THE UNIVERSITY OF CHICAGO

THE INTERACTION OF PROMINENCE, RHYTHM, AND TONE IN MEDɄMBA

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE HUMANITIES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF LINGUISTICS

BY

KATHRYN HANNON FRANICH

CHICAGO, ILLINOIS

DECEMBER 2017

Copyright © 2017 by Kathryn Hannon Franich

All rights reserved

To J.P.S.J. and D.J.S.

CONTENTS

LIST OF FIGURES vii

LIST OF TABLES xi

ABSTRACT xiii

1 INTRODUCTION 1
1.1 Definitions of Key Terms 2
1.2 Assessing Prominence 4

2 MEDƟMBA: HISTORICAL CONTEXT & THEORETICAL SIGNIFICANCE 6

3 PROMINENCE AND RHYTHM IN LANGUAGE: THE VIEW(S) FROM PHONOLOGY AND PHONETICS 14
3.1 Rhythm and Prominence in Phonological Theory 14
3.1.1 Early Theories of Prominence and Metrical Structure 14
3.1.2 Studies of Tone and Accent and the 'Stress to Tone Continuum' 16
3.1.3 Alternatives to Traditional Metrical Theory 19
3.1.4 Prominence at Higher Levels of Prosodic Structure 21
3.1.5 Prominence Asymmetries in Bantu: Non-Tonal Evidence 23
3.2 Quantitative Approaches to Rhythm 26
3.2.1 Isochronic and Non-Isochronic Approaches 26
3.2.2 Discrete Patterning in Time: the Dynamical View 30

4 THE SPEECH CYCLING TASK: THEORETICAL UNDERPINNINGS AND APPLICATION 32
4.1 Parallels with Other Motor Behaviors 34
4.1.1 Synergistic Behavior in a Reaching Task 34
4.1.2 Coupling and Phase Shifts in a Finger Wagging Task 35
4.2 Dynamical Patterns of Gestures and Syllables: Articulatory Phonology 37
4.3 The Prosodic Hierarchy as a System of Coupled Oscillators 38
4.4 Evaluating Stable Modes of Coordination: Gaussian Mixture Models 43
4.4.1 Gaussian Mixtures: Basic Principles 45
4.4.2 Choice of Components: Parametric Bootstrapping of Likelihood Ratio Statistics 46

5 EXPERIMENTS 1 AND 2: THE RELATIONSHIP BETWEEN STEM POSITION AND PROMINENCE 47
5.1 Evidence for Stem-Initial Prominence in Medʉmba 48
5.2 Experiment 1: The Relative Prominence of Stem-Initial versus Stem-Final Syllables 53
5.2.1 Stimuli 53
5.2.2 Participants, Procedure, and Data Processing 54

5.2.3 Analysis Plan and Hypotheses . . . 56
5.2.4 Results . . . 59
5.2.5 Discussion . . . 72
5.3 Experiment 2: The Relative Prominence of Stem-Initial Syllables versus Prefixes and Enclitics . . . 73
5.3.1 Stimuli . . . 73
5.3.2 Participants, Procedure, and Data Processing . . . 74
5.3.3 Analysis Plan and Hypotheses . . . 76
5.3.4 Results . . . 77
5.3.5 Discussion . . . 89
5.4 General Discussion: Implications for Medʉmba Phonology . . . 90

6 EXPERIMENTS 3 AND 4: THE RELATIONSHIP BETWEEN TONE AND PROMINENCE . . . 94
6.1 Experiment 3: The Influence of Tone on Relative Prominence of Pronominal Enclitics 95
6.1.1 Stimuli . . . 98
6.1.2 Participants, Procedure, and Data Processing . . . 99
6.1.3 Analysis Plan and Hypotheses . . . 99
6.1.4 Results . . . 100
6.1.5 Discussion . . . 112
6.2 Experiment 4: Evaluating Phrase-Level Prominence and the Role of Tone . . . 113
6.2.1 Stimuli . . . 116
6.2.2 Participants, Procedure, and Data Processing . . . 117
6.2.3 Analysis Plan and Hypotheses . . . 118
6.2.4 Results . . . 119
6.2.5 Discussion . . . 130
6.3 General Discussion: Relative Prominence and the Role of Tone . . . 130

7 PATTERNS OF INDIVIDUAL VARIATION AND RHYTHMIC SYNCOPATION IN THE SPEECH CYCLING TASK . . . 134
7.1 Individual Level Variation in the Task . . . 136
7.1.1 Demographics of Six Speakers . . . 138
7.1.2 Quantifying Individual Variation . . . 140
7.2 Explaining Patterns of Variation: Rate-Driven Shifts in Metronome Alignment . . . 144

8 COUPLING IN RELATION TO PHONOLOGY AND SOUND CHANGE: THE CASE OF TONE SHIFT AND REVERSAL . . . 150
8.1 Tone Shift and Tone Spread: A Role for Phase Alignment? . . . 150
8.1.1 Remaining Issues: Tone Spread vs. Tone Shift and Attraction to a ‘Weak’ Position . . . 162
8.2 Language and Music: A Shared Timekeeper? . . . 165

9 CONCLUSION . . . 171

REFERENCES 173

LIST OF FIGURES

2.1 Southern Bantoid languages, from Watters & Leroy (1989) 7

4.1 Common speech cycling patterns in musical notation (adapted from Cummins & Port 1998) 33
4.2 Illustration of finger wagging task (Kelso *et al.*, 1979) 36
4.3 Changes in stability of attractors with increasing movement frequency based on Haken *et al.* (1985) 37
4.4 Oscillators at foot (red) and syllable (blue) levels at 1:2 and 1:3 frequency locking ratios (adapted from Salzman et al. 2008) 40
4.5 Distribution of phase alignment patterns for *Dig for a duck* in speech cycling from Cummins & Port, 1998, as shown in Port 2003: 601 41
4.6 Comparison of external phase patterns with increasing speech rate: English (stress-timed) vs. Japanese (mora-timed) (from Tajima 1998: 35) 42
4.7 Parameters of a single Gaussian distribution (Left) and a Gaussian mixture (Right) with three components (corresponding to colored areas) 44

5.1 Medʉmba consonant inventory 48
5.2 Medʉmba vowel inventory 48
5.3 Consonant and vowel distributions by stem position and affix type 48
5.4 Sample measure of External Phase. External Phase for Syllable 2 is the interval of *b* divided by the interval of *a*. 56
5.5 Scatter plot of phase angle by metronome speed (speech rate), Experiment 1 (red=Syllable 2, green=Syllable 3, blue=Syllable 4). Each point represents a single observation for a given speaker. 60
5.6 Distribution by syllable with regression lines, Experiment 1 (red = Syllable 2, green=Syllable 3, blue=Syllable 4) 61
5.7 Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), SISI condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position 62
5.8 Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), SISF condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position 63
5.9 SISI Syllable 3: 2 Component Gaussian mixture (dotted red lines indicate SHP positions of .333 and .500) 65
5.10 SISF Syllable 3: 3 Component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667) 65
5.11 SISI Syllable 4: 3 Component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667) 66
5.12 SISF Syllable 4: 3 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667) 66

5.13 Patterns for Syllable 3 (circles) and 4 (squares), SISI Condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3. 68
5.14 Patterns for Syllable 3 (circles) and 4 (squares), SISF Condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3. 69
5.15 Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 1 70
5.16 Scatter plot of phase angle by metronome speed (speech rate), Experiment 1 (red=Syllable 2, green=Syllable 3, blue=Syllable 4) 77
5.17 Distribution by syllable with regression lines, Experiment 2 (red = Syllable 2, green=Syllable 3, blue=Syllable 4) 78
5.18 Heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), PreSI condition, Experiment 1. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position 79
5.19 Heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), SIEnc condition, Experiment 1. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position 80
5.20 PRESI Syllable 3: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .333 and .500) 81
5.21 SIENC Syllable 3: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .335 and .500) 82
5.22 PRESI Syllable 4: 3 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667) 82
5.23 SIENC Syllable 4: 3 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667) 83
5.24 Patterns for Syllable 3 (circles) and 4 (squares), PRESI condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3. 84
5.25 Patterns for Syllable 3 (circles) and 4 (squares), SIENC condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3. 86
5.26 Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 2 88

6.1 Scatter plot of phase angle by metronome speed (speech rate), Experiment 3 (red=Syllable 2, yellow=Syllable 3, green=Syllable 4, blue=Syllable 5, purple=Syllable 6). Each point represents a single observation for a given speaker. 101
6.2 Distributions by syllable with regression lines, Experiment 3 (red=Syllable 2, yellow=Syllable 3, green=Syllable 4, blue=Syllable 5, purple=Syllable 6) 102
6.3 Heat map of syllable alignment (Proportion Phase by Speed), Syllable 5 (left pane) and Syllable 6 (right pane), HIGHPRO condition, Experiment 4. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position 103

6.4 Heat map of syllable alignment (Proportion Phase by Speed), Syllable 5 (left pane) and Syllable 6 (right pane), LOWPRO condition, Experiment 4. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position . . . 104
6.5 HIGHPRO Syllable 5: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500 and .667) . . . 106
6.6 LOWPRO Syllable 5: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .500 and .667) . . . 106
6.7 HIGHPRO Syllable 6: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .500 and .667) . . . 107
6.8 LOWPRO Syllable 6: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .500 and .667) . . . 107
6.9 Patterns for Syllable 5 (triangles) and 6 (diamonds), HIGHPRO Condition. Red = Pattern 1, blue = Pattern 2, green = Pattern 3. . . . 108
6.10 Patterns for Syllable 5 (triangles) and 6 (diamonds), LOWPRO Condition. Red = Pattern 1, blue = Pattern 2.. . . . 110
6.11 Comparison of dominant components for Syllable 5 (left) and Syllable 6 (right), Experiment 3 . . . 111
6.12 Distributions by syllable with regression lines, Experiment 4 (red=Syllable 2, yellow=Syllable 3, green=Syllable 4, blue=Syllable 5, purple=Syllable 6) . . . 120
6.13 Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), HIGH/IP-INITIAL condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position . . . 121
6.14 Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), LOW/IP-MEDIAL condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position . . . 122
6.15 HIGH/IP-INITIAL Syllable 3: 2 component Gaussian mixture (Dotted blue lines indicate half SHP positions of .165 and .250) . . . 124
6.16 LOW/IP-MEDIAL Syllable 3: 2 component Gaussian mixture (Dotted blue lines indicate half SHP positions of .165 and .250) . . . 125
6.17 HIGH/IP-INITIAL Syllable 4: 2 component Gaussian mixture (Dotted blue lines indicate half SHP positions of .165 and .250) . . . 125
6.18 LOW/IP-MEDIAL Syllable 4: 2 component Gaussian mixture (Dotted blue line indicates half SHP Position of .25; dotted red line indicates SHP position of .330) . 126
6.19 Patterns for Syllable 3 (circles) and 4 (squares), HIGHPRO Condition. Red = Pattern 1, blue = Pattern 2. . . . 126
6.20 Patterns for Syllable 3 (circles) and 6 (squares), LOWPRO Condition. Red = Pattern 1, blue = Pattern 2, green = Pattern 3. . . . 127

6.21 Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 4. Black lines represent closest half SHP positions of .167 (left panel; Syllable 3) and .250 (right panel; Syllable 4). Red dotted lines indicate component means for the HIGH/IP-INITIAL condition, and blue dotted lines indicate component means for the LOW/IP-MEDIAL condition. 129

7.1 Sample measure of External Phase. External Phase for Syllable 2 is the interval of *b* divided by the interval of *a*. 134
7.2 Schematic representation of in-phase and anti-phase alignment between metronome beats and syllables in phrase repetitions 136
7.3 In-phase (left) and anti-phase (right) coupling of oscillators at the metronome and phrase levels 136
7.4 Scatter plot of phase angle by metronome speed (speech rate) by individual, Experiment 1 (red=Syllable 2, green=Syllable 3, blue=Syllable 4) 137
7.5 Comparison of Syllable 4, Experiment 1, PRESI Condition, Subjects s6 and s11 (Red dotted line = SHP position of .5) 138
7.6 Comparison of Syllable 4, Experiment 1, PRESI Condition, 6 subjects 141
7.7 Results of linear and nonlinear model comparison for 6 subjects, Syllable 4, PRESI condition 143
7.8 Alignment of Syllable 4 with respect to metronome beat, 6 subjects (Black dotted line indicates metronome beat) 145
7.9 Changes in stability of attractors with increasing movement frequency (Haken *et al.*, 1985). Movement frequency increases left to right, and top to bottom. 146

8.1 Autosegmental representations of anticipatory tone spread (left) and tone shift (right) 151
8.2 Autosegmental representations of anticipatory tone spread in Medʉmba 154
8.3 Standard bell rhythm (Agawu 2006: 1) 166

LIST OF TABLES

5.1 Subset of Medʉmba possessive pronouns and object pronouns (Voorhoeve 1967, adapted according to modern IPA conventions) . . . 51
5.2 Stimuli for Experiment 1. Syllables hypothesized prominent are in bold. . . . 54
5.3 Results of Anderson-Darling test for normality, Experiment 2 . . . 63
5.4 Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K. . . . 64
5.5 Model parameters for Syllable 3 and Syllable 4: Experiment 2 . . . 64
5.6 Component strengths, Syllables 3 and 4, SISI condition . . . 69
5.7 Component strengths, Syllables 3 and 4, SISF condition . . . 70
5.8 Results of linear mixed effects models Syllable 3 and Syllable 4, Experiment 1 . . . 71
5.9 Stimuli for Experiment 2. Syllables hypothesized prominent are in bold. Clitic boundaries indicated with '=' and affix boundaries with '-' . . . 74
5.10 Results of Anderson-Darling test for normality, Experiment 1 . . . 80
5.11 Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K. . . . 81
5.12 Model Parameters For Syllable 3 and Syllable 4: Experiment 1 . . . 81
5.13 Component strengths, Syllables 3 and 4, PRESI condition . . . 85
5.14 Component strengths, Syllables 3 and 4, SIENC condition . . . 87
5.15 Results of linear mixed effects models Syllable 3 and Syllable 4, Experiment 2 . . . 88

6.1 Stimuli for Experiment 3. Syllables hypothesized prominent are indicated in bold. Clitic boundaries indicated with '=' . . . 98
6.2 Results of Anderson-Darling test for normality, Experiment 3 . . . 103
6.3 Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K, Experiment 3. . . . 105
6.4 Model Parameters For Syllable 5 and Syllable 6: Experiment 3 . . . 105
6.5 Component strengths, Syllables 5 and 6, HIGHPRO condition . . . 109
6.6 Component strengths, Syllables 5 and 6, LOWPRO condition . . . 110
6.7 Results of linear mixed effects models Syllable 5 and Syllable 6, Experiment 3 . . . 111
6.8 Stimuli for Experiment 4. Syllables hypothesized prominent are in bold. Clitic boundaries indicated with '=' . . . 117
6.9 Results of Anderson-Darling test for normality, Experiment 4 . . . 122
6.10 Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K, Experiment 4. . . . 123
6.11 Model Parameters For Syllable 3 and Syllable 4: Experiment 4 . . . 123
6.12 Component strengths, Syllables 3 and 4, HIGH/IP-INITIAL condition . . . 128
6.13 Component strengths, Syllables 3 and 4, LOW/IP-MEDIAL condition . . . 128
6.14 Results of linear mixed effects models Syllable 3 and Syllable 4, Experiment 4 . . . 129

7.1 Demographics of six Medʉmba speakers showing varied behavior in speech cycling 139
7.2 Model parameters for best-fit models, Syllable 4, PRESI condition, 6 subjects . . . 142

7.3 AIC and BIC Values for comparison of linear, quadratic, and quadratic with plateau models for Syllable 4, PRESI condition, 6 subjects; values for winning models for each subject are bolded . . . 142

ABSTRACT

This dissertation aims to examine the following questions: Do all languages have asymmetries in relative syllable prominence? And, how best can a linguist determine relative syllable prominence in a language they do not speak natively, particularly if acoustic correlates of prominence are sparse? To answer these questions, I present results from a series of speech production experiments utilizing the speech cycling paradigm (Cummins & Port 1998), in which subjects repeat sentences in time to a metronome at progressively faster rates. Previous results using the paradigm have shown that speakers from a wide variety of languages align prominent syllables in their respective languages with certain positions within the repetition cycle, namely lower-order fractions of the cycle such as the halfway point. This work has shown that stressed syllables in languages like English, Arabic, and Portuguese behave similarly to "accented" syllables in languages like Korean and Japanese in terms of relative alignment in phrase repetition. But what about languages which display no clear phonetic evidence of stress or accent? The present work addresses this question by examining speech cycling results in the Medʉmba, a Grassfields Bantu language with a robust system of lexical tone but which displays little phonetic evidence of stress or accent. Findings indicate that Medʉmba does display evidence of prominence asymmetries at the word and phrase levels. Specifically, stem-initial syllables are found to exhibit relatively greater prominence than stem-final and non-stem syllables, and syllables occurring in the head position of an intonational phrase are found to exhibit relatively greater prominence than phrase-medial syllables. No differences in the behavior of high tone vs. low tone syllables was found; syllables of either tone could display relatively greater or lesser prominence depending on the prosodic context. Rather, the relationship between tone and prominence is shown to be indirect: as suggested by prior work, high tones can be attracted to positions which are metrically prominent. Crucially, the present work shows that metrical prominence displays similar behavior across languages regardless of whether they have stress. Thus, it is argued, a single notion of metrical prominence (independent of stress) should be applied across languages.

CHAPTER 1

INTRODUCTION

The goal of this thesis is to better understand the nature of rhythm in language, and specifically, to answer the question: do all languages have rhythmic prominence asymmetries? And, relatedly, *how do we know?* In order to begin to answer these questions, I present experimental studies of Medʉmba, a Grassfields Bantu language, which investigate relative prominence of syllables based on their alignment in repeated phrases, a method known as SPEECH CYCLING. Despite years of in-depth study of their tonal systems, Bantu languages have been largely left out of discussions of rhythm. Indeed, very little is known about the nature of syllable prominence in these languages. This work seeks to move beyond many of the theoretical and methodological obstacles which have prevented in-depth study of prominence asymmetries in Bantu so as to build a more complete general theory of rhythm in language, while also highlighting aspects of Medʉmba (and potentially other Bantu languages) which push us to reconsider longstanding assumptions about, for example, the role of stress and the nature of prosodic structure.

In one respect, this work will make a contribution to the typological study of prominence, in that it will highlight similarities and differences in the elements which are treated as prominence-bearing across languages. Much interesting work has taken place in the domain of prominence typology, and has typically proceeded by grouping languages based on properties they share or do not share in *marking* prosodic prominence. The present study seeks to examine prominence from a slightly different perspective, one that will shed light on the role it plays in constraining (and being constrained by) our language production and perception systems. In doing so, this work also aims to understand general properties of linguistic prominence in terms of how it is shaped by cognitive, articulatory, and perceptual limits. An additional goal of the thesis is to extend these generalizations to map out possible paths of change in prosodic systems and the mechanisms which underlie them.

Key to understanding generalizations across languages which exhibit apparently different types

of prominence asymmetries (e.g. languages with word-level stress, such as English, vs. those with only phrase-level ‘accent’, such as Korean) is the ability to compare elements which behave in some sense ‘the same’ in terms of reflecting speaker intuitions of prominence. As will be discussed in §1.2, while metalinguistic judgments about prominence constitute a fascinating phenomenon to study in their own right, they are often highly variable within a speech community (Kimball & Cole, 2014), making assessments of prominence difficult to gauge within a single language, never mind across different languages. It is for this reason that the present work relies on the speech cycling paradigm of Cummins (1997) and Cummins & Port (1998) and Tajima (1998), which has subjects repeat short sentences in time to a metronome at increasingly faster speech rates. Speech cycling experiments have now been carried out on a number of languages including English (Cummins, 1997a; Cummins & Port, 1998), Japanese (Tajima, 1998), Jordanian Arabic (Zawaydeh *et al.*, 2002), Brazilian Portuguese (Barbosa, 2006), and Korean (Chung & Arvaniti, 2013). The benefit of the speech cycling task is that it allows for the evaluation of speaker intuitions of prominence without relying on metalinguistic judgments from speakers. The task and its predicted outcomes are also firmly grounded within a powerful theoretical framework, Dynamic Systems Theory, which will be covered in more depth in Chapter 4, though the reader is encouraged to consult Haken *et al.* (1985); Kelso (1995, 2009) and references therein to supplement this introduction.

1.1 Definitions of Key Terms

By now, much of the vocabulary surrounding the study of prosody and suprasegmentals has come to be used differently by different authors, sometimes due to theoretical concerns, but often simply due to the tradition surrounding the study of a particular language or family of languages. Therefore, in order to avoid confusion, I will first lay out some basic definitions of the terms I will use throughout the dissertation. Justification for some of these choices in vocabulary will become clear in the rest of this section. Some aspects of these definitions may still be problematic, as there are

rarely cases in which clear-cut typological boundaries can be drawn along prosodic lines. These definitions are meant largely as a starting point, as it is my hope that, through the course of the dissertation, a different perspective on prominence will take shape which will allow for a more nuanced understanding of the terms to be presented.

Firstly, I will draw on the work of van der Hulst (2011) in defining ACCENT as "the locational aspect of prominence" (p. 2). Accent, for our purposes, may refer to the word level, the phrase level, or any higher level of prosodic structure. Separate from this notion (in van der Hulst's terms) is the *phonetic realization of prominence*, which can be conveyed through a number of cues cross-linguistically, including f_0 height, syllable duration, amplitude, and vowel quality. This type of prominence has also alternately been referred to by Lerdahl & Jackendoff (1983) and later Hyde (2016) as "phenomenal" prominence.

Since f_0 may be one phonetic correlate of accent, we will define TONE separately as the *distinctive use of pitch at the lexical level*, where *relative pitch values of individual syllables or moras can be determined by the word itself*. Tone and accent can of course interact in various ways, with high tones across languages often attracted to positions or relative prominence within a word or phrase (de Lacy, 2002).

I will define PROSODY or *prosodic structure* as the units of structure implicated in traditional analyses of the prosodic hierarchy, such as Selkirk (1980, 1984) and Nespor & Vogel (1982). As will be discussed, the term RHYTHM has a complex history within the field of linguistics, and notions of rhythm have varied between researchers. In this dissertation, we will use the term to broadly refer to the principled organization of strong and weak elements—or PROMINENCE ASYMMETRIES—within a particular prosodic unit. METRICAL STRUCTURE, on the other hand, will be used specifically to refer to formal grid- and foot-based theories, such as Liberman & Prince (1977); Hayes (1980, 1995).

1.2 Assessing Prominence

One of the greatest difficulties linguists are faced with in trying to understand prominence asymmetries across language is in assessing exactly which elements in a language are prominent. For many years, the lack of availability of high-quality recording equipment—particularly where linguists were working in the field, far from the amenities of a laboratory—meant that transcriptions of impressionistic judgments of phonetic and phonological patterns constituted the only available method for data collection. Such transcriptions still inform a great deal of work in phonology and language typology. As de Lacy (2014) notes, however, particularly where judgments of stress are concerned, individuals who are not native speakers of the languages they study are prone to make inaccurate generalizations based on incomplete data or variable patterns of use among speakers, and may also make transcription errors due to problems perceiving what the speaker has said. For this reason, he and others (see also, for example, Gordon (2014)) highlight the importance of collecting instrumental phonetic data from a wide variety of prosodic contexts to verify phonological analysis of stress.

Another issue, however, is that phonological patterns which suggest the presence of prominence asymmetries may not always map to speaker intuitions of prominence. For example, resistance to vowel reduction, a common pattern found on syllables bearing stress, may not reflect prominence at all, but rather properties of word structure which favor (perhaps for perceptual reasons) the maintenance of contrasts in certain positions, such as word-initial position (Beckman, 1998; Smith, 2002). In such instances, it is quite difficult to assess whether or not phonological patterns directly map to speaker intuitions of prominence. Furthermore, even among languages which show relatively 'clear' evidence for phonetic stress, correlates of stress can vary (Llisterri *et al.*, 2003; de Jong & Zawaydeh, 1999; Williams, 1985), making it difficult to assess, at first look, which acoustic cues are associated with greater prominence. Relative prominence is also found in speech perception studies to be highly influenced by one's language (Peperkamp & Dupoux, 2002; Rosenberg *et al.*, 2010; Chrabaszcs *et al.*, 2014).

In an ideal world, researchers would verify their analysis by confirming their perceptions of relative prominence with the intuitions of a native speaker. This, too, can prove quite difficult in practice, however. For example, it is difficult to identify exactly how to frame questions of relative prominence for metalinguistic judgment. As noted by Kimball & Cole (2014), asking subjects to identify 'beats' or 'prominences' in a sentence may evoke different responses depending on subjects' interpretation of these terms. An alternative approach is to try to access speakers' intuitions indirectly, without the use of metalinguistic judgments. Though it was not designed for this specific purpose, the speech cycling task has proven useful in this regard. The task, which will be described in greater depth in Chapter 4, provides a window into speakers' intuitions about locations of prominence by highlighting common timing patterns displayed by prominent syllables and moras across languages. Thus, it will be argued, the speech cycling task is a very useful tool for allowing researchers to compare their own impressions of prominence with native speaker intuitions. In Chapters 5-7, I will apply the speech cycling task to investigate patterns of relative prominence in the Medʉmba language. In the next chapter, I give a brief overview of the language, its historical and theoretical context, and the aspects of the language which make an especially good fit for exploring prominence asymmetries in Bantu languages.

CHAPTER 2

MEDʉMBA: HISTORICAL CONTEXT & THEORETICAL SIGNIFICANCE

The empirical contribution from the present study comes from a series of experiments carried out with speakers of the Medʉmba language, a Grassfields Bantu language spoken in Cameroon. While descriptions of specific phonological patterns thought to be associated with prominence asymmetries will be introduced in Chapters 5-7, this chapter aims to give a brief overview of the language, its place within the study of Bantu more generally, and some aspects of its phonology and morphology which have been of interest to linguists. As will be discussed, many of the characteristics of Medʉmba an other Grassfields languages which have drawn attention from historical linguists in understanding the Bantu language family are those which make Medʉmba an especially good language for investigation through the speech cycling method.[1]

Medʉmba is estimated to be spoken by around 210,000 speakers (Simons & Fennig, 2017) in Nde division, West Province, Cameroon. The language belongs to the Mbam-Nkam (otherwise known as Eastern Grassfields) family, and is subclassified as Bamileke (Watters, 2003) (Figure 2.1).

1. For additional information on the structure of the Medʉmba language, the reader is encouraged to consult work by Voorhoeve (1971) and Danis (2011) for description and analysis of the Medʉmba tone system, Voorhoeve (1967, 1968) for description of nominal morphology, Kouankem (2012) for syntactic analysis of DP structure, Keupdjio (Under review) for analysis of plural constructions, Keupdjio (2014) for studies of A-bar movement and focus, Keupdjio & Wiltschko (Under review) for work on polar questions, Nganmou (1991) for discussion of tense and modality, andMucha (2017) for formal semantic analysis of 'graded' tenses.

I. Northern Bantoid	G. Wide Grassfields	H. Tikar
II. Southern Bantoid	1. Menchum	I. Ndemli
(Wide Bantu)	2. Western Momo	J. Mbam
A. Tivoid	3. Narrow Grassfields	1. West
B. Jarawan	a. Momo	2. Yambasa
C. Mbe	b. Ring	3. Sanaga
D. Ekoid	1) West	K. (Narrow) Bantu
E. Mamfe (Nyang)	2) Center	
F. Beboid	3) East	
1. West	4) North	
2. East	c. Mbam-Nkam (Eastern Grassfields)	
	1) North	
	2) Nun	
	3) Ngemba	
	4) **Bamileke**	

Figure 2.1: Southern Bantoid languages, from Watters & Leroy (1989)

In many ways, Medʉmba appears to differ from other, more 'canonical' Bantu languages, most notably in terms of its mainly mono- and disyllabic lexical inventory, its relative lack of productive inflectional noun class morphology (Voorhoeve, 1968; Hyman *et al.*, 1970; Good, 2012; Goldman *et al.*, 2013), and, similar to other Grassfields Bantu languages, largely isolating morphology (Hyman, 2007). This is in contrast to, for example, Southern Bantu languages such as Zulu or Sotho which make use of highly agglutinating morphology, including elaborate patterns of inflectional morphology.[2] Indeed, early Bantu classifications, including the system of referential zones for Bantu languages proposed by Guthrie (1948), did not include Medʉmba and other Grassfields languages due to these apparent differences. Later descriptions by Greenberg (1963) and Hyman (2007) describe Niger-Congo as a whole as being variably affected by 'drift' towards simplification of morphology, resulting in a proposed division of languages into more 'Bantu-like' (e.g. with more agglutinating morphology) versus more 'Kwa-like' (e.g. more isolating) profiles. As Good (2012) shows, the situation is quite a bit more complex, especially if one takes into account, for example, the many different ways that noun class systems can pattern across Niger-Congo.

As will be discussed at the end of this section, the relatively reduced and isolating morphology of Medʉmba actually makes it an especially good language for investigating prominence in speech

2. Though see Demuth *et al.* (2009) for evidence that Sotho, too, has undergone some loss of noun class prefixes.

cycling. First, however, I will discuss the complex but compelling evidence that Medʉmba can, indeed, be traced to Proto-Bantu. A demonstration for this relation comes from Voorhoeve's 1971 work showing tonal evidence for historically more complex morphological structure in Medʉmba. The story begins among high tone nouns, where Voorhoeve demonstrated that some nouns could trigger downstep on another following high tone word (here, following Danis 2011, I use the example of the polar question particle *kí*) and others could not. In (1), as indicated by the red downward arrow (ꜜ), the high tone on the polar question particle is lower in pitch compared with the preceding high tone on the word *mɛ́n* 'child'.[3]

(1) Variation in downstep after high tone words (Danis 2011:24)

a. Ó ꜜjə́n ꜜmɛ́n ꜜkí
2.sg see child Q
'Have you seen the child?'

b. Ó ꜜjə́n ꜜʧʉ́ kí
2.sg see tree Q
'Have you seen the tree?'

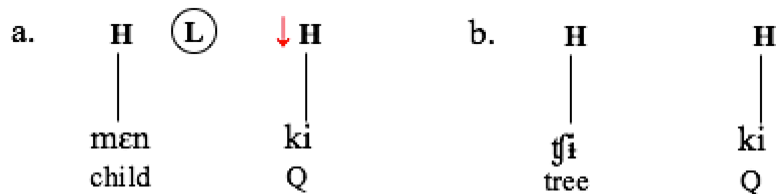
Voorhoeve proposed that this difference in behavior could be explained by positing underlying 'nonsegmental' tones, or what would later become known in much of the Bantu literature as *floating tones* in accordance with terminology from Goldsmith's 1976 autosegmental theory of tone.[4] Specifically, Voorhoeve proposed that words like *mɛ́n* which could trigger downstep carried a 'suffixal' floating L tone, and that words like *ʧʉ́* either carried no additional floating tone, or carried a high floating tone. In this way, downstep could be explained as the effect of a floating low tone

3. Note that while an orthographic system has been developed for Medʉmba by the Centre d'Étude et de Promotion des Oeuvres Mə̀dʉ̂mbɑ̀, the orthography is not transparent for understanding phonological processes. For the sake of clarity, I will use the International Phonetic Alphabet for all examples.

4. Goldsmith (1990) points out, importantly, that the term 'floating tone' has been used to refer to two potentially very different phenomena: on the one hand, the term has been used to refer to a morpheme which has only a tone but no segments; on the other, it has been used to refer to a tone which (perhaps in only certain phonological contexts, such as in the coalescence of two vowels) is not associated with a segment. As will be described, Medʉmba displays both such phenomena. To differentiate, we will refer to the former case as a 'floating tone morpheme' and the latter simply as a 'floating tone'. We retain the term 'floating' in both cases, as it seems appropriate for both types of tones. For example, tones of both varieties can 'dock' to a nearby segment (thereby ceasing to 'float') under the right conditions.

intervening between two high tone words. Autosegmental representation of this pattern is shown in (2).

(2) Underlying floating tones proposed by Voorhoeve (1971) for Medʉmba high tone words



Voorhoeve proposed (and Hyman 2003 later elaborated on) a theory of downstep in Medʉmba in which patterns were predictable based on diachronic tonal contexts: in those cases where the process was triggered, the noun in question had possessed a final low tone syllable in Proto-Bantu, which is not found in Medʉmba. In the cases where downstep does not take place, the nouns in question had either been monosyllabic (such as 'tree') or had had a final syllable which was high tone. Voorhoeve posited that the floating tones for high tone nouns (as well as for low tone nouns, which for the sake of brevity are not presented here) might be remnants of the original tones from Proto-Bantu (3).

(3) Relating Medʉmba downstep to tones from Proto-Bantu

a. Ó ꜜjə́n ꜜmɛ́n ꜜkí *mù-jánà 'child'
2.sg see child Q
'Have you seen the child?'

b. Ó ꜜjə́n ꜜʧʉ́ kí *mù-té 'tree'
2.sg see tree Q
'Have you seen the tree?'

Voorhoeve's work went beyond the description of basic downstep in Medʉmba, and showed that one could actually observe *double downstep* under the right conditions, such as in the associative 'noun of noun' construction. In (4), we see that each noun is associated with a preceding floating low tone (in red), traced to Proto-Bantu noun class prefixes, and a following floating low

tone (also in red), traced to a second syllable of the noun stem in Proto-Bantu. In between the two nouns is a floating tone *morpheme* (in green), which is the associative marker. This morpheme never bears segmental content; rather, it either surfaces on nearby segments (often forming a contour with tones already associated with that segment) or, like other floating tones, is only observed as downstep. The H(L)(H)(L)H sequence in (4) results in double downstep: the pitch of *mɛ́n* is lowered roughly twice as much as it would be when targeted for only a single downstep.[5]

(4) Double downstep found in the Medʉmba associative construction (Voorhoeve 1971)

(5) ` ʒú ` ´ ` mɛ́n ` → ʒú ꜜꜜ mɛ́n

thing AM child

'thing of child'

As Hyman (2003:) shows, this pattern, too, can be linked diachronically to segmental content in Proto-Bantu.

(6) Proto-Bantu roots of double downstep (Hyman 2003:9)

Proto-Bantu		Medʉmba	
*kì-júmà + kí-á + mù-jánà	→	ʒú ꜜꜜ mɛ́n	'thing of child'
L H L H L H L			

In (6) we see the reconstructed associative construction for 'thing of child', including the low tone noun class prefixes on each noun, the disyllabic noun stems with high-low melodies, and the high tone associative marker with accompanying noun class prefix, which is concordant with the class of the first (head) noun. Hyman posits that the loss of the vowels in the stem-final syllables of nouns, as well as those of the noun class prefixes and the associative marker, resulted in a sequence of floating tones forming the observed H(L)(H)(L)H sequence. Each low floating tone can trigger an occurrence of downstep, so double downstep results.

5. Hyman & Tadadjeu (1976) point out that variation exists within speakers such that constructions such as 'thing of child' can either be produced with a simple doubly downstepped pitch drop on the second noun (e.g. $\acute{N}_1$ ꜜꜜ$\acute{N}_2$) or with a high to downstepped high contour on the first (head) noun, and an additional level of downstep applied to the second noun (e.g. $\acute{\bar{N}}_1$ ꜜN_2, where the acute macron [́̄] represents a high to downstepped high contour).

As Hyman notes, from a representational standpoint, there are various ways one might consider treating the effects of floating tones. One possibility would be to include a [+ downstep] diacritic on high tones which can trigger downstep, and [- downstep] on those that cannot. Another would be to create a tiered feature geometry such that high and low tones could be present at multiple levels of structure, providing for more combinations of tones and possible outputs. In the case of Medʉmba, however, evidence would seem to support Voorhoeve's claim that downstep was the floating tone analogue to *downdrift*, whereby lowering occurs on the second of two high tones separated by an associated low tone. Indeed, there are many environments where one sees parallel downstep and downdrift behavior between H Ⓛ H sequences (two high tones with an intervening low tone) and H L H sequences (two high tones separated by a non-floating, associated low tone). For example, verbs in the unmarked recent past (or present for stative verbs) can either have a Ⓛ-H or L-H tonal profile. When they follow a high tone subject, the high tone on both types of verbs is downstepped with respect to the subject. This can be seen in (7a), where the floating low tone is represented as an unassociated grave accent ` and the relevant downstep is marked in red. Note that downstep is marked before the low tone, though it really targets the high tone in each sequence.

(7) Downstep targets both Ⓛ-H and L-H verbs after a H subject

a. Ó ꜜ ` jə́n ꜜmɛ́n
2.sg see child
'You have seen the child.'

b. Ó ꜜ lɛ̌n ꜜmɛ́n
2.sg know child
'You know the child'

As noted by Danis (2011), a floating tones analysis also explains a number of other behaviors of tones in Medʉmba which other analyses cannot adequately capture, including tone docking behavior on toneless particles. Thus, both synchronic and diachronic facts seem to converge, supporting proposals by Voorhoeve and Hyman that Medʉmba is descended from Proto-Bantu,

and has lost much of its segmental content, leaving only tones remaining.[6] It is interesting to note that, in the case of nouns, stem-initial syllables remained fully intact, while prefixes, and, in many cases, stem-final syllables, were reduced to floating tones. While there are various reasons why segmental reduction may have patterned in the way that it did, these results are consistent with a situation in which stem-initial syllables are more prominent (and thus less resistant to reduction) than other syllables, a possibility we will explore more in Chapter 5.

All in all, there is very little work examining rhythmic properties of Bantu languages generally, and there has been no phonetic investigation of syllable prominence in Medʉmba previously. Franich (2014), the only work to date investigating relative prominence in Medʉmba, does so from the perspective of trying to understand high tone spreading, a common tone sandhi process in Bantu languages which we will return to in some depth in Chapter 8. Franich finds, among other things, that Medʉmba does not show clear evidence for mora timing, as has been found for other Bantu languages (Goldsmith 1992; Ntihirageza, 2001; Downing & Kadenge 2015). For example, typical features of quantity sensitive languages, such as contrastive vowel length, are not found, nor is the phonological behavior of open and closed syllables found to differ. However, Franich does describe a number of distributional patterns and phonological processes which suggest that certain syllables—namely stem-initial syllables, and heads of intonational phrases—behave as more prominent than other syllables. We investigate these processes more in Chapters 5-7.

Finally, there are a number of reasons why Medʉmba makes an excellent test case for a study in Bantu prominence asymmetries. For example, the more isolating morphology, lack of class prefixes, and shorter words in the language allow for testing of a great many combinations of structures and tonal combinations while constraining the overall length of sentences. This is ideal for speech cycling, as the task of repeating sentences to a metronome gets increasingly difficult as sentences get longer and more cumbersome to articulate. In addition, it would be difficult to speak of prominence in Bantu without discussing tone, which Medʉmba has in abundance. Traditionally,

6. Indeed, subsequent work on other closely related Eastern Grassfields languages such as Fe'fe' (Hyman, 1972) and Yemba (Dschang) (Hyman & Tadadjeu, 1976) has concluded the same for these languages.

many linguists have treated stress and tone as existing on opposite ends of a prosodic continuum (Chen, 2000; Hualde *et al.*, 2002; Yip, 2002; Gussenhoven, 2004), with the implicit idea that tone and stress cannot exist in the same language unless it lies somewhere in the 'middle' of the continuum. In other words, it is not predicted that a language with a prototypical, nonrestricted tone system will also show evidence of stress-like prominence asymmetries. As seen from the examples above, Medʉmba has a vibrant system of lexical tone in which both high and low tones, as well as falling and rising contour tones, are active, making it an ideal language with which to test this prediction. In Chapters 5-7, I will show that Medʉmba does, indeed, show evidence for prominence asymmetries, and will argue for a more nuanced connection between stress and tone which is crucially mediated by patterns of temporal alignment between the two. Before that, I will discuss previous work in phonetics and phonology on the nature of prominence in language.

CHAPTER 3

PROMINENCE AND RHYTHM IN LANGUAGE: THE VIEW(S) FROM PHONOLOGY AND PHONETICS

The study of rhythm in language has a long tradition in linguistics, and rhythm has been studied from many different perspectives. The aim of this chapter is to situate the approach to prominence asymmetries put forth in this thesis within the broader literature on prominence in phonology and phonetics, highlighting important work across different domains of rhythm studies in language. I first provide an overview of earlier theoretical approaches to the study of English stress, and how these theories have impacted the study of prominence cross-linguistically, including within Bantu languages. I then discuss more quantitative approaches to the study of speech rhythm, and the insights they have provided as to the typology of rhythm classes across languages.[1] Finally, I introduce prominence from the perspective of dynamical models, which capture many of the observations from both theoretical and quantitative traditions.

3.1 Rhythm and Prominence in Phonological Theory

3.1.1 Early Theories of Prominence and Metrical Structure

Since at least the work of Pike (1967) and Fudge (1969) (though see Goldsmith (2009) for additional references), it has been widely recognized that a huge amount of explanatory power is gained by assuming that phonological structure, like syntactic structure, is hierarchical in nature. Unlike the stress rules laid out in Chomsky & Halle's 1968 *Sound Patterns of English* which referred exclusively to segments assigned a feature value of [±stress], Pike's and Fudge's work recognized the importance of the syllable as a unit of representation and of phonological rule application.

1. Note that the labels 'theoretical' and 'quantitative' are used here only for ease of exposition, and are not meant to imply that the two domains are mutually exclusive: theoretical work very often involves a quantitative component, and quantitative work very often involves a strong theoretical basis.

Building on this notion of syllable structure and drawing on notions of meter and stress-tune pairings in music, Liberman (1975) and later Liberman & Prince (1977) proposed that stress patterns in language should be viewed in terms of relative prominence relations resulting from the grouping of phonological constituents (in this case, syllables) into binary-branching metrical 'trees'. These trees were then aligned on a grid structure which allowed for readjustments to rhythmic patterns (such as the well-known English 'rhythm rule' which prevents phrase-level stress clash) that would otherwise be predicted based on constituent structure alone. Later work by Vergnaud & Halle (1979) and Halle & Vergnaud (1987) attempted to dispense with the notion of a [±stress] feature, instead accounting for stress patterns in terms of more general organizational properties of syllables, introducing the binary stress foot as a basic unit of phonological analysis. This work also moved beyond the study of English stress and began to investigate metrical patterns in languages such as Polish and Spanish, as well as non-Indo-European languages including Dakota, Odawa, and Chamorro, and even exploring metrical patterns of tone assignment in Creek. Hayes (1980, 1995) later added considerable cross-linguistic data to the study of metrical feet, now introducing a distinction between iambs, or feet with stress on the right ('right-headed') and trochees, or feet with stress on the left ('left-headed'). Hayes's work also was the first to highlight apparent phonetic asymmetries found between syllables in iambs versus trochees: while trochees were argued to display greater amplitude on stressed syllables but similar duration between stressed and unstressed syllables, iambs were characterized as having longer stressed syllables, but similar amplitude between stressed and unstressed syllables. Hayes predicted based on this generalization that languages with iambic stress systems—but not those with trochaic systems—would always be *quantity-sensitive*, with stress assignment dependent on the internal structure of syllables. Subsequent work within the framework of Optimality Theory (Prince & Smolensky, 1993) has recast many of these generalizations in terms of violable constraints. For example, Kager's (2007) influential work developed constraints such as FtType=Trochee and FtType=Iamb, whose ranking would determine whether feet in a given language would present as trochaic or iambic, and

PARSE-SYL, which requires that syllables be parsed into feet. There is still ongoing debate within the field of metrical theory about whether or not the foot is a necessary unit of analysis, with some researchers arguing that binary stress relations can be adequately characterized based on a grid system alone (Prince, 1983; Gordon, 2012). One shortcoming of these accounts, however, is in accounting for apparent non-stress patterns which are sensitive to foot-level constituency (Gordon, 2011). We will discuss some such processes later in this section.

3.1.2 Studies of Tone and Accent and the 'Stress to Tone Continuum'

During the period of the 1970s and 80s when extensive theoretical investigation into stress systems was taking place, major advances were also being made in theoretical approaches to lexical tone. While Yip's 1980; 1989; 1995 work provided important and detailed investigation of tone in East Asian languages, the focus of much of the work on tone in the 70s and 80s was on Bantu languages. Descriptions of tone systems provided insight after insight into the complex behavior of tone in Bantu (Leben, 1973; Hyman & Schuh, 1974; Goldsmith, 1976; Clements & Ford, 1979; Clements & Goldsmith, 1984), and were central in the development of many areas of linguistic theory including the nature of distinctive features (Clements, 1985; Sagey, 1986), the phonology-morphology interface (McCarthy, 1981; Marantz, 1982; Kiparsky, 1982; Mohanan, 1986), and, with the publication of Goldsmith's 1976 dissertation on Autosegmental Phonology, the very nature of phonological representations. Despite the great focus on tonal properties of Bantu languages, relatively little work sought to understand prominence asymmetries in these languages. This is likely in large part due to 1) the tendency to view prominence as exclusively tied to stress and languages with clear phonetic correlates of stress (Hayes 1995, for example, absolutely equates stress with rhythm); and 2) the (related) assumption that tone, rather than constituting a separate phenomenon, was somehow an alternative to stress, perhaps residing on the other end of a prosodic continuum from stress (Hyman, 2009). The latter bias seems to have surfaced as early as Martinet (1960) and Garde (1967) and was also furthered through the study of so-called 'pitch accent systems',

which many have characterized as an intermediate step on the stress-to-tone continuum (McCawley, 1970; Beckman, 1986; Hualde *et al.*, 2002; Elordieta & Hualde, 2003; Gussenhoven, 2004), but which have since been recognized as a very heterogenous group of languages which display quite different prominence-related behavior from one another (McCawley, 1978; van der Hulst, 1999; van der Hulst & Ewen, 2001; Hyman, 2001, 2006, 2009).

What early work did investigate prominence in Bantu was largely aimed at explaining the complexities of tonal patterns, using the notion of accent as a theoretical device rather than exploring it as an object of study in its own right. Meeussen (1963); Carter (1971); Goldsmith (1981) all take different approaches in exploring the role of accent in conditioning tonal patterns in Tonga, a Bantu language spoken primarily in Zambia. In these investigations, the distribution of high tone, in particular, is argued to be linked to an underlying abstract accent (or 'determinant' in Meeussen's terms). Goldsmith (1984) provides a diachronic account of what he characterizes as the rise of accent in two Lacustrine Bantu languages, exploring evidence from patterns of high tone shift towards what he characterizes as a 'metrically strong position', though it is not entirely clear what independent evidence exists for metrical strength. Goldsmith (1987, 1991) proposes parallel systems of tone and accent assignment which are both dependent on a metrical grid or tree structure. In particular, Goldsmith (1987) proposes the TONE-ACCENT ATTRACTION CONDITION (TAAC), a well-formedness condition which says that no tone-bearing syllable should have a lower level of accent than a toneless syllable. This idea is applied by Bickmore (1995) in an analysis of tone in Lamba in which he proposes that stress is assigned via trochaic footing within certain morphological domains, and that high tones are then attracted to stressed syllables, ultimately leading to alternating patterns of high and low syllables. In the data that Bickmore presents, stressed syllables always attract high tones, leaving open the question of what, if anything, distinguishes a stressed syllable other than its ability to attract high tones. Later work by de Lacy (2002) builds on notions from Goldsmith's and Bickmore's work within Optimality Theory, proposing a set of markedness constraints which motivate the attraction of tones (high tones, in particular) to heads of metrical

feet. As is apparently the case with Bickmore's account, de Lacy invokes an abstract notion of stress which equates it with foot structure.

Other work has drawn a distinction between binary foot-like structures and stress, instead invoking the notion the TONAL FOOT, a unit which, according to Leben (1997, 2002), operates over a binary domain on the tonal tier, but can in principle operate independently of stress-based foot structure on the segmental level. Leben applies this unit in the analysis of tone in loan words in Hausa and Bambara, showing how it captures generalizations about high and low tone alternations. Subsequent work by Pearce (2006) argues for quantity sensitive iambic tonal feet in Kera, using foot structure (crucially defined in terms of metrical feet at the segmental level) to explain patterns of tone spreading and vowel harmony. Despite stress-like behavior of foot heads in Kera (including increased duration and intensity), Pearce argues that the language does not have stress or accent. Further work by Rose & Jenks (2011) appeals to the tonal foot in explaining high tone placement and spreading in Moro, a Kordofanian language, labeling such feet as 'metrical' but arguing, again, that Moro lacks stress. Downing (1990) argues for two tone spreading rules in Nguni, including a foot-based metrical spreading rule and an additional non-metrical 'local' spreading rule. By ordering these two rules in a derivational framework, tone spread can be constrained to targeting high tones underlyingly associated with the penultimate position, but not those which occur on the penult as a result of tone spreading. Once again, metrical structure is invoked, but stress is argued not to be present in the language, and no direct claim is made about the existence of accent.

While potentially important connections between tone and prominence across languages are highlighted in the work summarized above, many questions still remain about the more general nature of prominence and how it is determined in these languages. As mentioned, in some cases, analyses of tone languages have invoked the notion of accent but have not provided a clear theoretical motivation for its use, other than illustrating its utility in explaining tonal phenomena. Even when foot-based structures are introduced, it is not always clear how such structures should be evaluated in relation to, for example, more traditional stress-based foot structures from metri-

cal theory. Hyman (1989) points out that the proliferation of work evaluating 'accentual' versus 'purely tonal' analyses of tonal phenomena is mired by a lack of generality in the analysis of the structure and basis of accent. As will be seen later in this section, efforts to refocus work on accent in tone languages has revealed many sources of evidence for accent-like prominence asymmetries in these languages which may prove useful in understanding the nature of accent and how it arises. Before turning to this evidence, we examine some theoretical models of metrical and prosodic structure which provide additional flexibility for analyzing prominence alternations in the absence of stress.

3.1.3 Alternatives to Traditional Metrical Theory

While earlier grid- and foot-based metrical theories analyzed rhythmic structure in terms of stress patterns and posited a one-to-one correspondence between stress, metrical prominence, and foot structure, some more recent formal theories of prominence have eschewed the view that stress and prominence are synonymous, leaving open the possibility for greater variety of rhythmic systems to be characterized. One example is van der Hulst's 1984; 1996; 2009 'Accent First' theory. Rather than assigning stress first and interpreting relative prominence (e.g. primary vs. secondary stress) based on grid or tree alignment of stressed syllables, van der Hulst proposes that abstract accent is primary, and selects the syllable which will carry primary stress. In this view, stress is simply a phonetic exponent of accent, and nonprimary stresses, rather than being assigned at the same time as primary stress, are said to be calculated separately but with reference to the primary stressed syllable. It is suggested by van der Hulst that stress and other phonetic exponents of accent 'respect' the location of accent by aligning with it. As this theory was developed within a rule-based framework, alignment is stipulated to occur across the board. However, under this view, an accented syllable may or may not display phonetic correlates of stress, or may display specific phonetic or phonological properties which differ from typical stress, such as interactions with tonal behavior. This type of system thus better captures metrical generalizations which may operate in the absence

of stress. In general, the approach is designed to allow for languages without typical phonetic correlates of stress (e.g. increased duration and amplitude, fluctuations in f_0, etc.) to be treated equally to those with typical stress correlates, assuming evidence for abstract accent can be identified in both. If, indeed, all language are found to display properties of abstract accent, the only difference between a language with stress and a language without stress would be the phonetic correlates associated with accent.

By invoking principles of constraint violability within Optimality Theory, others have attempted to maintain fundamental tenets of traditional metrical theories while allowing for flexibility in the relation between stress, feet, and grid structure. Crowhurst (1996), for for example, recasts in OT terms the principle of INSEPARABILITY assumed in classic metrical theories which requires that feet assigned in a given language must manifest metrical heads. This allows for headless feet, and accounts for data, such as in Cairene Arabic, where head-based phonetic prominence is absent but foot-level phonological processes are evident. Hyde (2002) takes a different tack by positing that the requirement for feet to bear phonetic stress is violable; this is implemented through a lack of grid entries for stressless feet. Crucially, both of these approaches differ from that of van der Hulst in that they still assume an implicational relation between foot-based prominence and phonetic stress.

Yet another approach to prominence, proposed by Hyde (2016), advances a distinction between abstract and phonetic accent, similar to van der Hulst. Hyde's account posits a difference between what he terms *metrical accent*, or abstract grid-based prominence, and *phenomenal accent*, which is phonetic in nature. Hyde's account differs from van der Hulst's in that Hyde's system allows for misalignment between abstract and phonetic prominence: as Hyde points out, this type of misalignment very frequently occurs in music. Similar to van der Hulst, Hyde also insists that the presence of abstract metrical accent does not depend on the presence of phenomenal accent, though the former can, in many cases, inform the location of the latter.

3.1.4 Prominence at Higher Levels of Prosodic Structure

Selkirk (1978b, 1980, 1981) as well as Nespor & Vogel (1982, 1983) proposed to enrich the set of category labels available for analysis of the accentual system, introducing the first iterations of the *prosodic hierarchy* which contained, in addition to the syllable (σ) and the foot (Σ or FT), the prosodic word (ω), the phonological phrase (ϕ), and the intonational phrase (ι). Nespor & Vogel (1986) also later introduced the Clitic Group (C) to account for elements appearing to form a prosodic unit between the levels of the prosodic word and phonological phrase. A central advantage of this prosodic hierarchy was that it allowed for an elegant account of the apparent close, though non-isometric relation between syntactic and phonological structures: through a set of mapping rules and prosodic well-formedness conditions, prosodic constituents could be formed in terms of syntactic constituents without having to match them exactly. Subsequent adaptations of prosodic theory in Optimality Theory have captured the violability of syntax-prosody mappings in terms of alignment constraints (Selkirk, 1986, 1996; Truckenbrodt, 1995, 1999). While the original formulations of prosodic theory did not allow for recursive prosodic domains (requiring instead for STRICT LAYERING of prosodic constituents), later work by Selkirk (2011) in Match Theory has incorporated some level of recursivity. Additionally, growing evidence shows that the specific units of structure represented in the prosodic hierarchy can vary across languages (Downing, 1999; Schiering *et al.*, 2010; Downing & Kadenge, 2015), though overall, nested units at similar timescales are thought to occur fairly universally (Leong, 2012).

An important tenet of theories of prosodic structure was that the strong element of a higher level of structure (say, the phonological phrase) was inherently more prominent than the strong element of a lower level of structure (say, the prosodic word). This proposal has inspired many to explore how such differences in strength levels of prominent positions might be manifested phonetically across languages. One of the most notable studies to investigate the phonetics of prominence comes from Beckman & Pierrehumbert (1986), who showed that Japanese and English not only show evidence for similar levels of prosodic phrasing, but that these languages use similar strate-

gies for the organization of pitch accents, with pitch accents in both languages triggering downstep (or 'catathesis', as it is referred to by B&H) within similar prosodic domains (termed the 'Major Phrase' or MaP). Later work by Pierrehumbert & Talkin (1992), Jun (1993), Hawkins & Smith (2001), Keating & Shattuck-Hufnagel (2002), and Keating *et al.* (2003) demonstrated how the realization of individual segments in terms of such measures as glottal closure/voicing, VOT, vowel quality, and duration depended to a large degree on the specific position of that segment within the prosodic structure. Work by Fougeron & Keating (1997), Cho & Keating (2001) and Lee *et al.* (2006) has also found these effects to be cumulative when a segment occurs a location of prominence at multiple levels of the prosodic hierarchy. The notion that certain types of phonological contrasts are in fact licensed only in more prominent positions (and fully or partially neutralized elsewhere) has been observed by many (Itô & Mester, 1986; Itô, 1989; Goldsmith, 1990; Rubach, 1990) and has been implemented within Optimality Theory by Beckman (1998) through a series of *positional faithfulness* constraints.

Recently, researchers have become increasingly aware of the need to disentangle the phonetic cues associated with word- and phrase-level prominence. Liu (2009) provides an in-depth analysis of interactions between tone, sentence type, and focus in elucidating the role of each in contributing to surface intonational patterns in English and Mandarin, concluding that these elements are encoded parallel to one another in the acoustic signal. Gordon's 2014 typology of accentual systems concludes that much existing word-prosodic typology is problematic since words being tested are often elicited either in focus positions, or in isolation/phrase finally. Such approaches, Gordon argues, inevitably conflate word-level stress with phrase-level accent. Remijsen & van Heuven (2005) addresses this issue by examining f_0 patterns of stress and tone independently, providing evidence for the existence of lexical tone and lexically contrastive stress in the Curaçao dialect of Papiamentu. Hyman & Monaka (2008) examine words in different sentential contexts to tease apart the effects of word-level tone and f_0 cues at the level of the Intonational Phrase in Shekgalagari. Gordon & Nafi (2012) explore acoustic properties of word- and phrase-level prominence

separately in Tashlhiyt Berber, demonstrating that intensity is the primary correlate of word-level stress, while raised f_0 is the primary correlate of phrasal accent. Such studies have proven enormously enlightening and represent important steps toward understanding how different prosodic levels may interact.

These findings highlight two important facts about the current state of affairs in studies of prominence asymmetries: 1) due to a long-standing reliance on data from secondary and tertiary sources in the construction of phonological theories (often coming from primary sources using distinct elicitation contexts, as Gordon points out), much of what has been expressed regarding differences between 'lexical' versus 'post-lexical' or 'phrase-level' prominence must be considered with caution, and in many cases, re-examined on the basis of primary data collection; and 2) going forward, we must be very careful to be sure that comparisons of prominence across languages are carried out using comparable measures across linguistic contexts. While the present study does not provide in-depth phonetic analysis of prominence across prosodic positions, it does employ a methodology which allows for as direct as possible a comparison between languages in evaluating phrase-level prominence.

3.1.5 Prominence Asymmetries in Bantu: Non-Tonal Evidence

As discussed earlier in this chapter, tonal evidence for prominence asymmetries abounds in many Bantu languages. But what other types of corroborating evidence for prominence asymmetries might we expect in these languages, assuming typical phonetic correlates of stress (e.g. increased amplitude and duration) are absent? Hayes (1995) notes that it is common for vowel contrasts to be fully realized in the strong syllable of a foot, but neutralized to some degree in the weak syllable. Patterns of positional privilege such as this have been documented for many languages around the world, and have been explained through a variety of mechanisms. One approach, termed *Licensing-by-Cue* (Steriade, 1997), holds that perceptual factors drive the observed positional restrictions. Specifically, the theory holds that contrastive cues are better perceived in certain

phonetically-prominent positions (such as in stressed positions) due to the phonetic enhancements which accompany those positions. As a result, contrasts are more likely to be realized in those positions. A crucial observation made by Steriade was that patterns of contrast neutralization almost always seem to co-occur with phonetic patterns that would logically give rise to them. As an example, unstressed vowel reduction most often targets contrasts concerning vowel height (rather than, say, backness), and perceptual cues to vowel height are also the most vulnerable in environments where vowel duration is decreased (a typical correlate of unstressed vowels). Similarly, for tone languages, Gordon (1999a, 2001) and Zhang (2002) highlight cross-linguistic phonetic pressure for tonal contrasts to be realized on more sonorous syllables; this fact is argued to have a perceptual basis, as tonal information is more easily recovered by listeners in these environments.

Another approach, advocated by Beckman (1998) and couched within the framework of Optimality Theory, assumes that certain 'prominent' positions are granted greater licensing capacity through the universal high ranking of a set of *Positional Faithfulness* constraints which shield against contrast reduction in these positions. Prominence in this approach, particularly in the case of stem-initial syllables, is linked to psycholinguistic factors. For example, stem-initial syllables play a greater role in lexical access, word recognition, and speech production than non-initial syllables (Cutler *et al.*, 1985; Nooteboom, 1981; Hawkins & Cutler, 1988). Under this view, stem-positional effects and stress-based effects are treated as instances of the same phenomenon, stemming from similar functional and grammatical principles, though they are indexed by different constraints and generally apply in different groups of languages.

The need for a notion of psycholinguistic prominence has emerged independently in research at the phonology-morphology interface. Yu (2007) proposes that psycholinguistic prominence is the unifying factor underlying the various possible *pivot points*, or morphophonological units, to which infixes may attach (see also Nevins & Vaux 2003). Yu separates these units into *edge pivots* (such as the first vowel in a stem) and PROMINENCE PIVOTS (such as a stressed syllable). Once again, stem-initial syllables and stressed syllables appear to pattern together in the grammar.

Many African languages show segmental distribution asymmetries similar to the ones described above. In a large proportion of these languages, the initial syllable of the stem seems to be privileged in that it is able to display the greatest number of segmental contrasts (Bennett, 1978; Hyman, 1990; Akinlabi & Urua, 2003; Downing, 2004; Harris, 2004; Hyman, 2008; Downing, 2010; Lovegren, 2012; Franich, 2013; Hyman, 2015). Bennett (1978) shows that the Proto-Bantu seven-vowel contrast was only realized in stem-initial syllables, and only stem-initial vowels could contrast for tone. Hyman (1990) notes that Ibibio and Efik, two closely-related Cross-River languages of Nigeria, allow thirteen different consonants in onset position of a stem-initial syllable; only seven of these consonants are allowed in onset position of a non stem-initial syllable. Similarly, Lovegren (2012) shows that of the twenty-two contrastive consonants in Mungbam, a Bantoid language of Northwest Cameroon (and a language which shows many similarities morphologically and tonologically with Medʉmba), only eight occur outside of the stem-initial syllable position. In addition, the feature [ATR] is only contrastive in stem-initial vowels. Beckman (1998) shows that vowel height in Shona verbs may vary freely in stem-initial positions, but that it is restricted non-initial syllables. In Punu, a Bantu language of Gabon, five vowels contrast in root-initial syllables, while only three contrast in non-initial syllables. Long vowels are also only allowed in stem-initial syllables (Kwenzi, 1980; Fontaney, 1980; Hyman, 2015). Hyman (1998) provides an analysis of Yaka vowel height harmony which crucially relies on prominence peaks separated by what he refers to as a 'prosodic trough'.

As discussed previously, tones also show various positional restrictions which seem to indicate greater prominence of stem-initial syllables. Lovegren (2012) shows that non-initial stem syllables in Mungbam always carry a copy of the stem-initial tone in both nouns and verbs, suggesting that non-initial stem syllables are underlyingly toneless and inherit their tones through tone spreading from the stem-initial syllable. Similar patterns have been attested in a number of other languages, including Shona (Odden & Bickmore, 2014) and Sotho (Zerbian, 2006). Lovegren also shows that non-initial stem syllables are not able to bear contour tones, while initial stem syllables may carry

all possible contours.

One of the appeals of Positional Faithfulness is that it allows for factors other than phonetics to determine prominence. If prominence can, indeed, occur independent of phonetic cues, then this is certainly a desired outcome of the theory. The notion that prominence should be treated separately from phonetic correlates of stress is far from new, and was in fact recognized in much earlier work on Bantu languages by Goldsmith (1982). Goldsmith proposed that prominence might be best understood in terms of an abstract accent which could occur independent of stress, and which could also crucially influence tonal alignment patterns. This view has been brought up again many times since, most notably in the work of van der Hulst (1984, 1996, 2011).

If we allow ourselves to separate out the notion of prominence or 'accent' from its phonetic correlates, it becomes even more plausible to view positional asymmetries in phonological contrast across stress and non-stress languages as stemming from the same underlying factors. Beyond this, it opens up the possibility for a multifaceted view of prominence in which prominent syllables could, in principle, take many different forms.

3.2 Quantitative Approaches to Rhythm

We now turn to examine rhythm from a different perspective, one that is more focused on quantifiable phonetic manifestations of rhythm and rhythm perception. Indeed, the phonetic nature of rhythm is still hotly debated, in large part because instrumental studies of speech have been mixed in providing support for regular rhythmic patterns. This disparity between intuitions of rhythm and phonetic evidence for it have fueled much debate about the proper characterization of rhythm both in quantitative and theoretical terms.

3.2.1 Isochronic and Non-Isochronic Approaches

The idea that language was spoken as a regular pattern of temporally equal units was likely first proposed by Steele (1779). Much later, Pike (1945) formally proposed a term for this: *isochrony*.

Pike went on to propose that languages could belong to one of two categories, labeled *stress-timed* and *syllable-timed*. Stress-timed languages (such as English and Russian) were defined as those for which intervals between stresses (called 'feet') were of equal temporal length, despite the fact that different numbers of syllables might occur in this intervals. Syllable-timed languages (such as Spanish and French) were defined as those languages for which syllables were all of roughly equal length. Abercrombie (1967) built on this notion and proposed that the difference between stress- and syllable-timed languages was rooted in physiology, claiming that the contraction of the intercostal muscles when it occurred with a stressed syllable ('stress pulse') was stronger than when it occurred with an unstressed syllable ('chest pulse'). The observed rhythmic difference between languages was said to relate to which type of pulse occurred periodically.

A third category termed *mora-timed* was later introduced by Bloch (1950), and then by Han (1962) and Ladefoged (1975) in order to account for such languages as Japanese where rhythmic timing appeared to be tied to a unit smaller than the syllable. Similar to syllable-timed languages, mora-timed languages were defined as those in which successive moras were nearly equal in duration.

Despite the intuitive appeal of such characterizations of stress- and syllable-timed languages, phonetic data has continually failed to support the existence of isochrony in either type of language. Several authors (Shen & Peterson, 1962; Bolinger, 1965; Delattre, 1966; Faure *et al.*, 1980; Pointon, 1980; Wenk & Wioland, 1982; Roach, 1982; Dauer, 1983; Manrique & Signorini, 1983; Nakatani *et al.*, 1981; Dauer, 1987; Eriksson, 1991) have shown that interstress intervals are not of equal duration in stress-timed languages (and nor are syllables), and evidence also abounds that syllables and moras do not occur at equal durations (nor do interstress intervals) in syllable- and mora-timed languages (Pointon, 1980; Wenk & Wioland, 1982; Roach, 1982; Dauer, 1983, 1987).

Though the precise definitions of stress- and syllable/mora-timed languages proposed by Pike and Abercrombie may not be fully supported by phonetic data, the intuition that a meaningful difference exists between these two groups of languages has continued to fuel a great deal of

research on rhythmic structure. In an effort to reconcile such intuitions with the phonetic data, Lehiste (1977) proposed that isochrony in language could exist as a purely perceptual phenomenon. Just as listeners have been shown to normalize (or 'compensate') for variation in the speech stream attributable to such phenomena as coarticulation (Mann, 1980; Mitterer, 2006), Lehiste proposed that speakers might also be able to normalize for variation in rhythmic patterns, enabling them to perceive isochrony even where physical evidence for it was weak. Indeed, the author showed that listeners were better able to identify small durational differences in inter-stress intervals in non-speech stimuli than in speech stimuli, suggesting that speakers normalized more for duration when listening to the latter. Work by Donovan & Darwin (1979) also supports the existence of perceptual isochrony in speech. In this experiment, individuals were asked to imitate the stress pattern of an English utterance by tapping, and then perform a similar task imitating a sequence of non-speech noises. Taps in the speech condition were found to exhibit less temporal variability than taps in the non-speech condition, supporting the idea that speakers perceived the temporal spacing of beats in the speech stimuli to be more regular than that of the non-speech stimuli.

Another dominant view of speech rhythm, proposed by Dasher & Bolinger (1982), dispenses entirely with the notion of isochrony, and instead attributes perceived differences between stress- and syllable-timed languages to phonological properties. Based on observations by Dauer (1983), Dasher and Bolinger propose that impressions of rhythmic differences between languages stem from a combination of a) the greater number of syllable types and weights in stress-timed vs. syllable-timed languages (with the former more likely to exhibit quantity-sensitive stress) and b) a greater acoustic contrast (both through vowel quality and duration) between stressed and unstressed syllables in stress-timed languages vs. syllable-timed languages. Dauer (1987) further proposed a property-driven approach to rhythm typology citing the potential for properties such as syllable inventory and stress reduction to be both independent and cumulative, allowing for languages to fall along a continuum between 'canonically stress-timed' and 'canonically syllable-timed' end-points. Nespor (1990) provides concrete examples of so-called 'intermediate languages', including

languages which possess low variation in syllable structure but a high degree of unstressed vowel reduction (Catalan) and those which possess a high level of syllable structure variation and a low amount of unstressed vowel reduction (Polish).

With this phonologically-based perspective in mind, researchers have begun to re-examine possible acoustic correlates of rhythm associated with syllable inventory and vowel reduction so as to more concretely evaluate where languages would be predicted to fall on a continuum of rhythmic structure. Working from the perspective of infant speech perception, Ramus *et al.* (1999) proposed a set of phonetic variables crucial for rhythm perception based on the idea that infants largely rely on energy fluctuations over vocalic intervals to extract rhythmic patterns (Mehler *et al.*, 1996). Examining corpus data from eight languages of varying rhythmic varieties (including both typical 'stress-timed' and 'syllable-timed' languages), the authors established three key variables including proportion of vocalic intervals within a sentence (%V), standard deviation of the duration of vocalic intervals within a sentence (ΔV), and standard deviation of the duration of consonantal intervals within a sentence (ΔC), which they took to be crucial for evaluating rhythmic structure. Using these correlates, researchers have been able to model rhythmic differences between languages within a three-dimensional space, in many cases supporting prior impressionistic judgments about rhythm class. For example, Gut (2005) utilizes the %V and ΔC measures to investigate timing in two African languages, Yoruba, Hausa, as well as Nigerian English. Findings suggested, in line with descriptions from previous work which characterized African languages as syllable-timed (Abercrombie, 1967; Auer, 1991), that Nigerian English had a higher percentage of vocalic duration (e.g. less vowel reduction) than British English, but a lower percentage of vocalic duration than either Hausa or Yoruba.

Subsequent work by Low *et al.* (2000) and Grabe & Low (2002) moves to replace the acoustic measures of %V, ΔC and ΔV with what they refer to as *pairwise variability index* (PVI). The formula for the PVI, as it is applied to vowels (known as the Normalized Pairwise Variability Index, or nPVI) is given in (3.1).

$$nPVI = 100 \times \left[\sum_{k-1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| /(m-1) \right] \quad (3.1)$$

Here m represents the number of vocalic intervals, and d represents the duration of the kth interval. The formula in Equation 3.1 thus calculates the absolute value of the difference in duration d between vowels of adjacent syllables divided by the mean duration of the two vowels. Values for all pairs of vowels in an utterance are then summed and divided by the total number of differences and multiplied by 10. The authors argue that the PVI is a more reliable measure of linguistic rhythm due to the fact that it is cumulative, controls for speaking rate within an intonational phrase, and contains a component designed to normalize for articulation rate. Indeed, the authors show that the PVI more reliably distinguishes between language varieties such as British English and Singaporean English which have consistently been reported to have different rhythmic structures (the former typically categorized as stress-timed, and the latter as syllable-timed) and clearly different patterns in vowel reduction. The PVI has since been adopted and tested by many other researchers on many other languages (Ramus, 2002b; Lee & Todd, 2004; Dellwo, 2006), and has even been applied to musical rhythm, allowing for the successful differentiation of, for example, English and French musical rhythm, and even for the examination of musical 'contact' phenomena such as the waning influence of Italian music on German music from the 250 year period spanning the Baroque era to the Romantic era (Patel & Daniele, 2003a).

3.2.2 Discrete Patterning in Time: the Dynamical View

The last perspective on rhythm to be introduced describes what I will refer to as *dynamical models* of rhythm. Such models, exemplified by O'Dell & Nieminen (1999, 2008), Barbosa (2007), Saltzman *et al.* (2008), Cummins & Port (1998) and Port (2003), posit the existence of a system of coupled oscillators which are crucial for regulating speech timing in production and perception. Work from the dynamical perspective is similar in some respects to other quantitative approaches

to rhythm described above in that it attempts to capture aspects of speech timing, such as differences in relative vowel durations within syllables across languages. A fundamental difference in the two approaches is that dynamical models are focused more on the mechanisms which underlie timing differences across languages, rather than classifying languages based on these differences. Dynamical models have also been highly useful in explaining the emergence of discrete patterns in language, such as patterns of stress shift, which can vary as a function of speaking rate (Quené & Port, 2003). Iskarous (2017) highlights a number of other patterns at the interface of phonetics and phonology for which dynamical models successfully capture the relation between continuous variables and discrete units in language. We will discuss dynamical models in more depth in Chapter 4 when we discuss the speech cycling paradigm.

CHAPTER 4

THE SPEECH CYCLING TASK: THEORETICAL UNDERPINNINGS AND APPLICATION

This chapter provides an overview of the experimental paradigm, speech cycling, which will be used to study prominence asymmetries in the remaining chapters. The paradigm is of interest both for its ability to capture speaker intuitions about prominence, but also for the theoretical assumptions on which it is built. The task was originally used by Cummins (1997a) in a study of English rhythm. In it, Cummins asked speakers to repeat a set of sentences (e.g. *Take a pack of cards*) in time to a metronome, decreasing the metronome period with each successive ‘trial’, each of which consisted of 8 sentence repetitions. Over the course of the experiment, what was found was that speakers showed a very limited set of syllable alignment patterns within a repetition cycle, particularly with respect to the final stressed syllable of the sentence. Specifically, the final stressed syllable was preferentially aligned at one of three positions: one-third, one-half, or two-thirds of the way through the phrase repetition cycle (measured as the time between subsequent utterances of the first syllable of each phrase). Alignment with the one- and two- thirds phase positions reflects more of a ‘waltz-like’ pattern, with ternary rhythmic organization, whereas alignment with the halfway point gives the impression of ‘cut-time’ musical notation (Figure 4.1), organized with two main beats.

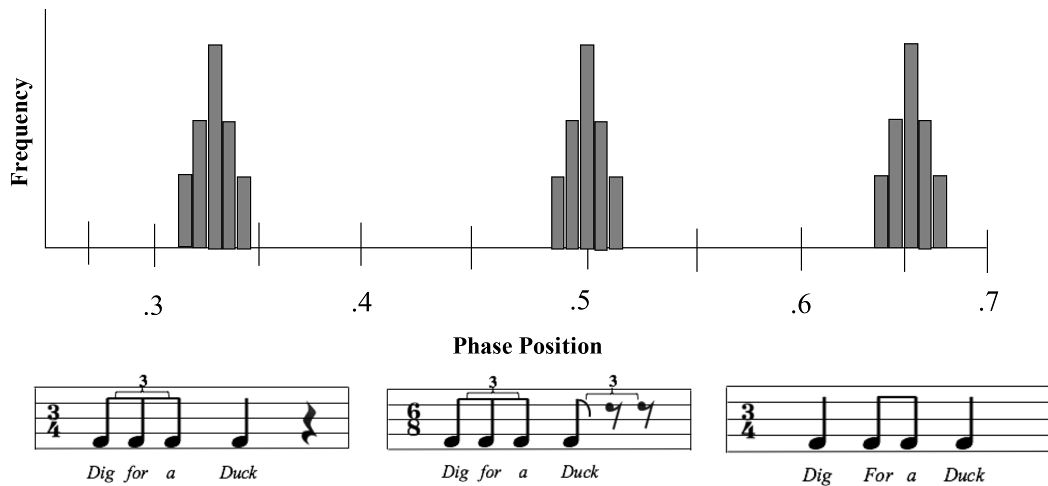


Figure 4.1: Common speech cycling patterns in musical notation (adapted from Cummins & Port 1998)

In another version of the task (known as 'targeted speech cycling'), Cummins & Port (1996, 1998) used a different approach, instead utilizing two alternating metronome beats at different f_0 frequencies. The period of the high tone beat was held constant, while the second, low tone beat was manipulated to occur at five different intervals between the high tones: .3, .4, .5, .6, and .7 of the high tone cycle. Subjects were asked to align the final stressed word (e.g. *duck* in *Dig for a duck*) with the low tone beat. As would be predicted from results of the more unstructured task, subjects had a difficult time maintaining alignment at all of these intervals except the .5 interval. For any given pattern of beats, subjects gravitated toward aligning the final stressed syllable with the halfway, one-third (.333), or two-thirds (.667) intervals of the high tone cycle, regardless of their intention to align with the low tone.

Subsequent cross-linguistic work in speech cycling has found similar results for other languages. Within Indo-Eurpean, findings from Jordanian Arabic (a stress-timed language, like English) show that, similarly to English, the final stressed syllable of a phrase/sentence will align with the same, lower-order fractions of the repetition cycle (Zawaydeh *et al.*, 2002). We will henceforth refer to these positions, following Cummins & Port, as SIMPLE HARMONIC PHASES,

or SHPs. Similar results were reported for Brazilian Portuguese, a language which is relatively more syllable-timed, by Barbosa (2006). Speech cycling studies on languages outside of Indo-European, including Japanese and Korean, have yielded different results, but consistent with predictions about relative prominence in those languages. In Japanese, phrase-final syllables were found to be prominent (Tajima, 1998), in line with other work by Fujimura (2003) and Kawahara *et al.* (2014); Kawahara (2015) examining articulatory correlates of prominence in Japanese. Tajima also found evidence for greater attraction of pitch accented moras to SHP positions over nonaccented moras. In Korean, a language argued to have no word-level prominence, but rather phrase-level accent, accented syllables of Accentual Phrases (APs) have been found to be attracted to SHP positions (Chung & Arvaniti, 2013).

4.1 Parallels with Other Motor Behaviors

The above evidence from a wide variety of languages suggests that prominent events such as stressed syllables in speech are preferentially aligned at simple harmonic phases; in other words, periodic events in speech at the phrase and stress (or foot) level exhibit ENTRAINMENT, or coordination, with one another. Why should this be? Cummins & Port (1998) take inspiration from experimental work done by Kelso *et al.* (1979) and Kelso (1995) which examined patterns of coordinative behavior in limb and finger movements. These studies highlight dynamical principles which can apply across different motor behaviors, including speech.

4.1.1 Synergistic Behavior in a Reaching Task

An important principle in dynamical systems is that of SYNERGY, which refers to the linking together of variables within a system such that they are controlled collectively. Synergy allows for economy of movement within a system, as it reduces the degrees of freedom within that system, limiting the ways in which individual elements can can vary with respect to one another. An often-cited metaphor for synergy within a system is a car: though a car has four separate wheels, each

cannot operate independently to drive in a different direction. Rather, the wheels are yoked together to a single steering chassis and constrained to operate collectively. The direction of the car is thus entirely determined by a single variable, that of the steering wheel. This makes steering the car—a task that at the most basic level requires many individual moving parts—eminently manageable. In a similar manner, the movement patterns of our bodies are constrained by the ways in which individual body parts can relate to one another in coordination. Kelso *et al.* (1979) demonstrate this nicely in a manual reaching task. In the task, an individual must reach each hand toward a different target, with the two targets at different distances from the individual hands. If the hands were working completely independently from one another in this task, we would expect the hand with the closer target to reach its target sooner. Rather, both hands are found to reach their targets at the same time, suggesting the limbs are coordinated together or COUPLED in order to act as a unit for the purpose of this task. Of course, not all tasks require the hands to work together in this way. It is for this reason that synergy and coupling are characterized in terms of task-specific dynamics: the functional goal of movement serves as an organizing and constraining influence over the individual components, or EFFECTORS, within a system.

4.1.2 Coupling and Phase Shifts in a Finger Wagging Task

Another important line of research carried out by Kelso and colleagues which demonstrates further dynamical principles is the 'finger-wagging task', in which individuals were asked to move their fingers repetitively toward and away from the midline of the body (Figure 4.2). The subjects were asked to coordinate finger movement such that the two fingers moved in the same direction (left or right) at the same time, similar to the pattern of windshield wipers on a car. Then, the speed of finger movement was paced using a metronome, such that movement was led to increase in frequency incrementally with each trial of the task. The striking finding from this study was that after the subjects reached a certain frequency of finger movement, they were unable to maintain same-direction movement of the two fingers; instead, they shifted to moving the fingers toward and

away from one another. They were able to maintain this pattern for the remainder of the task, but were not able to revert back to the windshield wiper pattern. Such PHASE SHIFTS are elegantly accounted for in a model of motor coordination developed by Haken *et al.* (1985) which is based on the notion that bimanual and inter-limb coordination is governed by a system of COUPLED OSCILLATORS. Coupled oscillators are oscillators (systems of masses on springs) which are connected such that energy can be transferred back and forth between them. A powerful example of coupling comes from pendulum experiments in which two pendulums are set to swing side by side, out of sync with one another. If transferred to a moving plane, such as a board atop two parallel tubes or soda cans, energy is easily passed between the two pendulums, and they will quickly fall into synchrony (1:1 frequency-locking) with one another.

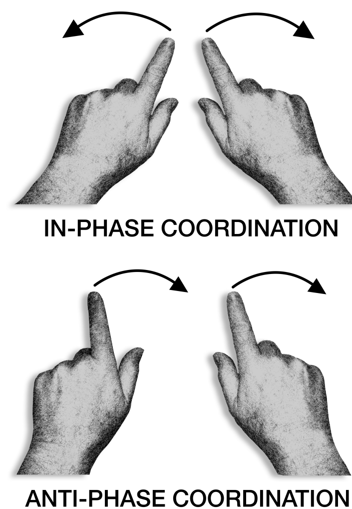


Figure 4.2: Illustration of finger wagging task (Kelso *et al.*, 1979)

What Haken et al. showed was that the abrupt transitions would occur from windshield wiper-like anti-phase (180°) finger wagging to in-phase (0°) wagging, suggesting the former was less stable than the latter under certain conditions. More specifically, while both anti-phase and in-phase coordination are quite sustainable at lower movement frequencies (and both far-more stable than, say, 90° or 130° phase angle coordination), only in-phase coordination is sustainable at higher frequencies. Abrupt transition from anti-phase to in-phase coordination is characteristic of

a nonlinear dynamical system, in which small changes in movement frequency yield qualitative changes in behavior of the system, such that certain ATTRACTORS within the system (e.g. 180°, 0°) become less stable. Haken et al. model this behavior in terms of a potential function $V(\mathrm{u}) = -a\cos(\phi) - b\cos(2\phi)$, which corresponds to the superimposition of two cosine waves at a 1:2 frequency ratio. The CONTROL PARAMETER (movement frequency) is said to correspond to a ratio of the amplitudes of the components of the potential function b/a (Figure 4.3). As the control parameter is increased, this ratio changes and the once-stable anti-phase coordinative mode loses stability; this can be seen in Figure 4.3 as the 'basins of attraction' at $\phi = 180°$ (π) become shallower with increasing frequency, eventually inverting and losing stability.

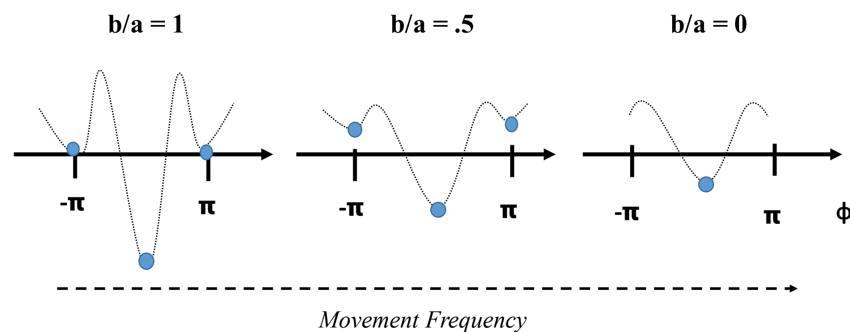


Figure 4.3: Changes in stability of attractors with increasing movement frequency based on Haken *et al.* (1985)

In what follows, we discuss ways in which similar notions of coupling, phase shifts, and entrainment have been proposed to account for various aspects of speech.

4.2 Dynamical Patterns of Gestures and Syllables: Articulatory Phonology

One well-known application of dynamical principles to language has been within the framework of Articulatory Phonology (AP) (Browman & Goldstein, 1986, 1989, 1990). In Articulatory Phonology, which is originally couched within the task dynamic model developed by Saltzman (1986) and Saltzman & Kelso (1983), individual articulators—the lips, tongue, teeth, etc.—are said to form

GESTURES, or task-oriented action units. As an example, the lip closure gesture for a voiceless aspirated bilabial plosive [pʰ] involves the reduction in space between the two lips two zero. This labial gesture is also coordinated with a glottal opening gesture which facilitates aspiration. This is not the whole story, however. In the context of a syllable or word such as *peak* ([pʰik]), articulatory evidence suggests that a different coordination pattern exists between the initial consonant and the following vowel than between the vowel and the coda consonant. Specifically, the gestures involved in the initial consonant and the following vowel are initiated simultaneously, or IN-PHASE, while those between the vowel and the coda consonant are coordinated in sequence, or ANTI-PHASE (Goldstein, 2007). For this reason, utterance units in AP are described as *constellations* of (potentially overlapping) gestures. These units are context-independent, but their dynamical nature—in which the temporal and spatial properties of individual articulator trajectories can vary in principled ways based on factors such as speaking rate—allows for the natural emergence of phonetic variation. With enough perturbation, the system can also demonstrate qualitative changes in organization. For example, syllable repetition studies have found that anti-phase coupling between vowels and codas in VC syllables such as *ab* becomes unstable at high speaking rates for English speakers. Just as with the finger wagging task, once they reach a certain rate of repetition, speakers exhibit an abrupt shift in production towards in-phase CV *ba* syllables (Stetson, 1951; Tuller & Kelso, 1991).

4.3 The Prosodic Hierarchy as a System of Coupled Oscillators

Cummins & Port (1998) take a similar dynamically-based approach to the study of speech rhythm, in which the prosodic hierarchy is modeled as a set of PLANNING OSCILLATORS which operate at different frequencies and which are coupled together in time. In the case of English, for example, the highest frequency might correspond to a mora- or syllable-level oscillator, then at progressively lower frequencies, the foot level, prosodic phrase level, etc. Crucially, cross-frequency coupling is argued to lead to modulation of lower-level oscillations such that they occur in-phase with higher-

level oscillations. In a related model, O'Dell & Nieminen (1999, 2008) demonstrate how changes in the relative coupling strength of 'stress group' vs. 'syllable' level oscillators can reflect the change in duration of a stress group as a function of number of syllables in the group. Coupled oscillator models of prosodic structure capture the cross-linguistic observation that levels within the prosodic hierarchy appear nested within one another: syllables within feet, feet within phrases, and so on. The model also captures the widely-held observation that prominent elements at one level—such as the head of a foot and an intonational phrase—often occur in-phase with one another, frequently leading to cumulative effects of acoustic prominence such as increased syllable or segment lengthening and f_0 modulation (Byrd & Saltzman, 1998; Byrd *et al.*, 2006; Krivokapić & Byrd, 2012).

In the speech cycling task, lower-order (e.g. $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$) fractions of the repetition cycle are thought to correspond with prominent events—such as stressed syllables—due to the fact that attentionally salient acoustic events are coordinated in-phase with periodic neurocognitive attentional pulses which guide speech perception and motor coordination (Port, 2003; Jones & Boltz, 1989; Large & Jones, 1999). Thus, foot-level oscillators in English occur at equal intervals, either phase-locked one and two thirds of the way through the higher phrase-level cycle (for ternary feet) or halfway through the phrase (for binary feet) (Figure 4.4).

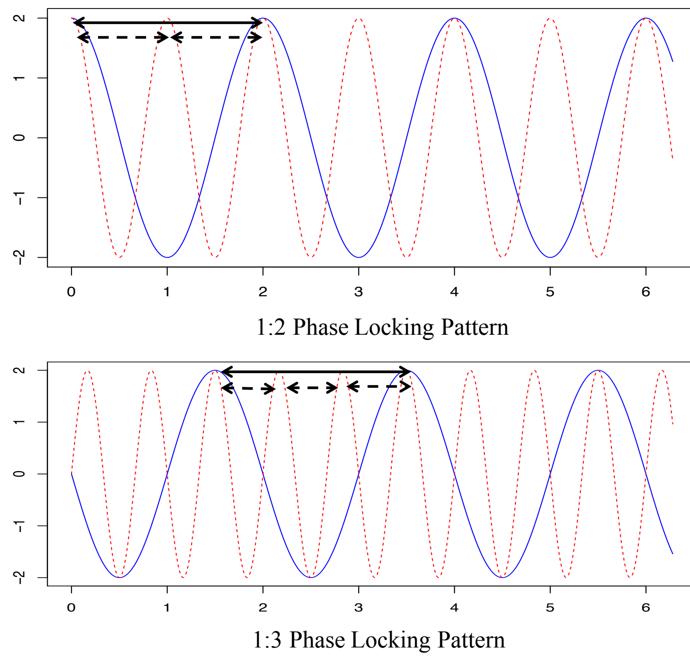


Figure 4.4: Oscillators at foot (red) and syllable (blue) levels at 1:2 and 1:3 frequency locking ratios (adapted from Salzman et al. 2008)

Just as increased speech rate leads to shifts in stable patterns of coordination at the syllable level, so, too, does it affect coordination at the phrase-level. For example, Cummins & Port found that the final stressed syllable of an English phrase such as *big for a <u>duck</u>* would typically occur two thirds of the way through the repetition cycle (the waltz-like pattern, ternary) at slower speech rates, but for all participants, it would shift to the halfway point in the cycle (the 'cut-time', two-beat pattern) once speech rate was increased sufficiently. Often subjects were found to exhibit periods of instability or PHASE DRIFT (Vorberg & Wing, 1996) once speech rate became too fast to maintain the initial waltz-like pattern. Before too long, however, subjects would resume in-phase coordination once they switched over to the two-beat pattern. Histograms such as those in Figure 4.5 show clear patterns of entrainment such that individuals were pulled back towards these lower-order, harmonic intervals.

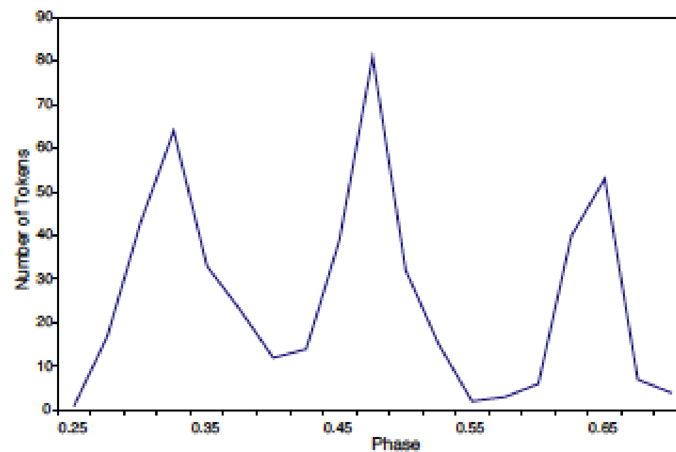


Figure 4.5: Distribution of phase alignment patterns for *Dig for a duck* in speech cycling from Cummins & Port, 1998, as shown in Port 2003: 601

The most comprehensive investigation of speech cycling in a language other than English is Tajima (1998), which explored speech cycling patterns in Japanese. In theoretical literature, Japanese has been hypothesized to have a mora-timed rhythmic structure (Bloch, 1950; Han, 1962; Ladefoged, 1975), in contrast with English, which is hypothesized to be stress-timed (Pike, 1945).

As discussed in Chapter 3, the main phonetic correlate which differentiates these rhythm classes is relative syllable or mora duration. English has quite variable syllable duration between stressed and unstressed syllables (Grabe & Low, 2002), and allows for a large amount of temporal compression of syllables as a function of word length. Japanese moras are relatively evenly timed (Grabe & Low, 2002) and do not undergo much compression with increasing word length (Han, 1962; Hoequist, 1983; Port *et al.*, 1987).

Relative rhythm class has been found to have a clear impact of speech cycling results, particularly when it comes to the position prominent syllables within the phrase repetition cycle as speech rate increases. Tajima found that, in contrast with English, final syllables in Japanese would move steadily later within the phrase repetition cycle as speech rate increased. This difference can be clearly seen in Figure 4.6.

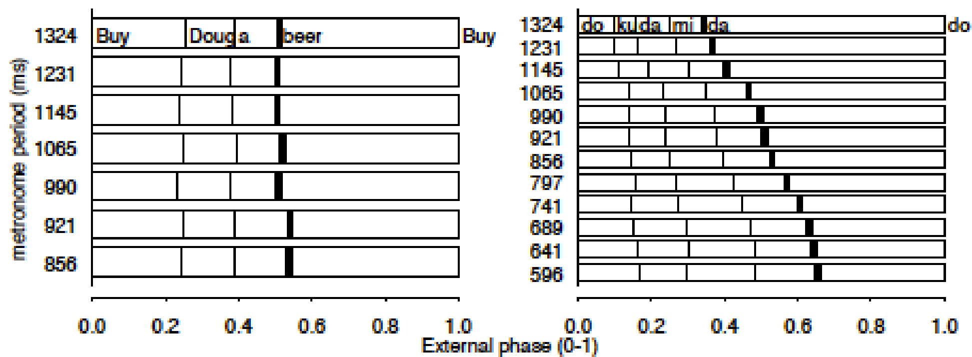


Figure 4.6: Comparison of external phase patterns with increasing speech rate: English (stress-timed) vs. Japanese (mora-timed) (from Tajima 1998: 35)

Thus, while English speakers typically allow for considerable syllable compression in order to time the final stressed syllable with harmonic positions, Japanese speakers do not exhibit the same level of compression. Instead, the Japanese speakers exhibited relatively more frequent transitions in and out of stable modes of coordination with harmonic positions. Despite this, cluster analyses Tajima performed revealed that the phase positions Japanese speakers coordinated with most stably were the same as in English: 1/3, 2/3, and 1/2. As will be seen in the chapters that follow, Medʉmba, similar to Japanese, demonstrates relatively little compression of syllables as

speech rate increases, indicating that it behaves as a canonical syllable timed language. Also like Japanese, Medʉmba speakers will be shown to display relatively longer transition periods between stable modes of coordination, as compared with English speakers.

4.4 Evaluating Stable Modes of Coordination: Gaussian Mixture Models

As has been discussed, key to understanding entrainment behavior in speech cycling is the ability to establish stable modes of syllable coordination within the task as speech rate is manipulated. We can make the prediction based on extensive previous work that syllable phase position does not show a perfectly linear relationship with speech rate; rather, relative phase position will likely reflect a set of stable modes of coordination. We can therefore test the hypothesis that each distribution is better modeled in terms of a mixture (or weighted sum) of distributions, each associated with a coordinative mode. Furthermore, based on previous work, these distributions are likely to be roughly normal. In order to leave open the possibility that Medʉmba speakers may show more (or fewer) modes of coordination than speakers of other languages, rather than specify *a priori* the location of stable modes, we can infer them from the data using GAUSSIAN MIXTURE MODELS (GMMs). GMMs allow for probabilistic modeling of previously unidentified subgroups (for our purposes, phase positions representing stable coordination) within the data and are thus well-suited to the kind of problem we are faced with.

The reader is encouraged to consult Lindsay (1995) and Marin *et al.* (2011) for more comprehensive background to supplement the following brief introduction to GMM modeling. GMMs have been used since as far back as the 19th century, their properties elaborated most thoroughly in the work of Pearson (1894). GMMs are known for being both powerful and flexible when applied as a tool for probabalistic modeling of univariate and multivariate data. GMMs can be distinguished from other clustering algorithms such as *K*-means in that, rather than separating data into discrete groups via hard assignment, they assign a probability to each datapoint corresponding to the likelihood of its membership within a component distribution (Press *et al.*, 2007). Distribu-

tions may overlap, indicating possible membership in either cluster, for points located toward the tails of distributions. Figure 4.7 compares a single Gaussian distribution (left pane) with mean (μ), standard deviation (σ) and probability density (ϕ) parameters with a mixture model (right pane) composed of three component Gaussian distributions, each with a separate set of parameters. Mixing weights associated with component distributions sum to 1.

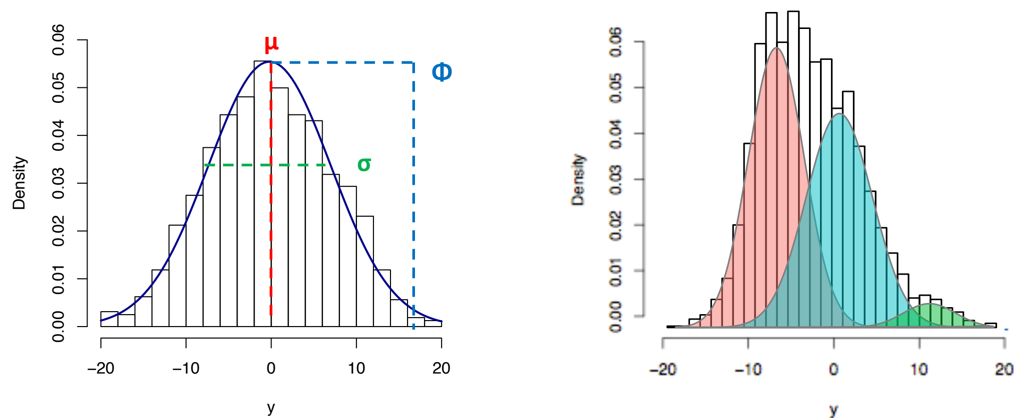


Figure 4.7: Parameters of a single Gaussian distribution (Left) and a Gaussian mixture (Right) with three components (corresponding to colored areas)

GMMs have a fairly long history for modeling linguistic behavior of various types, though they have not been applied as frequently to the study of speech timing. Exceptions include Tilsen (2006) who used GMMs in a similar way to the present study to model phase alignment patterns in 'threepeat' disfluencies, where individuals show ternary repetitions (e.g. *We went to the park* *and and and watched the birds*) found to occur at harmonic intervals with respect to one another and the larger utterance. Heo & Sethares (2015) and Rouas *et al.* (2005) apply GMMs to classify speakers and languages based on speech rhythm. GMMs have also been used frequently to model patterns of phonetic cue weighting in speech perception (Toscano & McMurray, 2008, 2010), as well as patterns of variation and change (Kirby, 2010) and acquisition of phonetic categories (de Boer & Kuhl, 2003; Vallabha *et al.*, 2007).

GMMs also take into consideration covariance structure among datapoints in clustering. The

method thus enhances our understanding of how each datapoint relates to the overall distribution. For the present research, GMMs will allow us to observe relative strength of entrainment of syllables within phrase repetitions in terms of their phase alignment, represented by the mean μ of each component distribution, and temporal variability, represented by the standard deviation σ of each component distribution.

4.4.1 Gaussian Mixtures: Basic Principles

The mixture model framework assumes a vector of random variables $\mathbf{X}_1,\ldots,\mathbf{X}_n$ sampled from a finite mixture m of arbitrary distributions (components), where $m > 1$. Each distribution $\mathbf{X}_i$ has the probability density function in (4.1), where λ_j stands for the mixing weights, or the probability that a randomly selected observation comes from component j.

$$g_\theta(\mathrm{x}_i) = \sum_{j=1}^{m} \lambda_j \phi_j(\mathrm{x}_i), \quad \mathrm{x}_i \in \mathbb{R}r \tag{4.1}$$

Here, $\theta = (\lambda, \phi) = (\lambda_1, \ldots, \lambda_m, \phi_1, \ldots, \phi_m)$ denotes the parameter and the mixing weights λ_j must be positive and sum to 1. We will assume here that the ϕ_j are drawn from a parametric family $\mathcal{F}$ of Gaussian distributions. The univariate Gaussian probability density function can be written as in (4.2).

$$\mathcal{F} = \{\phi(\cdot|\mu, \sigma^2) = \text{density of } \mathcal{N}(\mu, \sigma^2), (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}*^+\} \tag{4.2}$$

Model parameters can then be reduced to

$$\theta = (\lambda, (\mu_1, \sigma_1^2), \ldots, (\mu_m, \sigma_m^2)) \tag{4.3}$$

Mixture models were fitted using the `mixtools` package for *R* statistical software (Benaglia *et al.*, 2009). The methodology used in the package represents mixture modeling as a case of maximum likelihood estimation (MLE) in which observations are treated as incomplete data resulting

from non-observed complete data. Starting with an initial guess about the parameter values θ, an expectation-maximization (EM) algorithm is applied iteratively 1) to calculate posterior probabilities of component inclusion, conditional on the data and the current model parameters (referred to as the *E*-step) and 2) to re-estimated model parameters based on the conditional probabilities obtained (referred to as the *M*-step). Eventually, convergence is obtained, and data is assigned to a particular component based on a specified probability threshold. The mixing proportion associated with each component will them indicate the proportion of the input data assigned to that component.

4.4.2 Choice of Components: Parametric Bootstrapping of Likelihood Ratio Statistics

In order to evaluate the tradeoff between goodness of fit and model complexity (which increases as the number of components increases), an additional step involving parametric bootstrapping of likelihood ratio statistics (LRT) were carried out on nested models involving components $k = k_0$ versus $k = k_0 + 1$, up to a total of 4 possible components. The benefit of using LRT (over, for example, Akaike's Information Criterion or Bayesian Information Criterion) for model comparison is that one can evaluate at a chosen confidence level (here, $p < .005$) whether to reject the hypothesis that restrictions on model complexity are warranted. It is thus an appropriate method to apply when we have some sense about where different modes should emerge in the data and we want to test our hypotheses about the presence of such modes. LRT is also preferable in our case to another popular method for model comparison, cross-validation, in that it does not require evaluation based on 'real world' test data with associated outcomes predictions.

CHAPTER 5

EXPERIMENTS 1 AND 2: THE RELATIONSHIP BETWEEN STEM POSITION AND PROMINENCE

This chapter reports results from two speech cycling experiments used to evaluate the relative prominence of stem-initial syllables in Medʉmba. As discussed in Chapter 3, prominence in many languages is associated with stress, which can correlate with various acoustic cues, including increased duration, greater intensity, and higher (or lower) fundamental frequency. In addition to these suprasegmental cues to stress, languages often show effects at the segmental level. Allophonic processes affecting consonants and vowels often lead to patterns in which unstressed syllables display only a subset of the contrasts that stressed syllables do. Such processes typically (though not always; see Gordon 2011 for an overview of exceptions) involve some form of reduction such that vowels become more centralized and consonants undergo lenition. In English, for example, vowels in pronouns reduce to [ə] in unstressed positions; similar patterns have been attested in many different languages (see Crosswhite 2004 for an overview). Likewise, consonants at the onsets of unstressed syllables often have 'weaker' forms: plosives weaken to fricatives or taps, and voiceless segments sonorize and become voiced, or otherwise increase in sonority. As an example, in English, /t/ and /d/ are both realized as the tap /ɾ/ when they occur intervocalically and preceding an unstressed vowel, as in words like *city* [ˈsɪɾi].

Segment distribution asymmetries related to reduction processes have helped to motivate the presence of stress in many languages, and indeed serve as an important acoustic cue to stress in some languages (Chrabaszcs *et al.*, 2014; Correia *et al.*, 2015). In what follows, we examine just such asymmetries in Medʉmba noun and verb stems and attempt to evaluate whether or not such asymmetries are associated with prominence relations.

5.1 Evidence for Stem-Initial Prominence in Medʉmba

Medʉmba stem-initial syllables realize more consonantal and vocalic contrasts than syllables in any other position. The full consonant and vowel inventories for the language are given in Figures 5.1 and 5.2, respectively.

	Labial	**Dental**	**Palatal**	**Velar**	**Labiovelar**	**Glottal**
Prenasalized stops	mʙ mb	nt nd		ŋk^{h} ŋg		
Prenasalized affricates		nts ndz	ntʃ ndʒ			
Stop	b	t^{h} t d		k^{h} k	k^{w}	ʔ
Nasals	m	n	ɲ	ŋ		
Trills	ʙ					
Fricatives	f v	s z		ɣ		
Affricates		ts dz	tʃ dʒ			
Approximants			j			
Laterals		l				

Figure 5.1: Medʉmba consonant inventory

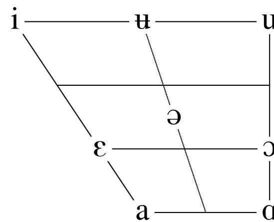


Figure 5.2: Medʉmba vowel inventory

Consonants				
Initial Stem (33)	Stem-medial (7)	Stem-final (7)	Prefixes (1)	Suffixes (1)
mʙ, mb, nt, nd, nts, ndz, ntʃ, ndʒ ŋk, ŋg, b, t^{h}, t, d k^{h}, k, k^{w} m, n, ɲ, ŋ, ʙ, f, v s, z, ʒ, ts, dz, tʃ, dʒ j, l	b, ʔ, l, ɣ, m, n, ŋ	p, t, k, ʔ, m, n, ŋ	n	d
Vowels				
Initial Stem (8)	Stem-medial	Stem-final (1)	Prefixes (1)	Suffixes (1)
a, ɑ, ə, ɛ, ɔ, u, ʉ, i	N/A	ə	ə	ə

Figure 5.3: Consonant and vowel distributions by stem position and affix type

The majority of non-compound stems are either (N)CV or (N)CVC monosyllabic or (N)CVCV disyllabic. In non-initial stem positions, only 7 of the 33 consonants and 1 of the 8 vowels are allowed (Figure 5.3). While stops are always realized as voiced stem-internally between vowels (and, in the case of velar and coronal stops, lenited to [ɣ] and [l], respectively), stops devoice in final position in (N)CVC forms. Vowels all reduce to schwa in non-initial syllables. Medʉmba displays few segmental affixes, but those that do exist exclusively contain the vowel [ə].

The distributional asymmetries of consonants have been shown by Danis (2011) to stem from at least in part from a prosodic word-internal lenition process, as demonstrated in examples like (8). One class of nouns which includes words such as *mʙɔ́ɣə́* 'fire' are realized disyllabically in isolation or in phrase-final position (8a), but monosyllabically phrase-initially or -internally (8b-8d).

(8) a. mʙɔ́**ɣ**ə́
'fire'

b. mʙɔ́**ɣ** ʉ̀m
'my fire'

c. mʙɔ́**k** ꜜʒúmə́ (*mʙɔ́ɣ ꜜʒúmə́)
The fire is hot'

d. mʙɔ́**k** ꜜÀriâne (*mʙɔ́ɣ ꜜÀriâne)
'Ariane's fire'

As seen in (8a), the velar stop /k/ is voiced and spirantized word-internally to [ɣ] within a stem and preceding a possessive pronoun, but does not undergo these processes (collectively referred to henceforth as 'lenition') word-finally, even before another vowel-initial word (8d). Danis proposed the prosodic structures in (9) for (8a-8d).

(9) a. $_{\omega}$(mʙɔ́ɣə́)$_{\omega}$

b. $_{\omega}$(mʙɔ́ɣ àm)$_{\omega}$

c. $_{\omega}$(mʙɔ́k)$_{\omega}$ $_{\omega}$(ꜜʒúmə́)$_{\omega}$

d. $_{\omega}$(mʙɔ́k)$_{\omega}$ $_{\omega}$(ꜜÀriâne)$_{\omega}$

As can be seen, lenition, per Danis' analysis, occurs intervocalically within a prosodic word, but not across prosodic words. Or, more precisely, lenition does not apply to consonants *at the right edge of a prosodic word.* This same generalization holds for verbs: compare (10a) and (10b), in which /k/ is lenited stem-medially and preceding a VC object pronoun, with (10c), in which /k/ occurs at the right edge of a prosodic word and does not undergo lenition.

(10) a. ká**ɣ**ə́ $_{\omega}$(káɣə́)$_{\omega}$
'release'

b. ká**ɣ** ɑ́m $_{\omega}$(káɣ ɑ́m)$_{\omega}$
'release me'

c. ká**k** ꜜÀriâne $_{\omega}$(kák)$_{\omega}$ $_{\omega}$(ꜜÀriâne)$_{\omega}$
'release Ariane'

An additional generalization not discussed by Danis is that consonants also do not lenite stem-initially after a CV prefix (11a), even when they are contained within the same prosodic word as the preceding and following vowels (11b).

(11) a. nə̀-**k**áɣə́ (*nə̀-**ɣ**áɣə́)
INF-release
'to release'

b. $_{\omega}$(nə̀-káɣə́)$_{\omega}$

The generalization thus seems to be that lenition can target intervocalic consonants within a prosodic word, but not if they are stem-initial. In order to avoid having a generalization which makes reference to morphological categories, we might propose that the right edge of a trochaic

Possessive Pronouns		Object Pronouns	
ɑ̀m	1.sg	ɑ́m	1.sg
ù	2.sg	ú	2.sg
í	3.sg	í	3.sg
ják	1.pl	ják	1.pl exclusive
zín	2.pl	bə̀n	1.pl inclusive
júp	3.pl	zín	2.pl
ʧám	1.pl, plural possessum	júp	3.pl
ʧú	2.pl, plural possessum		
tsə́	3.pl, plural possessum		

Table 5.1: Subset of Medʉmba possessive pronouns and object pronouns (Voorhoeve 1967, adapted according to modern IPA conventions)

foot is aligned with the right edge of the prosodic word, and that lenition is banned from targeting segments within the foot head (underlined) (12).

(12) $_{\omega}(\sigma \ _{FT}(\underline{\sigma} \ \sigma)_{FT})_{\omega}$

Note that pronominal enclitics such as the possessive *ɑ̀m* in (8b) (which varies in shape based on the class of the noun it modifies) and the object pronoun ɑ́m in (10b) form an interesting middle ground between stem-initial and non-initial syllables, as they follow the consonant lenition patterns of non-initial stem syllables, but are able to display nearly the full range of vowel qualities available for initial stem syllables. An abbreviated list of pronominal forms is given in Table (5.1) to demonstrate variation in vowel quality. Furthermore, as seen in (10b), downstep is blocked from occurring between a verb and pronominal enclitic, while it does occur between a verb and a full noun complement (10c). Downstep is in fact another diagnostic Danis uses for characterizing the prosodic word domain: while the process may occur across words, it may not occur within words. Thus, pronominal enclitics appear to behave similarly to stem-final syllables in terms of lenition and downstep behavior (evidencing their lack of status as independent prosodic words), but do not undergo vowel reduction in the way that stem-final syllables do. They thus form an intriguing structure to investigate as far as their relative prominence.

Finally, some positional restrictions on tone also suggest stem-initial syllables in Medʉmba are

associated with greater prominence. Medʉmba contrasts high and low tones, as well as falling contour tones on loanwords and some native words[1]. Though high and low tones can occur freely on stem-initial and non-initial syllables, contour tones are more restricted. Since few native Medʉmba words carry lexical contours, we must turn to loanwords from English for evidence of this restriction. Typically, CVN syllables from the source language are borrowed as HL contours (13a-13b).

(13) a. ŋgɔ̂n
'gun'

b. tâm
'time'

Where a nasal-final syllable occurs in a non-initial stem position, however, an epenthetic vowel is added to the end of the word to create an extra syllable, and the HL contour is distributed over the final two syllables (14).

(14) dɔ́sínì
'dozen'

A similar process occurs in cases where a compound is borrowed as a single stem 15.

(15) fríbánə̀
'frying pan'

Since contours appear to be perfectly fine before a pause as shown in (13a) and (13b), the generalization seems to be that HL contours are dispreferred specifically on non-initial stem syllables.

In sum, I have presented evidence for phonological and distributional asymmetries between stem-initial and non-initial/non-stem syllables. In the experiments that follow, I will use the speech cycling paradigm to investigate whether or not these asymmetries may correlate with differences

1. Contour tones can also form when a floating tone morpheme (either high or low) docks to a syllable bearing a different tone. As Franich (2015) shows, these instances of contour tones, in contrast to falling tones found on loanwords, behave as a sequence of two level tones.

in relative prominence, similar to the difference between stressed versus unstressed syllables in English or accented versus unaccented moras in Japanese. As outlined in Chapter 4, greater prominence from a speech cycling perspective is viewed in terms of syllable alignment in repeated phrases. Specifically, if stem-initial syllables are indeed more prominent than non-initial syllables and clitics, we should predict them to occur more frequently at *Simple Harmonic Phases* (SHPs) of the repetition cycle, which correspond to lower-order fractions such as the $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{2}{3}$ fractions of the cycle. Starting with Experiment 1, I will first present the stimuli to be investigated, as well as the participants and procedures involved, and then will review key hypotheses. Subsequently, I will present a description and discussion of experimental results. A similar format will follow for Experiment 2.

5.2 Experiment 1: The Relative Prominence of Stem-Initial versus Stem-Final Syllables

Experiment 1 aimed to look at the relative alignment of stem-initial syllables in nouns and verbs as compared with syllables of similar shapes occurring in stem-final position.

5.2.1 Stimuli

Stimuli for Experiment 1 consisted of sentences containing four syllables. The first syllable of each phrase was one of three nouns of shape CVN, including *mɛ́n* 'child', *bún* 'children', or *sáŋ* 'bird', and the second syllable was always the same, the recent past tense marker *ʧák*. Manipulations for Experiment 1 were concentrated on the fourth syllable. As seen in Table 5.2, two Form conditions were contrasted: **Stem Initial Stem Final** (SISF), in which the final word of the phrase was a disyllabic verb ending in a schwa, and **Stem Initial Stem Initial** (SISI), in which the penultimate syllable was a monosyllabic verb and the final syllable was a monosyllabic noun, the object complement of the verb. Thus, the penultimate syllable of each condition was a verb stem-initial

syllable, but the final syllable differed in whether it was stem-initial (for the SISI condition) or final (for the SISF condition). The prediction is that Syllable 4 of the SISI condition will behave as the most prominent syllable in the phrase, occurring close to Simple Harmonic Phase positions. In the SISF condition, we predict that Syllable 3, instead, will be the most prominent syllable, since Syllable 4 for that Form is in stem-final position.

Experiment 1

Form	IPA Transcription	Gloss	Translation
SISF	mɛ́n ʧák **mbi**bə́	child TNS wait	'The child waited.'
SISI	mɛ́n ʧák **mʙi ba**	child TNS lose bicycle	'The child lost the bicycle.'
SISF	bún ʧák **nʃú**mə́	children TNS go.out	'The children went out.'
SISI	bún ʧák **ntum mɛ́n**	children TNS accuse child	'The children accused the child.'
SISF	sáŋ ʧák **mfʉ́**lə́	bird TNS fly	'The bird flew.'
SISI	sáŋ ʧák **njʉ́ lá**	bird TNS eat pineapple	'The bird ate the pineapple.'

Table 5.2: Stimuli for Experiment 1. Syllables hypothesized prominent are in bold.

5.2.2 Participants, Procedure, and Data Processing

All data were collected in Bangangté, the central city around which Medʉmba is spoken, and Yaoundé, the capital of Cameroon (roughly 250 km southeast of Bangangté), during field trips in the winter and fall of 2015. Participants in Experiment 1 included 14 Medʉmba speakers, seven male, seven female. Ages of participants ranged from 19 years to 52 years. All were native speakers who had spent the majority of their lives in Bangangté or the villages surrounding it, where Medʉmba is the primary language of communication and commerce.

Participants were seated at a table in a quiet hotel room in front of a Macbook Pro 13" laptop. All were fitted with a Shure SM35 head-mounted condenser microphone which rested just over their ears and attached around the back of the head. Over this they wore a pair of Sony MDR 7506 studio headphones through which the metronome beats were played. The metronome sound consisted of a synthetic drumbeat created in Audacity, an open-source program for sound editing. A sample of this sound can be found at `http://home.uchicago.edu/kfranich/drumbeat`.

Each of the six target sentences were elicited at 15 different speech rates, from slowest (Speed

1) to fastest (Speed 15): this resulted in a total of 90 trials per participant. The slowest speech rate corresponded to a 1600 ms metronome period, which was estabished after pilot testing with several subjects indicated they could not speak slower than this and maintain coordination with the metronome. Following Tajima 1998, the period was reduced by 3% for each subsequent speed, such that the fastest speed corresponded to a 579 ms metronome period. For each trial, subjects heard a total of twelve clicks of the metronome. They were asked to listen to the first four beats so as to acclimate to the new speed and then begin repeating on the fifth beat, saying the sentence once per beat.

The target sentence was displayed on the computer screen in Powerpoint in white font against a black background in both French and Medʉmba, in the native Medʉmba orthography. Most subjects could not read Medʉmba, and many struggled to read the French as well. Thus, for each new sentence, the experimenter read the sentence in both French and Medʉmba and asked the subject to repeat both forms back to verify they had understood what the target sentence was.

Subjects underwent several practice trials prior to starting the experiment. Practice trials involved a separate set of sentences, and were all conducted using the slowest metronome rate. Once the subject felt comfortable with the task, the experimenter advanced to the experimental trials. During the course of the experiment, subjects were given periodic breaks to rest and drink water.

Data were annotated semi-automatically using the beat extractor method developed by Cummins (1997b) and Scott (1993) and implemented in Praat using the BeatExtractor script written by Barbosa (2003). The script is designed to insert boundaries at each *perceptual center* (p-center), or the instantaneous 'beat' where listeners perceive a syllable to occur (Morton *et al.*, 1976); this point typically lies close to the vowel onset in Medʉmba (Franich, under review). The script works by applying a second-order Butterworth filter to the speech signal, after which the signal is rectified and low-pass filtered. 'Beats' are then inserted, in the form of TextGrid boundaries, at points corresponding to the local maxima of the first derivative of each amplitude envelope. Annotations were subsequently hand-corrected, and spurious boundaries removed. Datapoints corresponding

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

P-center locations were extracted using Praat. Timing of each syllable (referred to henceforth as the syllable's *relative phase*) was calculated relative to the phase repetition cycle (PRC), or the time of the interval spanning successive repetitions of the target sentence, measured from the p-center of the first syllable in each repetition. This measurement is demonstrated in Figure 4: the interval of *b*, which extends form the beginning of the repetition cycle to the start of the second syllable, is divided by the interval of *a*, the duration of the entire repetition cycle (beginning of Syllable 1 to the beginning of the following Syllable 1). This gives the External Phase measure for Syllable 2; similar measures are taken for Syllable 3 and Syllable 4.

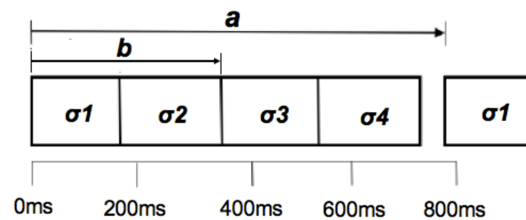


Figure 5.4: Sample measure of External Phase. External Phase for Syllable 2 is the interval of *b* divided by the interval of *a*.

5.2.3 Analysis Plan and Hypotheses

To review from the discussion in Chapter 4, there are various patterns which were of interest in analyzing the data from the speech cycling task. First, it was of interest to see whether or not a linear relationship held between speaking rate (as determined through metronome speed) and relative phase of target syllables, as this relationship can shed light on the status of a language as being more 'stress-timed' versus more 'syllable-timed' (or 'mora-timed'). Specifically, mora- and stress-timing were expected to associate with a stronger linear relationship between speed and phase position, particularly for syllables occurring later in the sentence (Tajima, 1998).

Second, repetitions of target syllables—in this case, Syllables 3 and 4—were expected to cluster at certain points in the phrase repetition cycle, including Simple Harmonic Phase positions of .333, .500, and .667. Subjects were also expected to show PHASE SHIFTS, such that more than one stable mode of alignment would be found for a given syllable. It was expected, then, that data for each syllable might deviate from a normal distribution, and instead tend towards a multi-modal distribution. To evaluate whether data for a given syllable deviated from a normal distribution, Anderson-Darling (A-D) tests of normality were run for each syllable in each Form condition. The A-D test of normality evaluates whether a normal distribution can adequately describe a given set of data. The test is powerful enough to detect most types of deviations from normality (Stephens, 1974), with larger A-D statistics associated with greater deviation from normality. A-D tests were implemented using the `nortest` package for *R* (Gross & Ligges, 2015). The reader is directed to documentation for this package for more information.

If data for a given syllable deviated from a normal distribution, Gaussian Mixture Models were used to evaluate whether the data was better fit using a mixture of normal distributions. Parametric bootstrapping of log likelihood ratio statistics was carried out as described in §4.4.2 to evaluate tradeoffs between model fit and complexity for models composed of between 1 and 4 components; resulting models were then compared. While all component distributions for target syllables 3 and 4 are reported, direct evaluation of differences in alignment for each pair of components across syllables would be quite cumbersome. Assuming even only 2 components per mixture, this amounts to 4 separate analyses per syllable (2 components $\times$ 2 Form conditions). For this reason, in order to quantify differences in phase position of target syllables across Form conditions, a value was calculated which will be referred to to as *component strength* (or *CS*) for each component distribution in the mixture, which consists of the component's standard deviation divided by its mixing weight (5.1).

$$CS = \sigma k_i / \lambda k_i \tag{5.1}$$

CS is a measure of tradeoff between variability (measured as standard deviation) and data coverage (measured as mixing weight ratio) and is meant to highlight those components which account for the most amount of data while maintaining a relatively low level of variability, making them less likely to be fitting noise in the form of e.g. repetition disfluencies or marginal patterns utilized only a few times.[2] Components with lower *CS* values are considered 'stronger'. DOMINANT COMPONENTS for each distribution—or those components with the lowest *CS*—were then extracted and differences in phase position between these distributions evaluated using linear mixed effects modeling. Note that the goal of this analysis was to compare the position of syllables occurring in relatively *comparable* positions across forms to see if syllables in corresponding positions genuinely showed different alignment patterns across Form conditions. Thus, if dominant components to be compared across Form conditions were not the most comparable in alignment across the two conditions, additional comparisons of non-dominant (but more closely-aligned in terms of relative phase position) distributions were carried out. Finally, based on findings from Tajima (1998), in a syllable- or mora- timed language, syllables which occur later in the sentence generally have higher standard deviations. This reflects the fact that coordination of syllable is stronger closer to the metronome beat itself, and becomes weaker as syllables to repeat become farther from the beat. Thus, *CS* values were also predicted to be larger for later syllables than for earlier ones. Since we are primarily concerned with comparing component strength *within* a given syllable position across Form conditions, this difference in relative *CS* values across syllable positions is not a cause for concern.

Specific hypotheses as to alignment patterns for target syllables in Experiment 1 can be summarized as follows:

1. Syllable alignment was predicted to cluster around a relatively small number of phase positions

2. Note that (Cummins & Port, 1998) utilize a similar concept of tradeoff measure in evaluating 'optimal' clustering for English speech cycling data.

2. Alignment of target syllables was predicted to show certain patterns. Specifically:

   (a) Stem-initial syllables in Syllable 4 position of the SISI condition were predicted to align more closely ($<.03$ phase[3]) with SHP positions of .333, .5, or .667 than were stem-final syllables of the SISF condition.

   (b) Assuming variation in relative prominence across stem-initial syllables might occur based on speech rate, it was predicted that verb stem-initial syllables occurring in penultimate position (Syllable 3) might also show close alignment with SHP positions.

5.2.4 Results

One subject displayed great difficulty in coordinating with the metronome beat, and did not appear to vary her speech rate according to the metronome period. Since it was not clear whether the subject fully understood the task, the subject's data were excluded from analysis. What follows is an analysis of the data of the 13 remaining subjects.

The first pattern to note in Figure 5.5 is the relatively high rate of change in relative phase across speech rates. This is further confirmed by looking at the by-syllable and by-form plots with regression lines and coefficients in Figure 5.6. Though a significant linear relationship was found for all syllables across forms ($p < .001$ for all), In other words, as the metronome speed increased, relative phase proportion increased, as well. The strongest effect of speed could be found for the syllables occurring later in the utterances, Syllables 3 and 4 ($R^2 > .40$). This is reminiscent of findings from Japanese from Tajima (1998) who compared speech cycling behavior between speakers of Japanese, a prototypical mora-timed language, and speakers of English, a prototypical stress-timed language. As mentioned in Chapter 4, Tajima found that English speakers were more

3. While some deviation in alignment from specified SHP positions is expected, alignment of prominent syllables is predicted to fall close to these positions. 'Closeness' is a subjective notion, however. Since a primary goal here is to see whether speakers of Medʉmba behave similarly to speakers of other languages on the speech cycling task, I draw on previous speech cycling results (Cummins, 1997a; Tajima, 1998; Cummins & Port, 1998), which suggest that a sufficiently conservative cutoff for 'closeness' appears to be around .03 phase.

likely than Japanese speakers to adjust the duration of syllables in order to consistently align the final stressed syllable of a phrase with the same position. Japanese speakers, on the other hand, showed incremental changes in phase position with increasing metronome speeds, especially for syllables occurring later in target sentences. Still, evidence showed that changes in phase were not equal across speech rates: Japanese speakers still preferred SHP positions to other phase positions when aligning accented syllables.

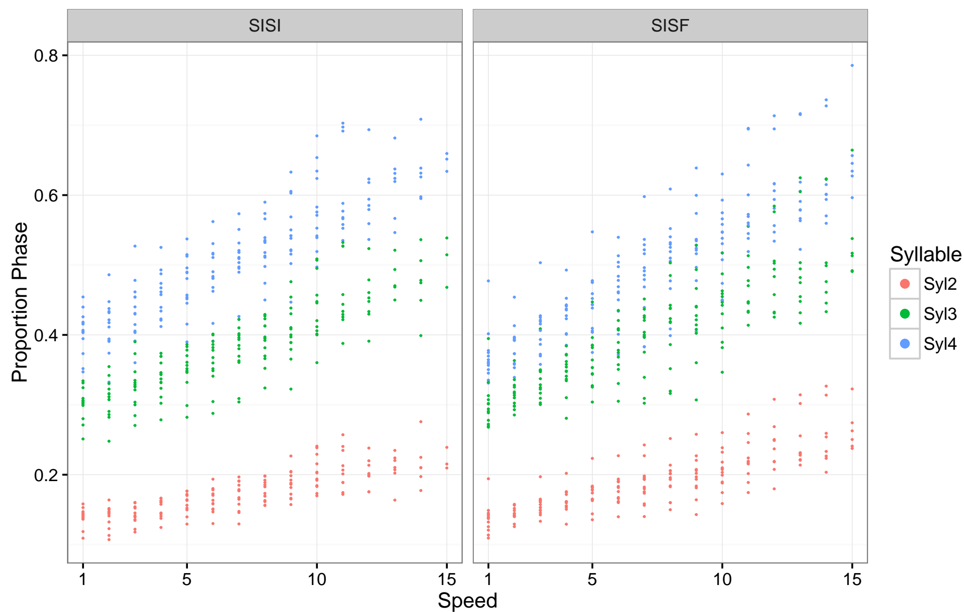


Figure 5.5: Scatter plot of phase angle by metronome speed (speech rate), Experiment 1 (red=Syllable 2, green=Syllable 3, blue=Syllable 4). Each point represents a single observation for a given speaker.

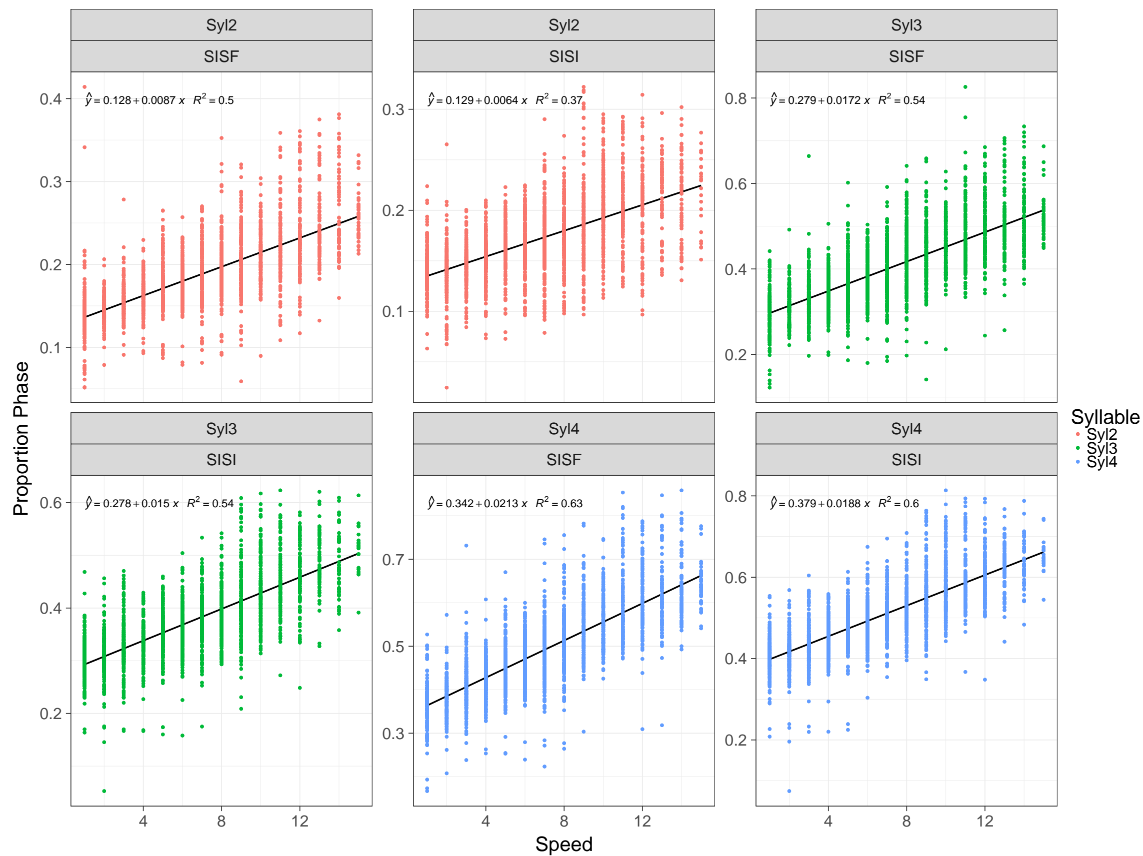


Figure 5.6: Distribution by syllable with regression lines, Experiment 1 (red = Syllable 2, green=Syllable 3, blue=Syllable 4)

Heatmaps in Figure 5.7 and 5.8 represent two-dimensional histogram showing distribution of values in the data across both metronome speed and phase position. Higher concentrations of repetitions at a given phase position and speed are indicated by warmer colors, such as red, orange, and yellow. Lower concentrations are represented by cooler colors, such as blue and green. As can be seen, despite the general trend we see of change in phase position by speed, syllable alignment appears to have clustered around certain points in the repetition cycle, in line with Hypothesis 1. This was especially true at metronome speed levels 8 and below, which were the speeds speakers reported being the most comfortable for the task.

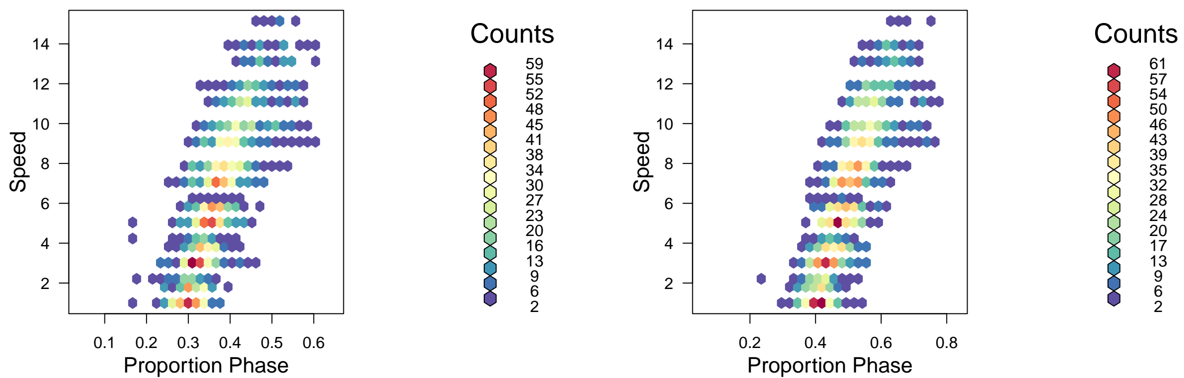


Figure 5.7: Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), SISI condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position

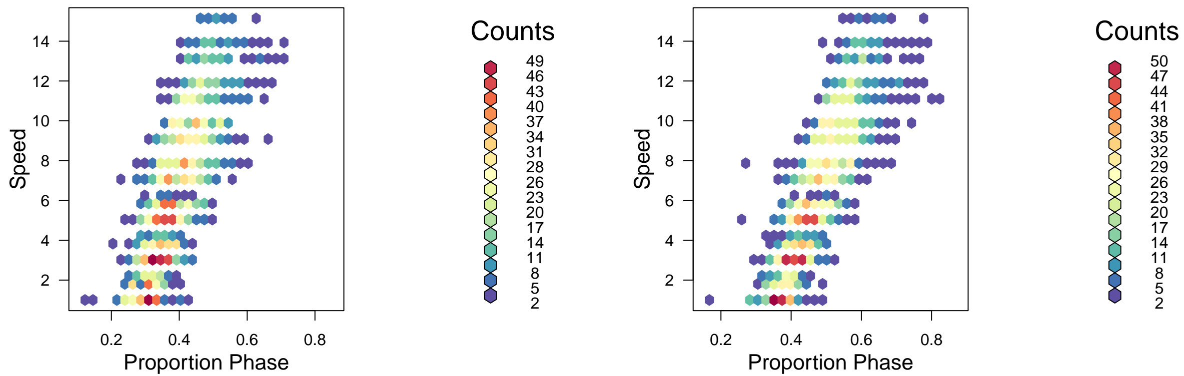


Figure 5.8: Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), SISF condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position

In order to verify that the distributions of target syllables were non-normal, Anderson-Darling tests of normality were conducted for each syllable in each condition. Results of these tests are presented in Table 5.3; as shown, target syllables in both Form conditions significantly deviated from normality ($p < .0001$).

Form	Syllable	A-D Statistic	p-value
SISI	Syl 3	49.559	$< .0001$
	Syl 4	3.985	$< .0001$
SISF	Syl 3	33.227	$< .0001$
	Syl 4	35.76	$< .0001$

Table 5.3: Results of Anderson-Darling test for normality, Experiment 2

Gaussian mixture models were fit separately for each Form (SISI vs. SISF), for Syllable 3 and Syllable 4, the two syllables of interest. In each case, bootstrapping of likelihood ratio statistics was done to test model fit against complexity for models with different numbers of components. For Syllable 3, two components were found to be justified at the $p < .005$ significance level in the SISI condition, and three components in the SISF condition. For Syllable 4, three components were found to be justified for both conditions. Observed likelihood ratio statistics and corresponding

p-vales from the parametric bootstrap procedure are provided in Table 5.4.

		SISI		SISF	
		Obs. LRS	p-value	Obs. LRS	p-value
Syllable 3	$K = 1$ vs. $K = 2$	217.766	$< .005$	151.273	$< .005$
	$K = 2$ vs. $K = 3$	1.375	$= .15$	3.6201	$< .005$
	$K = 3$ vs. $K = 4$	N/A	N/A	3.052	$=.53$
Syllable 4	$K = 1$ vs. $K = 2$	113.375	$< .005$	135.381	$< .005$
	$K = 2$ vs. $K = 3$	29.685	$< .005$	25.260	$< .005$
	$K = 3$ vs. $K = 4$	26.749	$=.04$	17.011	$=.63$

Table 5.4: Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K.

Final mixing weights λ (proportion of data attributable to a particular component), component means μ and standard deviations σ are provided for syllable (Syllable 3 and Syllable 4) in each Form condition in Table 5.5.

		SISI			SISF		
		μ	σ	λ	μ	σ	λ
Syllable 3	Component 1	.352	.056	.843	.343	.048	.510
	Component 2	.491	.056	.157	.453	.045	.213
	Component 3	N/A	N/A	N/A	.446	.112	.277
Syllable 4	Component 1	.474	.069	.813	.426	.063	.619
	Component 2	.528	.007	.019	.559	.057	.329
	Component 3	.627	.070	.168	.711	.059	.052

Table 5.5: Model parameters for Syllable 3 and Syllable 4: Experiment 2

Plots of mixture models by syllable and condition are shown in Figures 5.9-5.12. Means for each component (Component 1 in red, Component 2 in blue, Component 3 in green) are displayed to the right of each plot, and dotted red lines are drawn to indicate the nearest (within .15 of any given distribution) Simple Harmonic Fractions of the cycle.

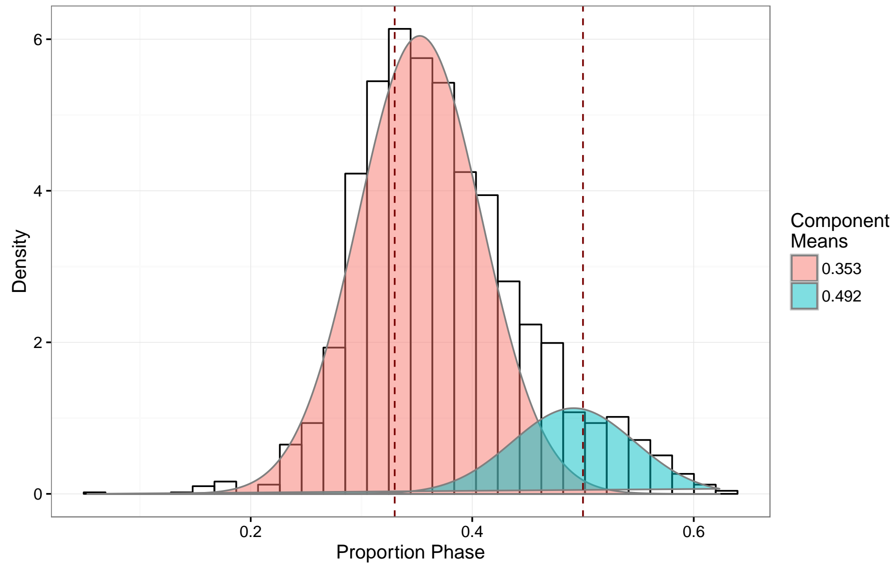


Figure 5.9: SISI Syllable 3: 2 Component Gaussian mixture (dotted red lines indicate SHP positions of .333 and .500)

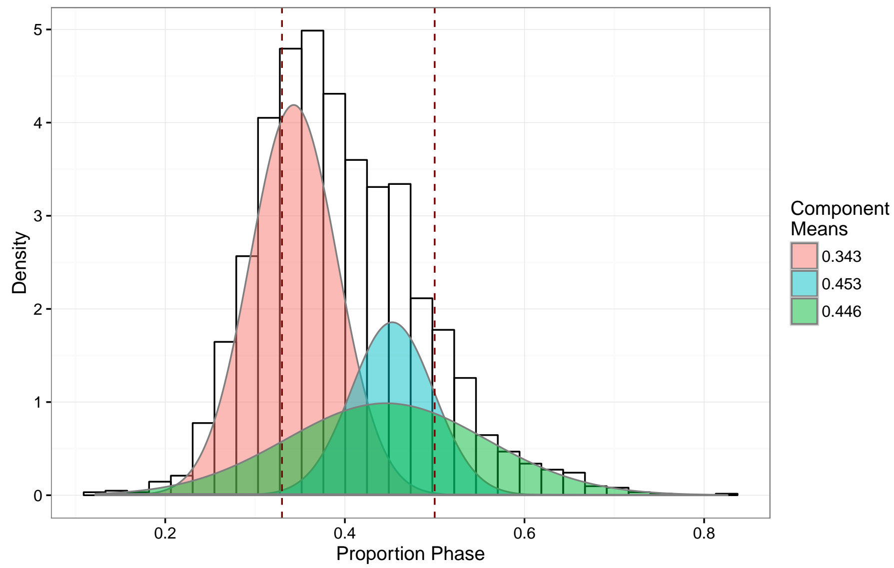


Figure 5.10: SISF Syllable 3: 3 Component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667)

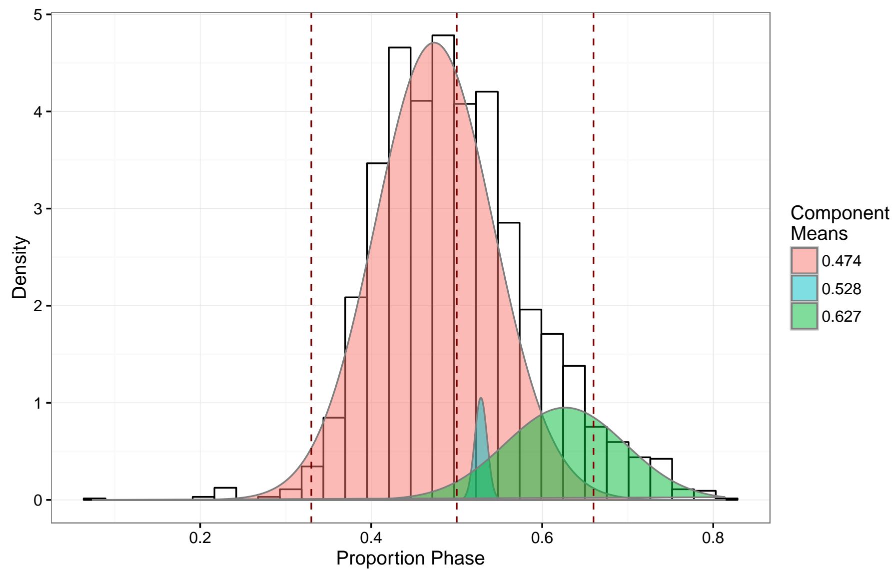


Figure 5.11: SISI Syllable 4: 3 Component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667)

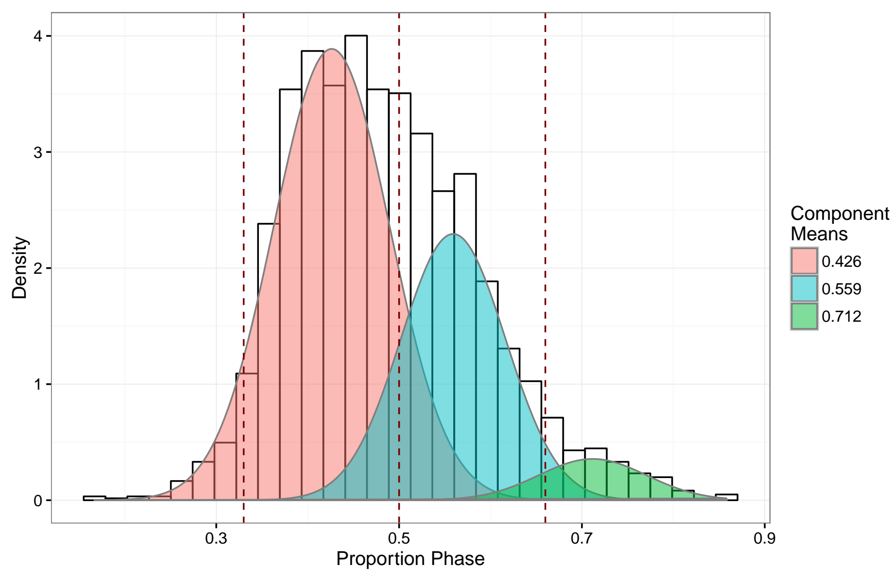


Figure 5.12: SISF Syllable 4: 3 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667)

Hypothesis 2, that subjects' behavior on the speech cycling task should reflect a preference for a relatively small number of stable modes of alignment, appears to be borne out from results of Experiment 1. Hypothesis 3 was that stem-initial syllables would occur at Simple Harmonic Phases (SHPs), or lower order fractions of the repetition cycle, such as $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{2}{3}$. In order to understand whether or not this prediction was supported, we will need to examine the means of each component across the two syllables of interest in each condition. In particular, the phase positions between the two syllables *relative to one another* will be important to understand. The reason for this is that, depending on the rhythmic mode speakers were using, there should be variation in the strength that certain positions exert in attracting prominence. For example, if the rhythmic mode involves a 'waltz-like' pattern in which the subject is subdividing the phase in three, phase positions of $\frac{1}{3}$ and $\frac{2}{3}$ (and not $\frac{1}{2}$) should be the strongest attractors of prominence. If, however, the subject is subdividing the phase in 2 for a 'cut-time' pattern, the phase at $\frac{1}{2}$ will be strongest. The difficulty arises in assessing which mode the subject was in, as the occurrence of a syllable at certain phase positions may be consistent with either subdivision strategy; thus, examining the position of Syllable 3 and Syllable 4 relative to one another will be more informative than considering the phase position of either syllable on its own. We consider the relative position of these syllables by Form condition, beginning with the SISI condition.

SISI Condition

In the SISI condition, Syllable 3 showed alignment modes with means of .352 (Component 1) and .491 (Component 2). These positions are relatively close ($<$.02 phase) to SHP conditions of $\frac{1}{3}$ and $\frac{1}{2}$, respectively. Syllable 4 in the SISI condition showed alignment modes with means .474 (Component 1), .528 (Component 2), and .627 (Component 3). While the first two components occurred relatively close ($<$.3 phase) to the $\frac{1}{2}$ SHP condition, the third component was farther away from an SHP position ($>$.35 phase).

The first component for both Syllable 3 and Syllable 4 accounted for between 80-80% of the

data, suggesting that subjects were relatively consistent in their placement of Syllable 3 around the $\frac{1}{3}$ SHP and Syllable 4 around the $\frac{1}{2}$ position. I will refer to this pattern of relative positioning as Pattern 1 (Figure 5.13). The second component for Syllable 4 in the SISI condition accounted for less than 2% of the data, but represented a small rightward shift in alignment for this syllable (though it still occurred close to the $\frac{1}{2}$ SHP position), while Syllable 3 evidently stayed around the $\frac{1}{3}$ SHP position. This pattern will be referred to as Pattern 2. Finally, the second component for Syllable 3 and the third component for Syllable 4 each accounted for the final roughly 16% of the data, and represented a shift of Syllable 3 from around the $\frac{1}{3}$ SHP position to the $\frac{1}{2}$ position, and Syllable 4 from the $\frac{1}{2}$ position to a non-SHP position (Pattern 3). Thus, all in all, Syllable 3 and 4 both showed attraction to prominence-attracting positions.

Crucially, the patterns shown in Figure 5.13 are not meant to reflect patterns of variation in alignment across subjects, but rather shifts in alignment positions with increasing speaking rate *within* subjects. Indeed, data from all subjects is represented in each of the components described modeled above; these patterns thus represent more general behavior across subjects.

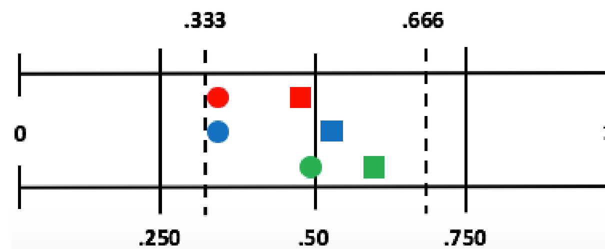


Figure 5.13: Patterns for Syllable 3 (circles) and 4 (squares), SISI Condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3.

Component Strengths were calculated for each syllable in each form by dividing each component's mixing weight λ by its standard deviation σ. Dominant components are indicated in bold in Table 5.6. As seen, Component 1 was the dominant component for both syllables in the SISI condition.

We now move on to discuss syllable alignment patterns in the SISF condition.

	Syllable 3	Syllable 4
Component 1	**.066**	**.085**
Component 2	.357	.368
Component 3	N/A	.417

Table 5.6: Component strengths, Syllables 3 and 4, SISI condition

SISF Condition

In the SISF condition, Syllable 3 showed alignment modes with means of .343 (Component 1), .453 (Component 2), and .446 (Component 3). While the first component occurred close (<.01 phase) to the SHP position of $\frac{1}{3}$, the other two component means were far (>.04 phase) from SHP positions. Syllable 4 showed mean alignments at .426 (Component 1), .559 (Component 2), and .712 (Component 3), all of which were far (>.05 phase) from SHP positions.

Component 1 for Syllable 3 and 4 accounted for just over half of the data for the SISF condition, indicating this was a relatively consistent pattern (Pattern 1; Figure 5.14) at lower speech rates. Here, Syllable 3 (but not Syllable 4) occurred at an SHP, that of $\frac{1}{3}$. Moving from Pattern 1 to Pattern 2, large shifts occurred in both Syllable 3 and Syllable 4, with neither landing near an SHP position. For Pattern 3, while Syllable 3 stayed in essentially the same position as for Pattern 2, the average phase position of Syllable 4 shifted quite a bit later, to yet another non-SHP position.

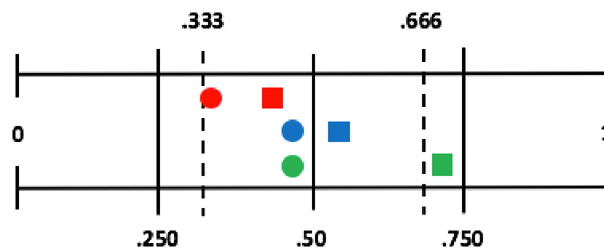


Figure 5.14: Patterns for Syllable 3 (circles) and 4 (squares), SISF Condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3.

In looking once more at component strengths in the SISF condition, we see that Component 1 was the dominant component for both Syllable 3 and Syllable 4 (Table 5.7). These dominant

components form Pattern 1, where Syllable 3 occurred at the $\frac{1}{3}$ SHP, and Syllable 4 occurred far from any SHP.

	Syllable 3	Syllable 4
Component 1	**.094**	**.102**
Component 2	.211	.173
Component 3	.404	1.135

Table 5.7: Component strengths, Syllables 3 and 4, SISF condition

Comparing Dominant Components Across Conditions

Comparing now across the SISI and SISF conditions, in Figure 5.15, distributions for each syllable are presented, with dotted lines indicating distribution means (blue for SISI condition, red for SISF condition). We see that the distributions of Syllable 3 stem-initial syllables in SISI and SISF occur quite close together, and both occur close to the SHP value of $\frac{1}{3}$, indicated with the solid black line. For Syllable 4, the stem-final syllable in the SISF condition appears to have a much earlier phase position than the same syllable in the SISI condition, and also appears to occur farther from the closest SHP position of $\frac{1}{2}$, again indicated with the solid black line.

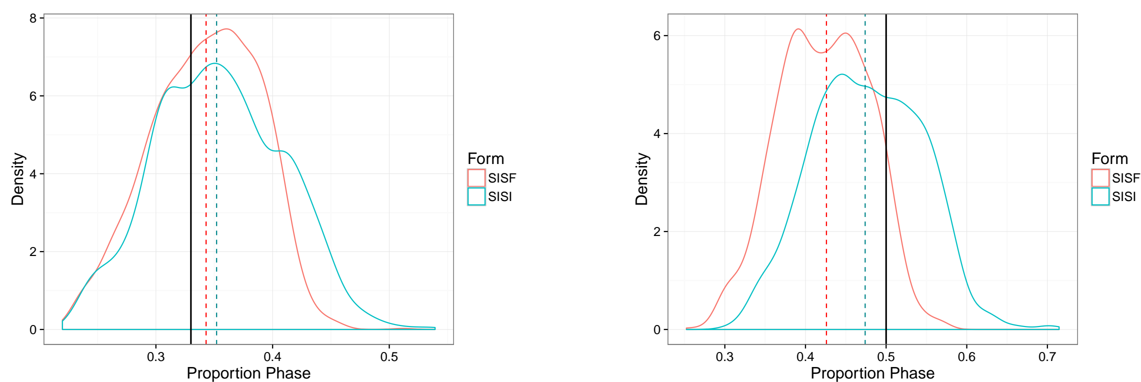


Figure 5.15: Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 1

To compare alignment of dominant components across conditions, a linear mixed effect model

was constructed with PROPORTION PHASE as the response variable, and which included fixed effects of FORM and METRONOME SPEED (as well as their interaction), as well as by-subject random slopes for FORM and METRONOME SPEED. The factor FORM was sum-coded and the variable SPEED was treated as continuous and was mean-centered. Results revealed no significant difference in alignment of Syllable 3 between SISI and SISF conditions (t = .418, df = 10, p = .684). For Syllable 4, the stem-final syllable in the SISF condition was found to occur significantly earlier than the stem-initial syllable in the SISI condition (t = -13.070, df = 12, $p < .0001$). A significant interaction was also found between FORM and SPEED for Syllable 4 (t = -5.665, df = 3357, $p < .0001$), reflecting the fact that phase position varied more as a function of metronome speed in the SISF condition than in the SISI condition.

Full results are provided in Table 5.8.

	Factor	**Estimate**	**Std. Error**	**df**	**t-value**	**p-value**
Syllable 3	Form	.001	.0014	10	.418	.684
	Speed	.002	.0018	11	8.797	$< .0001$
	Form × Speed	.000	.0005	2536	-.409	.683
Syllable 4	Form	-.021	.0016	12	-13.070	$< .0001$
	Speed	.039	.0032	12	12.276	$< .0001$
	Form × Speed	-.004	.0007	3357	-5.665	$< .0001$

Table 5.8: Results of linear mixed effects models Syllable 3 and Syllable 4, Experiment 1

5.2.5 Discussion

Several interesting patterns emerged in Experiment 1. The positive linear relationship between metronome speed and relative phase position, and in particular, the quite large regression coefficients for target syllables 3 and 4, suggest that Medʉmba behaves more like a prototypical mora- or syllable-timed language than a stress-timed language. As was shown in Figure 4.6 in Chapter 4 from Tajima (1998), English speakers tended to show less variability in phase alignment patterns with increasing speech rate, likely stemming from the increased availability in English of the process of polysyllabic shortening, which enables speakers to shorten the duration of syllables within a foot and prioritize regular timing of stress feet (Kim & Cole, 2005; O'Dell & Nieminen, 1999, 2008). While this pattern is not itself sufficient to conclude that Medʉba is a syllable-timed language, it is suggestive of such a pattern, and warrants further investigation. Since this pattern is not a central focus for the present research, I leave this for future work.

As far as alignment patterns were concerned, it was hypothesized that stem-initial syllables would display evidence of greater prominence than stem-final syllables in terms of their relative position with respect to Simple Harmonic Phase positions. Results were consistent with this hypothesis: stem-final syllables, similar to prefix syllables in Experiment 1, occurred significantly earlier than stem-initial syllables in the same position, and significantly farther from SHP positions. Experiment 1 has also shown that speech rate also plays an important role in determining which repetition patterns subjects settle on and which syllable behaves as more prominent. In particular, in the SISI condition, Syllable 3 occurred around the SHP position of $\frac{1}{3}$ where metronome speed/speech rate was slower (consistent with a three beat waltz-like pattern), but shifted up to the $\frac{1}{2}$ position at higher speech rates (consistent with a two beat 'cut-time' pattern).

Of further interest was the fact that, in Pattern 1 for SISI, both Syllables 3 and 4 occurred near SHP positions, suggesting subjects were treating both syllables as prominent simultaneously. This was an unexpected pattern, as it was thought that behavior within a 'waltz-like' mode should preclude alignment of prominent syllables at the $\frac{1}{2}$ position, and that behavior within a 'cut-time'

mode should preclude alignment of syllables at the $\frac{1}{3}$ and $\frac{2}{3}$ position. This finding will be taken up in more detail in §5.4.

5.3 Experiment 2: The Relative Prominence of Stem-Initial Syllables versus Prefixes and Enclitics

Experiment 2 aimed to look at the relative alignment of stem-initial syllables in nouns and verbs as compared with syllables of similar shapes occurring as prefixes and enclitics, two elements which have been shown to exhibit evidence of relatively lower prominence.

5.3.1 Stimuli

Stimuli used for the experiment (presented in Table 5.9) included three sets of four-syllable sentences paired for tonal and syllabic structure. The first syllable of each phrase was one of two nouns of shape CVN, including *mɛ́n* 'child' or *sáŋ* 'bird', and the second syllable was always the same, the verb *lɛ̌n*, 'know'. The final two syllables in each sentence varied by Form condition: in the **Prefix-Stem Initial** (PRESI) condition, a monosyllabic infinitival verb prefix occurred in penultimate position, followed by a monosyllabic verb stem in final position. In the **Stem Initial-Enclitic** (SIENC) condition, the penultimate syllable was a monosyllabic noun stem, followed by a monosyllabic possessive pronominal enclitic in final position. As described in §5.1, these pronouns look to form a single prosodic word with the preceding noun stem. Likewise, the infinitival prefix and verb stem in the PRESI condition share a single prosodic word. Thus, these constitute constructions which vary only in the hypothesized relative prominence of penultimate and final syllables (henceforth, Syllable 3 and Syllable 4, respectively): while Syllable 4 is predicted to behave as prominent in the PRESI condition (and Syllable 3 as non-prominent), Syllable 2 is predicted to behave as prominent in the SIENC condition (and Syllable 4 as non-prominent). Table 5.9 provides a list of sentences in each condition.

Experiment 2

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

Table 5.9: Stimuli for Experiment 2. Syllables hypothesized prominent are in bold. Clitic boundaries indicated with '=' and affix boundaries with '-'

5.3.2 Participants, Procedure, and Data Processing

All data were again collected in Bangangté and Yaoundé, Cameroon, during field trips in the winter and fall of 2015. Participants in Experiment 2 included 14 Medʉmba speakers, seven male, seven female. Ten of these subjects (s1, s2, s4, s5, s6, s8, s10, s11, s12, and s13) were from the same subject pool as was included for Experiment 1. Three of the original subjects (two men and one woman) were not available to participate in the second experiment. Three new subjects, two men and one woman were recruited to replace these subjects, and one additional female subject was recruited to replace the subject whose data was not able to be used for Experiment 1. Ages of participants ranged from 19 years to 58 years. All were native speakers who had spent the majority of their lives in Bangangté or the villages surrounding it, where Medʉmba is the primary language of communication and commerce.

Participants were seated at a table in a quiet hotel room in front of a Macbook Pro 13" laptop. All were fitted with a Shure SM35 head-mounted condenser microphone which rested just over their ears and attached around the back of the head. Over this they wore a pair of Sony MDR 7506 studio headphones through which the metronome beats were played. The metronome sound consisted of a synthetic drumbeat created in Audacity, an open-source program for sound editing. A sample of this sound can be found at `http://home.uchicago.edu/kfranich/drumbeat` .

Just as with Experiment 1, each of the six target sentences were elicited at 15 different speech rates, from slowest (Speed 1) to fastest (Speed 15): this resulted in a total of 90 trials per par-

ticipant. The slowest speech rate corresponded to a 1600 ms metronome period, and the period was reduced by 3% for each subsequent speed, such that the fastest speed corresponded to a 579 ms metronome period. For each trial, subjects heard a total of twelve clicks of the metronome. They were asked to listen to the first four beats so as to acclimate to the new speed and then begin repeating on the fifth beat, saying the sentence once per beat.

The target sentence was displayed on the computer screen in Powerpoint in white font against a black background in both French and Medʉmba, in the native Medʉmba orthography. Once again, in case subjects were not able to read the target sentences, the experimenter read the sentence in both French and Medʉmba and asked the subject to repeat both forms back to verify they had understood what the target sentence was.

Subjects underwent several practice trials prior to starting the experiment. Practice trials involved a separate set of sentences, and were all conducted using the slowest metronome rate. Once the subject felt comfortable with the task, the experimenter advanced to the experimental trials. During the course of the experiment, subjects were given periodic breaks to rest and drink water.

Data Processing

Data were annotated semi-automatically using the beat extractor method developed by Cummins (1997b) and Scott (1993) and implemented in Praat using the BeatExtractor script written by Barbosa (2003). Annotations were subsequently hand-corrected, and spurious boundaries removed. Datapoints corresponding to more than 2 standard deviations from the mean for a given syllable and metronome speed were removed as they most often reflected disfluencies in repetitions where subjects stumbled over a word.

P-center locations were extracted using Praat. Timing of each syllable (referred to henceforth as the syllable's *relative phase*) was calculated relative to the phase repetition cycle (PRC), or the time of the interval spanning successive repetitions of the target sentence, measured from the p-center of the first word in each repetition.

5.3.3 Analysis Plan and Hypotheses

Analysis for Experiment 2 proceeded as for Experiment 1. First, patterns of change in relative phase position by speed were evaluated for each syllable of each target sentence using linear regression. Second, distribution of data for each target syllable—in this case, Syllable 3 and Syllable 4—in each Form condition were evaluated to see whether or not they deviated from normality using Anderson Darling tests. Third, if the data for each syllable showed deviation from a normal distribution, patterns of clustering in the data were examined via Gaussian mixture models. Specifically, parametric bootstrapping of log likelihood ratio statistics was carried out to evaluate tradeoffs between model fit and complexity for models composed of between 1 and 4 components; resulting models will then be compared.

Hypotheses for Experiment 2 are summarized as follows:

1. Syllable alignment was predicted to cluster around a relatively small number of phase positions
2. Alignment of target syllables was predicted to be quantitatively different across Form conditions. Specifically:
   (a) Stem-initial syllables in Syllable 3 in the SIENC condition were expected align more closely (<.03 phase) with SHP positions of .333, .500, or .667 than prefix syllables of the PRESI condition.
   (b) Syllables in Syllable 4 position of the PRESI condition were expected to align more closely with SHP positions of .333, .500, or .667 than were pronominal enclitic syllables of the SIENC condition.

5.3.4 Results

Once again, as with Experiment 1, Figure 5.16 shows there was a linear change in phase across speech rates; regression analysis revealed this relationship to be significant across all syllable ($p < .0001$). Once again, the strongest linear trends associated with target syllables 3 and 4 (Figure 5.17), where $R^2 \geq .40$.

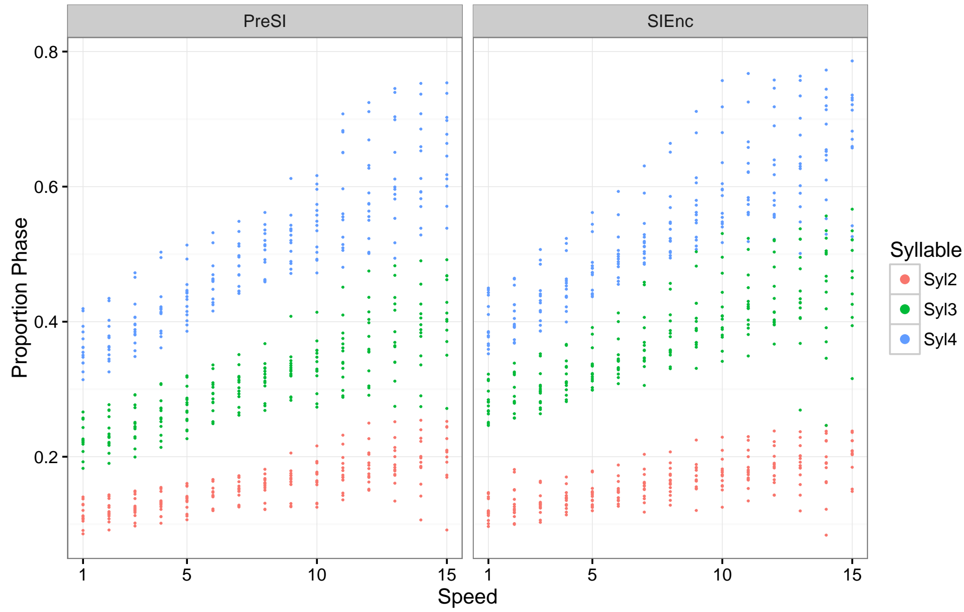


Figure 5.16: Scatter plot of phase angle by metronome speed (speech rate), Experiment 1 (red=Syllable 2, green=Syllable 3, blue=Syllable 4)

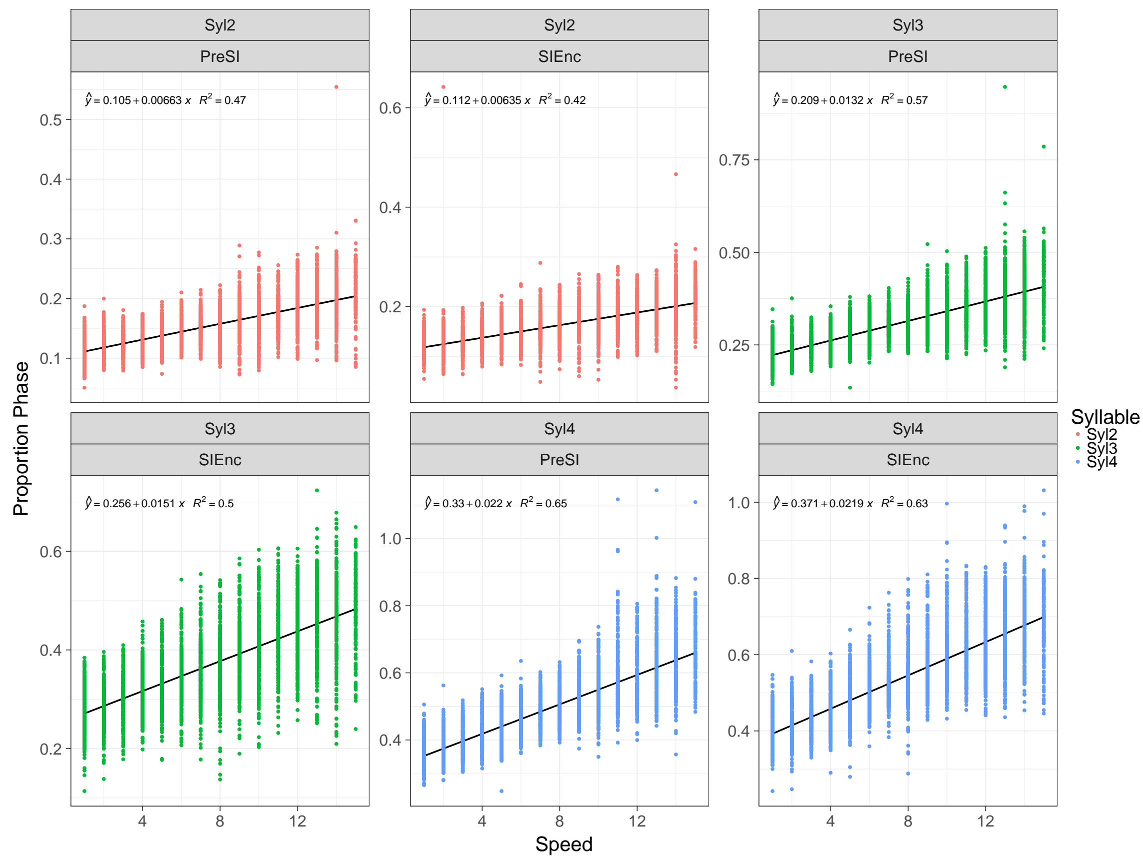


Figure 5.17: Distribution by syllable with regression lines, Experiment 2 (red = Syllable 2, green=Syllable 3, blue=Syllable 4)

In order to verify that the distributions for the target syllables were non-normal, Anderson-Darling tests of normality were conducted for each syllable in each condition. Results of these tests are presented in Table 5.10; as shown, target syllables for both Form conditions failed the normality test ($p < .0001$). Heatmaps of alignment patterns by speed of each syllable by Form condition are provided in Figures 5.18 and 5.19, and show the relative frequency of alignment at different phase positions by speed. Cooler colors are associated with lower frequencies, and warmer colors with higher frequencies. As can be seen in the patterns of orange and red dots, frequency of alignment of syllables was greater at certain phase positions than at others, particularly at metronome speed levels 8 and below, which were the speeds speakers reported being the most comfortable for the task.

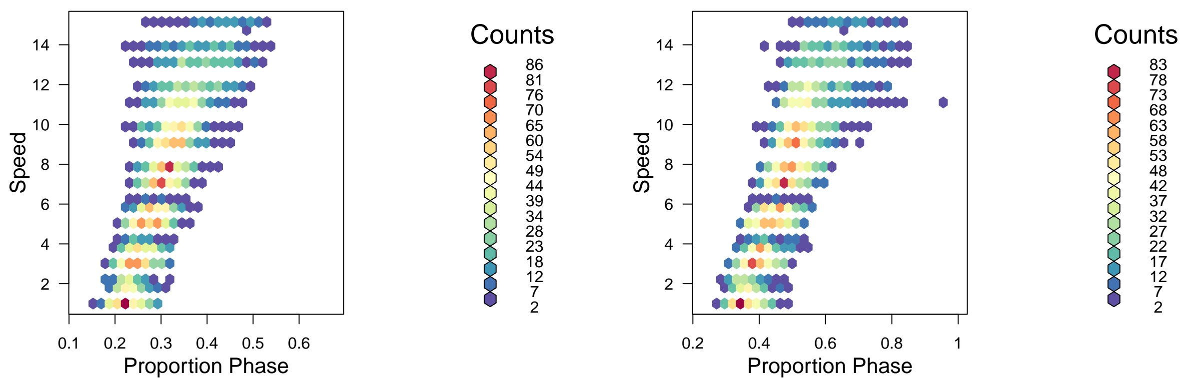


Figure 5.18: Heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), PreSI condition, Experiment 1. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position

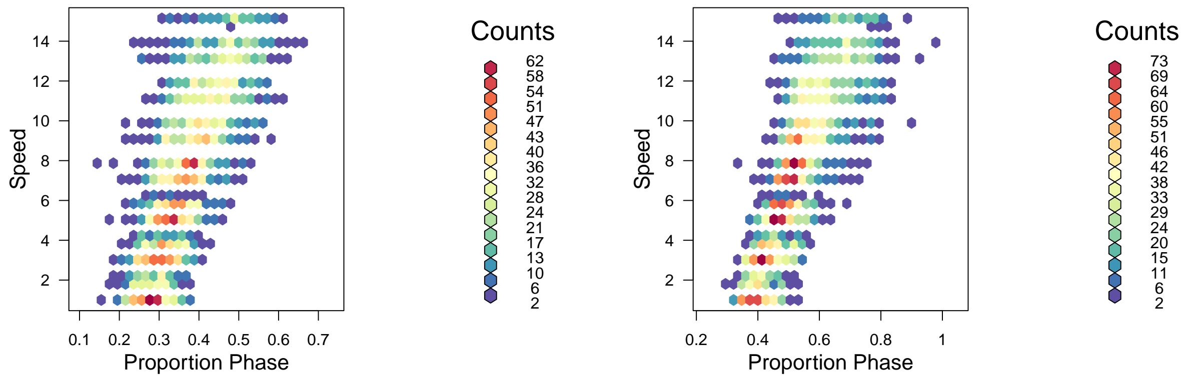


Figure 5.19: Heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), SIEnc condition, Experiment 1. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position

Form	**Syllable**	**A-D Statistic**	**p-value**
PRESI	Syl 3	64.685	$< .0001$
	Syl 4	47.212	$< .0001$
SIENC	Syl 3	48.307	$< .0001$
	Syl 4	58.604	$< .0001$

Table 5.10: Results of Anderson-Darling test for normality, Experiment 1

Gaussian mixture models were fitted separately for each Form (SIENC vs. PRESI), for Syllable 3 and Syllable 4, the two syllables of interest. In order to evaluate the tradeoff between goodness of fit and model complexity (which increases as the number of components increases), an additional step involving parametric bootstrapping of likelihood ratio statistics was carried out on models involving components $k = k_0$ versus $k = k_0 + 1$, up to a total of 4 possible components. Resulting observed log likelihood estimates for these tests and corresponding p-values are provided in Table 5.11.

For Syllable 3 in each condition, it was found that 2 components could be justified at the $p < .005$ level of significance. For Syllable 4 in each condition, 3 components were found to be justified. Final mixing weights, mean parameters and standard deviations are provided for each

		PreSI		SIEnc	
		Obs. LRS	p-value	Obs. LRS	p-value
Syllable 3	$K = 1$ vs. $K = 2$	403.304	$< .005$	291.045	$< .005$
	$K = 2$ vs. $K = 3$	87.711	=.01	9.876	= .03
Syllable 4	$K = 1$ vs. $K = 2$	364.268	$< .005$	377.094	$< .005$
	$K = 2$ vs. $K = 3$	129.168	$< .005$	71.296	$< .005$
	$K = 3$ vs. $K = 4$	.176	=.83	17.448	=.14

Table 5.11: Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K.

component of each model in Table 5.12. Plots of mixture models by syllable and condition are shown in Figures 5.20-5.23.

		textscPreSI			SIEnc		
		μ	σ	λ	μ	σ	λ
Syllable 3	Component 1	.278	.047	.682	.321	.056	.620
	Component 2	.373	.070	.318	.450	.072	.380
Syllable 4	Component 1	.372	.038	.221	.384	.027	.083
	Component 2	.487	.059	.528	.493	.067	.607
	Component 3	.621	.102	.251	.662	.090	.310

Table 5.12: Model Parameters For Syllable 3 and Syllable 4: Experiment 1

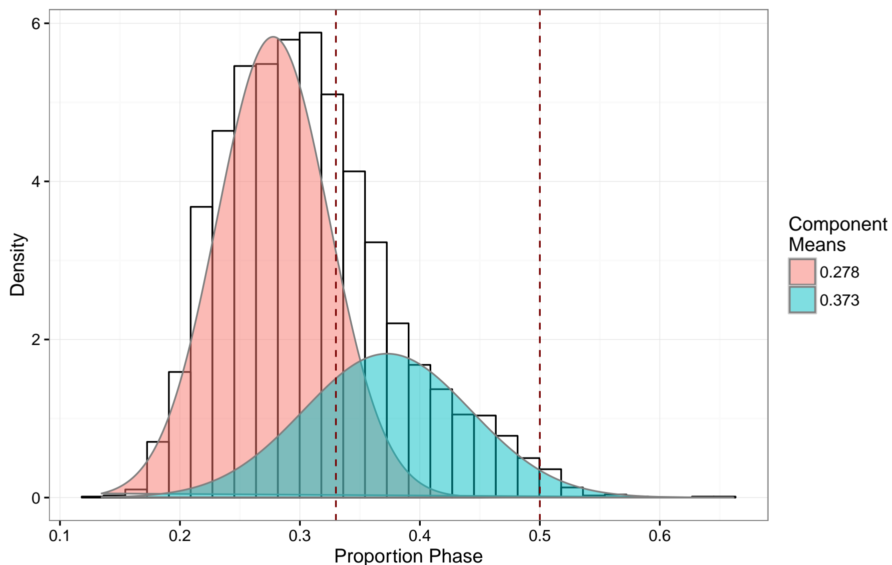


Figure 5.20: PreSI Syllable 3: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .333 and .500)

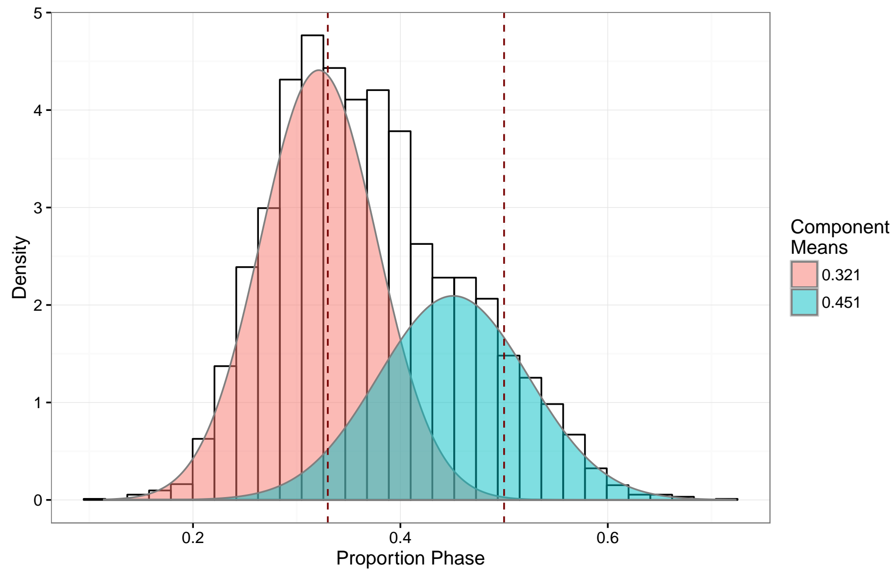


Figure 5.21: SIENC Syllable 3: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .335 and .500)

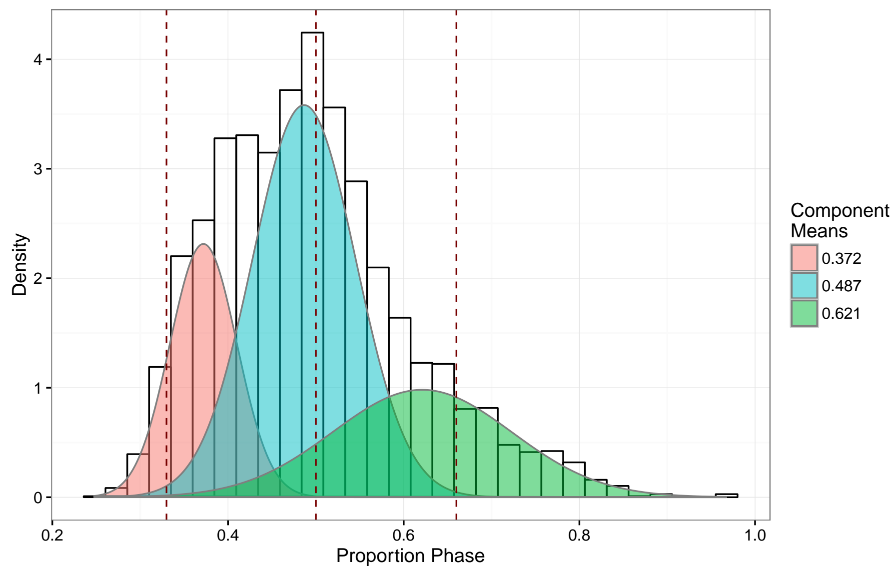


Figure 5.22: PRESI Syllable 4: 3 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667)

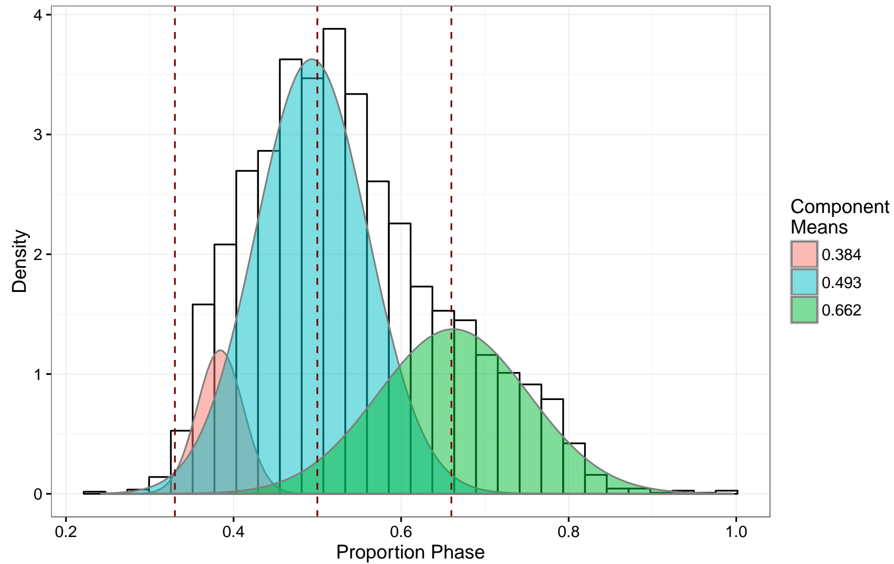


Figure 5.23: SIENC Syllable 4: 3 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500, and .667)

Hypothesis 1, that subjects' behavior on the speech cycling task should reflect a preference for a relatively small number of stable modes of alignment, appears to be supported by the from the results of Experiment 2. Hypothesis 2 made the further prediction that stem-initial syllables would occur at Simple Harmonic Phases (SHP), or lower order fractions, of the Phrase Repetition Cycle. In order to understand whether or not this prediction was supported, we will need to examine the means of each component across the two syllables of interest in each condition. Again, as with Experiment 1, the phase positions between the two syllables *relative to one another* will be important to understand. To reiterate, the reason for this is that, depending on the rhythmic mode speakers were using, there should be variation in the strength that certain phase positions exert in attracting prominence. We consider the relative position of these syllables by Form condition, beginning with the PRESI condition.

PRESI Condition

Hypothesis 2a was that Syllable 4 in the PRESI condition, a stem-initial syllable, would display greater prominence than Syllable 3, a prefix. Referring back to Table 5.12, starting with Syllable 3,

we see that there were two modes of coordination, one at the proportion .278 of the repetition cycle, and one at .373. As seen in Figure 5.20, neither of these proportions aligns very closely with any of the predicted SHP positions of $\frac{1}{3}$ (.333), $\frac{2}{3}$ (.667), or $\frac{1}{2}$ (.500). The closest SHP for either of these components would be $\frac{1}{3}$, which is .056 phase away from .278, Component 1, and .040 phase away from .373, Component 2. For Syllable 4, coordination modes with means of .372 (Component 1), .487 (Component 2) and .621 (Component 3) were found. While the mean of Component 2 was quite close ($<$.02 phase) to SHP position of $\frac{1}{2}$, means for the other two components were quite far ($>$.04) from SHP positions.

We see that the first two components for Syllable 4 have a combined mixing proportion λ of .749, meaning that together the accounted for nearly 75% of the data for Syllable 4. This was only slightly more than the mixing proportion assigned to the first component alone for Syllable 3, at .682 (68%). We can conclude, then, that Syllable 3 remained relatively stable at .278 for the bulk of experimental trials, while Syllable 4 shifted between .372 (Pattern 1) and .487 (Pattern 2) phases for these same trials, depending on metronome speed/speech rate. For the remainder of the trials, Syllable 3 occurred around .373 and Syllable 4 shifted to .621 (Pattern 3); neither of these positions reside close to an SHP position. These three alignment patterns are presented in Figure 5.24.

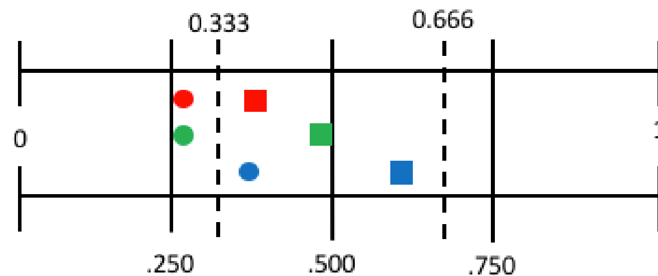


Figure 5.24: Patterns for Syllable 3 (circles) and 4 (squares), PRESI condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3.

Thus far, we see that Syllable 3 occurred quite far from SHP positions overall, whereas Syl-

lable 4 occurred quite close to SHP positions for at least some of the time, as indicated by the mean alignment for Component 2. Once again, it is important to keep in mind that not all of the patterns shown in Figure 5.24 had the same probability of occurrence (as reflected in mixing weights of each component in the distributions listed in Table 5.12). Furthermore, the shape of the distributions of each of the model components varied widely, with some displaying much larger standard deviations and some much smaller. Thus, once again, in order to ascertain which of the distributions accounted for the most data while displaying the least amount of variation, Component Strength was computed, with smaller values representing 'better' ratios of variation to data coverage. Dominant components for each syllable (those with the smallest CS values) are bolded in Table 5.13.

	Syllable 3	Syllable 4
Component 1	**.069**	.172
Component 2	.220	**.112**
Component 3	N/A	.402

Table 5.13: Component strengths, Syllables 3 and 4, PRESI condition

As seen in Table 5.13, Component 1 was identified as the dominant component for Syllable 3, and Component 2 for Syllable 4. Recall that Component 2 for Syllable 4 was also the distribution whose mean lay closest to an SHP position ($\frac{1}{2}$), in line with Hypotheses 2a and 2b, that Syllable 4 would behave as more prominent than Syllable 3 in the PRESI condition.

SIENC Condition

Based on evidence that pronominal enclitics are prosodically dependent on stem-initial syllables, forming prosodic words with them, it was predicted that Syllable 4 enclitics should display less syllable prominence than Syllable 3 stem-initial syllables in the SIENC condition. Looking first at Syllable 3, which displayed two modes of coordination, means of .321 (Component 1) and .450 (Component 2) were found. The mean for Component 1 was quite close to the $\frac{1}{3}$ SHP position, while the mean for Component 2 was fairly far (=.05 phase) away from the closest SHP position

of $\frac{1}{2}$. Turning to Syllable 4, modes of alignment appeared with means of .384 (Component 1), .493 (Component 2), and .662 (Component 3). Means for Components 2 and 3 both occurred quite close (<.01 phase) to SHPs of $\frac{1}{2}$ and $\frac{2}{3}$, respectively, while the mean for Component 1 remained far (>.05 phase) from an SHP position.

Mixture weights of Components 1 and 2 for Syllable 4 had a combined value λ of .690, meaning they accounted for 69% of the data. Component 1 for Syllable 3 accounted for almost this much data, with a value of .620 (62%). Thus, it seems, similar to the PRESI condition, that Syllable 3 remained relatively fixed in its mean position of .321 for around two-thirds of the trials while Syllable 4 occurred for a brief time at .384 (Pattern 1) before shifting to .493 (again, close to the $\frac{1}{2}$ SHP position—Pattern 2). Finally, for the last third or so of the data, Syllable 3 shifted to .450 and Syllable 4 to .662 (again, close to the $\frac{2}{3}$ SHP position—Pattern 3). The three alignment patterns can be seen in Figure 5.25.

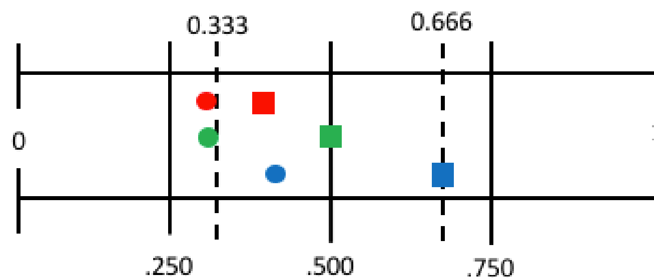


Figure 5.25: Patterns for Syllable 3 (circles) and 4 (squares), SIENC condition. Red = Pattern 1, green = Pattern 2, blue = Pattern 3.

For Pattern 1, Syllable 3 appears to have been drawn to a lower-order phase than Syllable 4 at slower speech rates: the first component for Syllable 3 occurred at .321, close to the $\frac{1}{3}$ phase, and Syllable 4 occurred at .384, not clearly approximating any of the SHPs. For Pattern 2, the second component for Syllable 4, had a mean of .493, squarely aligned with the lowest-order SHP of $\frac{1}{2}$. In this situation, similar to one of the patterns found for the SISI condition in Experiment 1, Syllable 3 and Syllable 4 are both occurring at SHPs ($\frac{1}{3}$ and $\frac{1}{2}$, respectively). One possibility is

that subjects were split in terms of which syllable they were treating as most prominent within this pattern, with some subjects treating Syllable 3 as most prominent, and others treating Syllable 4 as most prominent. However, closer inspection of the data reveals that all subjects were fairly equally represented in their simultaneous treatment of Syllable and Syllable 4 as prominent. Indeed, Pattern 2 reflects the dominant components for both Syllable 3 and Syllable 4 in the SIENC condition (Table 5.14). We discuss this pattern further in §5.4.

	Syllable 3	Syllable 4
Component 1	**.090**	.325
Component 2	.189	**.110**
Component 3	N/A	.290

Table 5.14: Component strengths, Syllables 3 and 4, SIENC condition

Comparing Dominant Components Across Conditions

Up to this point, we have only compared alignment of Syllable 3 and Syllable 4 within each condition. Also of interest is how alignment varied across the PRESI and SIENC conditions. First, it so happens that dominant components across conditions were also the components which happened to most closely align in mean phase proportion within each syllable. Despite this, as seen in Figure 5.26 (left side of the page), alignment of Syllable 3 dominant components was quite different across conditions, with alignment of the prefix syllable in the PRESI condition (blue line) occurring much earlier than the stem-initial syllable in the SIENC condition (red line). The means for Syllable 3 are indicated with dotted lines. We see that the mean for the PRESI condition (blue dotted line) occurred much earlier than that of the SIENC condition (red dotted line), which occurred very close to the $\frac{1}{3}$ SHP position (solid black line). To confirm that the difference in alignment of Syllable 3 was significant across the two conditions, a linear mixed effect model was constructed with Proportion Phase as the response variable, and which included fixed effects of FORM and METRONOME SPEED (as well as their interaction), as well as by-subject random slopes for FORM and METRONOME SPEED. Results revealed alignment for the PRESI condition

	Factor	Estimate	Std. Error	df	t-value	p-value
Syllable 3	Form	-.025	.0022	16	-11.837	$< .0001$
	Speed	.008	.0008	13	11.053	$< .0001$
	Form $\times$ Speed	.000	.0001	6175	3.478	$< .001$
Syllable 4	Form	-.005	.0021	13	1.286	.24
	Speed	.013	.0008	24	-2.426	$< .0001$
	Form $\times$ Speed	.001	.0002	4852	-7.379	$< .0001$

Table 5.15: Results of linear mixed effects models Syllable 3 and Syllable 4, Experiment 2

occurred significantly earlier than that of the SIEnc condition (t = -11.837, df = 16, $p < .0001$). There was also a significant interaction between Metronome Speed and Form. Referring back to Figure 5.17, change in phase position by speed was greater in the PreSI condition than in the SIEnc condition. All results from mixed effects models for Experiment 2 can be found in Table 5.15.

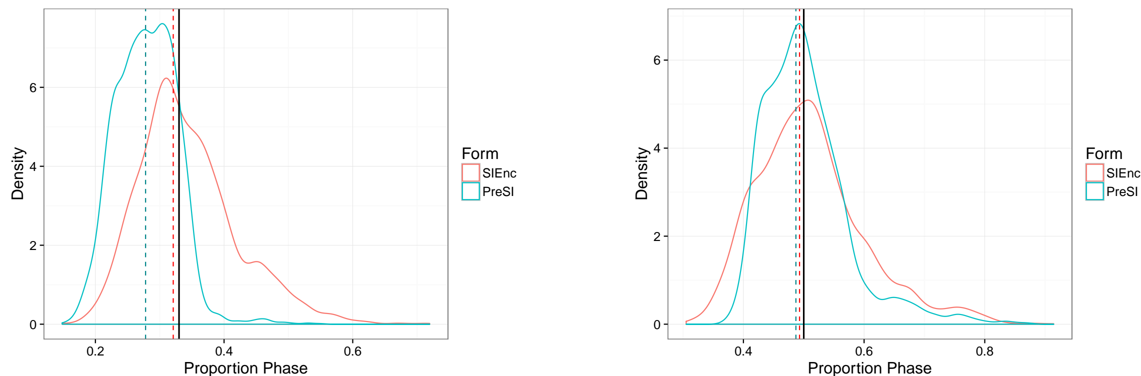


Figure 5.26: Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 2

Looking now at Syllable 4 (right side of the page), alignment of dominant components was quite similar between the PreSI (blue) and SIEnc (red) conditions, with means both quite close to the SHP position of $\frac{1}{2}$. Indeed, results of mixed effects modeling indicated no significant difference in alignment position between the two conditions (t = 1.286, df = 13, p = .24). Once again, a significant interaction was found between Metronome Speed and Form, such that the change in phase position by speed was slightly greater for the PreSI condition than for the SIEnc condiiton.

5.3.5 Discussion

As with Experiment 1, a linear relationship was found between phase position and relative phase for syllables in Experiment 2, and particularly those occurring toward the end of the sentence. Furthermore, as predicted, speakers once again showed a preference for a limited set of stable modes of rhythmic organization. Our initial hypothesis was that stem-initial syllables would display greater prominence than non-stem syllables, as evidenced from their closer proximity to SHP positions. For the PRESI condition, evidence of greater alignment of the Syllable 4 stem-initial syllable to SHP positions as compared to the Syllable 3 prefix syllable supports the hypothesis that stem-initial syllables are more prominent than non-stem syllables. However, alignment patterns in the SIENC condition indicate that both stem-initial and pronominal enclitic syllables display elevated prominence. That pronominal enclitics should behave as more prominent than prefixes is not entirely surprising given the patterning of vowels across these elements: while prefixes (and suffixes) are only ever realized with the vowel [ə], enclitics can be realized with the full array of vowel qualities.[4][5] Still, the fact that enclitics are subject to the same lenition processes and restrictions on downstep as are found stem-internally suggests they behave as a single prosodic word with their hosts, making our results somewhat surprising.

Finally, mode shifting behavior showed subjects shifting the position of syllables—such as Syllable 4 in the SIENC condition—from one SHP position to another (in this case, from $\frac{1}{2}$ to $\frac{2}{3}$). Perhaps even more interesting, there were also circumstances under which subjects maintained alignment of Syllable 3 near the $\frac{1}{3}$ SHP position, consistent with a three beat waltz-like pattern, and Syllable 4 with $\frac{1}{2}$, consistent with a two beat 'cut-time' pattern, *within the same set of utterances*. This type of pattern has not been reported in other speech cycling studies, and in fact is predicted

4. The lack of observable vowel alternations (as are found based on stress in English, e.g. in focused vs. unfocused versions of function words) in these environments prohibits investigation of patterns of phonetic reduction to confirm whether ə is behaving in such environments as a 'reduced' vowel similar to in English.

5. Note that the range of vowel qualities includes ə, as in the 3.pl/plural posssessum form *tsə́*. As no differences were found in behavior between enclitics bearing ə versus other vowels, it is unlikely that vowel quality itself plays a role in Medʉmba for marking prominence asymmetries.

not to occur, since subjects are generally thought to behave in accordance with one pattern or the other (Cummins & Port, 1998). I will discuss this pattern further in the next section.

5.4 General Discussion: Implications for Medʉmba Phonology

Looking at the results from the two speech cycling experiments, we can conclude several things. First, Medʉmba behaves like a typical syllable-timed language in speech cycling, as an overall quite strong positive relationship was observed between relative phase position and speech rate. This would indicate that subjects did not modify the durations of syllables in order to align particular syllables with specific phase positions, as has been observed for English speakers (Cummins & Port, 1998; Tajima, 1998). Second, speakers displayed a multi-modal pattern of phase-alignment strategies, consistent with previous speech cycling research, and consistent with our initial prediction that speakers should display entrainment and gravitate toward specific coordinative modes in speech.

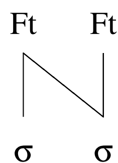
Furthermore, patterns of entrainment were largely found to support the hypothesis that stem-initial syllables are associated with greater prominence than non-initial syllables and than certain types of non-stem syllables, namely, prefixes. Specifically, across both experiments, stem-initial syllables were found to be attracted more strongly to lower-order harmonic phase positions than were stem-final syllables and non-stem syllables such as prefixes. These results are consistent with an account in which a trochaic foot is right-aligned with the prosodic word in Medʉmba, such that the foot head (underlined in (16)) aligns with the stem-initial syllable.

(16) $_{\omega}$(nə̀- $_{FT}$(<u>ká</u>ɣə́)$_{FT}$)$_{\omega}$

From this perspective, then, it is surprising that pronominal enclitics, which also appear to form a foot with stem-initial syllables, display similar prominence to stem-initial syllables in speech cycling. As discussed, the phonological patterning of segments within enclitics is intermediate between that of stem-initial and stem-final syllables in that they display consonant lenition, but

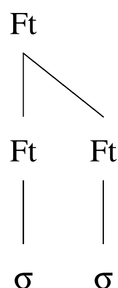
no vowel reduction. Downstep patterns with enclitics seem to indicate they are not independent prosodic words. How might we account for this pattern in a foot-based account? One possibility is to assume, following Hyde (2002), that feet are allowed to overlap, such that a given syllable may be shared by multiple feet. In the structure in (17), the rightmost syllable, the enclitic, forms the head of its own monosyllabic foot, while forming the dependent node of another foot headed by the leftmost stem-initial syllable.

(17) Overlapping Foot Structure



Another approach, originally put forth by Rice (1992) and subsequently advocated for by Bennett (2012); Martínez-Paricio (2012, 2013); Bennett (2013), is to allow for recursive footing. Note that, in order to capture prominence-related generalizations from the present experiment, it would be necessary to posit two monosyllabic *minimal feet* which are then nested within a larger trochaic binary *maximal foot*, allowing us to account for lenition patterns between the stem-initial syllable and enclitic (18).

(18) Recursive Foot Structure



In either of these approaches, at least one syllable must form on its own a monosyllabic foot. While this may seem undesired given crosslinguistic evidence for the prevalence of binary feet (as mentioned in Chapter 2, Medʉmba does not show clear evidence of mora timing, making bimoraic

monosyllables unlikely), the preponderance of monosyllabic words in Medʉmba seems to at least leave open the possibility for commonness of monosyllabic foot structures. As Hyde (2016) points out, however, the structure in (18) would be prohibited on entirely different grounds due to the fact that the proposed maximal foot has dependents which aren not syllables, but other feet. Hyde discusses the fact that, while theories of prosodic recursion are generally quite flexible in the types of recursion they allow at the prosodic word level and above (Itô & Mester, 2009, 2012, 2013), the range of possibilities for recursive structure tend to be more limited at the foot level. Thus, while the recursive foot has been useful in characterizing, for example, ternary stress patterns, (Rice, 1992; Bennett, 2012; Martínez-Paricio, 2012, 2013), the recursive foot as it has been instantiated in recent work appears to be insufficient for characterizing the data at hand, unless a foot can be allowed to contain multiple heads.

Also of interest in the results of the two experiments was that tone did not seem to play a role in influencing relative prominence of stem-initial syllables or stem-final/non-stem syllables. Despite the fact that the stem-initial syllable in the SIENC condition was always low tone (in order to match with the low tone prefix syllable in the PRESI condition), it behaved in a similar way as the high tone stem-initial syllables in Experiment 2. Likewise, the low tone prefix syllable in the PRESI condition from Experiment 1 behaved in a similar way to the high tone stem-final syllable in the SISF condition in Experiment 2: in both cases, these syllables tended to occur earlier than comparable stem-initial syllables and farther from SHP positions. However, since these two syllables occurred in different positions (the prefix in penultimate position and the stem-final syllable in final position), direct comparison is not possible between the two. In Chapter 6, in order to further explore possible effects of tone on relative prominence, we examine structures which are matched for morphological structure and which vary only in tone.

Finally, for the SIENC condition, the pattern which was most prevalent (Pattern 2, Figure 5.25) involved positioning of (stem-initial) Syllable 3 at the $\frac{1}{3}$ SHP position, consistent with a three beat waltz-like pattern, and (enclitic) Syllable 4 at the $\frac{1}{2}$ position. Recall that a similar phenomenon

was found in the SISI condition in Experiment 1, where, again, the prevalent pattern of alignment involved Syllable 3 (a stem-initial syllable) at the $\frac{1}{3}$ SHP position and Syllable 4 (also a stem-initial syllable) at the $\frac{1}{2}$ position (Pattern 1, Figure 5.13). The presence of prominent syllables occurring in SHP positions spanning two rhythmic modes at once has not been reported in other speech cycling studies, and in fact is predicted *not* to occur, since subjects are generally thought to behave in accordance with one pattern or the other (Cummins & Port, 1998). As discussed, the pattern does not appear to derive from individual differences in choice of alignment pattern, as data for all subjects was represented within each pattern. While it is possible that subjects were only feeling one of those two positions as prominent at any given time, it raises the question of whether simultaneous prominence between two syllables may be possible. For example, it is possible that subjects were subdividing the phase in 6 parts, which is consistent with either a 3-pulse rhythm (where each pulse is then subdivided in 2) or a 2-pulse rhythm (where each pulse is subdivided in 3, as in a 6/8 rhythm in music). The opposition of these two rhythms within a six or twelve beat pattern is in fact a very common occurrence in West African music, often referred to as 'polyrhythm' (Locke, 1982; Toussaint, 2003; Agawu, 2006). It is possible, then, that subjects were feeling both the third and fourth syllables as prominent within these utterances, but at different rhythmic 'levels'. We return to this possibility and what it could mean for the metrical structure of Medʉmba in Chapter 8.

CHAPTER 6

EXPERIMENTS 3 AND 4: THE RELATIONSHIP BETWEEN TONE AND PROMINENCE

In this chapter, we will take a closer look at the relationship between tone and prominence in Medʉmba. As discussed in Chapter 3, there is ample evidence that tone placement in many Bantu languages is influenced by relative syllable prominence, and that tone may also be able to exert an influence on stress placement (de Lacy, 2002). For example, high tones in some languages exhibit attraction to positions hypothesized to be prosodically or metrically strong (Goldsmith, 1987; Sietsma, 1989; Odden, 1990; Downing, 1990; Bickmore, 1995; de Lacy, 2002). In the previous chapter, we examined patterns of relative prominence between stem-initial and non-initial/non-stem syllables. It was found that stem-initial syllables of both high and low tones behaved as prominent. We also found that both high and low tone non-initial/non-stem elements can exhibit evidence prosodic weakness, as both high tone stem-final and low tone prefix syllables occurred far from SHP positions. Are there any cases in Medʉmba where tone can interact with other prosodic or structural factors in shaping prominence? In what follows, I present results from speech cycling experiments aimed to test for the role of tone in shaping relative prominence in two environments where it may be more likely to do so. The first environment is one we have explored previously: pronominal enclitics. It was found in Chapter 5 that a pronominal enclitic, despite showing evidence of prosodic dependence on a preceding noun root, shows evidence of elevated syllable prominence in speech cycling. It was argued that pronominal enclitics may act as both prosodically strong and dependent at once. In that experiment, all enclitics were high tone. Here, we explore whether or not the tone of a pronominal enclitic can influence its relative prominence in speech cycling. The second environment we examine is the intonational phrase, where high tones are argued to associate to the head position of the phrase. As will be seen, the evidence points to a role for tone in associating with relative prominence at the phrase level, but not within smaller domains, such as the word level. This further highlights the importance of studying prominence

directly, as opposed to relying on tonal patterns to understand the nature of prominence.

6.1 Experiment 3: The Influence of Tone on Relative Prominence of Pronominal Enclitics

Experiment 3 attempts to expand on findings from Experiment 2 to investigate how tone may interact with relative prominence of pronominal enclitics in Medʉmba. It is hypothesized that high tone enclitics may display relatively higher prominence than low tone enclitics. Before describing the details of the experiment, I will explore in more detail into the relationship between clitics and prominence crosslingusitically.

Clitics are recognized as having a kind of 'intermediate' status between independent words and affixes (Zwicky, 1977; Zwicky & Pullum, 1983). Results from Experiment 2 support this notion in that, while pronominal enclitics in Medʉmba are, on the whole, phonologically dependent on a preceding word, their relative prominence appears to be more in line with that an independent word. For example, the speech cycling results for pronominal enclitics in Experiment 1 mirrored those for monosyllabic words in Experiment 2. Indeed, previous accounts of clitics have argued for their chameleon-like morpho-prosodic status, showing more or less dependency on a potential host due to parsing of prosodic phrasing. Selkirk (1995) argues, for example, for a difference between *free clitics*, which are incorporated direction into a phonological phrase (ϕ) without having to also be parsed in a nested prosodic word (ω), and *internal clitics*, in which the clitic is incorporated into the same phonological word as a preceding or following lexical word (Selkirk considers specifically the case of 'function words' in English). Still another type of clitic, an *affixal clitic*, forms a recursive prosodic word with a preceding or following lexical word.[1] Crucially, the same sequence

1. Selkirk's account relies on the mapping of differences in prosodic parsing among different clitic types to differences in syntactic structure, which Zec (2002) argues is not always possible, or desirable. Zec proposes an alternative analysis in which the relative dependency of a clitic is determined by its status as either *bound* or *free*, but that this status is independent of the syntax, and instead governed by prespecification of prosodic frames for individual lexical items.

of words can be parsed in these different ways.

(19) Different prosodic parses for free clitics and internal clitics proposed by Selkirk (1995)

a. Free clitic: $_{\phi}($ $_{\omega}($ fnc $_{\omega}($ lex))

b. Internal clitic: $_{\phi}($ $_{\omega}($ fnc lex))

c. Affixal clitic: $_{\phi}($ $_{\omega}($ fnc $_{\omega}($ lex)))

Selkirk's analysis was designed, among other things, to explain why clitics can sometimes carry stress, and sometimes not. Other analyses, such as Anderson (2005), have proposed that the very status of an element as a clitic entails that it is prosodically deficient, such that clitics are never expected to be independently prominent. Anderson (2005) argues that cases in which a clitic appears to bear prominence can be attributed to the application of more general prosodic rules or constraints which assign or align foot structure such that clitics acquire prominence post-lexically. Good & Yu (2010) also provide evidence from Turkish that enclitics differ from suffixes in being 'accentually dependent' in that the former cannot receive word-level final stress, whereas the latter can.

Accounts barring clitics from independently bearing accent would predict that the patterns found in Experiment 2 with Medʉmba clitics should be impossible, as the evidence clearly points to enclitics as possessing prominence independent from their hosts. Therefore, it seems a more flexible accentual status for clitics is preferable to account for the Medʉmba data. In fact, there have been cases noted where clitics in a language can apparently come in both unaccented and accented forms. For example, Sideltsev (2016) discusses second position enclitics in Hittite, where one class of second position clitics encompasses both accented and unaccented forms. Similar results are noted for Sanskrit by Lowe (2011, 2014). Heath (2005) also discusses the presence of both underlyingly accented and unaccented clitics in Tuareg as it is spoken in Mali.

The case of Medʉmba is therefore not unique, but certainly seems to be typologically rare, at least according to current knowledge. One question that arises is whether or not pronominal enclitics in Medʉmba always behave as prominent, or whether they may show some variation in relative

prominence. One obvious place to investigate possible variation in relative prominence of enclitics is in relation to tone. Given that high tone seems to be the active tone and Medʉmba (participating, for example, in tone spreading processes, where low tone does not), and that relative prominence asymmetries associated with high and low tones across languages associate high tone with greater prominence, it is possible that low tone enclitics—as opposed to high tone enclitics—will display relatively lesser prominence. This may especially be the case where the preceding element to which the clitic attaches to is high. To investigate this possibility, I examine two pronominal enclitic forms which are phonologically identical except for their tone: *ɑ́m* and *ɑ̀m*. The distribution of the high versus low tone enclitics is determined, according to Voorhoeve (1968, 1976), by the class of the head noun of the pronominal construction. Specifically, singular nouns of class I and class II carry low tones, and singular nouns of class III and class IV carry high tones (20).[2]

(20) Possessive pronominal phrases with ɑ́m, and ɑ̀m

a. ŋgà=ɑ̀m 'my prairie'

(Voorhoeve class I; Proto-Bantu class 6)

b. mɛ́n=ɑ̀m 'my child'

(Voorhoeve class II; Proto-Bantu class 15)

c. kʉ̀=ɑ́m 'my foot'

(Voorhoeve class III; Proto-Bantu class 7)

d. mbɑ́d=ɑ́m 'my hill/mountain' (/d/ →[l], so realized as [mbɑ́l ɑ́m])

(Voorhoeve class IV; Proto-Bantu class 18)

I will now describe the specific stimuli and method used for Experiment 3.

2. Drawing on data from Dunstan (1966) for the related language Ngwe, Voorhoeve shows how the five noun classes he outlines can be related to the noun classes of Proto-Bantu, demonstrating, how the former noun class system likely came to be reduced to the system Medʉmba currently displays. Voorhoeve as well as Goldman *et al.* (2013) also show how the system of noun class agreement in Medʉmba operates quite separately from the concord system.

6.1.1 Stimuli

Stimuli for Experiment 3 are provided in Table 6.1. Since some results from Experiments 1 and 2 (namely, for the SIENC condition in Experiment 1 and the SISI condition in Experiment 2) indicated results consistent with both 3-beat and 2-beat patterns for portions of the data, the number of syllables for Experiments 3 and 4 was increased from 4 to 6, in the hopes that being able to compare alignment for a greater number of syllables might shed light into some of the preferred overall timing timing patterns used by Medʉmba speakers on the task. Experiment 3 looked specifically at alignment patterns of Syllable 5 and Syllable 6. Once again, two Form conditions were compared, each including three target sentences. For the HIGHPRO, Syllable 6 was a high tone enclitic pronoun, and for the LOWPRO condition, Syllable 6 was a low tone enclitic pronoun. The penultimate syllable (Syllable 5) in across both conditions was a high tone monosyllabic noun. Syllable 4 had the same makeup across Form conditions, always consisting of a high toned verb. Syllables 1-3, while similar in segmental content, varied in morphological structure and, for Syllable 3 only, in tone. As will be discussed in §6.2, this tonal manipulation was the focus of Experiment 4. For three sentences, Syllable 1 was a monosyllabic (N)CVC noun and Syllable 2 was the second person singular pronominal enclitic consisting of a single low toned vowel *ù* followed by a high tone negation marker on Syllable 3. For the other three sentences, Syllable 1 was a monosyllabic (N)CV noun and Syllable 2 was the low tone recent past tense marker *lù*, followed by a low tone negation marker on Syllable 3. As can be seen in Table 6.1, the entire segmental string encompassing Syllables 1 and 2, regardless of morphological makeup, was (N)CVCV.

Experiment 3

Form	IPA Transcription	Gloss	Translation
HighPro	mʙʉ́ lù kə̀ béb ꜜ**bán ám**	dog RPST NEG wait couscous=1SG.POSS	'The dog didn't wait for my couscous.'
LowPro	nʧó lù kə̀ kəʔ ꜜ**mvɛ́l** àm	thief RPST NEG cut brother=1SG.POSS	'The thief didn't cut my brother.'
HighPro	shút ù kə́ ꜜkúʔ **mbát ám**	ant=2SG.POSS NEG climb hill=1SG.POSS	'Your ant will not climb my hill.'
LowPro	wʉ́t ù kə́ ꜜkwág **mɛ́n** àm	body=2SG.POSS not touch child=1SG.POSS	'Your body will not touch my child.'
HighPro	mʙwə́ lù kə̀ vwáʔ ꜜ**ʧwɛ́n ám**	goat RPST NEG throw wood=1SG.POSS	'The goat did not throw my wood.'
LowPro	mvɛ́t ù kə́ ꜜzwí **ʃún** àm	brother=2SG.POSS NEG hurt friend=1SG.POSS	'Your brother did not hurt my friend.'

Table 6.1: Stimuli for Experiment 3. Syllables hypothesized prominent are indicated in bold. Clitic boundaries indicated with '='

6.1.2 Participants, Procedure, and Data Processing

All data were collected in Bangangté and Yaoundé, Cameroon. Participants in Experiment 3 included ten Medʉmba speakers, 4 male, 6 female. Three of these subjects (s3, s4, s9 and s11) were from the same subject pool as was included for Experiment 2, and subjects s4 and s11 had also participated in Experiment 1. Six new subjects, three men and three women, were newly recruited for Experiment 3. Ages of participants ranged from 20 years to 60 years. All were native speakers who had spent the majority of their lives in Bangangté or the villages surrounding it.

The experiment proceeded in the same way as Experiments 1 and 2: subjects repeated the six target sentences, which were elicited at 15 different speech rates, from slowest to fastest, for a total of 90 trials per subject. For Experiment 3 (as well as for Experiment 4, which will be presented in §6.2), the range of metronome periods was adjusted upwards so as to make repetitions of the longer, 6-syllable sentences more manageable. This time, slowest speech rate corresponded to a 1720 ms metronome period. As with Experiments 1 and 2, the period of the metronome was reduced by 3% for each subsequent speed. The fastest speed thus corresponded to a 623 ms metronome period. Subjects were again given several practice trials to acclimate to the experiment; all subjects were able to complete the task after practice. During the course of the experiment, subjects were given periodic breaks to rest and drink water.

Data were annotated in a similar way as for Experiments 1 and 2: P-centers were marked semi-automatically and hand-corrected; relative phase positions were then calculated for each syllable.

6.1.3 Analysis Plan and Hypotheses

As with Experiments 1 and 2, analysis of the data for Experiment 3 will involved several steps. First, linear regression was used to evaluate the degree to which phase position changed with metronome speed for each syllable in each condition. Second, deviations from normality of the distributions for the two target syllables, Syllable 5 and Syllable 6, were evaluated using Anderson Darling tests. Third, if the data for each syllable showed deviation from a normal distribution,

Gaussian mixture models were applied to these distributions to evaluate patterns of clustering in the data around specific phase positions. To quantify differences in phase position of target syllables across Form conditions, component strength was evaluated for mixture components of each target syllable, and dominant components (those components which account for the most data while exhibiting the smallest standard deviation) compared across Form conditions using linear mixed effects modeling. Again, as with Experiments 1 and 2, the goal of this analysis was to compare the position of syllables occurring in relatively *comparable* positions across forms to see if clustering genuinely showed different alignment patterns across Form conditions. Thus, if dominant components to be compared across Form conditions were not the most comparable in alignment across the two conditions, additional comparisons of non-dominant (but more closely-aligned) distributions were carried out.

Hypotheses for Experiment 3 can be outlined as follows:

1. Syllable alignment was predicted to cluster around a relatively small number of phase positions
2. Alignment of target syllables was predicted to follow certain patterns. Specifically:
    (a) High tone pronoun enclitic syllables were predicted to align more closely ($<.03$ phase) with SHP positions of .333, .5, or .667 than low tone pronoun enclitic syllables.
    (b) Assuming variation in relative prominence across syllables might occur based on speech rate, it was predicted that noun stem-initial syllables occurring in penultimate position (Syllable 5) might show close alignment ($<.03$ phase) with SHP positions.

6.1.4 Results

Figure 6.1 shows the overall distribution of data across the two Form conditions for Experiment 3. Phase alignment of syllables appears to increase with speed; linear regression confirmed a

significant linear relationship between speed and phase position for all syllable ($p < .0001$). As shown in Figure 6.2, stronger linear trends were found for the last two (target) syllables, Syllable 5 and Syllable 6 ($R^2 > .40$)

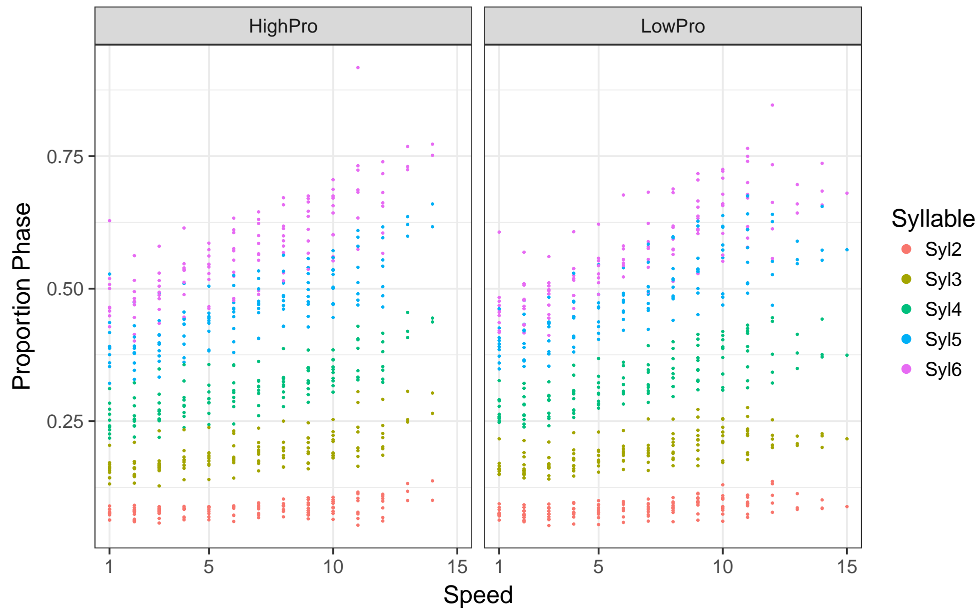


Figure 6.1: Scatter plot of phase angle by metronome speed (speech rate), Experiment 3 (red=Syllable 2, yellow=Syllable 3, green=Syllable 4, blue=Syllable 5, purple=Syllable 6). Each point represents a single observation for a given speaker.

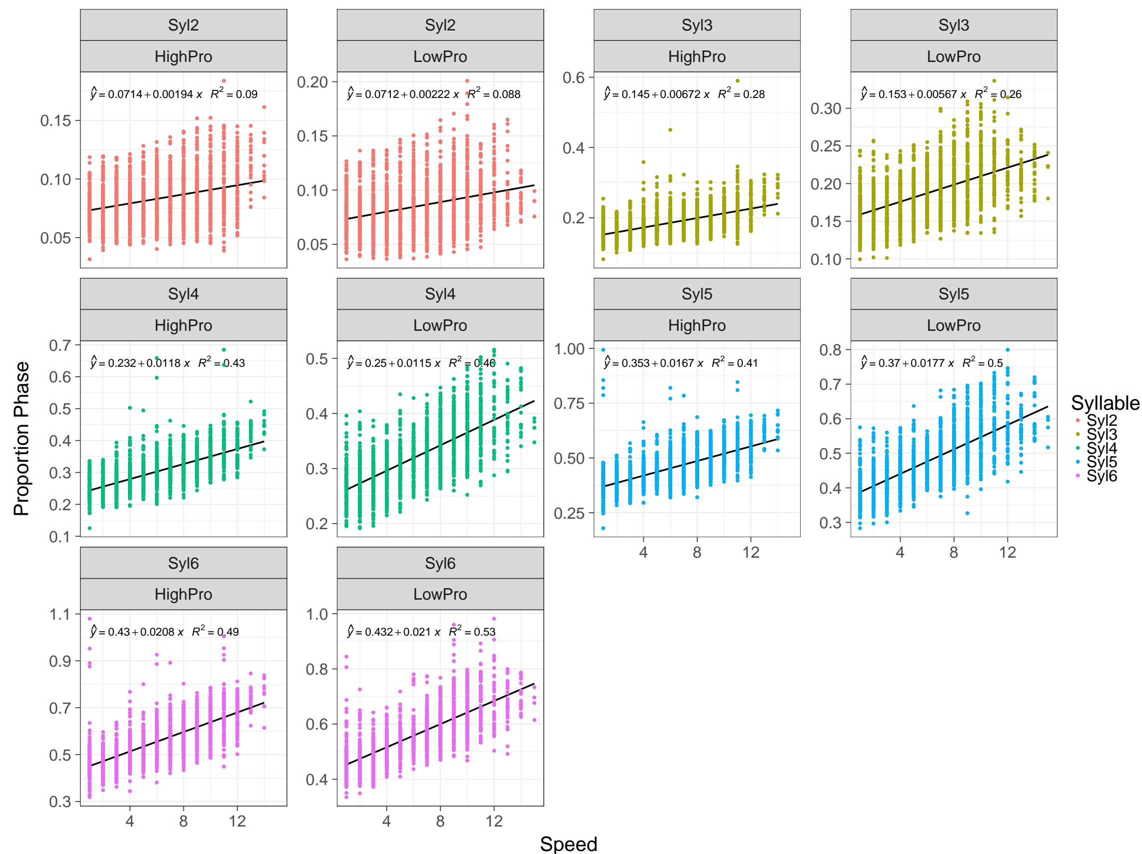


Figure 6.2: Distributions by syllable with regression lines, Experiment 3 (red=Syllable 2, yellow=Syllable 3, green=Syllable 4, blue=Syllable 5, purple=Syllable 6)

Form	Syllable	A-D Statistic	p-value
HIGH PRONOUN	Syl 5	9.760	< .0001
	Syl 6	8.440	< .0001
LOW PRONOUN	Syl 5	7.790	< .0001
	Syl 6	1.060	< .0001

Table 6.2: Results of Anderson-Darling test for normality, Experiment 3

Heatmaps of alignment patterns by speed of each syllable by Form condition are provided in Figures 6.3, 6.4. These plots show the relative frequency of alignment at different phase positions by speed. Cooler colors are associated with lower frequencies, and warmer colors with higher frequencies. As can be seen in the patterns of orange and red dots, and consistent with Hypothesis 1, frequency of alignment of syllables was greater at certain phase positions than at others, particularly at metronome speed levels 8 and below, which were the speeds speakers reported being the most comfortable for the task.

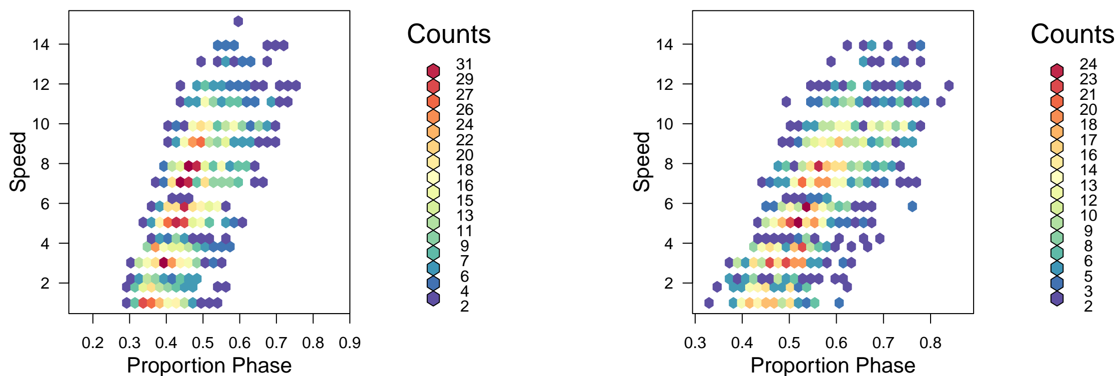


Figure 6.3: Heat map of syllable alignment (Proportion Phase by Speed), Syllable 5 (left pane) and Syllable 6 (right pane), HIGHPRO condition, Experiment 4. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position

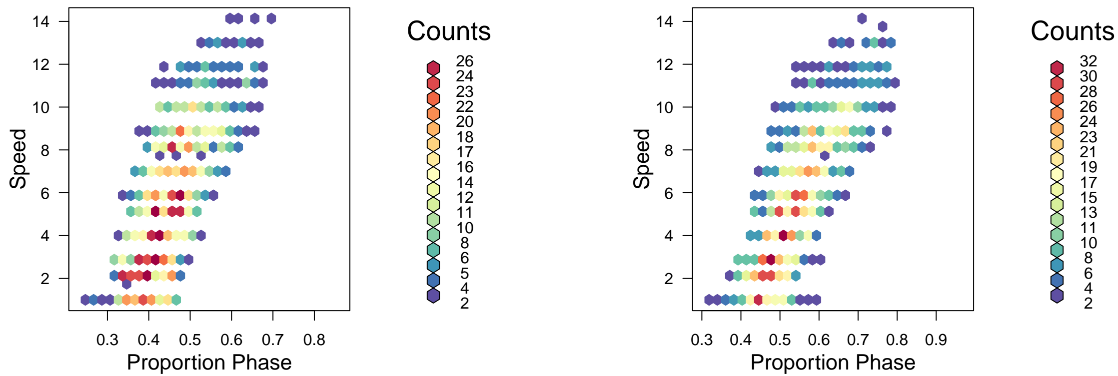


Figure 6.4: Heat map of syllable alignment (Proportion Phase by Speed), Syllable 5 (left pane) and Syllable 6 (right pane), LOWPRO condition, Experiment 4. Greater counts/warmer colors (e.g. yellow and red) indicate greater instance of syllable alignment at a particular position

Results of Anderson-Darling tests of normality indicated that all syllable distributions deviated significantly from normality ($p < 0.0001$) (Table 6.2). Assuming these deviations might represent the expected trend towards multi-modality, Gaussian mixture modeling was used to test the prediction that syllables would cluster around a relatively small number of phase positions. Specifically, it was predicted that a model involving a mixture of Gaussians with small number of components, the mean of each reflecting a preferred alignment pattern in the data, would be a better fit to the data than a model involving only a single Gaussian.

First, a parametric bootstrap of log-likelihood ratio statistics was done to evaluate the optimal number of components for the mixtures. As can be seen in Table 6.3, each of the target syllables across both Form conditions was best modeled with a mixture of two Gaussian distributions. The fact that only two peaks in syllable alignment were found to be justified for each target syllable is in line with predictions made by Hypothesis 3, that data should cluster around a relatively small number of phase positions.

Parameters of the resulting mixture models, including means (μ), standard deviations (σ), and mixing weights (λ) can be found in Table 6.4.

		HIGHPRO		LOWPRO	
		Obs. LRS	p-value	Obs. LRS	p-value
Syllable 5	$K = 1$ vs. $K = 2$	15.202	$< .005$	94.098	$< .005$
	$K = 2$ vs. $K = 3$	11.123	=.04	31.830	= .01
Syllable 6	$K = 1$ vs. $K = 2$	96.665	$< .005$	121.345	$< .005$
	$K = 2$ vs. $K = 3$	4.487	= .22	15.099	= .01

Table 6.3: Observed log likelihood ratio statistics and p-values for parametric bootstrap test for number of components K, Experiment 3.

		HIGHPRO			LOWPRO		
		μ	σ	λ	μ	σ	λ
Syllable 5	Component 1	.431	.062	.861	.439	.058	.639
	Component 2	.590	.048	.139	.547	.080	.361
Syllable 6	Component 1	.506	.062	.636	.512	.062	.600
	Component 2	.644	.078	.364	.634	.094	.400

Table 6.4: Model Parameters For Syllable 5 and Syllable 6: Experiment 3

Graphs of each mixture are provided in Figures 6.5 - 6.8. Means for each component are shown on the right side of each graph (red indicating the mean for Component 1, and blue indicating the mean for Component 2). For each graph, SHP positions that fell within .15 phase of the resulting components are displayed with dotted red lines.

Results for Experiment 3 confirm Hypothesis 1, that subjects would display a relatively small set of stable modes of production across target syllables and conditions. Hypothesis 2 predicted that differences in alignment would be found for Syllable 6 across Form conditions, with high pronoun enclitic syllables occurring more frequently near SHP positions of .333, .500, and .667. No differences were predicted between Syllable 5, the antipenultimate syllable, which was a noun stem-initial syllable in both Form conditions. We now explore whether Hypothesis 2 was supported by examining data within and across Form conditions.

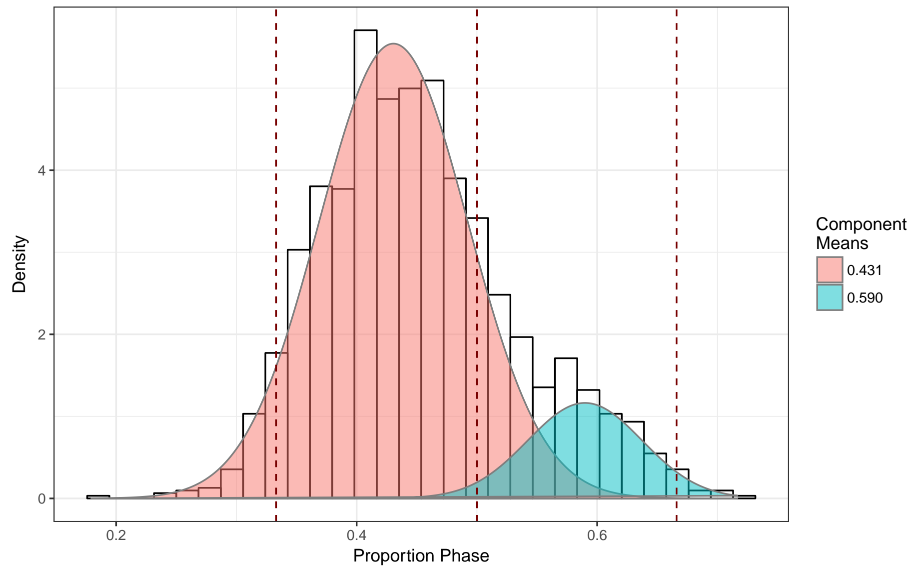


Figure 6.5: HIGHPRO Syllable 5: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .333, .500 and .667)

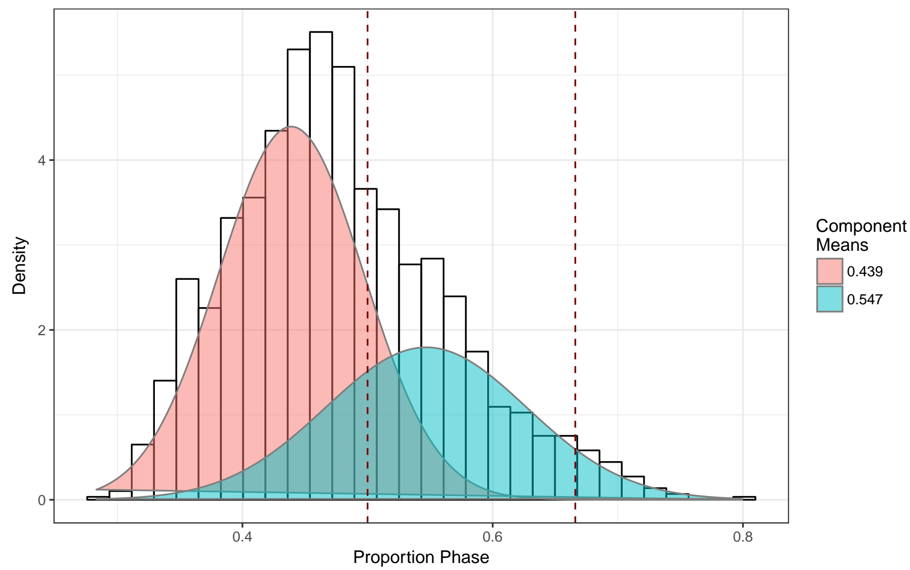


Figure 6.6: LOWPRO Syllable 5: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .500 and .667)

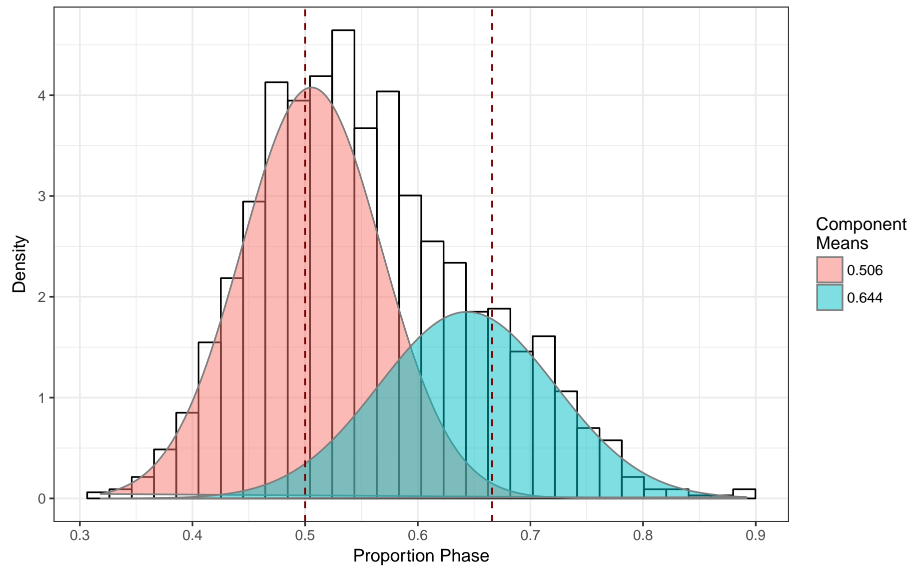


Figure 6.7: HIGHPRO Syllable 6: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .500 and .667)

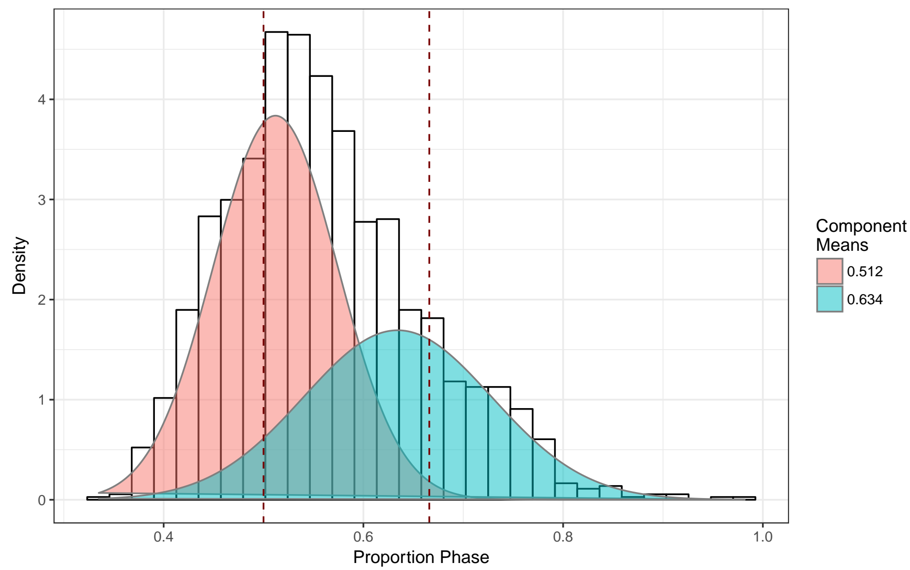


Figure 6.8: LOWPRO Syllable 6: 2 component Gaussian mixture (dotted red lines indicate SHP positions of .500 and .667)

HIGHPRO Condition

Means of Components 1 and 2 for Syllable 5 in the HIGHPRO condition were $\mu = .431$ and $\mu = .590$, respectively (Table 6.4). This means that the mean for Component 1 was .069 phase away from the closest SHP position of .500, and Component 2 was .076 phase away from the closest SHP position of .667. All in all, alignment of Syllable 5 was therefore quite far from SHP positions. Looking now at Syllable 6, we see that resulting components occurred much closer to SHP positions than was found for Syllable 5. Means for Components 1 and 2 for Syllable 6 in the HIGHPRO condition were $\mu = .506$ and $\mu = .644$ respectively, meaning that Component 1 occurred only .006 phase away from the closest SHP position of .500, and Component 2 occurred only .022 phase away from the closest SHP position of .667.

We see from Table 6.4 that Component 1 accounted for around 86% of the data for Syllable 5 and for about 64% of the data for Syllable 6. Component 2 accounted for about 14% of the data for Syllable 5, and 36% of the data for Syllable 6. Thus, while Syllable 5 was consistently (over 75% of the time) aligned around the Component 1 mean, alignment of Syllable 6 was slightly more evenly divided between the Component 1 mean of .506 and Component 2 mean of .644. Therefore, it seems three overall patterns can be posited for these syllables in the HIGHPRO Form condition: Syllable 5 around phase position of .431 and Syllable 6 around .506 (Pattern 1), Syllable 5 around .431 and Syllable 6 around .644 (Pattern 2), and Syllable 5 around .590 and Syllable 6 around .644 (Pattern 3). A diagram of these patterns is shown in Figure 6.9.

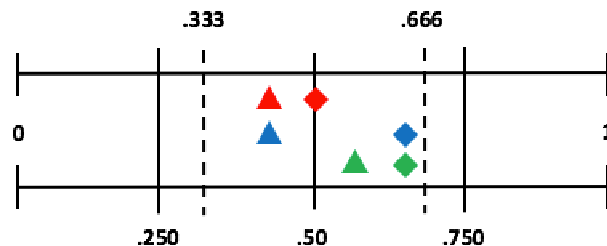


Figure 6.9: Patterns for Syllable 5 (triangles) and 6 (diamonds), HIGHPRO Condition. Red = Pattern 1, blue = Pattern 2, green = Pattern 3.

Evaluation of Component Strength of the HIGHPRO condition revealed Component 1 to be the Dominant Component for both Syllable 5 and Syllable 6 (Table 6.5).

	Syllable 5	Syllable 6
Component 1	**.072**	**.097**
Component 2	.345	.214

Table 6.5: Component strengths, Syllables 5 and 6, HIGHPRO condition

LOWPRO Condition

We see from Table 6.4 that, for the LOWPRO condition, means of Components 1 and 2 for Syllable 5 were μ = .439 and μ = .547, respectively, meaning Component 1 had a mean .061 phase away from the closest SHP position of .500 and Component 2 had a mean .47 phase away from the closest SHP position of .500; both components were thus quite far from SHP positions. In the LOWPRO condition, Components 1 and 2 for Syllable 6 were μ = .512 and μ = .634, respectively. Component 1 was thus only .012 phase from the closest SHP of .500 and Component 2 was .032 phase from the closest SHP position of .667.

For the LOWPRO condition, Component 1 across both syllables accounted for nearly the same amount of data (between 60 and 64 %), and Component 2 also accounted for similar proportions of data across conditions (between 36 and 40 %). We can conclude, then, that there were two primary patterns of alignment, the first with Syllable 5 at around phase position of .439 and Syllable 6 around .512 (Pattern 1), and the second with Syllable 5 around phase position of .547 and Syllable 6 around .634 (Pattern 2). Alignment patterns for the LOWPRO condition can be found in Figure 6.10.

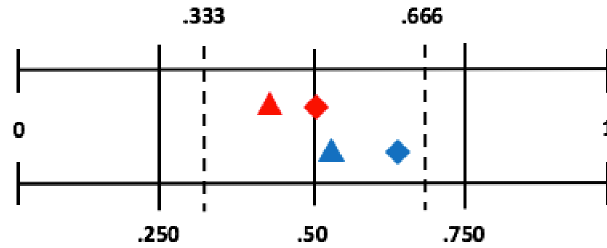


Figure 6.10: Patterns for Syllable 5 (triangles) and 6 (diamonds), LOWPRO Condition. Red = Pattern 1, blue = Pattern 2..

Component 1 was found to be the Dominant Components for both Syllable 5 and Syllable 6 in the LOWPRO condition.

	Syllable 5	Syllable 6
Component 1	**.091**	**.103**
Component 2	.222	.235

Table 6.6: Component strengths, Syllables 5 and 6, LOWPRO condition

Comparing Dominant Components Across Conditions

All in all, it appears that, contrary to Hypothesis 4, Syllable 6 showed similar alignment across both Form conditions, in both cases showing close alignment with SHP positions. This in contrast with findings from Syllable 5, in which it appears that resulting components were far form SHP positions across both Form conditions. In order to confirm quantitatively whether differences in alignment were significant for each syllable across Form conditions, dominant components were compared using linear mixed effects models. Again, tables 6.5 and 6.6 show that the dominant component for Syllable 5 was Component 1 across both Form conditions, and the dominant component for Syllable 6 was also Component 1 across both conditions. For both syllables, it was the case that the resulting dominant components were also the closest-aligned components across conditions. Thus, if results indicate that these components occurred at significantly different positions from one another across Form conditions, this means that any other comparison of components will show an even greater difference.

A visual comparison of dominant components for Syllable 5 and Syllable 6 across conditions is provided in Figure 6.11. As can be seen, alignment across conditions was quite similar for both Form conditions in both Syllable 5 and Syllable 6. Linear mixed effects models were run for each syllable including fixed effects of FORM and METRONOME SPEED (as well as their interaction), as well as by-subject random slopes for both of these variables. Results (displayed in Table 6.7) indicate there was no significant effect of Form for alignment in either Syllable 5 (β = -.008, t = -1.652, p = .13) or Syllable 6 (β = .002, t = .593, p = .566), though there was a significant interaction between METRONOME SPEED and FORM for Syllable 6, such that the influence of speed on phase position was greater in the LOWPRO condition than in the HIGHPRO condition (β = -.001, t = -3.275, $p < .01$).

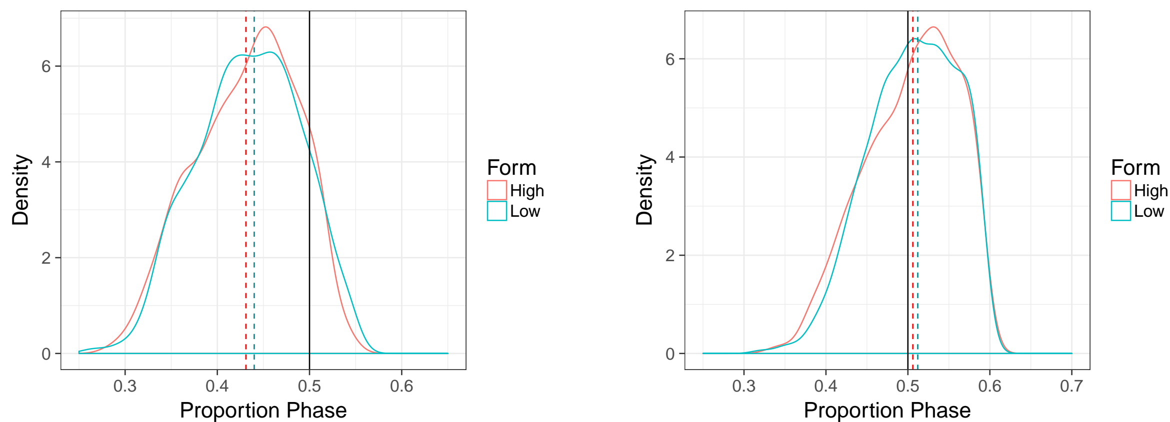


Figure 6.11: Comparison of dominant components for Syllable 5 (left) and Syllable 6 (right), Experiment 3

	Factor	Estimate	Std. Error	df	t-value	p-value
Syllable 5	Form	-.008	.0049	9	-1.652	.1330
	Speed	.030	.0027	8	1.901	$< .0001$
	Form $\times$ Speed	-.001	.0007	2186	-1.941	.0524
Syllable 6	Form	.002	.0049	10	.593	.566
	Speed	.013	.0014	8	9.644	$< .0001$
	Form $\times$ Speed	-.001	.0003	2102	-3.275	$< .01$

Table 6.7: Results of linear mixed effects models Syllable 5 and Syllable 6, Experiment 3

6.1.5 Discussion

Results of Experiment 3 were in line with both Hypothesis 1, that syllable alignment would be found to cluster around a relatively small number of phase positions. However, results did not support either part of Hypothesis 2. Contrary to Hypothesis 2a, there was no difference in alignment between Form conditions for Syllable 6, meaning that tone did not influence relative phase position of pronominal enclitics. Instead, Syllable 6 in both conditions occurred consistently close to SHP positions of .500 and .667. This finding echoes the findings from Experiments 1 and 2 in which tone did not play a role in influencing relative syllable alignment for either stem-initial or non-initial/non-stem syllables. It seems that enclitics, regardless of tone, behave as prominent.

More surprising was the fact that Syllable 5 did not occur near SHP positions of .333, .500 or .667, as had been predicted in Hypothesis 2b. Since Syllable 5 was a noun stem-initial syllable, it was predicted, especially given results from the SIENC condition of Experiment 2, that this syllable should behave as prominent. Recall that dual prominence was assessed for the SIENC condition between the noun stem-initial syllable in penultimate position and the enclitic in final position. Instead, Syllable 5 consistently occurred quite far ($>$.045 phase) from SHP positions in both the HIGHPRO and LOWPRO conditions.

One possibility is that the greater number of syllables included in the stimuli led subjects to adopt a narrower range of alignment strategies as compared with Experiments 1 and 2. The fact that both Syllable 5 and 6 only showed two modes of alignment for Experiment 3 seems to support this notion, as target syllable alignment in many cases for Experiments 1 and 2 tended to display 3 modes of alignment. However, the fact that subjects opted to consistently align the enclitic and not the noun stem-initial syllable with SHP positions is surprising. Another possibility is that the relative prominence of syllables leading up to Syllables 5 and 6 had an influence on the timing of Syllable 5 and 6. We will investigate this possibility further in §6.3. First, results from Experiment 4 will be presented, which examines the same stimuli as Experiment 3, but instead looking at Syllables 3 and 4. Experiment 4 turns to investigate the possibility of phrase-level prominence in

Medʉmba, and its relation to tone.

6.2 Experiment 4: Evaluating Phrase-Level Prominence and the Role of Tone

Experiment 4 examined the interactions between prosodic prominence and tone the level of the intonational phrase. Franich (2014) argues that the left edge of an intonational phrase (to be abbreviated ι) in Medʉmba typically (though not always) aligns with the right edge of the subject, and encompasses the remainder of the clause. One piece of evidence for this comes form placement of a 'left edge'[3] high floating tone which typically associates as the rightmost tone on the subject. The vowel to which this high tone associates is underlined in (21a, 21b). In the case of a toneless subject,[4] the high tone simply docks to the subject (21c). Note that the left edge high tone is also the trigger for downstep, which, in the case of (21), targets the high tone of the verb *jə́n*.

(21) High tone attachment toward the left edge of an intonational phrase in Medʉmba

a. ι[mvə̰̌n ꜜ`jə́n í]
chief see 3SG
'The chief has seen him.

b. ι[mvə̀n ǎm ꜜ` jə́n í]
chief 1SG.POSS see 3SG
'My chief has seen him.

c. ι[mə̰́ ꜜ`jə́n í]
1SG see 3SG
'I have seen him."

Franich (2014) shows that pitch is reset at the start of each new intonational phrase, which typically aligns with a syntactic clause. Thus, rather than continuing with downstep across a clause

3. Though this tone is not always exactly at the left edge of the phrase, it seems to reside somewhere *close* to that position; for this reason, I will continue to refer to that tone as a 'left edge high tone'.

4. Subject pronouns are analyzed as toneless due to the great amount of variability they display tonally based on the contexts where they occur. The reader is directed to Voorhoeve (1976) for further discussion of this variability.

boundary, no downstep occurs, as pitch is set back roughly to the level it was at the start of the preceding intonational phrase.[5] Keupdjio (2015, 2014) shows that Medʉmba demonstrates syntactic evidence for A-bar and focus movement. For example, subjects can be extracted from their original position and moved to a higher clause when in a wh-question, when relativized, or when focused. Franich (2014) shows that, once a subject undergoes any of these processes, two things happen. First, the high tone which associates with the subject can no longer trigger downstep on a following high tone. Second, an additional high tone is introduced, associating to the particle to the right of the subject any relativizers.[6] In (22), the new high tone attaches to the right edge of the recent past marker *lù*, which underlyingly bears a low tone. Franich (2014) argues that these patterns stem from the fact that, after extraction/focus movement, the subject forms its own intonational phrase.[7] Thus, two intonational phrases result: one encompassing the subject, and the other encompassing everything to the right of the subject and any relativizers. The vowels which bear the left edge high tones of each of these intonational phrases are marked with an underline in (22).

(22) Subject movement results in high tone on right edge of tense marker and no downstep triggered by subject on either tense marker or verb

a. ɩ[wʉ̠́ zà] ɩ[lʉ̠̌ njə́n í ɔ̀] (*wʉ̠́ zà ꜜlʉ̠̌ njə́n í ɔ̀)
wh REL.3SG RPST see 3SG Q
'Who saw him (yesterday)?'

b. ɩ[mvə̠̌n zà] ɩ[lʉ̠̌ njə́n mvə̀n ꜜlá]
chief REL.3SG RPST see chief REL
'The child who saw the chief (yesterday)'

c. ɩ[mə̠́] ɩ[lʉ̠̌ njə́n í]
1SG RPST see 3SG

5. In the case of coordinative clauses, Franich shows that pitch is still reset, but not quite as high as the initial pitch of the matrix clause left-edge high tone.

6. Keupdjio 2016 analyzes the correlation between the occurrence of this high tone and movement processes as a case of a high tone agreement marker. Such an analysis is compatible with the present prosodically-based account, as it could be that tone is doing 'double-duty' between syntactic and prosodic marking.

7. See also Hamlaoui & Szendröi (2017) for interesting parallels in prosody and tone in Bàsàá, a Northwest Bantu language also spoken in Cameroon.

‘I$_{FOC}$ saw him (yesterday)’

d. ι[mvə̰̌n] ι[lṵ̌ njə́n í]
chief RPST see 3SG
‘The chief$_{FOC}$ saw him (yesterday)’

There is one additional generalization which applies to intonational phrase formation, as Franich (2014) points out, which is that certain particles—such as negation markers—tend to attract the left edge high tone, such that the high tone attaches to the right of the subject, instead to the particle (23). Thus, for a low tone negation marker *kə̀ʔ* as is found in the recent past, a HL contour results (23a, 23b). For a high tone negation marker, such as *kə́*, the future tense negation marker, a single high tone is realized (23c). In cases where the subject is toneless (as in the case of a pronoun), it is left to carry a default low tone (23b, 23c). Again, the vowel bearing the left edge high tone is underlined for all examples in (23).

(23) Left edge high tone attaches to the left edge of a following negation marker, rather than to the subject

a. ι[mvə̀n kə̰̂ʔ ↓`jə̰́n í]
chief NEG see 3SG
‘The chief has not seen him.’

b. ι[mə̀ kə̰̂ʔ ↓`jə̰́n í]
1SG NEG see 3SG
‘I have not seen him.’

c. ι[mə̀ kə̰́ ↓`jə̰́n í]
1SG NEG see 3SG
‘I will not see him.’

Finally, where another particle intervenes between the subject and a negation marker, the subject will still bear the left edge high tone of the intonational phrase. In (24), we see that the placement of the low tone recent past marker *lù* leads to the left edge high tone attaching to the subject once again.

(24) Low tone tense marker intervening between subject and negation: high goes to right edge of subject

a. ɩ[mvə̲̌n lù kə̀ ↓jə̲́n í]
chief RPST NEG see 3SG
'The chief saw him (yesterday).'

b. ɩ[mə̲́ lù kə̀ ↓jə̲́n í]
1SG RPST NEG see 3SG
'The chief did not see him (yesterday).'

To summarize, intonational phrases in Medʉmba are found to exhibit a high tone which occurs toward the left edge, either on the subject or on a particle that follows it. Franich (2014) argues that the mobility of this high tone has to do with the relative position of the head of the intonational phrase, which can either be aligned with the subject or the following particle. In the experiment that follows, I will explore whether words—and, specifically, negation markers—which coincide with this head position demonstrate relatively greater prominence than words which occur in intonational phrase-medial position.

6.2.1 Stimuli

Stimuli for Experiment 3 are provided in Table 6.8. These were the same sentences elicited for Experiment 4, but this time, the target syllables of interest are Syllable 3 (the negation marker) and Syllable 4 (the following verb). Three sentences were included for each Form condition, with the HIGH/IP-INITIAL condition featuring a high, intonational phrase-initial negation marker in Syllable 3 position, and the LOW/IP-MEDIAL condition featuring a low, intonational phrase-media negation marker. The verb in Syllable 4 position across Form conditions was always high. As for the other syllables, as with the preceding experiment, the penultimate syllable in every sentence was a high tone monosyllabic noun, and the final syllable was either a high or a low tone pronominal enclitic. The initial syllable of every utterance was a high tone noun. As can be seen in the phonetic transcriptions in Table 6.8, the first and second syllables for the sentences had a similar

segmental structure across the two Form conditions, but, as indicated in the morphoglosses to the right, varied in morphological structure. This difference in morphological structure was necessary in order to achieve the correct positioning of the target third syllable negation markers. For HIGH/IP-INITIAL condition, the second syllable was the low tone second person singular pronominal enclitic consisting of a single low toned vowel *ù* followed by a high tone negation marker on the third syllable, and for the LOW/IP-MEDIAL condition, the second syllable was the low tone recent past tense marker *lù*, followed by a low tone negation marker. Where a pronoun occurred in Syllable 2 position, Syllable 1 was a (N)CVC syllable; where a tense marker occurred in Syllable 2 position, Syllable 1 was a (N)CV syllable. Thus, the entire segmental string encompassing Syllables 1 and 2 was (N)CVCV, regardless of their morphological makeup.

Experiment 4

Form	IPA Transcription	Gloss	Translation
High/IP-Medial	shúl ù **kə́** ꜜ**kúʔ** mbál ə́m	ant=2SG.POSS NEG climb hill=1SG.POSS	‘Your ant won’t climb my hill.’
Low/IP-Medial	nʧó lù kə̀ ꜜ**kəʔ** ꜜmvɛ́l ə̀m	thief RPST NEG cut brother=1SG.POSS	‘The thief didn’t cut my brother.’
High/IP-Medial	wʉ́l ù **kə́** ꜜ**kwág** mɛ́n ə̀m	body=2SG.POSS not touch child=1SG.POSS	‘Your body won’t touch my child.’
Low/IP-Initial	mʙʉ́ lù kə̀ **béb** ꜜbə́n ə́m	dog RPST NEG wait couscous=1SG.POSS	‘The dog didn’t wait for my couscous.’
High/IP-Initial	mvɛ́l ù **kə́** ꜜ**zwí** ʃún ə̀m	brother=2SG.POSS NEG hurt friend=1SG.POSS	‘Your brother didn’t hurt my friend.’
Low/IP-Medial	mʙwə́ lù kə̀ **vwáʔ** ꜜʧwɛ́n ə́m	goat RPST NEG throw wood=1SG.POSS	‘The goat didn’t throw my wood.’

Table 6.8: Stimuli for Experiment 4. Syllables hypothesized prominent are in bold. Clitic boundaries indicated with ‘=’

6.2.2 *Participants, Procedure, and Data Processing*

Data for Experiment 4 are drawn from the same recordings as were used for Experiment 3. To review, data were collected in Bangangté and Yaoundé, Cameroon. Participants in Experiment 4 included the same ten Medʉmba speakers, 4 male, 6 female, as were included in Experiment 3. Ages of participants ranged from 20 years to 60 years. All were native speakers who had spent the majority of their lives in Bangangté or the villages surrounding it. Subjects repeated the six target sentences were elicited at 15 different speech rates, from slowest to fastest, for a total of 90 trials per subject. Subjects were again given several practice trials to acclimate to the experiment; all subjects were able to complete the task after practice. During the course of the experiment, subjects

were given periodic breaks to rest and drink water. P-centers were marked semi-automatically and hand-corrected; relative phase positions were then calculated for each syllable.

6.2.3 Analysis Plan and Hypotheses

Alignment distributions of repetition alignment for each target syllable in each Form condition will be evaluated using the following steps. First, within each Form condition, deviation of each syllable's distribution from normality will be evaluated using Anderson Darling tests. Second, assuming alignments for each syllable show deviation from a normal distribution, distributions will be modeled using Gaussian mixture models. Results of Gaussian mixture models will be compared visually, and then, to quantify differences in phase position of target syllables across Form conditions, dominant components for each distribution—or those components which account for the most data while exhibiting the smallest standard deviation—will be evaluated for differences in relative phase across Form conditions using linear mixed effects modeling. Again, as with Experiments 1-3, the goal of this analysis is to compare the position of syllables occurring in relatively *comparable* positions across forms to see if clustering genuinely showed different alignment patterns across Form conditions. Thus, if dominant components to be compared across Form conditions are not the most comparable in alignment across the two conditions, additional comparisons of non-dominant (but more closely-aligned) distributions will also be carried out.

Hypotheses for Experiment 4 can be outlined as follows:

1. As with Experiments 1-3, syllable alignment was predicted to cluster around a relatively small number of phase positions
2. Alignment of target syllables was predicted to follow certain patterns. Specifically:
    (a) For Syllable 3, intonational phrase-initial syllables were predicted to align more closely with SHP positions of .333, .5, or .667 than phrase-medial syllables.

(b) Verb stem-initial syllables occurring in Syllable 4 were predicted not differ across Form conditions, and were predicted to show close alignment (<.03 phase) with SHP positions.

6.2.4 Results

Results for individual syllables are plotted in Figure 6.12, including regression lines with corresponding equations and R^2 values. Note that plots represent the same data as was presented for Experiment 3, but organized by the Form conditions relevant for Experiment 4. As can be seen, alignment for all syllables increased with speed in both Form conditions ($p < .0001$), but the degree of change by speed was greatest for syllables occurring the later in the sentence, with Syllables 5 and 6 showing the strongest linear trend ($R^2 > .40$)

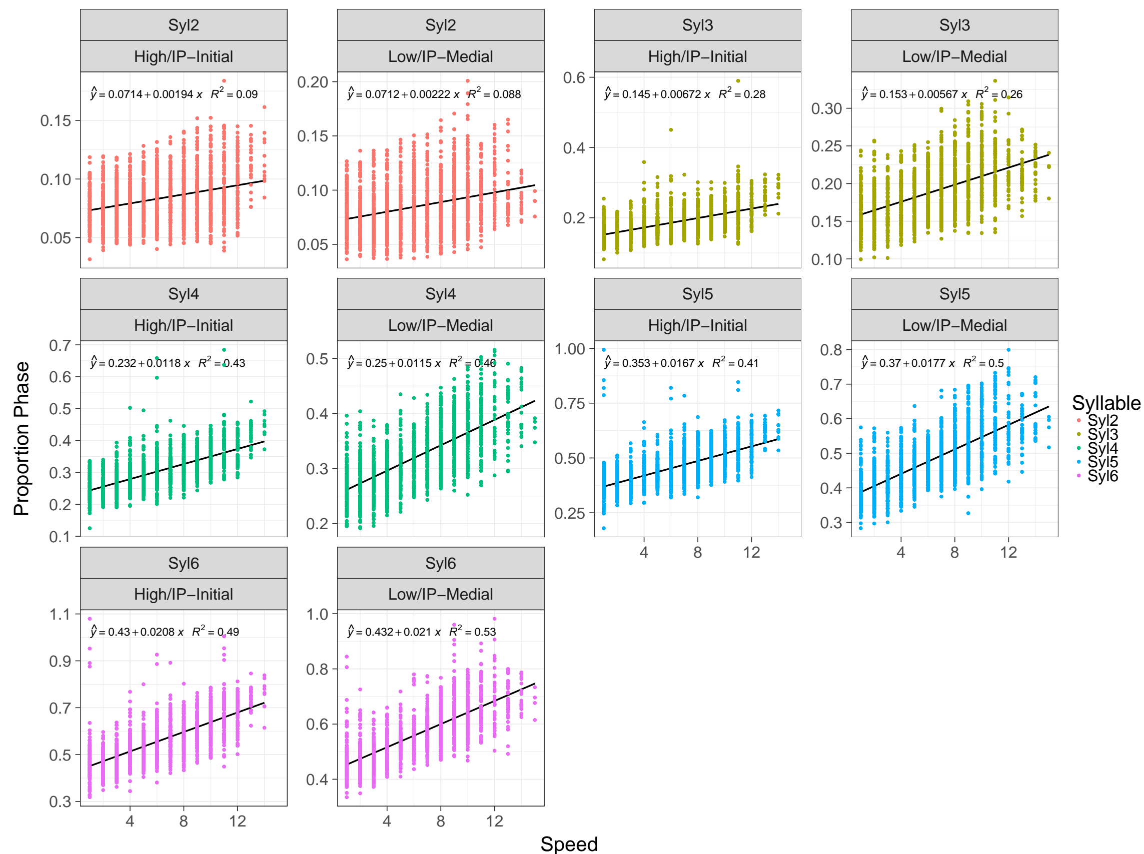


Figure 6.12: Distributions by syllable with regression lines, Experiment 4 (red=Syllable 2, yellow=Syllable 3, green=Syllable 4, blue=Syllable 5, purple=Syllable 6)

Heatmaps shown in Figures 6.13 and 6.14. Once again, these plots display frequencies of different syllable phase alignment positions by speed, with yellow, orange, and red dots indicating the positions with the largest concentrations of repetitions. The presence of concentrations of repetitions around certain phase positions even as metronome speed is in line with predictions from Hypothesis 1.

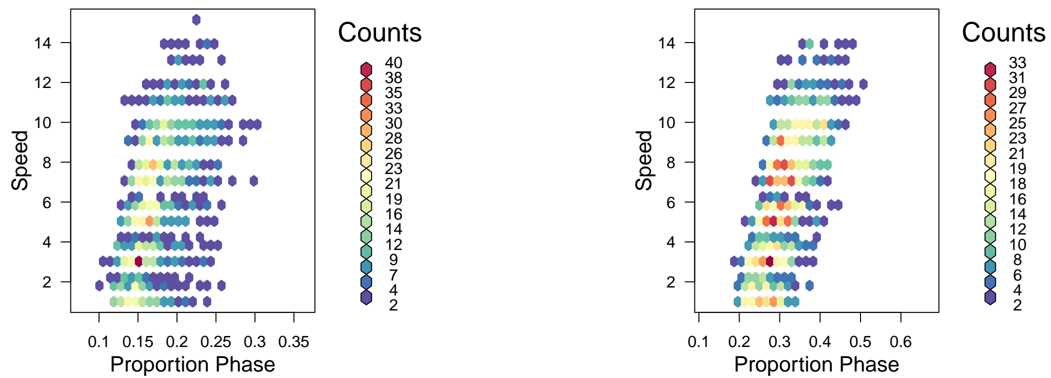


Figure 6.13: Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), HIGH/IP-INITIAL condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position

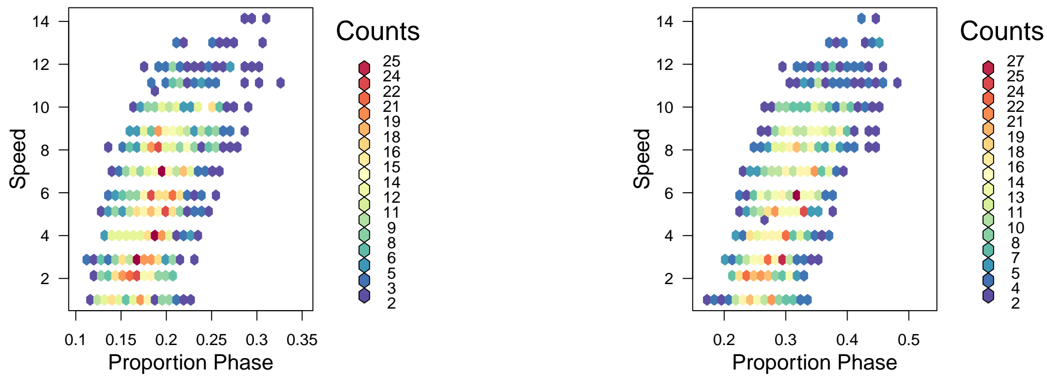


Figure 6.14: Hexagonal heat map of syllable alignment (Proportion Phase by Speed), Syllable 3 (left pane) and Syllable 4 (right pane), LOW/IP-MEDIAL condition, Experiment 4. Greater counts/warmer colors (e.g. yellow, orange, and red) indicate greater instance of syllable alignment at a particular position

To evaluate whether syllable distributions deviated from normality, Anderson-Darling tests were conducted for each syllable. Results of these tests are presented in Table 6.9, confirming that all distributions were significantly non-normal ($p < .0001$).

Form	**Syllable**	**A-D Statistic**	**p-value**
HIGH/IP-INITIAL	Syl 3	18.106	$< .0001$
	Syl 4	7.136	$< .0001$
LOW/IP-MEDIAL	Syl 3	7.553	$< .0001$
	Syl 4	5.905	$< .0001$

Table 6.9: Results of Anderson-Darling test for normality, Experiment 4

Once again, data for each syllable in each Form condition was modeled as a Gaussian mixture. First, parametric bootstraps of log likelihood ratio statistics were carried out to evaluate the tradeoff between model fit and complexity for models consisting between 1 and 4 component Gaussian distributions. Results of the bootstrap procedure are provided in Table 6.10 and indicate that 2 components were optimally fit for each target syllable in each condition. Thus, Hypothesis 1—that target syllables would be found to cluster around a relatively small set of phase positions—was borne out. Model parameters, including means (μ), standard deviations (σ) and mixing weights

(λ) for each distribution are shown in Table 6.11. Histograms of resulting mixtures are shown in Figures 6.15-6.17.

		HIGH/IP-INITIAL		LOW/IP-FINAL	
		Obs. LRS	p-value	Obs. LRS	*p*-value
Syllable 3	$K = 1$ vs. $K = 2$	219.072	$< .005$	103.552	$< .005$
	$K = 2$ vs. $K = 3$	6.458	=.06	3.110	= .44
Syllable 4	$K = 1$ vs. $K = 2$	87.398	$< .005$	73.594	$< .005$
	$K = 2$ vs. $K = 3$	4.507	= .25	11.063	= .04

Table 6.10: Observed log likelihood ratio statistics and *p*-values for parametric bootstrap test for number of components *K*, Experiment 4.

		HIGH/IP-INITIAL			LOW/IP-MEDIAL		
		μ	σ	λ	μ	σ	λ
Syllable 3	Component 1	.156	.018	.485	.185	.029	.875
	Component 2	.203	.035	.515	.252	.032	.125
Syllable 4	Component 1	.280	.036	.526	.285	.040	.643
	Component 2	.351	.053	.474	.354	.054	.357

Table 6.11: Model Parameters For Syllable 3 and Syllable 4: Experiment 4

In Figure 6.15, we see the resulting mixture for Syllable 3 in the HIGH/IP-INITIAL position. Recall that, per Hypothesis 3, this syllable, the head of an intonational phrase, should display greater prominence in terms of phase alignment close to SHP positions. As it turns out, component distributions show that alignment of this syllable—as well as Syllable 3 for the LOW/IP-MEDIAL condition—was timed quite a bit earlier than specified SHP positions of .333, .500, and .667. Instead, Syllable 3 in the HIGH/IP-INITIAL condition occurred with mean phase alignments of .156 (Component 1) and .204 (Component 2). As shown in Figure 6.16, for Syllable 3 in the LOW/IP-MEDIAL condition, alignment was at .185 (Component 1) and .252 (Component 2). If we compare results for Syllable 3 in the present experiment with those found for Experiments 1 and 2, it is clear that the introduction of additional syllables to the utterance for Experiments 3 and 4 has had the effect of shifting alignment of the first few syllables earlier within the phase. Since Syllable 3 resides far from the stated SHP positions of $\frac{1}{3}$, $\frac{1}{2}$, or $\frac{2}{3}$ positions, we can explore the

subdivision of these positions, the $\frac{1}{2}$ (.250) and $\frac{1}{6}$ (.167) positions. I will refer to these positions as HALF SHP POSITIONS. Dotted blue lines representing half SHP positions are drawn on Figures 6.15 and 6.16. As can be seen, Component 1 in the HIGH/IP-INITIAL position, which carries a mean of .156, occurs quite close ($< .011$) phase from the $\frac{1}{6}$ half SHP position. Component 1 for the LOW/IP-MEDIAL condition falls further away ($> .017$), but still close to the $\frac{1}{6}$ position. Component 2 for the LOW/IP-MEDIAL condition also falls exactly at the $\frac{1}{4}$ position.

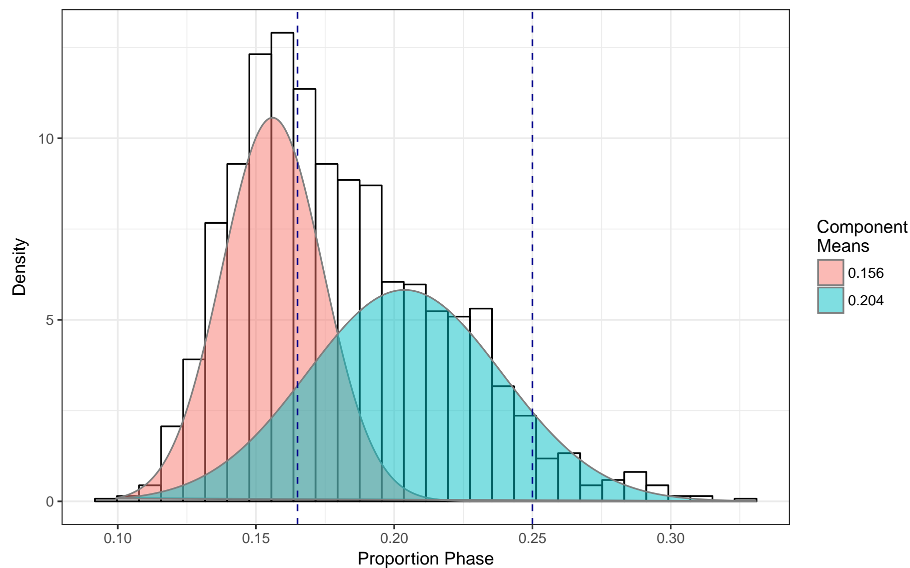


Figure 6.15: HIGH/IP-INITIAL Syllable 3: 2 component Gaussian mixture (Dotted blue lines indicate half SHP positions of .165 and .250)

Examining now Syllable 4, means for component Gaussians were quite similar across conditions. As seen in Figures 6.17 and 6.18, the means for Component 1 in both the HIGH/IP-INITIAL condition and the LOW/IP-MEDIAL condition occurred relatively close ($> .035$) to the $\frac{1}{4}$ (.250) SHP position (indicated with a dotted blue line), and Component 2 for both conditions was relatively close ($> .030$) to the $\frac{1}{3}$ (.333) SHP position (indicated with a dotted red line).

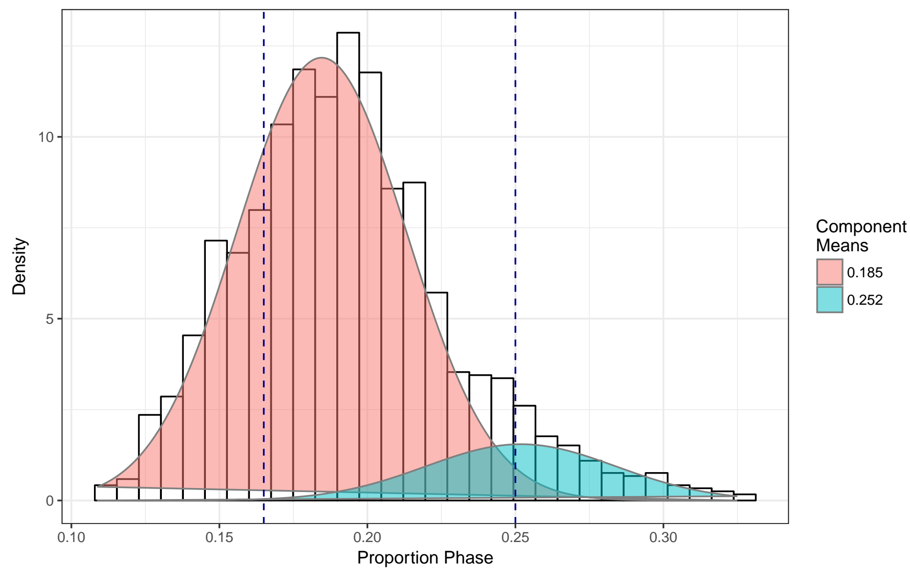


Figure 6.16: LOW/IP-MEDIAL Syllable 3: 2 component Gaussian mixture (Dotted blue lines indicate half SHP positions of .165 and .250)

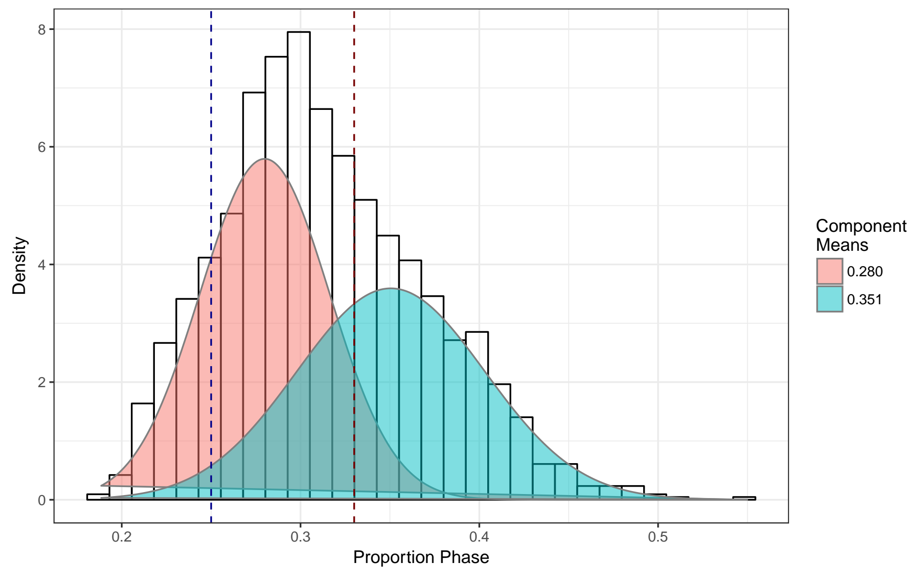


Figure 6.17: HIGH/IP-INITIAL Syllable 4: 2 component Gaussian mixture (Dotted blue lines indicate half SHP positions of .165 and .250)

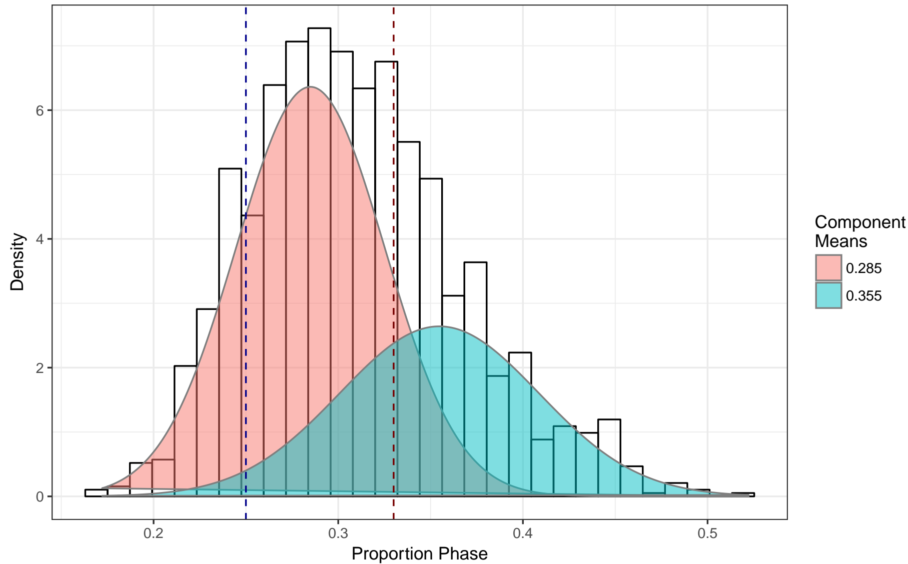


Figure 6.18: LOW/IP-MEDIAL Syllable 4: 2 component Gaussian mixture (Dotted blue line indicates half SHP Position of .25; dotted red line indicates SHP position of .330)

HIGH/IP-INITIAL Condition

Focusing now on patterns of alignment within each Form condition, we see from Table 6.11 that mixing weights (λ) for component distributions were fairly similar for each syllable in the HIGH/IP-INITIAL position, with Component 1 accounting for roughly half of the data for each syllable and Component 2 accounting for the other half. We can thus deduce that two overall patterns of alignment held for the data in this condition. Figure 6.19 provides a diagram of the two alignment patterns, Pattern 1 (in red) and Pattern 2 (in blue).

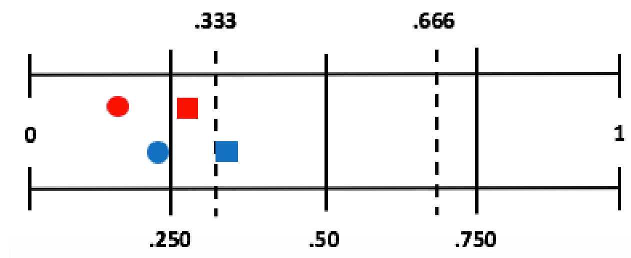


Figure 6.19: Patterns for Syllable 3 (circles) and 4 (squares), HIGHPRO Condition. Red = Pattern 1, blue = Pattern 2.

LOW/IP-MEDIAL Condition

For the LOW/IP-MEDIAL condition, mixing weights were not as evenly distributed between components. Rather, for Syllable 3, Component 1 accounted for over 85% of the data, while Component 1 for Syllable 4 only accounted for just under 65% of the data. Thus, it seems three overall patterns are necessary to capture the preferred alignments of subjects, all shown in Figure 6.20: Pattern 1 (in red) has Syllable 3 occurring around .185 and Syllable 4 occurring around .285, Pattern 2 (in blue) has Syllable 3 still occurring around .185 and Syllable 4 now occurring around .354, and finally, Pattern 3 (in green) has Syllable 3 occurring around .252 and Syllable 4 occurring still around .354.

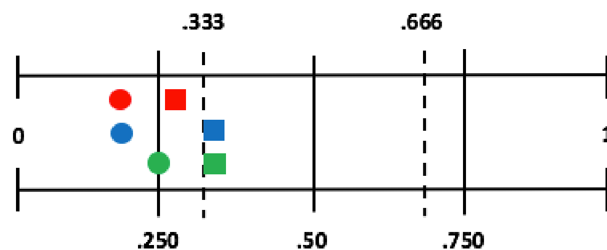


Figure 6.20: Patterns for Syllable 3 (circles) and 6 (squares), LOWPRO Condition. Red = Pattern 1, blue = Pattern 2, green = Pattern 3.

Comparing Dominant Components Across Conditions

Component 1 was assessed to be the dominant component (bolded in Tables 6.12 and 6.13) for both syllables in both Form conditions. It also happened to be the case that dominant components were the components which most closely aligned in mean phase proportion across Form conditions for each syllable, meaning that a significant difference in position across Forms for these components would entail differences for any other comparison across Form conditions. Plots of dominant components comparing between Form conditions are shown in Figure 6.21.

To evaluate whether a difference in relative phase position emerged across Form conditions for

Syllable 3 and Syllable 4, a linear mixed effects model was constructed for each syllable, including fixed effects of FORM and SPEED as well as their interaction, and by-subject random slopes for each variable. Full results are presented in Table 6.14. A significant effect of FORM was found for Syllable 3, indicating that this syllable in the HIGH/IP-INITIAL condition occurred significantly earlier and closer to the closest half SHP position of .167 than in the LOW/IP-MEDIAL condition (β = -.008, t = -4.129, $p < .01$). There was also a significant interaction between FORM and SPEED (β = -.001, t = -7.502, $p < .0001$). Referring back to Figure 6.12, we can see that the rate of change in alignment by speed for Syllable 3 was higher for the LOW/IP-MEDIAL condition than for the HIGH/IP-INITIAL condition. No significant effect of FORM was found for Syallable 4 (β = -.002, t = -.611, p = .55), nor was there a significant interaction (β = .006, t = -1.965, p = .0524).

	Syllable 3	Syllable 4
Component 1	**.037**	**.068**
Component 2	.068	.110

Table 6.12: Component strengths, Syllables 3 and 4, HIGH/IP-INITIAL condition

	Syllable 3	Syllable 4
Component 1	**.033**	**.062**
Component 2	.256	.151

Table 6.13: Component strengths, Syllables 3 and 4, LOW/IP-MEDIAL condition

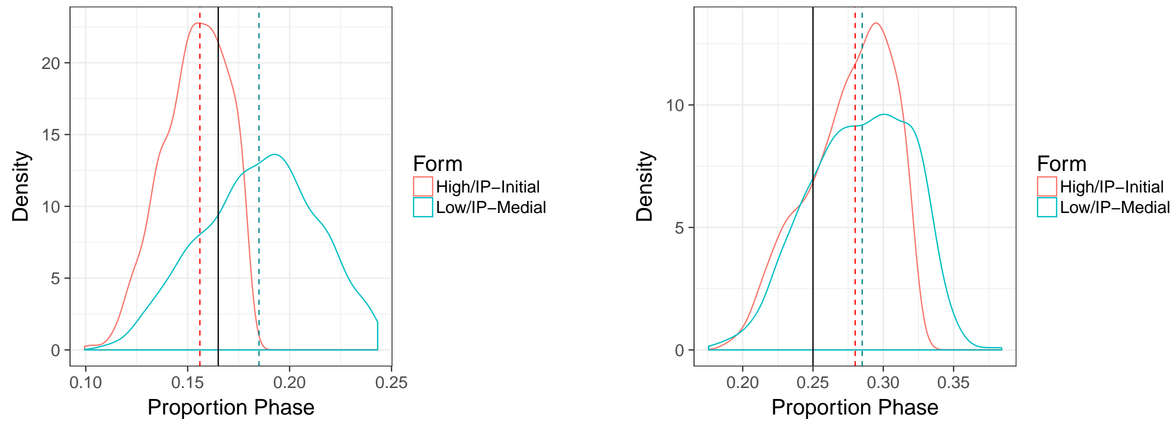


Figure 6.21: Comparison of dominant components for Syllable 3 (left) and Syllable 4 (right), Experiment 4. Black lines represent closest half SHP positions of .167 (left panel; Syllable 3) and .250 (right panel; Syllable 4). Red dotted lines indicate component means for the HIGH/IP-INITIAL condition, and blue dotted lines indicate component means for the LOW/IP-MEDIAL condition.

	Factor	**Estimate**	**Std. Error**	**df**	**t-value**	**p-value**
Syllable 3	Form	-.008	.0020	8	-4.129	$< .01$
	Speed	.003	.0005	8	7.225	$< .0001$
	Form × Speed	-.001	.0001	2319	-7.502	$< .01$
Syllable 4	Form	-.0018	.0026	12	-.611	.55
	Speed	.006	.0078	8	8.256	$< .0001$
	Form × Speed	.006	.0002	2141	-1.965	.0524

Table 6.14: Results of linear mixed effects models Syllable 3 and Syllable 4, Experiment 4

6.2.5 Discussion

To reflect, since Syllable 3 was found to occur quite early in terms of relative phase position, half SHP positions—predicted to be the next strongest phase positions after SHPs themselves—were incorporated into the analysis. If we use these half SHP positions for comparison, Hypothesis 2 appears largely to be supported: per Hypothesis 2a, the dominant component for the intonational phrase-initial syllables were found to align more closely with half SHP positions than the intonational phrase-medial syllable occurring in the same position, Syllable 3. It is worth noting, too, that a significant interaction between FORM and SPEED indicated that the intonational phrase-initial syllable underwent less change in phase position as speed increased as compared to the phrase-medial syllable. Finally, the intonational phrase-initial syllable also had a standard deviation which was smaller (σ = .018) than that of the phrase-medial syllable (σ = .029). The phrase-initial syllable's closeness to a half SHP position, its relatively small standard deviation, and its resistance to perturbations with changing speed are all good indicators that this syllable was treated with greater prominence than the corresponding phrase-medial syllable.

And per Hypothesis 3, the dominant components for Syllable 4, the verb, occurred relatively close to the half SHP position of $\frac{1}{4}$ (.250). For the LOW/IP-MEDIAL condition, the dominant component was slightly farther than the stated cutoff for closeness of .03, coming in at .285. Given that Component 2 for the verb in both conditions also occurred quite close to an SHP position, this time the $\frac{1}{3}$ (.333) position, it seems clear that the verb was regularly treated by speakers a s prominent syllable.

6.3 General Discussion: Relative Prominence and the Role of Tone

Overall, results from Experiments 3 and 4 showed several interesting patterns. First off, pronominal enclitics were once again shown to have elevated prominence, regardless of tone, and were even found to act as more prominent than the head nouns which preceded them. This pattern differs from

that found in Experiment 1, where both the head noun and the enclitic very clearly showed alignment at SHP positions. What could be driving these different patterns across experiments? One possibility is that what I have been referring to as pronominal enclitics in fact form a heterogenous group such that the specific forms tested in Experiment 3, *àm* and *ám*, rather than bearing their own accent, are perhaps more affix-like and can in fact induce a kind of 'stress shift' similar to affixes like English *-ion* which shift stress away from its original position (e.g. *accómmodate* → *accommodátion*), or similar to the case of Turkish *k*-paradigm endings discussed by Good & Yu (2010). This would, however, be an indication that Medʉmba had word-final or iambic metrical alternations at the word level, which, given the results of Experiment 2 for the SISF condition which showed stem-final syllables in disyllabic forms to be less prominent, seems unlikely. Treatment of Medʉmba as an iambic language also contradicts a host of other types of patterns discussed in Franich (2014), including from tone spreading, loanword adaptation, and restrictions on contour tone placement, which point to, if anything, trochaic structures.

Another possibility is that the specific pattern of syllables in the target sentences for Experiment 3 and 4 led to a situation where the noun in penultimate (Syllable 5) position was in fact deaccented. Looking at the overall patterns, it is clear that the verb in Syllable 4 position regularly exhibited elevated prominence, occurring at SHP or half SHP positions consistently. What about the rest of the syllables? Looking at dominant components for Syllables 3, in the HIGH/IP-INITIAL condition, this syllable occurred around phase position .156 which, as mentioned, is very close to the $\frac{1}{6}$ position in that phase (a half SHP position of $\frac{1}{3}$. The dominant component for Syllable 4 across HIGHPRO and LOWPRO conditions was around .280, just .030 away from the half SHP position (this time for the $\frac{1}{2}$ SHP) of $\frac{1}{4}$. Measuring the mean value for the distribution of Syllable 2, whose distribution, incidentally, was best modeled as a single Gaussian (rather than a mixture), it falls at .075, which is very close to .082, or the $\frac{1}{12}$ position in the phase. It seems, then, that the most popular alignment strategy resembled a subdivision of the phrase in 12, which in western musical notation could be written in 12/8 meter. This subdivision strategy can also help us to place

overall positioning of Syllables 5 and 6, which, in the HIGHPRO condition, occurred at .431, a sight bit later than .416 corresponding to the $\frac{5}{12}$ position, and .501, corresponding to the $\frac{1}{2}$ SHP.

Thus, it seems the noun in Syllable 5 position occurred at the $\frac{5}{12}$ position, which is, in fact, the least prominent position for a syllable to occur. Why would this be, given that stem-initial syllable, including in nouns, have been found to align with prominent positions? It is possible that increased number of syllables in Experiments 3 and 4 may have led some syllables which are generally capable of aligning with prosodically prominent positions to essentially become 'de-accented', allowing neighboring syllables—such as the preceding verb and the following pronominal enclitic—to align with prominent positions. Of course, this is quite a speculative claim, and further data will be needed to evaluate whether or not this is a viable explanation. However, in principle, it is of course possible for words crosslinguistically to vary in relative prominence depending on how they are parsed prosodically in a given string, and various factors, including syllable count, syntactic structure, and focus/communicative function may play a role in shaping prosodic structure. Further exploration of these influences must be left for future work.

What the above breakdown of syllable timing patterns also tells us is that, at least some of the time, speakers were adopting a subdivision pattern of 12 'beats' which is compatible with SHP positions which form both binary (four patterns of 3 beats) and ternary (3 patterns of of 4 beats). Thus, is possible that earlier findings regarding the apparent 'dual prominence' of of sequential stem-initial syllables (or stem initial syllables followed by enclitics) were, indeed, representative of simultaneous prominence in both the $\frac{1}{3}$ and $\frac{1}{2}$ SHP positions.

With regards to the comparison of Syllable 3 across intonational phrase-initial and medial-positions, evidence from comparison of dominant components did support the hypothesis that the high, phrase-initial negation marker occurred at relatively stronger phase positions than did the low, phrase-medial negation marker. This supports the notion that phrase-initial position is more prominent than non-initial position, though the fact that even the high negation marker did not occur near one of the stated SHP positions of .333, .500, or .667 suggests that this position is not

as prosodically prominent as, say, the position where the verb or the final pronoun occurred.

All in all, the patterns discovered suggest that tone is associated with prominence in Medʉmba only insofar as high tone is attracted to positions of prominence at the phrase level. We have seen ample evidence from stem-initial syllables, prefixes, stem-final syllables and enclitics that tone does not play a role in mitigating relative syllable prominence at the word level.

CHAPTER 7

PATTERNS OF INDIVIDUAL VARIATION AND RHYTHMIC SYNCOPATION IN THE SPEECH CYCLING TASK

In this chapter, we turn to another novel result from the experiments presented in Chapters 5 and 6 which has not yet been described, and which concerns alignment patterns used by speakers with respect to the metronome beats of the speech cycling task. Recall that, for the general speech cycling task, relative phase position is measured in terms of the distance between successive repetitions of the first syllable of each target phrase. A diagram for calculation of relative phase (what Tajima 1998 refers to as External Phase) is again given in Figure 7.1.

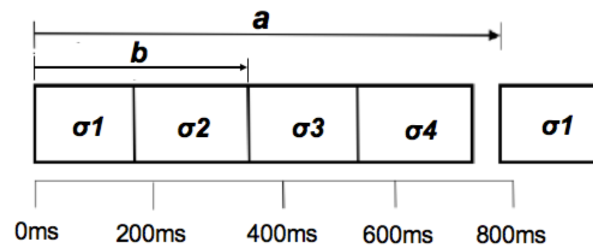


Figure 7.1: Sample measure of External Phase. External Phase for Syllable 2 is the interval of *b* divided by the interval of *a*.

As can be seen in Figure 7.1, the metronome beat itself does not form a part of the calculation of External Phase. Thus, while the metronome serves as a guide for subjects in terms of regulating speaking rate, its position relative to the utterance is not crucial to determine for the purposes of phase position calculation. However, in order to provide clear guidelines to subjects at the start of the task, researchers typically instruct subjects to align the first syllable of each utterance with the beat itself Cummins (1997a); Tajima (1998); Zawaydeh *et al.* (2002). This pattern has evidently been intuitive enough to subjects in previous speech cycling research as in no case was it reported that subjects attempted to align sentences in any other way, and, for at least the experiments on English and Japanese, speakers are confirmed to have naturally gravitated to the on-beat metronome alignment pattern (Fred Cummins, personal communication). In Targeted Speech Cycling, where

alternating high and low metronome beats are provided, alignment of the first syllable to the initial metronome beat is in fact essential, as phase manipulation on the task crucially takes into account the distance between beats of the metronome. Therefore, it came as some surprise when Medʉmba speakers completing the general task, rather than aligning their first syllables with the metronome beat, aligned them slightly *after* the beat, apparently on the 'off-beat'. Indeed, all subjects across all experiments showed this pattern consistently, without having seen one another complete the task. This was in spite of the fact that subjects were instructed to align initial syllables with the metronome beat.

In some ways, it might not come as much of a surprise that in the Medʉmba context, there should be a preference for off-beat alignment. Music familiar to these speakers—both traditional and modern—involves heavy use of syncopated rhythms, similar to the situation in much of the rest of Africa (Waterman, 1948; Temperley, 2000). One possibility, then, is that speakers were simply opting for a more musical approach, perhaps breaking up the monotony of the task by taking some creative license and adding their own rhythmic flourishes. However, if this were the case, we would likely expect some variation among subjects in the specific patterns they chose. This was not the case, however: all subjects displayed, for at least the initial trials of the experiment, off-beat 'anti-phase' alignment with the metronome.[1] A schematic diagram of in-phase versus anti-phase alignment of syllables and metronome beats is provided in Figure 7.2, and corresponding in-phase and anti-phase couplings of sine wave oscillations are shown in Figure 7.3. The reader is encouraged to listen to sound files of the raw data at http://home.uchicago.edu/kfranich/alignmentpatterns demonstrating anti-phase metronome alignment from four speakers, along with versions of those speakers' same repetitions with the metronome beats re-aligned for an in-phase pattern. Note that an additional beat has been inserted halfway through the larger period to demonstrate that realignment to in-phase results in the final syllable occurring exactly between successive beats of the

1. Though many factors may contribute to the precise phasing relation subjects displayed at any point in time (see, for example, Chapter 4 for a discussion of *phase drift*, which can lead to patterns which are neither perfectly in-phase nor anti-phase), subjects demonstrably showed anti-phase or close to anti-phase behavior on some trials.

	In-Phase	Anti-Phase
SYLLABLE LEVEL	> > > >	> > > >
METRONOME LEVEL	x x x x	x x x x

Figure 7.2: Schematic representation of in-phase and anti-phase alignment between metronome beats and syllables in phrase repetitions

Figure 7.3: In-phase (left) and anti-phase (right) coupling of oscillators at the metronome and phrase levels

original, larger period. Also note that no alteration to the larger period of the metronome beat has been made; the beat as it was played for subjects during the task was simply shifted later with respect to the phrase repetition.

The consistency of behavior among subjects on the task suggests that subjects were not simply taking creative license; rather, it seems the anti-phase alignment pattern chosen by subjects reflects a deeper intuition about how the task should be done. What I will suggest is that this pattern may reflect the specific way in which Medʉmba speakers align prominent elements in speech with underlying metrical rhythms. To begin to understand the influence that different alignment patterns might have on behavior in the speech cycling task, I first take a look at some of the individual-level differences exhibited by Medʉmba speakers on the task. I then use these patterns as a starting point to discuss another phenomenon—tonal anticipation—which is common in African languages and which I will argue may relate to some of the same principles which drive anti-phase alignment with metronome beats in the speech cycling task among speakers.

7.1 Individual Level Variation in the Task

Looking at overall patterns in the data from Experiments 1-4, one is struck by the different ways subjects patterned at increasingly faster speech rates on the speech cycling task. For example, looking at the data as a whole for Experiment 2 in Figure 7.4, while all subjects appear to have

deviated somewhat from a perfectly linear relationship between metronome speed and proportion phase, some subjects displayed considerably more linear patterns than others. For example, looking at Syllable 4 in the PRESI condition, we see that subjects s1 and s11 maintained a quite linear pattern of phase change as metronome speed increased, while others, such as s6 and s13, showed what looks to be a more quadratic relationship between speed and proportion phase, with changes in proportion phase decreasing as speech rate got faster.

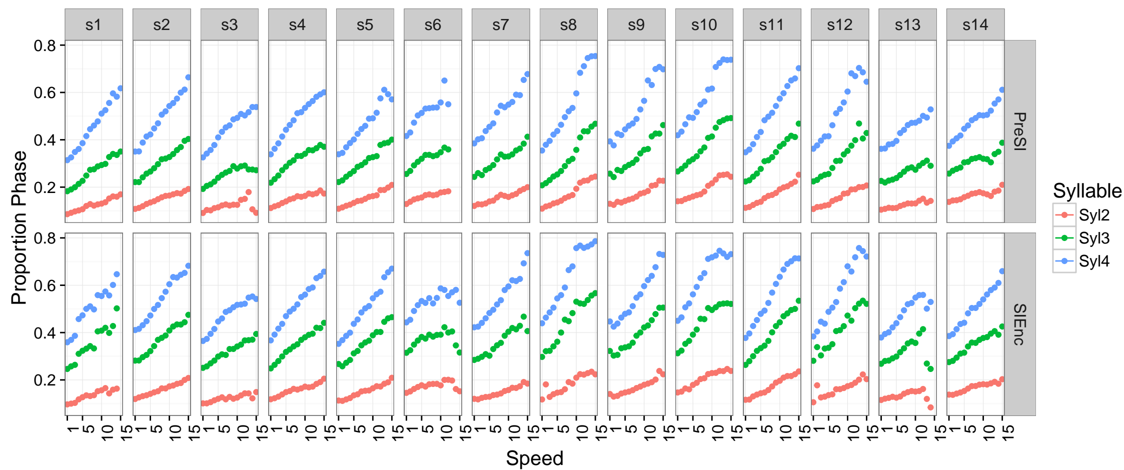


Figure 7.4: Scatter plot of phase angle by metronome speed (speech rate) by individual, Experiment 1 (red=Syllable 2, green=Syllable 3, blue=Syllable 4)

To allow a more direct visual comparison, data for (stem-initial) Syllable 4 in the PRESI condition is displayed for subjects s11 and s6 in Figure 7.5. As can be seen, s11 shows some evidence of clustering around the .5 phase position (SHP position of $\frac{1}{2}$) around speeds 8 and 9, but continues with a relatively linear pattern for the remainder of the task. Subject s6, on the other hand, displays considerable plateauing in phase proportion starting at around metronome speed 7, and then maintains relatively close alignment with the .5 position for several more trials, and remains relatively close to this position for the rest of the task.

To gain a better understanding for where these patterns of individual variation may stem from, I now present a more in-depth look at data from a subset of subjects. Since space constraints do not permit me to go in-depth into the behavior of all speakers across all experiments, I will limit

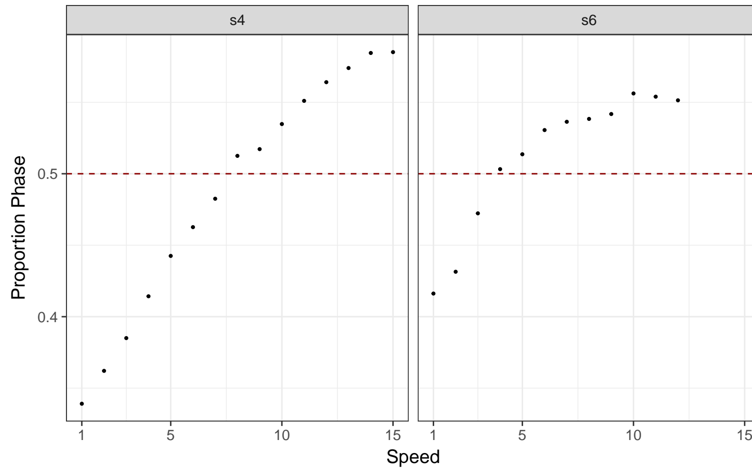


Figure 7.5: Comparison of Syllable 4, Experiment 1, PRESI Condition, Subjects s6 and s11 (Red dotted line = SHP position of .5)

discussion to six subjects who appear to exemplify some of the most diverse patterns of alignment in the task, and will also limit discussion to data from Experiment 2, and specifically the final stem-initial high tone syllable in the PRESI condition. Since the fourth syllable in this condition very clearly showed evidence of prominence and was not preceded by another syllable bearing prominence, it provides the clearest picture of alignment strategies with prominent syllables. Given that individual subjects used similar alignment strategies across experiments, we can assume that the generalizations to be discussed hold more generally for these subjects across subexperments with the task.

7.1.1 Demographics of Six Speakers

Demographic information for the six subjects of interest, 3 female and 3 male, is provided in Table 7.1. Recent work documenting the complex sociolinguistic dynamics in the Cameroonian Grassfields (Di Carlo & Good, 2014; Di Carlo, 2015, In press.,I) has highlighted both the extensive multilingualism and multilectalism among Cameroonians from that geographical area as well as the intricate ways in which language is used to express group affiliation, most notably with one's native village (or 'village-chiefdom', to use Di Carlo's 2017 term). Though far more work would

Subj.	Sex	Age	Langs	Village	Spouse's Langs	Mother's Langs	Father's Langs Langs	Lived Outside Bangangté?
s1	M	22	Meʉmba French Pidgin	Bangangté	N/A	Medʉmba French	Medʉmba French	Bangangté area whole life
s3	F	45	Medʉmba French Pidgin	Bangoulap	Medʉmba	Balengou Medʉmba	Medʉmba French	7 years in Bandjoun as a child
s4	M	27	Medʉmba French	Bangoulap	N/A	Medʉmba French	Medʉmba French	Bangangté area whole life
s6	F	43	Medʉmba Bafang Ghomalá	Bangoulap	Medʉmba	Ghomalá	Ghomalá English	Born in Bangangté but lived in Bafoussam; moved to Bangoulap during childhood
s11	F	21	Meʉmba French Pidgin	Bahouoc	Medʉmba French	Medʉmba French	Medʉmba French Pidgin	Bangangté area whole life
s13	M	21	Meʉmba French Pidgin	Babou	N/A	Medʉmba French	Medʉmba French Pidgin	Bangangté area whole life

Table 7.1: Demographics of six Medʉmba speakers showing varied behavior in speech cycling

be needed to develop a comprehensive understanding of these Medʉmba speakers' language backgrounds, I have attempted to provide some clarity on this issue by eliciting, in addition to standard demographic information such as age and sex, information on languages spoken by subjects, their spouses (if applicable), and their parents, the villages where they grew up, and any amount of time they have lived outside of the greater Bangangté area.[2][3]

As can be seen from Table 7.1, subjects ranged in age from 21 to 45, and are all multilingual (all speaking French, and some Cameroonian Pidgin), though they all consider themselves to be native Medʉmba speakers.[4] All subjects had lived the majority of their lives in Bangangté or one of the surrounding villages, though two subjects had lived for some period of time slightly

2. It is common, for example, for children from villages around Bangangté to go and stay with family members in the city while studying or working, and for children born in the city to return to the village for a period of time to stay with grandparents or other relatives.

3. Additional demographic information, such as subjects' *ndab*, or traditional names, marriage status, spouse's native village(s), as well as relative fluency in the languages they report speaking, was elicited, but is not reported here.

4. The question of nativeness and/or language dominance is a complex one, particularly in a deeply multilingual society such as Cameroon, and especially where nativeness is tightly tied into notions of group affiliation as much as functional fluency (see Di Carlo & Good 2014 for further discussion). In addition to being asked which languages they spoke, individuals were asked to rate on a five point scale how fluent they were in each language as far as speaking, understanding, and writing. While only one subject reported being able to write in Medʉmba (it is only since the 90s that Medʉmba and other mother tongue languages in the region are being formally taught in schools), all reported strong speaking and understanding skills.

outside of the area. Subject s3, for example, had lived for 7 years during childhood in Bandjoun, a village north and slightly west of Bangangté off of the N4 highway toward Bafoussam, a relatively large city of about 250,000 people located in the Mifi department. Bafoussam is a major trade center, in which coffee, tobacco, and tea are farmed and two major markets draw people from many surrounding villages. The primary ex-colonial language of Bafoussam is French (as it is in Bangangté), but it is notable that Bafoussam lies only 77 kilometers from Bamenda, the center of the English speaking zone in Cameroon. Subject s3 reported her mother's dominant language was Balengou, which is considered to be a dialect of Medʉmba (Breedveld & Mous, 1086). Another subject, s6, was born in Bangangté but lived in Bafoussam until she was around 10 years old, when she moved back to the village of Bangoulap outside Bangangté. This subject reported that her parents' native language was not Medʉmba but Ghomalá, another Bamileke language, and one of Bafoussam's primary languages. We will return to the details of speakers' language backgrounds after we discuss quantitative patterns of variation in the speech cycling task in the next section.

7.1.2 Quantifying Individual Variation

We now move on to discuss patterns of inter-speaker variation more in-depth, and to attempt to characterize them in a quantitative way. Figure 7.6 displays syllable alignment patterns for the six subjects of focus for Syllable 4 in the PRESI condition. As can be seen, some subjects, such as s1 (red) and s11 (darker blue) showed relatively linear changes in proportion phase with changes in speed. For others, such as s3 (yellow), s6 (lighter blue), and s13 (pink), the rate of change in proportion phase by speed seems to have tapered off at higher speech rates, leading to what appears to be a more quadratic pattern of results. Subject s4 (green) also seems to show a similar pattern, though not as strong as for subjects s3, s6, and s13.

In order to verify different patterns in change of phase proportion by speed across subjects, each subject's data were fitted to three different models, a linear model, a linear model with a quadratic term, and a linear model with a quadratic plateau term, in the `easynls` package for *R*. Model

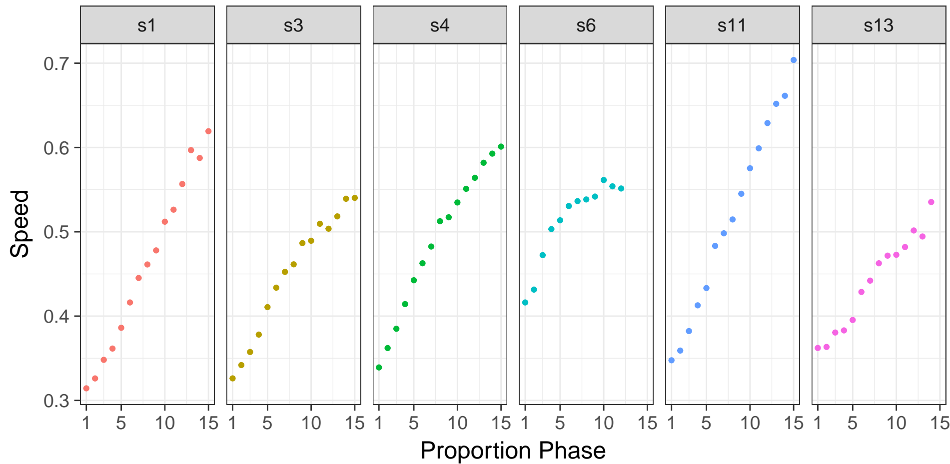


Figure 7.6: Comparison of Syllable 4, Experiment 1, PRESI Condition, 6 subjects

fits were then compared for both Akaike's Information Criterion (AIC) and Bayes Information Criterion (BIC); results are displayed in Table 7.3 (the lowest value for both AIC and BIC, even when negative, is preferred).

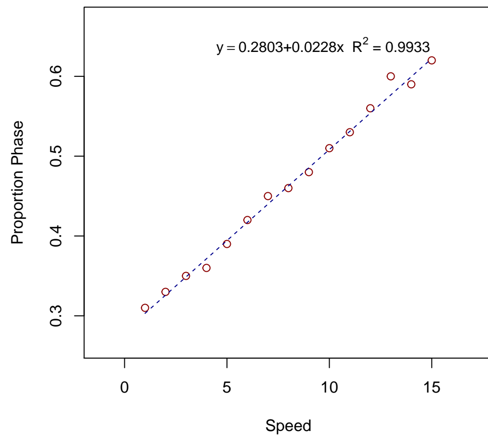
Graphs of best-fit models are shown in Figure 7.7 with corresponding R^2 values, and full model parameters are provided in Table 7.2. As can be seen, the distribution of phase proportion by speed for subjects s1 and s11 was best fit with a linear model, while subjects s3, and s4 were best fit with quadratic models. For subject s13, AIC and BIC values gave conflicting results: BIC, which penalizes models with greater numbers of parameters more than AIC, favored the linear model, while AIC favored the quadratic model. Given the quadratic model had a numerically higher R^2 value and appeared upon visual inspection to be a better fit to the data, this model was chosen. Finally, data for subject s6 was best fit with a linear with plateau model, indicating that this subject's alignment of the syllable ceased to change at higher speech rates.

Subject	Model	Coef$_a$	Coef$_b$	Coef$_c$	p-value$_a$	p-value$_b$	p-value$_c$	adjusted R^2
s1	Linear	.280	.023	N/A	$< .0001$	$< .0001$	N/A	.84
s3	Quadratic	.291	.027	-.001	$< .0001$	$< .0001$	$< .0001$	.81
s4	Quadratic	.306	.030	-.001	$< .0001$	$< .0001$	$< .0001$	.90
s6	Quad. Plateau	.379	.037	-.002	$< .0001$	$< .0001$	$< .0001$	.60
s11	Linear	.313	.026	N/A	$< .0001$	$< .0001$	N/A	.85
s13	Quadratic	.338	.015	-.001	$<.0001$	$< .0001$	$< .001$	.79

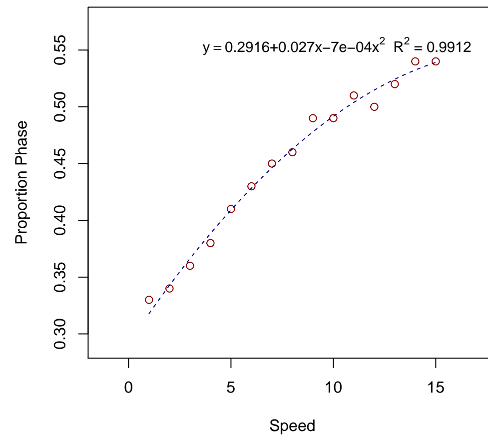
Table 7.2: Model parameters for best-fit models, Syllable 4, PRESI condition, 6 subjects

	Linear		**Quadratic**		**Quadratic w/ Plateau**	
	AIC	BIC	AIC	BIC	AIC	BIC
s1	**-1209.094**	**-1197.572**	-1207.653	-1192.291	-1207.653	-1192.291
s3	-1248.958	-1237.662	**-1282.588**	**-1267.527**	-1282.588	-1267.527
s4	-1447.684	-1436.179	**-1498.800**	**-1483.460**	-1498.800	-1483.460
s6	-825.609	-815.360	-863.428	-849.763	**-863.915**	**-85.251**
s11	**-1123.984**	**-1112.542**	-1122.086	-1106.829	-1122.086	-1106.829
s13	-1031.0842	-102.379	**-1031.121**	**-1016.848**	-1031.121	-1016.848

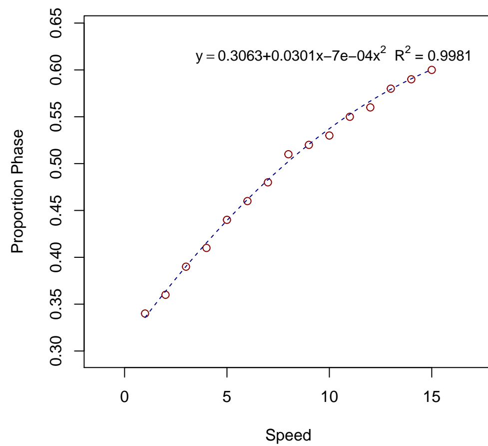
Table 7.3: AIC and BIC Values for comparison of linear, quadratic, and quadratic with plateau models for Syllable 4, PRESI condition, 6 subjects; values for winning models for each subject are bolded



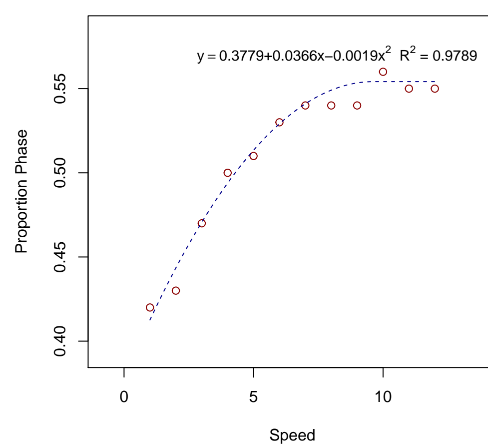
s1: Linear model



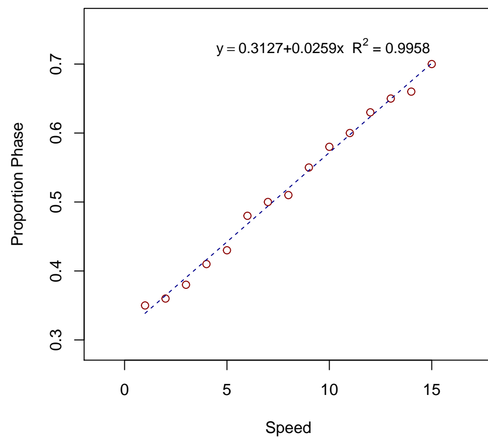
s3: Quadratic model



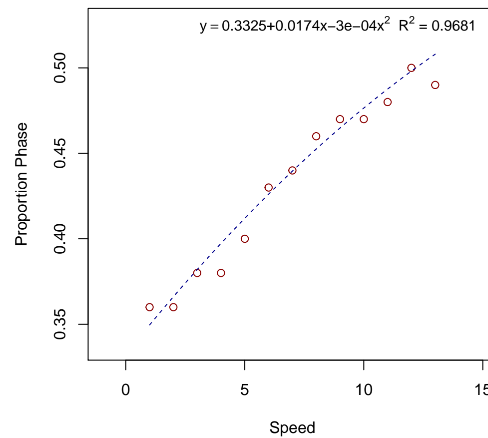
s4: Quadratic model



s6: Quadratic with plateau model



s11: Linear model



s13: Quadratic model

Figure 7.7: Results of linear and nonlinear model comparison for 6 subjects, Syllable 4, PRESI condition

7.2 Explaining Patterns of Variation: Rate-Driven Shifts in Metronome Alignment

What could explain the different patterns of alignment across these six subjects? One important clue lies in examining the alignment of phrase repetitions *with respect to the metronome*. As mentioned at the start of this chapter, subjects all aligned their repetitions with the first syllable on the offbeat of the metronome beat, forming an out-of-phase or even perfectly anti-phase relationship between repetitions and the metronome cycle. While all subjects maintained this pattern at slower speech rates, subjects showed different patterns at higher speech rates, with some shifting their alignment pattern to be in-phase with the metronome once they reached a certain metronome speed. This can be observed in Figure 7.8, which displays phase position relative to the metronome (rather than relative to the repetition cycle) at odd numbered speeds, from 3-15.[5] In this figure, the dotted black line represents the click of the metronome. All subjects can be seen to start out using a near perfect anti-phase coordination strategy, clustering at speed 3 between phase positions of .55 and .7. As the speed of the metronome increases, however, subjects show quite different trajectories. Subject s1, for example (in red), can be seen to maintain distance between his repetition of Syllable 4 and the metronome beat for all speeds of the task, though his repetitions do move somewhat closer to the beat as speed increases. Subject s6, on the other hand (in light blue), shows a rapid change to in-phase pattern with the metronome, which she maintains for the rest of her repetitions. Subjects s3, s4, and s13 show more gradual changes in phase, but all show a plateau once their repetitions of Syllable 4 get close to the positinon of the metronome beat. Finally, subject s11 (in dark blue) shows a zigzag pattern, such that she seems to be pulled closer to the metronome (especially at speed 11) and then moves farther away. This could be an indication that the subject was attempting to maintain anti-phase alignment through PHASE CORRECTION, a phenomenon which has often been noted in the context of sensorimotor synchronization tasks such

5. Data for speed 1 was omitted due to high variability shown at this speed—some subjects showed great difficulty coordinating with the metronome at this speed, as it was evidently too slow relative to their average speaking rate

as finger tapping (Repp, 2001, 2002a,b, 2008).

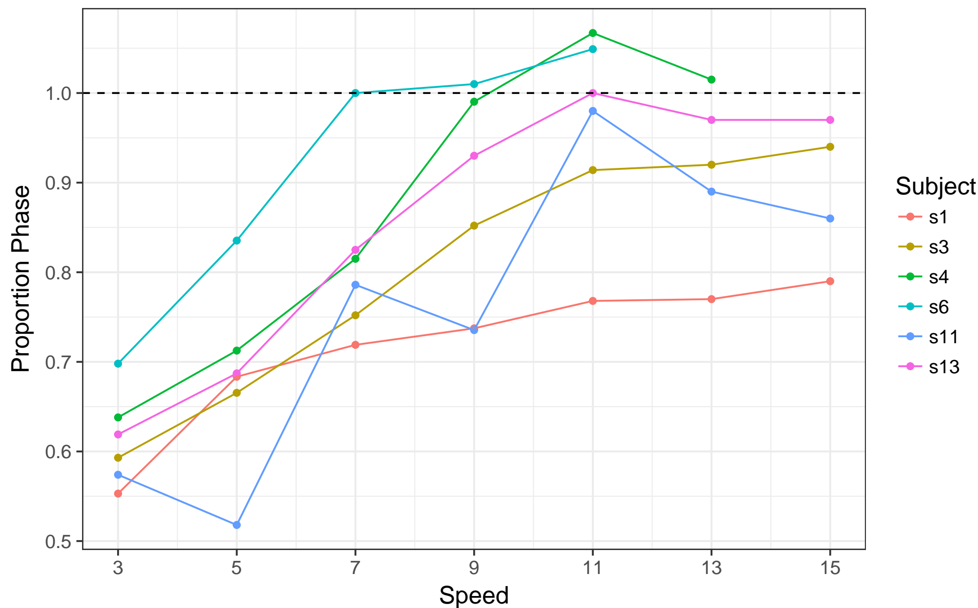


Figure 7.8: Alignment of Syllable 4 with respect to metronome beat, 6 subjects (Black dotted line indicates metronome beat)

These differing patterns bring up a number of questions. First off, why should it be that subjects such as s6 who fell in-phase with the metronome should only do so at higher speech rates? Recall from Chapter 4 that the Haken Kelso Bunz (henceforth, HKB) model of coordination predicts that in-phase patterns of coordination are inherently more stable than anti-phase patterns, and that this increased stability is best observed as the control parameter—in this case, speech rate—is raised. Specifically, as speech rate increases, anti-phase attractors (represented by the basins of attraction from Figure 4.3, Chapter 4, reproduced below as Figure 7.9) become less stable, while in-phase attractors grow in strength. Thus, anti-phase coordination with the metronome becomes less stable as speech rate increases. This finding mirrors work in motor coordination which shows that transitions from the anti-phase mode to the in-phase mode are often observed at faster movement frequencies (Kelso, 1984; Kelso *et al.*, 1990; Schmidt *et al.*, 1998).

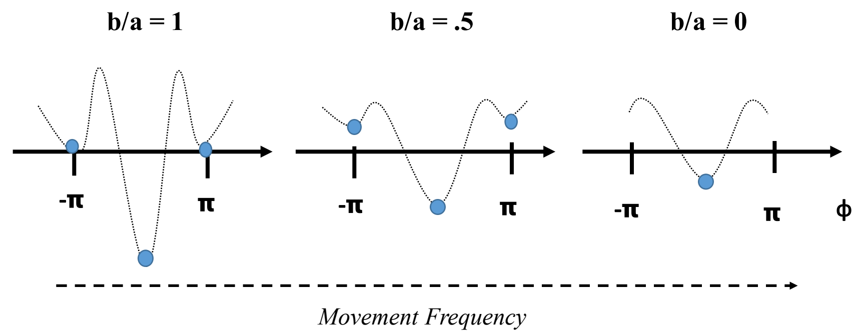


Figure 7.9: Changes in stability of attractors with increasing movement frequency (Haken *et al.*, 1985). Movement frequency increases left to right, and top to bottom.

A further interesting question is, why should shifting to in-phase coordination with the metronome lead to greater clustering—or PHASE LOCKING—of prominent syllables around SHP positions? Research by Keller & Repp (2008) on multilevel coordination may shed some light on the answer to this question. Keller and Repp (henceforth, K & R) had subjects tap their hands in anti-phase coordination to a metronome tone whose period was progressively shortened (which they refer to as *inter-agent coordination*) while alternating between their two hands (which they refer to as *intra-personal coordination*). Subjects heard feedback tones when they successfully tapped anti-phase with the metronome. The tones could come in one of five frequency levels, and could either match or mismatch between hands, and either match or mismatch with the metronome tone. The researchers found that simultaneous coordination with the metronome and between hands was most facilitated at higher rates where tones triggered by the two hands were distinct from one another and close in frequency—but not identical—to the metronome tone. The authors suggest that the perceptually close but still distinct nature of the optimal feedback tones makes them more easily to integrate within a single task-goal representation, as opposed to two separate task-goal representations (one linked to the inter-agent task and the other linked to the intra-personal task). They suggest that integrating task-goal representations in complex tasks may facilitate multilevel coordination by promoting a single locus of auditory and perceptual control for both levels of coordination.

The speech cycling task involves a very similar type of multilevel coordination to K & R's metronome task. Just as subjects in K & R's task had to coordinate in-phase the tapping gestures of their two hands, subjects in the speech cycling task performed an intra-personal coordination task by attempting to maintain in-phase coordination between phrase- and foot-level structures within phrase repetitions. Also similar between the two tasks was the dimension of inter-agent coordination, which in the speech cycling task consisted of coordination anti-phase with the metronome (though speech cycling subjects were not required to use anti-phase coordination, as they were by the task guidelines in K & R's experiment). In order to maintain simultaneous coordination in both tasks and to reduce overall attentional demands, speech cycling participants may have, at higher speech rates, opted to coordinate behavior within each task in the in-phase mode and also to coordinate tasks in-phase with one another. Per K & R's analysis, this would allow for a single overall coordinative mode between the two tasks. This change in coordinative modes was also accompanied by an increase in phase-locking within the intra-personal phrase repetition task, likely due to the fact that the strength of the in-phase attractor was greatest at higher speech rates, possessing an even greater amount of potential energy, and leading prominent syllables to exhibit greater phase locking.

Should this analysis be on the right track, the shift to in-phase alignment between phrase repetitions and metronome beats and the concomitant phase-locking of prominent syllables in phrase repetition cycles demonstrates that use of a single, in-phase coordinative mode leads to very different alignment patterns among prominent syllables when compared with the use of distinct coordinative modes between phrase repetition and metronome alignment tasks. With universal in-phase alignment, the timing of prominent syllables appears to be far more constrained in speech. Based on patterns from speech cycling and other motor behaviors we have discussed which suggest that English speakers tend towards in-phase coordination, this may be one of the reasons that, generally speaking, timing of stressed syllables in a language like English is, while not perfectly isochronous, largely predictable in its timing. This is evidenced by, for example, metrical priming studies which

show that English speaker-listeners depend heavily on stress for speech planning and production (Dilley & McAuley, 2008; Shaw, 2013).

If in-phase alignment is inherently more stable than anti-phase alignment, why is it that Medʉmba speakers, similar to English speakers or Japanese speakers, do not opt for in-phase coordination with the metronome at all speech rates? I would like to argue that anti-phase alignment with the metronome in the speech cycling task reflects a deeper property of the grammar of Medʉmba speakers: specifically, unlike Japanese, English, or Arabic speakers, Medʉmba speakers have a *default anti-phase alignment grammar* in which syllable-level 'beats' in speech are preferably coordinated anti-phase with a metrical grid. As we have seen, constraints on the perceptuo-motor system dictate that anti-phase coordination is not always possible, just as similar constraints ensure that certain patterns of variation—such as final consonant deletion and vowel reduction—will be more prevalent (indeed, nearly unavoidable) above a given speech rate threshold for a speaker (Lindblom, 1963; Gay, 1977; Stetson, 1951; Tuller & Kelso, 1991; de Jong, 2001). Thus, the question of why Medʉmba speakers prefer anti-phase alignment is similar in spirit to age-old questions about why some languages display a greater range of rare and articulatorily complex consonant types, or why some languages display OVS word order when SVO is hypothesized to be easier to process (Gibson, 1998). Linguists have identified countless patterns which seem counterproductive from the standpoint of communicative efficiency, but there are many reasons to think that this diversity, on the whole, is positive and perhaps even necessary for human development (Fishman, 1982; Fitch, 2011).

To conclude this subsection, I would like to refer to one additional interesting pattern which emerges from the examination of individual variability. Specifically, it is interesting that two of the four subjects who showed clear in-phase alignment with the metronome beat at higher speech rates—namely, subjects s3 and s6—had lived outside of the greater Bangangté area for an extended period of time and spoke languages or dialects other than the Bangangté-area dialects of Medʉmba. In particular, subject s6 showed a rapid shift to in-phase alignment in comparison with other sub-

jects. While further data and a more in-depth understanding of speaker backgrounds would be necessary to say conclusively, this pattern is suggestive of a situation in which language experience—perhaps specifically related to the strength of native-like language abilities—can influence phase relations of the kind described above. I leave for future work a deeper investigation of this phenomenon.

CHAPTER 8

COUPLING IN RELATION TO PHONOLOGY AND SOUND CHANGE: THE CASE OF TONE SHIFT AND REVERSAL

Now that the dynamics of inter-subject variation in speech cycling in Medʉmba have been examined in some depth, I turn to some of the broader implications of the present analysis on phonological theory and, in particular, phonological patterns of tone distribution in African languages. Specifically, I will discuss patterns of anticipatory tone shift and tone spread, two related phenomena which are common in African languages and which have proven difficult to explain in terms of phonetic motivations. I suggest that these two patterns may stem from the dynamics of anti-phase alignment between syllable-level beats and metrical structure in African languages.

8.1 Tone Shift and Tone Spread: A Role for Phase Alignment?

Two of the best-described phonological processes affecting tone in African languages are TONE SHIFT and TONE SPREAD, the former serving to move a tone from its original underlying position to a different surface position, and the latter serving to linearly extend the range of syllables over which a tone is associated. An example of tone shift is demonstrated in data from Giryama from Philippson (1998) as cited in Hyman (2007) in (25), showing that a high tone underlyingly associated to the verb stem *-ón-* 'see' moves to the penultimate syllable of the phrase, the noun stem-initial syllable in *revu* 'beard'. The original position of the high tone in (25b) is indicated with an underline.

(25) Righward tone shift in Giryama (Philippson 1998: 321; Hyman 2014: 18)

a. ku-tsol-a ki-revu 'to choose a beard' /-tsol-/ 'choose'

b. ku-o̲n-a ki-révu 'to see a beard' /-ón/ 'see'

An example of tone spread from Yoruba as cited in Akinlabi & Liberman (2000) is given in

(26). Here, we see that high and low tones from the initial syllable are spread to the second syllable, leading to contour tones forming on the second syllable.

(26) Rightward tone spread in Yoruba (Akinlabi & Liberman 2000: 13)

a. /àlá/ → [àlǎ] 'dream'

b. /rárà/ → [rárâ] 'elegy'

(27) Haya tone shift (Hyman & Byarushengo 1984: 56)

a. /ó-mu-tí/ → o-mú-ti 'tree'

b. → o-mu-tí gwange 'my tree'

c. → ó-mu-tí gwange 'to see my tree'

Shifting and spreading are thought to be related phenomena, the latter typically represented in autosegmental terms as a multiply linked tone, and the former as a tone that was first multiply linked and then delinked from its original host (Figure 8.1).[1]

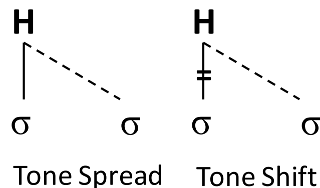


Figure 8.1: Autosegmental representations of anticipatory tone spread (left) and tone shift (right)

Hyman & Schuh (1974) and later Hyman (2007) note that, while spreading and shifting both occur quite commonly in a left-to-right pattern (e.g. tones move or shift to the right of where they are underlyingly associated, as in Figure 8.1), right-to-left spreading and shifting are quite a bit more constrained. Furthermore, while rightward tone spreading can be either *bounded*, or restricted

1. Neither process of anticipatory shifting or spreading (which are both assimilatory processes in the present case) seems directly comparable to other processes of anticipatory *dissimilation* found in many tone languages from different families, including Mandarin (Xu, 1994a, 1997), Vietnamese (Gandour *et al.*, 1992) and Triqui (DiCanio, 2014), though the possibility for such a dissimilatory process is certainly not ruled out for Medʉmba or African languages in general.

to affecting only neighboring syllables within a specified prosodic domain (form example, binary spreading within a foot), or *unbounded*, affecting a potentially unlimited number of syllables, the vast majority of leftward spreading is bounded and binary (Odden, 2015). Shifting shows a similar asymmetry: while a number of languages have been argued to display nonlocal rightward shifting of a tone to, for example, the word- or phrase-penultimate syllable, leftward tone shift, by-and-large, appears to be primarily local, targeting only the syllable immediately preceding the syllable to which a tone is underlyingly associated. Indeed, the study of tone spread in Bantu languages has inspired a great deal of research on the nature of locality (or 'noniterativity') and domains of application of phonological processes (Clements & Ford, 1979; Clements, 1985; Mtenje, 1987; Myers, 1987, 1997; Philippson, 1998; McCarthy, 2004; Key, 2007; Kisseberth, 2007; Kaplan, 2008).[2]

Hyman and Schuh 1974 and Hyman (2014) hypothesize that directional asymmetries in shifting and spreading arise due to different mechanisms underlying rightward versus leftward shifting/spreading. For example, Hyman points to multiple analyses which cite the late realization of tonal targets as the basis for the rightward bias; in other words, rightward 'carryover' coarticulation is stronger and more common than leftward 'anticipatory' coarticulation (Akinlabi & Liberman, 2000; Kingston, 2003). Indeed, phonetic evidence from studies by Xu (1997), Xu & Liu (2006), Li *et al.* (2003), Gandour *et al.* (1994), and Brunelle (2003) confirms a bias for carryover coarticulation.

So what, then, drives leftward spreading and shifting? Hyman (2007) points to three common patterns:

1. Attraction to a <u>strong position</u>

2. Pressure from the <u>right edge</u>

3. Attraction to a <u>weak position</u>

2. As noted by Kaplan 2008a, such processes have been especially challenging for implementation in Optimality Theory due to the need for input-output mapping constraints.

I discuss the first two in turn in the context of data from various African languages which Hyman and others have cited as evidence for these general trends. The third will be dealt with separately at the end of this section.

The first general pattern of high tone anticipation, ATTRACTION TO A STRONG POSITION, refers to a preponderance of left-oriented processes which target prosodically prominent positions. For example, in (28), we see a case of anticipatory high tone shift in standard Rundi cited by Philippson (1991) and later by Hyman (2007) in which a high tone originating on stem-final syllable moves leftward, reassociating with the stem-initial syllable *-go-*.

(28) Standard Rundi anticipatory high tone shift (Philippson 1991: 186, Hyman 2007: 15)
/ u-mu-gozí / $\rightarrow$ [umugózi] 'rope'
AUG-CL3-rope

A similar case is found in Haya, this time involving regressive high tone spreading. In (29), we see that a high tone on the final syllable of the stem *gòló* 'snuff' spreads leftward to the stem-initial syllable.

(29) Haya anticipatory high tone spread (Hyman 2007: 14)
/ bù-gòló / $\rightarrow$ [bùgóló] 'snuff'
CL14-snuff

Such a process of regressive high tone spreading has been argued by Franich (2014) to occur in Medʉmba, as well, though the situation is a bit more complex. Franich argues that falling (high-low) contour tones in Medʉmba are targeted for high tone anticipation when they occur in certain prosodically-prominent positions. In (30), we see that the distant past marker *nâʔ* is realized as high-low when it precedes a low tone negation marker *kə̀*, but as a high-to-downstepped-high contour (indicated with an acute macron and marked in red) when it occurs before the high tone verb *jə́n*, suggesting that the (downstepped) high originating on *jə́n* is realized within the contour

of the distant past marker.[3]

(30) Medʉmba Regressive high spreading onto a falling tone tense marker (acute macron in red represents a high-to-downstepped-high falling contour tone)

a. mvə̀n nâʔ kə̀ ꜜjə́n í
chief DPST NEG see 3SG
'The chief didn't see him (a long time ago)'.

b. mvə̀n ná̄ʔ jə́n í
chief DPST see 3SG
'The chief saw him (a long time ago)'.

Regressive high tone spreading is also observed to occur within demonstrative phrases in Medʉmba, as shown in (31). Here, we see that a high tone on the demonstrative is anticipated on the preceding noun.

(31) Medʉmba Regressive H spreading from demonstrative to head noun

a. mvə̀n 'chief'

b. mvə̌n ná 'that chief'

c. kə̀lò 'banana'

d. kə̀lǒ lá 'that banana'

Autosegmental representations of (30b) and (31b) are shown in Figure 8.2.

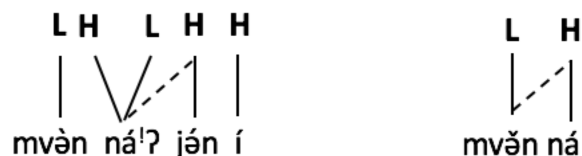


Figure 8.2: Autosegmental representations of anticipatory tone spread in Medʉmba

3. In terms of representation, one could traditionally treat this alternation either as a case of high tone spreading or low tone raising, a rule which raises a low tone in the presence of a high tone, as has been proposed for similar cases in Igbo by Goldsmith (1976) and for other West African languages by Hyman (1977). In either case, it is clear that the alternation arises from anticipatory assimilation. Franich treats the Medʉmba data as a case of tonal merger.

In Kinyarwanda and Kirundi, a similar type of tone shift to that described in (28), but instead involving low tones, has been posited to occur within the verbal system of some dialects, leading, in some cases, to a complete 'tone reversal'. As can be seen in (32), the first stage in the tone reversal involved anticipation of the low tone on the final vowel to the stem-initial syllables *bon* and *doːt*, such that a HL contour resulted on these syllables. Later, anticipatory spread of the high tone on the stem-initial syllable is posited to occur, leaving a high tone on the infinitive prefix *ku-*. Finally, the contour on the stem-initial syllable is reduced to a simple low tone, eventually resulting in full tonal reversal for verbs such as *kùbónà 'to see' and *kùdóːtà 'to dream'.

(32) Reconstructions of tonal changes in Kirundi and Kinyarwanda verbs (Hyman 1978: 264)

a. **kù-bón-à* > *kù-bôn-à* > *kú-bôn-à* > *kú-βòn-à* 'to see'

b. * *kù-dóːt-à* > *kù-tôːt-à* > *kú-dôːt-à* > *kú-rôːt-à* 'to dream'

Hyman (1978) illustrates several patterns resulting from interspeaker variation which demonstrate synchronic realization of the various steps in the reconstruction. In (33), we see patterns of interspeaker variation for the verb 'to see'.

(33) Variation in tone in Kinyarwanda and Kirundi verb 'to see' Hyman (1978)[4]

a. kù-βôn-à

b. kǔ-βôn-à

c. kú-βôn-à

d. kú-βòn-à

Such reversals are not limited to Kinyarwanda, and have in fact been cited for a number of Bantu languages from different subdivisions, including Ruwund (Nash, 1992,-), Tembo (Kaji, 1994), Luba (van Spaandonck, 1971), Tonga (Meeussen, 1963; Carter, 1971), Shi (Polak-Bynon,

4. Myers (2003) also provides instrumental evidence that Kinyarwanda tone anticipation is phonetically gradient (though his focus is on nouns). See also Kimenyi (1979), Furere & Rialland (1985), and Overdulve (1975) for descriptions of the phenomenon in Kinyarwanda.

1975), Sukuma (Batibo, 1976), Marachi (Marlo, 2007, 2008), Wanga (Mutonyi, 2000; Marlo, 2008), and Bukusu (Mutonyi, 2000; Marlo, 2008).

Given the data on regressive shift and spread outlined above as well as the known existence of tone reversals in many languages, it is clear that local anticipatory shifting and spreading—and, in particular, shifting and spreading *to a prosodically prominent syllable*—is quite common across African languages. As has been discussed, the anticipatory nature of these processes is not easily explained through coarticulatory processes, as tonal coarticulation shows a strong *rightward* bias across languages (Xu, 1997; Xu & Liu, 2006; Li *et al.*, 2003; Gandour *et al.*, 1994; Brunelle, 2003). Furthermore, work examining the influence of prosodic prominence on coarticulation finds that prominent syllables should be, if anything, more resistant to coarticulatory processes, not more susceptible (Cho 2004).

Hyman 1978 suggests that the relationship between regressive spreading and prominence may involve a reanalysis of tones (high tones, in particular) as accentual, going so far as to say:

> "The more accentlike a H(igh) tone is, the more likely tonal anticipation will occur."
>
> (Hyman 1978: 264)

He expands on this claim (and indeed echoes it in Hyman 2014) by arguing that high tones are likely to be treated as more 'accentlike' if, for example, the language tends towards CULMINATIVITY, or a limit to a single high tone per word.[5] As we have seen above, however, there are languages (including Medʉmba) which display anticipatory tone spread but which do not display clear evidence of culminativity. Furthermore, though there may be historical processes such as the loss of tonal contrasts which can accompany the development of systems with culminative tonal patterns, it is not immediately clear what, exactly, leads a tone to be reanalyzed as 'accentual'.

Another proposal that has been put forth by Hyman & Schuh (1974), as well as by Hyman (2007) and Köhnlein (2013), is that leftward spread and shift are motivated by PRESSURE FROM

5. Subsequent work by Hyman (2006) has also identified OBLIGATORINESS, or the requirement that a word carry at least one tone, and SYLLABLE-DEPENDENCY, or the requirement that the tone-bearing unit is the syllable, as factors which may contribute to the 'accentlikeness' of tone in a language

THE RIGHT EDGE. In other words, where information-rich tonal and intonational contours occur at the right edge of, say, a prosodic phrase, earlier tones are more likely to move leftward, away from these contours. It is argued that tones move away from the right edge in order to avoid coinciding with intonational contours, perhaps to preserve perceptual contrasts (Smiljanić, 2002; Crane, 2010). Such an account is appealing given work within the Autosegmental-Metrical (A-M) tradition (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988; Ladd, 1996) which shows that f_0 turning points on stressed syllables in various languages occur earlier when they precede another stressed syllable (Pierrehumbert, 1990; Prieto *et al.*, 1995; Arvaniti *et al.*, 2006), a phenomenon referred to as TONAL CROWDING. This effect serves to counteract the typical phonetic pressure of peak delay (Xu, 1999). Tonal crowding effects have been explained within the A-M framework in terms of the existence of intonational melodies which are sparsely specified and which are associated to elements within a segmental string; when the relative 'tonal density' (ratio of tones per segments/syllables) is high, tones may shift their alignment (Arvaniti *et al.*, 2006). However, recent evidence suggests that tonal crowding is likely not sufficient for explaining all tonal retraction processes. For example, Katsika *et al.* (2014) find that stressed words with and without pitch falls ('pitch accents' in A-M terms) affect boundary tone placement in Greek. Furthermore, greater tonal density was not found to correlate with differences in boundary tone placement in that study. The authors suggest that coordinative dynamics underlying articulation of boundary tones, vocalic gestures, and stress drive differences in peak delay. In Bantu, Myers (2003) finds variable effects of a following high tone on relative peak delay both within and across individuals, but, contrary to the tonal crowding approach, finds that higher tonal density leads to *less* reduction in peak delay, not more. Myers (2000) also shows that high tone retraction, a process with similar distribution to anticipatory high tone shift, does not show the kind of phonetic gradience relative to the distance between high tone syllables expected based on tonal crowding accounts. Myers instead proposes a phonological account of retraction in Optimality Theory, drawing on a constraint-based formulation of the Obligatory Contour Principle, to explain the patterns in Shona. This account of course

leaves open how retraction may have first arisen in the language.

To fully account for tonal retraction, a separate or additional mechanisms is needed. I would like to propose that, in the case of Medʉmba, and likely other African languages, relative phase relations between metrical and prosodic structure may provide a missing link in explaining how anticipatory tone spread and shift may arise, and, for some languages, how a tone reversal might take place. First, I return to the finding detailed at the beginning of this chapter showing that Medʉmba speakers preferred anti-phase alignment with the metronome beat in the speech cycling task at slower speaking rates, but some fell in synchrony with the metronome beat at higher speaking rates. When this transition to in-phase alignment with the metronome occurred, speakers also displayed a greater tendency to align prominent syllables at SHP positions. This finding was accounted for in terms of the dynamics of attractor layouts: the anti-phase alignment pattern with the metronome beat was available to speakers at slower speech rates due to the BISTABILITY of attractors: both anti-phase and in-phase attractors were available at these rates. Once speaking rate reached a critical speed, however, the attractor layout for some subjects shifted such that only the in-phase attractor was available. As speaking rate continues to increase, the width of the in-phase attractor basin continues to increase, leading to even greater stability of the in-phase attractor. Integration of the perceptual goals of metronome coordination and phrase repetition into a single goal was facilitated by the alignment of metronome beats with syllable repetitions, and by 2:1 (or 3:1, in the case of prominent syllables occurring in Syllable 3 position) phase-locking of prominent syllables with the phrase level and metronome beat. Thus, in-phase alignment with the metronome lead to more constrained temporal patterning of prosodically prominent syllables within the repetition cycle.

How does tone anticipation factor into such a model? After all, experiments 1-3 showed quite clearly that neither high nor low tones are directly associated with greater prominence in Medʉmba; neither appears to act, according to Hyman's terms, in an 'accentual way'. I would like to submit, however, that despite the fact that tones are not endowed with relatively greater or lesser promi-

nence, the semi-rhythmic timing of alternations of high and low tones in Medʉmba and other African languages may lead to a situation where they, too, are more effectively organized at higher speech rates such that certain tones occur phase-locked with prominent syllables. Support for such a proposal comes from work by Repp (2003, 2004, 2005) on so-called DISTRACTOR EFFECTS in sensorimotor synchronization tasks. For example, Repp (2003) shows that, where subjects are asked to coordinate finger taps in-phase with an isochronous sequence of target tones with an additional sequence of distractor tones interleaved with the main sequence, taps are attracted to the distractor tones, especially when these tones are timed such that they closely precede target tones. Furthermore, it was found that the more rhythmic the sequence of distractor tones, the more likely they were to 'capture' individuals' taps. Repp (2005) suggests this effect reflects a tendency for rhythmic movement to be attracted to auditory rhythms. As we have seen in the results of experiments 1-4, speech likely also relies on underlying rhythmic structures, and coordination of speech is subject to many of the same dynamical principles as tapping. If alternating high and low tones act as a perceptual 'distractor' of sorts where rhythmic coordination has already been greatly simplified at higher speech rates, it may be the case that subjects tend to coordinate prominent syllables with tones, and particularly with those tones which are marked, or most 'active', in their language. In many (but not all) Bantu languages, high tone is typically identified as the more active tone due to, for example, its greater ability to influence other tones in tone sandhi processes such as downstep (Stevick, 1969; Maddieson, 1978). In a great many cases, then, it will be the high tone which coordinates with a prominent syllable. We can also predict, based on Repp's findings, that the tones most likely to be timed with prominent syllables are those which occur closest to them. Thus, the finding that high tone anticipation tends to occur locally, only targeting an immediately preceding syllable, also seems to fit with this account.

The notion that relative tone height and timing should interact is far from new. A growing body of research has shown the important relationship between pitch and timing in the structuring of rhythmic representations in speech. For example, it has been shown that relative pitch

height and slope can bias duration perception of tones (Yu, 2010), and that duration and pitch height show an interdependence in cuing rhythmic groupings in speech (Cumming, 2011). This effect has been shown to be active in cuing listeners' perceptions of prosodic boundaries in English (Brugos & Barnes, 2014) and has also in fact been shown to influence timing of syllables in Medʉmba (Franich, under review). Thus, though tone may not be directly linked to prominence in Medʉmba, or indeed many other African languages, the link between pitch and timing is robust across languages, making relative tone height a potentially useful cue for recruitment in speech rhythm planning and production.

The notion that tonal alternations can act as rhythmic is also not novel. A number of studies have shown that tones in languages lacking stress show alternations which are predictable at the phrase-level (Kim & Cho, 2009; Warner *et al.*, 2010; Jun, 2012, 2014), and which aid in perception of prosodic boundaries and lexical segmentation (Kim & Cho, 2009; Warner *et al.*, 2010). Jun (2012, 2014) argues that, while 'nonrestricted' tone languages—e.g. those which make extensive use of lexical tonal contrasts—typically have relatively weaker phrase-level 'tonal rhythm', that this rhythm is nonetheless present, and in fact stronger in a language with only a binary high-low contrast, such as many African languages have, than in a contour tone language such as Mandarin. Lipski (2015) proposes that similarities in the timing of rhythmic alternations between high tones in Bubi and Fang—two Bantu languages spoken in Equatorial Guinea—and stressed syllables in Spanish (one of the ex-colonial languages spoken in that country) has facilitated the emergence of tonal effects in the Spanish of Guineans. Within research on motor coordination, Jones & Pfordresher (1997) show that synchronization to a stimulus is easier if melodic alternations reflect rhythmic alternations.

To summarize, I have proposed a model by which tone anticipation might be explained through metrical and prosodic coordination dynamics and relative phase transitions. First, it was observed that some subjects exhibited a transition to in-phase coordination between metronome beats and phrase repetitions at higher speech rates in the speech cycling task, and that this led to greater

phase-locking of prominent syllables at SHP positions. I then proposed that this situation likely arises through the decrease in strength of an anti-phase coordination mode at higher speech rates, and a concomitant increase in strength of the in-phase attractor. As the in-phase coordination strategy is adopted, I propose that subjects integrate perceptual goals of phrase repetition and coordination with the metronome, leading to a single perceptual goal and overall coordination strategy. While the in-phase alignment pattern is argued not to be the default strategy for Medʉmba speakers overall (or perhaps for speakers of other African languages), I suggest that, under high speech rate conditions, this pattern may spontaneously arise. More specifically, just as a given individual's speech rate can vary from moment to moment, so, too, could their underlying metrical-prosodic phase alignment. I have furthermore argued that regular alternating patterns in high and low tones, which form an additional rhythmic stream to the phrase repetition/metronome stream (possibly also governed by oscillatory dynamics), may also be incorporated into the same perceptual goal, leading to a situation where the active tone (in the case of Medʉmba, the high tone) is preferentially aligned in-phase with prosodically prominent syllables and metrically strong beats.

The above scenario provides a phonetically natural account of how anticipatory spreading or shifting of a tone can be triggered by coordination dynamics under certain speaking conditions, namely at high speech rate. Though the proposed mechanism for these processes differs fundamentally from coarticulation-based accounts of rightward tone spreading (Xu, 1997; Akinlabi & Liberman, 2000; Kingston, 2003; Xu & Liu, 2006; Li *et al.*, 2003; Gandour *et al.*, 1994; Brunelle, 2003), the essence of the present account is similar in spirit in that it predicts that, just as with coarticulation and other phonetically-based processes, anticipatory tone spreading and shifting arise from patterns of variation. Also note that, though we have been focused on explaining anticipatory tonal processes due to the relatively greater challenge of explaining their distribution in phonetic terms, the present analysis would also be equally-well suited to handling cases of binary rightward spreading. I leave it to future work to investigate whether evidence for both types of spreading from alignment changes versus f_0 lag can be found. Importantly, patterns of phase alignment can

also lead to distinct paths of variation and change across languages: in some cases, anticipatory spreading may lead subjects to reanalyze a syllable with a different tone, and in others (as in the case of Kinyarwanda), tonal anticipation can be source of dialectal or individual variation. In still other cases, anticipation can beget further changes, as in the case of Ruwund or Tembo, where a complete tone reversal results after contour leveling.

Crucially, this account does not require that tones themselves carry any sort of 'accent' or other diacritic which leads to their attraction to prominent positions. Rather, the naturally alternating patterns of tones lead them to be integrated into a single coordinative mode with accented and prosodically prominent syllables under conditions where articulatory and perceptual dynamics encourage or require it. This association opens the door for one other possible path of change, which is the reanalysis of high tones (or indeed low tones) as direct markers of prosodic and/or metrical prominence. Reanalysis of this sort has been argued to take place in languages such as Llogoori (Goldsmith, 1991). Thus, a crucial link still holds between (non-stress) tone and relative prominence, albeit not as direct a link as has been suggested in prior work. I now move on to discuss two remaining issues—tone anticipation leading to contour tone formation, and tone shift to a 'weak' position—and explain how these phenomena can be explained by the current model.

8.1.1 Remaining Issues: Tone Spread vs. Tone Shift and Attraction to a 'Weak' Position

A question which remains is, why should phase shifting as described above lead in so many cases to a contour tone forming on the prominent target syllable, rather than a complete shift? Or, put differently, why should anticipatory spreading occur at all, when anticipatory shifting would more directly reflect in-phase alignment of tones and prominent syllables? First off, it must be noted that there is clearly a dearth of instrumental phonetic data on anticipatory spread and shift, and a deeper understanding of their phonetic patterning (as e.g. Myers 2003 has undertaken) is necessary before we can fully understand the phenomenon. However, assuming that transcriptions

of these phenomena have been relatively accurate, what could explain the prevalence of contour formation? One possibility is that contour tones are in fact even more beneficial than level high tones for speech coordination across metrical, prosodic, and tonal levels. For example, much work has shown that f_0 turning points serve as an important locus of coordination between tonal and segmental gestures (van Santen & Hirschberg, 1994; Ladd *et al.*, 1999; Xu & Sun, 2002; D'Imperio *et al.*, 2003). In research on vision perception, it has also been found that moving stimuli—e.g. bouncing balls—facilitate sensorimotor synchronization better than static stimuli.

Finally, there is one additional pattern described by Hyman (2007) which we have not yet discussed and which commonly characterizes tone anticipation: ATTRACTION TO A WEAK POSITION. In such cases, a tone appears to be attracted to a position which shows no evidence of prosodic prominence, and even shows evidence for prosodic weakness. For example, Hyman cites an example discussed by Schadeberg (1977) from Lingala, in which a high tone spreads from the final vowel to the post-radical vowel, a position which Hyman points out is also targeted for vowel harmony in Punu, a related language. Hyman argues that stem-internal syllables typically act as prosodically weak (see, for example, Hyman 1998 for a discussion of 'prosodic trough' effects of these vowels in Yaka[6]), and suggests that spreading, in the case of Lingala, targets precisely those syllables. In (34), we see that the tone of the final vowel spreads leftward up to the second syllable of the stem.

(34) Lingala high tone shift to postradical vowels (Adapted from Hyman 2007: 26; Shadeberg 1977: 198)

6. As was discussed in Chapter 5, stem-initial syllables have been shown to behave as prominent in Bantu languages. Hyman (2010) also describes several cases where the penultimate syllable in Bantu seems to behave as prominent.

		pre-stem + stem	pre-stem + stem	final vowel
		‘to get confused’	‘to stagger’	
a.	infinitive:	kò + kàk-àt-àn-à	kò + tél-èng-àn-à	/-à/
	subjunctive:	ná + kàk- àt-àn-à	ná + tél-èng-àn-à	/-à/
b.	past:	nà + kàk-át-án-í̲	nà + tél-éng-án-í̲	/-í/
	remote past:	nà + kàk-át-án-á̲	nà + tél-éng-án-á̲	/-á/

In many ways, apart from the directionality of spreading, the tendency for tones to spread to prosodically weak syllables is not surprising in light of observations that coarticulation very often targets weak syllables (Cho, 2004). More puzzling are those cases where a tone is seen to shift away from a prosodically prominent position to a prosodically weaker one. Example (35) demonstrates such an effect with data from Totela from Crane (2010) which shows that high tones shift from the stem-initial syllable to the final syllable of an infinitival prefix (a) or to an object marker prefix (b).

(35) Totela High Tone Anticipation (Adapted from Crane 2010: 6)

a. Infinitive forms

Toneless root		H root	
oku-lwa	‘to bury’	okú-bi̲ika	‘to hide’
oku-ukuta	‘to shake’	okú-ye̲mbela	‘to herd’

b. Infinitives with toneless object markers

Toneless root		H root	
oku-mu-ziika	‘to bury him’	oku-mú-bi̲ika	‘to hide him’
oku-mu-ukusa	‘to shake him’	oku-mú-hu̲pula	‘to think of him’

A similar effect is found in Tonga, which Goldsmith analyzes in terms of underlying accents structurally associated with certain vowels (including verb stem-initial syllables) which then attract HL tone melodies, the low portion of which is also said to be accented. Principles of tone association and a decontouring rule then produce outputs in which accented low tones are associated to

underlyingly accented syllables, and then high tones associate to neighboring syllables which have no underlying accent. While treatment of low tones as the phonologically 'active' tones seems to work well for Tonga (and see Hyman 2007 for a proposal that Tonga has been reanalyzed from a privative high tone system to a privative low tone system), Crane points out that such an analysis is unlikely to work for Totela, where high tones are clearly the active tones (for example, triggering downstep and other tone rules). How, then, does a high tone come to shift one syllable to the left from what appears to be a prosodically prominent to a prosodically non-prominent position? One possibility is that prosodically prominent syllables in Totela are coordinated anti-phase with underlying metrical structure, such that metrically prominent positions tend to align with prosodically non-prominent positions. It could be, then, that high tones in Totela are being attracted to syllables which are aligned with *metrically* prominent positions, but not prosodically prominent ones. This is a similar account to what is provided for Medʉmba and other languages in §8.1, with the only difference being that alignment of metrical and prosodic prominence is essentially skipped, and tonal rhythms align only with underlying metrical structure.

Before concluding, I discuss one last implication of anti-phase rhythmic alignment in speech, which is its relation to music. As will be discussed, a growing body of research shows that language and music are likely processed using similar areas of the brain, and thus experience with one may affect functioning in the other, and vice versa. In §8.2, I lay out some ways in which the domains of language and music may show a shared influence of anti-phase rhythmic alignment, and discuss some ways we might further test such a prediction.

8.2 Language and Music: A Shared Timekeeper?

As mentioned at the start of this chapter, one reason we might not be entirely surprised by the anti-phase metronome alignment patterns preferred by the Medʉmba speakers is that these speakers are accustomed to highly syncopated forms of music, which are the norm in West Africa and indeed much of the continent. Scholars of musicology consistently reference 'off-beatness' as a

key feature of West African music Locke (1982); Toussaint (2003); Agawu (2006). The specific rhythmic structures utilized in African traditional music and drumming are quite complex, and still a subject of great debate in the field of musicology (Agawu 2006).[7] Locke (1982) describes the many layered 'polyrhythms' which can be found in dance drumming music of the E̱ve (otherwise known as Ewe or Èʋe) people of Ghana which are played with a combination of bells, rattles, wooden clappers, hand claps, and a variety of other drum types. These rhythms are said to be organized around one particular repeating pattern typically played on a bell, often referred to as the 'key pattern', or simply as 'the bell pattern' (Nketia, 1963; Montfort, 1987; Toussaint, 2005; Penalosa, 2009; Kubik, 2010).

Agawu (2006) provides an example of the basic bell pattern, which is recreated in Figure 8.3:

Figure 8.3: Standard bell rhythm (Agawu 2006: 1)

As can be seen in Figure 8.3, the standard bell pattern involves syncopation starting with the third beat, where an eighth note followed by a quarter note leads to a displacement of the attack of the third quarter note to the 'and' or 'offbeat' of the third beat. The bell pattern is typically transcribed either with a 12/8 or 6/8 'triple' meter or with a 4/4 or 2/4 'duple' meter, and westerners often feel the music as toggling between the two (Danielson, 2007), but this perception has been argued to vary based on previous musical experience (Vuust & Witek, 2014). The bell pattern is found to be common in musical traditions throughout Sub-Saharan Africa, but especially among speakers of Bantu and Kwa languages, and is thought to have spread with along with iron-working technology as part of the Bantu migration (Walton, 1955). The bell pattern is also common in music of many African diaspora communities, most notably in Afro-Cuban music, where a variant

7. As Agawu (2006) discusses, it is not only the geometric or 'structural' patterns of rhythm which are debated within African musicology, but also, increasingly, the role and experience of the listener and the culturally-influenced interpretation they bring to to perception and analysis of rhythmic patterns. This parallels ongoing discussions within the fields of linguistics and linguistic anthropology regarding the role of communicative function and culture in shaping linguistic structures, though relatively little debate has taken place, to my knowledge, as to the role of function in shaping rhythmic structures and interpretation in language.

of it is referred to as the 'clave' rhythm (Grenet, 1939; Jones, 1959).

In discussing the ways in which dancers and other drummers relate to the bell pattern in practice, Locke notes that clapping, for example, is used to "add distinct layers to the overall polyrhythmic texture" (Lock 1982: 218). Thus, rather than coordinating directly with the bell pattern (or any other pattern), individuals clap in response to those patterns[8].

A comprehensive description of musical rhythms in African music is clearly well beyond the scope of the present dissertation, but from this brief description, we can at least see how it is that Medʉmba subjects' approach to the speech cycling task may relate to their experience of rhythms more generally, and specifically, their experience coordinating rhythmic action (such as a clap or a dance step) with a beat. The question then becomes, is music directly influencing subjects' speech patterns? Or is language somehow shaping musical patterns? Rather than seeing one of these domains as 'shaping' the other in any direct manner, another way to conceptualize the relationship between musical and speech rhythms is to think of them as being mutually influenced by a shared TIMEKEEPER. The notion of a timekeeper in music is discussed both in abstract terms, as in reference to the underlying 'pulse' of a piece of music, and in more concrete terms, as we have seen from discussions of the bell pattern by various scholars. The notion of the timekeeper has also been implicated in studies of speech by way of research on motor control (see Turk & Shattuck-Hufnagel (2013) for an overview). Models of the timekeeper for speech and motor control can broadly be broken down into INTERVAL-BASED MODELS and ENTRAINMENT-BASED MODELS. Interval-based models typically involve some type of internal clock which provides a periodic pulse, and then a reference memory which stores information about duration so that past intervals of time can be compared with upcoming intervals. One very prominent interval-based model is the two-level model of Wing & Kristofferson (1973), which is best known for modeling tapping

8. It is interesting to note that rhythmic 'call-and-response' patterns are very common in many West African traditions more generally, and this practice is also often integrated into music and dance (Agawu, 1995, 2016), adding further intricacy to the overall rhythmic picture. Indeed, the notion of playing 'with' vs. 'in response to' a musical beat is one that may be important for characterizing African musical rhythms (Burns, 2010). It is also likely that some forms of music treat silences and emptiness themselves as musically significant (Bohlman, 1999)

behavior and error correction processes in sensorimotor synchronization tasks. The model involves an internal clock which controls inter-tap intervals by emitting a regular signal which triggers a separate process of motor response to rhythmic events. Another well-known interval-based model comes from Scalar Expectancy Theory (SET) (Gibbon, 1977; Church *et al.*, 1994) which was primarily designed to explain rhythmic patterns in animal behavior. The model involves a neural pacemaker which emits pulses which are accumulated and counted over the course of a stimulus period. More recent models (Church & Broadbent, 1990; Matell & Meck, 1999) posit that the pacemaker can be modeled as a set of neural oscillators with fixed periods which are synchronized with a stimulus onset, where maximal neural activity takes place.

An alternative to interval-based models of timekeeping mechanisms is the entrainment-based model of timing, best exemplified in Dynamic Attending Theory (DAT), proposed by Large & Jones (1999). DAT posits that the timekeeper is a system of continuous oscillators which can entrain to rhythmic stimuli in the environment and which are characterized by an *attentional pulse*, or a point of heightened attentional energy, which are coordinated with perceptually salient sensory events, allowing for expectations about upcoming rhythmic stimuli to be built. There is now substantial neurological research supporting the existence of multifrequency oscillating rhythms in the brain which facilitate production and perception of rhythms in speech and music, as well as other coordinative domains (see Grahn (2012) for an overview). Crucially, the DAT model allows for coupling between, for example, rhythms in motor behavior and attentional rhythms, which predicts that attentional rhythms should influence motor rhythms, and vice versa. Larger movement amplitudes have been found to elicit stronger entrainment of neural oscillators, lending support to this theory (Varlet *et al.*, 2012, 2014). Additional work has shown that entrainment and rhythm perception is not entirely stimulus-driven, and that individuals show neural evidence of beat perception where a beat would be expected (as in a continuous string of isochronous pulses where one pulse has been removed), providing further support for the existence of the continuous nature of neural oscillatory behavior (Snyder & Large 2005). Individuals have also been found to be able

to selectively phase-lock to one stimulus over the other, suggesting that entrainment also relies on directed attention (Kerlin et al. 2010). Furthermore, musical experience has been shown to influence entrainment behavior to musical rhythms, suggesting that experience and learning shape endogenous neural oscillations (Nozaradan *et al.*, 2016; Stupacher *et al.*, 2017).

Indeed, a growing body of research shows that past experience in the domain of music can influence neural behavior in language, and vice versa. Wong *et al.* (2007) find that auditory brainstem encoding of linguistic pitch in Mandarin, a tone language, is encoded more robustly in musicians than nonmusicians. Patel (2003) provides an overview of evidence that language and music may also rely on similar processing mechanisms for syntactic structures (see also Lerdahl & Jackendof 1983). Neural links between music, movement, and language have also been established: not only does listening to music engage areas of the brain utilized in motor control (Alluri *et al.*, 2011), but tapping and drumming certain types of complex rhythms has been found to activate language areas in the brain (Vuuset *et al.*, 2006; Vuust *et al.*, 2011).

Given the burgeoning research pointing to links in processing of motor function, language and music across many types of structures, it is not much of a stretch to imagine that movement, music, and language experience might collectively influence rhythm processing. If this is the case, then it is only natural that Medʉmba speakers, growing up in a culture in which dance plays a central role in community life, and where children learn to effortlessly coordinate dance movements with complex polyrhythms from a young age, should show influence of such rhythms in language, as well.

One specific way in which rhythm processing across all domains may be affected, as I have discussed previously, is in terms of the range of possible coupling relations individuals can exploit in coordinating with both endogenous rhythms (such as an internal timekeeping oscillator) and external stimuli, such as a drum, or even a conversation partner. As has been discussed, the ability to coordinate anti-phase with an external rhythm entails the ability to coordinate in-phase with it. Strength of sensorimotor coupling behavior is also shown to be influenced by previous experience

and practice (Repp, 2010). Thus, individuals who can coordinate anti-phase with a stimulus with relative ease will exhibit a greater range of coordinative possibilities than those who struggle to maintain anti-phase coordination. As was discussed in §8.1, greater flexibility of rhythmic phase relations may also shape the types of language variation and change expected to occur in a given language. One example I have given is that of anticipatory tone spreading. It is interesting to note, in light of our discussion of links between music and language, that anticipation of beats is also a well-known rhythmic strategy in West African music, and is through to arise through the syncopated relationship between an instrument's rhythm and the 'underlying main beats' (Gerstin 2017:9). A potentially fruitful area of future study may therefore be to investigate text-to-tune pairing in Medʉmba and other West African languages to see whether phrases in language which involve tone anticipation (such as demonstrative phrases in Medʉmba) are aligned in music such that anticipation of tone aligns with anticipation of the underlying beat of the music.

Another area of investigation which will be important going forward is to investigate how flexibility of phase alignment strategies may influence language processing. The theory of PREDICTIVE CODING (Friston, 2009; Feldman & Friston, 2010; Adams *et al.*, 2013), a model of motor behavior and neural processing which is gaining wide popularity, posits that the motor system facilitates temporal-predictive functioning, utilizing task-specific rhythms to enable us to build expectations about sensory input. Investigations of predictive coding have not been widely investigated in speech per se, but much of the literature showing facilitative effects of isochronous rhythms on speech perception and processing (e.g. Dilley & McAuley (2008); Dilley *et al.* (2010); Brown *et al.* (2011); Dilley *et al.* (2012) may in fact reflect the effects of predictive coding. An interesting question concerns how rhythmic experience influences the degree to which certain rhythms may or may not have a facilitative effect on processing.

CHAPTER 9

CONCLUSION

In the preceding chapters, I have attempted to provide empirical support for the existence of metrical prominence asymmetries in Medʉmba, a Grassfields Bantu language which, like many African languages, lacks canonical cues for stress. Using the speech cycling paradigm, I have shown that stem-initial syllables are more strongly attracted to prominence-attracting positions than non-initial syllables and prefix syllables. I have also shown some evidence, albeit somewhat weaker, for the attraction of intonational phrase-initial syllables to positions of prominence. Tone, on the whole, was not found to directly impact relative prominence of syllables. Rather, it seems the connection between prominence and tone is indirect: high tone is attracted to positions of relative prominence (e.g. the head of an intonational phrase), but high tone itself cannot be considered prominent, based on the experiments presented here.

Aside from the lack of direct effects of tone on syllable prominence, one surprising finding from the experiments was the consistently strong metrical behavior exhibited by pronominal enclitic syllables. These syllables, despite showing evidence for phonological dependence on the syllable that precedes them, consistently occurred in prominence-attracting positions. Furthermore, while the behavior of enclitic syllables in Experiment 1 suggested they might be best analyzed as bearing simultaneous head and dependent status in overlapping foot structures, evidence from Experiment 4 suggests that they can in fact bear greater prominence than the nouns that precede them in some cases, suggesting a process of accent retraction may also need to be posited in order to fully account for the data at hand. Still, these findings overall suggest that more in-depth investigation of the phonetic and phonological patterning of these pronominal enclitics is necessary to more thoroughly understand their behavior.

As detailed in Chapter 7, another unexpected pattern concerned the specific alignment pattern with the metronome that Medʉmba speakers adopted in doing the speech cycling task. Unlike subjects in previous speech cycling research, Medʉmba speakers preferred to align their repetitions

anti-phase with the metronome beat, instead of on the beat. It was also found that the substantial individual variability subjects showed in the level of clustering of syllables around prominence-attracting positions was related to alignment patterns with the metronome: where subjects fell in-phase with the metronome, clustering around prominence-attracting positions increased. Though still quite speculative, I have offered an account of how phase relations may contribute to variability in speech more generally, and how they may contribute to typologically unusual patterns of regressive tone spreading in African languages more specifically. I have outlined several concrete next steps which can be used to test some of the hypotheses put forth in that chapter.

Finally, on a practical note, this dissertation represents one of the first studies of speech cycling in which the method has been applied in an exploratory way for understanding patterns of prominence in a language (as opposed to, for example, confirming results which have been described based on phonetic and phonological studies). The success I have had in locating prominence in Medʉmba using the method suggests that speech cycling, given its ease of application, would provide a good tool for phonologists and phoneticians to use more generally in probing for or confirming impressions of prominence, particularly in cases where the language under investigation is not native to the researcher. The method thus provides another useful tool for a fieldworker's toolbox in mapping out the accentual patterns of the world's languages.

REFERENCES

Abercrombie, D. 1967. *Elements of General Phonetics*. Chicago: Aldine and Atherton.

Adams, R.A., S. Shipp, & K.J. Friston. 2013. Predictions not commands: Active inference in the motor system. *Brain Structure and Function* 218.611–643.

Agawu, K. 1995. *African Rhythm: A Northern Ewe Perspective*. Cambridge University Press.

—— 2006. Structural analysis or cultural analysis? competing perspectives on the 'standard pattern' of West African rhythm. *Journal of the American Musicological Society* 59.1–46.

—— 2016. *The African Imagination in Music*. Oxford University Press.

Akinlabi, A., & M. Liberman. 2000. Tonal complexes and tonal alignment. *Proceedings of NELS 31* 31.1–21.

——, & E.E. Urua. 2003. Foot structure in the Ibibio verb. *Journal of African Languages and Linguistics* 24.119–160.

Alluri, V., P. Toiviainen, I.P. Jaaskelainen, M.S. Glerean, & E. Brattico. 2011. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage* 59.3677–3689.

Anderson, S. 2005. *Aspects of the Theory of Clitics*. Oxford University Press.

Arvaniti, A., D.R. Ladd, & I. Mennen. 2006. Phonetic effects of focus and 'tonal crowding' in intonation: Evidence from Greek polar questions. *Speech Communication* 48.667–696.

Auer, P. 1991. "stress-timing" vs. "syllable-timing" from a typological point of view. In *Proceedings of the Conference on Linguistics and Phonetics: Prospects and Applications*, ed. by Bohumil Palek, 292–305.

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

—— 2006. *Incursões em torno do ritmo da fala*. Campinas: Pontes.

—— 2007. From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication* 49.725–742.

Batibo, H. 1976. A new approach to Sukuma tone. In *Studies in Bantu Tonology*, ed. by L.M. Hyman, 241–257, Los Angeles. University of Southern California.

Beckman, J. *Positional Faithfulness*. University of Massachusetts, Amherst dissertation.

Beckman, M., & J. Pierrehumbert. 1986. Intonational structure in japanese and english. *Phonology Yearbook* III.15–70.

Beckman, M.E. 1986. *Stress and Non-stress Accent*. Dodrecht: Foris Publications.

Benaglia, T., D. Chauveau, D. Hunter, & D. Young. 2009. mixtools: An *r* package for analyzing finite mixture models. *Journal of Statistical Software* 32.1–29.

Bennett, P. R. *The role of stress and intonation in Early Proto-Bantu*. University of Wisconsin-Madison dissertation.

Bennett, R. *Foot-conditioned phonotactics and prosodic constituency*. University of California at Santa Cruz dissertation.

—— 2013. The uniqueness of metrical structure: Rhythmic phonotactics in Huaripano. *Phonology* 30.355–398.

Bickmore, L.S. 1995. Stress in Lamba. *Phonology* 12.307–341.

Bloch, B. 1950. Studies in colloquial Japanese IV: Phonemics. *Language* 26.86–125.

Bohlman, P.V. 1999. Ontologies of music. In *Rethinking Music*, ed. by N. Cook & M. Everist, 17–34, New York. Oxford University Press.

Bolinger, D.L. 1965. *Forms of English: Accent, Morpheme, Order*. Cambridge, MA: Harvard University Press.

Breedveld, A., & M.P.G.M. Mous. 1086. A dialectometrical study of some Bantu languages (a.40-a.60). In *La Methode Dialectometrique Appliquee Aux Langues Africaines*, ed. by G. Guarisma & W. Mohlig, 177–241, Berlin.

Browman, C.P., & L. Goldstein. 1986. Towards and articulatory phonology. *Phonology Yearbook* 3.

——, & ——. 1989. Articulatory gestures as phonological units. *Phonology* 6.201–251.

——, & ——. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18.299–320.

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

Brugos, A., & J. Barnes. 2014. Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English. *Proceedings of Speech Prosody 7* 388–392.

Brunelle, M. 2003. Tonal coarticulation in Northern Vietnamese. In *15th International Congress of Phonetic Sciences*, 2673–2676.

Burns, J. 2010. Rhythmic archetypes in instrumental music from Africa and the diaspora. *Music Theory Online* 16.

Byrd, D., J. Krivokapić, & S. Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120.1589–1599.

——, & E. Saltzman. 1998. Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26.173–199.

Carter, H. 1971. Morphotonology of Zambian Tonga: Some developents of meeussen's system. *African Language Studies* 12,14.1–30 and 36–52.

Chen, M. 2000. *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge, UK: Cambridge University Press.

Cho, T. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32.141–176.

——, & Patricia Keating. 2001. Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics* 29.155–190.

Chomsky, N., & M. Halle. 1968. *The Sound Patterns of English*. New York: Harper and Row.

Chrabaszcs, A., M. Winn, C.Y. Lin, & W. Idsardi. 2014. Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language and Hearing Research* 1–12.

Chung, Y., & A. Arvaniti. 2013. Speech rhythm in Korean: Experiments in speech cycling. *Proceedings of Meetings on Acoustics* 19.1–9.

Church, R.M., & H.A. Broadbent. 1990. Alternative representations of time, number, and rate. *Cognition* 37.55–81.

——, W.H. Meck, & J. Gibbon. 1994. Application of scalar timing theory to individual trials. *Journal of Experimental Psychology: Animal Behavior Processes* 20.135–155.

Clements, G.N. 1985. Principles of tone assignment in Kikuyu. In *Autosegmental Studies in Bantu Tone*, ed. by G.N. Clements & J. Goldsmith, 281–340, Dordrecht. Foris Publications.

——, & K.C. Ford. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* 10.

——, & J. Goldsmith. 1984. Autosegmental studies in Bantu tone: An introduction. In *Autosegmental studies in Bantu tone*, ed. by G.N. Clements & J. Goldsmith, 1–17, Dodrecht. Foris.

Correia, S., J. Butler, M. Vigário, & S. Frota. 2015. A stress 'deafness' effect in European Portuguese. *Audiology & Speech-Language Pathology* 58.48–67.

Crane, T.M., 2010. High-tone anticipation in Totela: An unexpected system. Presentation at TIE4 Conference, Stockholm.

Crosswhite, K. 2004. Vowel reduction. In *Phonetically Based Phonology*, ed. by Bruce Hayes, Donca Steriade, & Robert Kirchner, 191–231, New York. Cambridge University Press.

Crowhurst, M. 1996. An optimal alternative to conflation. *Phonology* 13.409–424.

Cumming, R.E. 2011. The interdependence of tonal and durational cues in the perception of rhythmic groups. *Phonetica* 67.219–242.

Cummins, F., & R.F. Port. 1996. Rhythmic constraints on english stress timing. In *Proceedings of the Fourth International Conference on Spoken Language Processing*, ed. by H.T. Bunell & W. Isardi, 2036–2039, Wilmington, Deleware. Alfred duPont Intitute.

——, & ——. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26.145–171.

Cummins, Fred. *Rhythmic Coordination in English Speech: An Experimental Study*. Indiana University dissertation.

——. *Rhythmic Coordination in English Speech: An Experimental Study*. Indiana University dissertation.

Cutler, A., J. Hawkins, & G. Gilligan. 1985. The suffixing preference: A processing explanation. *Linguistics* 23.723–758.

Danielson, V. 2007. The canon of ethnomusicology: Is there one? *Notes* 64.223–231.

Danis, N. 2011. Downstep and contour formation in Medumba: A prosodic account. In *Proceedings of the 42nd Annual Conference on African Linguistics*.

Dasher, R., & D. Bolinger. 1982. On pre-accentual lengthening. *Journal of the International Phonetic Association* 12.58–69.

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

Dauer, R.M. 1983. Stress-timing and syllable-timing re-analysed. *Journal of Phonetics* 11.51–62.

de Boer, B., & P. Kuhl. 2003. Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters On-line* 4.129–134.

de Jong, K. 2001. Effects of syllable affiliation and consonant voicing on temporal djustments in a repetitive speech production task. *Journal of Speech, Language, and Hearing Research* 44.826–840.

——, & B.A. Zawaydeh. 1999. Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics* 27.3–22.

de Lacy, P. 2002. The interaction of tone and stress in optimality theory. *Phonology* 19.1–32.

—— 2014. Evaluating evidence for stress systems. In *Word stress: Theoretical and typological issues*, ed. by Harry van der Hulst, 149–193. Cambridge University Press.

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

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

Demuth, K., M. Machobane, & F. Moloi. 2009. Learning how to license null noun-class prefixes in Sesotho. *Language* 85.865–883.

Di Carlo, P. 2015. Multilingualism, solidarity, and magic. new perspectives on language ideology in the Cameroonian Grassfields. In *Plurilinguismo, sintassi. Atti del XLVI congresso internazionale di studi della Societfrm[o]–a di Linguistica Italiana (SLI)*, ed. by S. Casini, C. Bruno, F. Gallina, & R. Siebetcheu, 287–302, Roma: Bulzoni.

—— In press. Multilingualism, affiliation, and spiritual insecurity. from phenomena to process in language documentation. In *Language Documentation and Conservation*, ed. by M. Seyfeddinipur, volume African Language Documentation: New Data, Methods, and Approaches.

—— In press. Towards an understanding of African endogenous multilingualism. In *International Journal of Sociology of Language*, ed. by D. Smakman & P. Heinrich.

——, & J. Good. 2014. What are we trying to preserve? diversity, change, and ideology at the edge of the Cameroonian Grassfields. In *Endangered Languages: Beliefs and Ideology*, ed. by P. Austin & J. Sallabank, 231–264, Oxford, UK. Oxford University Press.

DiCanio, C. 2014. Triqui tonal coarticulation and contrast preservation in tonal phonology. *Proceedings from Sound Systems of Mexico and Central America* .

Dilley, L., S. Mattys, & L. Vinke. 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.

——, & D. McAuley. 2008. Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language* 59.294–311.

——, J. Wallace, & C. Heffner. 2012. Perceptual isochrony and fluency in speech by normal talkers under varying task demands. In *Prosodies: Context, Function, and Communication*, ed. by O. Niebuhr & H. Pfitzinger, Language, Context, and Cognition series, 237–258, Berlin/New York. Walter de Gruyter.

D'Imperio, M., N. Nguyen, & K.G. Munhall. 2003. An articulatory hypothesis for the alignment of tonal targets in Italian. In *Proceedings of the International Congress of Phonetic Science*, 253–256, Barcelona.

Donovan, A., & C.J. Darwin. 1979. The perceived rhythm of speech. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, volume 2, 268–274, Copenhagen.

Downing, L. 1990. Local and metrical tone shift in Nguni. *Studies in African Linguistics* 21.261–317.

—— 1999. Prosodic stem ≠ prosodic word in Bantu. In *Studies on the Phonological Word*, ed. by T.A. Hall & U. Kleinhenz, Current Issues in Linguistic Theory 174, 73–98, Amsterdam. John Benjamins.

——, & M. Kadenge. 2015. Prosodic stems in Zezuru Shona. *Southern African Linguistics and Applied Language Studies* 33.291–305.

Downing, Laura. 2004. What African languages tell us about accent typology. *ZAS Papers in Linguistics* 37.101–136.

——. 2010. Accent in African languages. In *A Survey of Word Accentual Patterns in the Languages of the World*, ed. by Rob W. N. Goedemans, Harry G. van der Hulst, & Ellen A. van Zanten, 381–427, Berlin. Mouton de Gruyter.

Dunstan, M.A. *Tone and concord systems in Ngwe nominals*. London, UK: University of London dissertation.

Elordieta, G., & J.I. Hualde. 2003. Experimental test of a hypothesized diachronic change in Basque accentuation. In *Proceedings of teh 15th International Congress of Phonetic Sciences*, ed. by M.-J. Solé, D. Recasens, & J. Romero, 67–70, Barcelona. Causal Productions Pty Ltd.

Eriksson, A. 1991. Aspects of Swedish speech rhythm. *Gothenburg Monographs in Linguistics* 9, University of Gōteborg.

Faure, G., D.J. Hirst, & M. Chafcouloff. 1980. Rhythm in English: Isochronism, pitch, and perceived stress. In *The Melody of Language*, ed. by L.R. Waugh & C.H. van Schooneveld, 71–79, Baltimore. University Park Press.

Feldman, H., & K. Friston. 2010. Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience* 4.215.

Fishman, J. 1982. Whorfianism of the third kind: Ethnolinguistic diversity as a worldwide societal asset. (the whorfian hypothesis: Varieties of validation, confirmation, and disconfirmation ii). *Language in Society* 11.1–14.

Fitch, W.T. 2011. The evolution of syntax: An exaptationist perspective. *Frontiers in Evolutionary Neuroscience* 3.

Fontaney, Louise. 1980. *Eléments de description du Punu*, volume CRLS. Université Lyon II.

Fougeron, C., & Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101.3728–3739.

Franich, K. under review. Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language.

Franich, Kathryn, 2013. High tone anticipation and prosodic structure in Medʉmba. Qualifying Paper, University of Chicago.

——. 2014. Contour tones and prosodic structure in Medʉmba. In *Proceedings of the 40th Annual Berkeley Linguistic Society*, ed. by Herman Leung, Zachary O'Hagan, Sarah Bakst, Auburn Lutzross, Jonathan Manker, Nicholas Rolle, & Katie Sardinha.

Friston, K. 2009. The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences* 13.293–301.

Fudge, Eric C. 1969. Syllables. *Journal of Linguistics* 3.253–286.

Fujimura, O. 2003. Stress and tone revisited: Skeletal vs. melodic and lexical vs. phrasal. In *Cross-Linguistic Studies of Tonal Phenomena: Historical Development, Phonetics of Tone, and Descriptive Studies*, ed. by Kaji Shigeki, 221–236, Tokyo. Tokyo University of Foreign Studies.

Furere, R., & A. Rialland. 1985. Tons et accents en Kinyarwanda. In *African Linguistics*, ed. by D. Goyvaerts, 99–166, Amsterdam. John Benjamins.

Gandour, J., S. Dechongkit, & S. Ponglorpisit. 1992. Tonal coarticulation in Thai disyllabic utterances: A preliminary study. *Linguistics of the Tibeto-Burman Area* 15.93–110.

——, S. Potisuk, & S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22.477–492.

Garde, P. 1967. *L'Accent*. Presses Universitaires de France, Paris.

Gay, T. 1977. Effect of speaking rate on vowel formant movements. *Haskins Laboratories Status Report on Speech Research SR-51/52* .

Gibbon, J. 1977. Scalar expectancy theory and Weber's law in animal timing. *Psychological Review* 84.279–325.

Gibson, E. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68.1–76.

Goldman, N., W. Orman, E. Paquette, K. Franich, R. Hawkes, A. Ngabeu, & C. O'Connor. 2013. Interspeaker variation in noun class realization in Medʉmba. In *Selected Proceedings of the 44th Annual Conference on African Linguistics*, ed. by R. Kramer, E.C. Zsiga, & O.T. Boyer, Somerville, MA. Cascadilla Proceedings Project.

Goldsmith, J. *Autosegmental Phonology*. Massachusetts Institute of Technology dissertation.

——, 1981. Towards an autosegmental theory of accent: The case of Tonga. Indiana University Linguistics Club.

—— 1982. Accent systems. In *The Structure of Phonological Representations*, ed. by H. van der Hulst & N. Smith, volume 1, 47–63, Dodrecht. Foris.

—— 1984. The rise of rhythmic structure in Bantu. In *Phonologica*, ed. by W. Dressler, 65–78, Cambridge, MA. Cambridge University Press.

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

—— 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

—— 1991. Tone and accent in llogori. In *Lacustrine Bantu Phonology*, ed. by F. Katamba, 7–30, Cologne. University of Cologne.

—— 2009. Syllables. In *The Handbook of Phonological Theory, 2nd Edition*, ed. by John Goldsmith, Jason Riggle, & Yu, 162–196. Wiley Blackwell.

Goldstein, L. 2007. Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. *Proceedings of the 16th International Congress of Phonetic Sciences* 241–244.

Good, J. 2012. How to become a 'Kwa' noun. *Morphology* 22.294–335.

——, & A.C.L. Yu. 2010. Morphosyntax of two Turkish subject pronominal paradigms. In *Clitic and Affix Combinations: Theoretical Perspectives*, ed. by L. Heggie & F. Ordóñez, 315–374, Amsterdam/Philadelphia. John Benjamins.

Gordon, M. *Syllable weight: Phonetics, phonology, and typology*. UCLA dissertation.

—— 2001. A typology of contour tone restrictions. *Studies in Language* 25.405–444.

—— 2011. Stress: Phonotactic and phonetic evidence. In *The Blackwell Companion to Phonology*, ed. by Marc van Oostendorp, Colin Ewen, Elizabeth Hume, & Keren Rice, 924–948. Wiley-Blackwell.

—— 2012. The acoustic correlates of pitch accent in tashlhiyt berber. *Journal of Phonetics* 40.706–724.

—— 2014. Disentangling stress and pitch accent. In *Word Stress: Theoretical and Typological Issues*, ed. by H. van der Hulst, 83–118, Oxford, UK. Oxford University Press.

——, & L. Nafi. 2012. The acoustic correlates of stress and pitch accent in Tashlhiyt Berber. *Journal of Phonetics* 40.706–724.

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

Grahn, J.A. 2012. Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics in Cognitive Science* 4.585–606.

Greenberg, J. 1963. *The languages of Africa*. The Hague: Mouton de Gruyter.

Grenet, E. 1939. *Popular Cuban Music*. New York: Bourne Inc.

Gross, J., & U. Ligges. 2015. nortest package for R statistical softward. Technical report, TU Dortmund University.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press.

Gut, U. 2005. Nigerian English prosody. *English World-Wide* 26.153–177.

Guthrie, M. 1948. *The Classification of the Bantu Languages*. New York/Toronto: Oxford University Press.

Haken, H., J. A. S. Kelso, & H. Bunz. 1985. A theoretical model fo phas transitions in human hand movements. *Biological Cybernetics* 51.347–356.

Halle, M., & J.-R. Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Hamlaoui, F., & K. Szendröi. 2017. The syntax-phonology mapping of intonational phrases in complex sentences: A flexible approach. *Glossa: A Journal of General Linguistics* 2.55.

Han, M.S. 1962. The feature of duration in Japanese. *Onsei No Kenkyuu* 10.65–80.

Harris, J. 2004. Release the captive coda: The foot as the domain of interpretation. In *Phonetic interpretation: Papers in Laboratory Phonology*, volume 6, 103–129, Cambridge, MA. Cambridge University Press.

Hawkins, J.A., & Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In *Explaning Language Universals*, ed. by J.A. Hawkins, Oxford. Basil Blackwell.

Hawkins, S., & R. Smith. 2001. Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Journal of Italian Linguistics-Rivista di Linguistica* 13.99–188.

Hayes, B. *A Metrical Theory of Stress Rules*. MIT dissertation.

—— 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Heath, J. 2005. *A Grammar of Tamashek (Tuareg of Mali)*. Berlin/New York: Walter de Gruyter.

Heo, I., & W. Sethares. 2015. Classification based on speech rhythm via a temporal alignment of spoken sentences. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23.2209–2216.

Hoequist, C.E. 1983. The perceptual center and rhythm categories. *Language and Speech* 26.367–376.

Hualde, J.I., G. Elorideta, I. Gaminde, & R. Smiljanić. 2002. From pitch accent to stress accent in basque. In *Laboratory Phonology 7*, volume 7, 547–584, Berlin. Mouton de Gruyter.

Hyde, B. 2002. A restrictive theory of metrical stress. *Phonology* 19.313–359.

—— 2016. The role of phenomenal accent. In *Dimensions of Phonological Stress*, ed. by J. Heinz, R. Goedemans, & H. van der Hust, 9–48. Cambridge University Press.

Hyman, Larry M. 1972. A phonological study of Fe'fe'-Bamileke. *Studies in African Linguistics* Supplement 4.

Hyman, L.M. 1977. On the nature of linguistic stress. In *SCOPIL 4*, ed. by Larry M. Hyman, volume 4, 37–82.

—— 1978. Tone and/or accent. In *Elements of Tone, Stress and Intonation*, ed. by D.J. Napoli, 1–20, Washington. Georgetown University Press.

—— 1989. Accent in Bantu: An appraisal. *Studies in the Linguistic Sciences* 19.115–134.

—— 1990. Boundary tonology and the prosodic hierarchy. In *The Phonology-Syntax connection*, ed. by S. Inkelas & D. Zec, 109–125. University of Chicago Press.

—— 1998. Positional prominence and the 'prosodic trough' in Yaka. *Phonology* 15.41–75.

—— 2001. Privative tone in Bantu. In *Cross-linguistic studies of tonal phenomena*, ed. by Shigeki Kaji, 237–257, Tokyo. Institute for the Study fo Languages and Cultures.

—— 2006. Word-prosodic typology. *Phonology* 23.225–257.

—— 2007. Universals of tone rules: 30 years later. In *Tones and tunes: Studies in word and sentence prosody*, ed. by Tomas Riad & Carlos Gussenhoven, 1–34, Berlin. Mouton de Gruyter.

—— 2008. Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46.

—— 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31.213–238.

—— 2010. Do all languages have word accent? or: What's so great about being universal? *UC Berkeley Phonology Lab Annual Report* 297–315.

—— 2015. Positional prominence vs. word accent: Is there a difference? *UC Berkeley Phonology Lab Annual Report* 86–99.

——, & K.C. Monaka. 2008. Tonal and non-tonal intonation in Shekgalagari. *UC Berkeley Lab Annual Report* 269–288.

——, & R.G. Schuh. 1974. Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry* 5.81–115.

——, & M. Tadadjeu. 1976. Floating tones in Mbam-Nkam. *Southern California Occasional Papers in Linguistics* 3.57–111.

——, E.F.K. Voeltz, & G. Tchokokam. 1970. Noun class leveling in bamileke. *Studies in African Linguistics* 1.185–209.

Iskarous, K. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. In *Journal of Phonetics Special Issue: Mechanisms of Regulation in Speech*, ed. by H. Müke, volume 64, 8–20.

Itô, J. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7.217–260.

——, & A. Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17.49–73.

Itô, J., & A. Mester. 2009. The extended prosodic word. In *Phonological domains: Universals and deviations*, ed. by J. Grijzenhout & Baris Kabak, 135–194, Berlin. Mouton de Gruyter.

——, & ——. 2012. Recursive prosodic phrasing in Japanese. In *Prosody matters: Essays in honor of Elisabeth Selkirk*, ed. by T. Borowsky, S. Kawahara, T. Shinya, & M. Sugahara, 280–303, London. Equinox.

——, & ——. 2013. Prosodic subcategories in Japanese. *Lingua* 124.20–40.

Jones, A.M. 1959. *Studies in African Music*. Oxford University Press.

Jones, M.R., & M. Boltz. 1989. Dynamic attending and responses to time. *Psychological Review* 96.459–491.

——, & P.Q. Pfordresher. 1997. Tracking melodic events using joint accent structure. *Canadian Journal of Experimental Psychology* 51.271–291.

Jun, S.-A. *The Phonetics and Phonology of Korean Prosody*. Ohio State University dissertation.

—— 2012. Prosodic typology revisited: Adding macro-rhythm. In *Proceedings of Speech Prosody 6*.

Jun, S-A. 2014. Prosodic typology: By prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II: The Phonology of Intonation and Phrasing*, ed. by S-A. Jun, 520–539, Oxford, UK. Oxford University Press.

Kager, R. 2007. Feet and metrical stress. In *The Cambridge Handbook of Phonology*, ed. by Paul de Lacy, Cambridge, UK. Cambridge University Press.

Kaji, S. 1994. Tone reversal in Tembo (Bantu j.57). *GENGO KENKYU (Journal of hte Linguistic Society of Japan)* 105.54–86.

Kaplan, A. *Noniterativity as an emergent property of grammar*. University of California at Santa Cruz dissertation.

Katsika, A., J. Krivokapić, C. Mooshammer, M. Tiede, & L. Goldstein. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44.62–82.

Kawahara, S. 2015. The phonology of japanese accent. In *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, ed. by H. Kubozono, 445–491. Mouton de Gruyter.

——, D. Erickson, J. Moore, A. Suemitsu, & Y. Shibuya. 2014. Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress. *Onsei Kenkyu* 18.77–87.

Keating, P., T. Cho, C. Fougeron, & C-S. Hsu. 2003. Domain-initial strengthening in four languages. In *Phonetic interpretation: Papers in Laboratory Phonology*, ed. by J. Local, R. Ogden, & R. Temple, 143–161. Cambridge University Press.

——, & S. Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101.112–156.

Keller, P.E., & B.H. Repp. 2008. Multilevel coordination stability: Integrated goal representations in simultaneous intra-personal and inter-agent coordination. *Acta Psychologica* 128.378–386.

Kelso, J. A. S. 1984. Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory Integrative and Comparative Physiology* 15.R1000–R1004.

—— 1995. *Dynamic Patterns*. Cambridge, MA: MIT Press.

—— 2009. Coordination dynamics. In *Encyclopedia of complexity and system science*, ed. by R. A. Meyers, 1537–1564, Heidelberg, Germany. Springer.

——, J.D. DelColle, & G. Schöner. 1990. Action-perception as a pattern formation process. In *Attention and Performance*, ed. by M. Jeannerod, 139–169, Hillsdale, NJ. Erlbaum.

——, D. Southard, & D. Goodman. 1979. On the nature of human interlimb coordination. *Science* 203.1029–1031.

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

—— Under review. The syntax and semantics of associative plurals in Bamileke medumba. *African Languages and Linguistics* .

——, & M. Wiltschko. Under review. Polar questions in Bamileke Medumba. *Journal of West African Languages* .

Keupdjio, Hermann. 2014. A-bar movement, a-bar agreement and resumption in Bamileke-Medumba. In *Proceedings of the Qualifying Papers Mini-conferences 2013-2014*, ed. by A. Anghelescu, J. Dunham, & N. Weber, volume 42. Proceedings of the Qualifying Papers Mini-conferences 2013-2014.

Key, M. 2007. Headed spans in Bantu tonology. *University of Massachusetts Occasional Papers in Linguistics* 32.187–207.

Kim, H., & J. Cole. 2005. The stres foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. In *Proceedings of the 9th European conference on speech communication and technology*, 2365–2368.

Kim, S., & T. Cho. 2009. The useof phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *Journal of the Acoustical Society of America* 125.3373–3386.

Kimball, A., & J. Cole. 2014. Avoidance of stress clash in perception of American English. *Proceedings of Speech Prosody 7* .

Kimenyi, A. 1979. *A Relational Grammar of Kinyarwanda*. University of California Press.

Kingston, J. 2003. Mechanisms in tone reversal. In *Cross-linguistic studies in tonal phenomena: historical development, phonetics of tone, and descriptive studies*, ed. by Shigeki Kaji, 57–120. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA).

Kiparsky, P. 1982. Lexical morphology and phonology. In *Linguistics in the morning calm*, ed. by I.-S. Yang, 3–91, Seoul. Hanshin.

Kirby, J.P. *Cue Selection and Category Restructuring in Sound Change*. Chicago, IL: Univeristy of Chicago dissertation.

Kisseberth, C.W. 2007. Tone. *Language* 83.662–664.

Köhnlein, B. 2013. Optimizing the relation between tone and prominence: Evidence from Franconian, Scandinavian, and Serbo-Croatian tone accent systems. *Lingua* 131.1–28.

Kouankem, C. *The syntax of the Medumba determiner phrase*. University of Yaounde I dissertation.

Krivokapić, J., & D. Byrd. 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics* 40.430–442.

Kubik, G. 2010. *Theory of African Music*, volume 2. University of Chicago Press.

Kwenzi, M.J. 1980. Esquisse phonologique du Punu. In *Eléments de description du Punu*, ed. by F. Nsuka-Nkutsi, volume CRLS, 7–18. Université Lyon II.

Ladd, D.R. 1996. *Intonational Phonology*. Cambridge Studies in Linguistics 79. Cambridge University Press.

——, D. Faulker, H. Faulkner, & A. Schepman. 1999. Constant 'segmental anchoring' of f0 mvements under changes in speech rate. *Journal of the Acoustical Society of America* 1543–1554.

Ladefoged, P. 1975. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Large, E., & M.R. Jones. 1999. The dynamics of attending: How we track time varying events. *Psychological Review* 106.119–159.

Leben, W. *Suprasegmental Phonology*. MIT dissertation.

—— 1997. Tonal feet and the adaptation of english borrowings into Hausa. *Studies in African Linguistics* 25.139–154.

—— 2002. Tonal feet and the adaptation of english borrowings into hausa. In *Studies in African Linguistics*, volume 25, 139–154.

Lee, C.S., & N.P.M. Todd. 2004. Towards an auditory account of speech rhythm: Application of a model of the auditory primal sketch of two multi-language corpora. *Cognition* 93.225–254.

Lee, S., D. Byrd, & J. Krivokapić. 2006. Functional data analysis of prosodic effects on articulatory timing. *Journal of the Acoustical Society of America* 119.1666–1671.

Lehiste, I. 1977. Isochrony reconsidered. *Journal of Phonetics* 5.253–263.

Leong, V. *Prosodic Rhythm in the Speech Amplitude Envelope*. Cambridge University dissertation.

Lerdahl, R., & R. Jackendoff. 1983. *A Generative Theory of Tonal Music*. MIT Press.

Li, Y., T. Lee, & Y. Qian. 2003. Analysis and modeling of f0 contours for Cantonese text-to-speech. *ACM Transactions on Asian Language Information Processing* 3.169–180.

Liberman, M. *The intonational system of English*. MIT dissertation.

——, & A. Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8.249–336.

Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 96.40–55.

Lindsay, B.G. 1995. Mixture models: Theory, geometry, and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, ed. by J. Collins, P. Kelly, & M. Gasko Donoho, volume 5, Hayward. Institute of Mathematical Statistics.

Lipski, J.M. 2015. Toned-up Spanish: Stress to pitch to tone (?) in Equatorial Guinea. In *Romance Linguistics 2013: Selected Papers from the 43rd Linguistic Symposium on Romance Languages (LSRL)*, ed. by C. Tortora, M. den Dikken, I.L. Montoya, & T. O'Neill, 233–256.

Liu, F. *Intonation Systems of Mandarin and English: A Functional Approach*. Chicago, IL: University of Chicago dissertation.

Llisterri, J., M. Machuca, C. de la Mota, M. Riera, & A. Ríos. 2003. The perception of lexical stress in Spanish. In *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. by M.-J. Solé, D. Recasens, & J. Romero, Barcelona.

Locke, D. 1982. Principles of offbeat timing and cross-rhythm in southern e̲ve dance. *Ethnomusicology* 26.217–246.

Lovegren, J. 2012. Stem-initial prominence in Mungbam. In *Selected Proceedings of the 42nd Annual Conference on African Linguistics*, ed. by Michael Marlo, Nikki B. Adams, Christopher R. Green, Michelle Morrison, & Tristan M. Purvis, 47–54, Somerville, MA. Cascadilla Proceedings Project.

Low, E.L., E. Grabe, & F. Nolan. 2000. Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. *Language and Speech* .

Lowe, J.J. 2011. R̥gvedic clitics and 'prosodic movement'. In *Proceedings of the LFG11 Conference*, ed. by M. Butt & T. Holloway King, 360–380. CSLI Publications.

—— 2014. Accented clitics in the R̥gveda. *Transac* 112.5–43.

Maddieson, I. 1978. Universals of tone. In *Universals of Human Language: Volume 2, Phonology*, ed. by J. Greenberg, 335–363, Stanford. Stanford University Press.

Mann, V. 1980. Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics* 28.407–412.

Manrique, A.M.B., & A. Signorini. 1983. Segmental reduction in Spanish. *Journal of Phonetics* 11.117–128.

Marantz, A. 1982. Re reduplication. *Linguistic Inquiry* 13.435–482.

Marin, J.M., K. Mengersen, & C.P. Robert. 2011. Bayesian modelling and inference on mixtures of distributions. In *Essential Bayesian Models*, ed. by D. Dey & C.R. Rao, volume 25 of *Handbook of Statistics: Bayesian Thinking - Modeling and Computation*. Elsevier.

Marlo, M. *The verbal tonology of Lumarachi and Lunyala-West: Two dialects of Luluyia (Bantu, J.30, Kenya)*. Ann Arbor, MI: University of Michigan dissertation.

—— 2008. Reversal and predictability in Luyia tone. In *Paper presented at the 38th Annual Conference on African Linguistics*, University of Georgia.

Martinet, A. 1960. *Eléments de Linguistique Générale*. Paris: Librarie Armand Colin.

Martínez-Paricio, V. 2012. Superfeet as recursion. In *Proceedings of the 30th West Coast Conference on Formal Linguistics*, ed. by N. Arnett & R. Bennett, 259–269.

—— *Maximal and minimal feet in Chugash*. University of Tromsø dissertation.

Matell, M.S., & W.H. Meck. 1999. Reinforcement-induced within-trial resetting of an internal clock. *Behavioral Processes* 45.159–171.

McCarthy, J. 1981. A prosodic theory of non-concatenative morphology. *Linguistic Inquiry* 12.373–418.

——, 2004. Headed spans and autosegmental spreading. Amherst, MA: University of Massachusetts Amherst.

McCawley, J.D. 1970. Some tonal systems that come close to being pitch accent systems but do not quite make it. In *Proceedings of the Chicago Linguisic Society*, volume 6, 526–531.

—— 1978. What is a tone language? In *Tone: A Linguistic Survey*, ed. by V.A. Fromkin, 113–131, New York. Academic Press.

Meeussen, A.E. 1963. Morphotonology of the Tonga verb. *Journal of African Linguistics* 2.72–92.

Mehler, J., E. Dupoux, T. Nazzi, & G. Dehaene-Lambertz. 1996. Coping with linguistic diversity: the infant's viewpoint. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. by J.L. Morgan & K. Demuth, 101–116, Mahwah, NJ. Lawrence Erlbaum Associates.

Mitterer, H. 2006. On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics* 68.1227–1240.

Mohanan, K. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.

Montfort, M. 1987. *Ancient Traditions Future Possibilities: Rhythmic Training Through the Traditions of Africa, Bali, and India*. Panoramic Press.

Morton, J., S. Marcus, & C. Frankish. 1976. Perceptual centers (p-centers). *Psychological Review* 83.405–408.

Mtenje, A.D. 1987. Tone shift principles in the Chichewa verb: A case for a tone lexicon. *Lingua* 72.169–207.

Mucha, A. 2017. Past interpretation and graded tense in medumba. *Natural Language Semantics* 25.1–52.

Mutonyi, N. *Aspects of Bukusu morphology and phonology*. Columbus, OH: The Ohio State University dissertation.

Myers, S. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15.847–892.

—— 2000. Boundary disputes: The distinction between phonetic and phonological sound patterns. In *Phonological Knowledge: Conceptual and Empirical Issues*, ed. by N. Burton-Roberts, P. Carr, & G. Docherty, 245–272, Oxford, UK. Oxford University Press.

Myers, Scott. 2003. F_0 timing in Kinyarwanda. *Phonetica* 60.71–97.

Myers, S.P. *Tone and Structure of Words in Shona*. The University of Massachusetts, Amherst dissertation.

Nakatani, L.H., J.D. O'Connor, & C.H. Aston. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38.84–105.

Nash, J. *Aspects of Ruwund grammar*. University of Illinois at Urbana-Champaign dissertation.

—— 1992-1994. Underlying low tones in Ruwund. *Studies in African Linguistics* 23.223–278.

Nespor, M. 1990. On the rhythm parameter in phonology. In *The Logical Problem of Language Acquisition*, ed. by I. Roca, 157–175, Foris. Dordrecht.

——, & I. Vogel. 1982. Prosodic domains of external sandhi rules. In *The Structure of Phonological Representations*, ed. by H. van der Hulst & N. Smith, 225–256, Dodrecht. Foris.

——, & ——. 1983. Prosodic structure above the word. In *Prosody: Models and Measurements*, ed. by Anne Cutler & D.R. Ladd, 123–140, Berlin. Springer Verlag.

——, & ——. 1986. *Prosodic Phonology*. Dodrecht, Holland and Riverton, NJ: Foris.

Nevins, A., & B. Vaux. 2003. Metalinguistic, schmetalinguistic: A phonology of shm-reduplication. *Proceedings of the 39th Chicago Linguistic Society* .

Nganmou, A. *Modalités Verbales*. University of Yaounde dissertation.

Nketia, J.H.K. 1963. *African Music in Ghana*. Northwestern University Press.

Nooteboom, S.G. 1981. Lexical retrieval from fragments of spoken words: beginnings vs. endings. *Journal of Phonetics* 9.407–424.

Nozaradan, S., I. Peretz, & P.E. Keller. 2016. Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Scientific Reports* 6.20612.

Odden, D. 1990. Tone in the Makonde dialects: Chimaraba. *Studies in African Linguistics* 21.61–105.

—— 2015. *Bantu Phonology*. Oxford University Press.

——, & L. Bickmore. 2014. Melodic tone in Bantu: An overview. *Africana Linguistica* 20.3–13.

O'Dell, M., & T. Nieminen. 1999. Coupled oscillator model of speech rhythm. *Proceedings of the International Congress on Phonetic Sciences* 1075–1079.

——, & ——. 2008. Coupled oscillator model for speech timing: Overview and examples. In *Proceedings of the Xth Conference, Helsinki*, ed. by M. Vainio, R. Aulanko, & O. Aaltonen, 179–191.

Overdulve, C.M. 1975. *Apprendre la langue Rwanda*. Mouton de Gruyter.

Patel, A.D. 2003. Language, music, and syntax in the brain. *Nature Neuroscience* 6.674–681.

——, & J.R. Daniele. 2003a. Stress-timed vs. syllable-timed music? a comment on huron and ollen. *Music Perception* 21.273–276.

Pearce, M. 2006. The interaction between metrical structure and tone in Kera. *Phonology* 23.259–286.

Pearson, K. 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A* 185.71–110.

Penalosa, D. 2009. *The Clave Matrix - Afro-Cuban Rhythm: Its Principles and African Origins*. Bembe Books, First Edition.

Peperkamp, S., & E. Dupoux. 2002. A typological study of stress 'deafness'. In *Laboratory Phonology 7*, ed. by C. Gussenhoven & Warner. N., 203–240.

Philippson, G. *Ton et accent dans les langues bantu d'Afrique Orientale: Étude comparative typologique et diachronique*. Université René Descartes, Paris dissertation.

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

Pierrehumbert, J. *The phonology and phonetics of English intonation*. Cambridge, MA: MIT dissertation.

—— 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I*, 72–106, Cambridge, UK. Cambridge University Press.

——, & M.E. Beckman. 1988. *Japanese Tone Structure*, volume Linguistic Inquiry Monograph of *15*. MIT Press.

——, & D. Talkin. 1992. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, ed. by G.J. Docherty & D.R. Ladd, 90–117. Cambridge University Press.

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

Pike, Kenneth. 1967. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*. The Hague: Mouton.

Pointon, G.E. 1980. Is Spanish really syllable-timed? *Journal of Phonetics* 8.293–304.

Polak-Bynon, L. 1975. *A Shi grammar*. Tervuren: Musée Royal de l'Afrique Centrale.

Port, R.F. 2003. Meter and speech. *Journal of Phonetics* 31.599–611.

——, J. Dalby, & M. O'Dell. 1987. Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81.1574–1585.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, & B.P. Flannery. 2007. *Numerical Recipes: The Art of Scientific Computing (3rd Ed.)*. New York: Cambridge University Press.

Prieto, P., J. van Santen, & J. Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23.429–451.

Prince, A. 1983. Relating to the grid. *Linguistic Inquiry* 14.19–100.

——, & P. Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Rutgers University Center for Cognitive Science Technical Report 2.

Quené, H., & R.F. Port. 2003. Produced speech rhythm depends on predictability of stress patterns. In *Proceedings of the International Congress on Phonetic Sciences XV*, ed. by M.-J. Solé, D. Recasens, & J. Romero, 2445–2448, Barcelona.

Ramus, F. 2002b. Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition, 2* .

——, M. Nespor, & J. Mehler. 1999. Correlates of linguistic rhythm. *Cognition* 73.265–292.

Remijsen, B., & V.J. van Heuven. 2005. Stress, tone, and discourse prominence in the curaçao dialect of papiamentu. *Phonology* 22.205–235.

Repp, B.H. 2001. Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance* 27.600–621.

—— 2002a. Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance* 28.410–430.

—— 2002b. Phase correction in sensorimotor synchronization: Nonlinearities in voluntary and involuntary responses to perturbations. *Human Movement Science* 21.1–31.

—— 2003. Phase attraction in sensorimotor synchronization with auditory sequences: effects of single and periodic distractors on synchronization accuracy. *Journal of Experimental Psychology: Human Perception and Performance* 29.290–309.

—— 2004. On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. *Human Movement Science* 23.389–413.

—— 2005. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review* 12.969–992.

—— 2008. Sensorimotor synchronization with adaptively timed squences. *Human Movement Science* 27.423–456.

—— 2010. Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science* 29.200–213.

Rice, C. *Binarity and ternarity in metrical theory: Parametric extensions*. University of Texas at Austin dissertation.

Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In *Linguistic Controversies*, ed. by D. Crystal, 73–79, London. Arnold.

Rose, S., & P. Jenks. 2011. High ton in Moro: Effects of prosodic categories and morphological domains. *Natural Language and Linguistic Theory* 29.211–250.

Rosenberg, A., J. Hirschberg, & K. Manis. 2010. Perception of English prominence by native Mandarin Chinese speakers. In *Proceedings of Speech Prosody*.

Rouas, J., J. Farinas, F. Pellegrino, & R. Andre-Obrecht. 2005. Rhythmic unit extraction and modelling for automatic language identification. *Speech Communication* 47.436–456.

Rubach, J. 1990. Final devoicing and cyclic syllabification in German. *Linguistic Inquiry* 21.79–94.

Sagey, E. *The representation of features and relations in non-linear phonology*. MIT dissertation.

Saltzman, E. 1986. Task dynamic coordination of the speech articulators: A preliminary model. In *Experimental Brain Research Series 15*, ed. by Heuer H. & C. Fromm, 129–144, New York. Springer-Verlag.

——, & J. A. S. Kelso. 1983. Skilled actions: A task dynamic approach. *Psychological Review* 94.84–106.

——, H. Nam, J. Krivokapić, & L. Goldstein. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In *Proceedings of Speech Prosody*, ed. by P. Barbosa, 175–184.

Schadeberg, T. 1977. Anticipation of tone: Evidence from Bantu. In *Language and Linguistic Problems in Africa*, ed. by P.F. Kotey & H. Der-Houssikian, 195–204, Colombia, SC. Hornbeam Press.

Schiering, R., B. Bickel, & K. Hildebrandt. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46.657–709.

Schmidt, R.C., M. Bienvenu, P.A. Fitzpatrick, & P. Amazeen. 1998. A comparison of intra-and interpersonal interlimb coordination: Coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance* 24.884–900.

Scott, Sophie K. *P-centers in Speech*. University College London dissertation.

Selkirk, E. 1986. On derived domains in sentence phonology. *Phonology* 3.371–405.

—— 1996. The prosodic structure of function words. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. by K. Demuth & J. Morgan, 187–214, New Jersey. Lawrence Erlbaum Associates.

Selkirk, E.O. 1978b. On prosodic structure and its relation to syntactic structure. In *Nordic Prosody II*, ed. by T. Fretheim, Trondheim. TAPIR.

—— 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11.563–605.

—— 1981. English compounding and the theory of word structure. In *The Scope of Lexical Rules*, ed. by M. Moortgat, H. van der Hulst, & T. Hoekstra, 229–277.

—— 1984. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.

—— 1995. *The Handbook of Phonological Theory*, chapter Sentence prosody: Intonation, stress and phrasing. London: Blackwell.

—— 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*, ed. by John Goldsmith, Jason Riggle, & Alan Yu. Wiley, John & Sons, Incorporated.

Shaw, J.A. 2013. The phonetics of hyper-active feet: Effects of stress priming on speech planning and production. *Laboratory Phonology* 4.159–189.

Shen, Y., & G.G. Peterson. 1962. Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers* 9.1–36.

Sideltsev, A.V. 2016. Accented clitics in Hittite? *Transactions of the Philological Society* 115.176–211.

Sietsma, B.M. *Metrical Dependencies in Tone Assignment*. MIT dissertation.

Simons, G.F., & C.D. Fennig. 2017. *Ethnologue: Languages of the World, Twentieth edition.*. Dallas, TX: SIL International.

Smiljanić, R. *Lexical, Pragmatic and Positional Effects on Prosody in Two Dialects of Croatian and Serbian: An Acoustic Study*. University of Illinois at Urbana-Champaign dissertation.

Smith, J.L. *Phonological augmentation in prominent positions*. University of Massachusetts, Amherst dissertation.

Steele, J. 1779. *Prosodia Rationalis*. London: J. Nichols.

Stephens, M.A. 1974. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* 69.730–737.

Steriade, D., 1997. Phonetics in phonology: The case of laryngeal neutralization. Ms., UCLA.

Stetson, R.H. 1951. *Motor phonetics*. Boston, MA: College-Hill Press.

Stevick, E.W. 1969. Pitch and duration in Ganda. *Journal of African Languages* 8.1–28.

Stupacher, J., G. Wood, & M. Witte. 2017. Neural entrainment to polyrhythms: A comparison of musicians and non-musicians. *Frontiers in Neuroscience* 11.208.

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

Temperley, D. 2000. Meter and grouping in African music: A view from music theory. *Ethnomusicology* 44.65–96.

Tilsen, S. 2006. Syllable timing in repetition disfluency. *UC Berkeley Phonology Lab Annual Report* 73–114.

Toscano, J.C., & B. McMurray. 2008. Using the distributional statistics of speech sounds for weighting and integrating acoustic cues. In *Proceedings of the Cognitive Science Society*, Mahwah, NJ. Erlbaum.

——, & ——. 2010. Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. In *Cognitive Science 34*, 434–464.

Toussaint, G.T. 2003. Classification and phylogenetic analysis of African ternary rhythm timelines. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*, 25–36, Granada, Spain.

—— 2005. The geometry of musical rhythm. In *Proceedings of the Japan Conference on Discrete and Computational Geometry*, ed. by J. Akiyama, M. Kano, & X. Tan, volume 3742, 198–212.

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

—— 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30.219–255.

Tuller, B., & J. A. S. Kelso. 1991. The production and perception of syllable structure. *Journal of Speech and Hearing Research* 34.501–508.

Turk, A.E., & S. Shattuck-Hufnagel. 2013. What is speech rhythm? a commentary inspired by Arvaniti & Rodriguez, Krivokapić, and Goswami & Leong. In *Laboratory Phonology*, volume 4, 93–118.

Vallabha, G.K., J.L. McClelland, F. Pons, J.F. Werker, & S. Amano. 2007. Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences* 104.13273–13278.

van der Hulst, H. 1999. Word accent. In *Word prosodic systems in the languages of Europe*, ed. by H. van der Hulst, 3–116, Berlin & New York. Mouton de Gruyter.

—— 2011. Pitch accent systems. In *The Blackwell Companion to Phonology*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, & Keren Rice, volume 2, 1003–1026. Wiley-Blackwell.

——, & C. Ewen. 2001. *The phonological structure of words: An Introduction*. Cambridge, UK: Cambridge University Press.

——, & R. Goedemans. 2009. Stresstyp: A database for word accentual patterns in the world's languages. In *The use of databases in cross-linguistics research*, ed. by M. Everaert & S. Musgrave, 235–282, New York/Berlin. Mouton de Gruyter.

van der Hulst, H.G. 1984. *Syllable structure and stress in Dutch*. Dordrecht: Foris Publications.

—— 1996. Separating primary accent and secondary accent. In *Stress Patterns of the World*, ed. by Rob W. N. Goedemans, Harry G. van der Hulst, & E. Visch, HIL Publications 2, 1–26, The Hague. Holland Academic Graphics.

van Santen, J.P.H., & J. Hirschberg. 1994. Segmental effects on timing and height of pitch contours. In *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan.

van Spaandonck, M. 1971. *L'analyse morphotonologique dans les langues Bantoues: Identification des morphotonemes et description de leurs representations tonologiques.*. Paris: SELAF.

Varlet, M., C.A. Coey, R.C. Schmidt, L. Marin, B.G. Bardy, & M.J. Richardson. 2014. Influence of stimulus velocity profile on rhythmic visuomotor coordination. *Journal of Experimental Psychology: Human Perception and Performance* 40.1849–1869.

——, C.A. Coey, R.C. Schmidt, & M.J. Richardson. 2012. Influence of stimulus amplitude on unintended visuomotor entrainment. *Human Movement Science* 31.541–552.

Vergnaud, J.R., & M. Halle, 1979. Metrical phonology. Unpublished manuscript, Department of Linguistics, MIT.

Voorhoeve, J. 1967. Personal pronouns in Bamileke. *Lingua* 17.421–430.

—— 1968. Noun classes in Bamileke. *Lingua* 21.584–593.

—— 1971. Tonology of the Bamileke noun. *Journal of African Languages* 2.44–53.

—— 1976. *Contes Bamileke*. Number 89 in 8. Tervuren, Belgium: Museé Royal de L'Afrique Centrale.

Vorberg, D., & A. Wing. 1996. Modeling variability and dependence in timing. In *Handbook of Perception and Action*, ed. by H. Heuer & S.W. Keele, volume 2, 181–262, London, UK. Academic Press.

Vuuset, P., A. Roepstorff, M. Wallentin, K. Mouridsen, & L. Ostergaard. 2006. It don't mean a thing...keeping the rhythm during polyrhythmic tension activates language areas (ba47). *Neuroimage* 31.832–841.

Vuust, P., M. Wallentin, K. Mouridsen, L. Ostergaard, & A. Roepstorff. 2011. Tapping polyrhythms in music activates language areas. *Neuroscience Letters* 494.211–216.

——, & M.A. Witek. 2014. Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology* 5.1111.

Walton, J. 1955. Iron gongs from the Congo and Southern Rhodesia. *Royal Anthropological Institute of Great Britain and Ireland* 55.20–23.

Warner, N., T. Otake, & T. Arai. 2010. Intonational structure as a word boundary cue in Japanese. *Language and Speech* 53.107–131.

Waterman, R.A. 1948. “hot” rhythm in Negro music. *Journal of the American Musicological Society* 1.

Watters, J.R. 2003. *Grassfields Bantu*. Routledge.

Wenk, B., & F. Wioland. 1982. Is French really syllable-timed. *Journal of Phonetics* 10.193–216.

Williams, B. 1985. Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics* 13.381–406.

Wing, A.M., & A.B. Kristofferson. 1973. Response delays and the timing of discrete motor responses. *Perception & Psychophysics* 14.5–12.

Wong, P.C., E. Skoe, N.M. Russo, T. Dees, & N. Kraus. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 10.420–422.

Xu, Y. 1999. F0 peak delay: When, where, and why it occurs. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 1881–1884, San Francisco.

——, & F. Liu. 2006. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Rivista di Linguistica* 18.125–159.

——, & X. Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111.1399–1413.

Xu, Yi. 1994a. Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95.2240–2253.

——. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25.61–83.

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

—— 1989. Contour tones. *Phonology* 6.149–174.

—— 1995. Tone in East Asian languages. In *The Handbook of Phonological Theory*, ed. by John Goldsmith, 476–494, Cambridge. Blackwell.

Yip, Moira. 2002. *Tone*. Cambridge & New York: Routledge.

Yu, A.C.L. 2007. *A Natural History of Infixation*. Oxford University.

—— 2010. Tonal effects on perceived vowel duration. In *Laborabory Phonology 10*, 151–168, Berlin. Mouton de Gruyter.

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

Zec, D. 2002. On the prosodic status of function words. *Working Papers of the Cornell Phonetics Laboratory* 14.

Zerbian, S. 2006. High tone spread in the Sotho verb. In *Selected Proceedings of the 35th Annual Conference on African Linguistics*, ed. by John Mugane, 145–157, Somerville, MA. Cascadilla Proceedings Project.

Zhang, J. *The Effects of Duration and Sonority on Contour Tone Distribution–Typological Survey and Formal Analysis*. UCLA dissertation.

Zwicky, A.M., 1977. On clitics. Indiana University Linguistics Club.

——, & G.K. Pullum. 1983. Cliticization vs. inflection: English N'T. *Language* 59.502–513.