ARTICLE IN PRESS

Journal of Phonetics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language

Kathryn Franich

Department of Linguistics, University of Chicago, 1115 E. 58th Street, Rosenwald Hall, Room 224, Chicago, IL 60637, USA

ARTICLE INFO

Article history:
Received 9 May 2017
Received in revised form 2 November 2017
Accepted 7 November 2017
Available online xxxx

Keywords:
Rhythm
Speech perception
Speech timing
Tone
Morphology
Phonology

ABSTRACT

Perceptual centers (or 'p-centers') correspond to the perceptual moment of occurrence of a syllable or word, and are crucial in the perception of speech rhythm. A metronome alignment task was used to investigate how tone and prenasalization—two elements which affect speech timing and which also interact acoustically—influenced p-center location in Medʉmba, a Grassfields Bantu language. Plain CV words bearing low tones were found to have p-centers which were later (farther from consonant releases and closer to vowel onsets) than those bearing high tones, but the observed effect was not present in prenasalized words. We attribute this difference to the effects of tone depression and slope leveling in prenasalized forms. While prenasalization generally led to earlier p-centers (mirroring effects found for onset clusters in other languages), forms with morphologically-derived prenasal onsets behaved more like plain CV forms, suggesting that nasal prefixes do not contribute to p-center timing. Our findings for derived prenasal sequences parallel similar articulatory findings for languages with simplex onset coordination, where consonant 'clusters' actually behave as separate timing units.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Rhythmic timing in speech has been of longstanding interest to linguists, though identifying precisely which elements in speech should be expected to behave as rhythmic, and in what way, remains a problem. Early approaches to rhythm analysis proposed a central role for *isochrony*, or equal timing between elements in speech, forming the basis of the distinction between hypothesized rhythm classes such as stress-timed and syllable- or mora-timed (Abercrombie, 1965, 1967; Bloch, 1950; Pike, 1945). Acoustic evidence for isochrony has been notoriously weak, however, with most studies providing little evidence for equal spacing of stress beats, syllables, or moras in any language under investigation (Bolinger, 1965; Dauer, 1983, 1987; Delattre, 1966; Roach, 1982; Shen & Peterson, 1962). This fact has led some to abandon the notion of isochrony altogether in characterizing speech rhythm, focusing instead, for example, on crosslinguistic differences in patterns of durational variability of consonant and vowel intervals in the speech stream (Dellwo, 2006; Grabe & Low, 2002; Ramus, Nespor, & Mehler, 1999). While these approaches have provided some empirical support for distinctions in rhythm class between languages (or at least for a continuum of rhythm types), different duration-based metrics have sometimes led to distinct rhythmic categorizations of the same language, or lack of a clear categorization for some languages. Furthermore, many factors, such as the type of speech elicited (e.g. isolated sentences versus conversational speech), have been found to influence results (see Arvaniti, 2009, 2012 for comprehensive overviews). Crucially, these approaches also fail to address the strong intuition among listeners of different languages that speech sounds isochronous (Lehiste, 1977). Indeed, a growing body of research shows that *perceived* isochrony in speech is beneficial for speech processing (Brown, Salverda, Dilley, & Tanenhaus, 2011; Dilley & McAuley, 2008; Dilley, Mattys, & Vinke, 2010, 2012). To better understand the facilitative role of perceived isochrony in speech processing, a clearer picture of how it relates to different acoustic and structural factors in language will be necessary.

The connection between perceived isochrony and phonetic properties has received a great deal of attention within studies of *perceptual centers*, or 'p-centers', which reflect the perceptual moment of occurrence of a syllable and serve as the locus of perceived isochrony (Fowler & Tassinary, 1981; Morton, Marcus, & Frankish, 1976; Rapp, 1971). P-center effects were first described as such by Morton et al. (1976) in relation to a

E-mail address: kfranich@uchicago.edu
URL: http://home.uchicago.edu/kfranich/

https://doi.org/10.1016/j.wocn.2017.11.001
0095-4470/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Franich, K. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* (2017), https://doi.org/10.1016/j.wocn.2017.11.001

memory experiment examining whether number recall could be improved if digits were presented at evenly-spaced temporal intervals in a list. In this experiment, it was found that digits manipulated into isochronous sequences based on onset of acoustic energy were not perceived to sound evenly-spaced, and in fact sounded quite arhythmic to listeners. When listeners were asked to adjust syllables to sound more regularly-timed, the adjustments made suggested that p-centers were tied neither exclusively to the syllable onset nor the vowel onset, but somewhere in between; the authors concluded that it must correspond to some function relating the two.

Other work has tried to identify more precisely the acoustic bases for the p-center effect. Marcus (1981) used a similar design—termed the *rhythm adjustment method* (see Villing, Repp, Ward, & Timoney, 2011 for further details)—where subjects evaluated temporal regularity between a set of alternating base and test sounds and adjusted test sounds such that they sounded more evenly timed with base sounds. Marcus found that a CV syllable's p-center tended to occur close to the vowel onset, but was influenced both by the duration of its onset consonant and, to a lesser degree, by its vowel and coda consonant durations. Most notably, increased onset consonant duration led to a shift in p-centers away from the vowel onset and to a point earlier in the syllable. Similar results were found using the same method by Pompino-Marschall (1989) and by Harsin (1997). Results of all of these studies also pointed to the importance of the role of the distribution of acoustic energy at critical frequency bands in determining p-center locations, a possibility which was tested explicitly in perception experiments and supported by Howell (1984, 1988), Scott (1998), Scott and Howell (1992).

Though the p-center effect is characterized as a perceptual phenomenon, various speech production studies examining speech timing have found similar results to those found in perception studies. Allen (1972) found that, when instructed to tap 'on the beat' of designated syllables within a sentence, subjects located their taps just before vowel onsets, and the precise duration between tap and vowel onset varied as a function of the duration of the prevocalic consonant. Similar results were found by Rapp (1971) when having subjects repeat noncewords to a regular metronome pulse (a method we refer to as the *metronome asynchrony procedure*). Fowler and Tassinary (1981) also had subjects repeat syllables to a metronome pulse, this time examining effects of different consonant clusters on p-center alignment. They found that, while speakers tended to align utterances such that vowel onsets occurred with the metronome pulse, as the number of segments in the syllable onset increased, the pulse occurred farther from the vowel onset, occurring instead somewhere within the onset consonant cluster. This finding was replicated by Šturm and Volín (2016) for the Czech language.

Additional studies have investigated the possible role of other aspects of phonology and syllable structure in affecting p-center location. Cooper, Whalen, and Fowler (1986), using the rhythm adjustment method, showed that onset cluster effects on p-centers are not influenced by the category of the segments involved, but rather depend exclusively on the duration of the cluster sequence as a unit. Further work has investigated contributions of the syllable rime to p-center location. Cooper et al. (1986, 1988) showed that duration of the syllable rime exerted a small but consistent influence on the location of p-centers, but not nearly as strong an effect as had been found with syllable onsets. The authors conclude from this asymmetry that syllable structure is an important factor in determining the effect of a particular segment on p-center timing. Similar results were obtained by Šturm and Volín (2016), who found that vowel duration and coda duration had a far weaker effect on p-center timing than onset duration in Czech.

Browman and Goldstein (1988) compare p-center results to those found with respect to an articulatory phenomenon known as the *C-center effect* (short for 'consonant center'). This refers to the strong tendency for the timing of onset consonantal gestures to correspond as a unit to other articulatory landmarks such as the target of the vowel gesture and the acoustic release of the syllable coda,[1] such that the midpoint of the onset consonant or cluster of consonants remains stably timed with these landmarks. Just as the p-center of a syllable moves consistently away from the syllable's vowel onset as onset duration increases, so, too, does the C-center. Thus, it would appear that p-centers may be tied to articulatory events, as has also been proposed by Fowler (1979, 1983). Also similar across the two phenomena is the fact that both appear far more dependent on properties of syllable onsets than syllable codas. These results are interpreted within the framework of Articulatory Phonology (Browman & Goldstein, 1990, 1998, 2000) as evidence that onset consonantal gestures, at least in English, are coupled *in-phase* with the following vowel, whereas codas are coupled 180° *anti-phase* to the vowel gesture. Thus, when the structure of the onset is changed through the addition of a consonant, the coordinative structure of the onset + vowel portion is affected as a whole, whereas addition of a coda consonant affects coda timing, but not vowel timing (at least not to a large degree).

Further work has set out to explicitly tie the p-center effect to articulatory events, but results have been mixed. de Jong (1994), by examining stimuli extracted from an articulatory database, found that while articulatory gestures predicted p-center location just as well as (if not better than) acoustic events, no single articulatory landmark (e.g. tongue tip minimum position or jaw maximum position) acted as the sole correlate of p-center location. Rather, the author concluded that p-centers correspond to a complex of articulatory events in the syllable. Likewise, Patel, Naito, and Löfqvist (1999) examined both acoustic and kinematic data in relation to p-center location, finding that no one landmark from either type of measure was exclusively tied to p-center location.

Despite the elusive nature of acoustic and articulatory landmarks associated with p-centers, the phenomenon remains remarkably robust across speakers of a language, and even across languages. Hoequist (1983), using the rhythm adjustment method, found that speakers of English, Spanish, and Japanese all tend to align p-centers around the onset of the vowel in monosyllables. Barbosa, Meireles, and Vieira (2005), using the metronome asynchrony method, found similar results for Brazilian Portuguese. Crosslinguistic results have revealed some interesting language-specific behavior of

[1] Note that different studies have used articulatory landmarks other than the release of the syllable coda as anchor points for measuring C-center effects; see Tilsen (2012) for an overview.

p-centers, however. In a recent metronome asynchrony study, Chow, Belyk, Tran, and Brown (2015) found that p-centers in Cantonese align more closely with the syllable onset, rather than the vowel onset, contrary to the results outlined above. The authors reason that, since Cantonese has fewer consonant clusters and overall less variability in syllable onset duration, onsets are a more reliable landmark around which to base the perceptual beat of the syllable. This study also investigated possible effects of tone on p-center location. Acoustic, articulatory, and phonological findings on East Asian tone languages have suggested that contour tones—which make up half of the Cantonese tonal inventory—have a more complex representation and timing structure than level tones (Duanmu, 1994; Gao, 2008; Gordon, 2001; Xu, 2004, 2009; Xu & Liu, 2006; Yip, 1991; Zhang, 2002). Despite this, no significant effect of tone was found in this study.

The present study attempts to examine tonal effects on p-center location once more, but this time in a language with a substantially different tonal inventory from Cantonese. As will be discussed, the Medʉmba language, a Bantu language, has a much smaller tone inventory, but the tones it does have exhibit acoustic properties which are interesting from the perspective of speech timing. In addition, we investigate properties of the word onset—namely, different types of prenasalization—to see how they may affect p-center location. Prenasalized sequences, which typically occur in word-initial position and involve a consonant and a preceding homorganic nasal segment (e.g. *ŋ̩gà*), are quite common in Bantu languages, and are known to interact with tone in interesting ways. Before discussing the design and results of our study, we outline details of tone and prenasalization in Medʉmba and how they might be expected to impact p-center location.

1.1. Tone and prenasalization in Medʉmba

Medʉmba is part of the Grassfields Bantu family spoken primarily in southwestern Cameroon, a language family which has not previously been studied with respect to the p-center effect. Apart from that, Medʉmba features a tone system which makes it an ideal test case for tonal effects on p-centers. Unlike Cantonese, which has six tones, Medʉmba has only a two tone inventory, comprising high and low. While both high and low tones are realized with a relatively flat pitch profile phrase-medially, low tones are realized with a falling contour both phrase-finally and in isolation, while high tones maintain level pitch in these environments (Voorhoeve, 1971). This very simple inventory thus features two tones which can pattern quite differently from one another both in terms of f_0 height and slope, two properties which have been found to have interesting effects on the perceived timing of syllables. Specifically, low tones are found to be perceived as shorter than high tones, even when duration is held constant, and contour tones are perceived as longer than level tones (Gussenhoven & Zhou, 2013; Yu, 2010). As already mentioned, contour tones—and falling tones, in particular—have also been found to have a more complex articulatory structure, at least in Mandarin Chinese (Gao, 2008). Within some models of speech timing, contour tones are also argued to differ from level tones in their tonal articulatory goals: contour tones have a dynamic tonal target, whereas level tones have a static target (Xu, 2004, 2009; Xu & Liu, 2006). If Medʉmba low falling tones behave as Mandarin falling tones and p-centers do reflect articulatory structure to some extent, we might expect tone to influence p-center location.

Of further interest here are effects of prenasalization and interactions between prenasalization and tone in affecting p-center location. Medʉmba, like many Bantu languages, has prenasalized consonant sequences, which can contain either a stop or a fricative with a preceding homorganic nasal. These sequences can occur either root-initially (some historically derived from noun class prefixes which are no longer productive) or in morphologically-derived forms in which a nasal prefix is concatenated with a word initiated with a plain oral stop or fricative. There are various ways in which prenasalization might be expected to interact with tone in affecting speech timing. It is known, for example, that nasal consonants are good 'carriers' of tone due to their highly sonorous profile. For example, contour tones are much more likely to occur on syllables with nasal codas than those with obstruent codas (Gordon, 2001; Zhang, 2002). Syllabic nasals are also known to bear tone in some Bantu languages (Hyman & Ngunga, 1997). It is possible, then, that the presence of the nasal segment would influence the timing of surrounding tones. Furthermore, prenasalized onset sequences are known to act as tone *depressors*, lowering f_0 of the following vowel (Cibelli, 2015; Hyman, 2008). If tone is found to affect p-center location, we might expect the effect to be mitigated in prenasalized forms, where height differences between high and low tones are minimized.

There are reasons to believe that all prenasalized sequences in Medʉmba act as complex entities. The first has to do with what Riehl (2008) refers to as *separability*: all segments which can form part of a prenasalized consonant (with the exception of allophonic [g]) can occur independently, as singleton consonants. The second is the fact that, as will be shown, derived and nonderived prenasalized sequences have similar durations, and are both much longer than plain segments in the language. Thus, prenasalized onsets may either behave as consonant clusters or as a sequence of a syllabic nasal plus an oral consonant. If they behave as consonant clusters, we would expect that p-centers should be pushed farther from vowel onsets and closer to word onset midpoints as compared with plain (non-prenasalized) words. If, however, the nasal portion of the sequence is syllabic, we might expect it to bear its own p-center, separate from that of the CV syllable it precedes. In this case, two additional outcomes are possible: (1) subjects will base their repetitions around the p-center of the nasal, such that metronome beats occur somewhere within the nasal segment; or (2) subjects will base their repetitions around the p-center of the following CV syllable, in which case metronome alignment should be similar between plain and prenasalized forms. We test for these various outcomes in Section 3.2.

Finally, despite apparent similarities between derived and nonderived prenasalized consonants, it is possible that timing differences with respect to the p-center might arise. It is known, for example, that differences in the timing of articulatory gestures exist between monomorphemic and bimorphemic words. Such effects have been accounted for in Articulatory Phonology in terms of the 'windows of variability' associated with

ARTICLE IN PRESS

timing and coordination strategies: while such windows are lexically-specified and therefore more constrained within a morpheme, they are unspecified and therefore more variable at morpheme boundaries (Cho, 2001). Thus, we might expect to find increased variability in p-center location in derived words. We investigate this possibility in Section 3.3.

1.2. Choice of method

As discussed in Section 1, there are various methods that have been used to estimate p-center location. Here, we adopt a method for measuring p-centers which we refer to as the metronome asynchrony procedure. This method has been used in several past studies (Barbosa et al., 2005; Chow et al., 2015; Fowler & Tassinary, 1981; Rapp, 1971; Šturm and Volín, 2016), and involves having subjects repeat words in time to a regular metronome pulse. This method differs from another commonly-used method, the rhythm adjustment method, in which subjects are instructed to adjust the timing of a test sound so that it sounds isochronous with preceding and following base sounds (i.e. to achieve the 'point of subjective isochrony' between sounds).[2] As discussed above, these methods have largely been found to yield similar results for p-center effects. The benefit of the production-based metronome asynchrony procedure is that it is quite easy to implement, and, as discussed by Villing et al. (2011), requires little effort or technical skill on the part of the participant to complete. This is ideal for a setting such as ours in which some participants did not have comprehensive experience using computers or participating in laboratory-based experiments. The metronome asynchrony procedure has also been used to examine p-center timing across a wide variety of languages (including English, Brazilian Portuguese, Cantonese, and Czech), allowing us to compare crosslinguistic results. Villing et al. (2011) point to one potential shortcoming of the method, which is that delays between metronome beat perception and performance of the motor task involved in speech production might lead to misalignment of a syllable's p-center with the metronome beat. However, since the goal of the present study is not necessarily to pinpoint the exact location of the p-center of a given word, but rather to assess the kinds of acoustic and structural factors which might affect its relative location, this shortcoming is not of grave concern here.

2. Participants, materials, and methods

2.1. Participants

Participants were 11 native speakers (5 female) of the Medʉmba language from around Bangangté, Cameroon. Subjects ranged in age from 20 to 53, with a median age of 27. Though the age range was fairly wide, participant ages skewed younger, with nearly two-thirds of subjects under the age of 40. No subjects reported or demonstrated speech or hearing problems. All subjects were provided with an informed consent form which was read aloud to them by the researcher (as some subjects did not read). The research was approved by the University of Chicago Internal Review Board. Subjects were paid the equivalent of $10 US for their participation.

2.2. Materials

Lexical items to be elicited consisted of 35 words beginning either in a plain oral stop /b/, /t/, or /k/, or a stop with a preceding homorganic nasal ([m], [n], or [ŋ]). Distributional factors suggest that voicing is not contrastive for either /b/ or /k/. However, /k/ often becomes voiced after a nasal: voicing occurs variably in derived environments and across the board in nonderived environments.[3] Words containing /d/ as an onset are few, and were thus not collected for this experiment. Most words were of (N)CV shape, though four words with coda consonants (/ʔ/ or /b/) were included (recall that a syllable's coda—especially if comprised of a single consonant—has not been found to have much effect on its p-center). Words were distributed across two tones, high and low. Words with nasals were also distributed across two morphological conditions, derived and nonderived. Derived forms were CV(C) noun or verb roots with nasal prefixes, either constituting a plural prefix (glossed as PL) or a tense-related verbal prefix (glossed as TNS). While we have alluded to the possibility that nasal prefixes could be syllabic, all words were judged to be monosyllabic by two native speakers. Note that we refer to derived and nonderived conditions collectively as 'prenasalized', despite the different status of nasal segments in each. Though efforts were made to create balanced word sets across conditions, lexical gaps in the language made this infeasible. As discussed in Section 2.5, we specifically chose a statistical approach—linear mixed effects models—which would be robust to imbalance in data between conditions. A full list of words can be found in Appendix A.

2.3. Procedure

Participants were seated at a table in a quiet hotel room in Bangangté and 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 connected to a 2010 Macbook Pro through which the metronome beats were played. The metronome sound consisted of a synthetic drumbeat created in Audacity sound editing software, version 1.3.8.[4] Metronome beats were spaced 770 ms apart, as determined by pilot testing with three subjects to assess a generally comfortable speaking rate. For each trial, the experimenter read the target word in both Medʉmba and French (the locally-spoken ex-colonial language) in order to ensure the subject would repeat the correct word. When subjects were ready and knew the word they were to repeat, the experimenter initiated the sound of the drumbeat by pushing a button on the laptop. Twelve drumbeats were played for each target word. Subjects were instructed to listen to the first four beats and begin repeating on the fifth, speaking each word once per beat in synchrony with the metronome click, for a total of eight repetitions of each word. Subjects rested briefly after each trial, and then moved on to the next. Longer breaks occurred every 10 trials. Repetitions

[2] See Villing et al. (2011) for discussion of a third method, known as the *phase correction response method*, which measures subjects' adaptation behavior to perturbations in an otherwise isochronous sequence.

[3] Voicing in words in Appendix A is transcribed according to the pattern most frequently produced across speakers.

[4] A sample of this sound can be found at https://github.com/kfranich/Medumba-p-center/.

Please cite this article in press as: Franich, K. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* (2017), https://doi.org/10.1016/j.wocn.2017.11.001

and metronome beats were recorded simultaneously on two separate channels on a Zoom H4nSP recorder. Prior to analysis, first and last repetitions for each trial were removed to avoid transient effects. Thus, in total, the maximum number of analyzable tokens uttered per subject was 210 (35 words × 6 repetitions), for a total of 2310 utterances. The study took approximately 30 min to complete. One male subject's data was excluded entirely due to the fact that he had trouble aligning his repetitions with the metronome and often forgot to wait for the first four beats to play to assess the repetition tempo. Due to some dialectal variation, some words (or plural forms of words) were not recognized by some speakers. A total of 7% of the remaining tokens were not produced due to this reason; a summary of missing trials can be found in Appendix B. During the course of the experiment, subjects would occasionally forget to begin repeating the word on the correct beat, or would mispronounce a word and attempt to quickly repeat it. This often led to a drastic misalignment of repetitions with the metronome beat. In order to avoid having these clear errors affect statistical analysis, prior to model fitting, datapoints corresponding to vowel, consonant, or nasal distances/durations which were more than 2 standard deviations from mean values were excluded as outliers. This resulted in trimming of an additional 3.5% of the data.

2.4. Data annotation and preparation

Data were annotated in Praat 6.0.19. Onsets of metronome beats were automatically detected and marked using an in-house Praat script. Subjects' repetitions were annotated in Praat by the author based on the acoustic landmarks listed in Table 1. Intervals that were analyzed, including the initial nasal segment (where applicable), stop closure, release burst, and vowel, are indicated for the words *bə́* 'be' and *mbə́*, a complementizer, in Fig. 1.

Subsequent to annotation, TextGrids for metronome and repetition data were combined into a single TextGrid so that timing of each segment and metronome beat could be extracted and distance (in ms) of each segment from the metronome beat could be calculated. Distances were calculated for each repetition of each word. Fig. 2 shows sample measurements of the difference between the metronome beat and the nasal onset (N_{Diff}), the consonant onset (taken as the onset of the release burst; C_{Diff}), and the vowel onset (V_{Diff}).

2.5. Statistical analysis

Data were analyzed using linear mixed effects models, implemented with the `lmer` package for *R* statistical software from Bates, Mächler, Bolker, and Walker (2015). Linear mixed effects models allow for the inclusion of both fixed and random effects and are well-suited for analysis of unbalanced data. Where multiple comparisons were carried out, we report model results with Bonferroni-corrected p-values.

3. Results

Results of statistical models are presented in three parts: in Section 3.1, we look at onset and segment duration between plain words and the two types of prenasalized words; in Section 3.2, we examine effects of prenasalization and tone on p-center location; and in Section 3.3, we look at the effect of morphological status of the nasal on p-centers.

Table 1
Acoustic landmarks marked for analysis.

Segment	Acoustic landmark
Prenasalization/nasal prefix	Onset of nasal murmur
Stop release	Onset of release burst
Stop closure (prenasalized forms only)	Silence between nasal and release burst
Stop closure (/b/ only)	Onset of prevoicing
Vowel onset	Onset of vowel formants

3.1. Onset and segment duration

The first set of analyses evaluated differences in duration of word onsets, their component nasal and oral stop segments, and vowel duration. Models were constructed for each of these dependent variables including a fixed effect of Onset Type which had, for whole onset and consonant duration measures, three levels: plain, prenasalized-nonderived and prenasalized-derived, and for nasal duration measures, just two levels: derived and nonderived. In examining differences in onset and segment duration, the variable was contrast-coded to evaluate those differences which appeared to be significant from visual inspection of data. The models also included by-subject random slopes for Onset Type.

In comparing the duration of word onsets (for plain forms, duration of release burst; for prenasalized forms, duration of the nasal, closure, and release burst combined), a significant effect of Onset Type was found between plain and prenasalized conditions ($t = -10.108$, $p < 0.001$, df = 9); as can be seen in Fig. 3[5] plain onsets were significantly shorter than prenasalized onsets, by an average of 125 ms. No significant difference in word onset duration was found between the two prenasalized conditions. In comparing just burst duration between the three conditions, nonderived forms were found to have significantly shorter burst durations than plain and derived forms collectively ($t = -5.172$, $p < 0.001$, df = 8), but no significant difference was found between plain and derived forms (Fig. 4). Plain and derived forms had burst durations which exceeded those of nonderived forms by an average of 16 ms. Of the three plain stops, only the segment /b/ occurs with prevoicing in Medʉmba and thus is the only plain stop for which acoustic evidence of closure duration is present. Duration of closure was therefore compared across Onset Type conditions for this consonant only. No significant difference in closure duration was found between plain and prenasalized forms, or between derived and nonderived prenasalized forms. Finally, comparing between the two prenasalized conditions, nasal duration was found to be longer in the nonderived condition than in the derived condition ($t = 2.663$, $p < 0.05$, df = 7), by an average of 24 ms (Fig. 5). No significant differences were found in vowel duration between plain and prenasalized conditions or between derived and nonderived conditions.

In sum, despite differences in segment duration within word onsets, nonderived and derived prenasalized forms had similar onset durations overall, and had much longer onset durations

[5] Dots in all graphs represent the mean and error bars the standard deviation.

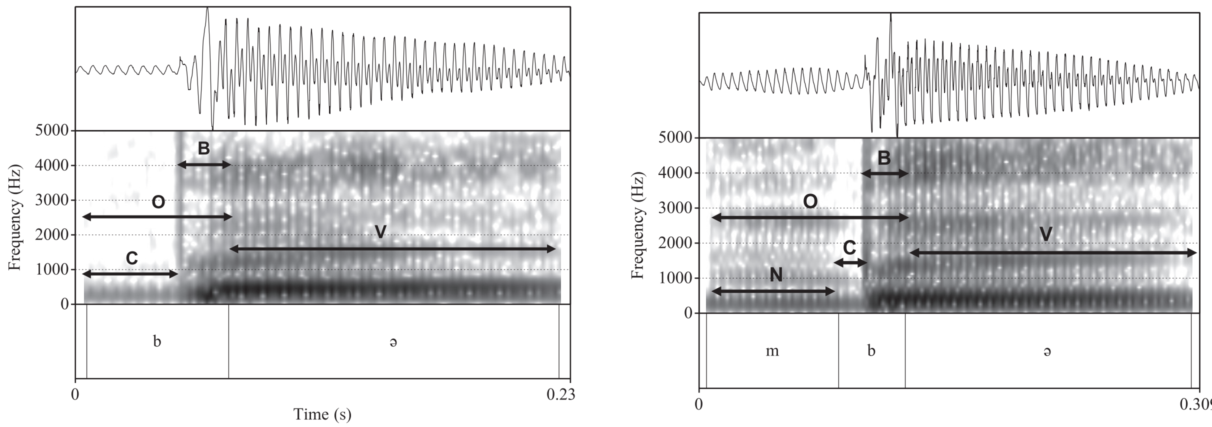

Fig. 1. Annotated intervals for the words *bə́* ('be') and *mbə́* (complementizer). N = 'nasal', C = 'closure', B = 'burst', O = 'onset', V = 'vowel'.

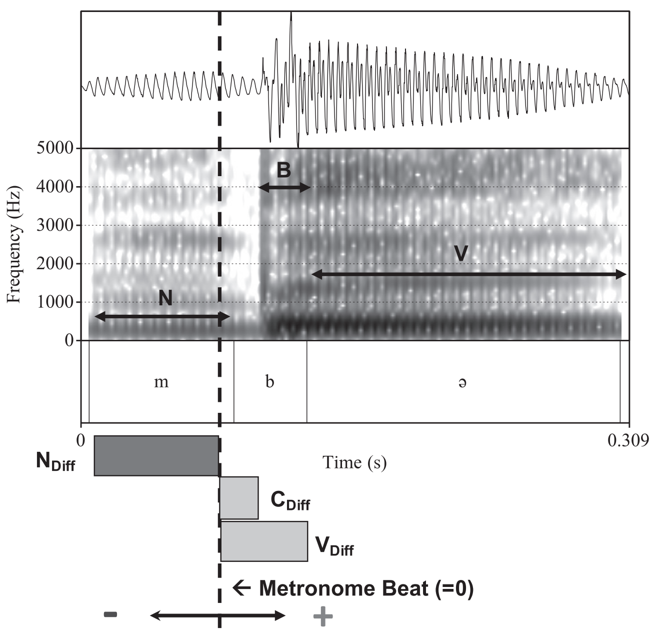

Fig. 2. Sample segment measurements for the word *mbə́*. Vertical dashed line represents the metronome beat. N_{Diff} represents the distance from the nasal onset to the beat, C_{Diff} the consonant/burst onset to the beat, and V_{Diff} the vowel onset to the beat.

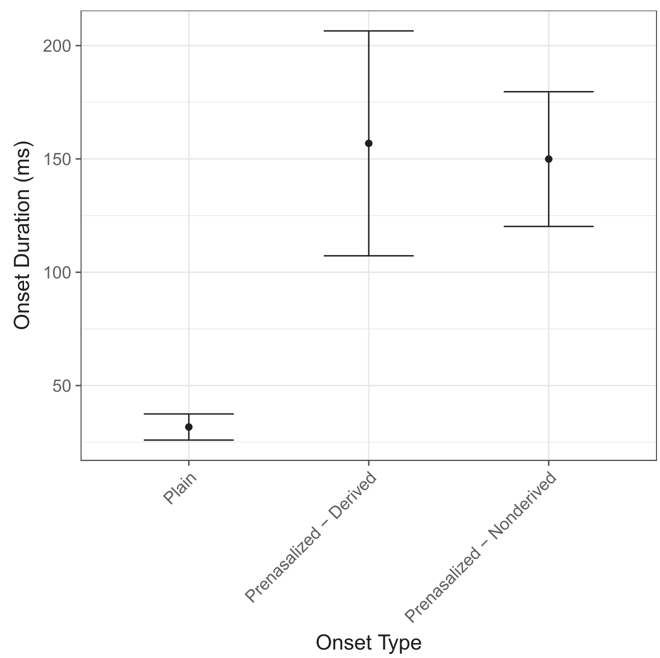

Fig. 3. Onset duration across onset types.

(even when closure duration was taken into account) than plain forms.

3.2. Effect of tone and prenasalization on p-center timing

The second set of analyses were meant to examine the overall effect of tone and prenasalization on the perceptual center of words. The initial model was applied to each of two dependent variables, Consonant Release, which measured the distance between stop consonant release and metronome beats, and Vowel Onset, which measured the distance between vowel onsets and metronome beats. This model included fixed effects for three categorical variables, Tone (two levels: high vs. low), Prenasalization (two levels: prenasalized vs. plain), and Place (three levels: labial, coronal, or dorsal), and one continuous variable, Burst Duration. The final variable Burst Duration was included as a measure of onset consonant duration because, as mentioned in Section 1, onset consonant duration has been found to have a large effect on the perceptual center of a syllable, and burst duration was a measure applicable for both plain and prenasalized conditions. Several three- and two-way interactions between variables were also examined; these are listed in Table 2.

The model included by-subject random slopes for Prenasalization, Tone, Place, and Burst Duration. Burst Duration was mean-centered to ameliorate the effects of collinearity, and categorical variables were contrast-coded to compare differences which appeared significant from visual inspection of data. Model selection proceeded with likelihood ratio tests, starting with the maximal model, and using stepwise backward elimination of each term until the optimal model was reached. Kappa values for the resulting models were all <4, indicating that collinearity between predictors was low and unlikely to affect model results (Baayen, 2008). Only significant results and other results of interest will be detailed below.

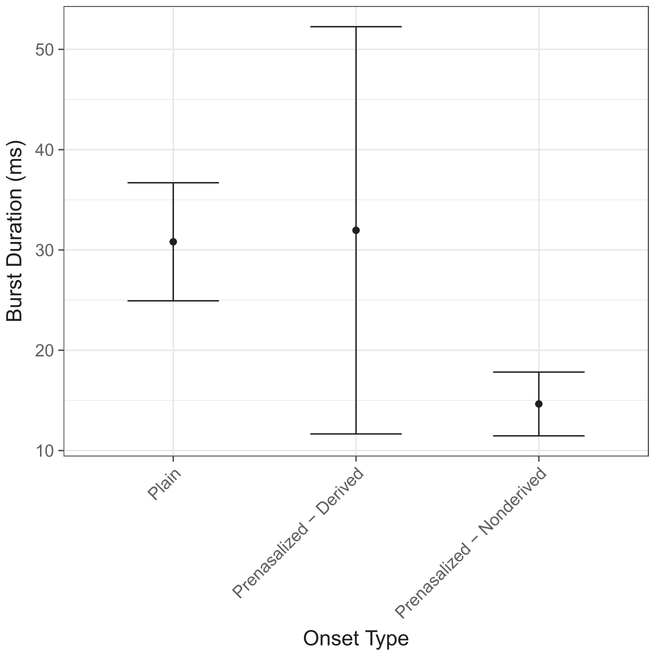

Fig. 4. Burst duration across onset types.

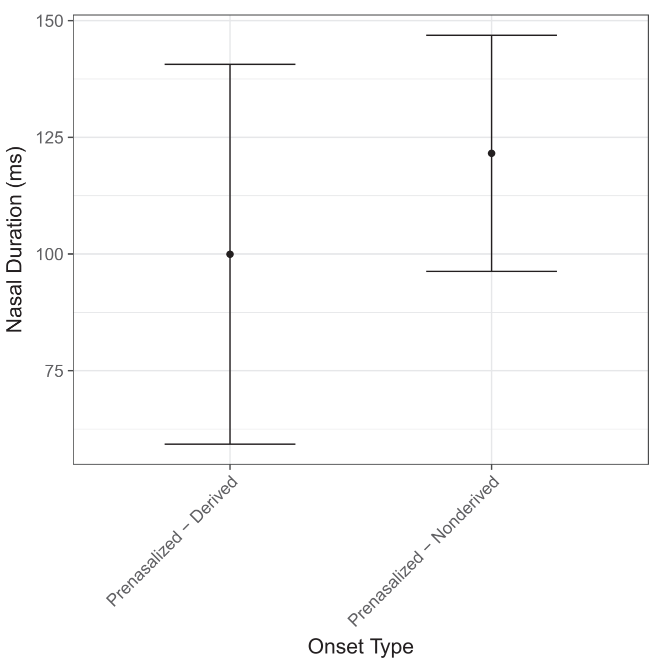

Fig. 5. Nasal duration across onset types.

Table 2
Interactions included in Model 2.

Three-way interactions	Two-way interactions
Tone × Prenasalization × Place	Tone × Prenasalization
Tone × Prenasalization × Burst Dur	Prenasalization × Place
	Tone × Place
	Prenasalization × Burst Dur
	Tone × Burst Dur

As expected, Burst Duration significantly impacted both consonant timing ($t = -3.438$, $p < 0.05$, df = 9) and vowel timing ($t = 3.072$, $p < 0.05$, df = 8). Similar to previous findings from Fowler and Tassinary (1981), where onset consonant duration was longer, consonant releases occurred earlier relative to the metronome beat, and vowel onsets occurred later relative to the beat.

Model results revealed a significant effect of Prenasalization on consonant timing ($t = -6.873$, $p < 0.001$, df = 9), such that consonant releases occurred earlier with respect to the beat in the plain condition (Fig. 6). No significant effect of Prenasalization was found for vowel timing; however, we note that vowels also occurred slightly earlier with respect to the beat in the plain condition than in the prenasalized condition. While no effect of Tone was found, there was a significant interaction between Prenasalization and Tone for both consonant ($t = -4.606$, $p < 0.001$, df = 1746) and vowel timing ($t = -4.032$, $p < 0.01$, df = 1752). As shown in Fig. 7, in plain forms, both consonants and vowels were initiated earlier for low tone words as compared with high tone words; this effect of tone disappeared or was even slightly reversed in prenasalized forms.

Franich (2016) found that vowel duration is an important cue in the perception of tonal contrasts in Medʉmba, with longer vowel duration found to bias listeners to hear syllables as low-toned. To investigate whether the interaction found between tone and prenasalization could be related to vowel duration differences across prenasalization conditions, we constructed one additional model investigating Vowel Duration as a dependent variable, including Tone and Prenasaliation (as well as their interaction) as fixed effects, each with corresponding by-subject random slopes. No significant effect of either Tone or Prenasalization was found, nor was there a significant interaction between these two variables.

In sum, prenasalization led to later timing of consonants and slightly later timing of vowels with respect to the metronome. In plain forms, low tone syllables had earlier consonant and vowel timing with respect to the metronome beat than high tone syllables, but this difference did not extend to prenasalized forms. The interaction between Tone and Prenasalization does not appear to be attributable to vowel duration differences across conditions.

3.3. Effect of morphophonology on p-center timing

The third set of analyses were meant to examine the effect of morphological status of prenasalized stops—derived vs. nonderived—on timing of nasals, consonants, and vowels. As will be discussed later in this section and in Section 4, though we refer to this variable in terms of morphology, the observed effects may be better explained in terms of underlying phonological structure which varies depending on morphological status of the prenasalized sequence.

Three dependent variables were utilized: Nasal Onset (distance from the onset of nasalization to the metronome beat), Consonant Release, and Vowel Onset. The primary models included fixed effects for two categorical variables, Morphology (two levels: derived vs. nonderived) and Place (three levels: labial, coronal, and dorsal), and one continuous variable, Onset Duration (which included the nasal, closure, and burst portions of onsets). Onset Duration was mean-centered to minimize the effects of collinearity, and categorical variables were sum-coded. The model also included by-subject random slopes for

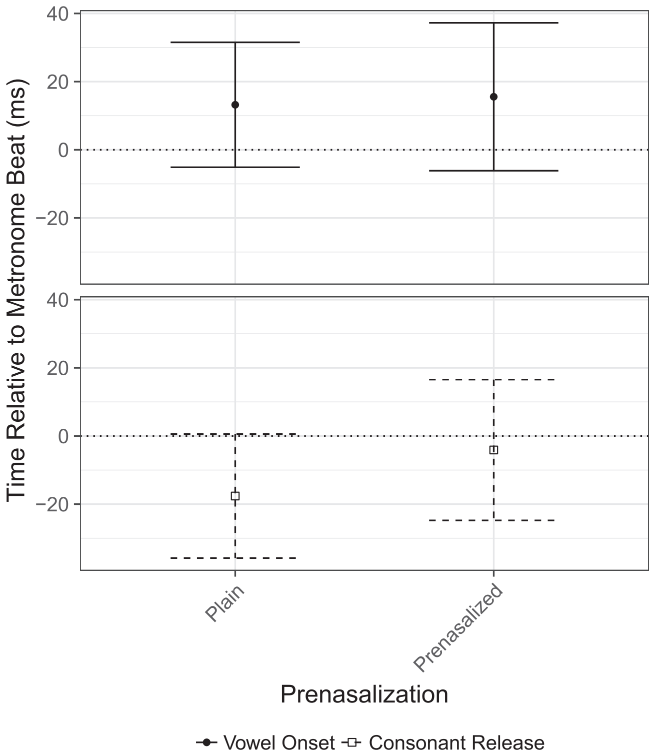

Fig. 6. Effect of prenasalization.

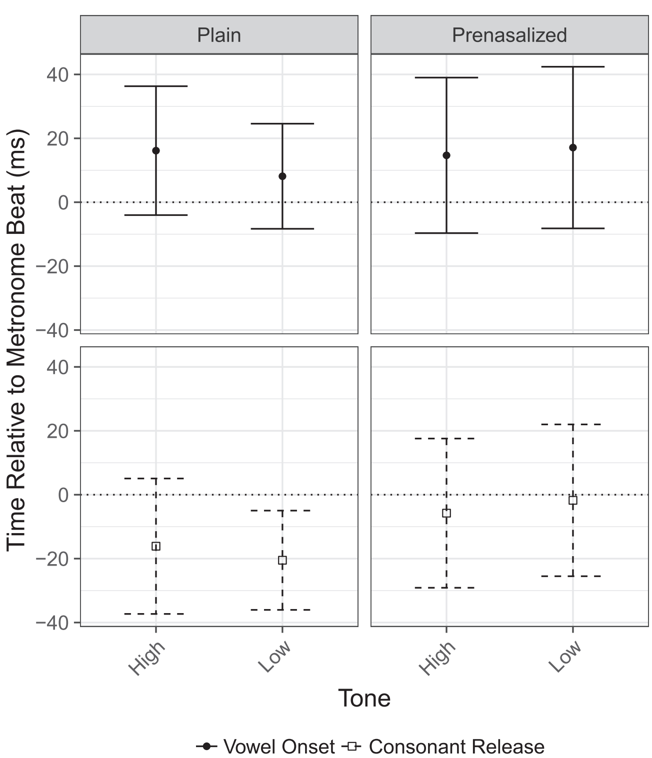

Fig. 7. Effect of tone by prenasalization.

Morphology, Place, and Onset Duration. Three- and two-way interactions are listed in Table 3.

Model selection proceeded with likelihood ratio tests, starting with the maximal model, and using stepwise backward elimination of each term until the optimal model was reached. Kappa values for the resulting models were all <4, indicating that collinearity between predictors was low and unlikely to affect model results.

Inter-quartile range in timing of all segments was found to be greater for derived forms (vowels: 53 ms; consonants: 70 ms; nasals: 100 ms) than for nonderived forms (vowels: 46 ms; consonants: 42 ms; nasals: 71 ms), indicating greater variability in timing in derived forms. A significant effect of Onset Duration was found for both vowel timing ($t = 3.313$, $p < 0.05$, df = 10) and nasal timing ($t = -13.484$, $p < 0.001$, df = 8), indicating vowels were timed later with respect to the metronome beat and nasals earlier with respect to the beat as onset duration increased. No significant effect of onset duration was found for consonant timing. Model results also revealed a significant effect of morphology for consonant timing ($t = 3.993$, $p < 0.01$, df = 8), indicating that consonants were timed later and closer to the metronome beat in the nonderived condition than in the derived condition (Fig. 8).

While no effect of Morphology was found for either nasal or vowel timing, a significant interaction was found between Morphology and Place for both measures. A comparison of the effects of morphology on nasal timing across Place conditions revealed a significant difference in patterning of coronal and labial consonants compared with dorsal consonants ($t = -4.112$, $p < 0.001$, df = 651), but no significant difference between coronal and labial consonants. As can be seen in Fig. 9, similar to the pattern found for consonant timing, nasals were timed earlier with respect to the metronome beat in the derived condition vs. the nonderived condition for labials and coronals. Though nasals also appear to have been timed slightly earlier for dorsal consonants in derived forms, the difference was not as great as for other consonants. As with nasal onsets, timing of vowel onsets was not found to differ significantly between labial and coronal conditions, but these two conditions showed a significant difference in patterning from dorsal consonants ($t = -5.425$, $p < 0.001$, df = 651). Once again, Fig. 9 shows slightly earlier timing of vowel onsets in the derived condition than the nonderived condition for both labial and coronal consonants, but no such difference for dorsal consonants. As discussed in Section 1.1, /k/ is always voiced after a nasal in nonderived (but not derived) forms, which leads the stop release burst in such environments to be quite short. Dorsal stops in derived forms also had a longer mean burst duration ($\mu = 38$ ms) than coronal ($\mu = 23$ ms) or labial stops ($\mu = 16$ ms). It is likely due to these facts that the overall pattern of earlier timing in derived forms versus nonderived forms did not obtain for vowels following dorsal consonants.

In sum, findings overall indicate earlier timing of segments with respect to the metronome beat in derived forms than in nonderived forms, though dorsal consonants showed some deviation from this overall pattern, particularly with respect to the timing of vowel onsets.

3.3.1. Similarity between plain, derived, and nonderived forms

At this juncture, we compare results from Section 3.2, where p-centers were found overall to occur farther from vowel onsets in prenasalized words than in plain words, and Section 3.3, where p-centers were found to occur overall farther from vowel onsets in nonderived as opposed to derived

Table 3
Interactions included in Model 3.

Three-way interactions	Two-way interactions
Morphology × Onset Duration × Place	Morphology × Onset Dration Morphology × Place Onset Duration × Place

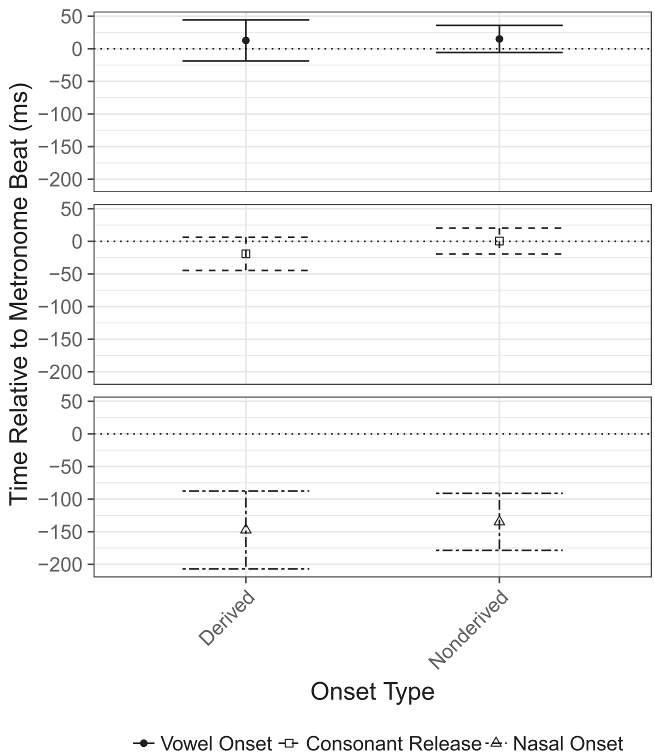

Fig. 8. Effect of morphology.

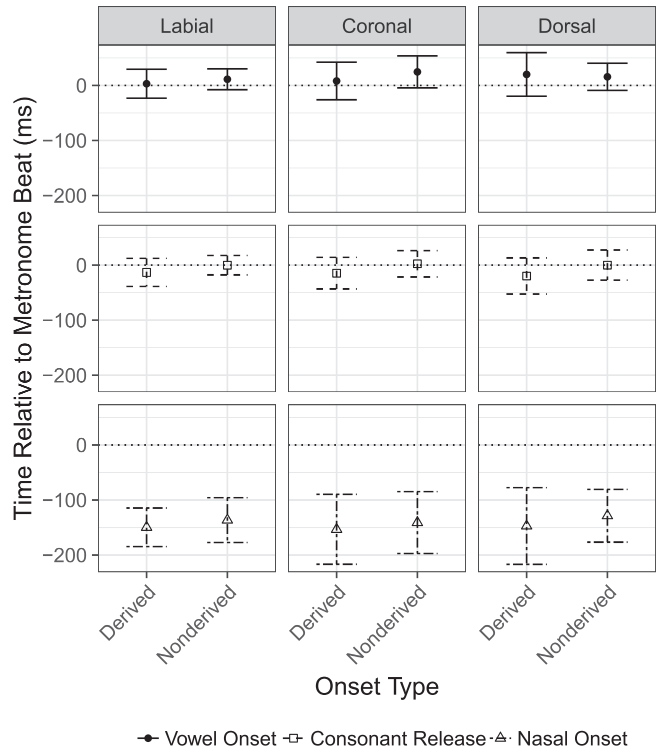

Fig. 9. Effect of morphology by consonant place.

prenasalized words. This latter finding, in particular, suggests that derived words may be behaving more similarly to plain words in terms of consonant and vowel timing, perhaps even indicating a separate timing unit associated with the nasal prefix for derived forms. If this is the case, we would expect minimal differences in p-center timing between plain and derived forms, and a larger difference in p-center timing between these two forms and nonderived prenasalized forms. To test this hypothesis, we conducted one final set of analyses. Our model included categorical fixed effects of Onset Type (3 levels: plain, prenasalized - derived, and prenasalized – nonderived) and Place (3 levels: labial, coronal, and dorsal), as well as a continuous fixed effect of Burst Duration. Two-way interactions between Onset Type and Place as well as between Onset Type and Burst Duration were included. Onset Type was contrast-coded to investigate differences between plain and derived forms on the one hand, and plain and derived forms against nonderived forms, on the other; Place was sum-coded, and Burst Duration was mean-centered.

As predicted according to the hypothesis that nasal prefixes constitute separate timing units for p-center calculation, results indicated no significant difference in consonant or vowel timing between plain and derived forms, but a significant difference between plain and derived forms collectively when compared with nonderived forms (consonants: $t = 6.518$, $p < 0.001$, df = 12; vowels: $t = 3.768$, $p < 0.01$, df = 10) (Fig. 10).

4. Discussion

Results from the present study in large part align with those from previous work on perceptual centers: just as has been found previously, p-centers in basic CV syllables in Medʉmba reside close to vowel onsets, and move earlier within the syllable as onset duration increases. Thus, the Medʉmba data presented here provide further evidence for the crosslinguistic prevalence of the p-center effect. Overall, our results are consistent with an analysis of prenasalized stops in Medʉmba as complex entities. In contrast with findings from other languages where prenasalized stops are analyzed as unary segments (Maddieson, 1989; Welmers, 1978), prenasalized stops in Medʉmba are significantly longer in overall duration than plain stops, regardless of their morphological status. For the most part, prenasalization was found to lead to a shift in p-center location away from the vowel onset and towards the word onset, similar to the effect found from consonant clusters crosslinguistically. However, a closer look at differences between nonderived and derived prenasalized environments revealed that p-center location varied considerably across the two conditions. Not only were words initiated earlier with respect to the metronome beat in the derived condition than in the nonderived condition, but the location of p-centers in the derived condition was relatively closer to the vowel onset than the consonant release, whereas in the nonderived condition p-centers occurred relatively closer to the consonant release than the vowel onset. In comparing the two prenasalized conditions with the plain condition (Section 3.3.1), it was found that p-centers in derived forms pattern much more

ARTICLE IN PRESS

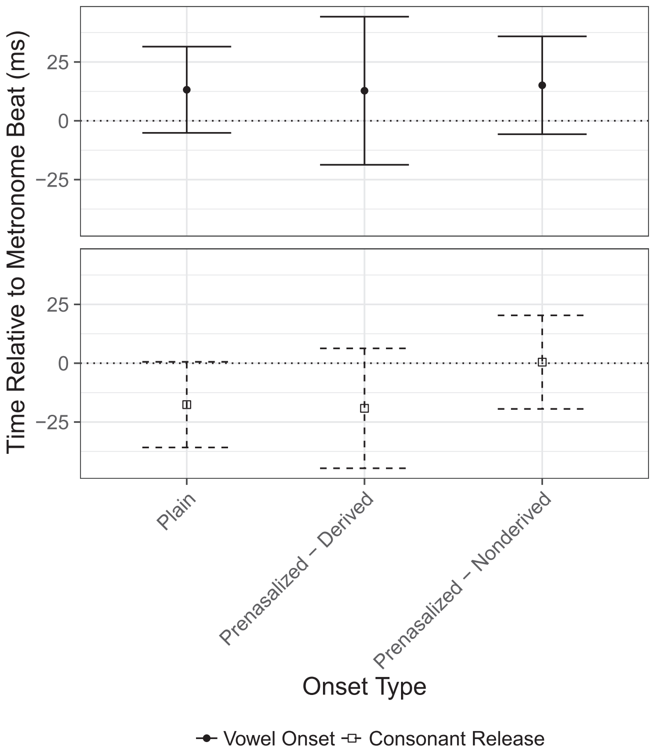

Fig. 10. Onset type comparison.

closely with those in plain forms. This suggests that the nasal prefix in derived forms is not contributing to p-center timing at all, likely indicating that these prefixes constitute a separate timing unit from the rest of the word. One possibility is that nasal prefixes are themselves syllabic, perhaps bearing their own separate p-center from that of the CV portion of target words. This outcome could account for the higher variability of p-centers in derived forms, assuming p-centers reflect the greater variability of gestural timing across a syllable boundary. That nasal prefixes are syllabic would be somewhat surprising, however, as speaker intuitions do not indicate a syllabic status for nasal prefixes, nor do other phonological factors such as tone bearing ability support a syllabic analysis. One other possibility is that nasal prefixes are simply 'unaffiliated' segments forming a simplex onset with the following stop, leaving them free to syllabify as codas to a preceding syllable (Goldstein, Chitoran, & Selkirk, 2007). From an articulatory standpoint, this could mean that the nasal gesture is not coupled with the following consonant and vowel gestures in any way, unlike a typical complex cluster, in which both consonants of the cluster are competitively coupled to the vowel and coupled anti-phase to one another (Fig. 11). Further articulatory data would be useful in evaluating these possibilities.

An interaction between tone and prenasalization indicated that p-centers were timed later in low falling tone words than high tone words, but that this effect disappeared (or was slightly reversed) for prenasalized words. Given there was no significant difference in vowel duration between prenasalized and plain forms, the demonstrated effect of tone does not seem to be tied to differences in vowel duration across forms with different onset types. According to the Target Approximation Model proposed by Xu and Liu (2006), contour tones differ from level tones in that the articulatory goal of the latter is associated with a static pitch target, whereas that of the former is associated with the pitch movement of the contour. One possibility is that p-centers are aligned with the articulatory goals of different tones, and that prenasalization leads to earlier initiation of the falling pitch movement in low tones. This would mean that the falling pitch movement would occur latest in plain low words, accounting for the later occurrence of p-centers in these words. However, examination of the timing of pitch movements between plain and prenasalized words does not align with predictions made from such a hypothesis: timing measures of f_0 peak indicate that contour pitch movements initiate quite close to vowel onsets in both plain and prenasalized forms.

Closer inspection of f_0 data from different forms reveals several patterns: (1) prenasal onsets condition pitch depression (as has been reported for other Bantu languages), (2) prenasalized words bearing low tones have much more restricted pitch ranges than do their plain counterparts, and (3) plain low tone words have a steeper f_0 slope than prenasalized low words (Fig. 12). Thus, another possible explanation for why plain low tone words displayed later p-centers overall could be related to the very dynamic nature of their f_0 profiles. It is known that contour tones are perceived as longer in duration than level tones (Yu, 2010). Perhaps listeners timed p-centers later where they perceived vowel duration to be longer as a result of increased pitch slope (even though vowel duration, objectively, was not longer for these forms). Indeed, these findings mirror those of Cooper et al. (1988) who found that increasing vowel duration could lead to perception of p-centers as occurring later in words.

The current study is thus the first to show a direct connection between tone and p-center location. More broadly, this finding suggests that tone can impact rhythm perception in language, in that it influences the point in a syllable around which perception of isochrony is likely based. It is interesting, in light of these findings, that Chow et al. (2015) found no effects of tone on p-center location in Cantonese, especially given that tonal slope has been found to influence vowel duration perception among speakers of Cantonese (Yu, 2010). However, a primary finding of Chow et al. (2015) was that Cantonese speakers rely on onset consonantal landmarks for p-center location more than vocalic landmarks, which is reflected in generally earlier p-centers for Cantonese syllables. It makes sense, then, that tonal information should be less influential in p-center location, as tones in Cantonese are aligned with vowels. The authors apparently did not investigate interactions between tone and onset type in Cantonese p-center location: one interesting question concerns whether the VC syllables examined in Chow et al.'s study showed tonal effects, being that p-centers were closely aligned with vowel onsets in these syllables according to their results.

Findings from the current study also highlight interesting connections between articulatory and perceptual properties of speech. Specifically, our finding that nasal prefixes in Medʉmba do not influence p-center location is reminiscent of recent work in Articulatory Phonology showing that onset clusters across languages can vary in their coordinative properties. While articulatory data have shown robust effects of cluster status on C-center effects, demonstrating, for example, that onset consonant sequences in Taslhiyt Berber and Moroccan

Please cite this article in press as: Franich, K. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* (2017), https://doi.org/10.1016/j.wocn.2017.11.001

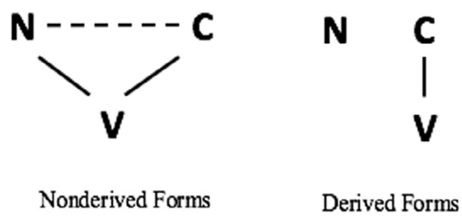

Fig. 11. Coupling relations in nonderived and derived prenasalized forms. Solid lines indicate in-phase coupling, while dashed lines indicate anti-phase coupling. Absence of a line between segments indicates there is no coupling relation between them.

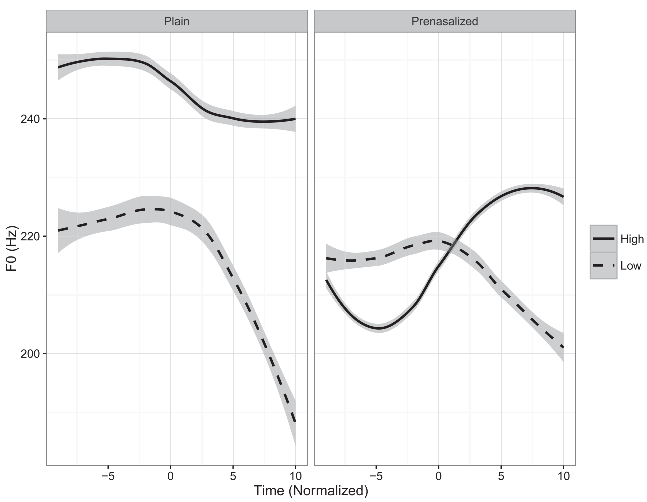

Fig. 12. Pitch tracks of Plain and Prenasalized forms by Tone (Time = 0 represents vowel onset).

Arabic behave as heterosyllabic (Goldstein et al., 2007; Hermes, Ridouane, Mucke, & Grice, 2011; Shaw, Gafos, Hoole, & Zeroual, 2009), ours is the first comparable finding with regard to p-center effects. Though additional articulatory data from Medʉmba would be needed to confirm whether the observed p-center effect does, indeed, reflect similar structural differences at the articulatory level, our finding could have important ramifications for fieldworkers trying to evaluate syllable structure properties in underdocumented languages, especially where conditions preclude the use of equipment for articulatory data analysis.

As far as more practical applications, our research may have important implications for ongoing attempts to simulate human-to-human conversational interaction with machines. As previously mentioned, perceived isochrony has been found to aid in word segmentation and lexical access (Brown et al., 2011; Dilley & McAuley, 2008; Dilley et al., 2010, Dilley, Wallace, & Heffner, 2012). Recent work by Kawasaki, Yamada, Ushike, Miyauchi, and Yamaguchi (2013) also suggests that neuronal entrainment during conversation is enhanced when rhythmic speech coordination between conversation partners is highest, and that this entrainment contributes to more fluent conversation between humans as opposed to between a human and a machine. Being able to more accurately simulate rhythmic coordination in speech between humans is thus a key to improving human-machine interaction, an increasingly important form of communication. Accurate estimation of p-centers for the modeling of synthetic speech rhythms provides an important step towards this goal.

4.1. Limitations

One possible limitation of the present study is the inclusion of a relatively wide age range of speakers, as older speakers might be expected to exhibit greater motor control deficits which may have affected timing of words with respect to the metronome. Ideally, a production experiment such as the present one which relies heavily on timed speech motor responses would exclude subjects who might exhibit such deficits. We note that additional research on the effect of age on p-center alignment (or on speech timing, more generally) is necessary, and leave that for future work. Furthermore, as mentioned, the production method used may not be the best way of measuring the exact location of p-centers, and rather gives a relative indication of p-center location across syllable types. In order to more precisely estimate p-center location in the forms tested, a perception study using something like the rhythm adjustment method or the phase correction response method would be optimal.

5. Conclusion

This study has shown that the p-center phenomenon—defined broadly as a lack of alignment between a syllable's acoustic onset and its perceived moment of occurrence—is present in the Medʉmba language, providing further crosslinguistic support for the phenomenon. As found for many other languages, p-centers in Medʉmba occur within the transition between onset consonant and vowel, following the release of the consonant and prior to the onset of vowel formants. P-centers shift farther from vowel onsets as syllable onset duration increases. In contrast with previous findings from Cantonese, we found that lexical tone had an effect on p-center location: CV words bearing low falling tones had later p-centers than those bearing high tones. We explain these effects in terms of the psychoacoustic effect of dynamic tones on perceived vowel duration. We also provide evidence that the morphological status of nasal segments in Medʉmba influences their ability to affect p-center location, with nasal prefixes apparently treated as separate timing units from the CV (C) portion of the word they concatenate with. This may be indicative of the status of such prefixes as separate syllables, or simply unaffiliated segments which must syllabify as the coda of a preceding syllable. Our findings have implications both for speech synthesis and for the study of syllable structure in language more broadly.

Acknowledgments

This work was supported by National Science Foundation Linguistics Program Grant No. BCS-1423865 (co-PIs: Kathryn Franich and Alan C.L. Yu). The National Science Foundation does not necessarily endorse the ideas and claims in this paper. I thank the Medʉmba speakers who participated in the study for generously sharing their time and linguistic knowledge. I am deeply grateful to Dr. Ange Bergson Lendja for helping me to develop stimuli and recruit subjects. Thanks to Josh Falk, Jacob Phillips, Betsy Pillion, Anisia Popescu, and Alan Yu for helpful feedback. Thanks also to Fred Cummins and two anonymous reviewers for very insightful comments and suggestions. All remaining errors are of course my own.

ARTICLE IN PRESS

Appendix A

Fig. a.

		Plain		Prenasalized			
				Derived		Nonderived	
Tone	High	bə	'be'	m-bib	'TNS-wait'	mba	'bicycle'
		bu	'3rd person plural'	n-to	'PL-tamtam'	mba	'pot'
		ta	'haggle'	n-tu	'TNS-ask'	mbə	'clay'
		tə	'drop off/deposit'	ŋ-ka	'PL-plate'	ŋga	'no'
		to	'neck/hole'	ŋ-ko	'PL-mountain'		
		to	'tamtam'				
		ka	'plate'				
		kə	'what'				
		ki	question particle				
		ko	'mountain'				
		ku	'peel'				
	Low	bà	'ripe'	ŋ-kà	'PL-line'	mbà	'palm kernel'
		bə̀	'fall'	ŋ-kùʔ	'PL-traditional chair'	mbə̀	'complementizer'
		tò	'subtly steal a handful of			ntù	'traditional religion'
			(e.g.) unharvested peanuts'			ntà	'fanfare'
		tà	'difficult'			ŋgà	'root'
		kà	'Maggi (seasoning)'			ŋgò	'country'
		kò	'like/love'				
		kùʔ	'traditional chair'				

Fig. a. Complete list of words elicited.

Appendix B

Fig. b.

Subject 1	ntu 'TNS-ask'; mbà 'palm kernel'; kà 'Maggi'; nto 'PL-tamtam'
Subject 2	ŋka 'PL-plate'; ŋko 'PL-mountain'
Subject 3	ŋko 'PL-mountain'
Subject 4	ntu 'TNS-ask'; mbib 'TNS-wait'
Subject 5	mbib 'TNS-wait'
Subject 6	mbib 'TNS-wait'
Subject 7	bə̀ 'fall'; kà 'Maggi', ŋko 'PL-mountain'; nto 'PL-tamtam'
Subject 8	ta 'haggle'; ŋkà 'PL-line; kà 'Maggi'; ŋka 'PL-plates'; nto 'PL-tamtam'
Subject 9	ntu 'TNS-ask'; ŋka 'PL-plates'; mbib 'TNS-wait; ŋko 'PL-mountain'; nto 'PL-tamtam'
Subject 10	mbà 'palm kernel'; ŋko 'PL-mountain'; nto 'PL-tamtam'

Fig. b. List of missing trials by subject.

References

Abercrombie, D. (1965). *Studies in phonetics and linguistics*. Oxford, UK: Oxford University Press.

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine and Atherton.

Allen, G. (1972). The location of rhythmic stress beats in English: An experimental study. *Language and Speech, 15*, 72–100.

Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica, 66*, 46–63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*, 351–373.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Barbosa, P. A., Meireles, A. R., & Vieira, J. M. (2005). Abstractness in speech-metronome synchronization: P-centers as cycle attractors. *Proceedings of Interspeech*, 1441–1444.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26*, 86–125.

Bolinger, D. L. (1965). *Forms of English: Accent, morpheme, order*. Cambridge, MA: Harvard University Press.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Plumetiaz, 45*(140–155).

Browman, C., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18*, 299–320.

Browman, C. P., & Goldstein, L. (1998). Modeling syllable structure as constraints on gestural coordination. In *Presentation at NSF Workshop on syllable structure and gesture timing*. Ohio State University.

Browman, C., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée, 5*(25–34).

Brown, M., Salverda, P., Dilley, L., & Tanenhaus, M. (2011). Distal prosody influences lexical intepretation in on-line sentence processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Austin*. .

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica, 58*, 129–162.

Chow, I., Belyk, M., Tran, V., & Brown, S. (2015). Syllable synchronization and the P-center in Cantonese. *Journal of Phonetics, 49*, 55–66.

Cibelli, E. (2015). The phonetic basis of a phonological pattern: Depressor effects of prenasalized consonants. In J. Romero & M. Riera (Eds.), *The phonetics-phonology interface: Representations and methodologies* (pp. 171–192). Amsterdam: John Benjamins.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception and Psychophysics, 39*, 187–196.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics, 16*, 231–241.

Dauer, R. M. (1983). Stress-timing and syllable-timing re-analysed. *Journal of Phonetics, 11*, 51–62.

Dauer, R. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the XIth International Congress of Phonetic Sciences*, 447–450.

de Jong, K. (1994). The correlation of P-center adjustments with articulatory and acoustic events. *Perception & Psychophysics*, 15.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching, IV*(3), 183–198.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltac. In P. Karnowski & I. Szigeti (Eds.), *Language and language-processing: Proceedings of teh 38th Linguistics Colloquium* (pp. 231–241). Frankfurt am Main: Peter Lang.

Please cite this article in press as: Franich, K. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language. *Journal of Phonetics* (2017), https://doi.org/10.1016/j.wocn.2017.11.001

Dilley, L., Mattys, S., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*, 274–294.

Dilley, L., & McAuley, D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*, 294–311.

Dilley, L., Wallace, J., & Heffner, C. (2012). Perceptual isochrony and fluency in speech by normal talkers under varying task demands. In O. Niebuhr & H. Pfitzinger (Eds.), *Prosodies: Context, function, and communication Language, Context, and Cognition series* (pp. 237–258). Berlin/New York: Walter de Gruyter.

Duanmu, S. (1994). Against contour tone units. *Linguistic Inquiry, 25*, 555–608.

Fowler, C. A. (1979). Perceptual centers in speech production and perception. *Perception and Psychophysics, 25*, 375–388.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General, 112*, 386–412.

Fowler, C. A., & Tassinary, L. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. IX). New Jersey: Erlbaum.

Franich, K. (2016). Internal and contextual cues to tone perception in Medʉmba. *Journal of the Acoustical Society of America Express Letters, 140*(1), 107–112.

Gao, M. (2008). Gestural coordination among vowel, consonant and tone gestures in Mandarin Chinese. In *Presentation at the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers*. .

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrucken, Germany* (pp. 241–244).

Gordon, M. (2001). A typology of contour tone restrictions. *Studies in Language, 25*, 405–444.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Papers in Laboratory Phonology* (Vol. 7). Berlin: Mouton.

Gussenhoven, C., & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. In *Proceedings of interspeech* (pp. 1365–1369). Lyon, France.

Harsin, C. A. (1997). Perceptual-center modeling is affected by including acoustic rate-of-change modulations. *Perception & Psychophysics, 59*(2), 243–251.

Hermes, A., Ridouane, R., Mucke, D., & Grice, M. (2011). Kinematics of syllable structure in Tashlhiyt Berber: The case of vocalic and consonantal nuclei. In *9th International Seminar on Speech Production* (pp. 1–6). Montréal, Canada.

Hoequist, C. E. (1983). The perceptual center and rhythm categories. *Language and Speech, 26*, 375–388.

Howell, P. (1984). An acoustic determinant of perceived and produced anisochrony. In M. P. R. van den Broecke & A. Cohen (Eds.), *Proceedings of the 10th International Congress of Phonetic Sciences* (pp. 429–433). Dordrecht: Foris.

Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope. *Perception & Psychophysics, 43*, 90–93.

Hyman, L. M. (2008). Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics, 46*(2).

Hyman, L. M., & Ngunga, A. (1997). Two kinds of moraic nasals in Ciyao. *Studies in African Linguistics, 26*, 131–163.

Kawasaki, M., Yamada, Y., Ushike, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports, 3*. 1692 EP.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253–263.

Maddieson, I. (1989). Prenasalized stops and speech timing. *Journal of the International Phonetic Association, 19*(2), 57–66.

Marcus, S. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics, 30*, 247–256.

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

Patel, A. D., Naito, W., & Löfqvist, A. (1999). The acoustics and kinematics of regularly timed speech: A database and method for the study of the P-center problem. *Proceedings of the 14th International Congress of Phonetic Sciences*, 405–408.

Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics, 17*, 175–192.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm. *Cognition, 73*, 265–292.

Rapp, K. (1971). A study of syllable timing. *Papers from the Institute of Linguistics, 8*, 14–19.

Riehl, A. K. (2008). *The phonology and phonetics of nasal obstruent sequences* [dissertation]. Cornell University.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Arnold.

Scott, S. (1998). The point of P-centres. *Psychological Research, 61*, 4–11.

Scott, S., & Howell, P. (1992). Perceptual centers in speech: An acoustic analysis. *Journal of the Acoustical Society of America, 94*(4), 2443.

Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. (2009). Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability in articulation. *Phonology, 26*, 187–215.

Shen, Y., & Peterson, G. G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers, 9*, 1–36.

Šturm, P., & Volín, J. (2016). P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment. *Journal of Phonetics, 55*, 38–52.

Tilsen, S. (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology*, 51–81.

Villing, R. C., Repp, B. H., Ward, T. E., & Timoney, J. M. (2011). Measuring perceptual centers using the phase correction response. *Attention, Perception & Psychophysics, 73*, 1614–1629.

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

Welmers, W. (1978). *African language structures*. Berkeley, CA: University of California Press.

Xu, Y. (2004). Understandign tone from the perspective of production and perception. *Language and Linguistics, 5*, 757–797.

Xu, Y. (2009). Timing and coordination in tone and intonation: An articulatory-functional perspective. *Lingua, 119*, 906–927.

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Rivista di Linguistica, 18*(1), 125–159.

Yip, M. (1991). *The tonal phonology of Chinese*. New York: Garland Publishing.

Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. *Laborabory Phonology, 10*, 151–168.

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution–Typological survey and formal analysis* [dissertation]. UCLA.