstem syllables occurring in corresponding ordinal positions (e.g. σ1 position across prosodic form conditions), such that stem-initial syllables occur with greater frequency at SHP positions, and stem-final syllables, affixes, and perhaps enclitics occur farther from these positions. The study hypotheses are formally summarized in 11.

(11) a. **H0**: Syllables of all types (stem-initial, stem-final, affix, enclitic) will be equally likely to occur at SHP positions (e.g. 0.33, 0.5, and 0.67 proportions of the repetition cycle). Differences in relative timing for syllables in corresponding ordinal positions are not necessarily predicted.

b. **H1**: 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, noninitial/nonstem syllables should be parsed with the stem/foot-initial syllables that 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

---

usable for comparison. As there are relatively few CV noun roots initiated with a nasal, some forms (notably *tsí=gú* ‘your in-laws’ and *t̃n=ú* ‘your ear’) used in the SilEnc condition 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), *tsí* and *t̃* 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 *t̃ñ* ‘ear’ was syllabified with the vowel of the pronominal enclitic, giving us a comparison between *nú* and *pú̃* for the final syllable in set C.
repetition cycle in two (aligning key foot heads with the 0.5 phase position) or in three (aligning key foot heads with the 0.33 or 0.67 phase positions, or both).

Figure 4 shows several possibilities for how syllable alignment might be expected to pattern for experiment 1 and experiment 2. Pattern (a) in Fig. 4 shows, for the SISI condition, subdivision in two with four equally timed feet (one for each syllable of the sentence), with the final stem-initial/foot-initial syllable occurring at the 0.5 phase position. Pattern (b) for experiment 1 shows subdivision in three, so that the final two syllables (both stem-initial) each head their own foot, one at the 0.33 position, the other at the 0.67 position (the tense marker in $\sigma_2$ 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 two with the final stem-initial syllable ($\sigma_3$) aligned with the 0.5 position. Pattern (d) shows subdivision in three with this same syllable aligned with the 0.33 position. Finally, pattern (e) shows subdivision in three with $\sigma_2$ (the tense marker) and $\sigma_3$ (the final stem/foot-initial syllable) aligned with SHP positions 0.33 and 0.67, 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, $\sigma_4$ 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 $\sigma_4$ (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 $\sigma_2$ position and the infinitival prefix that occurs in $\sigma_3$ position), with the final foot initiated at the 0.5 position (subdivision in two). Pattern (g) shows these same three feet with the final foot initiated at the 0.33 condition. Pattern (h) shows three equal feet with the penultimate and final feet aligned with the 0.33 and 0.67 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 $\sigma_3$ and $\sigma_4$ are predicted to differ in their relative phase positions, with $\sigma_3$ occurring at SHP positions in the SIEnc condition and $\sigma_4$ occurring at SHP positions in the PreSI condition, consistent with the alternative hypothesis.

![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 two (dashed) and subdivision in three (dotted) strategies. MP stands for ‘metrically prominent’.](image-url)
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 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: (i) $\sigma_3$ and $\sigma_4$ will occur later in the SISF condition than in the SISI condition, with $\sigma_3$ occurring in an SHP position in the SISF condition parallel to that of $\sigma_4$ in the SISI condition (assuming participants opt for patterns (a) and/or (b) with patterns (c) and/or (e)); or (ii) $\sigma_3$ will be timed similarly across prosodic form conditions, and $\sigma_4$ 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 $\sigma_3$ to occur earlier in the SISF condition than in the SISI condition, or for $\sigma_4$ 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, $\sigma_4$, in the SIEnc condition. Results should either show that (i) $\sigma_3$ occurs later in the PreSI condition than the SIEnc condition (assuming participants opt for patterns (f) or (h)), with $\sigma_4$ occurring in parallel SHP positions across conditions (assuming the enclitic is prominent and participants opt for patterns (a) or (b)), or (ii) $\sigma_3$ occurs in parallel harmonic phase positions in the SIEnc condition to $\sigma_4$ in the PreSI condition, with $\sigma_4$ in the SIEnc condition falling farther from harmonic phase positions than in the PreSI condition (assuming the enclitic is not prominent and participants opt for patterns (c), (d), or (e)). Should participants opt for pattern (g), $\sigma_3$ and $\sigma_4$ will both occur earlier in the PreSI condition than in the SIEnc condition. What we should not expect is for $\sigma_3$ 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 (six female) and fourteen in experiment 2 (seven female). Three male participants were unavailable for experiment 2 and were replaced with three other male participants of the same ages, plus one additional female participant. Ages of participants ranged from nineteen to fifty years old (mean age of thirty-two). All were native speakers of Medumba 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 was elicited at fifteen different speech rates, from slowest (speed 1) to fastest (speed 15). This resulted in a maximum of ninety total trials per participant (though see below for repetition success rates for each subject), each of which consisted of eight 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 fifteen metronome speeds).
The target sentence was displayed on the computer screen in PowerPoint in white font against a black background in both French and Medumba, in the native Medumba orthography. In case participants were not able to read the target sentences, the experimenter read the sentence in both French and Medumba 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 in Medumba only. 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 (σ₃ or σ₄) 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,¹¹ 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 σ₄, 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 σ₄ (estimate = −0.02, t = −9.00, p < 0.001), with σ₄ occurring earlier in the SISF condition than in the SISI condition. No effect of Condition was found for σ₃ across prosodic conditions (estimate = 0.001, t = 0.64, p = 0.53). Figure 5 shows that the mean alignment position for σ₄ in the SISI condition was right at 0.5 and slightly earlier, at 0.47, in the SISF condition. Mean alignment of σ₃ for both the SISI and SISF conditions was somewhat farther from an SHP position (0.37 and 0.38, respectively), but 0.33 was the closest position in both conditions. This pattern is discussed further in §4.2.

As has been found in previous work (Cummins & Port 1998, Tajima 1998, Zawaydeh et al. 2002), the proportion phase of syllable repetitions did increase with speech rate for both σ₃ (estimate = 0.01, t = 12.21, p < 0.001) and σ₄ (estimate = 0.02, t = 14.23, p < 0.001). As has also been found in previous work, variability in syllable alignment was slightly greater for phrase-final syllables (σ₄) than phrase-medial syllables (σ₃), with interquartile ranges (IQRs) at 0.12 and 0.10, respectively. This finding is in line with observations that phrase-final syllables tend to occur with phrase-final lengthening and,  

¹¹ 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.
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: the IQR for $\sigma_3$ was 0.11 in the SISF condition and 0.09 in the SISI condition, while for $\sigma_4$ there was found to be an IQR of 0.13 in the SISF condition and 0.11 in the SISI condition.

Experiment 2. For experiment 2, it was unclear whether a difference in alignment would be found for $\sigma_4$ 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 $\sigma_4$, with this syllable timed later in the SIEnc condition (estimate = −0.01, $t = −4.20$, $p < 0.001$). A significant difference was also found in the alignment of $\sigma_3$ across prosodic conditions, with the stem/foot-initial syllable in the SIEnc condition aligned later than the prefix syllable in the PreSI condition (estimate = −0.02, $t = −6.55$, $p < 0.001$). Mean phase positions for the PreSI and SIEnc conditions were 0.30 and 0.35, respectively, for $\sigma_3$, and 0.50 and 0.53, respectively, for $\sigma_4$ (Figure 6).

Once again, proportion phase was found to increase with metronome speed for both $\sigma_3$ (estimate = 0.01, $t = 11.69$, $p < 0.001$) and $\sigma_4$ (estimate = 0.02, $t = 15.05$, $p < 0.001$). As was found for experiment 1, the IQR for syllable alignment was found to be greater...
for $\sigma_4$ (0.14) overall than for $\sigma_3$ (0.11). Going by condition, variability was greater for $\sigma_3$ in the SIEnc condition than in the PreSI condition, with IQRs of 0.09 and 0.14, respectively. Variability was more comparable for $\sigma_4$ across conditions, but the IQR was still slightly higher at 0.15 in the SIEnc condition than in the PreSI condition, where it was found to be 0.14.

4.2. Proximity of syllable alignment to SHP positions and effects of metronome speed. Results from experiment 1 showed patterns that accorded with predictions, in that $\sigma_4$ occurred significantly earlier in the SISF condition than in the SISI condition. The lack of difference in alignment of $\sigma_3$ also followed from our predictions for alignment strategies (a) and (d) in Fig. 4. In experiment 2, differences in syllable alignment were observed for both $\sigma_3$ and $\sigma_4$ across prosodic conditions; however, the direction of this difference was different from what was predicted for $\sigma_4$. 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 Medumba, 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 0.33 phase position at slower speech rates to aligning that same syllable to the 0.50 or 0.67 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 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 $\sigma_3$) show a clear secondary peak, suggesting that the distributions are multimodal. This multimodality may, then, be associated with distinct alignment strategies at different speech rates, or for individual speakers.

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 Fig. 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 sometimes to 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 σ4 (in gray) between speeds 1 and 4, then shows a larger change in alignment position at around speeds 5 and 6, before stopping off again at the 0.5 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 0.33 SHP position for σ3 (in black), and then another plateau around the 0.5 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 deter-
mined SHP positions of 0.33, 0.5, and 0.67. A question now arises: if we examine clustering in the data across conditions within each experiment, are participants equally likely to align stem-final/nonstem syllables with SHP positions as they are to align stem/foot-initial syllables with these positions? This question is investigated in §4.3.

4.3. Assessing multimodality: Gaussian mixture models. Based on the impressionistic patterns of multimodality displayed in Figs. 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 ($\mu$) and standard deviation ($\sigma$). Assuming that 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 mean and standard deviation, and each accounting for a particular proportion ($\lambda$) of the overall data set for that syllable.

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 nonobserved complete data. Starting with some initial parameter values $\theta$, an expectation-maximization (EM) algorithm is applied iteratively (i) 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 (ii) to reestimated 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 are assigned to a particular component based on the new estimated probabilities. The mixing proportion $\lambda$ 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 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 0.33, 0.5, and 0.67 at least some of the time. Standard deviations 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 not only that foot-initial syllables are more likely to be attracted to SHP positions in the speech cycling task, but also that temporal alignment of foot-initial syllables within a particular alignment mode is overall less variable (Tajima & Port 2003). Finally, the mixing proportion 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 and Port (1998) on which the current analysis is based predicts that participants should gravitate to alignment positions of 0.33, 0.5, and 0.67 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 two or three. 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 online supplementary materials that accompany this article. Results were found to match our predictions: the bootstrap procedure revealed that distributions for most syllables were optimally modeled with a three-component model, with an exception for σ₃ in experiment 2, which was optimally modeled with a two-component model for both prosodic conditions.

Model fitting. With the optimal component numbers for each syllable of interest having been established, 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 there were 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. 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

12 The supplementary materials are available at http://muse.jhu.edu/resolve/122.
13 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.
forms for which a two-component distribution was optimal), the two SHP positions were chosen with which the data were most likely to overlap. So, for example, since the overall means for $\sigma_3$ in experiment 1 were 0.38 and 0.39, the means were set to 0.33 and 0.5 (since there would be few data points occurring close to the 0.67 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 that ensures generalizability of a model within the data set. In this case, a three-fold leave-one-out procedure was performed,14 in which the data are 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 1,000 iterations was permitted for convergence to be obtained in the EM 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 online supplementary materials. For the MU models, with few exceptions, results of the chi-squared tests were not significant at the $p < 0.05$ 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 that provided the best fit to the testing data. This model was then fit to the overall data set for each condition. Results from those models are presented in §4.4 below.

4.4. GMM model results.

Experiment 1. Estimated model parameters for each syllable and prosodic form condition are provided in Table 5. 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 $\sigma_3$ showed a component with a

---

14 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 are sufficiently large to be statistically representative of the data set 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 data set contained (a maximum of) six repetitions for each condition per participant, a three-fold approach was deemed optimal so that each fold could in principle have an equal number of two observations per condition per participant.
mean at the 0.34 phase position to which 71% of the data would most likely belong, another smaller component with a mean of 0.43 to which 21% of the data would most likely belong, and a third component with a mean of 0.53 to which 8% of the data would most likely belong. In the SISF condition, $\sigma_3$ showed a component with a mean around the 0.34 phase position to which 60% of the data points would most likely belong, a smaller component with a mean around the 0.45 phase position to which around 22% of the data would most likely belong, and a third component with a mean around 0.44 and a large standard deviation, to which 18% of the data would most likely belong. Thus, across both conditions, repetitions clustered most heavily around the 0.33 SHP position, consistent with strategies (a), (b), and (d) for $\sigma_3$ from Fig. 4.

<table>
<thead>
<tr>
<th>CONDITION</th>
<th>MODEL</th>
<th>MEAN</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>MIXING</th>
<th>LOG LIKELIHOOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syll. 3,</td>
<td>MU</td>
<td>0.34</td>
<td>0.43</td>
<td>0.53</td>
<td>0.04</td>
<td>0.04 0.04 0.71</td>
</tr>
<tr>
<td>SISI</td>
<td>MS</td>
<td>0.33</td>
<td>0.50</td>
<td>0.67</td>
<td>0.05</td>
<td>0.06 0.02 0.82</td>
</tr>
<tr>
<td>Syll. 3,</td>
<td>MU</td>
<td>0.34</td>
<td>0.45</td>
<td>0.44</td>
<td>0.05</td>
<td>0.05 0.11 0.60</td>
</tr>
<tr>
<td>SISF</td>
<td>MS</td>
<td>0.33</td>
<td>0.50</td>
<td>0.67</td>
<td>0.06</td>
<td>0.07 0.13 0.72</td>
</tr>
<tr>
<td>Syll. 4,</td>
<td>MU</td>
<td>0.41</td>
<td>0.49</td>
<td>0.61</td>
<td>0.03</td>
<td>0.06 0.08 0.14</td>
</tr>
<tr>
<td>SISI</td>
<td>MS</td>
<td>0.33</td>
<td>0.50</td>
<td>0.67</td>
<td>0.05</td>
<td>0.07 0.06 0.27</td>
</tr>
<tr>
<td>Syll. 4,</td>
<td>MU</td>
<td>0.41</td>
<td>0.52</td>
<td>0.54</td>
<td>0.05</td>
<td>0.05 0.11 0.49</td>
</tr>
<tr>
<td>SISF</td>
<td>MS</td>
<td>0.33</td>
<td>0.50</td>
<td>0.67</td>
<td>0.04</td>
<td>0.09 0.10 0.08</td>
</tr>
</tbody>
</table>

Table 5. Final model parameters, experiment 1.

Turning to $\sigma_4$ in the SISI condition, the optimal model had a component with a mean phase position of 0.41 to which 14% of the data points most likely belonged, another component with a mean at 0.49 to which 68% of the data most likely belonged, and a
third component with a mean of 0.61, 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 0.41 to which 49% of the data was most likely assigned, another component with a mean of 0.52 to which 27% of the data was most likely assigned, and a third component with a mean of 0.54 to which 24% of the data was most likely assigned. Thus, while $\sigma_4$ showed similar mean alignment positions across conditions, this syllable gravitated much more consistently to the 0.5 SHP position in the SISI condition than in the SISF condition, where the bulk of repetitions occurred around the 0.41 phase position, far from any of the theoretically determined SHP positions. The dominant patterns for $\sigma_4$ are therefore consistent with strategies (a) and (d) from Fig. 4.

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 0.41 phase position for $\sigma_4$ in both the SISI and SISF conditions, a position that is far from either the 0.33 or 0.5 SHP positions. Similarly, $\sigma_3$ in both the SISI and SISF conditions showed alignment between the 0.43 and 0.45 phase positions for some portion of the task. Overall, model estimates of mixing proportions indicate that these alignment positions were relatively marginal compared with the 0.33 and 0.5 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 $\sigma_4$ in the SISF condition, the one syllable that did not represent a stem/foot-initial syllable. Thus, as predicted, $\sigma_4$ in the SISF condition demonstrated the greatest difference in model fits between the MU and MS 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 two, with the final syllable attracted to the 0.5 SHP position, and the SISF condition showed subdivision in three, with $\sigma_3$ attracted to the 0.33 SHP position, as demonstrated in strategies (a) and (d) in Fig. 4 for experiment 1.

**Experiment 2.** Estimated parameters for the 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 6 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 $\sigma_3$ in the SIEnc condition yielded a component with a mean phase position of 0.32 to which 76% of the data would most likely be assigned, and another component with a mean position of 0.46 to which 24% of the data would most likely be assigned. In the PreSI condition, the best-fit model showed a component with a mean of 0.28 to which 74% of the data would most likely be assigned, and another component with a phase position of 0.37 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 0.33 SHP position, the infinitival prefix in the corresponding syllable position occurred regularly at a phase position quite a bit earlier than the 0.33 position and far from any of the identified SHP positions. While these results are consistent with strategies (a) and (b) for $\sigma_3$ from Fig. 5 in the SIEnc condition, $\sigma_3$ in the PreSI condition did not align with any of the predicted positions in Fig. 4.
Turning to $\sigma_4$, the best-fit model for the SIEnc condition had a component with a mean phase position of 0.38 to which 9% of the data was most likely assigned, another component with a mean phase position of 0.49 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 0.62 to which 42% of the data was most likely assigned. For the PreSI condition, a component was identified with a mean of 0.38 to which 16% of the data was most likely assigned, another component with a mean of 0.48 to which 67% of the data was most likely assigned, and a third component with a larger standard deviation with a mean of 0.67 to which 17% of the data was most likely assigned. Based on these parameter estimates, it seems $\sigma_4$ did occur consistently at the 0.5 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, 0.67, in the PreSI condition than in the SIEnc condition. These results are consistent with strategies (a) and (f) for $\sigma_4$ in Fig. 4.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Model</th>
<th>Mean $\mu$</th>
<th>Standard Deviation $\sigma$</th>
<th>Mixing Proportion $\lambda$</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syl. 3,</td>
<td>MU</td>
<td>0.32</td>
<td>0.06</td>
<td>0.76</td>
<td>3274.70</td>
</tr>
<tr>
<td>SIEnc</td>
<td>MS</td>
<td>0.33</td>
<td>0.05</td>
<td>0.54</td>
<td>3253.26</td>
</tr>
<tr>
<td>Syl. 3,</td>
<td>MU</td>
<td>0.28</td>
<td>0.05</td>
<td>0.74</td>
<td>3814.45</td>
</tr>
<tr>
<td>PreSI</td>
<td>MS</td>
<td>0.33</td>
<td>0.07</td>
<td>0.99</td>
<td>3402.35</td>
</tr>
<tr>
<td>Syl. 4,</td>
<td>MU</td>
<td>0.38</td>
<td>0.03</td>
<td>0.09</td>
<td>2420.52</td>
</tr>
<tr>
<td>SIEnc</td>
<td>MS</td>
<td>0.33</td>
<td>0.02</td>
<td>0.01</td>
<td>2368.60</td>
</tr>
<tr>
<td>Syl. 4,</td>
<td>MU</td>
<td>0.38</td>
<td>0.04</td>
<td>0.16</td>
<td>2540.92</td>
</tr>
<tr>
<td>PreSI</td>
<td>MS</td>
<td>0.33</td>
<td>0.02</td>
<td>0.05</td>
<td>2417.92</td>
</tr>
</tbody>
</table>

Table 6. Final model parameters, experiment 2.

![Density curves for model-theoretic components plotted against histograms of raw data, experiment 2. Dashed lines mark locations of SHPs at the 0.33, 0.5, and 0.67 phase positions.](image1)

a. Syllable 3, SIEnc.  
b. Syllable 3, PreSI.  
c. Syllable 4, SIEnc.  
d. Syllable 4, PreSI.
Once again, model parameter estimates for the MU models suggest that participants adopted alignment strategies that fell far from the theoretically determined SHP positions for at least part of the task. In particular, repetitions of $\sigma_4$ in both the SIEnc and PreSI conditions occurred with a phase position of around 0.38 for some portion of the task. The drop in log likelihood between the 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 $\sigma_3$ in the SIEnc condition showed less than a 1% drop in log likelihood from the MU model, the drop was at 11% for $\sigma_3$ in the PreSI condition. The drop in log likelihood was at around 5% for $\sigma_4$ 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 $\sigma_3$ occurs close to the 0.33 phase position and $\sigma_4$ occurs close to the 0.5 position, is consistent with pattern (a) in Fig. 4, alignment positions in the PreSI condition did not match any of the predicted patterns from Fig. 4. While $\sigma_4$ did consistently occur close to the 0.5 SHP position in this condition, consistent with strategy (f), $\sigma_3$ occurred quite early, consistently surfacing around a phase position of 0.28. This was considerably earlier than the predicted position given strategy (f) in Fig. 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 $\sigma_1$ and with $\sigma_2$ and $\sigma_3$ as dependents.

### 4.5. Results Synthesis

The goal of the experiments presented in §4 was to test the hypothesis that foot-initial syllables in Medumba, 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 SHP positions in the task. Meanwhile, we predicted that less prominent stem-final syllables and nonstem syllables, such as prefixes, should occur less consistently at these positions. With the overall differences in alignment of foot/stem-initial and noninitial/nonstem syllables highlighted, 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 found 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 noninitial/nonstem syllables to positions other than harmonic phase positions. These findings are consistent with our hypothesis that stem-initial syllables bear greater metrical prominence than nonstem/noninitial 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 Medumba are either absent or lacking in true metrical prominence asymmetries. Heads of feet in Medumba, 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 that lack unambiguous phonetic cues to stress, or where stress patterns may not align with metrical structure. 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 Medumba, as mentioned in §2, the confound of domain edges makes this type of analysis unreliable as a measure of metrical prominence in Medumba, 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 metrical prominence asymmetries highlighted in the present work may be one of the most important unifying factors underlying metrical prominence crosslinguistically.

That being said, 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’ (p. 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 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 e.g. Tilsen 2011 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 alignment strategies for enclitic syllables that were mostly similar to the ones they showed for stem-initial syllables. In particular, the drop in log likelihood between models with means specified to harmonic phase positions vs. without means specified was small—the smallest for any syllable across the two experiments—indicating that the more restrictive model actually fit the data for the pronominal enclitics quite well. Given the variability in lenition patterns across dialects for pronominal enclitics, one possibility is that lenition constitutes a change in progress that is further along in the Bangoulap dialect than in 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 metrical weak stem-final syllables. Prosodic weakening of functional elements such as pronouns is of course quite common crosslinguistically (Zwicky 1977, Jeffers & Zwicky 1980), particularly in posttonic 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 that 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 article to variation in metrical prominence at the word level, while 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 borne not only lexical stress, but also nuclear accent at the phrase level. The present work provides clear evidence that not all syllables within a word in Medumba have equal potential for rhythmic prominence; if the opposite were true, we would expect any syllable in a word to be equally likely to be attracted to the theoretically determined SHP 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 in Franich 2019 has shown that syllables in Medumba occurring earlier within a sentence (e.g. the second or third syllable within a longer six-syllable utterance) that bear acoustic evidence of phrase-level accent are also drawn more consistently to SHP positions than are 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 crosslinguistically. 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 (Özçelik 2014), and where rhythmic stress patterns may be conditioned by higher-level prosodic constituents (Kabak & Vogel 2001).

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 that assume articulatory gestures as phonological primitives have also demonstrated the importance of relative timing as an index of phonological structure (Browman & Goldstein 1988, 1992, Gafos 2002, Krakow 1989, 1999, Shaw et al. 2009, 2011, Tuller & Kelso 1991). 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, Nam & Saltzman 2003), 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 that 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. Nevertheless, there is still much work to be done in terms of understanding the underlying dynamics that shape speech behaviors in these striking
ways. Future work in phonology would benefit greatly from a deeper understanding of the physical laws that govern speech timing.

5.4. Links between speech cycling and musical structure. The 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 inquire, 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. Hayes 1983, Kiparsky 1975, Liberman 1975). In the case of English and other languages that have been examined, the speech cycling paradigm appears to tap into the same notion of beat strength that 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 multijoint 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. In the musical domain, by contrast, 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 into 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 that 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 that 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 Medumba 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 that are extraordinarily detailed: not only do we expect overall timing differences between prominent and nonprominent syllables, but we also 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 crosslinguistically, 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, Medumba has only a binary tonal contrast between high and low tones; 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, nonfocused contexts with one of two tone melodies, H or LH, and can be either monosyllabic or disyllabic. As can be seen in A1, 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 A1a is a single tone linked to both stem syllables (Hyman & Tadadjeu 1976).

(A1) Mono- and disyllabic verbs with H and LH tone melodies

a. zi ‘sleep’ ziná ‘walk’
   ʒuí ‘eat’ ʒuímá ‘be dry’
b. bá ‘be ripe’ báɣá ‘split’
   só ‘press’ sóɣá ‘throw’

Thus, while contour tones are permitted on stem/foot-initial syllables in some contexts, such as A1b, contours never occur on a noninitial syllable. In fact, evidence from loanwords from English suggests that contours are explicitly disallowed from occurring on syllables that 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

15 While researchers largely agree that both high and low tones must be specified in the tonal inventory of Medumba (Danis 2011, Franich 2017, Hyman 2003, Voorhoeve 1971, 1976), Keupdjio (2020) argues that lexical categories are limited to a low vs. unmarked distinction, in which the unmarked tone surfaces with a default high tone. I maintain that an 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 (cf. Myers 1998), with both tones undergoing processes such as tone spread and downstep.
only syllable, if monosyllabic), and a low tone assigned to the unstressed syllable (A2a), though HL falling contours are found in certain loanword contexts, such as words with a CVN syllable shape (A2b).

<table>
<thead>
<tr>
<th>Medumba loanword</th>
<th>English source IPA</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tî</td>
<td>[t̚i]</td>
<td>‘tea’</td>
</tr>
<tr>
<td>bá</td>
<td>[baj̚]</td>
<td>‘bike’</td>
</tr>
<tr>
<td>hàmá</td>
<td>[ˈhæmo]</td>
<td>‘hammer’</td>
</tr>
<tr>
<td>tìlò</td>
<td>[ˈt̚el̚o]</td>
<td>‘tailor’</td>
</tr>
<tr>
<td>b. pìn</td>
<td>[p̚i̚]</td>
<td>‘pin’</td>
</tr>
<tr>
<td>tám</td>
<td>[t̚æm]</td>
<td>‘time’</td>
</tr>
<tr>
<td>ˈgùmmá</td>
<td>[ˈɡavna]</td>
<td>‘governor’</td>
</tr>
<tr>
<td>sụ̀gĩ</td>
<td>[ˈs̚n̚ɡlat̚]</td>
<td>‘singlet’</td>
</tr>
</tbody>
</table>

Interestingly, HL contours are permitted only 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 [ɛ], 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 (A3).

The above patterns can be accounted for if we assume that (i) where possible, Medumba builds maximally disyllabic feet starting from the left edge of the stem; (ii) syllables bearing contour tones are banned from occurring in the dependent (rightmost) position of a foot; and (iii) vowel epenthesis is enacted as a repair strategy to enable the second syllable in the forms in A3 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 Medumba 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 (A4).

<table>
<thead>
<tr>
<th>Medumba loanword</th>
<th>English source IPA</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>dósíni</td>
<td>(*d̚os̚i̚n)</td>
<td>‘dozen’</td>
</tr>
<tr>
<td>f̚lebáná</td>
<td>(*f̚leb̚án)</td>
<td>‘fry pan’</td>
</tr>
</tbody>
</table>

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 contour 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 Medumba (A5, A6).

<table>
<thead>
<tr>
<th>Medumba loanword</th>
<th>English source IPA</th>
<th>English translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>kámára</td>
<td>[ˈk̚əm̚əra]</td>
<td>‘camera’</td>
</tr>
<tr>
<td>kábíɗáɗá</td>
<td>[ˈkap̚inta]</td>
<td>‘carpenter’</td>
</tr>
<tr>
<td>ˈɛ̱mb̚ási</td>
<td>[ˈɛ̱m̚b̚asi]</td>
<td>‘embassy’</td>
</tr>
<tr>
<td>bitáli</td>
<td>[b̚ita_li̚]</td>
<td>‘bitter leaf’</td>
</tr>
</tbody>
</table>

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 contour 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 Medumba (A5, A6).

Recall that stem-internal tone spreading is also common to native words in Medumba.

16 The reason for this pattern likely has 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 Medumba.

17 Note that the epenthesis which results in the examples in A3 cannot be attributable to the illicitness of CVCVC as a foot shape, since we do find other loanwords of this shape, for example, [slǐp̚t̚] ‘slipper’ and [t̚òv̚èt̚] ‘towel’. Native Medumba roots of the shape CVCVN can also be found, such as bålì̌g̚ ‘potato’; low-tone spreading in such forms seems to indicate that they also form a single foot.

18 Superscript nasals in examples in A4 indicate that the nasal is parsed as part of a ‘prenasalized’ t̚C onset, rather than as a coda-onset sequence.
Trisyllabic examples such as sásídè ‘Saturday’, kísímì ‘kitchen’, kúbátì ‘cupboard’, and líbátì ‘liberty’ demonstrate that lenition patterns discussed in §2.1 and tone spreading converge on the same foot patterns (A7).

\[
\begin{align*}
(A7) \ F_t(sásí) & F_t(\text{dè}) F_t \\
F_t(kísí) & F_t(mì) F_t \\
F_t(kúbá) & F_t(tì) F_t \\
F_t(líbá) & F_t(tì) F_t
\end{align*}
\]

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 Medumba (Franich 2018).\(^1\) 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 experiments 1 and 2, respectively) in Figure A1.

\(^1\) Note that the word-initial nasal for the verb in both prosodic conditions in experiment 1 is a prefix. Franich 2018 showed that these prefixes do not influence p-center location of a CV sequence that they prefix to, suggesting that they form their own separate timing unit. This could indicate that the nasal is itself syllabic, though, as noted in Franich 2018, Medumba speakers do not have the intuition that they are. Another possibility is that the prefix is ‘extrasyllabic’ (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.
Timing of each syllable (referred to henceforth as the syllable’s relative phase) was calculated in terms of the phrase 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 A2: 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 \( \sigma_1 \) to the p-center of the following repetition of \( \sigma_1 \)). This gives the relative phase measure for \( \sigma_3 \); similar measures are taken for other syllables.

\[
\text{Figure A2. Relative phase for } \sigma_3 \text{ in a four-syllable utterance is interval } b \text{ divided by interval } a.
\]

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 compared for each subject only data 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 online supplementary materials.

Finally, data points 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 the 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 A8, where \( \lambda_j \) stands for the mixing weights, or the probability that a randomly selected observation comes from component \( j \).

\[
(A8) \quad g_d(x_i) = \sum_{j=1}^{m} \lambda_j \phi_j(x_i), \quad x_i \in \mathbb{R}^r
\]

Here, \( \theta = (\lambda, \phi) = (\lambda_1, \ldots, \lambda_m, \phi_1, \ldots, \phi_m) \) denotes the parameter, and the mixing weights \( \lambda_j \) must be positive and sum to 1. We assume here that the parameters \( \phi_j \) are drawn from a parametric family \( \mathcal{F} \) of Gaussian distributions. The univariate Gaussian probability density function can be written as in A9.

\[
(A9) \quad \mathcal{F} = \{\phi(\cdot|\mu, \sigma^2) = \text{density of } \mathcal{N}(\mu, \sigma^2), (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^*_+\}
\]

Model parameters can then be reduced to A10.

\[
(A10) \quad \theta = (\lambda, (\mu_1, \sigma_1^2), \ldots, (\mu_m, \sigma_m^2))
\]

**REFERENCES**


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 32.1–29. DOI: 10.18637/jss.v032.i06.


Katsika, Argyro; Jelena Krivokapic; Christine Mooshammer; Mark Tiede; and Louis Goldstein. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44.62–82. DOI: 10.1016/j.j Phonetics.2014.03.003.


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. *Proceedings of the National Academy of Sciences* 111(40).14559–64. DOI: 10.1073/pnas.1406219111.


[kfranich@udel.edu] [Received 28 February 2020; revision invited 13 July 2020; revision received 27 September 2020; revision invited 24 November 2020; revision received 17 December 2020; accepted 28 December 2020]