

**DECONSTRUCTING THE SOUTHEAST ASIAN SESQUISYLLABLE:
A GESTURAL ACCOUNT**

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Becky Ann Butler

January 2014

© 2014 Becky Ann Butler

**DECONSTRUCTING THE SOUTHEAST ASIAN SESQUISYLLABLE:
A GESTURAL ACCOUNT**

Becky Ann Butler, Ph.D.

Cornell University 2014

This dissertation explores a purportedly unusual word type known as the *sesquisyllable*, which has long been considered characteristic of mainland Southeast Asian languages. Sesquisyllables are traditionally defined as ‘one and a half’ syllables, or as one *major* syllable preceded by one *minor* syllable, which is phonologically reduced in terms of segmental inventory, prosodic prominence and syllable shape. The goal of the dissertation is to deconstruct the notion of the sesquisyllable via empirical acoustic investigation of the minor syllable, the results of which are interpreted in light of Articulatory Phonology. These results show that purported sesquisyllables can be reanalyzed as two different types of words: (i) monosyllables with word-initial consonant clusters that have excrescent transition states and (ii) maximally disyllabic iambs. I argue that only the latter of these two should be considered a sesquisyllable.

The dissertation begins with a description of the cross-linguistic properties of sesquisyllables. Based on these characteristics, I propose both a structural/prosodic model of the sesquisyllable and an articulatory model of the minor syllable, which focuses on mid central (schwa-like) vocalic elements. Throughout the dissertation I maintain that an integrated phonological approach which relies on both of these models is necessary to adequately account for the sesquisyllable.

My analysis is supported by phonetic evidence from three purportedly sesquisyllabic languages: Khmer, Bunong (Mnong) and Burmese. Minor syllable “vowels” in Khmer are shown to be excrescent transition states whose voicing is dependent on neighboring consonants, while minor syllable vowels in Bunong are determined to be phonological. I also present a pilot study of Burmese trisyllabic ‘extended sesquisyllables’ which broadens the scope of word types that might be considered sesquisyllables. The dissertation concludes with a discussion of how the disyllabic nature of sesquisyllables suggests that prosody and articulation might further be integrated in terms of oscillation.

BIOGRAPHICAL SKETCH

Becky was born in Danville, Virginia, in January 1982. After graduating from Tunstall High School in 2000, she attended the University of Mary Washington for one year and then transferred to the University of Virginia, where she completed her B.A. in linguistics and Spanish in 2003, graduating with High Distinction. She subsequently received her M.A. in linguistics at the University of North Carolina in 2005 and began doctoral studies at Cornell University in 2007. She currently lives in Toronto with her husband Neil Ashton, along with their son Idris and cat Yao Ming.

To my parents,
Kate and Bill,
and my grandparents,
Mary, Dick, Delores and Jack

ACKNOWLEDGMENTS

First off, I would like to thank my committee chair Abigail Cohn for both her thoughtful criticisms of my work and her encouragement throughout the writing process. Her dedication to my success as a graduate student did not go unnoticed, most particularly in her willingness to have regular Skype meetings with me from halfway around the world. Her simultaneous dedication to her work, her family and herself is something to which many people aspire but which few attain. I am very grateful to have had the opportunity to be mentored by her.

I also offer my thanks to my committee members: Draga Zec and Sam Tilsen. Draga's detailed critiques of my work have undoubtedly made it stronger, and she has consistently challenged me to be a better scholar. I thank Sam for his willingness to listen and to engage with new ideas and for his contagious enthusiasm about phonetics. I could not have completed this dissertation without them.

In addition to my committee members, I owe a debt of gratitude to other academic mentors who have all helped me grow over the years. In particular, I wish to thank Jennifer Smith and Elliott Moreton at UNC. I am also grateful for Christina Kakava, my first linguistics professor during my freshman year of college; I will always remember her kindness and the joy she brought to the classroom.

This work would also not have been possible without a number of people who helped me create and check wordlists: Hannah Phan for Khmer, Becky and Todd Bequette for Bunong and Tom Patton for Burmese. I am further indebted to the Bequettes for their assistance in Cambodia,

from helping set me up with consultants to sharing linguistic resources to lending me a motorbike helmet and a lot of things in between! I cannot thank them enough.

I would also like to thank my language consultants. First off, thank you to the MA students at the Royal University of Phnom Penh who were so excited about my project they organized themselves so that I could record them. I am also grateful to my Bunong consultants for welcoming me into their homes and workplaces and for their kindness to a stranger. Thanks also to my Burmese participants who were willing to help out on short notice.

I am also grateful to Cornell's Southeast Asia Program. When I began studying at Cornell in 2007, I did not consider myself a Southeast Asianist, but the overwhelming intellectual richness of the SEAP community transformed my outlook as a linguist and as a scholar more generally. While I am thankful to SEAP for their grants and fellowships which have enabled me to do my research, I am even more grateful for the amazing sense of community they have provided during my time at Cornell.

I wish to thank my fellow graduate students who have provided intellectual engagement and friendship over the years and who are, for better or for worse, more knowledgeable about sesquisyllables than they ever hoped to be. In particular, I would like to thank my Southeast Asian linguist predecessors who have served as mentors to me, including Marc Brunelle, Pittayawat "Joe" Pittayaporn and John Phan. Thanks also to a number of the graduate students who preceded me and showed me the ropes, most particularly Johanna Brugman and Adam Cooper. I am also grateful to all the members of the Phonetics Lab, especially those who provided me a delicious chocolate cake on the day of my defense.

I am especially indebted to Cara DiGirolamo and Peggy "Hank" Renwick, who not only both took time out of their schedules to read and comment on parts of my dissertation but have

also provided friendship along the way. My thanks go out to Cara in particular for making me ginger syrup and offering plenty of hugs. In Hank, not only did I find an engaging colleague but I made a friend for life, always willing to encourage and always willing to help out when I need it most.

I am also appreciative of the city of Ithaca, which really is a lovely place to have spent the past six years. It has provided a wealth of non-academic stress-relieving activities, and I'm particularly thankful for the Viva softball team.

Lastly, I would like to thank Neil Ashton, who has shown me that it's possible to feel safely grounded and yet so absolutely free at the same time.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH	iii
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS.....	viii
LIST OF FIGURES	xiii
LIST OF TABLES.....	xvii
CHAPTER 1: AN INTRODUCTION TO THE SESQUISYLLABLE.....	1
1.1 The Sesquisyllable in Southeast Asia	1
1.2 Properties of Sesquisyllables	3
1.2.1 An Overview of Descriptions of the Sesquisyllable	4
1.2.2 The Essential Properties of the Sesquisyllable	5
1.2.2.1 Minor Syllables.....	6
1.2.2.2 Language Properties.....	8
1.2.2.3 Shared Properties	9
1.3 A Prosodic Model of the Sesquisyllable.....	12
1.3.1 The Prosodic Hierarchy and Violability	12
1.3.2 The Sesquisyllable as a Maximally Disyllabic Iamb.....	14
1.3.3 Phonological Alternatives to the Disyllabic Iamb	18
1.3.3.1 Maori and *Ft.....	20
1.3.3.2 Indonesian and Non-Foot(ə)	22
1.3.4 Summary	24

1.4 An Phonetic Model of the Minor Syllable	24
1.4.1 Epenthetic and Excrescent Mid Central Vowels	25
1.4.2 The Effect of Stress on Mid Central Vocalic Elements	31
1.4.3 Mid Central Vocalic Elements as Gestures	36
1.4.4 The Model	41
1.5 Evaluating Alternative Models of Sesquisyllables	43
1.5.1 Synchronic Models	44
1.5.2 Diachronic Models	48
1.5.3 Summary	54
1.6 The Predictions of the Model	55
CHAPTER 2: “MINOR SYLLABLES” ARE NOT SYLLABLES: PHONETIC EVIDENCE FROM KHMER	61
2.1 Introduction	61
2.2 Background of the Minor Syllable in Khmer	64
2.3 Goals and Hypotheses	70
2.4 Methods	74
2.4.1 Participants	74
2.4.2 Stimuli	74
2.4.3 Measurements	76
2.5 Results	77
2.5.1 Token Distribution	77
2.5.2 Non-sibilant C1s	80
2.5.3 Sibilant C1s	81
2.5.4 Sonority and Place of Articulation	85

2.5.5 Effects of Place on Underlap	89
2.5.6 Formant Analysis	91
2.6 Discussion	94
CHAPTER 3: THE PHONOLOGICAL REALITY OF MINOR SYLLABLES: PHONETIC EVIDENCE FROM BUNONG	
100	
3.1 Introduction.....	100
3.2 Bunong Phonology.....	101
3.2.1 General Background	101
3.2.1.1 Bunong Vowels.....	102
3.2.1.2 Bunong Consonants	103
3.2.2 Word Shapes and Minor Syllables.....	104
3.3 Goals and Hypotheses.....	108
3.4 Methods.....	111
3.4.1 Participants.....	111
3.4.2 Stimuli and Task	112
3.5 Results.....	114
3.5.1 <u>CCVC</u> Transitions and <u>C_ΔCVC</u> Vowels.....	114
3.5.1.1 <u>CCVC</u> Distributional Results.....	115
3.5.1.2 Non-sibilant C1s in <u>CCVC</u> Words.....	118
3.5.1.3 Sibilant C1s in <u>CCVC</u> Words	121
3.5.1.4 Comparison of <u>CCVC</u> Transitions with <u>C_ΔCVC</u> Vowels.....	122
3.5.2 Comparison of <u>C_ΔCVC</u> Vowels with <u>CVC</u> Vowels	126
3.5.3 Syllabic Nasal Minor Syllables.....	133
3.6 Discussion.....	133

3.6.1 Summary of Results.....	133
3.6.2 Minor Syllables as Gestures.....	134
3.6.3 Conclusion	139
CHAPTER 4: MULTIPLE MINOR SYLLABLES IN BURMESE	140
4.1 Introduction.....	140
4.2 Minor Syllables in Burmese.....	142
4.3 Hypotheses	149
4.4 Methods.....	151
4.4.1 Participants.....	151
4.4.2 Stimuli.....	151
4.4.3 Measurements	152
4.5 Results.....	153
4.5.1 Reduction from Compounding	153
4.5.2 Comparison across Minor Syllables	156
4.5.3 Summary of Results.....	160
4.6 Discussion.....	160
CHAPTER 5: CONCLUSION	164
5.1 Summary of Themes and Structure.....	164
5.2 Descriptive Contributions	165
5.3 Theoretical Contributions	168
5.3.1 Overview of Contributions	168
5.3.2 Prosody and Oscillation.....	170
5.3.2.1 Alternation as a Fundamental Property of Speech.....	171

5.3.2.2	Binarity as a Stipulation of Grammar	173
5.3.2.3	Oscillation as Motivation for Binarity and the Sesquisyllable	175
5.3.2.3.1	Introduction to Oscillators	176
5.3.2.3.2	Frequency Locking	178
5.3.2.3.3	Unarity	179
5.3.2.3.4	Functionally Motivated Constraints.....	181
5.3.3	Summary	183
5.4	Future Directions	184
5.4.1	Articulatory Experiments.....	184
5.4.2	Other Sesquisyllable Types.....	186
5.4.2.1	Super Extended Sesquisyllables	186
5.4.2.2	Sesquisyllables with Gradient Markedness	188
5.4.2.3	Trochaic Sesquisyllables.....	189
	REFERENCES	191

LIST OF FIGURES

Figure 1.1: Component parts of the sesquisyllable, where σ_{\min} = minor syllable and σ_{maj} = major syllable	5
Figure 1.2: Prosodic hierarchy	13
Figure 1.3: Prosodic model of the sesquisyllable	24
Figure 1.4: CVC monosyllable with a mid central vowel	32
Figure 1.5: Iamb ([dʌ. 'dʌd]) with two mid central vowels.....	33
Figure 1.6: Trochee (['dʌ.ʌd]) with two mid central vowels	33
Figure 1.7: Prosodic model of sesquisyllables.....	43
Figure 1.8: Scale of Thomas' (1992) sesquisyllable types	44
Figure 1.9: Proposed phonological structure of Thomas' (1992) sesquisyllable types.....	47
Figure 1.10: Compounding-Prefixation Cycle (Matisoff 2003)	48
Figure 1.11: Compounding-Prefixation Cycle overlaid with MCVE types in boxes.....	49
Figure 1.12: Path to monosyllables (Brunelle and Pittayaporn 2012, 424), with MCVE types	51
Figure 1.13: Schematized spectrographic predictions of MCVE types.....	58
Figure 2.1: [Khmer] Potential MCVE types	62
Figure 2.2: Khmer vowels (Registers separate and combined)	64
Figure 2.3: [Khmer] Underlying stressed wedge in [phʌŋ]	71
Figure 2.4: [Khmer] Underlying unstressed wedge in [khʌm. 'khat].....	71
Figure 2.5: [Khmer] Voiced underlap (a) and voiceless underlap (b)	72
Figure 2.6: [Khmer] Possible gestural configurations for sibilant-initial clusters, without underlap (a) and with underlap (b).....	72

Figure 2.7: Predictions about lexical wedge and the intrusive vocoid	73
Figure 2.8: [Khmer] Distributions of underlap durations (voiceless and voiced)	80
Figure 2.9: [Khmer] Center of gravity measurements in Hz for [tʃ]	83
Figure 2.10: [Khmer] Center of gravity measurements in Hz for [s]	83
Figure 2.11: [Khmer] Voiced underlap in /tʃbah/ 'be clear'	84
Figure 2.12: [Khmer] Possible voiceless underlap in /tʃbah/ 'be clear'.....	84
Figure 2.13: [Khmer] Center of Gravity in Hz for [tʃ] in four clusters	85
Figure 2.14: [Khmer] C(ɤ/ɤ̄)C durations by sonority type.....	86
Figure 2.15: [Khmer] Relationship between underlap and consonants in Oɤ̄S sequences	88
Figure 2.16: Degrees of overlap and underlap.....	90
Figure 2.17: [Khmer] Vowel plot of average values and standard deviations for the intrusive vocoid (◊), stressed wedge (◻) and unstressed wedge (◊) following alveolar (A) and labial (L) consonants	92
Figure 2.18: Possible gestural representations of MCVEs	93
Figure 2.19: [Khmer] Prosodic structure and MCVE types of monosyllables and disyllables	95
Figure 2.20: Gestural landmarks.....	96
Figure 2.21: Schema of G2 onset alignment with G1 C-center, with underlap (left) and without underlap (right)	98
Figure 3.1: Location of Bunong speakers	100
Figure 3.2: [Bunong] Proposed sesquisyllable type	101
Figure 3.3: Bunong vowels.....	103
Figure 3.4: Word-initial consonant sequences in Bunong.....	108
Figure 3.5: Bunong words and their proposed gestural representations: (a) [kʌl] CVC, (b) [klʌŋ] CCVC, (c-1) [mɤ̄ric] CCVC, (c-2) [mʌric] CVCVC	110
Figure 3.6: [Bunong] Percentage underlap (top) and percentage voiced underlap (bottom)....	116

Figure 3.7: [Bunong] Underlap in /kl/ cluster.....	117
Figure 3.8: [Bunong] Underlap in /ŋl/ cluster.....	117
Figure 3.9: [Bunong] Total duration (left) and residual total duration (right) for consonant sequences with no transition, voiced transition and voiceless transition.....	119
Figure 3.10: [Bunong] Transition duration (left) and residual transition duration (right) for consonant sequences with voiced and voiceless transitions.....	120
Figure 3.11: [Bunong] Center of Gravity in Hz for the 's' portion of [sr-], [sɜr-] and [sar] sequences	122
Figure 3.12: [Bunong] Vocalic transition in [ŋɜrʌŋ] 'hammock'.....	123
Figure 3.13: [Bunong] Purported minor syllable vowel in [lʌ.hat] 'tightly fitting'.....	123
Figure 3.14: [Bunong] Normalized durations for CCVC transitions and CʌCVC vowels.....	124
Figure 3.15: [Bunong] Formant values for word-initial CCVC vocalic transitions and CʌCVC vowels	124
Figure 3.16: [Bunong] Correlation between the duration of the CʌCVC vowel and F1	125
Figure 3.17: [Bunong] Purported minor syllable vowel in [kʌna:r] 'wing'.....	126
Figure 3.18: [Bunong] Underlying /ʌ/ in [kʌl] 'big turtle'	126
Figure 3.19: [Bunong] Underlying /a/ in [kal] 'chop'.....	127
Figure 3.20: [Bunong] Vowel durations for CaC, CʌC and CʌCVC vowels.....	127
Figure 3.21: [Bunong] Formant values for CʌCVC, CʌC and CaC vowels.....	128
Figure 3.22: [Bunong] F2 by Vowel Duration for CʌCVC vowels.....	128
Figure 3.23: [Bunong] Prenasalized stop in [ᵐpʌt].....	130
Figure 3.24: [Bunong] Purported syllabic nasal in [ŋ.lap].....	131
Figure 3.25: [Bunong] Simplex nasal onset in [nʌr].....	131
Figure 3.26: [Bunong] Duration distributions of nasals	131
Figure 3.27: [Bunong] Average normalized nasal intensity	132

Figure 3.28: Gestural representation of Bunong monosyllables and sesquisyllable	137
Figure 3.29: Derivation of epenthetic minor syllable vowels in Bunong	137
Figure 4.1: Burmese vowels (adapted from Green 2005).....	141
Figure 4.2: Possible footing of extended sesquisyllables in Burmese	146
Figure 4.3: [Burmese] Spectrogram of [càN] ‘floor’	154
Figure 4.4: [Burmese] Spectrogram of [cΛ.bò:] 'bug'	154
Figure 4.5: [Burmese] Vowel space for non-reduced [à] (circle) and reduced [Λ] (triangle).....	155
Figure 4.6: [Burmese] Durations of vowels in the final syllable of disyllables (V), the initial syllable of disyllables (Λ), the final syllable of /cεʔ.shíN/ (i) and the initial syllable of /cεʔ.shíN/ (ε)	156
Figure 4.7: [Burmese] Purported multiple minor syllables in [kΛ.lΛ.bjé:] 'India'.....	156
Figure 4.8: [Burmese] Vowel space for minor syllable nuclei in [kΛ.lΛ.bjé:]	157
Figure 4.9: [Burmese] Durations (ms) for vowels in [kΛ.lΛ.bjé:].....	157
Figure 4.10: [Burmese] Purported multiple minor syllables in [thΛ.mΛ.jé:] 'rice water'	158
Figure 4.11: [Burmese] Vowel space for minor syllable nuclei in [thΛ.mΛ.jé:].....	159
Figure 4.12: [Burmese] Durations (ms) for vowels in [thΛ.mΛ.jé:].....	159
Figure 4.13: [Burmese] Possible foot structures for [kΛ.lΛ.bjé:]	161
Figure 4.14: [Burmese] Possible foot structures for [thΛ.mΛ.jé:].....	161
Figure 4.15: The sesquisyllable on the pathway of language change.....	163
Figure 5.1: Traditional differences in sesquisyllables	167
Figure 5.2: Integrated model of sesquisyllables	169
Figure 5.3: CVCVV and CVVCV words in Finnish	177
Figure 5.4: 1:1 and 1:2 frequency-locking relations.....	179
Figure 5.5: Redundant prosodic structure in Vietnamese.....	180

LIST OF TABLES

Table 1.1: Notational conventions for mid central vocalic elements.....	28
Table 1.2: Preliminary model of phonological content in various word types.....	31
Table 1.3: Model of phonological content in and the role of stress on various word types	35
Table 1.4: Possible gestural representations of mid central vocalic elements (MCVEs).....	38
Table 1.5: Model of gestural representations associated with MCVE types	42
Table 1.6: Aheu correspondences to Vietnamese monosyllables (tones omitted).....	54
Table 2.1: Khmer consonants	63
Table 2.2: Khmer word types (adapted from Henderson 1952)	65
Table 2.3: Classes of word-initial consonant clusters in Khmer, according to Huffman 1972.....	67
Table 2.4: Khmer word-initial consonant clusters (Based on Huffman 1972).....	68
Table 2.5: Khmer wordlist	75
Table 2.6: [Khmer] Word-initial C1C2 combinations – Percentage of sequences with underlap.....	78
Table 2.7: [Khmer] Word-initial C1C2 combinations – Percentage of underlap tokens with voiced underlap.....	79
Table 2.8: [Khmer] Percentage of [t̪]- and [s]-initial tokens with underlap visible on a spectrogram.....	84
Table 2.9: [Khmer] CC durations (maximum, median and minimum) by sonority type	86
Table 2.10: [Khmer] Durations (ms) of C(ɛ/ɛ̃)CVC and CΛC.' CVC sequences for OO, OS and SO sequences	89
Table 2.11: [Khmer] Durations (ms) of C(ɛ/ɛ̃)CVC and CΛC.' CVC sequences for OO, OS and SO sequences	89

Table 2.12: [Khmer] Median formant values (Hz) for female and male productions of C ₁ VC, C ₁ VC and C ₁ VC, where C ₁ is alveolar and C ₂ is velar	93
Table 3.1: Bunong consonants	104
Table 3.2: Word-initial C ₁ C ₂ VC and C ₁ VC.C ₂ VC sequences in Bunong.....	107
Table 3.3: Bunong wordlist.....	113
Table 3.4: [Bunong] Prenasalized stops, purported minor syllable nasals and simplex nasal onsets	130
Table 4.1: Burmese stops (adapted from Green 2005)	140
Table 4.2: Representation of Burmese tone	141
Table 4.3: Properties of minor and major syllables in Burmese (Green 2005)	143
Table 4.4: Various prohibitions against sequences of unfooted syllables	149
Table 4.5: Burmese wordlist	152
Table 4.6: [Burmese] Mean durational differences (ms) between syllables by speaker for [k ₁ l ₁ .b ₁ jé:]. Significant differences across the columns in each row are indicated by different shading.	158
Table 4.7: [Burmese] Mean durational differences (ms) between syllables by speaker for [th ₁ m ₁ .jé:]. Significant differences across the columns in each row are indicated by different shading.	160

CHAPTER 1

AN INTRODUCTION TO THE SESQUISYLLABLE

1.1 THE SESQUISYLLABLE IN SOUTHEAST ASIA

The sesquisyllable is an unusual word type consisting of a final heavy syllable preceded by a prosodically light and phonologically reduced first syllable. It is generally considered an areal feature of Southeast Asia, particularly mainland Southeast Asia, including Myanmar, Laos, Thailand, Cambodia and Vietnam (Diffloth and Zide 1992, Enfield 2005), although many languages of southern China are reported to contain sesquisyllables as well (Matisoff 2003). However, given a revised understanding of the sesquisyllable, we can find evidence for it in India (Pnar, Turung), West Papua (Hatam) and even in places as distant from Southeast Asia as central Mexico (San Martín Itunyoso Trique). Despite the pervasive use of the sesquisyllable as a diagnostic for Southeast Asian type languages, there is little agreement as to what the sesquisyllable actually is. This dissertation investigates the phonological validity of this word type and proposes a more nuanced understanding of it based on well-established phonological categories.

There are two main research goals motivating this dissertation. The first is determining the nature of sesquisyllables both descriptively and structurally. Despite the strong tradition of language description within Southeast Asian linguistics, current analyses of sesquisyllables suffer from the lack of an integrated model which, in addition to being descriptively adequate, is also theoretically informed. And yet a number of these descriptive accounts have been used as evidence to support various phonological analyses, a fact which hints at a larger issue in the phonological literature more generally, i.e. the use of erroneous data to support theoretical

arguments (cf. de Lacy 2011). In contrast, I maintain that a prosodic analysis of the sesquisyllable and an articulatory description of the content of the minor syllable are complementary. Both are equally necessary and together sufficient for an understanding of this word type.

Second, because most languages that have been described as sesquisyllabic, i.e. containing sesquisyllabic word types, are relatively morphologically depleted, the studies presented here are necessarily phonetic ones, which could not rely on morpho-phonological alternations, for example, as a source of evidence. Therefore, throughout the work, I endeavor to determine what and how much phonetic evidence can tell us about phonological structure. I evaluate these questions and test my structural and articulatory model of the sesquisyllable by conducting three phonetic case studies of purported sesquisyllabic languages – Khmer, Bunong and Burmese.

To investigate these three languages, I take an analytic approach by applying phonological theory to empirical phonetic data. The results of these experiments are then used to refine the theory. Specifically, the set of purported sesquisyllables in Khmer are shown to crucially differ from purported sesquisyllables in Bunong and Burmese in that the former are monosyllables while the latter are disyllables. Combined, the results suggest that the term *sesquisyllable* is a cover term for a number of word types, a fact which I suggest obscures its contribution to phonological theory. Consequently, I propose a narrower definition of this word type.

This type of approach, although not novel, is important. An example of an effective model which addresses both of my stated goals is found in Cho's (2011, 367) work on the role of phonetics within the prosodic hierarchy. She argues that phonetic details play "linguistically

significant roles” in terms of boundary and prominence marking, contrast enhancement, speech comprehension and lexical processing. She also refers to a number of cognitive and phonological models of speech that necessarily integrate phonetic data. In the same vein, I argue that each of these components – descriptive/analytic and phonetic/phonological – is important in a descriptively accurate and linguistically informed model of the sesquisyllable.

In Section 1.2 of this chapter, I propose that the descriptive cross-linguistic properties of sesquisyllables can be characterized by their loci of prosodic prominence and phonological reduction. I argue for both a structural model of the sesquisyllable as a maximally disyllabic iamb in Section 1.3 and an articulatory model of the minor syllable which holds that minor syllable nuclei must have an associated tongue gesture in Section 1.4. Only by maintaining an integrated phonological approach, i.e. structural/prosodic and articulatory, can this word type be adequately explained. After this framework is established, I use it to evaluate other synchronic and diachronic analyses of the sesquisyllable in Section 1.5. I end the chapter in Section 1.6 with the proposal of an empirically testable model and the rationale for the case studies that follow in the subsequent three chapters.

1.2 PROPERTIES OF SESQUISYLLABLES

The sesquisyllable has been described in a number of ways, which has led to ambiguity in what the term actually means. This section first provides an overview of properties that cannot be used to cross-linguistically characterize the sesquisyllable and subsequently posits a set of properties which can. These properties are then taken as the basis for the structural/prosodic model.

1.2.1 *An Overview of Descriptions of the Sesquisyllable*

The sesquisyllable was first defined by Matisoff (1973, 86) in a discussion of the tonal features of Austroasiatic languages. He states, “Proto-[Austro-Asiatic] had what one might call a ‘sesquisyllabic’ structure, with morphemes that were ‘a syllable and a half’ in length. That is, the prevocalic consonant was often preceded by a ‘pre-initial’ consonant... It is perhaps no accident that these ‘halfway tonal’ [Mon-Khmer] languages also have a syllabic structure intermediate between the truly monosyllabic [Sino-Tibetan] and truly polysyllabic [Austronesian] types.” Although this was the first explicit use of *sesquisyllable*, the component parts of the sesquisyllable – the major syllable and the minor syllable – were discussed twenty years earlier by Henderson (1952) in her work on Khmer.

Henderson (1952) described Khmer as having four types of syllables: simple monosyllables, extended monosyllables, minor disyllables and major disyllables (1.1a-d). The disyllabic word types are named after their non-final syllables, which Henderson refers to as *minor* and *major*, respectively. She notes, “Disyllables of this type [i.e. minor disyllables] are intermediate structurally between the extended monosyllable and the full, or major disyllable” (150).

(1.1) Khmer

- | | | | |
|----|------------------------|----------------------|-----------------|
| a) | Simple monosyllable: | [kait] | ‘to wake up’ |
| b) | Extended monosyllable: | [p ^h dek] | ‘to put to bed’ |
| c) | Minor disyllable: | [səm.naɪm] | ‘humidity |
| d) | Major disyllable: | [kait.laiŋ] | ‘to grow’ |

(Henderson 1952, transcription mine)

Over time, the sesquisyllable has come to be understood as a word-final “major” syllable, which in all respects is a canonical and phonotactically well-formed syllable, and which is preceded by a “minor” syllable, sometimes called a pre-syllable. Michaud (2012, 2) provides a typical contemporary definition of the minor syllable as “a simple consonant... plus an optional nucleus, V: either a vowel, or a sonorant (nasal or liquid) serving as nucleus. In the Austroasiatic domain, the most frequently encountered situation is one in which there can be no vowel contrast in the presyllable: the nucleus consists simply in a *schwa*, a noncontrastive, optional vowel.” Similarly, Diffloth and Zide (1992, 3) state that in contrast to the major syllable in which final-syllable stress and lack of suffixation converge to make it the “richest and most stable part of the word,” the minor syllable has a poor consonant inventory as well as a “vocalism”, which they suggest reduces to a single possible vowel, i.e. [ə], in most cases. Two instantiations of this structure are given in Figure 1.1. Although we will see that this characterization does not capture the complexity of how sesquisyllables have been described, I take this as a starting point.

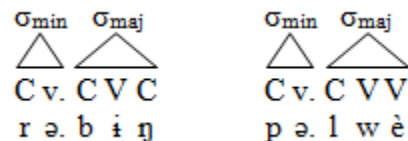


Figure 1.1: Component parts of the sesquisyllable, where σ_{\min} = minor syllable and σ_{maj} = major syllable

1.2.2 *The Essential Properties of the Sesquisyllable*

What stands out in the above descriptions of the sesquisyllable are the differences between the minor and major syllables. I propose that the cross-linguistic descriptive properties that are common to all sesquisyllables can be expressed as a small set of characteristics, which can then be used as the basis for a structural model of the sesquisyllable. As a starting point for this analysis, we can look to a similar scenario, in which Hyman (2009) addresses the issue of the validity of so-called *pitch-accent* languages. Just like sesquisyllables are claimed to fall

between monosyllables and disyllables, pitch-accent languages are often understood to lie on continuum between stress and tone languages. In deconstructing the notion of pitch-accent languages, Hyman (2009) poses two questions: Are there any characteristics which apply to all pitch-accent languages, and are any of these characteristics unique to pitch-accent languages? Similarly, in determining whether the sesquisyllable is a word type unto itself, these two questions must be addressed. Is there a set of properties common to all sesquisyllabic languages? Are any of these properties exclusive to sesquisyllabic languages? I suggest that before we can determine which properties are shared among sesquisyllables cross-linguistically, we must first take a step back and determine which are not.

1.2.2.1 Minor Syllables

While descriptions of the major syllable, i.e. the word-final syllable, are generally consistent, much of the definitional variation in previous work on sesquisyllables relates to the minor syllable. As Brunelle and Pittayaporn (2012, 414) note, “While many authors take sesquisyllables to be any disyllabic words with a reduced number of contrasts in initial syllables (Larish 1999; Thurgood 1999), others take the more restrictive position that the syllabicity of the minor syllable is carried by a neutral vowel or a syllabic consonant (Diffloth 1976: 232; Svantesson 1983: 27)”. The examples in (1.2) display the wide array of minor syllables in words which have been described as sesquisyllabic: (1.2a) and (1.2c) have schwa vowels, while (1.2b) has a non-schwa vowel, and (1.2d) is a syllabic nasal. (1.2c) also contains a coda while the other examples do not.

(1.2)

a)	[rə.biŋ]	‘gourd’	Bunong	(Butler 2012)
b)	[ti.j̥k]	‘to point’	Jahai	(Burenhult 2001)
c)	[tər.pah]	‘to slap each other’	Pacoh	(Watson 1964)
d)	[m.ləm]	‘one’	Stieng	(Haupers 1969)

These examples suggest a number of ways minor syllables cannot be differentiated from other word types. First, many languages allow codas in the minor syllable, although some allow only nasals while others include other liquids as well. In the examples below, we note that Burmese, the national language of Myanmar, does not allow any codas in minor syllables (1.3a). However, in So, a Mon-Khmer language of Thailand (1.3b), nasal codas are allowed in minor syllables.

(1.3)

a)	No codas: Burmese (Green 2005)			
	[təə.bó]	<	[təáN + pó]	‘bed-bug’
	[ŋə.lə]	<	[ŋi? + lə]	‘two months’
b)	Codas: So (Migliazza 2003)			
	[baŋ.pɛc]			‘to work sorcery’
	[sam.lo:ŋ]			‘slipknot’

In addition, cross language descriptions of minor syllables vary in the types of vowels allowed. These might include [ə] exclusively or a small set of peripheral vowels which may or may not be subject to vowel harmony. There can even be variability in the phonetic realization of a minor syllable within one language. For example in Turung (a Tibeto-Burman language spoken in India) the word /ljung/ ‘finger’ can be realized in three different ways, i.e. with no intervening

vowel (1.4a), with a schwa (1.4b) or with a vowel that is either harmonized with that of the main syllable or the C2 glide, if applicable (1.4c).

(1.4) Turung

- | | | | |
|----|----------|----------|-------------------------|
| a) | [ljung] | ‘finger’ | |
| b) | [ləjung] | ‘finger’ | |
| c) | [lijung] | ‘finger’ | (Morey 2005, Example 6) |

1.2.2.2 *Language Properties*

There are two additional properties that do not address the minor syllable in particular which must be taken into account of a description of the sesquisyllable. These are the type of prosodic prominence found in a language, whether stress or tone, and the notion of a “sesquisyllabic language” itself. In the first place, sesquisyllables are found in both stress and tone languages (1.5), indicating that neither system is required for or excludes the presence of sesquisyllables. The examples below are from Cua, a Mon-Khmer language spoken in Vietnam, and Thai, a Tai-Kadai language and the national language of Thailand.

(1.5)

- | | | |
|----|---------------------------|----------------|
| a) | Stress: Cua (Maier 1969) | |
| | [ka. 'laat] | ‘hunk of meat’ |
| | [ta. 'rʌk] | ‘unison call’ |
| b) | Tone: Thai (Bennett 1995) | |
| | [la.mút] | ‘sp. fruit’ |
| | [sa.nùk] | ‘fun’ |

Finally, the notion of a “sesquisyllabic language” (Diffloth and Zide 1992) does not adequately explain the cross-linguistic properties of the sesquisyllable as a word type.

Sesquisyllables can be found in languages in which they are the maximal word shape, i.e. languages which contain no (or almost no) words larger than sesquisyllables, as well as in languages that have words far larger than sesquisyllables. Examples of the latter are given in (1.6).

(1.6)

a) Thai (Bennett 1995)

[ma.nút.sa.ja.c^hon] ‘human’

[sǎa.thāa.ra.náʔ] ‘public, common’

b) Khmer (Henderson 1952, transcription mine)

[bɔ̄.raw.hit] ‘family priest’

[kɔ̄.ru.na] ‘pity’

1.2.2.3 *Shared Properties*

Despite these differences, however, there is a set of properties shared cross-linguistically by all so-called sesquisyllables. Bennett (1995) suggests minor and major syllables should be understood as a convergence of four properties. These include a more restricted inventory in minor syllables than in major syllables, prosodically lighter minor syllables than major syllables, tonal contrasts on major syllables but not minor syllables, and the necessity of major (but not minor) syllables for the phonotactic well-formedness of a word. With some slight modifications, I suggest that this description is largely correct.

First, prosodic prominence (i.e. stress or tone) must be word-final. Although some alternative definitions of the sesquisyllable suggest this may not be the case (cf. Bennett 1995 on Chinese), I take the position that if this property does not hold, a word cannot be considered a sesquisyllable. One potential counter example to this property is found in Northern Kammu, a

Mon-Khmer language spoken in Laos, in which tonal contrasts are present on penultimate syllables. However, this contrast is a marginal one. As Svantesson and Karlsson (2004, 2) state, “The functional load of the minor syllable tone contrast is low.” There are only ten documented minimal pairs in which minor syllable tone is contrastive. However, many of the words in these pairs are formed by infixation, causing the tone which was previously contrastive on the stem to shift to the minor syllable (1.7). Because the tonal contrast on the minor syllable is the result of a morphological process, an argument could be made that the locus of the contrast is still the major syllable.

(1.7) N. Kammu

- | | | | | |
|----|-----------|------------------|-------|-----------|
| a) | [kóm.nòh] | ‘cutting board’ | < kóh | ‘to cut’ |
| b) | [kəm.nòh] | ‘weeding period’ | < kòh | ‘to weed’ |

(Svantesson and Karlsson 2004)

However, there are other words in Northern Kammu, although not minimal pairs, in which the tone on the minor syllable is specified by its own onset and is clearly not derived from the major syllable. For example, if the minor syllable onset is a liquid, its tone must be low, regardless of the tone on the major syllable (1.8). There is therefore no possibility that the minor syllable tone is derived from the major syllable. Yet although the minor syllable bears its own tone, because the tone is predictable from the onset consonant, tone on minor syllables is not contrastive.

(1.8) N. Kammu

- | | | |
|-----------|----------|--------------------------------|
| [rə.háaŋ] | ‘bamboo’ | (Svantesson and Karlsson 2004) |
|-----------|----------|--------------------------------|

The second property of sesquisyllables is that non-final syllables are phonologically reduced. This includes segmental properties, syllable shape and weight. Bennett (1995) separates

these into distinct characteristics, but because these are all a type of reduction, I group them together as one property. Although the types of reduction vary across languages, sesquisyllables always display a smaller inventory of possible segments in the minor syllable than in the major syllable. For example, in Bunong, implosives are allowed in major syllables but not in minor syllables. The shape of the minor syllable is always reduced. Even if codas are allowed, for example, their inventory is restricted.

Finally, the minor syllable is always light, which leads to Bennett's (1995) final criterion that the minor syllable on its own is insufficient for word minimality requirements. A major syllable is always necessary for phonological well-formedness while a minor syllable is not. Because this is both implied by the property of phonological reduction and is dependent on language specific requirements on word minimality, I do not list it as a property unto itself. These properties are summarized in (1.9).

(1.9) Properties of sesquisyllables

a) Prosodic prominence is word-final

b) Non-final syllables are phonologically reduced¹

i) Non-final syllables have a reduced segmental inventory

ii) Non-final syllables have a reduced syllable shape

iii) Non-final syllables are light

→ Non-final syllables do not constitute well-formed prosodic words on their own

¹ The one possible exception to this descriptive property is that of syllabic nasal minor syllables since consonantal nuclei are more marked than vocalic nuclei (See Zec (1988, 1995) for the relationship between sonority and prosodic structure). Syllabic nasals are addressed in Chapter 3.

1.3 A PROSODIC MODEL OF THE SESQUISYLLABLE

Given the descriptions laid out in the section above, the goal of this section is to propose a prosodic structural model of the sesquisyllable as a word which is maximally a disyllabic iamb. First, I provide what I take to be a standard account of the prosodic hierarchy, on which this proposal is based. Then I show that the category of sesquisyllable is not co-extensive with the disyllabic iamb but is instead unique because of its entailed property of maximality. Finally, I address some residual issues that this proposal creates and then argue against a number of phonological alternatives.

1.3.1 *The Prosodic Hierarchy and Violability*

The prosodic hierarchy is the set of structural constituents out of which phonological structures are constructed in all languages. A phonological structural description in terms of the prosodic hierarchy is a rooted directed acyclic graph whose leaves are phonological segments and whose nonterminals are elements of the prosodic hierarchy. In its most canonical form, it is built with constituents that get progressively larger beginning at the mora and proceeding up through the utterance. My account here is mostly concerned only with what Inkelas (1989) refers to as the *metrical hierarchy*, comprising all levels at and below the foot, as well as how these lower-level constituents interact with the prosodic word.

First, moras are motivated by the need to account for syllable weight, or the difference between light and heavy syllables (Hayes 1989). The role of the mora as it relates to syllable weight is addressed by Prince (1980), Hyman (1985) and Zec (1988), among others. Next, the syllable is useful largely because it can account for a number of phenomena which would otherwise require constructing overly powerful segment-based rules (cf. Kahn 1976, Kiparsky 1979, Zec 2007, *inter alia*). Finally, feet are necessary for the placement of metrical stress. While

syllables are actually the stress-bearing units, sub-word level minimal bracketed groups, i.e. feet, determine where stress falls. Feet were first formally proposed by Hayes (1985), based on work by Prince (1976) and Selkirk (1980). This work was preceded and motivated by Liberman and Prince (1977) who claimed that stress should be understood as a matter of relative prominence between syllables. Figure 1.2 below is an example of the prosodic hierarchy as it relates to English. Note that a number of these constituents may be considered degenerate in a standard metrical theory (cf. Hayes 1995).

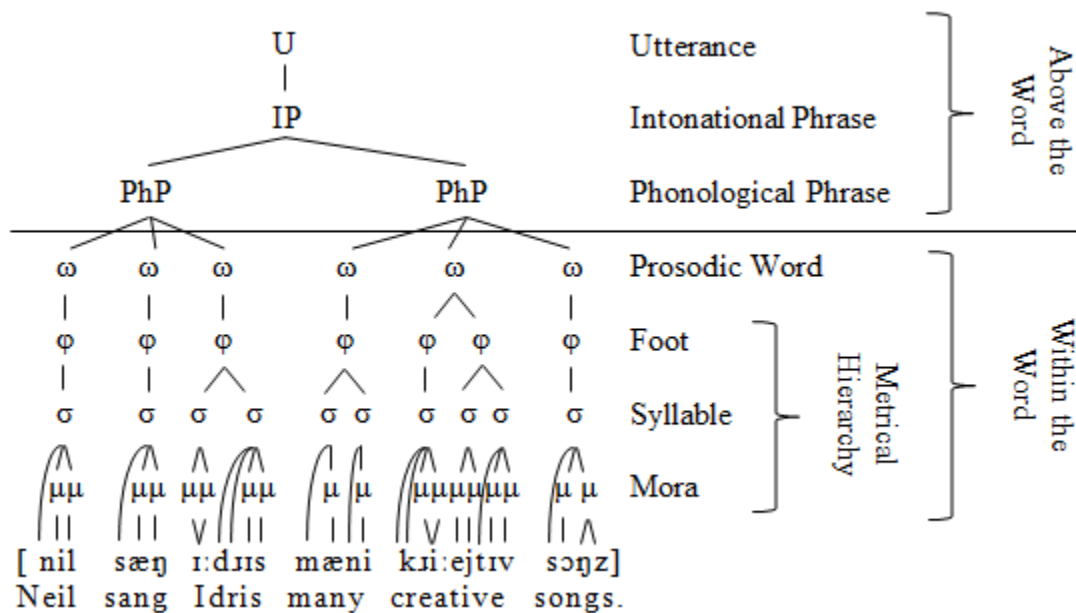


Figure 1.2: Prosodic hierarchy

Phonological accounts of sesquisyllables based on the prosodic hierarchy must also take into consideration the principles which govern the relationships between its constituents. In its original formulation, the prosodic hierarchy was said to be governed by four inviolable principles which together compose the Strict Layer Hypothesis (Selkirk 1995, 6):

- (i) Layeredness: No C^i dominates a C^j , $j > i$,
e.g. "No σ dominates a Ft."

(ii) Headedness: Any C^i must dominate a C^{i-1} (except if $C^i = \sigma$),

e.g. "A PWd must dominate a Ft."

(iii) Exhaustivity: No C^i immediately dominates a constituent C^j , $j < i-1$,

e.g. "No PWd immediately dominates a σ ."

(iv) Nonrecursivity: No C^i dominates C^j , $j = i$,

e.g. "No Ft dominates a Ft."

However, since the development of Optimality Theory (OT) (Prince and Smolensky 1993, 2004; McCarthy and Prince 1993b), constraints on grammar have been understood to be violable. In fact, a number of studies have shown that Exhaustivity and Nonrecursivity do not universally hold (Inkelas 1989, McCarthy and Prince 1993a, among many others; cf. also Selkirk 1995 and references therein). In addition, Schiering et al. (2010) suggest that at least for Vietnamese, Headedness may not hold either, in that redundant or non-branching constituents may not be represented in the grammar, and as a result, a constituent C^i may immediately dominate a C^{i-2} , C^{i-3} and so forth. The one property that does seem to hold without exception is Layeredness. Yet despite the violability of these principles, relationships between constituents display quite remarkable consistency cross-linguistically. Additional explanations for this consistency are explored in Chapter 5.

1.3.2 The Sesquisyllable as a Maximally Disyllabic Iamb

Having laid out the cross-linguistic descriptive properties by which sesquisyllables can be characterized, we are now in a position to address the second question posed in Section 1.2.2, i.e. are any of the definitional properties of sesquisyllables exclusive to sesquisyllables? Given that the properties I established are word-final prosodic prominence and phonological reduction of the penultimate syllable, the answer is clearly no. There is nothing about the properties laid out

here which differentiates sesquisyllables from words that are maximally disyllabic iambs that exhibit a weak-strong pattern. This is not to say that every iamb is a sesquisyllable since heavy monosyllables may be well-formed iambs and even light monosyllables may be considered iambs, albeit degenerate iambs (Hayes 1995). In other words, all sesquisyllables are iambs but not all iambs are sesquisyllables.

However, the terms *disyllabic iamb* and *sesquisyllable* do not define one and the same thing. A priori, disyllabic iambs can be iterative, and the uniqueness of the sesquisyllable is that it characterizes a word which is maximally a disyllabic iamb so that the word is coextensive with the prosodic foot. This raises a potential problem for my analysis in that the basis for the distinction between trochees and iambs is iterativity, or the repetition of stress patterns, i.e. feet, within prosodic words. However, in shorter words, like disyllables and monosyllables, diagnosing stress patterns is far more difficult. And when languages are maximally di- or monosyllabic, assigning a foot type to the language in its entirety becomes much more difficult. Although stress placement can be indicative of foot type, other confounding factors like extrametricality and edge prominence also play a role. In other words, without prosodic iterativity, we really cannot say whether a language is trochaic or iambic. This ambiguity is particularly problematic for many so-called sesquisyllabic languages in which words are maximally disyllabic.

Indeed, while Hayes (1995) presents the distinction between trochees and iambs as a rather concrete one as proposed in the Iambic/Trochaic Law, a number of alternative proposals have been considered. The Iambic/Trochaic Law states that elements contrasting in intensity naturally form groupings with initial prominence (trochees) while elements contrasting in duration naturally form groupings with final prominence (iambs). However, Kager (1993, 381)

argues that the iambic-trochaic law is unnecessary as a fundamental property of language but instead the patterns that emerge from it can be accounted for by “asymmetries with respect to lengthening and shortening, and an asymmetry with respect to directionality of iambic parsing.” Yet another striking alternative is put forth by van de Vijver (1998) who argues against underlying iambic feet altogether, casting them instead as the emergent result of constraint interactions.

Ambiguity in the classification of systems is found in a number of languages in which iterativity can be used as a diagnostic, i.e. words are longer than just disyllables. For example, Hayes (1995) notes that iambic lengthening is avoided in several languages which are otherwise considered iambic due to their stress placement (e.g. Hixkaryana, Choctaw, Yidiñ, etc.). This means that maximally disyllabic words would actually appear to be durationally even, i.e. trochaic. This somewhat false dichotomy of languages as either iambic or trochaic is highlighted in Cohn and McCarthy’s (1998) discussion of parallelism in Indonesian stress. They define parallelism as “the idea that the constraint hierarchy evaluates candidates that are fully-formed output representations, with the effects of various phonological and morphological processes all under consideration at once” (55). At a superficial level, this is clearly problematic given that, as noted in Section 1.3.1, since the advent of Optimality Theory, constraints are understood to be violable. In other words, there is no reason to expect that a language is strictly trochaic or iambic, but instead we should expect that mixed stressed systems are far more common than they have been claimed to be.

The (near) exclusivity of iambicity in languages whose word types are maximally disyllabic is striking. They are characterized by the descriptive properties given above, in which initial syllables are more phonologically restricted in terms of segmental inventory, syllable

shape and weight than are final syllables. Initial syllables are also shorter in duration than final syllables, which suggests that these systems are iambic (although there is the confounding factor of final lengthening). Indeed this pervasive tendency is what motivates the usefulness of the sesquisyllable as a word type. In other words, it captures the striking correlation between disyllabic maximality and the placement of prosodic prominence at the word level.

Crucially, my definition does not limit the term *sesquisyllable* to refer to words whose first syllable may only contain a mid central vowel. As shown above, many languages exhibit other vowels in minor syllables. This is addressed in a study on Moken (Pittayaporn 2005), in which he notes that previous research separates the categories of “presyllable”, i.e. a weak initial syllable with a schwa vowel, and “minor syllable”, i.e. a weak initial syllable with a reduced set of non-schwa vowels. The author conflates them into a single category because of a lack of phonological evidence that they should be treated separately. Until such evidence is found, there is no reason to treat iambs whose initial syllables contain schwas differently than iambs whose initial syllables contain a reduced set of vowels. If, however, for historical or typological purposes it becomes prudent to definitionally restrict the minor syllable to contain only the schwa vowel, we could then consider the sesquisyllable to be a sub-type of the disyllabic iamb but an iamb all the same. To my knowledge, though, no such evidence yet exists, and an investigation into this issue is beyond the scope of this dissertation.

From this arise two residual but important issues that will be addressed again later. First, historical models show that most if not all maximally sesquisyllabic languages represent a stage of language change in which apparently trochaic languages are moving toward monosyllabicization. However, although diachronically this process of change suggests that so-called sesquisyllabic languages are unstable, they are, at least for some generations of speakers,

synchronically stable, so they should not be dismissed simply as change in progress. This issue is revisited in Section 1.5.2. Second, the ambiguity inherent in metrical systems presents the problem of how to understand the stress patterns in languages in which either prosodic words or prosodic feet are strictly monosyllabic. In Chapter 4 I investigate whether phonetic evidence can provide insight on this point by looking at compounds in Burmese.

Given my definition that sesquisyllables are maximally disyllabic, we must address the fact that although most descriptions of the sesquisyllable refer to one major and one minor syllable only, several (generally Sino-Tibetan) languages have been described as having more than one minor syllable (1.10). Morey (2005, 158) calls them “double sesquisyllables,” and I refer to them as *extended sesquisyllables*.

(1.10)

- | | | | | |
|----|--------------|---------|---------|---------------|
| a) | [kə.lə.bjè:] | ‘India’ | Burmese | (Green 2005) |
| b) | [rə.kər.taʔ] | ‘loom’ | Palaung | (Shorto 1960) |

Diffloth and Zide (1992) state that “most” languages only allow one minor syllable to precede the major syllable, suggesting that others allow multiple. In addition, Matisoff (2003, 149) discusses a process of “reprefixation” in which a word is prefixed twice. He states, “Sometimes each of the two prefixes has schwa vocalism, so that a form is ‘doubly sesquisyllabic’.” Although far less work has been done on these word shapes than canonical sesquisyllables, I suggest they still fit into the proposal I have laid out thus far. These will be addressed further in Chapter 4 with an in-depth discussion of Burmese.

1.3.3 Phonological Alternatives to the Disyllabic Iamb

In addition to the prosodic/structural model laid out above, sesquisyllables and minor syllables have been (or could be) accounted for in a number of other ways. In most previous

phonological analyses, sesquisyllables have been used as evidence to argue for various phonological processes. However, as these analyses are based on once widely held theoretical assumptions that are no longer maintained, they do not contribute much to a more accurate understanding of the sesquisyllable. I review them briefly here before moving on to two phonological analyses which might serve as alternatives to my proposal.

Itô (1989) uses minor syllables in Temiar, described by Benjamin (1976) and Diffloth (1976a), to argue for a theory of epenthesis as a prosodic phenomenon based on syllabification instead of a process based on skeletal slot insertion rules. Shaw (1994) uses minor syllable data also from Temiar as well as Semai (Diffloth 1976a,b) and Kammu (Svantesson 1983) in order to argue for a template-based theory of epenthesis. Drawing heavily on Sloan (1988) who tacitly assumes that minor syllables comprise one or two consonants but no vowel, Shaw (1994) makes a more substantive claim, suggesting that minor syllable moras are degenerate. The Temiar and Semai data are also incorporated into OT analyses of reduplication by Gafos (1998) and Hendricks (2001), respectively. Finally, Yap (2009) argues against Shaw's (1994) proposal directly suggesting that syllabification in Temiar is not exhaustive. In doing so, she claims that what have been considered epenthetic segments in Temiar are actually excrescent transition states. In summary, while minor syllables have been used as evidence for various phonological frameworks, most of these arguments provide little new in the way of descriptive clarification. The two possible exceptions are Shaw's (1994) claims about the moraic structure of minor syllables and Yap's (1999) claims regarding the excrescent, i.e. non-phonological, nature of minor syllable nuclei in Temiar.

In addition to these mentions of minor syllables, there are two possible alternatives which might usefully address the structure of sesquisyllables. Although neither of these two proposals

deals with minor syllables directly, they do address issues relevant to the structure of minor syllables and should be considered as alternatives to the structural analysis I have suggested thus far. In both, minor syllable vowels are assumed to be fully well-formed phonologically, but they are targeted for reduction due to lack of prosodic prominence. Similar proposals have been made to directly account for minor syllables. For example, Green (1995, 2005) has proposed that minor syllables in Burmese are a result of non-exhaustive parsing (See Chapter 4 for more on this proposal in particular). In the following proposals, de Lacy (2004) explores a lack of footing and Cohn and McCarthy (1998) analyze the relationship between metrical projections and schwa.

1.3.3.1 *Maori and *FT-*

Although constraints on minimality are pervasive, particularly at the level of the prosodic word (Prince 1980, Crowhurst 1991, Itô et al. 1996, among many others), considerations of constraints on maximum word size are considerably rarer. However, they are a necessary means of understanding sesquisyllables. De Lacy (2004) proposes a constraint *FT- which limits the size of the prosodic word to maximally one prosodic foot by prohibiting non-head feet (1.11). This constraint is necessary for his analysis of Maori, an Austronesian language of New Zealand. A non-head foot is any foot without primary stress; therefore, this constraint effectively limits a prosodic word to at most one prosodic foot. For example, words in Maori may have up to four moras but are maximally one prosodic foot, as seen in (1.12).

(1.11)

*FT-: Incur a violation for every non-head foot. (de Lacy 2004, 3)

(1.12)

- a) ta('mai)ti ‘child’
b) ('ko:re)ro ‘tie’ (de Lacy 2004)

A constraint such as this often results in unparsed material, as can be seen in (1.12). In neither word is it possible to parse the remaining material into feet because those feet would be degenerate. This leaves three possibilities: (i) surface material is left unparsed, (ii) non-footed material is deleted or (iii) a combination of these two occurs. In the case of Maori, the result is that non-head feet are partially deleted and partially unparsed (1.13b). Some unparsed material is allowed so that MAX violations are minimized.

(1.13)

- a) /karaŋata/ → *('ka.ra).(ŋa.ta)
b) /karaŋata/ → ('ka.ra).ŋa (de Lacy 2004)

Similarly, the idea that sesquisyllables are composed of one monosyllabic footed syllable preceded by one unfooted syllable is worth considering. If sesquisyllables are maximally monopodal and feet are maximally monosyllabic, sesquisyllabic languages would align well with monosyllabic – and therefore implicitly monopodal – languages like Vietnamese, for example, which are known to have been sesquisyllabic at one point (Ferlus 1982). In addition, a lack of footing could help explain the phonological reduction in terms of weight and segmental inventory which I take to be a definitional property of sesquisyllables. If this is correct, the Bunong word /rɔbɪŋ/ ‘gourd’ would necessarily be parsed as (1.14b) instead of (1.14a), contra my analysis of sesquisyllables as disyllabic iambs.

(1.14)

- a) Light-Heavy iamb: [(rʌ. 'bɪŋ)]
- b) Maximally monosyllabic foot: [rʌ. '(bɪŋ)]

Without additional phonological evidence, there is no way to determine which structure is correct. However, there are a number of reasons to prefer the definition of a sesquisyllable as a disyllabic iamb over the monosyllabic analysis. The category *disyllabic iamb* entails a number of properties which are descriptively true of sesquisyllabic words. These are (i) that sesquisyllables tend to have exactly two syllables², (ii) that the word-final syllable is prosodically prominent and (iii) that the first syllable is phonologically reduced. In contrast, if we assume the *FT- analysis, each of these properties has to be stipulated separately. Therefore, without direct evidence to motivate the more marked word type that the *FT- constraint would create, I assume that a less marked word structure, i.e. the disyllabic iamb, is preferable.

1.3.3.2 Indonesian and NON-FOOT(ə)

If we limit ourselves to considering those languages which only allow schwa-like vowels in the nucleus of the minor syllable (e.g. Bunong and Burmese), another potential foot-based analysis is found in studies of word minimality in Indonesian (Cohn and McCarthy 1998, Cohn 2005). Schwa in Indonesian does not usually count for stress placement, nor does it generally count toward minimality constraints on word size. However, there are exceptions in which schwa clearly counts toward a disyllabic word minimum although it cannot be stressed. To account for these differences, Cohn and McCarthy (1998) propose two different constraints on schwa, given as (1.15a) and (1.15b) here.

² I address the issue of sesquisyllables larger than two syllables in Chapter 4.

(1.15)

- a) NON-FOOT(ə): Schwa-headed syllables have no metrical projection.
- b) NON-HEAD(ə): Stressed [ə] is prohibited.

Either of these constraints might be applicable to minor syllables, depending on which footing analysis is correct. The NON-FOOT(ə) constraint would further motivate an account like de Lacy's (2004). On the other hand, the NON-HEAD(ə) constraint fits in well with a simpler iambic analysis of sesquisyllables. However, both of these constraints fail to account for languages which allow vowels other than schwa in minor syllable nuclei. Furthermore, in an iambic language, we expect non-stressed, non-word-final syllables to be reduced. Such reduction is motivated by the iambic grouping of the syllables, so whereas a constraint preventing stress on schwa is necessary to account for Indonesian, it is redundant in an analysis of sesquisyllables.

Sesquisyllabic languages allow schwa vowels in major syllables. Bunong and Khmer, for example, both have underlying, phonological schwa-like vowels. It has long been the case in the Americanist transcription tradition to differentiate between unstressed [ə] and stressed [ʌ]. In such a tradition, it is certainly possible to claim that a language like Bunong allows [ʌ] in stressed syllables and [ə] in unstressed syllables, but this characterization is uninformative. Instead, I suggest that if the schwa-like vocalism is indeed not phonologically present, it cannot bear stress because it is not part of the phonological representation. If, on the other hand, we assume that all schwa vowels have the same phonological status, i.e. that they are phonological, then although a constraint like NON-HEAD(ə) is motivated for a language like Indonesian, it cannot be the case that this constraint is active in a language like Bunong.

In summary, while a maximality-based analysis which often results in unparsed surface material can potentially capture the descriptive facts about sesquisyllables, it also requires

additional constraints on the grammar. Furthermore, constraint-based approaches that target the phonological constituency of the minor syllable are problematic in a number of ways and fail to account for the large array of possible minor syllable types. Therefore, the most useful and accurate account of the sesquisyllable is one in which sesquisyllables are considered disyllabic iambs with minor syllables as the reduced first syllables of those iambs.

1.3.4 Summary

Now I present the prosodic model in Figure 1.3, which is revised from Figure 1.1. I have defined sesquisyllables thus far as words which are maximally disyllabic and monopodal. Due to both lack of morphological and phonological evidence as well as a lack of iterativity because of limited word size, this structural model is based on the descriptive characteristics laid out in the previous section. In the next section, I address the minor syllable more directly and provide the other half of the model which is laid out in gestural terms.

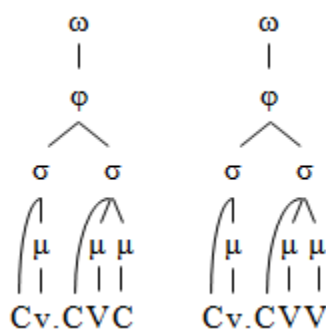


Figure 1.3: Prosodic model of the sesquisyllable

1.4 A PHONETIC MODEL OF THE MINOR SYLLABLE

After having explored the descriptive generalizations of the sesquisyllable and having proposed a phonological model based on structural prosodic properties, I now turn to the phonetic attributes – both acoustic and articulatory – of sesquisyllables with a particular focus on minor syllables. Although minor syllable nuclei may contain vowels other than schwa, those

with schwa-like (or mid central) vocalic material are the most common yet least understood cases. For this reason, I will focus on them for the remainder of the dissertation. In this section, I lay out the differences between excrescent and phonological mid central vowels with a focus on epenthesis, and I show how this has led to ambiguity in descriptions of minor syllables (Section 1.4.1). Then in Section 1.4.2, I discuss the effects of stress on these vowels, demonstrating how prosodic structure plays a role in my conceptualization of the sesquisyllable. In Section 1.4.3, I present an articulatory model of the minor syllable, and in Section 1.4.4, I bring together the roles of stress and articulation to present an integrated model which draws upon both phonological status (i.e. excrescent, epenthetic or underlying) and stress placement.

1.4.1 Epenthetic and Excrescent Mid Central Vowels

Excrescent schwa-like sounds differ from mid central vowels in that the former are surface-level effects of speech timing in which, for example, one consonant ends before the next begins so that a vowel-like transition is produced. I consider these to be phonetic. In contrast, mid central vowels, whether underlying or epenthetic, are understood to be part of the grammar of a language. The difference between these two entities has direct bearing on prosodic structure because a phonological vowel can serve as a syllable nucleus where an excrescent transition cannot (although cf. Grice et al. 2011 for a potential counterexample regarding Berber tone). This, in turn, determines whether or not a word can be considered a sesquisyllable, which is necessarily disyllabic. The most confusion within these categories is found between excrescent transitions and epenthetic schwas, due to the fact that neither is present in the underlying representation.

Epenthesis is a well-studied phonological process (Itô 1989, Fleischhacker 2001, *inter alia*). As used here, epenthesis refers to a phonological process by which a sound is inserted to

prevent the violation of some constraint on a language. These restrictions which require vowel epenthesis can be one of two types. First, a language may not allow strings of consonants in certain positions, like onsets or codas. Second, a language may not tolerate sonority violations.

Many languages do not allow strings of consonants in certain positions, no matter their sonority. Within OT, these types of restrictions are captured by a number of constraints. Word-initially, for example, this may be enforced by a constraint *COMPLEX_{ONS}: “Syllables must not have more than one onset segment” (Zec 2007, 168). Such a constraint is operative in languages such as Arabic (Mascaró 2004, McCarthy 2007), Japanese (Itô and Mester 2004) and Finnish (Keyser and Kiparsky 1984, Prince 1984), and see also Blevins (1995) for more examples and references.

Regarding sonority, a number of axioms have been introduced to account for cross-linguistic sonority patterns, including the Core Syllabification Principle (CSP) and the Sonority Sequencing Principle (SSP). The CSP states that for each syllable node, adjacent sounds should be syllabified with that node if they are less sonorous but should be syllabified independently if they are not (Clements and Keyser 1981, 1983; Clements 1990). Another constraint on sonority is the SSP, which states that as one moves from a syllable boundary to a nucleus, segments should increase in sonority (Clements 1990).

Excrescence refers to a process by which a short, vowel-like sound is produced when the tongue is in a transition state between two other sounds. Hall (2011) provides a number of properties regarding excrescent schwa-like transitions, including that they are often phonetically weaker than other vowels and do not always appear in the phonemic vowel inventory of a language. In addition, they tend to be short, do not bear stress and have no effect on the satisfaction of phonotactic constraints of a language. Hall (2011) notes that they occur in a

number of languages including Piro (Matteson and Pike 1958), Finnish (Harms 1976), Sanskrit (Allen 1953) and South Hamburg German (Jannedy 1994).

While some studies recognize the differences between excrescence and epenthesis (Levin 1987; Davidson and Stone 2004; Hall 2004, 2011; *inter alia*), most works focus on one or the other. Since the advent of Articulatory Phonology, a number of authors have taken a reductionist approach, suggesting that many if not all instances of epenthesis be reanalyzed as excrescence (Gick 1999, Ridouane and Fougeron 2011, *inter alia*). I argue instead that both exist and are relevant in different ways. For example, many languages are known for having complex strings of onsets and do not require epenthesis to break up SSP-violating sequences, e.g. Polish (Rubach and Booij 1990), Tsou (Hsin 2000), Georgian (Butskhrikidze 2002), cf. Kreitman (2008) for a fuller list. Often, in these languages that allow SSP-violating clusters, excrescent vocoids are present, e.g. Georgian (Chitoran et al. 2002). On the other hand, many languages do not allow SSP-violating clusters on the surface, so epenthesis is required, e.g. Vimeu Picard (Auger 2001). Both are necessary for an adequate account of the minor syllable.

Although epenthesis and excrescence are structurally distinct, they have often been confounded, at least in part due to a lack of experimental evidence. In order to disambiguate the notation of excrescent transitions and epenthetic vowels from one another and from underlying mid central vowels, the following conventions will be used throughout the remainder of the dissertation. Excrescent vowel-like elements will be represented by [ɛ] and referred to as ‘vocoids’. Epenthetic mid central vowels will be notated by [ə] and referred to as ‘schwa’, and underlying mid central vowels will be given as /ʌ/ or [ʌ] as appropriate and will be referred to as ‘wedge’. A summary of these is provided in Table 1.1.

Table 1.1: Notational conventions for mid central vocalic elements

Symbol	Name	Description
ɤ	ex crescent vocoid	Phonetic transition; not phonological
ə	schwa	Epenthetic; phonological but not underlying
ʌ	wedge	Underlying; may be fully specified or not

Given my definition of the sesquisyllable as a disyllabic iamb, it follows that minor syllable nuclei must be phonologically real, i.e. epenthetic or underlying but not ex crescent. If there is no nucleus present phonologically, then the purported sesquisyllable should be considered a monosyllable with a word-initial consonant cluster instead. This distinction is critical, as monosyllables with word-initial consonant clusters that have inter-consonantal transitions have historically been considered a type of sesquisyllable, e.g. Kuay (Preecha 1988), Piro (Lin 1998), Mon-Khmer languages in general (Cho and King 2003). Regarding the following data from the Senoic languages (1.16), Diffloth (1976a, 233) even states, “The vowels on the one hand and the /h/ or /ʔ/ on the other require articulators whose movements are totally independent of each other, so that it is perfectly possible to superpose both articulations.” This suggests that the putative minor syllables in these words might be the result of phonetic (mis-)timing and are therefore ex crescent.

(1.16) Senoic

- a) /bhi:p(m)/ → [bihi:p(m)] ‘blood’
- b) /kʔɛ:p/ → [kɛʔɛ:p] ‘red centipede’

I suggest that much of the confusion between phonological vowels and ex crescent vocoids is due to (i) the fact that grammars are often based on impressionistic transcriptions and (ii) the lack of both a phonological model which could adequately account for this difference as

well as the instrumental investigation necessary to verify such a model. Despite the impressive level of language descriptions by Southeast Asian field linguists, the acoustic differences between phonological vowels and excrescent vocoids can be both difficult to hear and easy to misinterpret because of L1 bias.

A likely example of mistranscription comes from Svantesson's (1983) work on Northern Kammu (see Section 1.2.2.3 above), which has been widely cited and used to argue for various phonological positions (e.g. Shaw 1994). In this example, excrescent vocoids are not interpreted as phonological. Instead, phonological mid central vowels are either interpreted as excrescent and not transcribed or they are simply not heard at all, exemplifying that mid central vowels can be difficult to perceive. In 1983, Kammu minor syllables were described as consisting of one or two consonants but no vowel, so that the nucleus of the minor syllable was consonantal and tone bearing in those minor syllables with tonal contrasts (1.17).

(1.17) Northern Kammu

- | | | |
|----|-----------|--------------------------|
| a) | [sr̩.kàr] | 'straight' |
| | [l̩.mà:c] | 'get stuck (expressive)' |
| b) | [s.kár] | 'cause to be straight' |
| | [p.ká:y] | 'cause to return' |

However, in more recent work in which spectrographic evidence is provided, Svantesson and Karlsson (2004, 1) state, "The phonological representation of a Kammu minor syllable consists of either one or two consonants, and in addition *there is a vowel nucleus* which is not phonemic, but can be regarded as an epenthetic schwa vowel" (emphasis mine). Indeed, all tone-bearing minor syllables in Northern Kammu have a nuclear schwa, which, although not

underlying, is phonological and which was likely inadvertently omitted in the 1983 transcriptions. Therefore, the minor syllable consonants are not moraic and do not bear tone.

Even the small set of non-sonorant consonant minor syllables which are claimed to lack tone by Shaw (1994) and Svantesson (1983) are reanalyzed by Svantesson and Karlsson (2004) as both having an epenthetic schwa and being tone-bearing, although their tone is always identical to that of the major syllable. Some examples are given in (1.18a), and suggested reanalyzed structures of (1.17b) are given in (1.18b).

(1.18) Northern Kammu

- | | | | |
|----|-----------|------------------------|--------------------------------|
| a) | [kó.mú:l] | ‘silver’ | |
| | [cə.məʔ] | ‘rope’ | |
| | [só.cá:ŋ] | ‘elephant’ | (Svantesson and Karlsson 2004) |
| b) | [só.kár] | ‘cause to be straight’ | |
| | [pá.ká:y] | ‘cause to return’ | (adapted from Svantesson 1983) |

In summary, the conceptualization of the sesquisyllable as a disyllabic iamb necessitates that the minor syllable nucleus be phonologically present and not just a transitional state. The potential for confusion in descriptions of sesquisyllabic languages underlies the need for a more structurally accurate model. Table 1.2 shows the beginnings of such a model. As indicated, both monosyllables with word-initial consonant clusters and disyllabic iambs have traditionally been considered sesquisyllables. What differentiates them is the phonological status of the mid central vocalic element (or MCVE), where MCVEs in monosyllables are phonetic and MCVEs in the word-initial syllables of disyllables are phonological, regardless of whether they are underlying or epenthetic.

Table 1.2: Preliminary model of phonological content in various word types

	Word Type	MCVE Type
	Monosyllable with simplex onset	None
	Monosyllable with consonant cluster (no excrescence)	
Traditionally a Sesquisyllable	Monosyllable with consonant cluster and optional excrescent vocoid	Phonetic
	Disyllabic iamb	Phonological
	Disyllabic trochee	

1.4.2 The Effect of Stress on Mid Central Vocalic Elements

The previous section showed that monosyllables can be differentiated from disyllables by the phonological status of their MCVEs. This section explores the differences between types of disyllables with phonological MCVEs, i.e. iambs and trochees, which result from stress placement. To understand these differences, I first establish some facts about mid central vowels more generally. Figure 1.4 shows an example of an underlying mid central vowel in the English word [dʌd] *dud*. Based on a source-filter model, an idealized representation of this vowel produced by a male speaker with no obstruction in the vocal tract should have formants around 500Hz, 1,500Hz and 2,500Hz (Johnson 2012), and these are roughly the formant values represented by the white bars here.

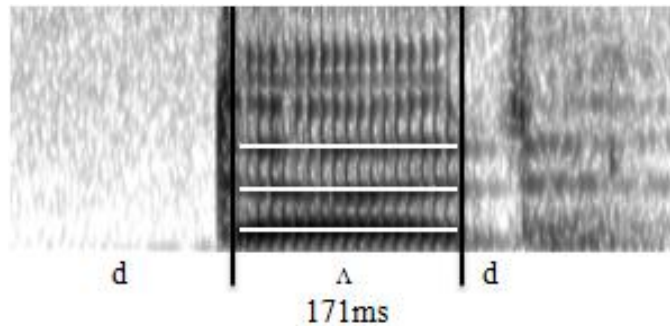


Figure 1.4: CVC monosyllable with a mid central vowel

Stress placement may affect both vowel quality and duration. First, the influence of stress on quality has been observed in English. As noted in Section 1.3.3.2, mid central vowels in English may either be stressed or unstressed and are traditionally referred to as ‘wedge’ and ‘schwa’, respectively. The matter of vowel reduction in English is further complicated by the fact that in most non-stem-final cases, reduced unstressed vowels are actually better transcribed as [ɪ] rather than [ə] (Flemming and Johnson 2007). In other words, the tongue is higher in the mouth and has a lower F1 than traditional transcriptions suggest. While this demonstrates the effect that stress can have on vowels in general and on mid central vowels in particular, it also highlights the problematic nature of describing these types of vowels, which is why I have opted for the transcription system given in Table 1.1.

In addition to the effects on vowel quality, (lack of) stress also affects the duration of mid central vowels. The following two spectrograms show examples of a nonce iamb [dʌ.ˈdʌd] (Figure 1.5) and a nonce trochee [ˈdʌ.ɾʌd] (Figure 1.6), both of which have two mid central vowels. Note that for both words, both the unstressed and stressed vowels are much shorter than the mid central vowel of the monosyllable in Figure 1.4 above. This is a direct result of the stress on the vowel in the monosyllable due to the property of culminativity as well as the effect of monosyllabic lengthening (White and Turk 2010 and references therein).

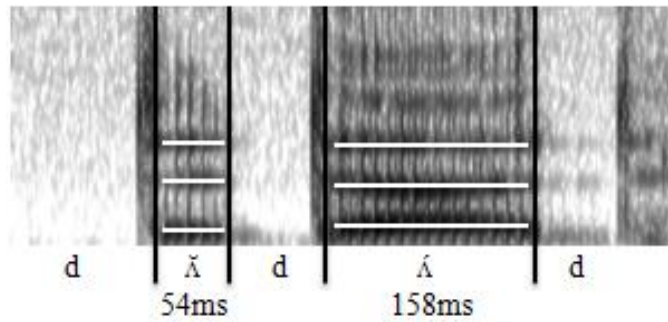


Figure 1.5: Iamb ([dʌ.ˈdʌd]) with two mid central vowels

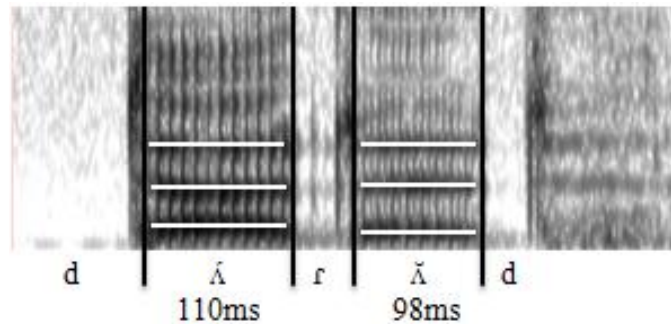


Figure 1.6: Trochee ([ˈdʌ.rʌd]) with two mid central vowels

In the iamb, the unstressed initial vowel is shorter than the stressed final vowel. In the trochee, the duration of both vowels is quite similar. This difference is due to a number of factors. Increased duration is often a phonetic manifestation of stress and final lengthening. In the iambic nonce word, these two factors converge, making the vowel in the word-final syllable much longer than vowel in the word-initial syllable. In the trochee, these two factors compete so that the nucleus of the word-initial syllable is lengthened due to stress while the nucleus of the word-final syllable is lengthened because it is word-final, and so the result is that the vowels in the nonce trochee are much more similar to each other in duration than are the vowels in the nonce iamb. The cross-linguistic patterns we find in iambs and trochees result from competing phonetic manifestations of prominence relations.

What differentiates the disyllables with phonological MCVes from one another is the placement of stress on the disyllable, i.e. word-initially or word-finally, as observed in Figure 1.5 and Figure 1.6 above. The difference between trochees and iambs as presented in this model is

clearly not dependent on whether or not the MCVE is epenthetic or underlying. Both epenthetic and underlying vowels may be either fully realized or undershot (Lindblom 1963). Instead, the phonetic realization of the phonological vowels is contingent on stress. Whether unstressed wedge is more fully attained than unstressed schwa is an open question and is not addressed here (cf. Smorodinsky 2002 for a similar investigation). For our purposes, what is important is that both unstressed wedge and schwa are less fully realized than stressed wedge and schwa. In other words, observed patterns suggest an inverse correlation between stress and reduction. While a lack of stress cross-linguistically tends to correlate with phonetic reduction (e.g. duration), phonological reduction due to stress is language-specific (cf. differences in unstressed vowels in English versus Spanish, for example); however, I consider such phonological reduction a requisite property of sesquisyllables. Table 1.3 below, which is expanded from Table 1.2 above, shows that whether a word is or is not a sesquisyllable is not contingent on the underlying representation of its MCVE but on its stress placement.

Table 1.3: Model of phonological content in and the role of stress on various word types. Sesquisyllabic types are given in the double outlined box.

Word Type	MCVE Type	Inserted?	Example
Monosyllable with simplex onset	None	N/A	/tak/ → [tak]
Monosyllable with consonant cluster (no excrescence)			/trak/ → [trak]
Monosyllable with consonant cluster – (ɜ)	Phonetic	Yes	/rtak/ → [rɜtak]
Disyllabic iamb – ə	Phonological	Yes	/rtak/ → [rə.'tak]
Disyllabic iamb – ʌ		No	/rʌtak/ → [rʌ.'tak]
Disyllabic trochee – ə	Phonological	Yes	/rtak/ → ['rə.tak]
Disyllabic trochee – ʌ		No	/rʌtak/ → ['rʌ.tak]

As in Table 1.2, Table 1.3 provides a description of each word type but expands the categories of disyllabic iambs and trochees to include both underlying and epenthetic vowels in the minor syllable position. The cases that I consider sesquisyllabic are found in the double-outlined box. These include disyllabic iambs with phonological MCVEs. Nonce examples of each type are also provided.

Finally, on the assumption that in the default case no syllabification is present in the underlying forms, the difference between monosyllables and disyllables with epenthetic vowels is determined by the surface realization of the words. Even if all vowels in a word are underlying, in cases where stress is regular or predictable, whether words are trochaic or iambic is still a surface property since no stress is present in underlying forms, either. This means, then, that the classification of a *word* as sesquisyllabic or as a maximally disyllabic iamb is only

relevant for surface forms and sesquisyllabicity is not necessarily a property of underlying representations of words. However, if a language has a preponderance of disyllabic roots due to a restriction on word or foot size, sesquisyllabicity is, in fact, a lexical property of that language.

Having laid out the differences between phonetic and phonological MCVes and the role that stress plays in their realization, we are now ready to consider how these facts interact within an articulatory framework. Such a framework provides a mechanism for understanding the difference between various types of vocalic material – in particular, whether a vocalic tongue gesture is present or not – differentiating monosyllables with word-initial consonant clusters from disyllabic iambs – and whether or not that gesture attains its target – differentiating disyllabic iambs from disyllabic trochees.

1.4.3 Mid Central Vocalic Elements as Gestures

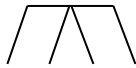

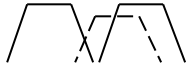

Traditional theories often divide speech into two different components – the physical/phonetic component and the cognitive/phonological component. In such theories, a mapping between the two is required. In contrast, Articulatory Phonology (or AP) (Browman and Goldstein 1986, 1989, 1990, 1992; Saltzman and Kelso 1987; *inter alia*) assumes that these components belong to the same system, in which speech is composed of gestures, or articulatory actions, that are related to each other both spatially and temporally. Because the primitives are actions, they are crucially dynamic, not static. In addition, utterances are conceived of as a set of gestural units which simultaneously involve a number of articulators. Relevant for our purposes here is the fact that utterances can contrast with one another through the presence or absence of a gesture. For example, [ɪd] and [kɪd] differ from one another in that the latter has an additional velar tongue body gesture that the former does not.

Because AP differs from segment-based frameworks in that it encodes temporal information, speech sounds can be understood as overlapping, and acoustic outputs can be understood as the result of varying degrees of target attainment. In addition to overlapping, gestures can also underlap, whereby the first gesture is released before the next begins, which often creates the appearance of schwa-like material between consonant gestures. Because the extent to which consonants overlap and underlap can be affected by speech rate, there can potentially be a large range of possible durations for phonological and excrescent sounds, which can make them appear to lie on a continuum instead of being discrete categories. AP provides a way to account for this variability in the acoustic correlates of MCVEs, which has most likely led to the range of previous accounts of minor syllables.

As can be seen in Table 1.4, there are three possible types of articulatory specifications for MCVEs. These correlate with the excrescent vocoid (\exists), the unstressed MCVE ([ə] or [ʌ]) and the stressed MCVE (again, [ə] or [ʌ]). In each case, the table shows whether the MCVE has an associated gesture, indicating whether it is phonological or not. I define the minor syllable MCVE as *phonological* in the case of the presence of a gesture and *phonetic* in the absence of a gesture. In addition, the table shows whether the vocal tract target is attained or not, along with a schematic gestural representation and a description of the manifesting acoustic properties of each. Crucially, while the presence or absence of a gesture can be considered a grammatical property of each type, the target attainment and associated acoustic properties are the resulting manifestations that stress has on each case. In other words, I assume that the unstressed MCVEs of iambs lack full target attainment and have different phonetic properties than the stressed MCVEs of trochees, which I model here as having full target attainment.

Aside from the three types of MCVEs, there is another possibility of gestural configuration for word-initial consonant clusters, given in the first row of Table 1.4. These are strings of segments which not only do not require phonological epenthesis but also do not exhibit an excrescent vocoid either. This is quite common in word-initial consonant clusters that do not violate the SSP. For example, the /kl/ sequence in the English word /klæp/ *clap* has no intervening excrescent material because the sounds do not underlap.

Table 1.4: Possible gestural representations of mid central vocalic elements (MCVEs). Sesquisyllabic types are given in the double outlined box.

Word Type	MCVE Notation	Gesture	Target Attainment	Gestural Representation	Phonetic Properties
Mono-syllable	None	✗	N/A		N/A
	ɜ	✗	N/A		Shorter Duration Lower F1 More variable F2
Disyllabic Iamb	ə or ʌ	✓	✗		Shorter Duration Equal or higher F1
Disyllabic Trochee	ə or ʌ	✓	✓		Longer Duration

In addition, note that the phonetic properties listed here are only meaningful inasmuch as they are relative to each other, and there is likely substantial variability in each category, as well. For example, while the stressed [ʌ] vowels in ['bʌt] *but* and [ʌ.'bʌt] *abut* would both be

represented by the schematic in the final row of Table 1.4, the former is no doubt longer than the former due to monosyllabic lengthening, as discussed in the previous section.

In stark contrast to the phonological stressed MCVEs are the vocoids found in the word-initial consonant clusters of monosyllables which have been so often described as sesquisyllables. These excrescent sounds, notated as [ɛ], are represented by the phonetic MCVE type and have no independent gesture (cf. Gafos 2002, Hall 2004). This is likely to be the case in the Nuxalk, Georgian and Polish clusters described in Cho and King (2003), as well. One well-documented case of gestural transition states is found in Tashlhiyt Berber. To test the phonological status of the intrusive MCVE in Tashlhiyt Berber, Ridouane and Fougeron (2011) conduct a series of articulatory and acoustic experiments to determine the phonetic conditioning necessary for the appearance of the intrusive MCVE as well as the durational differences across CC- and CɛC- sequences. These durational measures are interpreted as indicating whether or not [ɛ] has a “timing slot”, which they take to be evidence of its phonological reality. They find that the intrusive MCVE occurs in about 55% of their data, and is most common in CC- sequences in which C2 is voiced. They also find that CɛC- sequences are not significantly longer than CC- sequences, suggesting that [ɛ] does not have its own timing slot, which indicates that it is not phonological but that it is the result of gestural separation instead.

In between full target attainment by a vocalic gesture and the lack of a gesture altogether are the MCVEs I propose constitute the nuclei of minor syllables in true sesquisyllables. These are vowels with a gesture that do not have enough time to reach their targets. While unstressed vowels are demonstrably shorter than their stressed counterparts (barring higher level prosodic differences, of course), this effect is likely to be even more evident with mid central vowels whose duration is already shorter than other phonological vowels, all else being equal (Silverman

2011). Just as with stressed MCVEs, the gestural configuration for unstressed MCVEs can represent an underlying or epenthetic vowel. In the first case, using nonce words with unstressed MCVEs (which we assume are not epenthetic since they were provided in the stimuli), Browman and Goldstein (1992b) show that in English these sounds have a target and are not strictly an interpolation between neighboring vowels although they are highly affected by neighboring sounds. Davidson (2006a) observes a similar pattern with the second possibility – epenthetic [ə] – in her study of complex cluster repairs by non-native speakers. This study suggests that speakers are epenthesizing a vowel but that in order for the consonant sequences to sound as “native” as possible, they are shortening the duration of the epenthetic vowel, which prevents the gesture from being able to reach the target. Logically, for these unstressed epenthetic MCVEs, “shortening” must also imply that all else being equal, the post-vocalic gesture begins sooner relative to the onset of the preceding vowel, preventing the vowel gesture from fully attaining its target.

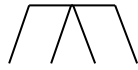

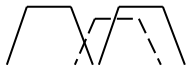

Each of these gestural representations tends to have different phonetic properties. First, after controlling for speech rate, consonant type, etc., the stressed phonological vowel should be the longest of the three. The unstressed phonological MCVE that does not attain its target and the phonetic transitional vocoid will both be shorter. In addition, because for excrescent vocoids the tongue is in a neutral position and is effectively interpolating between other gestures, it is usually higher in the mouth so its first formant (or F1) is lower. However, as it is dependent on both surrounding consonants and vowels, its F1 will be largely determined by its environment, so for example, since the tongue body is higher in the mouth for alveolars than for labials and velars, we expect the F1 of the vocoid to be lower when in the presence of alveolars than labials and velars. For unstressed vowels with an unreached target, the tongue is expected either to be in an

equivalent position to an underlying vowel with full target attainment or to be slightly lower, resulting in an equal or slightly higher F1 (Davidson 2006b). The second formant (or F2) of the vocoid should have a larger range than the F2 of the phonological vowels, again because it is completely dependent on surrounding consonants and vowels. Although this model captures the notion that unstressed MCVEs should be shorter than stressed MCVEs and have different formant specifications, I do not predict any phonological reduction between them. Indeed, they are assumed to have identical targets in the vocal tract; they are differentiated only by target attainment. These predictions are further explored in Chapters 2 and 3.

1.4.4 The Model

In summary, I have suggested that the ambiguity in characterizing the sesquisyllable can best be ameliorated through a theory of articulation. In using such a model, I have presented an empirically testable hypothesis regarding where the division between monosyllables and disyllables should be made. In particular, that is whether or not the minor syllable vowel has an associated gesture. Phonological vowels with gestures are further differentiated on a surface level by their target attainment (or lack thereof), which is dependent on whether or not they are stressed. I have also provided a list of acoustic correlates for three possible gestural configurations for the minor syllable vowel. These are fleshed out further in Section 1.6 below. I present Table 1.5 as a summary schematic of the facts discussed in this section. As exemplified in the table, throughout the remainder of the dissertation, phonetic excrescent vocoids are referred to as Type A MCVEs. Unstressed phonological vowels, whether [ə] or [ʌ], are referred to as Type B MCVEs. Finally, stressed phonological vowels are referred to as Type C MCVEs. I propose that of these types, only words with Type B MCVEs can be considered sesquisyllables.

Table 1.5: Model of gestural representations associated with MCVE types. Sesquisyllabic types are given in the double outlined box.

MCVE Type	MCVE Notation	Gesture	Target Attainment	Gestural Representation	Phonetic Properties
None	None	✗	N/A		N/A
A	∃	✗	N/A		Shorter Duration Lower F1 More variable F2
B	ə or ʌ	✓	✗		Shorter Duration Equal or higher F1
C	ə or ʌ	✓	✓		Longer Duration

Finally, this articulatory model of the minor syllable can be combined with the prosodic model of the sesquisyllable laid out in Section 1.3 to achieve a comprehensive integrated model. Figure 1.7 below shows prosodic structures in which words are maximally monopodal and all syllables are fully parsed. In this model, only iambic disyllables have Type B MCVEs and are considered sesquisyllables.

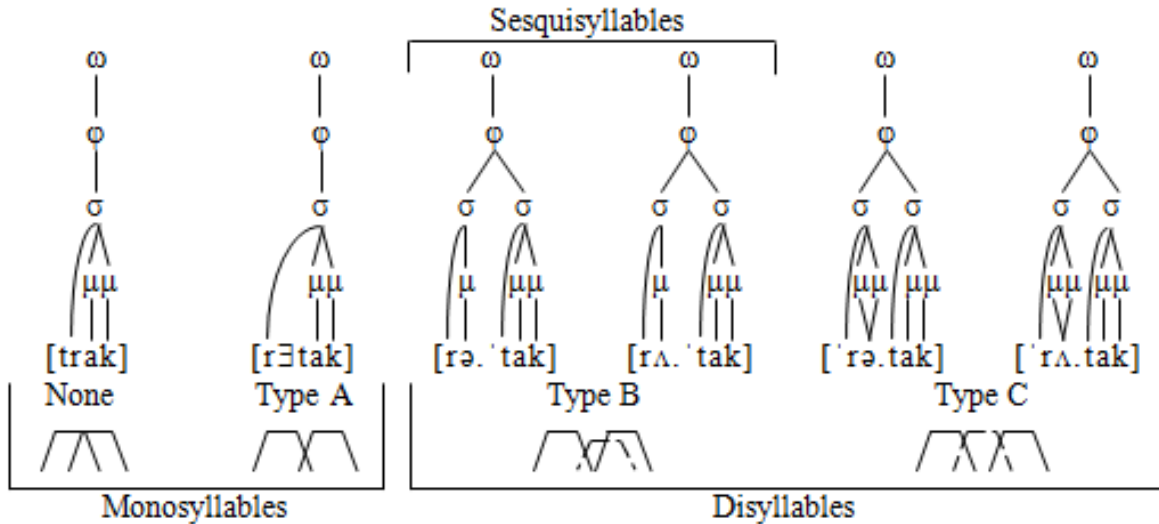


Figure 1.7: Prosodic model of sesquisyllables

1.5 EVALUATING PREVIOUS MODELS OF SESQUISYLLABLES

We are now in a position to examine previous models of the sesquisyllable. These are divided into two types – synchronic and diachronic – both of which are discussed below. While there are a number of descriptive synchronic accounts and case studies of sesquisyllables, very few actually present a cross-linguistic model of these word types. The most notable exception is Thomas (1992). I use Thomas’ analysis below to highlight both the richness of descriptive work that has been produced regarding the sesquisyllable as well as to demonstrate the need for a structural analysis in addition to such description. On the other hand, previous diachronic models, represented here by Matisoff (1990, 2003) and Brunelle and Pittayaporn (2012), while rich in structural analysis, tend to lack smaller-scale phonetic descriptions. While each of these approaches is a necessary contribution to the study of sesquisyllables, together they demonstrate the need for a more integrated descriptive and analytical approach.

1.5.1 Synchronic Models

Although a large amount of useful descriptive work has been carried out on a number of purportedly sesquisyllabic languages in Southeast Asia, the structural definition of the sesquisyllable has remained ambiguous. This is most particularly apparent in Thomas' (1992) account, which is one of the more comprehensive synchronic synopses of word types characterized as sesquisyllables. Thomas (1992) presents the sesquisyllable as four types of words, all intermediate between monosyllables and disyllables. I reconstruct his proposal here as a scale (Figure 1.8) and summarize it below.

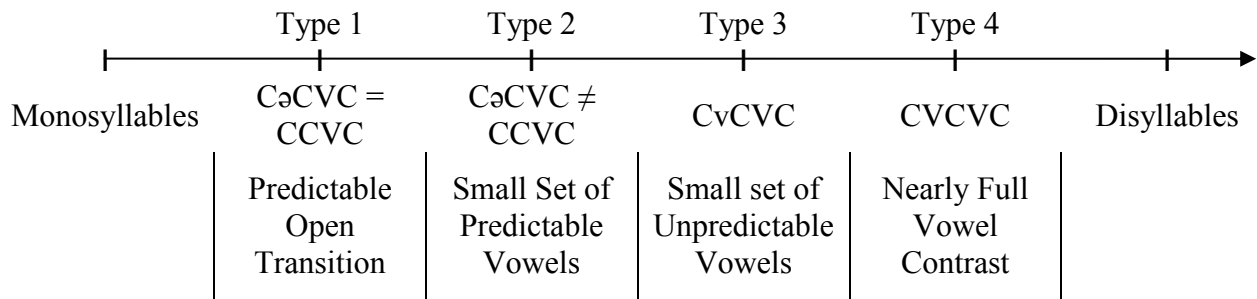


Figure 1.8: Scale of Thomas' (1992) sesquisyllable types

The minor syllable vowel in a Type 1 sesquisyllable is a “predictable open transition between consonants” (206), and the word itself is claimed to be a phonemic monosyllable. Thomas (1992) cites data from several languages for Type 1 sesquisyllables, in each of which there is one of two possible patterns. First, the optional schwa may be omitted when a nasal coda acts as the nucleus in its place (Halang, Cooper and Cooper 1966; Kuay, Preecha 1988; Northern Khmer, Thomas 1992). Second, the type of transition – aspirated, schwa or \emptyset – is predictable from the consonants in the minor syllable (Central Khmer, Jacob 1968; Stieng, Haupers 1969; etc.). Implicitly, Thomas (1992) considers the minor syllable non-contrastive with a CC-sequence because of this predictability, although he does not state this directly, nor is it stated in the source texts. An example of this lack of contrastivity was seen in Turung in example (1.4)

above, in which *finger* may be optionally realized as [ljung], [lɔjung] or [lijung], with no difference in meaning.

Next are Type 2 sesquisyllables which are contrastive between CəCVC and CCVC forms; in other words, minimal or near minimal pairs exist. For example, Chrau contrasts [plaj] ‘fruit’ with [pələj] ‘unfortunately’. These languages may have a small range of vowels in the minor syllable, but that range is usually limited to [i, ə, u]. In addition, the quality of the vowel is always conditioned by the environment. For example, in Chrau, the minor syllable vowel is realized as [ə] unless in the presence of a palatal or labial consonant, in which case it is realized as [i] or [u], respectively, as in (1.19).

(1.19) Type 2: Chrau

- | | | | |
|----|---------|------------|---------------|
| a) | [pədar] | ‘send’ | |
| b) | [sidac] | ‘king’ | |
| c) | [ruwəh] | ‘elephant’ | (Thomas 1971) |

The third type of sesquisyllable in Thomas’(1992) view are those in which the minor syllable has a reduced inventory, just as in the second type, but is crucially different in that the vowels are neither conditioned nor predictable by their environment. Pacoh (1.20) has /i, a, u/, and Kuay (1.21) has /i, ə, u/.

(1.20) Type 3: Pacoh

- | | | | |
|----|----------|------------|---------------|
| a) | [tinol] | ‘a post’ | |
| b) | [papi] | ‘converse’ | |
| c) | [kuchet] | ‘die’ | (Watson 1964) |

(1.21) Type 3: Kuay

- a) [kilek] 'a tree'
- b) [kəthiim] 'garlic'
- c) [sulin] 'Surin' (Preecha 1988)

In contrast to Type 3 minor syllables, Type 4 – represented as CVCVC – are claimed to have “nearly full vowel contrasts in a weakly stressed minor syllable” (209). However, all of the examples given – Halang (Cooper and Cooper 1966), Northeastern Thai (Preecha 1988), Kuay (Preecha 1988), etc. – have only a small set of these word types. The most extensive example is taken from Kensiw/Kensiu, a Mon-Khmer language of Thailand, a more expanded description of which is found in Bishop (1996). This fourth type of minor syllable is contrasted with a “presyllable” in Kensiw, which seems to be more like a different type, i.e. Type 1 or Type 3, according to Bishop (1996) and Thomas (1992), respectively. Given the rarity of Type 4 syllables, its usefulness as a category is somewhat dubious.

Thomas’ (1992) description is a welcome contribution to the study of sesquisyllabic languages because it attempts to deconstruct the notion of the sesquisyllable and even suggests differences between the role of the phonetics and the phonology in the composition of minor syllables. It also demonstrates that *minor syllable* is in actuality a cover term for a range of different linguistic entities. However, the model is still largely descriptive and does not lay out a clear analytic account of the sesquisyllable but continues to characterize what are very different linguistic entities as one type of unit. In response, we can evaluate Thomas’ (1992) descriptive model in terms of the more structural account I have provided thus far.

I suggest that Thomas’ (1992) types should be reconsidered in the following way, as represented in Figure 1.9. Type 1 sesquisyllables contrast with his “monosyllables” in one of two

possible ways. First, the former may have word-initial consonant clusters while the latter have simplex onsets. Alternatively, he may allow that “monosyllables” have word-initial consonant clusters but no intervening schwa-like material. Given his descriptions, it is unclear which of these he is assuming. Next, Types 2 – 4 are disyllabic iambs. While the vowels in the word-initial syllables are in both cases phonological, i.e. either underlying or epenthetic, in Type 2 sesquisyllables they are phonologically conditioned while in Type 3 sesquisyllables, they are not. Regarding Type 4 sesquisyllables, it is still unclear what their structure is. They seem to consist of only a small set of words in languages in which they are found, so there is insufficient evidence to definitively classify them. I include them here as disyllabic iambs, but perhaps they are more like trochees. More data are needed to make a strong conclusion. Finally, “disyllables” are understood to be disyllabic trochees.

Monosyllables Simplex Onsets (No MCVE)	Monosyllables Consonant Clusters (Type A)	Disyllabic Iambs (Type B)			Disyllabic Trochees (Type C)
		Penult Predictable	Penult Not Predictable	?	
	Type 1	Type 2	Type 3	Type 4	
Monosyllables	CəCVC = CCVC	CəCVC ≠ CCVC	CvCVC	CVCVC	Disyllables

Figure 1.9: Proposed phonological structure of Thomas' (1992) sesquisyllable types

While it is possible to interpret Thomas' (1992) categories in terms of prosodic structure, Thomas himself avoids doing so, in favor of a more descriptive approach. As a result, he considers a number of word types to be sesquisyllabic while my definition of the sesquisyllable excludes them. He also makes distinctions between word types, particularly Types 2 and 3 based on whether the minor syllable vowel is mid central or not, whereas I argue from a structural perspective that this distinction – albeit descriptively accurate – is not relevant for defining the

sesquisyllable. In other words, the lack of a one-to-one correlation between Thomas’ (1992) types and mine highlights the difference between underlying structure and surface properties.

1.5.2 Diachronic Models

Sesquisyllables often play an important role in historical reconstructions and diachronic models of Southeast Asian languages. In this section, I evaluate two such models. From a diachronic perspective, sesquisyllables are often understood as an intermediate step between disyllables and monosyllables as languages change over time. This idea is most explicitly laid out in Matisoff’s (1990, 2003) Compounding-Prefixation Cycle (Figure 1.10). In this approach, the sesquisyllable is a useful category, in that it represents a stage in which the minor syllable is characterized by a Type B MCVE instead of phonetically reduced vowels of any other quality.

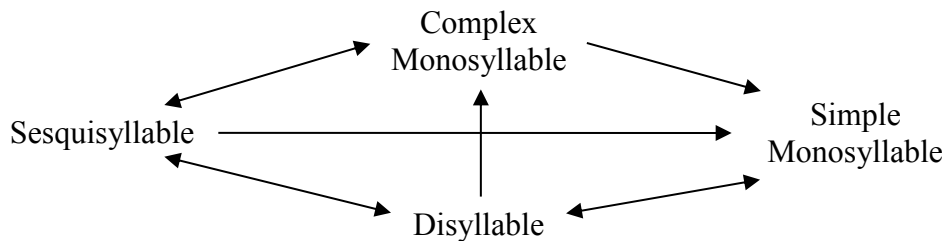


Figure 1.10: Compounding-Prefixation Cycle (Matisoff 2003)

In Figure 1.10, “simple monosyllable” refers to a monosyllable with a simplex onset, i.e. an onset consisting of exactly one pre-nuclear segment, in contrast to “complex monosyllable” with a word-initial consonant cluster. A “disyllable” is a compound word, made of two monosyllables, with no vowel reduction, whereas a “sesquisyllable” is a disyllabic word in which the first syllable contains [ə] or [ʌ]. Examples of each type in the history of Mizo (formerly referred to as Lushai), a Tibeto-Burman language primarily spoken in India, are given in (1.22) below in chronological order of evolution.

(1.22) Mizo

- a) Simple Monosyllable: *tsi ‘lung’; *-wap ‘spongy’
- b) Disyllable: *tsi.wap (from monosyllables)
- c) Sesquisyllable: *tʃə.wap (from disyllable)³
- d) Complex Monosyllable: *tʃuap (from disyllable) Matisoff (1990)

Matisoff’s (1990, 2003) model can be recast in structural terms of the type of MCVE represented in each stage. Figure 1.11 shows what such a model would look like. While these types fit quite well, it is unclear where iambs with non-mid central vowels and trochees with Type C MCVEs would lie. Nonetheless, note that Matisoff’s (1990, 2003) model has a much closer one-to-one relationship with MCVE types than does Thomas’ (1992) model, which highlights the difference that Matisoff’s model is largely based on structure while Thomas’ model is based on descriptive properties.

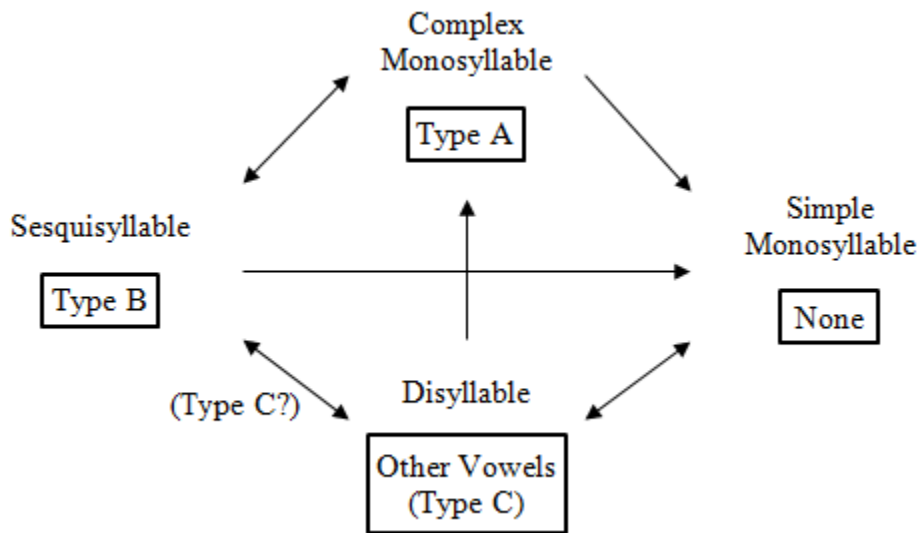


Figure 1.11: Compounding-Prefixation Cycle overlaid with MCVE types in boxes

Indeed, the diachronic role of the sesquisyllable in language change is not yet entirely understood. For example, in an effort to understand the processes by which sesquisyllables

³ [ə] here is based on Matisoff’s (1990) transcription and may be epenthetic or underlying.

become (simple) monosyllables, Michaud (2012) offers an example from Murong, a Mon-Khmer language spoken in Vietnam, in which */ksaŋ/ > */saŋ/ ‘tooth’. It is unclear whether the older reconstructed form was a sesquisyllable in the sense of having an epenthetic vowel (Type B), or whether it was just a complex monosyllable (Type A). Similarly, Sagart (1999) claims that in Chinese present-day monosyllables arose from two types of words – (i) monosyllables with “fusing” or tightly attached prefixes and (ii) iambic words with loosely attached prefixes. In other words, Old Chinese is claimed to have both monosyllables with complex onsets as well as sesquisyllables, which behaved differently from one another in the history of the language. Although this view is not universally accepted (cf. Ferlus 2009), it is nonetheless noteworthy that the subtleties of such a view are possible.

If, as I argue, sesquisyllables are in fact disyllabic iambs, then any model of language change which distinguishes disyllables as a group from sesquisyllables is problematic. Instead, what ought to be distinguished are disyllabic trochees and disyllabic iambs. For example, reconsider the Compounding-Prefixation Cycle and the examples for its categories given in (1.22b) and (1.22c) above, restated here as (1.23). Although the prominence relations between the syllables in this example are not provided (perhaps because they are not known), we may speculate that the form listed as disyllabic is a trochee, while the form listed as sesquisyllabic is an iamb.

(1.23) Mizo

- a) Disyllable: *tsi.wap (from monosyllables)
- b) Sesquisyllable: *tʃə.wap (from disyllable)

In more recent work, Brunelle and Pittayaporn (2012) have shown that although language change generally follows the pattern outlined by Matisoff (1990, 2003), reduction of

word/syllable shapes is restricted in an important way, i.e. that these changes are contingent on structural factors. Motivated by the Iambic-Trochaic Law (Hayes 1995), which as described in Section 1.3.2 states that elements contrasting in duration naturally form groupings with final prominence while elements contrasting in intensity naturally form groupings with initial prominence, they provide a more enunciated model of what is and is not possible in the types of sound changes found in mainland Southeast Asian languages (Figure 1.12).

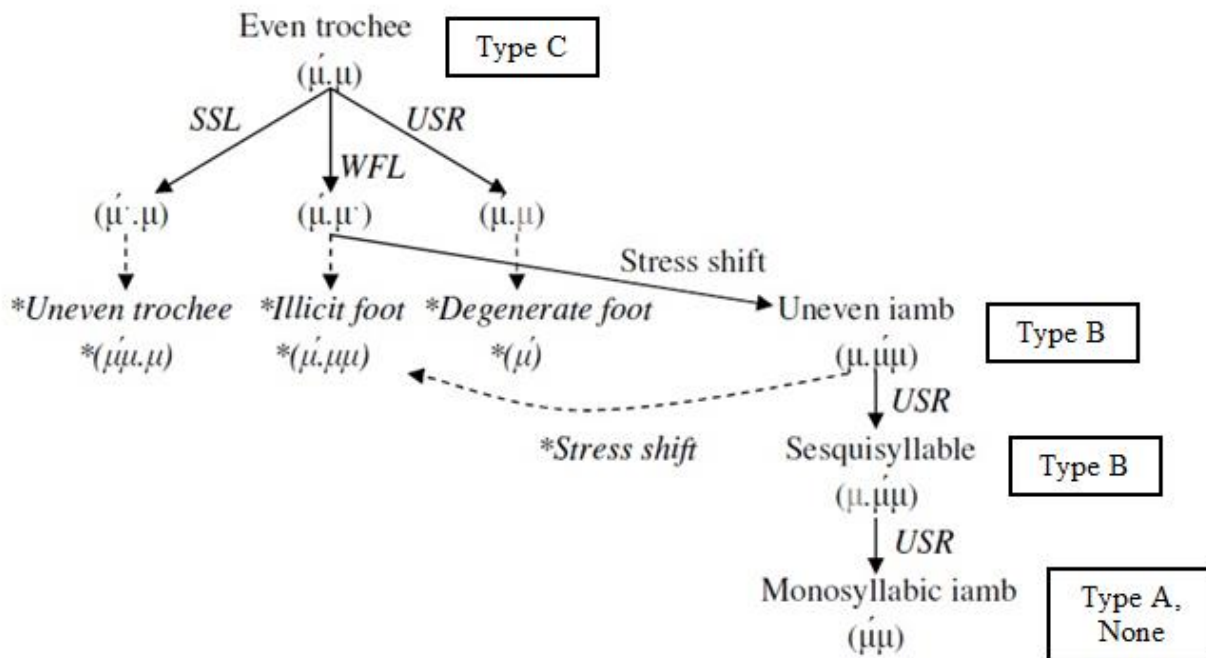


Figure 1.12: Path to monosyllables (Brunelle and Pittayaporn 2012, 424), with MCVE types

As laid out in this figure, disyllabic trochees with even weight are subject to a stress shift, which results in words in which the final syllable is longer than the initial syllable, i.e. iambs. During a later sesquisyllabic stage, the initial syllables are reduced, and finally the initial syllables are lost, resulting in monosyllables. Like the Compounding-Prefixation Cycle, this model has only four stable states, which are roughly associated with the MCVE types noted in boxes in Figure 1.12. It is crucially different than Matisoff's (1990, 2003) model in that it provides an intermediate stage in between the disyllable (trochee) and the sesquisyllable (iamb).

This model assumes that the sesquisyllable is limited to forms with schwa or wedge in the minor syllable, which is why the uneven iamb is separated from the sesquisyllable. In other words, the sesquisyllable is seen as a type of extra-reduced iamb. This model also conflates monosyllables with word-initial consonant clusters and monosyllables with simplex onsets. Because these types of monosyllables are phonologically identical in terms of weight, this lack of differentiation is reasonable here, given that this is a structurally-based analysis, although a distinction could be drawn between these two types of monosyllables within the model in more descriptive terms.

The sort of trajectory of language change we find in Figure 1.12 represents a large number of languages in Southeast Asia. One of the most well-documented examples of this type of change is found in Eastern Cham, an Austronesian language spoken in Vietnam. As laid out by Brunelle (2004), while Proto-Malayo-Chamic was trochaic, by the time of Proto-Chamic, the language was iambic (Thurgood 1996, 1999). Brunelle (2004, 44) states, “This new stress pattern was accompanied by a general neutralization of phonological contrast in the unstressed non-final syllables, resulting in what is called a *sesquisyllabic* canonical word shape.” He defines this neutralization as a reduction in the possible vowels and consonants in the unstressed syllable (1.24).

(1.24) Cham sesquisyllables

- | | | | |
|----|---------|-----------------|---------------------|
| a) | [kəpaw] | ‘water buffalo’ | |
| b) | [çəlan] | ‘road’ | |
| c) | [pətaw] | ‘stone’ | (Brunelle 2004, 44) |

This seems to suggest a conflation of the iambic and sesquisyllabic stages, as opposed to the separated stages in Brunelle and Pittayaporn’s (2012) model. Eventually in Cham, the vowels in these unstable unstressed syllables were reduced to schwa. Finally depending on the register

of the speech, either the unstressed vowels were deleted, leaving monosyllables with word-initial consonant clusters (1.25a), or the entire initial syllable was deleted, creating monosyllables with simplex onsets (1.25b).

(1.25) Cham monosyllables

a-1)	/palāj/	>	/plēj/	‘village’	
a-2)	/çəlan/	>	/klan/	‘road’	
b-1)	/tapa/	>	/pa/	‘to cross’	
b-2)	/rilo/	>	/lo/	‘many, a lot’	(Brunelle 2004, 46)

Similarly, Vietnamese, which was once a disyllabic language, has become monosyllabic, although the process is not so well documented as it is in Cham. Using comparative data and historical methods, Ferlus (1982) shows that in addition to reduction and deletion of the unstressed syllable as in Cham, monosyllables in Vietnamese were also created through a process of spirantization and eventual deletion of word-medial consonants. He provides a number of comparative examples with Aheu (or Thavung), a closely related Austroasiatic language spoken in Laos and Thailand, which still retains the “pre-syllables” (Table 1.6). As can be noted in these forms, the voiced onsets in Vietnamese correspond to the voiceless C2 onsets in the Aheu forms. The C1 onsets in Aheu are what Ferlus (1982) takes to be the minor syllables, and these are what are lost in the Vietnamese correspondences.

Table 1.6: Aheu correspondences to Vietnamese monosyllables (tones omitted)

Aheu/Thavung	Vietnamese	Gloss
kpaas	vaj	‘cotton’
ksang	raŋ	‘tooth’
pkʌ	gaj	‘female’
ckɔɔŋ	giŋ	‘ginger’

1.5.3 Summary

Previous synchronic and diachronic models of the sesquisyllable both reveal the need for an integrated descriptive and structural approach. While synchronic models tend to be more descriptive and diachronic models tend to rely more on structure, an ideal model will integrate both components. By teasing apart the different contributions associated with description and analysis, we are able to be more precise about the nature of sesquisyllables in synchronic grammars as well as how languages might change over time while passing through a phase of sesquisyllabicity.

Thomas’ (1992) synchronic model does rely on phonological structure inasmuch as monosyllables, sesquisyllables and disyllables are all distinguished from one another. However, the latter two word types would be categorized together under a structural account, recognizing, as noted in Section 1.4, that the difference between iambs (Types 2 and 3) and trochees (“disyllables”) is a surface distinction. In addition, a more detailed structural account may resolve the ambiguity of Thomas’ (1992) Type 4 sesquisyllables whose categorization at this point remains somewhat ambiguous.

In terms of diachrony, both of the models presented above, i.e. Matisoff (1990, 2003) and Brunelle and Pittayaporn (2012), view the minor syllable vowel as a result of reduction due to stress. Indeed Brunelle and Pittayaporn (2012) make a strong case that this is the pathway to sesquisyllabism in mainland Southeast Asian languages. However, it is worth noting that reduced vowels, which might be interpreted as minor syllable vowels, could *a priori* arise through excrescence reinterpreted as epenthesis. In other words, sesquisyllables could result from augmentation of Type A to Type B MCVEs rather than reduction of Type C MCVEs to Type B MCVEs. An account of what this might look like is presented in Chapter 3.

1.6 THE PREDICTIONS OF THE MODEL

At this point, our model provides a verifiable surface description of the sesquisyllable as well as an associated underlying structure. The model includes both prosodic and articulatory components, laid out in Sections 1.3 and 1.4, respectively, which capture the descriptive generalizations of sesquisyllables listed in (1.9) at the end of Section 1.2.2. In sum, sesquisyllables are non-iterative disyllabic iambs. Because they are disyllabic, the minor syllable must have a phonological nucleus, which I argue is most accurately represented as a Type B MCVE. These types of mid central vowels have gestural targets but their targets are not fully attained due to a lack of stress and a shorter duration. Thus my definition of the sesquisyllable is dependent on the phonological status of the minor syllable and is relevant only on surface forms of words, not underlying forms.

We are now in a position to lay out the specific predictions of the model. In testing those predictions, we will determine how and to what extent phonetic data can shed light on phonological analysis. As such, this study is couched in the paradigm of Laboratory Phonology (cf. Beckman and Kingston 1990), which seeks to understand how phonetics and phonology

interact with one another. To this end, the focus of the dissertation now turns to a series of phonetic experiments in which these ideas are explored with primary language data from three mainland Southeast Asian languages – Khmer, Bunong and Burmese. Khmer and Bunong are both Mon-Khmer languages spoken in Cambodia as well as Vietnam. While Khmer is the national language of Cambodia, Bunong is a minority language and is far less well documented. Burmese is a Sino-Tibetan language and is the official language of Myanmar. Khmer and Bunong data were collected in 2011 in field sites in Phnom Penh and in Monduliri Province, Cambodia. Burmese data were collected in Ithaca, NY.

Phonetic investigations of this sort provide a number of benefits. First, at least some of the variability in descriptions of the sesquisyllable is likely the result of impressionistic transcriptions. It can be quite difficult to determine simply from listening if a mid central vocalic element has an associated gesture or not. Alternatively, basing a description on phonetic data allows for a more verifiable description. Second, I also claimed that the previous phonological accounts of minor syllables rely heavily on second-hand phonetic data or on other phonological accounts. Given the discrepancies in descriptions of minor syllables and sesquisyllables, these previous phonological accounts of sesquisyllables and minor syllables often begin with the disadvantage of inaccurate or wrongly interpreted data (cf. de Lacy 2011).

Although the model I present is largely one of articulation and gestural configurations, the experiments themselves are acoustic in nature. To date no articulatory data have been collected for these languages. While gathering such data would be a useful further study, many articulatory events can be interpreted from the acoustic record (e.g. Davidson 2006b). Thus an acoustic study is a useful contribution in that specific predictions can be made about the correlates of articulation and acoustic realization. The results of these acoustic experiments lay

the groundwork for forming hypotheses regarding the articulatory configurations of minor syllables, which can then be empirically tested through articulatory measures in future studies.

In what follows, I lay out a number of phonetic properties which are suggested by my phonological model. While the verification of the following phonetic hypotheses is necessary if the phonological model is accurate, they cannot prove the phonological model correct. However, if the empirical data do not match the phonetic predictions, we can take this as evidence that the phonological model is somehow flawed. This is in part because these hypotheses represent only the phonetic manifestations of the phonological structure. In other words, they are the surface effects of underlying properties.

Figure 1.13 shows each of the possible MCVE types (including None) as what can be thought of as schematized spectrograms (based on Figure 1.4 – Figure 1.6 above). The dark bars represent F1 and F2 values. The word with no MCVE in its word-initial consonant cluster in (i) serves as a baseline, assuming that the F1 of the stressed vowel is that of an underlying /ʌ/ of 500Hz, and F2 is 1500Hz. While F2 is given as a reference point on the major syllable vowels, I have listed only F1 for the minor syllable vowels since we can be more specific about how the height of the tongue changes based on the phonological content of the MCVE and whether or not it is stressed than how the backness of the tongue changes. Nonetheless, in terms of F2, we can predict that its distribution should be most affected by surrounding consonants for Type A MCVEs and least affected for Type C MCVEs.

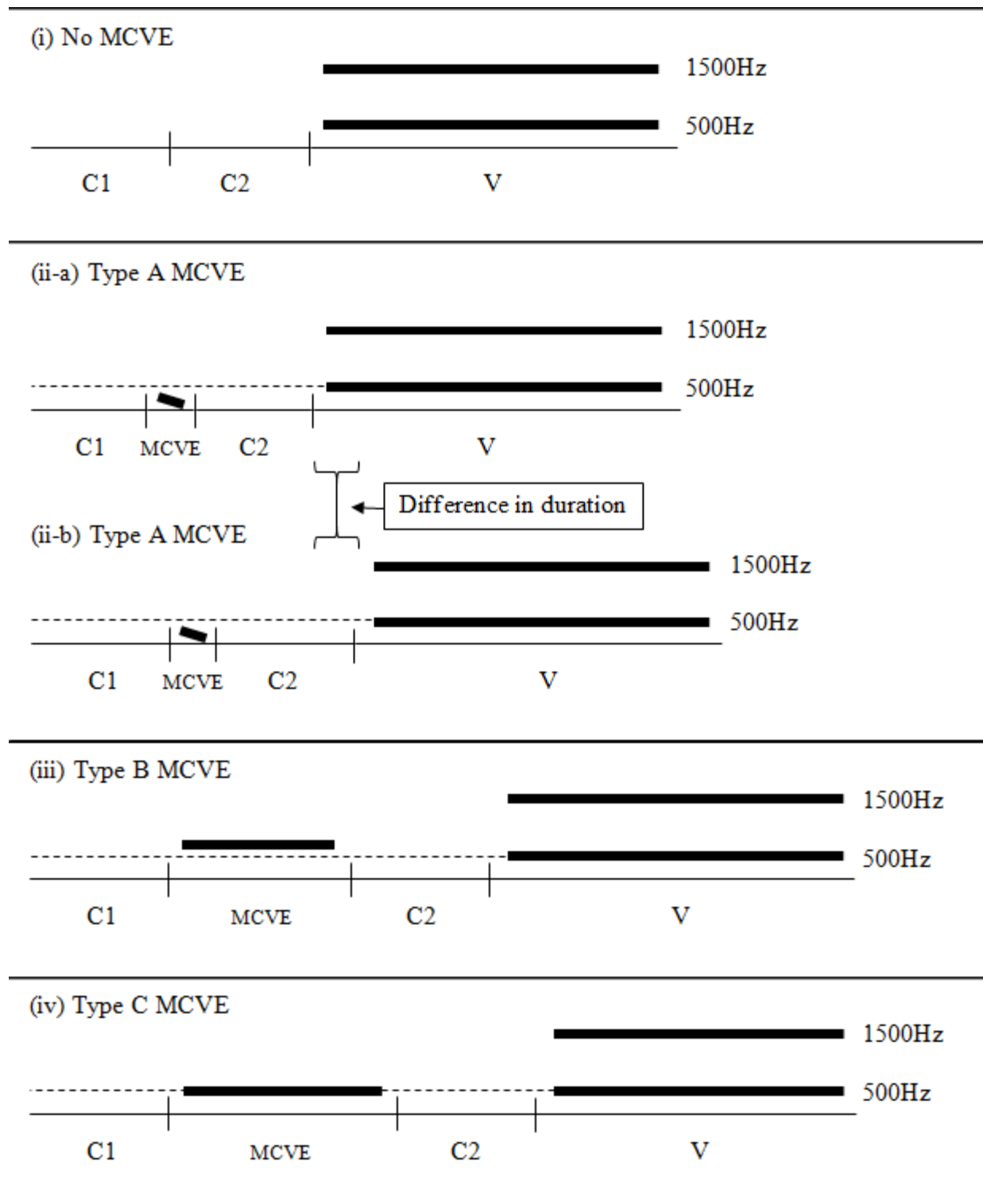


Figure 1.13: Schematized spectrographic predictions of MCVE types

In (ii), the Type A MCVE has two possible realizations, one in which the MCVE does not increase the duration of the CC sequence and one in which it does. First, note that the word-

final full vowel is equivalent in duration to the word with no MCVE in (i). Since each is a monosyllable, we do not expect any noticeable effects of shortening due to word length. Also note that in both possible realizations, F1 is a cline. This is because F1 of a Type A MCVE is contingent on the formant values of its surrounding consonants. All else being equal, it is also lower than that of underlying wedge, because it is just a transition state so the tongue does not lower between consonants to achieve a target. The difference between (ii-a) and (ii-b) is whether or not the transitional material contributes additional duration to the CC sequence. In the first representation (ii-a), the consonant gestures are achieved more quickly than they are in (i) – perhaps due to changes in speech rate or other normal variation – so that the MCVE does not contribute any duration, but in the second sequence (ii-b), the consonant gestures are simply farther apart so that the MCVE does contribute additional duration.

Type B MCVEs in (iii) differ from Type A MCVEs in that they are longer and their F1s are higher. Indeed, their F1s should be either a similar value as or higher than the F1 of underlying wedge because although they have a gestural target, they do not have time to fully attain it. The Type C MCVEs in (iv) are longer still than Type B MCVEs, and their F1s are equal to or lower than Type B MCVEs. Note that the words represented here are assumed to be said either in isolation or focused, so although disyllables with Type C vowels are taken to be trochees, the MCVE is actually not longer than the vowel in the final syllable.

These hypotheses are evaluated through the three case studies in Chapters 2, 3 and 4, the order of which is motivated by their relevance to the MCVE types laid out in Table 1.4. I begin in Chapter 2 by investigating a set of words in Khmer, which is claimed to have sesquisyllables with variable yet predictable minor syllable nuclei. Based on the acoustic results, I suggest that this type of word is not a true sesquisyllable. It is better understood as a monosyllable with a

word-initial consonant cluster whose purported minor syllable “vowel” is actually a transition between consonant gestures, or a Type A MCVE. In Chapter 3, I analyze Bunong. Results show that minor syllables in Bunong are structurally distinct from those in Khmer in that they are phonological entities with associated gestures. Bunong sesquisyllables are best understood as disyllabic iambs with Type B MCVEs. In Chapter 4, I present the results of a pilot study on Burmese, notable for having multiple minor syllables. I consider the Burmese data in light of two potential foot structures – both my iambic analysis and a trochaic analysis proposed by Green (1995, 2005) – to address the issue of ambiguity between iambic and trochaic systems and their implications for a theory of sesquisyllables. Finally, Chapter 5 presents general conclusions, highlighting the themes covered in the dissertation and the descriptive and theoretical contributions, with a particular focus on how principles commonly used within prosodic frameworks can be grounded in terms of oscillation.

CHAPTER 2

“MINOR SYLLABLES” ARE NOT SYLLABLES: PHONETIC EVIDENCE FROM KHMER

2.1 INTRODUCTION

Khmer, an Austro-Asiatic language spoken in Cambodia, is a fitting test case for the phonetic underpinnings of one type of minor syllable. Indeed, the term “minor syllable” was first coined in a description of Khmer by Henderson (1952). However, as we saw in Chapter 1, the definition of “minor syllable” has changed over time. I have argued that what are now standardly taken to be sesquisyllables in Khmer are in fact monosyllables with word-initial consonant clusters separated by excrescent vocoids, which I refer to as Type A mid central vocalic elements (MCVEs). The acoustic study presented here seeks to confirm this hypothesis and test the predictions made about the phonetic correlates of MCVE types laid out at the end of the previous chapter, repeated here in Figure 2.1.

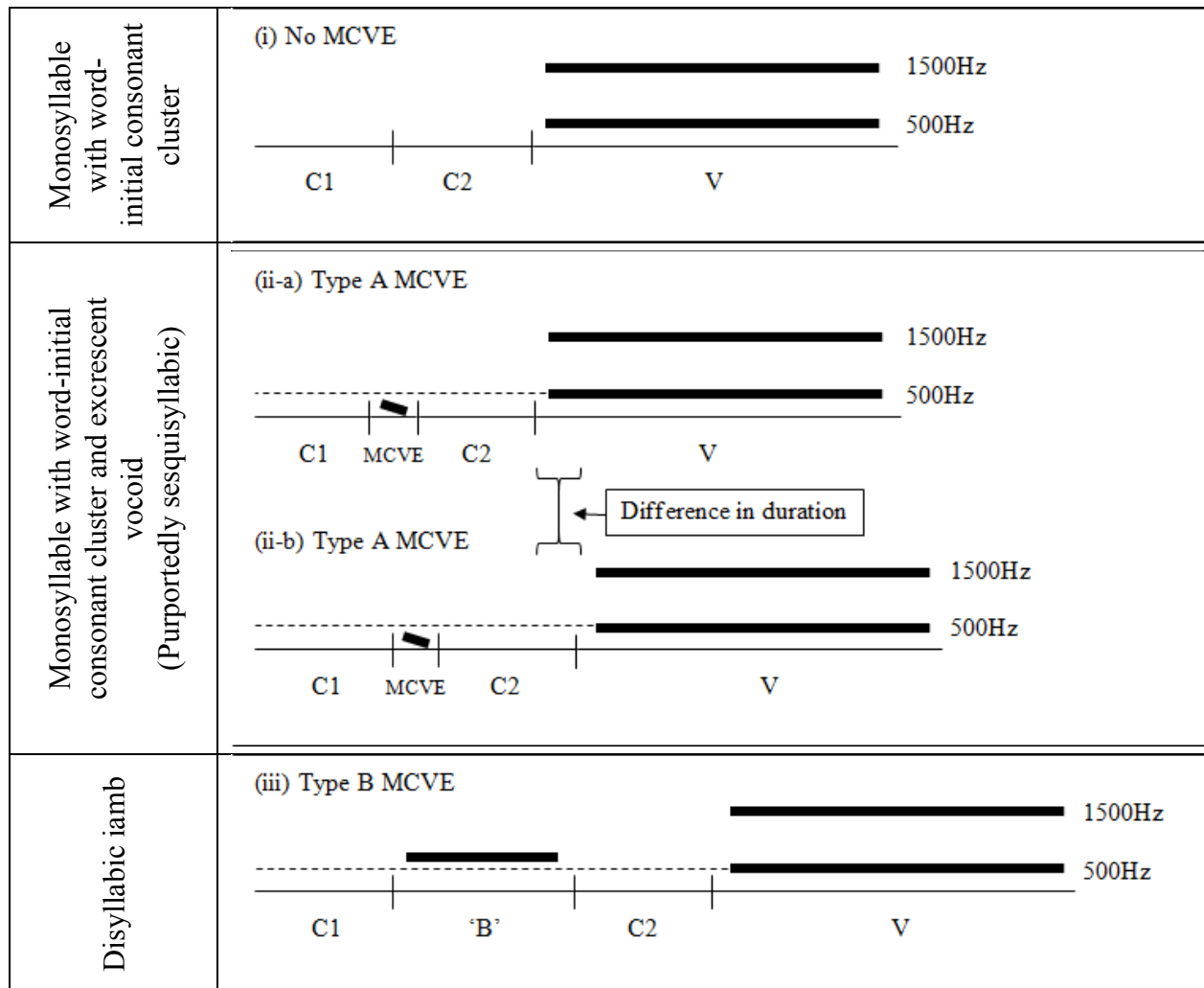


Figure 2.1: [Khmer] Potential MCVE types

This study evaluates as sesquisyllables only those forms listed in Figure 2.1 as having Type A MCVEs, which Henderson dubs *extended monosyllables*. However, Khmer does, in fact, contain a set of words (described by Henderson (1952) as *minor disyllables* and discussed below) which may be more in line with my proposed prosodic model of the sesquisyllable. I have chosen to focus exclusively on the words with Type A MCVEs because (i) they, more so than words with Type B MCVEs, are considered to be the sesquisyllables of Khmer (with the exception of DiCano 2005) and (ii) they provide a rich testing ground for the phonetic correlates of Type A MCVEs. Although an investigation of Khmer words with Type B MCVEs would be a valuable contribution to an analysis of sesquisyllables, it is left for future study.

Khmer is the national language of Cambodia and has between 12 and 13 million speakers (Lewis et al. 2013). It is one of the most well-studied Mon-Khmer languages in terms of register (Gaudes 1978, Wayland 2002, Wayland and Jongman 2003, *inter alia*), historical comparison (Huffman 1976, Ferlus 1992) and syllable shape (Henderson 1952, Pinnow 1979, Thach 1999, *inter alia*). While it has a fairly typical consonant inventory (Table 2.1), Khmer is known for having an abundance of vowels due to the instantiation of the historical register system as one of vowel quality. A register system is one in which a historical voicing contrast has been lost, and as a result the post-consonantal vowel has become the locus of contrast. Register can be phonetically manifested in numerous ways, including pitch, voice quality, etc., and in Khmer it is primarily exhibited through vowel quality differences. As a result, Khmer has two different sets of vowels – one for each register, i.e. one that follows historically voiceless consonants and one that follows historically voiced consonants. Figure 2.2 below provides a description of vowels in Standard Khmer (for alternatives, see Pinnow 1979).

Table 2.1: Khmer consonants

	Labial	Alveolar	Palatal	Velar	Glottal
Plain Stops	p	t	tʃ	k	ʔ
Aspirated Stops	p ^h	t ^h	tʃ ^h	k ^h	
Fricatives		s			h
Nasals	m	n	ɲ	ŋ	
Liquids		l, r			
Glides	w	j			

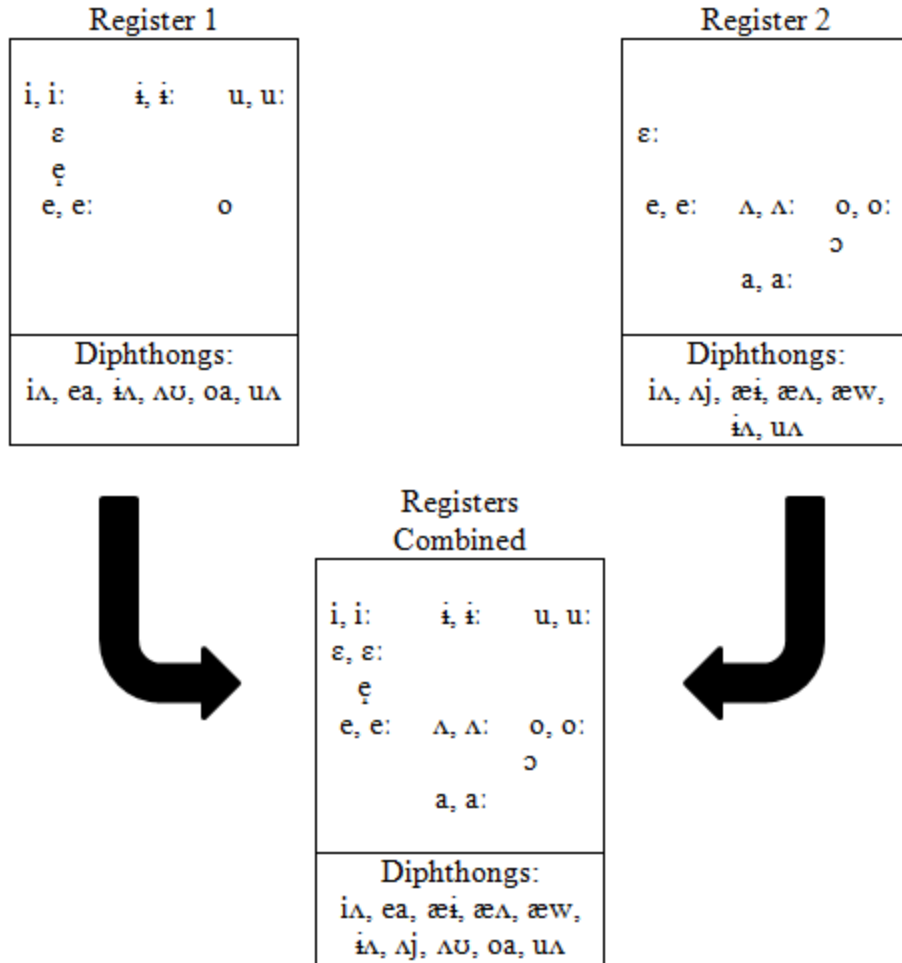


Figure 2.2: Khmer vowels (Registers separate and combined)

2.2 BACKGROUND OF THE MINOR SYLLABLE IN KHMER

Henderson (1952) divides word types in Khmer into four categories: simple monosyllables, extended monosyllables, minor disyllables and major disyllables (Table 2.2). Simple monosyllables have simplex onsets, and extended monosyllables have word-initial consonant clusters which often include a transition. Minor disyllables are composed of a minor syllable followed by an unreduced major syllable and are “intermediate structurally between the extended monosyllable and the full, or major disyllable” (150). Major disyllables are composed of two major syllables.

Table 2.2: Khmer word types (adapted from Henderson 1952)

Simple Monosyllables		Extended Monosyllables	
[kæʌt]	‘to be born, grow’	[khnæʌt]	‘waxing of moon’
[tʃam]	‘to await, keep watch’	[prʌtʃam]	‘to watch another one’
[sæʌm]	‘wet’	[psæʌm]	‘to wet’
Minor Disyllables		Major Disyllables	
[bɔŋkæʌt]	‘to produce, give birth’	[kæʌtlæʌŋ]	‘to grow’
[bɔŋtʃam]	‘to pledge’	[tʃammʌ]	‘to wait to see’
[sɔnsæʌm]	‘dew’		

In Henderson’s (1952) view, extended monosyllables, which would later be dubbed sesquisyllables by Matisoff (1973), and minor disyllables are differentiated in two important ways. First, the minor syllable vowel in minor disyllables can have a different register than the following major syllable. In contrast, when mid central vocalic elements (MCVEs) appear in extended monosyllables, they always match the register of the major syllable. Second, extended monosyllables are more permissive in their consonantal inventory than are minor disyllables, but they are more restrictive in their vowel inventory. In minor disyllables, C1 cannot be nasal, and a nasal coda must always be present, whereas there are no restrictions on the consonantal inventory of extended monosyllable onsets, aside from the language’s general phonotactic constraints. In terms of vowels, however, extended monosyllables allow only an MCVE (in addition to [h] or nothing), whereas minor disyllables may contain an [ʌ], [u] or [ɔ].

On the other hand, there are many similarities between extended monosyllables and minor disyllables. For example, in slow speech, Henderson (1952) suggests that the vocalic portion of the extended monosyllable can be identical in quality to the minor syllable vowel of

minor disyllables, assuming the registers are the same. In addition, both the MCVE of extended monosyllables and the nucleus of minor syllables are unstressed. Finally, differences are further obscured in fast speech, when minor syllable codas are dropped, as in (2.1). Note, however, that the orthographic forms represent the slower pronunciations, i.e. those including the coda.

(2.1)

/kɔ̃nla:t/	→	[kʌla:t]	‘type of insect’
/bɔ̃ŋko:l/	→	[pʌko:l]	‘stake’
/tʃɔ̃ŋkiʌŋ/	→	[tʃʌkiʌŋ]	‘lamp’

Henderson (1952, 170) notes there is a “gradual progression from simple monosyllable, through extended monosyllable and minor disyllable, to major disyllable. Between the stages there is only a relatively small structural difference. There is no sharp boundary between monosyllable and disyllable”. Nonetheless, some important difference must remain because as she states, speakers do differentiate between forms like those in (2.2), which reflect a fairly prevalent morphological process.

(2.2)

[khlah]	ḡḡ:	‘bolt (v)’
[kʌlah](< [kɔ̃nlah])	ḡḡ:	‘bolt (n)’

Subsequently, Huffman (1972) builds on and reinterprets Henderson’s analysis, further emphasizing the lack of a clear dividing line between monosyllables and disyllables. However, in his analysis, the focus of ambiguity is on Henderson’s (1952) extended monosyllables, not the minor disyllables. In fact, Huffman (1972) groups minor disyllables with major disyllables and does not address the differences between minor and major disyllables at all. However, he does

discuss at length the extensive reduction of disyllables, including both major and minor disyllables.

Huffman (1972) suggests that monosyllables with word-initial consonant clusters have predictable transition states depending on which consonants are involved. These are grouped into three classes: Class 1 has no transition, Class 2 has a voiceless aspirated transition, and Class 3 has an MCVE transition (Table 2.3).

Table 2.3: Classes of word-initial consonant clusters in Khmer, according to Huffman 1972

Class 1		Class 2				Class 3		
C1	C2	C1	C2			C1	C2	
p	s	p ^h	p	m	w	p	s	b
t	r	t ^h	t	n	j	t	m	d
t̂	h	t̂ ^h	t̂	ɲ	l	t̂	l	ʔ
k		k ^h	k	ŋ		k		

These class types are further exemplified in Table 2.4 below. C1s are listed along the y-axis, and C2s are listed along the x-axis. Each degree of shading represents a different class of word-initial consonant clusters. First, Class 1 sequences, which are the most darkly shaded, are not claimed to have any material intervening between the consonants. They are described by Huffman (1972, 55) as having “a relatively close transition” from C1 to C2. Next, Class 2 consonant sequences, which are lightly shaded, are claimed to have “slight aspiration” between C1 and C2 (represented here by [ɿ]). Finally, Class 3 sequences, which have no shading, are described as having a “weak intruded vocalism of a mid-central quality” (represented here by [ɿ]). Only a subset of possible clusters are reported on in this study; those are found in the double-outline boxes.

Table 2.4: Khmer word-initial consonant clusters (Based on Huffman 1972)

C1 \ C2	s	h	r	l	p	t	tʃ	k	m	n	ɲ	ʔ	b	d	
p	ps	ph	pr	pl		pt	ptʃ	pk		pn	pɲ	pʔ		pd	
t		th	tr	tl	tp			tk	tm	tn		tɲ	tʔ	tb	
tʃ		tʃh	tʃr	tʃl	tʃp			tʃk	tʃm	tʃn		tʃɲ	tʃʔ	tʃb	tʃd
k	ks	kh	kr	kl	kp	kt	ktʃ		km	kn	kɲ	kʔ			
s			sr	sl	sp	st		sk	sm	sn	sɲ	sɲ	sʔ	sb	sd
m	ms	mh	mr	ml		mt	mtʃ			mn	mɲ		mʔ		
l		lh			lp			lk	lm			lɲ	lʔ	lb	

Classes 1 and 2 remain phonetically separate from reduced disyllables because disyllables will always have some sort of vocalism in the unstressed syllable, as in (2.3).

(2.3)

- a) Class 1: [sla:] /sla:/ ‘to make stew’
 Reduced Disyllable: [sʌla:] /sʌla:/ ‘stew’
- b) Class 2: [pɛ̃teah] /pteah/ ‘house’
 Reduced Disyllable: [pʌteah] /pʌteah/ ‘to meet’

However, this is not true for Class 3 monosyllables, as can be seen in (2.4). Given these differences, Huffman (1972) concludes that there are three, not four, word types in Khmer. These are simple monosyllables, i.e. those with simplex onsets; complex monosyllables, i.e. those with word-initial consonant clusters; and disyllables, which includes both of Henderson’s minor and major disyllables.

(2.4)

Class 3:	[lɛbaːŋ]	/lbaːŋ/	‘to test’
Reduced Disyllable:	[lʌbaːŋ]	/lʌbaːŋ/	‘fence’

By the time the term *sesquisyllable* was introduced by Matisoff (1973), monosyllables with word-initial consonant clusters had begun to be considered a type of sesquisyllable. This is further evident when Thomas (1992) refers to several Khmer examples in his account of Type 1 sesquisyllables, which are characterized by the equivalence of word-initial CɛC- sequences with word-initial CC- sequences.

In addition, only the voiced variants like those in (2.5a) below have traditionally been considered sesquisyllables, while those in (2.5b) are not. There are no accounts of any language in which a voiceless vocalism between two consonants is considered a minor syllable vowel. Therefore, only words with Class 3 word-initial consonant clusters have been considered sesquisyllables, while words with Classes 1 and 2 word-initial consonant clusters have not. However, as this analysis will show, the data in (2.5) are fundamentally of the same structure and should be given a unified account.

(2.5)

a-1)	[mteh]		
		‘pepper’	ម្លើស
a-2)	[mɛ'teh]		
b-1)	[pʰoap]		
		‘attach’	ភ្ជាប់
b-2)	[pʰɛ'tʰoap]		

Following Huffman’s (1972) assessment, I treat minor disyllables and major disyllables as being structurally the same in their phonological forms. As such, in what follows, I consider these both as mid central nuclei of unstressed syllables in disyllables and transcribe them as [ʌ].

However, pursuant with the hypothesis being developed here, when a mid central vocalism appears in monosyllables with word-initial consonant clusters, it will be transcribed as [ɤ] when voiced and as [ɤ̥] when voiceless.

2.3 GOALS AND HYPOTHESES

Drawing on the various descriptions of minor syllables that exist for Khmer, the goal of this chapter is to gain a better understanding of Khmer monosyllables with word-initial consonant sequences. In order to determine the phonetic and phonological nature of MCVEs intervening in word-initial consonant clusters, i.e. transitional or intrusive vocoids, in Khmer, I investigate the duration of word-initial consonant sequences and the duration and formant values of their vocalic transitions. Based on previous descriptions of Khmer consonant clusters, the presence of a transitional MCVE is likely predictable. Although this suggests that it is not lexical, it does not tell us whether the MCVE is epenthetic with an associated gesture or excrescent with no gestural target. However, because sesquisyllables have traditionally been described as different from disyllabic words, in the sense that they do not comprise two full syllables, transitional vocoids are likely substantively different from not only stressed vowels (Figure 2.3) but also underlying vowels which are unstressed in minor syllables (Figure 2.4). These figures provide example spectrograms of both stressed and unstressed phonological vowels as well as schematics of both gestural and voicing configurations.

I hypothesize that the transitional vocoids are the result of gestural underlap – in which the release of one consonantal gesture occurs before the achievement of the target of a following one – and therefore word-initial consonant sequences with MCVEs should be durationally identical to or only slightly longer than word-initial consonant sequences without MCVEs but

shorter than unstressed syllables in disyllabic words, as exemplified in Figure 2.1, and their formant values should differ in predictable ways, as well.

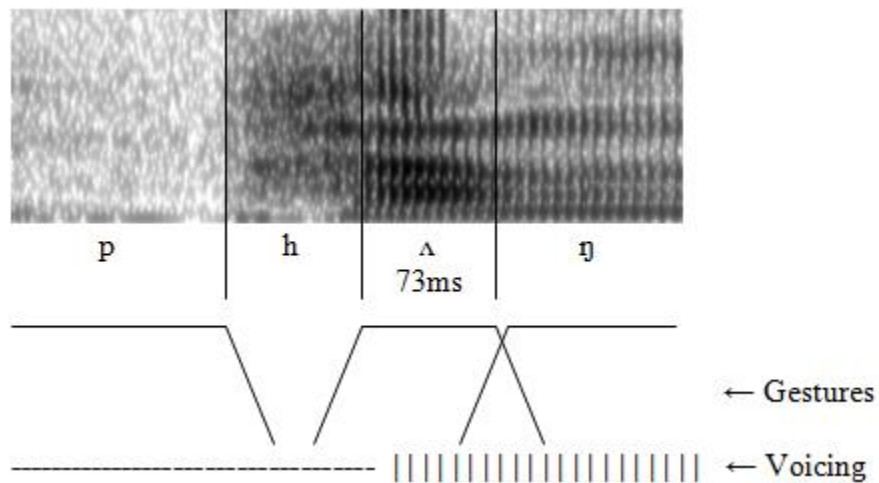


Figure 2.3: [Khmer] Underlying stressed wedge in [phʌŋ]. Total duration = 350ms.

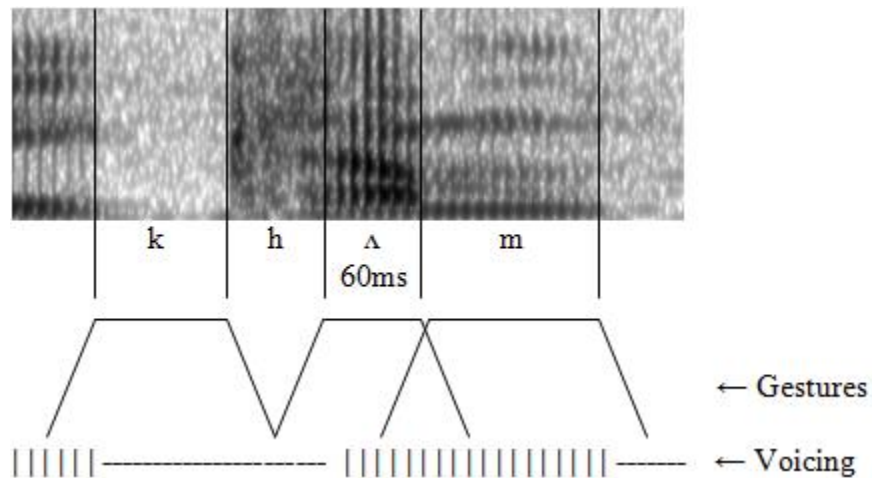


Figure 2.4: [Khmer] Underlying unstressed wedge in [khʌm.khat]. Total duration = 350ms.

There are three possible realizations of the word-initial consonant clusters in Khmer. First, they may be produced with an intrusive vocoid, i.e. [ɤ], between C1 and C2 (Figure 2.5a). This is characterized by a voiced period with formant structure between the two consonants. I will refer to this as voiced underlap. Second, they may be realized with non-harmonic material, i.e. [ɤ̃], between C1 and C2 (Figure 2.5b). This is likely what Huffman (1972) refers to as aspiration and what I will call voiceless underlap. Finally, they could be produced such that no

underlap is discernible (Figure 2.6). In this latter case, which is particularly relevant for sibilant-initial clusters (See Section 2.5.3), an absence of visible underlap may indicate either that there is no separation between consonants, i.e. no underlap, or that underlap *is* present but is being obscured by frication noise from C1.

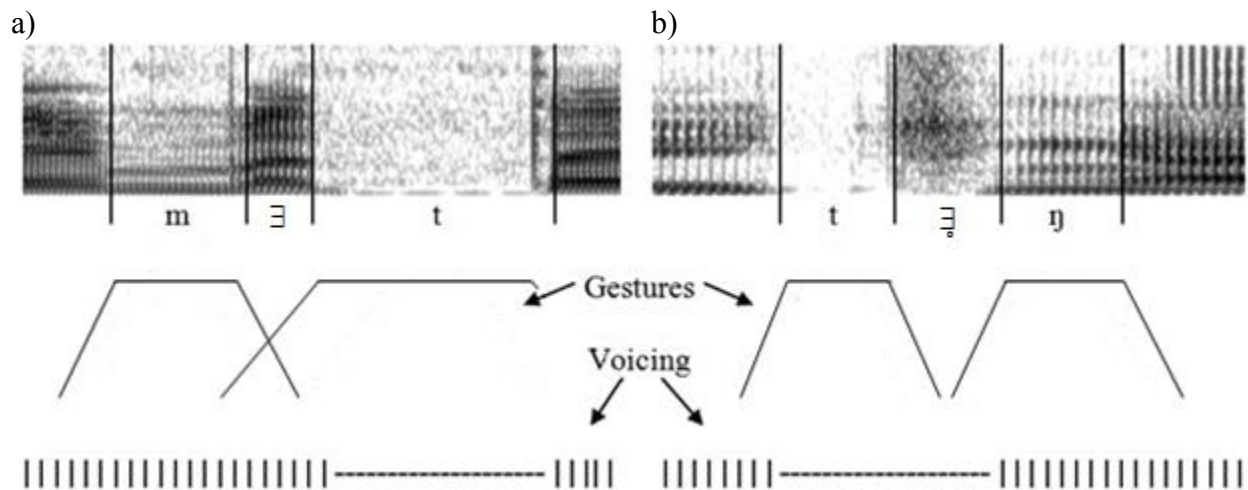


Figure 2.5: [Khmer] Voiced underlap (a) and voiceless underlap (b)

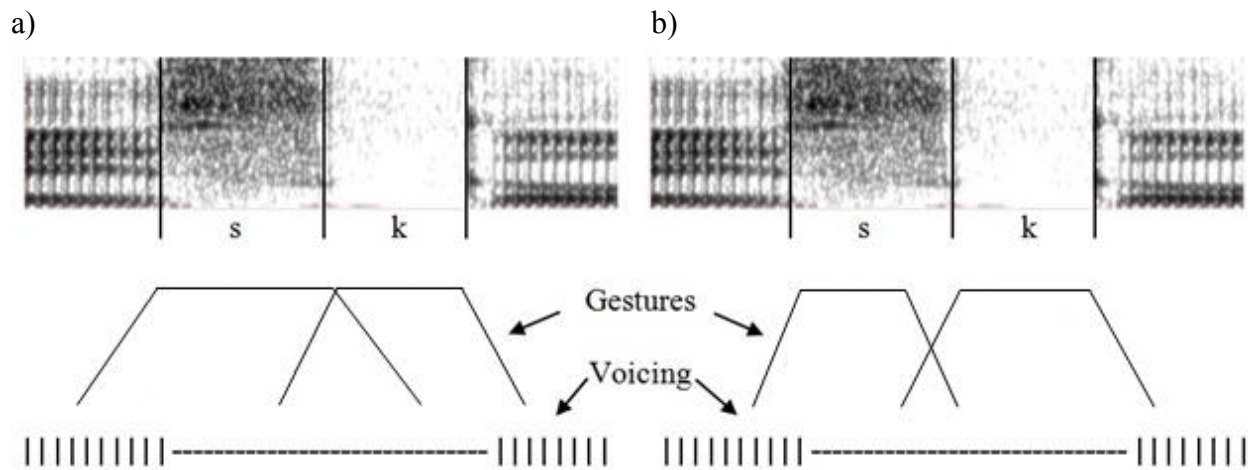


Figure 2.6: [Khmer] Possible gestural configurations for sibilant-initial clusters, without underlap (a) and with underlap (b)

These possibilities lead to several other questions. First, is there a structural difference between [ɤ] and [ɤ̚] in Khmer? In addition, are words with intrusive [ɤ] or [ɤ̚] different from words without visible underlap? What are the conditioning factors for their realizations and

distributions, and what role do manner and place of articulation play? Next, how does underlap, i.e. the transitional vocoid, compare with unstressed lexical wedge in disyllabic words?

Differences between transition states and phonological vowels should be apparent not only in durational differences but also in an analysis of formant values. Transitions are not predicted to have formant targets. Therefore, if the MCVE is a transition, we expect observed formants to differ from formants of underlying wedge or epenthetic schwa. There are at least two ways in which this could manifest. First, intrusive vocoids could be more strongly influenced by their consonantal context than lexical wedges, in which case their formant values will be highly dependent on neighboring consonants and vowels. Second, all else being equal, intrusive vocoids are likely to have a lower F1 than underlying wedges. This is because the tongue will not lower completely between consonant gestures since there is no vocalic target, particularly in the case of coronals. A schematic of what these possibilities might look like is given in Figure 2.7, which is a representation of the vowel space such that the axis values are inverted.

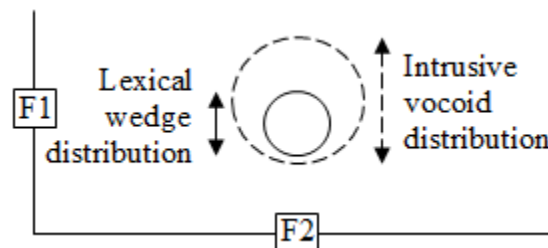


Figure 2.7: Predictions about lexical wedge and the intrusive vocoid

With these predictions in mind, we now turn to the experiment itself, particularly how the hypotheses are tested and how the acoustic results bear on our understanding of the sesquisyllable. The main goals are to determine what the nature of the MCVE in purported sesquisyllables actually is, how it patterns and how it compares to other types of MCVEs.

2.4 METHODS

2.4.1 *Participants*

Participants included eighteen native speakers of Khmer between the ages of 18 and 44 ($\mu = 27$). Although all recordings were made in Phnom Penh, many participants were from other provinces of Cambodia, including Pursat, Siem Riep, Kampot, Kampong Speu, Preah Vihear, Kompong Chhnang, Takeo, Kandal, Battambang and Kompong Cham. Seven of the participants were females and eleven were males.

2.4.2 *Stimuli*

Stimuli were randomized and presented to the participants one word at a time. Participants were instructed to read each aloud three times in the frame sentence [nijj ____ mdəŋ tiət] ('Say ____ one more time.'). Twenty words of type C(ɛ/ɛ̄)CVC were recorded, along with 4 disyllabic C₁C₂.CVC words and 13 monosyllabic C₁C words, as controls (Table 2.5). In monosyllabic words and in the unstressed syllables of disyllabic words, all vowels were phonologically short lexical wedges.

Table 2.5: Khmer wordlist

IPA	Gloss	Orthography	IPA	Gloss	Orthography
CCVC					
pʰoap	attach	ភ្ជាប់	stɨŋ	river	ស្ទឹង
pka	flower	ផ្កា	skoal	acquainted with	ស្គាល់
pnom	mountain	ភ្នំ	smæʌ	grass, hay	ស្មៅ
pɲut	bathe	ផ្ទុក	sɲiam	quiet	ស្ងៀម
præʌ	to use	ប្រើ	mteh	pepper	ម្រូស
tɲaj	day, sun	ថ្ងៃ	mnoah	pineapple	ម្នាស់
tʰha	to fry	ឆា	lkak	noisily	ល្អិត
tʰkaj	dog	ឆ្កែ	lbaijŋ	game	ល្បែង
tʰbah	be clear	ច្បាស់	lmom	sufficient	ល្មម
tʰɲaj	distant	ឆ្ងាយ	lɲajtʰ	afternoon	ល្ងាច
CVC					
pʌr	large lake	បឹង	tʰʌk	peck, nibble	ចឹក
pʌŋ	centimeter	ដង	sam	then, later	សឹម
pʌk	to drink	ផឹក	sʌk	worn out	សឹក
tʰʌb	pinch of something	ចឹប	sʌŋ	almost, practically	សឹង
tʰʌŋ	cook sugar, make syrup	នឹង	lim	vague, blurry	លឹម
min	to patch	មិន	lib	retract, pull in	លឹប
mit	dash away	មីត			
CVC.CVC					
tʌk.kaj	gecko	តឹកកែ	sʌk.pdək	liquor container	សឹកប្លឹក
khʌm.khat	fancy belt	ខឹមខាត់	mat.pot	stretch one's back	មីតពត់

2.4.3 Measurements

Stimuli were recorded on a Dell laptop with a Sennheiser headset microphone at a sampling rate of 44,100Hz, and all measurements were done in PRAAT (Boersma and Weenink 2012). Segmentation was completed by hand using spectrograms and waveforms. Three types of measurements were made: duration, formant values and spectral energy.

To obtain durations, the onset and offset of the second formant were considered to be the beginning and ending of vowels, where possible. Nasal-to-vowel and vowel-to-nasal transitions were demarcated at clear transitions in the amplitude of the second formant.

To calculate formant values, all sound files were downsampled to 10,000Hz and transformed into formant objects via a short-term spectral analysis with a window length of 25ms and a pre-emphasis of 50Hz. For males, the maximum value for three formants was set at 3,400Hz, and for females, 4,000Hz. Formant measurements were taken at vowel midpoints.

Finally, to determine the location of spectral energy in sounds labeled as voiceless sibilants, center of gravity was measured for affricate- and fricative-initial consonant sequences. To do so, spectra were generated using a fast Fourier transform. Sounds were also high-pass filtered at 1,500Hz to rule out any influence from voicing. Center of gravity measurements were taken over 20 equal windows for each sound. Because sounds were all of different lengths, window lengths varied by sound. However, windows for the majority of sounds were between 5ms and 10ms.

To obtain quantitative results, two different methods were used. For statistical analyses, only the second repetition of each word was used. However, to calculate distributional results, the second and third repetitions were both counted in order to make clear the general tendencies in the data, yielding 716 tokens.

2.5 RESULTS

In general the results show that voiced underlap and voiceless underlap, as seen in Figure 2.5 above, are not significantly different in duration, although sonority is shown to have an important effect on the duration of CC sequences. In addition, the voicing of the underlap is found to be predictably dependent on the voicing of C1. These results suggest that both voiced and voiceless underlap are a result of the same type of gestural configuration. In contrast, their durations are much shorter than both unstressed and stressed lexical wedge, supporting the hypothesis that they are in fact transition states. This is further confirmed by formant measurements, which show that vowel qualities differ significantly between voiced underlap and unstressed lexical wedge, suggesting that the latter have associated gestures while the former do not.

Throughout this section, box plots are presented to show the range of various results. The horizontal line in each figure represents the overall mean of all the data presented in that figure. For each category of data presented, the boxes contain the range of data which constitutes the 25th percentile through the 75th percentile. The horizontal line in each box represents the mean value for that category, and the whiskers on the boxes are calculated by the formulae (*1st quartile* – ($1.5 * \textit{interquartile range}$)) and ($3^{\textit{rd}}$ *quartile* + ($1.5 * \textit{interquartile range}$)).

2.5.1 Token Distribution

Two measures are important for the evaluation of consonant sequences: the percentage of consonant sequences that have underlap and the percentage of those instances of underlap which are voiced versus voiceless. First, of 716 C(∅/∅)CVC words (5 were omitted due to error), 442 tokens (62%) have some form of underlap, and 274 tokens (38%) do not. The percentage of instances of underlap in those sequences are given in Table 2.6. Again, the sequences tested in

the present experiment are double-outlined. The percentages range from 0%, i.e. there is no underlap, whether voiced or voiceless, to 100%, i.e. every repetition has some amount of underlap.

Table 2.6: [Khmer] Word-initial C1C2 combinations – Percentage of sequences with underlap

C1 \ C2	s	h	r	l	p	t	tʃ	k	m	n	ɲ	ʔ	b	d	
p	ps	ph	100	pl		pt	92	100		94	pn	94	pʔ		pd
t		th	tr	tl	tp			tk	tm	tn		94	tʔ	tb	
tʃ		6	tʃr	tʃl	tʃp			6	tʃm	tʃn		31	tʃʔ	56	tʃd
k	ks	kh	kr	kl	kp	kt	ktʃ		km	kn	kɲ	kʔ			
s			sr	sl	sp	0		0	6	sn	sɲ	0	sʔ	sb	sd
m	ms	mh	mr	ml		77	mtʃ			86	mɲ		mʔ		
l		lh			lp			97	97			97	lʔ	100	

Next, with respect to voicing, of the tokens with underlap, 123 (55%) have voiced underlap, and 99 (45%) have voiceless underlap, as represented in Figure 2.5 above. The percentage of repetitions with voiced underlap is given in the double-outlined boxes in Table 2.7. Sequences which show no underlap, i.e. those represented by 0% in Table 2.6, are labeled with an X here since they obviously have neither voiced nor voiceless underlap. While the percentage of voicing also ranges from 0% to 100%, the voicing results are much more categorical. That is, with only two exceptions ([pn] and [tʃb]), underlap for a given sequence is either always voiced or always voiceless.

Table 2.7: [Khmer] Word-initial C1C2 combinations – Percentage of underlap tokens with voiced underlap

C1 \ C2	s	h	r	l	p	t	tʃ	k	m	n	ɲ	ʔ	b	d
p	ps	ph	100	pl		pt	0	0		3	pn	0	pʔ	pd
t		th	tr	tl	tp			tk	tm	tn		0	tʔ	tb
tʃ		0	tʃr	tʃl	tʃp			0	tʃm	tʃn		0	tʃʔ	75
k	ks	kh	kr	kl	kp	kt	ktʃ		km	kn	kɲ	kʃ		
s			sr	sl	sp	X		X	0	sn	sɲ	X	sʔ	sb
m	ms	mh	mr	ml		100	mtʃ			100	mɲ		mʔ	
l		lh			lp			100	100			100	lʔ	100

With the exception of [pr] sequences, for sequences in which C1 is a non-sibilant obstruent, i.e. [p] or [t], 95% have some form of underlap. Of those 95% of sequences, 99% have voiceless underlap. Overall, when C1 is voiceless, 94% of sequences have voiceless underlap, and <1% have voiced underlap, which results from one repetition of [pn]. The voiced material in [pr] sequences is likely a result of the articulation of [r] which is realized as a tap or trill.⁴ Complementarily, for sequences in which C1 is voiced, i.e. [m] or [l], 93% have underlap. Of that 93%, 100% are voiced. In other words, voiceless underlap never occurs when C1 is voiced. These results indicate that the presence and type of underlap is highly contingent on the voicing and manner of C1.

⁴ Kirby (To appear) finds that C2 [r] is devoiced in some utterances. Whether or not [ɻ] is still present in these cases requires further investigation.

Finally, of sequences with sibilant C1s, i.e. [s] or [ʃ], only 10% appear to have underlap. Of those 10%, 58% have voiced underlap and 42% have voiceless underlap. Overall, of sequences in which C1 is a sibilant, 4% show voiceless underlap and 6% have voiced underlap. However, more than $\frac{3}{4}$ of these sequences with underlap – 77% – are [tʃb] sequences, suggesting that not only the voicing but also the manner of C2 affects underlap. As will be shown in Section 2.5.3, center of gravity measurements suggest that underlap is probably present in many more affricate-initial tokens than these cursory results suggest, but not in [s]-initial tokens.

2.5.2 *Non-sibilant C1s*

Voiced underlap and voiceless underlap vary not only by their context but also by their durational distributions. Figure 2.8 shows that voiceless underlap has a wider range of possible durations than does voiced underlap.

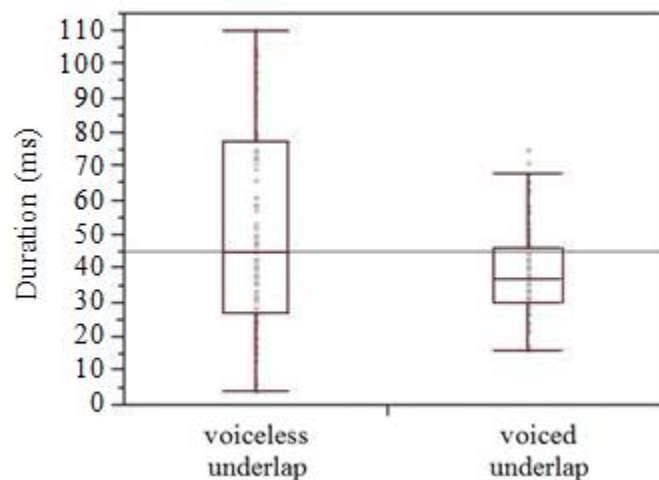


Figure 2.8: [Khmer] Distributions of underlap durations (voiceless and voiced)

To further explore this distributional difference in duration, comparisons between the durations of CC sequences with voiced underlap and voiceless underlap are necessary. However, because of the near-complementary relationship of the C1 contexts of voiced underlap and voiceless underlap, making meaningful comparisons across types, i.e. $C\exists CVC$ and $C\exists CVC$, is

not possible. Because of inherent durational differences between voiceless stops ([p] and [t]) and sonorants ([l] and [m]), there are two possible sources of durational variation in CC sequences. In other words, there are both durational differences between consonant types and possible durational differences between underlap types. Therefore, the variation of C1 and C2 types was removed by calculating the residuals of a regression of the total duration (or the underlap duration) with C1 and C2 types and Speaker as a random variable (2.6). Subsequently, those residuals were used to make interpretable comparisons.

(2.6)

$$\text{a) Total Duration} = \text{C1 Type} + \text{C2 Type} + [\text{Speaker}] + \varepsilon_{\text{TotalDur}}$$

$$\text{b) Underlap Duration} = \text{C1 Type} + \text{C2 Type} + [\text{Speaker}] + \varepsilon_{\text{UnderlapDur}}$$

Results show that once the variation in C1 and C2 type is removed via the residuals, underlap type, i.e. voiced underlap versus voiceless underlap, is not significantly correlated with the total duration ($p = 0.9934$) or with underlap duration ($p = 0.7565$). This indicates that neither the total duration of the CC sequences nor the duration of underlap alone (whether voiced or voiceless) is correlated with underlap type, suggesting that voiced underlap and voiceless underlap are not intrinsically distinct. Nonetheless, the total duration is correlated with the underlap duration ($p < 0.0001$), suggesting that the duration of the underlap contributes to the entire duration of the CC sequence.

2.5.3 *Sibilant C1s*

The lack of durational differences seen in Figure 2.8 again confirms the similarity in behavior of voiced underlap and voiceless underlap. However, it is not possible to directly test the durations of the consonants in the tokens with underlap against the tokens without visible underlap, again because of the near mutual exclusivity of their consonant types. In other words,

tokens with underlap generally have stop or liquid C1s, and stops with no distinguishable underlap have sibilant C1s. However, when word type – underlap versus no underlap – is regressed by the residuals of the total duration, as in (2.6) above, results show that the presence or absence of underlap is not correlated with the total duration of the sequence ($p = 0.1263$), indicating that sequences which clearly exhibit underlap are not significantly longer than those that do not, suggesting that the consonant gestures are actually shorter when underlap is present. This suggests that there might actually be underlap in the affricate- and fricative-initial sequences, which is obscured by the frication noise, such that there is no visible correlate of underlap on a spectrogram.

A number of studies have used a measurement of the center of gravity – or the locus of spectral energy – to determine differences in voiceless sounds (cf. Tsuchida 1994, Gordon et al. 2002, Niebuhr et al. 2011, *inter alia*). Based on these studies, we can make at least three predictions about the center of gravity as it relates to voiceless underlap. First, while voiceless underlap does not have an associated gesture, it should nonetheless have a concentration of energy at some frequency, depending on the position of the tongue in the mouth. Second, the center of its energy may differ from that of preceding fricatives; however this difference may not be observable by eye on a spectrogram. Third, if underlap is present in what appear to be C1 sibilants in CC sequences, its center of gravity should follow a different trajectory from that of simplex sibilant onsets to monosyllables. Therefore, I measured the center of gravity at 20 equidistant points throughout the duration of sibilants in both contexts.

In fact, measurements show that at least in the case of the palatal affricate, underlap appears to be present in some sibilant-initial consonant sequences. Euclidian distance from the peak intensity to the end of the fricative noise is significantly different for the sound labeled as

C1 [tʃ] in $\widehat{tʃ}VC$ words than the simplex onset of $\widehat{tʃ}VC$ monosyllables ($p < 0.0001$). The much more gradual slope toward the end of the sound, as seen in Figure 2.9, indicates a transitory underlap period. However, this difference is neither visible nor statistically significant for [s] in the same contexts ($p = 0.2828$), as see in Figure 2.10.

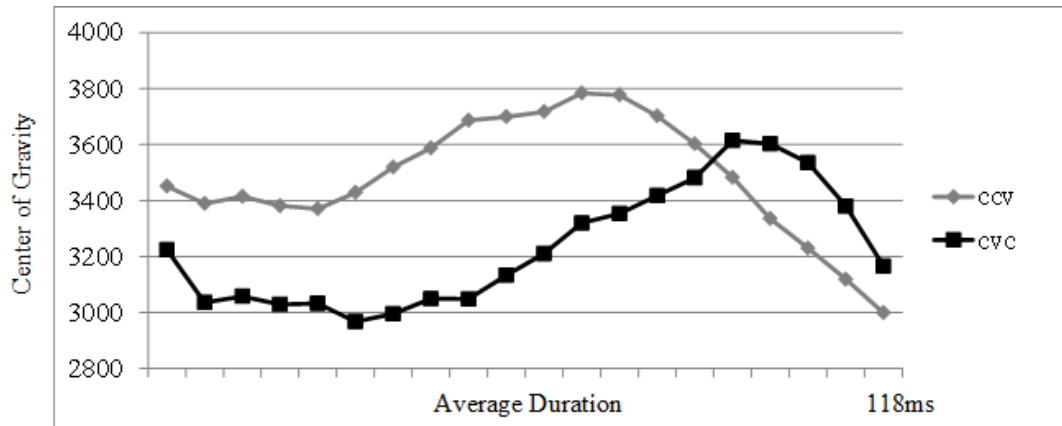


Figure 2.9: [Khmer] Center of gravity measurements in Hz for [tʃ]

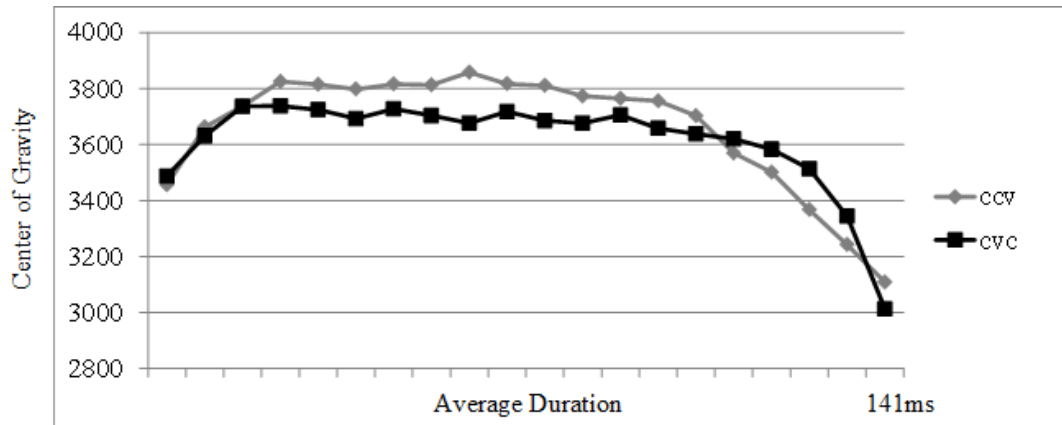


Figure 2.10: [Khmer] Center of gravity measurements in Hz for [s]

These results follow from the distributional results given in Table 2.6 and Table 2.7.

Affricate-initial clusters were shown to have visible voiced underlap in some cases, as opposed to [s]-initial clusters, which demonstrated almost no overlap across the board, as summarized in Table 2.8.

Table 2.8: [Khmer] Percentage of $[\widehat{tj}]$ - and $[s]$ -initial tokens with underlap visible on a spectrogram

$[\widehat{tj}\text{-h}]$: 6	$[s\text{-t}]$: 0
$[\widehat{tj}\text{-k}]$: 6	$[s\text{-k}]$: 0
$[\widehat{tj}\text{-ŋ}]$: 31	$[s\text{-m}]$: 6
$[\widehat{tj}\text{-b}]$: 56	$[s\text{-ŋ}]$: 0

For the $[\widehat{tj}]$ -initial clusters, the largest percentages of voiced underlap are present in clusters with a voiced C2, suggesting that, contrary to the non-sibilant-initial clusters, C2 does contribute some voicing to the transition period. Still, given the similarities in duration between $C\exists/\exists C$ and CC sequences, we may expect that underlap is present in the latter, even when it is not voiced. This is exemplified in Figure 2.11 and Figure 2.12 below. Figure 2.11 shows a $[\widehat{tj}b]$ cluster with voiced underlap, and Figure 2.12 shows a $[\widehat{tj}b]$ cluster without voiced underlap. In the latter, the suspected voiceless underlap portion is indicated by dashed lines.

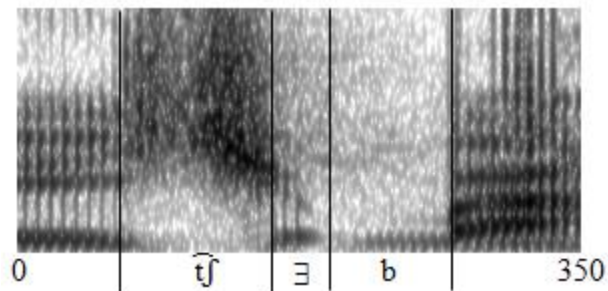


Figure 2.11: [Khmer] Voiced underlap in /t̂jah/ 'be clear'. Duration in ms.

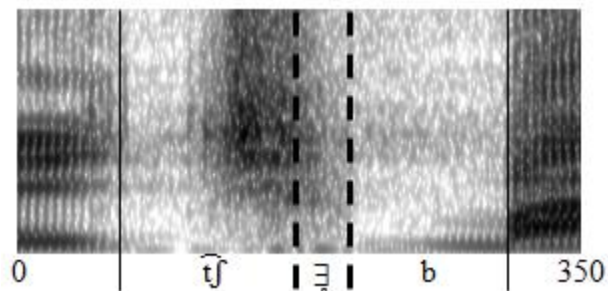


Figure 2.12: [Khmer] Possible voiceless underlap in /t̂jah/ 'be clear'. Duration in ms.

Finally, Figure 2.13 shows that the center of gravity measurements differ based on the place of articulation of C2, which is indicative of anticipatory coarticulation. In particular, when C2 is velar, the center of gravity peaks more quickly and drops off more gradually. However, all four cluster types have a similar trajectory to the overall average presented in Figure 2.9, as opposed to simplex onset $[\widehat{tj}]$.

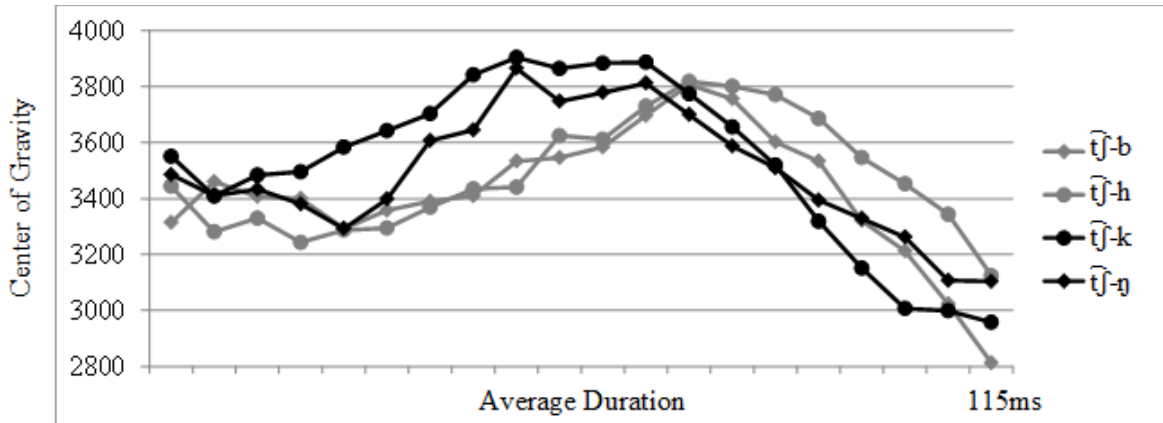


Figure 2.13: [Khmer] Center of Gravity in Hz for $[tj̄]$ in four clusters

While only an articulatory experiment can provide definitive results regarding voiceless underlap in sibilant-initial clusters, the durational and center of gravity results above suggest that voiceless underlap is likely to be present in at least some of the $[\widehat{tj}]$ -initial clusters. However, this does not seem to be the case for $[s]$ -initial clusters, which I suggest show no evidence of underlap.

2.5.4 *Sonority and Place of Articulation*

Because there is no correlation between underlap type and duration of CC sequence nor between underlap type and underlap duration, the large distributional difference in duration between voiced underlap and voiceless underlap, as seen in Figure 2.8, remains unaccounted for. Therefore the data are also analyzed according to sonority relationships. This is motivated by the cross-linguistic frequency in which epenthesis occurs to resolve sonority violations

(Fleischhacker 2001), as well as by the noted articulatory difficulties of producing clusters which match in sonority in addition to oral gesture (Gafos 2002).

Results show that consonant sequences which differ in sonority, i.e. whether obstruent or sonorant, are significantly longer than sequences that agree in sonority. In other words, Figure 2.14 and Table 2.9 show that S̄EO and ŌS̄ sequences are significantly longer than S̄ES and ŌEO sequences ($p < 0.0001$). In addition, OS sequences with no underlap are longer than OO sequences with no underlap ($p < 0.0001$). Indeed, even OS sequences with no underlap are significantly longer than ŌS̄O and S̄aS̄ sequences with underlap ($p < 0.0001$).

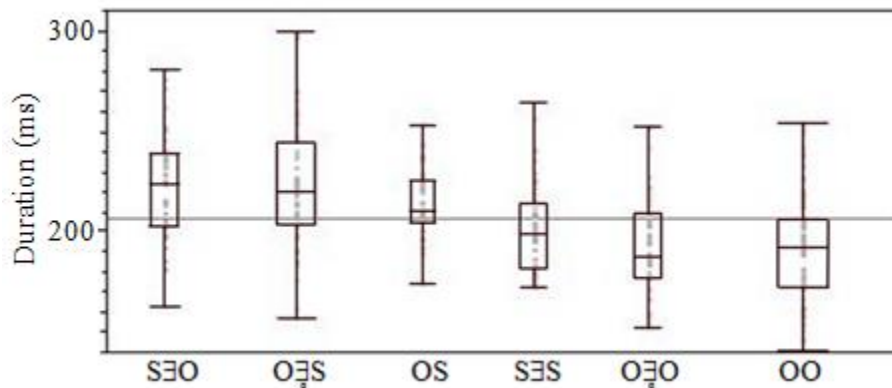


Figure 2.14: [Khmer] C(̄/̄)C durations by sonority type

Table 2.9: [Khmer] CC durations (maximum, median and minimum) by sonority type

Time (ms)	S̄EO	ŌS̄	OS	S̄ES	ŌEO	OO
Maximum	281	300	253	265	252	254
Median	224	220	211	199	187	192
Minimum	163	157	174	172	152	141

The result that sequences differing in sonority are longer than sequences that are alike in sonority is somewhat unexpected. Durationally longer sequences suggest that more gestural underlap is present. Because longer underlap is often motivated by the need for perceptual

recoverability, it is surprising that an obstruent-sonorant sequence (e.g. [pŋ]), which obeys the sonority sequencing principle (Clements 1990), is longer than an obstruent-obstruent sequence (e.g. [pk]). However, given the typological rarity of stop-nasal clusters, this could be the result of a more difficult articulation in a way that other obstruent-sonorant clusters, like stop-[l] clusters for example, would not be.

These sonority results prompt further investigations of two additional types. First, they merit a more detailed comparison of the sub-parts of CC sequences. In particular, is there any sort of durational trading relation between the duration of the consonants and the underlap? Second, what does it mean in terms of syllabification to say that CC sequences that differ in sonority are longer than CC sequences that are alike in sonority? Both of these types were compared to disyllabic words as a control.

Regarding subparts of the CC sequence, because the duration results above suggest that minor syllable “vowels” in Khmer are underlap, and because by definition underlap does not have an associated gesture but is instead the result of gestural spreading, we expect that there is no correlation between the duration of one or both of the consonants with the underlap. To test this, the data were separated by obstruent-sonorant type, and the underlap duration was compared to C1 and C2 for each of the four token types: SΞS, SΞO, OΞO and OΞS. None of the types show a significant correlation between underlap duration and C1 or C2 duration, except for the OΞS sequence. However, as can be seen in Figure 2.15, the variance is so large (C1: $R^2 = 0.17$; C2: $R^2 = 0.09$), that the significance of the correlation is not meaningful. In other words, on the surface, underlap duration and consonant duration do not seem to be inversely related. Note, however, that it is also possible that they are actually inversely correlated but that the correlation

is being counteracted by a positive correlation between duration and speech rate, such that no correlation – negative or positive – is apparent.

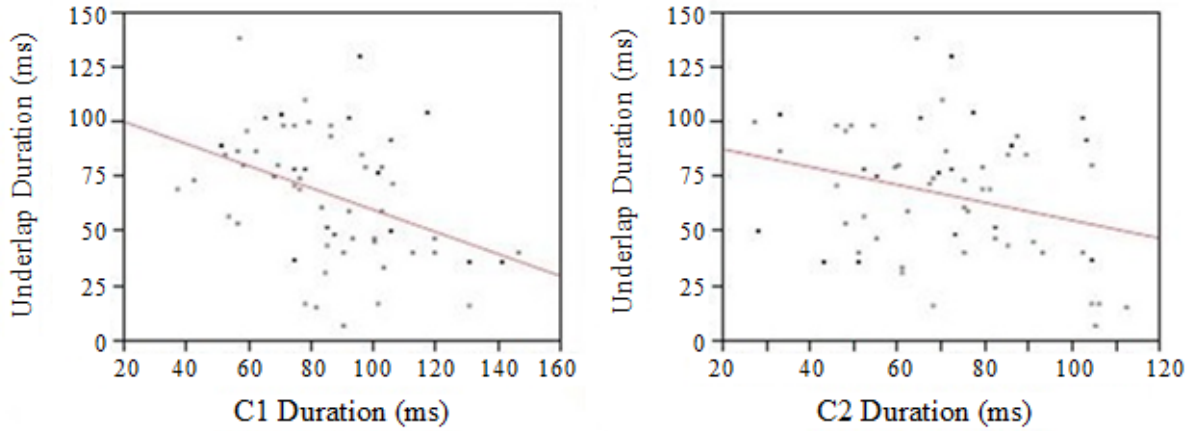


Figure 2.15: [Khmer] Relationship between underlap and consonants in OΞS sequences

Although these results show that there is no difference in how the internal parts of the CC sequence relate to each other, the result that sequences differing in sonority have a longer duration than sequences similar in sonority is still striking. To suggest that speakers syllabify these forms differently (as monosyllables versus disyllables) according to their sonority relations is quite odd, as is the particularly long duration of the OS sequence. Therefore the durations of CC sequences with both types of underlap and without underlap were compared to CΛC sequences in unstressed syllables of disyllabic words (e.g. [mΛt.'pot]). Results show that the total durations of OΛO, OΛS and SΛO sequences in the first half of disyllables are significantly longer ($p < 0.0001$) than the matching obstruent-sonorant sequences in word-initial CC sequences as seen in Table 2.10 (No SΛS syllables were recorded, so a comparison was not possible).

Table 2.10: [Khmer] Durations (ms) of C(ɤ/ɛ)CVC and C_ΔC.'CVC sequences for OO, OS and SO sequences

	OO		OS		SO	
	C(ɤ/ɛ)CVC	C _Δ C.'CVC	C(ɤ/ɛ)CVC	C _Δ C.'CVC	C(ɤ/ɛ)CVC	C _Δ C.'CVC
Minimum	141	219	157	245	163	191
Median	190	309	219	330	224	288
Maximum	254	516	300	488	281	371

In addition, durations of ɤ and [Δ] in these environments was also tested. For all conditions, [Δ] duration is significantly longer than underlap duration (OO and OS, $p < 0.0001$; SO $p = 0.0071$), as seen in Table 2.11. These results suggest that monosyllables with word-initial consonant clusters should not be grouped with disyllabic words, whether they have underlap or not.

Table 2.11: [Khmer] Durations (ms) of C(ɤ/ɛ)CVC and C_ΔC.'CVC sequences for OO, OS and SO sequences

	OO		OS		SO	
	C(ɤ/ɛ)CVC	C _Δ C.'CVC	C(ɤ/ɛ)CVC	C _Δ C.'CVC	C(ɤ/ɛ)CVC	C _Δ C.'CVC
Minimum	4	31	7	94	26	31
Median	27	58	72	134	40	48
Maximum	66	108	138	178	65	106

2.5.5 Effects of Place on Underlap

In addition to these sonority effects, place of articulation may also play a role in underlap duration. Indeed, Chitoran et al. (2002) show that because of perceptual recoverability, front-to-back sequences have more overlap than back-to-front sequences in Georgian. However, a simple

inverse relationship does not hold between overlap and underlap. As laid out in Figure 2.16, more overlap is not equivalent to less underlap. Whether two gestures are very overlapped or only slightly overlapped, neither has any amount of underlap. Therefore, in terms of perceptual recoverability, degree of overlap should not be conceived of as a negative degree of underlap. As a consequence, we cannot predict that the inverse of Chitoran et al.'s (2002) claim – that back-to-front sequences will have more underlap – will hold true.

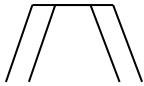
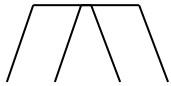

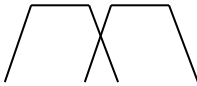

				
More overlap	Less overlap	No overlap No underlap	Less underlap	More underlap

Figure 2.16: Degrees of overlap and underlap

If back-to-front sequences did have more underlap than front-to-back sequences, we expect them to be longer, but a linear regression of the total duration of consonant sequences shows that back-to-front sequences are not significantly longer than front-to-back sequences ($p = 0.8593$). Furthermore, the duration of underlap alone is actually longer for front-to-back than back-to-front sequences ($p = 0.0132$). Although this result requires more investigation, it minimally shows that place relations do not have the same effect on underlap durations as they do on overlap durations. This is a reasonable result in terms of perceptual recoverability. When sounds are overlapped to some degree, perception can be affected because the offset of one gesture may obscure the onset of the next, but if two sounds are underlapped, neither has any amount of obstruction, no matter what the degree of underlap, although having only a very small degree of underlap may also jeopardize recoverability.

2.5.6 Formant Analysis

As we expect place of articulation to have a greater effect on the vowel formants than manner or voicing, all C1s and C2s were grouped according to their place of articulation to perform the formant analysis. Again because not all of the consonant environments were recorded for all the word types, just as for the duration measurements, the consonantal differences were accounted for by means of residuals, as seen in (2.7).

(2.7)

$$\text{a-1) } F1 = \text{C1 Type} + \text{C2 Type} + [\text{Speaker}] + \varepsilon_{F1}$$

$$\text{a-2) } F2 = \text{C1 Type} + \text{C2 Type} + [\text{Speaker}] + \varepsilon_{F2}$$

$$\text{b-1) } \varepsilon_{F1} = \text{Word Type} + [\text{Speaker}]$$

$$\text{b-2) } \varepsilon_{F2} = \text{Word Type} + [\text{Speaker}]$$

Results show that F1 and F2 of the intrusive vocoid pattern differently than underlying stressed and unstressed wedge. This is demonstrated in the vowel plot in Figure 2.17 below, which shows the values for [ɜ], stressed [ʌ] and unstressed [ʌ] following alveolar and labial C1s. Indeed, these results match the predicted MCVE patterns almost exactly as predicted. F1 for unstressed [ʌ], which I interpret as a Type B MCVE, is both equivalent to and higher than F1 of stressed [ʌ], for alveolar and labial C1s respectively. Moreover, F1 of [ɜ] is largely dependent on the place of the preceding consonant. When following alveolars, it is pulled higher and more forward in the mouth. When following labials, however, it is being pulled downward. Its backness is less easily explained, but more data are needed to confirm these results.

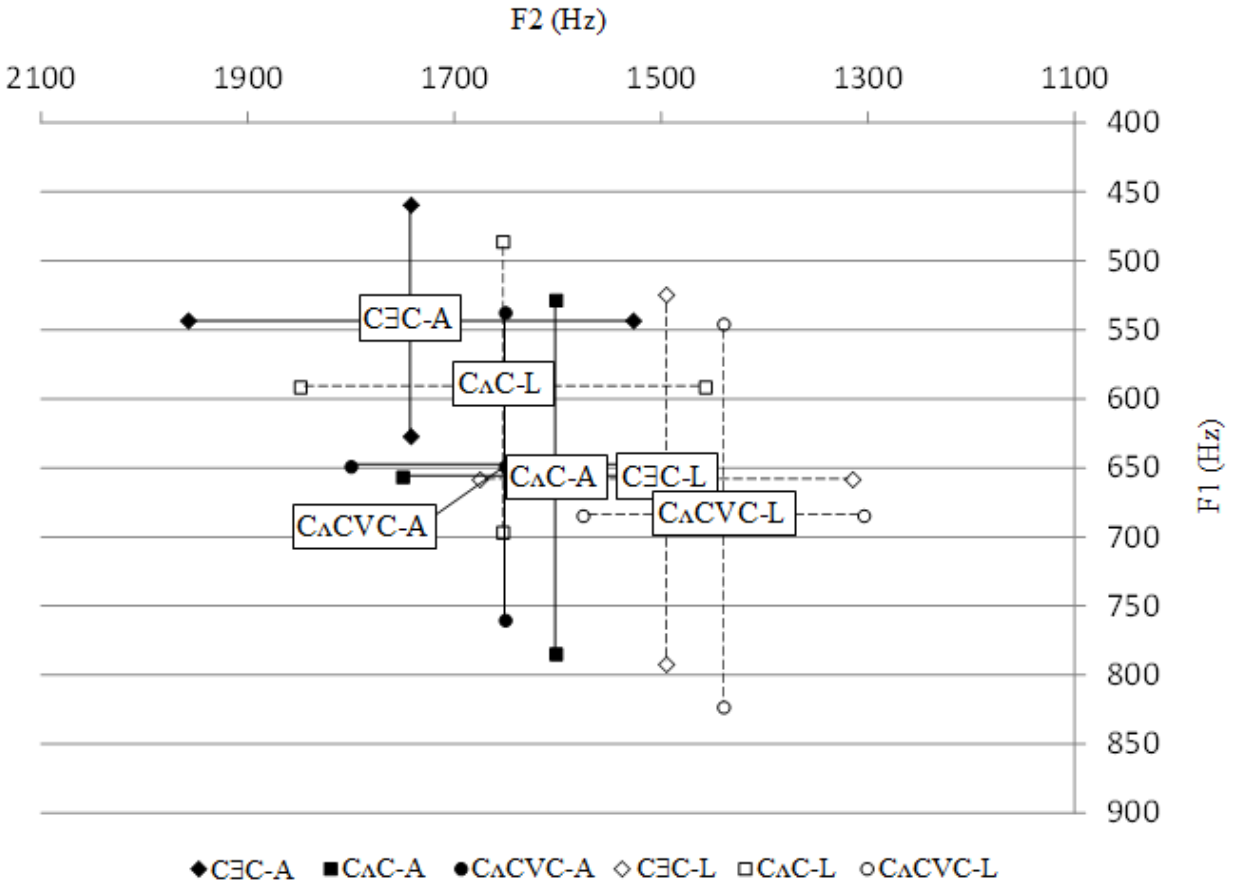


Figure 2.17: [Khmer] Vowel plot of average values and standard deviations for the intrusive vocoid (◊), stressed wedge (◻) and unstressed wedge (◊) following alveolar (A) and labial (L) consonants

An investigation of a subset of consonant types shows that for alveolar-velar CC pairs (chosen because they are present in all word types), the differences between the intrusive vocoid and underlying wedge are significant ($p \leq 0.0009$). The formant values in Table 2.12 suggest that [ɛ] is produced significantly higher and more forward in the mouth than both underlying unstressed and stressed wedge for both males and females (Statistical significance is represented by double solid lines). In addition, for males, F2 for underlying unstressed wedge is also significantly higher than stressed wedge.

Table 2.12: [Khmer] Median formant values (Hz) for female and male productions of C \exists CVC, C Δ C.'CVC and C Δ C, where C1 is alveolar and C2 is velar. Significant differences are represented by double solid lines.

	Female			Male		
	C \exists CVC	C Δ C.'CVC	C Δ C	C \exists CVC	C Δ C.'CVC	C Δ C
Median F1 (Hz)	562	771	821	477	651	639
Median F2 (Hz)	1976	1673	1714	1825	1588	1505

These formant results agree with our predictions about the differences between the excrescent vocoid and underlying wedge laid out in Chapter 1 and presented again here as Figure 2.18. The F1 results suggest that the tongue does not lower between consonants, which I take to be evidence that it lacks a gestural target. This is further supported by F2 measurements, which indicate that the intrusive vocoid is more susceptible to influence from surrounding consonants, which I again attribute to its lack of an associated gesture and target.



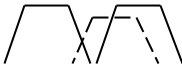

MCVE Type		Gesture	Target Attainment	Gestural Representation	Phonetic Properties
	No MCVE	✗	N/A		N/A
A	Excrescent \exists (gestural underlap)	✗	N/A		Shorter Duration Lower F1
B	Partial \varnothing	✓	✗		Shorter Duration Equal or higher F1
	Δ with undershoot				
C	Full \varnothing	✓	✓		Longer Duration
	Underlying Δ				

Figure 2.18: Possible gestural representations of MCVEs

Finally, before moving on to the discussion of the data, I suggest that all of these results are also applicable to the consonant sequences that were not tested. As indicated above, only a small subset of word-initial consonant sequences were investigated. These particular sequences were selected because the consonant combinations were also found in CVCVC and CVC words which could be used to make comparisons. Because the patterns found here are so robust, I propose that if the remaining consonant sequences are investigated in future work, these results would be supported. In other words, sequences with C2 [r] are likely to show some voiced underlap, along with sequences with C1 [m] and [l]. Still other sequences that begin with voiceless stops will most likely display voiceless underlap.

2.6 DISCUSSION

Although its presence is variable, the contexts in which the Khmer intrusive vocoid appears are highly predictable. In particular, when C1 is voiced, only [ɛ] may appear, and when C1 is voiceless, only [ɛ̥] may appear. While the above results have shown that voiced underlap and voiceless underlap are durationally indistinguishable from each other, they are significantly shorter than unstressed syllables with lexical wedge. Formant values also suggest underlap is more variable and qualitatively different than lexical unstressed wedge. These results together indicate that Khmer “sesquisyllables” are better understood as monosyllables with word-initial consonant clusters that have gestural underlap, instead of a separate word type, as I have suggested throughout. As indicated in the schema in Section 2.1, monosyllables with word-initial consonant clusters can only contain Type A MCVEs.

The Khmer results also suggest a further revision of our model of sesquisyllables. In particular, an accurate model should include both the potential variability in the presence of the excrescent vocoid as well as the predictable variation in the voicing of underlap. Note that while

the variability in the presence of the excrescent vocoid is generalizable across languages, the particular conditioning for the voicing of underlap is particular to Khmer (cf. Ridouane and Fougeron (2011) for a different type of conditioning in Tashlhiyt Berber). These revisions are reflected here schematically in Figure 2.19.

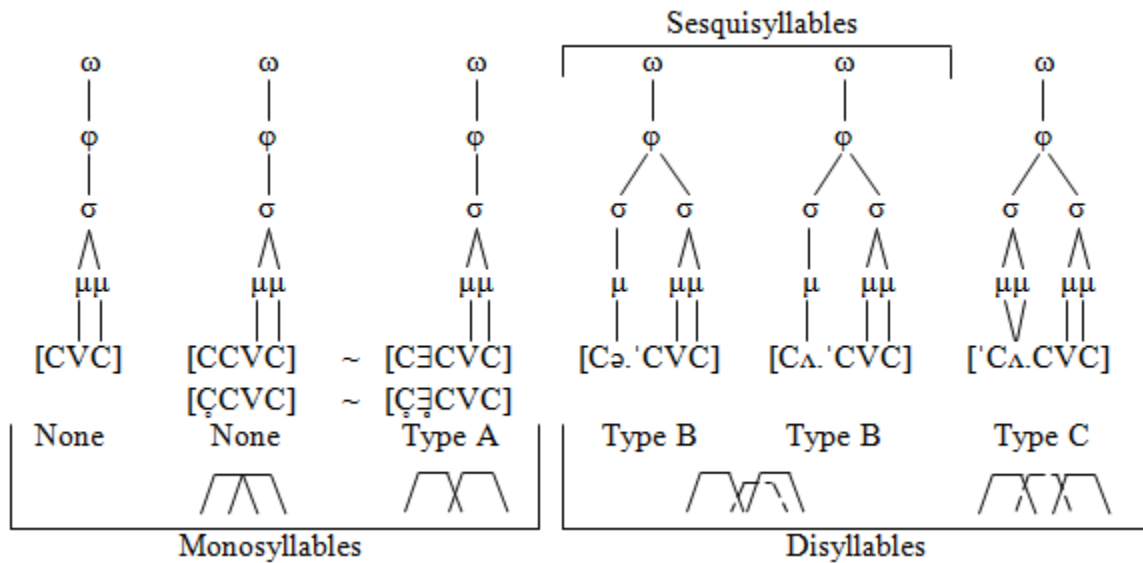


Figure 2.19: [Khmer] Prosodic structure and MCVE types of monosyllables and disyllables

Traditionally, only those forms with voiced underlap (not voiceless underlap) have been considered sesquisyllables. Consonant sequences with voiceless underlap or without audible underlap would not have been interpreted as minor syllables since they have no vowel-like material. In other words, what I have classed together as monosyllables with Type A MCVEs, i.e. [C \exists CVC] and [C \exists CVC], are considered distinct in older accounts of sesquisyllables and would not have been categorized as one entity, the former being considered a sesquisyllable and the latter a monosyllable because excrescent [\exists] was identified as being equivalent with epenthetic [ə] and even unstressed / Δ /, while excrescent [\exists] was only considered aspiration, distinct from [\exists].

However, duration measurements and the predictability of the voicing context suggest that voiced underlap and voiceless underlap are both the result of underlap, where the voicing between C1 release and C2 constriction in most cases derives from that of C1. In addition, because even sibilant-initial tokens without any visible underlap are durationally indistinguishable from the underlap tokens, I suggest that at least affricate-initial sequences also have underlap which is obscured by frication noise, a claim which is supported by the spectral distribution of their frication noise, and which can be tested further through articulatory means. These findings suggest that not only is the schematization in Figure 2.19 a correct characterization of sesquisyllables but also that understanding minor syllables in terms of gestures is a more accurate and informative way to conceptualize the sesquisyllable in Khmer and potentially in other languages as well.

Articulation of Khmer monosyllables can be modeled as gestural coordination relations in which onset clusters are conceptualized in terms of alignment constraints on gestures. Gafos (2002) presents such a model for MCVE excrescence in Moroccan Arabic, making use of landmarks, which are understood as points throughout the temporal span of a gesture, and which are aligned with landmarks of a different gesture to effect particular coordinations. Landmarks include the gestural onset of movement, the target (i.e. the point at which the gesture attains the target), the C-center (or the mid-point of the gesture) and the release (Figure 2.20).

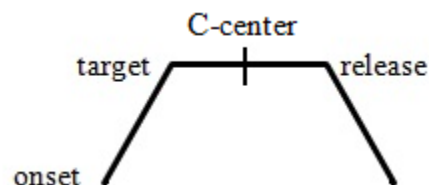


Figure 2.20: Gestural landmarks

The C-center effect is a well-documented phenomenon in Articulatory Phonology. Onset consonants are timed with vowel gestures, such that they are in a 0°, or in-phase, relation with

each other. Consonants, however, repel one another such that they are in a 180°, or anti-phase, relation. When multiple consonants occur in an onset cluster and both of them are phased to the following vowel, these phasing relations compete, yielding the C-center effect (Goldstein et al. 2007). Languages like English (Marin & Pouplier 2008), French (Kühnert et al. 2006) and, for some speakers, Georgian (Goldstein et al. 2007) exhibit the C-center effect, by which the middle point of the consonant gestures is timed in-phase with the vowel gesture. Other languages, however, have been documented to lack C-center effects and exhibit simplex timing instead. Goldstein et al. (2007) show that multiple word-initial pre-vocalic consonants in Tashlhiyt Berber do not function as onsets and therefore multiple consonants in word-initial position have no effect on the timing relation between the rightmost consonant and the following vowel. Hermes et al. (2008) also show that non-sibilant clusters in Italian display the C-center effect while sibilant-initial clusters do not. See Tilsen et al. (2012) for a full discussion of the differences between C-center timed languages and complex onset-timed languages.

Whether or not Khmer is a C-center/complex onset-timed language or a simplex-timed language is not a question that can be answered with the acoustic data presented here. In other words, we cannot tell whether C1 is timed in-phase with the following vowel or not. Indeed, in either case, excrescent vocoids may be present, whether because C1 and C2 gestures do not overlap at all (in a simplex-timed language) or because the strength of the repellant C1~C2 relation is stronger than that of the C1~V and C2~V relations. As a question for further research, comparing underlap durations for Khmer two-consonant clusters with underlap durations for three-consonant clusters may shed some light on this issue. Until that time, we can speculate as to what might be the most appropriate representation for complex onset-timed languages.

Gafos' (2002) model provides a way to formally state gestural timing relations by use of alignment constraints, like those laid out in McCarthy and Prince (1993b), whereby a landmark of one gesture is aligned with a landmark of a second gesture. These alignment constraints are evaluated gradiently, and the degree of the violation is measured from one landmark to the next. Crucially, the distances from onset to target attainment and from target attainment to release are considered equivalent. However, this equivalency is stipulative and probably not accurate. Indeed, if this model is to account for variability in the presence of excrescent vocoids based on speech rate, it must allow for variability in the duration of inter-landmark spans of a gesture. Even if gestures are aligned in the same way, such that the C-center of C1 is aligned with the onset of C2, for example, variation in speech rate may serve to compress or extend gestures so that portions of underlap or overlap are present to varying degrees, some of which will result in an audible transition state and some of which will not (Figure 2.21).



Figure 2.21: Schema of G2 onset alignment with G1 C-center, with underlap (left) and without underlap (right)

In conclusion, this study has shown that distinction between monosyllables and disyllables in Khmer is more clear-cut than previously thought. Although the issue of reduction in Henderson's (1952) minor disyllables still needs to be addressed, I have demonstrated that "extended" or "complex" monosyllables, i.e. monosyllables with word-initial consonant clusters, are indeed monosyllables. Given that our definition of the sesquisyllable presented in Chapter 1 states that sesquisyllables are maximally disyllabic iambs, then we must conclude that Khmer

“sesquisyllables”, and by extension all sesquisyllables with Type A MCVes, do not fit our criterion of sesquisyllabicity. Next, we turn to another Mon-Khmer language – Bunong – which is also claimed to be sesquisyllabic, although descriptions of sesquisyllables in Bunong suggest they are distinct from Khmer monosyllables in that they more similar to disyllables with Type B MCVes, which are phonological and have an associated gesture.

CHAPTER 3

THE PHONOLOGICAL REALITY OF MINOR SYLLABLES: PHONETIC EVIDENCE FROM BUNONG

3.1 INTRODUCTION

Bunong (also called Phnong or Mnong) is a Mon-Khmer language spoken in eastern Cambodia and in the Central Highlands of Vietnam (Figure 3.1). It is spoken by about 52,000 people worldwide (Lewis et al. 2013) and is considered vulnerable. Like many Mon-Khmer languages, Bunong has been claimed to have sesquisyllables, most recently by Phaen et al. (2012). Unlike the set of words addressed in the preceding Khmer chapter, descriptions of minor syllables in Bunong suggest they contain Type B MCVes (/ʌ/ or [ə]) instead of Type A MCVes.



Figure 3.1: Location of Bunong speakers

The goal of this chapter is to provide a better understanding of minor syllables in Bunong and to test the phonetic predictions regarding the correlates of Type B MCVes as shown in

Figure 3.2. Because Type B MCVEs may be either epenthetic or underlying, I explore the theoretical implications of each with regard to articulation. While historical evidence suggests that Bunong minor syllable nuclei are in fact underlying [ʌ], not epenthetic [ə], a consideration of epenthetic minor syllable nuclei is warranted here given its implications for the diachronic role of the sesquisyllable and for articulatory theories of speech.

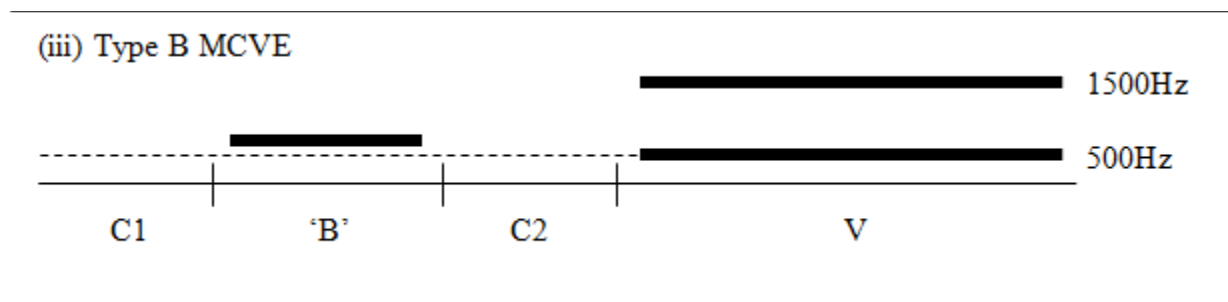


Figure 3.2: [Bunong] Proposed sesquisyllable type

The chapter is organized as follows. First, I review the small body of extant literature on Bunong phonology and the minor syllable in Bunong in particular. Then I present the results of an acoustic experiment, similar to the Khmer experiment reported on in the previous chapter. Results are once again interpreted in light of articulation, and differences between Bunong minor syllables and Khmer minor syllables are highlighted. In particular, I show that unlike the set of words investigated in Khmer, Bunong minor syllable vowels do have an associated gesture. In doing so, I also evaluate how articulatory configurations might change over time, dealing not only with the issues of excrescence and epenthesis but also deletion, suggesting multiple accounts for languages like Bunong.

3.2 BUNONG PHONOLOGY

3.2.1 *General Background*

Bunong, a South Bahnaric language, can be broken down into three dialect groups: Central, Southern, Eastern (Bequette 2008, Lewis et al. 2013). In addition to these three dialects

of Bunong, South Bahnaric also includes the languages Kraol, Koho, Maa, Stieng and Chrau (Lewis et al. 2013). Speakers of Central Bunong are divided by the political border between Cambodia and Vietnam, and each community uses a different writing system. In Vietnam, the Bunong script is Romanized, like the Vietnamese script, while in Cambodia, Bunong is written with a Khmer-based script. Nonetheless, the spoken dialect is still intelligible across the border. In this chapter, I report on the Central Bunong dialect as spoken in Cambodia, and indeed, most linguistic work to date has been undertaken on the Central dialect, either in Cambodia or Vietnam, with the notable exception of work by Blood (1966), who reports on the Eastern dialect.

Aside from the data presented here, there has been relatively little work on the phonology of Bunong. There are three linguistic studies, however, that at least mention Central Bunong phonology. The oldest of these is Phillips' (1973) study of registrogenesis in Bunong as it is spoken in Vietnam (see Section 2.1 above for a general discussion of register and also Butler (*accepted*) for an account of register in Bunong). Three subsequent studies focus on Bunong spoken in Cambodia. First, Vogel and Filippi (2006) following Vogel (2006) present a grammatical sketch of Bunong, which is preceded by a brief summary of the phonemic inventory. Second, Bequette's (2008) thesis on narrative discourse in Bunong also includes a review of Bunong phonology. Finally, Butler (*in progress*) also presents a grammatical sketch of Bunong with some consideration of the phonological system.

3.2.1.1 *Bunong Vowels*

There are 15 monophthongal vowels in Bunong and two diphthongs: [ai] and [oa]. Although each of the sources mentioned presents a slightly different inventory, each suggests

three contrastive vowel heights and three degrees of backness. I take Bequette's (2008) description to be the most accurate (Figure 3.3).

i	ĩ: / ĩ	u: / u
e	ɛ: / ɛ	o: / o
ɛ	a: / a	ɔ: / ɔ

Figure 3.3: Bunong vowels

In addition, length contrasts are relevant for at least the central and back vowel series. Bequette (2008) suggests that vowel length is contrastive for central and back vowels but not for front vowels. In contrast, Vogel and Filippi (2006) propose that there is a length distinction for front vowels as well; however, no minimal pairs are provided. Phillips (1973) suggests that a length distinction exists only for central vowels, and that instead there are two front lax vowels ([ɪ] and [ɛ]) and two back lax vowels ([ʊ] and [ɔ]). It is possible that both Phillips (1973) and Bequette (2008) are accurate and that the differences between their accounts represent either a change over time or a difference in dialects. However, because short vowels in Bunong are also produced in a more centralized and lax manner, Phillips (1973) and Bequette (2008) are likely reporting the same system.

3.2.1.2 *Bunong consonants*

The consonantal inventory for Central Bunong is given in Table 3.1. As with the vowels, there are a number of discrepancies in the consonantal inventory among previous phonological descriptions. In addition to the obstruent types listed in Table 3.1, Vogel and Filippi (2006) and Bequette (2008) also include a series of plain voiced stops. However, Butler (*accepted*) argues that aside from implosives, all stops in Bunong are now voiceless. In addition, Phillips (1973) includes a palatal and velar implosive that Vogel and Filippi (2006) and Bequette (2008) do not. Finally, Vogel and Filippi (2006) include a palatal fricative that Bequette (2008) does not, and

Bequette (2008) includes preglottalized glides that Vogel and Filippi (2006) do not. Each of these sounds is included here.

Table 3.1: Bunong consonants

	Lab.	Alv.	Pal.	Vel.	Glott.
Voiceless unaspirated stops	p	t	c	k	ʔ
Voiceless aspirated stops	p ^h	t ^h	c ^h	k ^h	
Implosives	ɓ	ɗ	ɟ	ɠ	
Fricatives			ç		h
Nasals	m	n	ɲ	ŋ	
Liquids		l, r			
Glides	w, ʔw		j, ʔj		

3.2.2 *Word Shapes and Minor Syllables*

Most sources say very little about word shape in Bunong. Phillips (1973) notes that the vowel and consonant inventories he provides are for the “main stressed syllable,” suggesting that the inventory for the minor syllable is distinct from the inventory for the major syllable. Vogel and Filippi (2006) provide slightly more detail, stating that Bunong has “pre-syllables” or “minor syllables” which consist of a consonant followed by an epenthetic schwa vowel. They suggest that the pre-syllable often derives from a prefix in which the vowel is neutralized due to lack of stress. However, there are two sources which provide a comprehensive analysis of word shapes in Bunong, including sesquisyllables. These are Phaen et al. (2012) and Bequette (2008) and I take them as the starting point for this discussion.

Bunong has a limited number of word shapes, including monosyllables, putative sesquisyllables and a very small number of disyllables. First, monosyllables must be heavy, either with a long vowel or a final consonant or both. The nucleus may be preceded by either one or two consonants (3.1).

(3.1) Bunong Monosyllables

- a) /bri:/ ‘forest’
- b) /plaj/ ‘fruit’
- c) /cuaj/ ‘offend’
- d) /khΛt/ ‘die’
- e) /kuʔ/ ‘sit’
- f) /ko:p/ ‘uncle’

Bunong minor syllables, i.e. the word-initial syllables in sesquisyllabic words, are claimed to be realized as CΛ, CΛN or N (Bequette 2008, Phaen et al. 2012). However, although Bunong does have an underlying /Λ/ vowel and an orthographic symbol to represent it, minor syllable vowels are always written as <a>.⁵ In an effort to avoid ambiguity, I have chosen to transcribe the minor syllable vowels as /Λ/ (Differences among stressed [Λ], unstressed [Λ] and [a] are investigated below). Examples are given in (3.2).

⁵ As is the case in many abugidas (i.e. segmental writing systems in which consonants and vowels are written as a consonant-based unit in which the vowel notation is secondary), the orthographic symbol for Bunong [a] is not written independently but must be attached to a consonant symbol, as in <ᦶᦺᦺᦹ> <kanar> [kΛnar] ‘well-worn path’.

(3.2) Purported Bunong Sesquisyllables

- a) /rʌ.la:w/ ‘more than’
- b) /lʌ.hat/ ‘tightly fitting’
- c) /tʌm.təl/ ‘to fight’
- d) /kʌ.tɔjç/ ‘hatchet’
- e) /m.lam/ CLASSIFIER

Finally, Bunong has a very small set of disyllabic words (3.3). Unstressed syllables in truly disyllabic words allow a greater variety of vowels than do minor syllables.

(3.3) Bunong Disyllables

- a) /ko.rap/ ‘lord’
- b) /ka.mək/ ‘knee’
- c) /tʃa.lot/ ‘fall over’

As presented by Bequette (2008) and Phaen et al. (2012), Bunong sesquisyllables ($C_1\Lambda.C_2VC$) can be distinguished from monosyllables with complex onsets (C_1C_2VC) by systematic differences in their consonant distributions. In particular, C_1 in complex onsets may be any consonant in the inventory of the language, but in minor syllables with vowel nuclei, i.e. excluding those formed by a syllabic nasal, C_1 must be one of /p p^h c^h k r l/. In addition, C_2 in complex onsets is limited to /r l w j/. A schema of the possibilities is given in Table 3.2.

Table 3.2: Word-initial C_1C_2VC and $C_{1\Delta}.C_2VC$ sequences in Bunong

C_1C_2VC		$C_{1\Delta}.C_2VC$		
C1	C2	C1	v	C2
any consonant	r			any
	l	p p ^h		consonant,
	w	c ^h k	ə	except
	j	r l		liquids or
				glides

This suggests that word-initial consonant sequences may be divided into two non-overlapping groups – (i) onsets of monosyllables or (ii) onsets of minor syllables – as seen in Figure 3.4.⁶

Onsets of Monosyllables include:

- Monosyllables with word-initial consonant clusters, in which C2 is either [r] or [l]. These include /pr, pl, tr, cr, kr, kl, khl, sr, sl, rl/. They are represented by the darkest shading and the structure CCVC.
- CC- sequences, which include a C1 nasal followed by a homorganic stop, i.e /mp, nt, nc, ŋk/. These should actually be considered unary prenasalized stops. They are represented by the lightest shading and the structure ^NCVC.

Onsets of Minor Syllables include:

- Any C1 which is /p p^h c^h k r l/ that does not precede an /r/ or /l/ C2, such as the C1s in /kp, kt, kc, km, kn, rp, rt, rc, rk, rs, rh, rm, rn, rj, rŋ, lh/. These are represented by the second darkest shading and the structure C_Δ.CVC.

⁶ Shading includes all consonant sequences predicted to exist by descriptions found in Bequette (2008) and Phaen et al. (2012). Consonant sequences for which the author has evidence are only those written in the table.

- Nasals preceding any consonant which is neither a homorganic stop nor /r/ nor /l/. In this case, nasals may be realized as syllabic – [ŋ] – or with an intervening vowel – [nə]. These include /mh, ns, nh, ɲh/. They are represented by the second lightest shading and the structures N.CVC and Nə.CVC.

In addition, some C1s are ambiguous. When nasals precede glides, /r/ and /l/, Phaen et al. (2012) report ambiguity in syllabification judgments across speakers. Some speakers consider the nasals as minor syllables while other speakers produce them as C1s in word-initial consonant clusters. These ambiguous sequences, i.e. /mr, ml, ɲr, ɲl/, are highlighted with diagonal stripes in the figure below.

C1 \ C2	p	t	c	k	s	h	m	n	ɲ	r	l
p	*									pr	pl
ph											
t		*								tr	
th											
c			*								
ch										cr	
k	KP	KT	KC	*			KM	KN		kr	kl
kh											khl
s					*					sr	sl
h						*					
m	mp						mh	*		mr	ml
n		nt					ns	nh			
ɲ			ɲc				ɲh				
ɲ				ɲk						ɲr	ɲl
r	RP	RT	RC	RK	RS	RH	RM	RN	RN	*	rl
l						LH					*

Figure 3.4: Word-initial consonant sequences in Bunong

3.3 GOALS AND HYPOTHESES

Descriptions of sesquisyllables in Bunong differ from descriptions of Khmer “sesquisyllables”, i.e. monosyllables with word-initial consonant clusters, in an important way.

In particular, in Khmer they have been defined as a set of words distinct from disyllables (and indeed, the analysis in Chapter 2 showed them to be monosyllabic). Bunong, however, is not claimed to have disyllabic words, save a few exceptions. Therefore, there is no meaningful way in which Bunong sesquisyllables can differ from non-existent Bunong disyllabic words. This leaves two possibilities as to the phonological structure of sesquisyllables in Bunong. First, if there is transitional vocalic material present between CCVC sequences as there is in Khmer, then minor syllable vowels in Bunong, i.e. C_ΔCVC vowels, might be durationally and qualitatively similar to that transitional vocalic material, in which case Bunong would essentially be a strictly monosyllabic language like Vietnamese. Alternatively, unstressed minor syllable vowels in Bunong might be more similar to stressed vowels in CVC monosyllables, in which case Bunong sesquisyllables should be considered disyllables. Regarding the latter comparison, because Bunong minor syllable nuclei are written with an ⟨a⟩ and are suggested to be reduced to [Λ] in their phonetic realization (Phaen et al. 2012), they should be compared to both /a/ and /Λ/ in monosyllables.

Figure 3.5 provides a schema of what the Bunong minor syllable vowel might look like as compared to the underlying vowel in monosyllables. Figure 3.5a provides gestural representations of CVC monosyllables and Figure 3.5b provides an example of a monosyllable with a word-initial consonant cluster. These represent two ends of a continuum, regarding possible vocalic material intervening between [k] and [l]: In (a) there is a vowel with full target attainment, whereas in (b) two consonants, i.e. [kl], are directly adjacent with no intervening material. We might expect Bunong minor syllables to fall somewhere in between these two. Possible gestural representations of the putative minor syllable [mΛ] are given in (Figure 3.5c-1) and (Figure 3.5c-2).

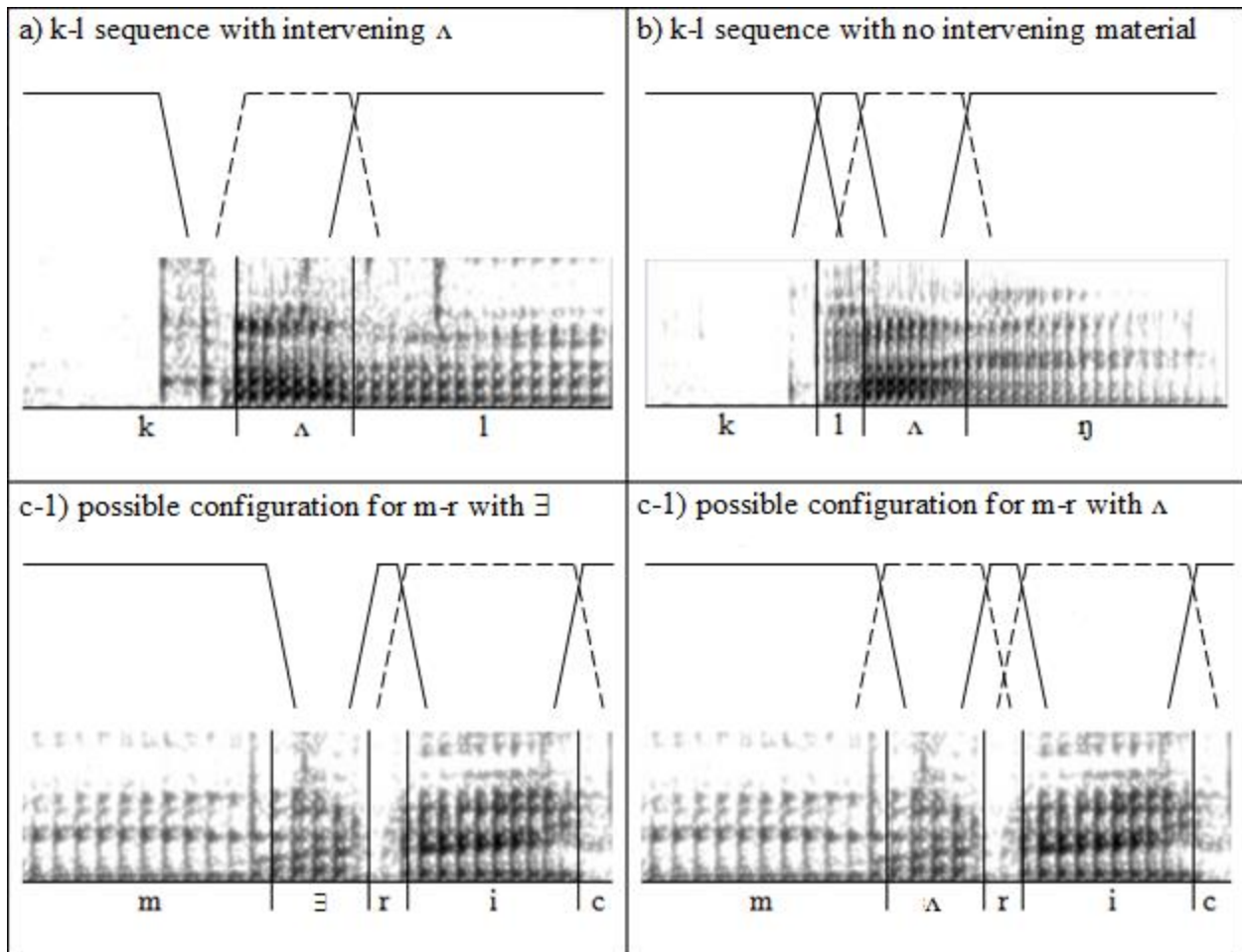


Figure 3.5: Bunong words and their proposed gestural representations: (a) [k Δ l] CVC, (b) [k Δ l η] CCVC, (c-1) [m ϵ r i c] CCVC, (c-2) [m Δ r i c] CvCVC. Consonant gestures are solid; vowel gestures are dashed.

In order to shed light on which gestural representation is more accurate, several issues are addressed. First, CCVC word-initial sequences are analyzed to confirm the presence of intervening harmonic or non-harmonic material which may be the result of underlap, similar to what was found for Khmer “sesquisyllables”. Next, underlap in CCVC sequences is compared to the minor syllable vowel C Δ CVC. If they are durationally and/or qualitatively different, this might suggest differences in their phonological statuses. In particular, I hypothesize that C Δ CVC vowels will be longer and their vowel quality will be less variable or reduced than vowel-like material found in CCVC sequences. Second, minor syllable vowels are compared to underlying vowels in monosyllables. Because minor syllable vowels are always unstressed, they will

undoubtedly be shorter than CVC vowels. However, a comparison with both underlying /a/ and stressed /ʌ/ might suggest whether the minor syllable vowel has a gestural target.

Subsequently, C1 nasals in NCVC sequences are evaluated in an attempt to determine if they are actually syllabic. There are at least two possible points of variation with regards to minor syllable nasals. First, there is a set of prenasalized stops, in which the nasal and following consonant agree in place of articulation. We may expect these nasals to be less intense and shorter than other C1 nasals. Second, Phaen et al. (2012) suggest there is individual variation in the pronunciation of certain words. For example, /mrol/ ‘lie’ may be produced either as a sesquisyllable [mərəl] or as a monosyllable with a complex onset [mrol]. As Parker (2002) has shown a correlation between intensity and sonority, and because syllabic segments should be more sonorous than non-syllabic segments, we expect syllabic segments to be not only longer than non-syllabic segments but also more intense.

3.4 METHODS

3.4.1 *Participants*

Twelve native Bunong speakers from Bou Sra village in Mondulhiri province, Cambodia, participated in the study. Ten participants were recorded in quiet locations (usually homes) in Bou Sra or Sen Monorom (a slightly larger city outside of Bou Sra), and two were recorded in a home in Phnom Penh. Participants were between the ages of 22 and 36 years old ($\mu = 28$). All participants were bilingual in Khmer, and some also spoke English, Vietnamese, French and/or Rade. Because the experiment was a reading task, it was necessary that all participants be literate in Bunong. Unfortunately, no literate female speakers could be located, so all participants were male.

3.4.2 *Stimuli and Task*

The wordlist (Table 3.3) consists of 21 tokens of type CVC, 7 of type CCVC, and 11 of type C Δ CVC . Because speech rate can affect the realization of underlap as well as the target attainment of shorter vowels (Davidson 2006b) each word was recorded in two blocks, which differed by speed. Speakers were instructed to read at a slow, comfortable pace during the first block and then quickly during the second. Other than the speech rate differentiated trials, the Bunong experiment followed the same methodology as the Khmer experiment in Section 2.4, i.e. stimuli were randomized and presented to the participants one word at a time, and three repetitions of each word were recorded. The frame sentence was [lah nau ____] ('Say the word ____'). Like the Khmer experiment, both the 2nd and 3rd repetitions were used to determine distributional results, i.e. the percentage of repetitions that have underlap and the percentage type of underlap. For all other results presented, including all statistical analyses regarding duration, formant values and spectral energy, only the second repetition was used.

Table 3.3: Bunong wordlist

IPA	Gloss	Orth.		IPA	Gloss	Orth.
CVC						
lɑr	varnish	លីវ		rɑt	tied tightly	រឹត
lɑŋ	section of bamboo	លីង		rɑk	sound of crowd	រឹក
rɑŋ	quorum	រឹង		nɑl	to prop up	នីល
nɑr	fishing basket	នីវ		ɕɑr	overgrown path	នីវ
mɑt	unripe	មីត		ɕɑt	to return	នីត
kɑl	big turtle	កីល		kɑt	frog	កឹត
lɑ:h	to do (neg.)	លើស		rɑ:h	generation	រើស
ŋɑ:r	bamboo for baskets	ងើវ		mɪ:r	liar	មីវ
kal	chop	កាល់		kat	cold	កាត់
la:h	to say	លាស		ka:ŋ	jaw	កាង
ka:l	well-worn path	កាល				
CCVC						
ŋrɑŋ	hammock	ងរឹង		ntit	crazy	នទឹត
mpɑt	skewer	មពីត		ŋla:p	to put inside	ងលាប
klɑŋ	to miss	ក្លឹង		mriɕ	black pepper	មរិច
ɕrɑŋ	to bait a hook	ឆ្រឹង				
CɔCVC						
rɑlɑŋ	type of bird	រាលីង		kɑna:r	wing	កានាវ
rɑkɑt	small raft	រាកឹត		kɑmɔk	knee	កាម៉ក់
rɑhɑ:n	show affection	រាហើន		katɔh	skin irritation	កាតិស
rɑnɑ:l	stool	រានើល		ɕɑlot	fall over	ចាលោត
lɑhat	tightly fitting	លាហាត់		kɑtojɕ	hatchet	កាតុយ្យ
rɑdɸl	reciprocate	រាដើល				

3.5 RESULTS

The results are organized in order to highlight comparisons between sesquisyllables and other word types in Bunong. First, purported minor syllable vowels are compared with transition states in word-initial consonant sequences in monosyllables. In order to make this comparison, CCVC sequences are analyzed for underlap in the context of both non-sibilant C1s and sibilant C1s. Next, C_ΔCVC vowels are compared to CVC vowels in monosyllables. Finally, nasal minor syllables are investigated.

Finally, because rate of speech was not found to have a significant effect for any of the statistical tests in the analysis, the results presented below are for fast speech, unless otherwise indicated. In addition to the raw fast and slow speech data, both speeds were also normalized relative to the duration of the frame sentence for each utterance. Statistical tests run on the normalized data yielded the same results as for the raw data. Results of tests run on normalized data are presented in lieu of fast speech data in Section 3.5.1.4 (CCVC vs. C_ΔCVC) because there were too few tokens for the tests run on fast speech alone to be reliable.

3.5.1 CCVC Transitions and C_ΔCVC Vowels

The goal of the first part of this experiment is to compare minor syllable [ʌ] with word-initial CCVC sequences. In order to do so, the latter were investigated to determine if they have any harmonic mid central vocalic material or any non-harmonic material intervening between Cs and, if so, to investigate the durational and qualitative properties of that material. Distributional results are presented first. As in Khmer, because frication noise may obscure visibility of voiceless underlap, results for CCVC sequences with non-sibilant C1s and CCVC sequences with sibilant C1s are presented separately. This is followed by a comparison of CCVC transitions with C_ΔCVC vowels.

3.5.1.1 CCVC Distributional Results

Figure 3.6 shows the distribution of underlap found in word-initial consonant clusters of monosyllables. These include /kl, sr, mr, ɲr, ŋl/, as well as the prenasalized stops /mp/ and /nt/ (double outlined) for which we do not expect any underlap. Two repetitions of each cluster were measured for each speaker at two rates of speech, for a total of 45 to 48 tokens for each form, once pronunciation errors were removed. Speech rates and repetitions are combined here because there is very little difference due to rate or repetition. When speakers are pooled, there is no difference larger than one between repetitions, by which I mean that one of the repetitions of a particular token has only one more instance of underlap than the other repetition. In terms of speech rate, the largest discrepancy is for /mr/, in which the faster rate has two more instances of underlap than the slow rate.

C1 \ C2	p	t	r	l
k				98 0
s			80 100	
m	0 X		89 100	
n		6 100		
ŋ			98 100	13 0

Figure 3.6: [Bunong] Percentage underlap (top) and percentage voiced underlap (bottom)

The top number in the boxes in Figure 3.6 shows the percent of tokens with some form of intervening material. The bottom number represents the percentage of those underlap tokens in which that material is voiced [ɤ] as opposed to voiceless [ɤ̥]. For example, 98% (or 47 of 48) /ŋr/ tokens have some intervening material, and of those tokens, in 100% of them it is voiced. Note that in all sequences, intervening material is either voiced (100%) or voiceless (0%), but none have both types.

These distributions show several tendencies within the data. First, the prenasalized stops /mp/ and /nt/ usually do not have any material intervening between C1 and C2, which is expected assuming that they are unary segments, as has been proposed in previous literature (cf. Phaen et

al. 2012).⁷ Second, in contrast, sequences with C2 /r/, i.e. /sr/, /mr/, /ŋr/, generally do have some intervening material, and that material is always voiced. Given that /r/ in Bunong is usually realized as a tap or trill with some pre-voicing, this is not surprising (Ladefoged and Maddieson 1996). Third, there is almost always a period of non-harmonic voiceless material between the consonants of the /kl/ sequence, indicating that /l/ voicing does not contribute to transition voicing.

In addition, a small percentage of /ŋl/ tokens (6 of 45) are recorded as having a transition period. It is important to note that when a transition does occur in these, it is voiceless. This is due to a small burst of air released simultaneously with the nasal closure, suggesting that there is a short but complete oral closure as well. Although this burst of air is counted as intervening material, it is shorter than the intervening material found in /kl/ sequences (Figure 3.7 and Figure 3.8).

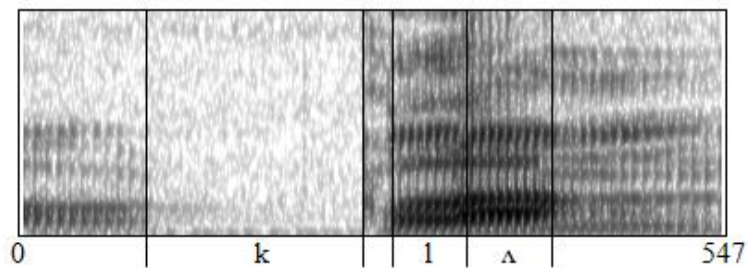


Figure 3.7: [Bunong] Underlap in /kl/ cluster. Duration in ms.

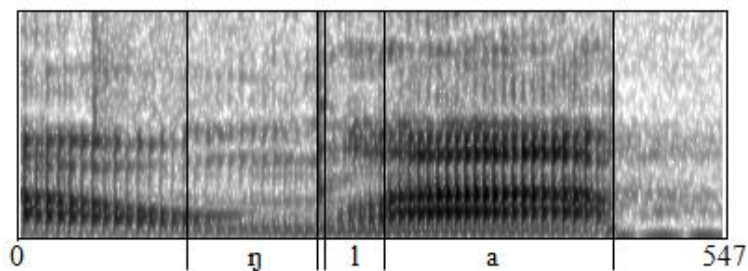


Figure 3.8: [Bunong] Underlap in /ŋl/ cluster. Duration in ms.

⁷ Unexpectedly, 3 of 47 repetitions did have some intervening material. These three may be speech errors or may be the result of hyperarticulation due to the experimental setup; however, it is highly doubtful that these intervening MCVEs are simply transition states since both consonants have the same place of articulation.

Finally, the most variation in the presence of underlap is found in the /sr/ clusters, in which 37 of 46 have voiced underlap. This likely derives from the anticipatory voicing of [r]. However, as in the Khmer data, there is a possibility that voiceless underlap is also present in the remaining tokens but is being obscured by frication noise. Therefore, these clusters are presented separately in the results below.

3.5.1.2 *Non-sibilant C1s in CCVC Words*

Having noted the general tendencies of the distributions of voiced and voiceless transitions periods in word-initial CC sequences, I now take a more systematic approach to determine if there are meaningful differences in their durations. First, just as in the Khmer data, a linear regression shows that the voicing of the transitions, i.e. [ɹ] vs. [ɹ̥], is significantly correlated with the total duration of CC sequences ($p < 0.0001$), which indicates that either (i) duration of the sequence depends on the voicing of the transition or (ii) duration of the sequence depends on the consonants present, with which the voicing of the transitions is correlated. Indeed, just as in Khmer, C1 and C2 types are correlated with the total duration of the CC sequences ($p = 0.0408$ and $p = 0.0001$, respectively), making it impossible to distinguish which of these analyses is correct.

In addition, different consonant types have inherent durational differences that should be accounted for. Therefore, a linear regression was run in which total duration of the CC sequence was regressed by C1 and C2 types (3.4). Subsequently, the residuals of this regression were regressed by transition voicing, i.e. [ɹ], [ɹ̥] or no apparent transition (3.5), the results of which show that the total duration of CC sequences is not correlated with the voicing of the transition ($p = 0.1402$).

(3.4)

$$\text{Total Duration} = \text{C1 Type} + \text{C2 Type} + [\text{Speaker}] + \varepsilon_{\text{TotalDur}}$$

(3.5)

$$\varepsilon_{\text{TotalDur}} = \text{Transition Voicing} + [\text{Speaker}] + \varepsilon$$

Figure 3.9 shows distributions of the total durations by transition voicing and distributions of the residuals of the total duration by transition voicing, respectively. Here and throughout, box plots are presented to show the range of the results and should be interpreted as explained in Section 2.5 above. A post-hoc Tukey HSD test shows that both C \exists C and C \exists C sequences are significantly longer than CC sequences ($p = 0.0046$ and $p = 0.0388$, respectively). There is no significant difference among the residuals; however, Figure 3.9 (left) shows that C \exists C and C \exists C sequences are still on average longer than CC sequences, although the distribution of the CC sequences is also noticeably larger than the distribution of either the C \exists C or C \exists C sequences.

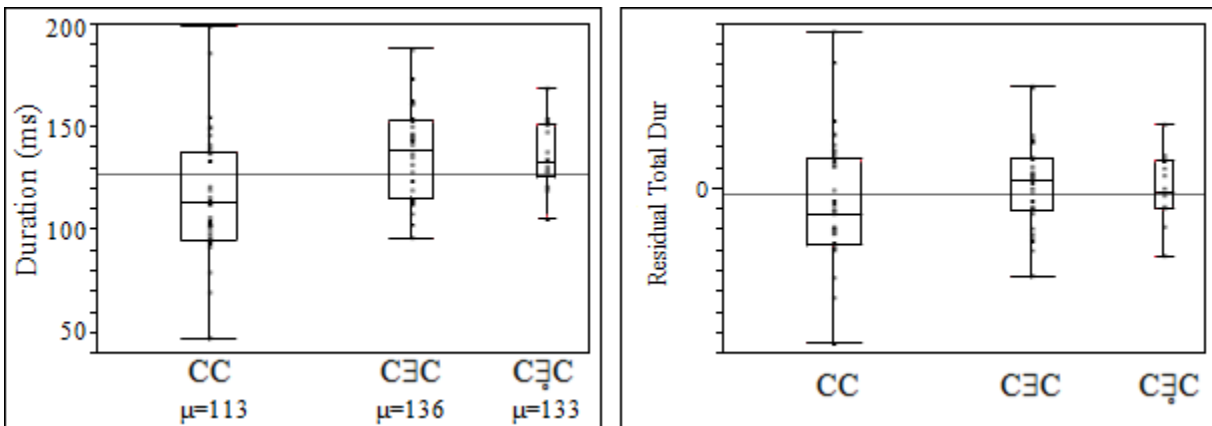


Figure 3.9: [Bunong] Total duration (left) and residual total duration (right) for consonant sequences with no transition, voiced transition and voiceless transition

In addition to comparing durations of the total sequences, durations of the transitions can be compared directly between C \exists C and C $\bar{\exists}$ C types to determine if there are qualitative differences between them. A linear regression shows that the duration of the transition is

significantly correlated with its voicing specification ($p < 0.0001$). Just as for the total duration, the duration of the transition is also significantly correlated with C1 and C2 types ($p = 0.0113$ and $p = 0.0001$, respectively). Therefore, the same procedure to calculate residuals for total duration in (3.5) above was followed for transition duration.

Figure 3.10 shows the distributions of transition duration by transition voicing and the distributions of residuals of transition duration by transition voicing, respectively. A post-hoc Tukey HSD test shows that [ɜ] is significantly longer than [ɝ] ($p < 0.0001$); however, there is no significant durational difference between the residuals of voiced and voiceless underlap, meaning that consonantal context has an important effect on underlap duration.

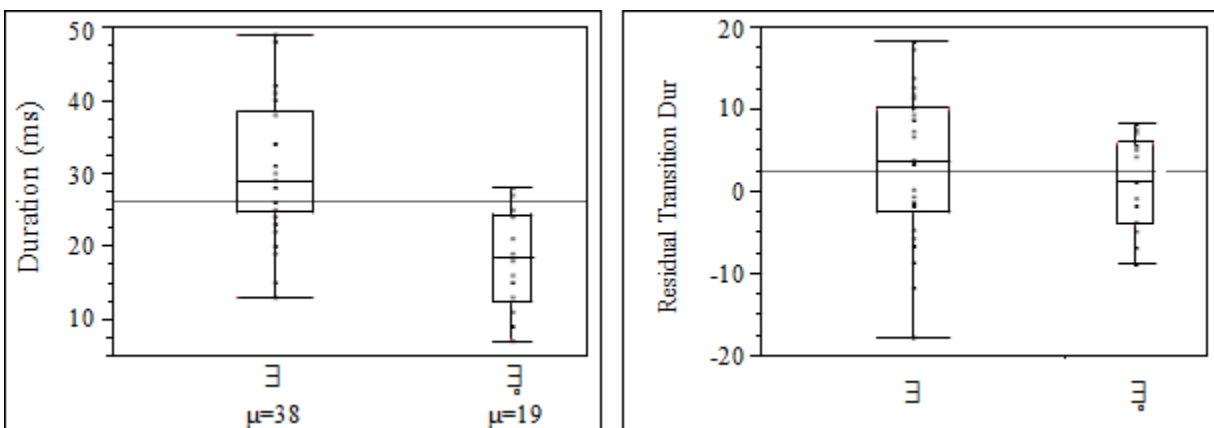


Figure 3.10: [Bunong] Transition duration (left) and residual transition duration (right) for consonant sequences with voiced and voiceless transitions

In sum, comparisons of the total duration of CC sequences and of transition durations indicate that transition voicing and its duration are highly correlated with consonant type. Indeed, transition duration is dependent on voicing specification, and voicing specification is dependent on consonant type. In general, consonant sequences can be classed into one of three categories: (i) prenasalized stops/no transition sequences, (ii) C-r/voiced transition sequences and (iii) C-l/voiceless transition sequences. Although a larger sample of sequences will shed more light on

the total picture, these results suggest that (i) C2 [r] requires a longer transition period than C2 [l] and that (ii) the voicing of the transition is conditioned by anticipatory voicing of the [r].

3.5.1.3 Sibilant C1s in CCVC Words

As was noted in the Khmer experiment (Section 2.5.3), transitional periods can be visually obscured by frication noise. Therefore, additional measures were made for Bunong [sr] sequences with no overt voiced harmonic transition period to determine if any transition period occurs. The energy was measured using the center of gravity of the fricative noise for the duration of the entire sequence labeled as fricative, as described in Section 2.4.3, and data were pooled across speakers.

Unfortunately, there were so few tokens that the results are largely unreliable. For fast and slow speech, there were 12 repetitions of the monosyllabic /sʌr/ token, i.e. one per speaker. For the C(ɤ)C forms, there were also 12 tokens; however, the breakdown differed between speech rates. As observed in Figure 3.6 above, 80% of /sr/ sequences have voiced underlap. In fast speech, 8 of 12 tokens have voiced underlap ('sɤr-'), while 4 do not ('sr-'). In slow speech, 10 tokens have a voiced transition, while 2 do not. Even, when durations are normalized and speech rates are pooled, this allows for the following distribution of types ((3.6):

(3.6)

a) [sʌr]: n = 24

b) [sɤr-]: n = 18

c) [sr-]: n = 6

When the tongue is in a state of transition, it is expected to have a wider distribution of energy than when a fricative is being produced. If [sr-] sequences have a transition period, we expect a different center of gravity measurement toward the end of the period labeled /s/ in [sr-]

sequences that we do not expect in the fricative portion of [sɛr-] sequences or /sʌr/ words.

Although a linear regression with Speaker as a random variable does show a significant difference between the Euclidean distance of the ‘s’ portion of [sʌr], [sɛr-] and [sr-] sequences ($p = 0.0014$), the erratic distribution of Center of Gravity measurements shown in Figure 3.11, suggest that this value is not meaningful. Indeed, more data are needed to determine if any voiceless underlap is present in the ‘s’ of [sr-] sequences.

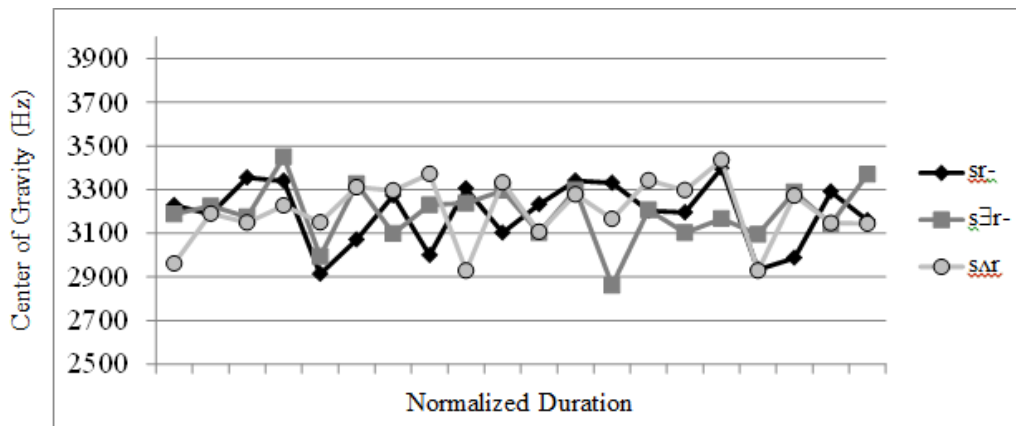


Figure 3.11: [Bunong] Center of Gravity in Hz for the 's' portion of [sr-], [sɛr-] and [sʌr] sequences

3.5.1.4 Comparison of CCVC Transitions with C_ΔCVC Vowels

Thus far, I have established that a voiced transition period is usually present in consonant sequences with C2 [r], whether C1 is a sibilant or not. Other consonant sequences are either prenasalized stops without underlap or have C2 [l], in which case voiceless underlap occurs but only rarely. We are now in a position to compare these transitions with purported minor syllable vowels in C_ΔCVC sequences. This comparison comprises two parts: duration and formant values. Such a comparison allows us to evaluate whether voiced transition periods are significantly different from what have been described as minor syllable vowels. If these transition sequences are equivalent to purported minor syllable vowels, we can conclude that just as in Khmer, Bunong minor syllables are not syllables at all. In contrast, if differences are found,

this would suggest that purported minor syllables in Bunong are in fact syllables. Examples of CCVC transitions and C Δ CVC vowels are given in Figure 3.12 and Figure 3.13, respectively.

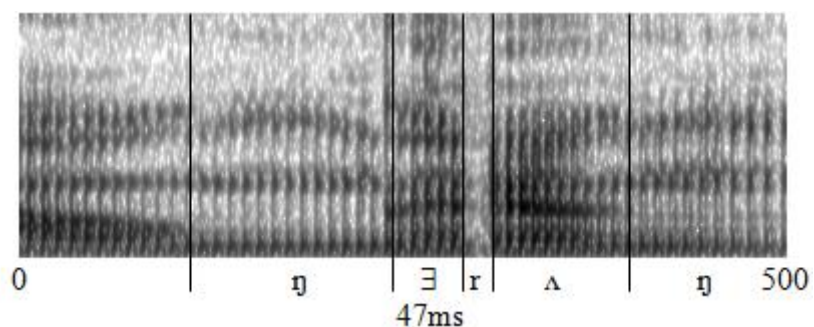


Figure 3.12: [Bunong] Vocalic transition in [ŋɛrʌŋ] 'hammock'. Duration in ms.

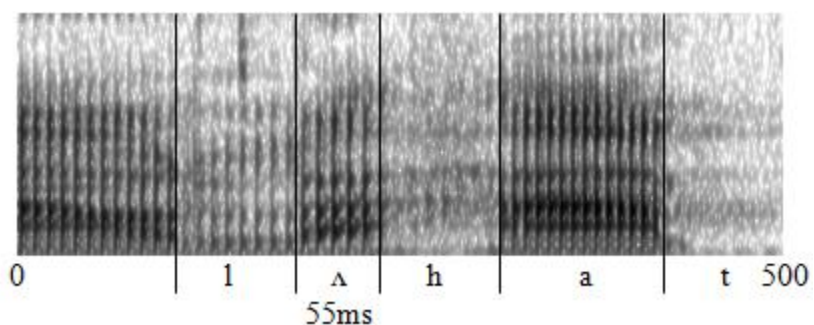


Figure 3.13: [Bunong] Purported minor syllable vowel in [lʌ.hət] 'tightly fitting'. Duration in ms.

To compare the durations of these CCVC transitions and C Δ CVC vowels, a linear regression of transition and vowel durations normalized by the frame sentence was performed. Normalized values were obtained by dividing the duration of the transitions and vowels by the duration of the frame sentence. Results reveal that vowels ($n = 248$) in C Δ CVC sequences have a normalized mean of 145ms with a range from 20ms to 318ms. Vocalic transition periods ($n = 57$) in CCVC sequences have a mean normalized duration of 62ms, with a range from 12ms to 183ms. C Δ CVC vowels are significantly longer than transition periods in CCVC sequences ($p < 0.0001$), as seen in Figure 3.14.

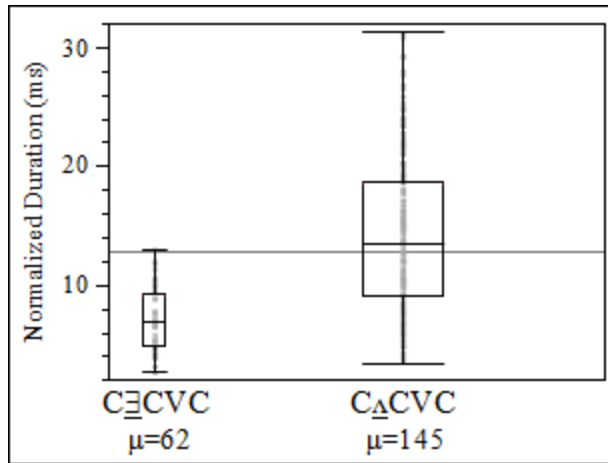


Figure 3.14: [Bunong] Normalized durations for CCVC transitions and C Δ CVC vowels

With regards to vowel quality, F2 values are not significantly different between CCVC vocalic transitions and C Δ CVC vowels ($p = 0.9540$), indicating that both have approximately the same amount of backness. F1 values, however, are higher for C Δ CVC vowels than for CCVC vocalic transitions ($p < 0.0001$), as can be seen in Figure 3.15. Just as in the Khmer data, F1 is lower for the vocalic transition than for the minor syllable vowel. These acoustic results suggest that this is because in the former, the tongue does not lower between consonants since it lacks a gestural target, i.e. the tongue position for CCVC vocalic transitions should directly reflect the surrounding consonants.

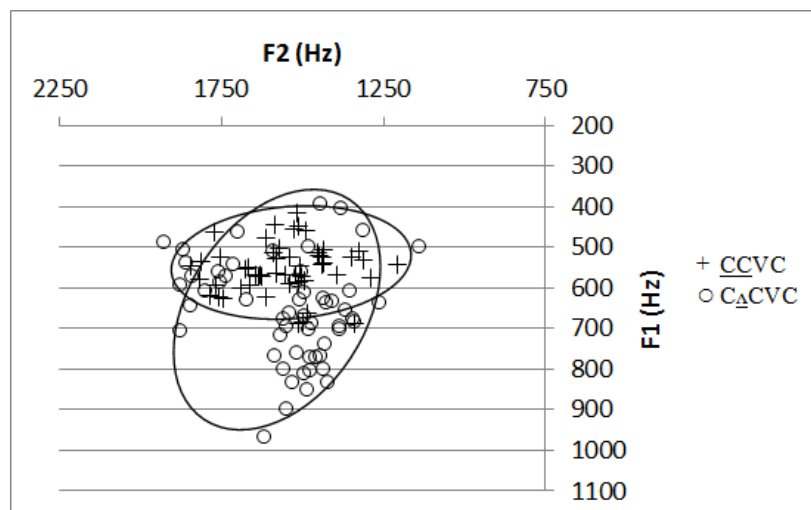


Figure 3.15: [Bunong] Formant values for word-initial CCVC vocalic transitions and C Δ CVC vowels

If this interpretation is accurate, we might then wonder why in Figure 3.15 there is a noticeable amount of overlap between the CCVC transitions and the C_ΔCVC vowels at lower values of F1. Why are these categories not more separated? In fact, the most overlap between these groups is found in shorter C_ΔCVC tokens. As seen in Figure 3.16, there is a significant correlation between the duration of the C_ΔCVC vowel and the value of F1 ($p = 0.0013$, $R^2 = 0.18$). In particular, the longer the vowel, the higher the F1. This suggests that lower F1 values for C_ΔCVC are due to insufficient time for complete target attainment. This fact, along with the formant measurements above, suggest that the vocalic transition period present in CCVC sequences is excrement and should be interpreted as underlap between consonant gestures.

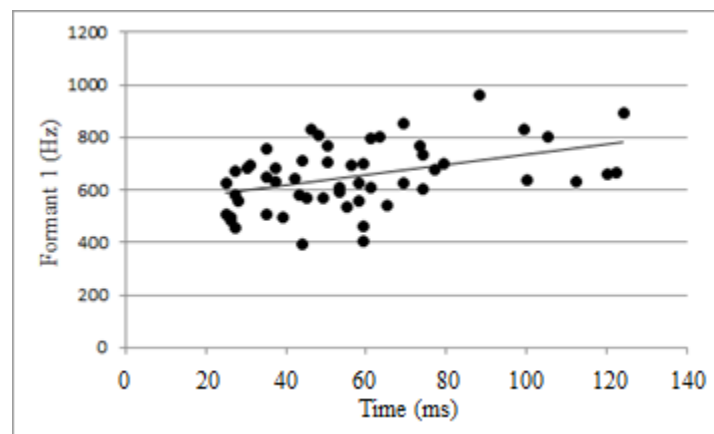


Figure 3.16: [Bunong] Correlation between the duration of the C_ΔCVC vowel and F1

In summary, Bunong clusters have transition states which should be interpreted as underlap although the environments in which underlap is found are far fewer than in Khmer because Khmer contains far more consonant clusters than does Bunong. In addition, duration and formant measurements suggest that underlap in Bunong is both quantitatively and qualitatively different from the purported minor syllable vowel nucleus, i.e. C_ΔCVC. With this in mind, I now turn to a comparison of C_ΔCVC vowels with vowels in CVC monosyllables.

3.5.2 Comparison of C_ΔCVC Vowels with CVC Vowels

Having established that unstressed C_ΔCVC vowels are durationally and qualitatively different from CCVC transitions, we can now evaluate them in light of underlying vowels in CVC monosyllables. In Bunong, C_ΔCVC vowels have been claimed to be underlyingly /a/, and are written as such in the orthography (Phaen et al. 2012). However, impressionistically, they are in most cases pronounced in a more centralized way. Therefore, in the following analysis, they are compared to the vowels of both C_ΔC and CaC monosyllables. Examples of C_ΔCVC, C_ΔC and CaC are given in Figure 3.17, Figure 3.18 and Figure 3.19, respectively.

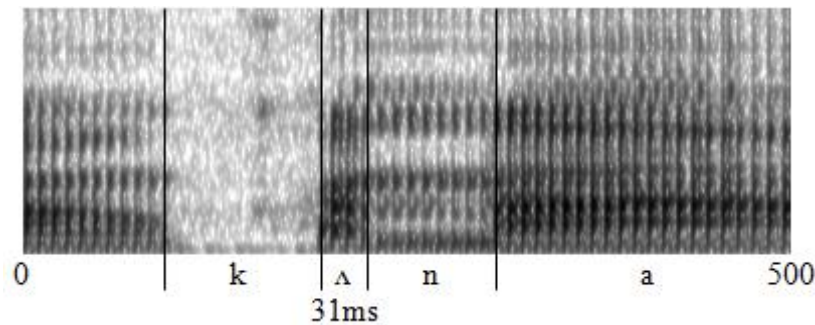


Figure 3.17: [Bunong] Purported minor syllable vowel in [kʌn:aɪ] 'wing'. Duration in ms.

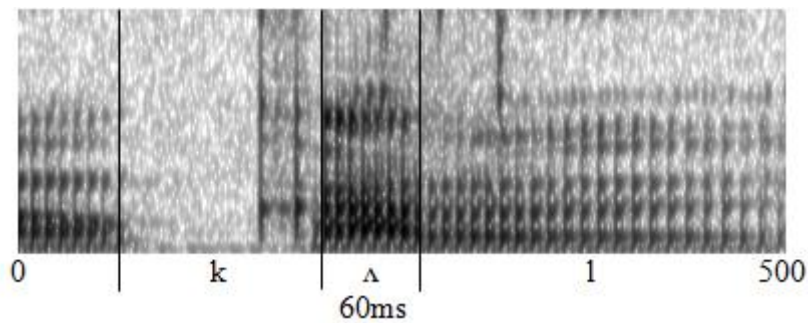


Figure 3.18: [Bunong] Underlying /ʌ/ in [kʌl] 'big turtle'. Duration in ms.

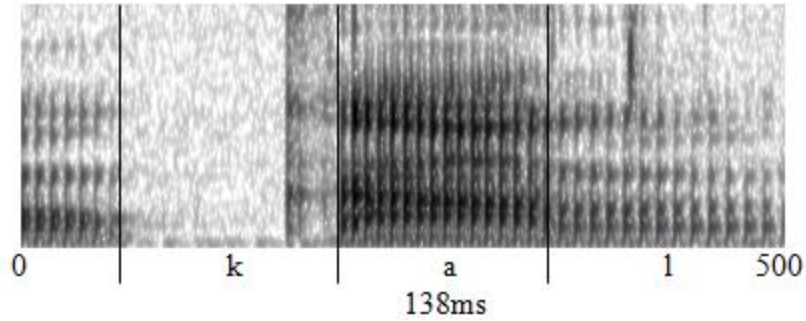


Figure 3.19: [Bunong] Underlying /a/ in [kal] 'chop'. Duration in ms.

Because $C_{\Delta}CVC$ vowels are unstressed and are found in disyllabic words while CVC vowels are stressed, we expect the latter to be longer than the former, and indeed they are. Both CaC and $C_{\Delta}C$ vowels are significantly longer than $C_{\Delta}CVC$ vowels ($p < 0.0001$), and CaC vowels are actually significantly longer still than $C_{\Delta}C$ vowels ($p < 0.0001$), as seen in Figure 3.20, which is not unexpected given the correlation between duration and vowel height. However, duration alone does not allow us to draw conclusions about the status of these vowels since many factors are known to greatly influence vowel duration (Klatt 1973, 1976).

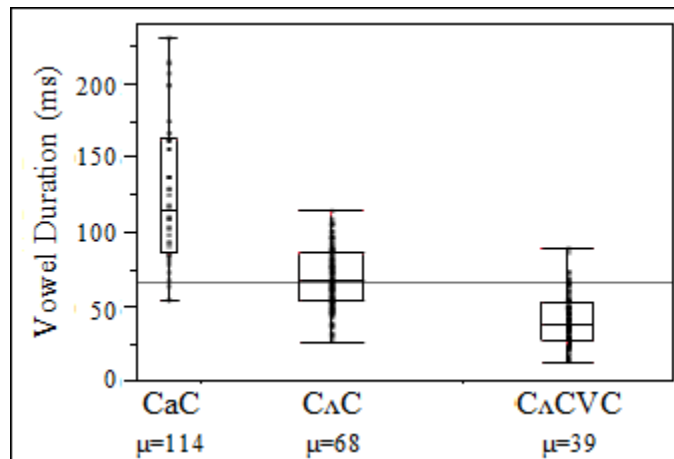


Figure 3.20: [Bunong] Vowel durations for CaC , $C_{\Delta}C$ and $C_{\Delta}CVC$ vowels

Formant values, however, are more telling. Just as was the case for unstressed $C_{\Delta}CVC$ vowels and $CCVC$ transitions, there is no significant difference in F2 among any of the categories tested here. As Figure 3.21 shows, the $C_{\Delta}CVC$ vowel as well as /a/ and stressed / Λ / are central. However, the distribution of $C_{\Delta}CVC$ is noticeably wider than either /a/ or stressed / Λ /.

This is likely a result of its short duration, which prevents it from being able to fully reach its gestural target and which also makes it more susceptible to influence from neighboring sounds. Indeed, Figure 3.22 shows that all the more peripheral C_ΔCVC vowels, i.e. those that lie outside the F2 range of stressed /Λ/, which is approximately 1350Hz – 1800Hz, have a duration less than 56ms, with one exception.

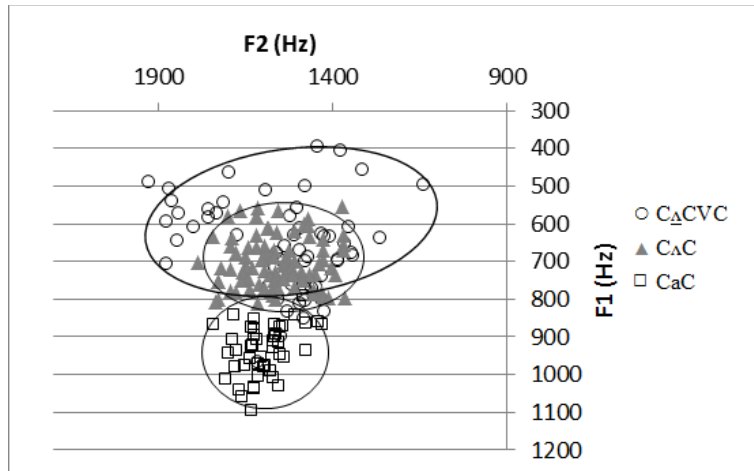


Figure 3.21: [Bunong] Formant values for C_ΔCVC, C_ΔC and CaC vowels

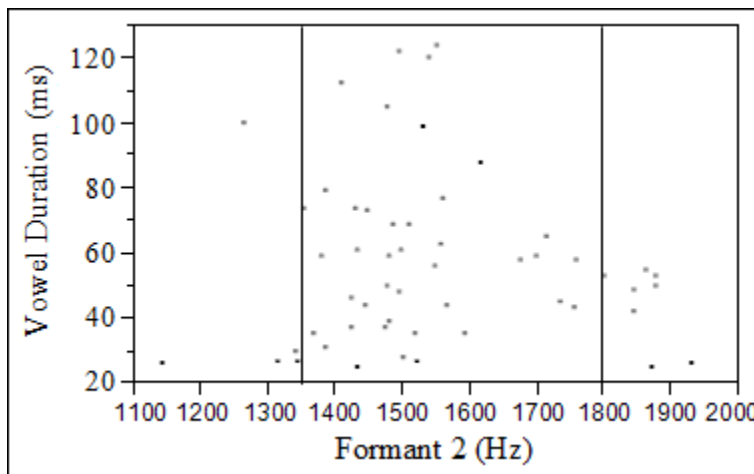


Figure 3.22: [Bunong] F2 by Vowel Duration for C_ΔCVC vowels

F1 differs significantly between each of the three types. As expected, /a/ has a higher F1 than stressed /Λ/ ($p < 0.0001$). In addition, despite the substantial amount of overlap seen in Figure 3.21, the C_ΔCVC vowel has a significantly lower F1 than stressed /Λ/ ($p = 0.0036$). Therefore, the phonetic realization of this unstressed vowel is certainly not [a], and although it

may be /a/ underlyingly, it is realized as something much closer to stressed [ʌ]. Still, the difference in F1 between stressed /ʌ/ and C_ΔCVC is unexpected if both unstressed and stressed /ʌ/ have the same gestural target. However, this difference is actually neutralized in slow speech, suggesting that the longer speakers take to produce unstressed /ʌ/, the more similar to stressed /ʌ/ it becomes, and as was demonstrated in Figure 3.16, F1 is inversely correlated with speech rate. Indeed, the average duration for C_ΔCVC vowels whose F1 is equal to or less than 550Hz is 42ms, while the average duration for C_ΔCVC vowels whose F1 is greater than 550Hz is 62ms.

In sum, results suggest that the unstressed C_ΔCVC vowel is quantitatively distinct from the stressed nuclei of monosyllables but qualitatively quite similar. First, C_ΔCVC is durationally shorter than C_aC and C_ΔC, which is expected given the effects of stress on duration (cf. Lehiste and Peterson 1959, Liberman 1960). Next, although it has a more widely distributed F2 and a lower F1 than either /a/ or stressed /ʌ/, these differences are speculatively due to a shortened duration, which prevents the tongue from fully reaching its gestural target. When combined with the results in Section 3.5.1.4, these results suggest that what are traditionally considered minor syllable vowels in Bunong have an associated gesture, as opposed to the underlap in word-initial consonant clusters, which does not.

3.5.3 Syllabic Nasal Minor Syllables

In addition to minor syllables of the shape Cv-, Bunong is claimed to have nasal minor syllables, in which the entire syllable is composed of a syllabic nasal – [ṃ, ṅ, ṇ, ṅ̃] (Phaen et al. 2012). Syllabic nasal minor syllables are common in descriptions of sesquisyllabic languages, e.g. Kammu (Svantesson 1983). In addition to nasal minor syllables, Bunong is also claimed to have prenasalized stops, which are always homorganic. In contrast, nasal minor syllables are not. In the data investigated here, [ṃp] and [ṅt] compose the set of prenasalized stops, and [ṃr], [ṅr]

and [ŋ] are purported to be syllabic nasals followed by onset consonants. A list of the forms analyzed is given in Table 3.4.

Table 3.4: [Bunong] Prenasalized stops, purported minor syllable nasals and simplex nasal onsets

Prenasalized Stops ^N CVC	Purported Syllabic Minor Syllables Ṇ.CVC	Simplex Nasal Onsets NVC	
/ ^m pʌt/	/ṃ.riç/	/mʌt/	/nʌl/
/ ⁿ tʰit/	/ŋ̣.la:p/	/mi:r/	/nʌr/
	/ŋ̣.rʌŋ/	/naç/	/ŋʌr/

To determine whether or not the nasals in these heterorganic consonant sequences are syllabic, they are compared in duration and intensity to the nasal portion of prenasalized stops as well as to simplex nasal onsets in NVC monosyllables, including [m], [n] and [ŋ]. If these nasals are syllabic, we expect them to have a longer duration and greater intensity than prenasalized stops and simplex onsets (Byrd 1993, Parker 2002). Examples are shown in Figure 3.23, Figure 3.24 and Figure 3.25.

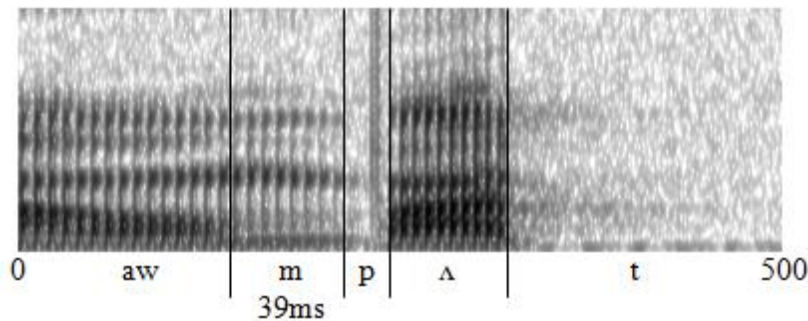


Figure 3.23: [Bunong] Prenasalized stop in [^mpʌt]. Duration in ms.

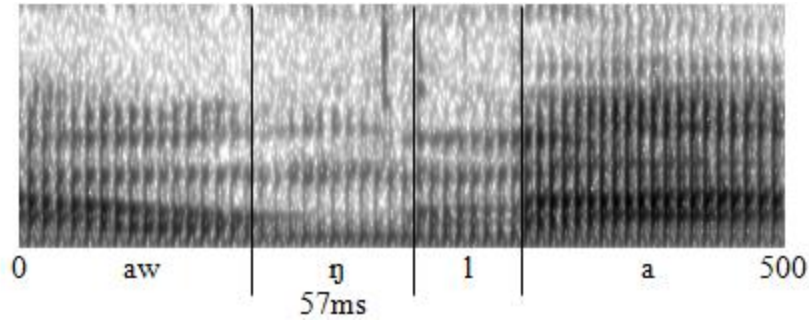


Figure 3.24: [Bunong] Purported syllabic nasal in [ŋ.lap]. Duration in ms.

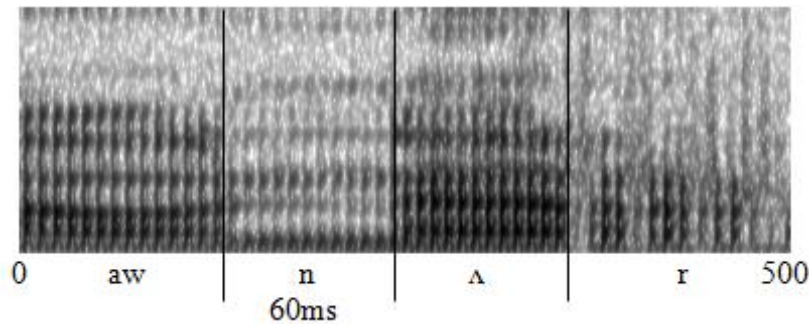


Figure 3.25: [Bunong] Simplex nasal onset in [nɹ]. Duration in ms.

Instead, duration results show that purported syllabic nasals are neither significantly longer than NVC nasals ($p = 0.2432$) nor the nasal portion of prenasalized stops ($p = 0.7282$). NVC nasals are also not significantly longer than the nasal portion of prenasalized stops, although at a level $p = 0.0644$, their difference is much closer to statistical significance than the difference between either of them and the purported syllabic nasals (Figure 3.26).

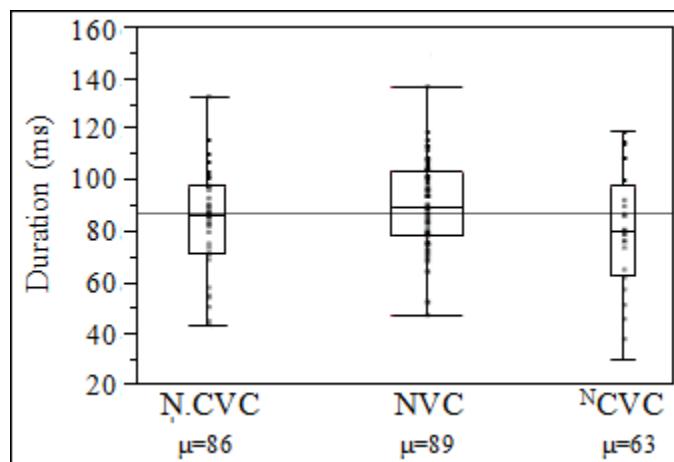


Figure 3.26: [Bunong] Duration distributions of nasals

Just as there is little durational difference among categories of nasals, there is also little difference in intensity. Normalized nasal intensity was obtained by subtracting the mean intensity of the nasal from the peak intensity of the preceding frame sentence, which was always located on the /a/ in /lah/. Because this resulted in the lower intensity nasals having larger associated values, the intensity measurements were then subtracted from 100 to invert the differences. Normalized intensity values for each nasal type can be seen in Figure 3.27. A post-hoc Tukey HSD test shows that while unexpectedly the nasals of prenasalized stops are significantly more intense than simplex onset nasals ($p = 0.0008$), purported syllabic nasals are not significantly different in intensity from either simplex onset nasals ($p = 0.2261$) or prenasalized stop nasals ($p = 0.2644$). In sum, nasals that have been claimed to be syllabic are neither longer nor more intense than other types of nasals.

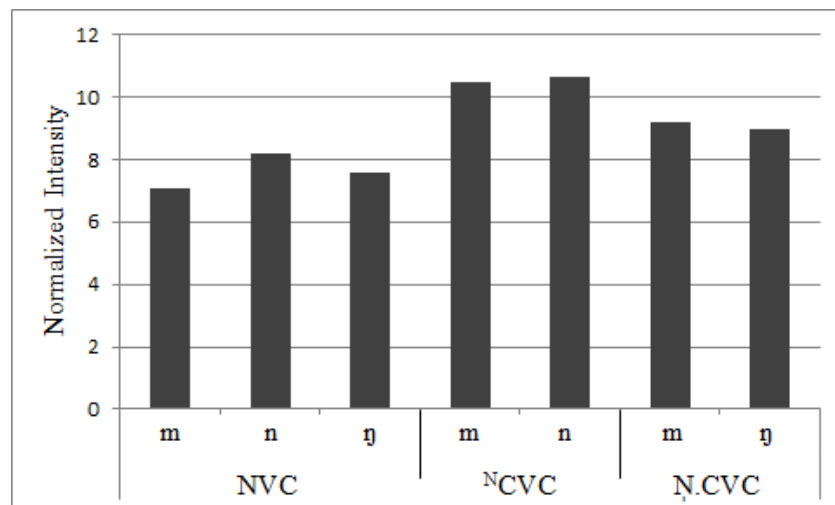


Figure 3.27: [Bunong] Average normalized nasal intensity

Although the above results suggest C1 nasals in heterorganic consonant sequences produced by the participants of this experiment are not syllabic but are instead part of word-initial consonant clusters, these results must be interpreted with caution. A number of studies have sought to determine the phonetic correlates of syllabic consonants and nasals in particular (cf. Riehl 2008, Pouplier and Benus 2011, Cohn and Riehl 2012, *inter alia*), but duration and

intensity alone – while suggestive – are not sufficient criteria on which to base concrete conclusions. In addition, it is possible, as suggested by Phaen et al. (2012), that productions may differ across speakers.

Nonetheless, these results suggest that Bunong does not contain minor syllables consisting of syllabic nasals only. Furthermore, as was discussed in Section 3.5.1.1, in cases where a vocalic transition appears in either ^NCVC words or in ^N.CVC words, the transition noise is substantially different from unstressed C_ΔCVC vowels, indicating that minor syllables in Bunong are neither syllabic nasals nor Cə- or C_Λ- sequences, which begin with nasals. In other words, the results suggest that all word-initial nasals followed by Cs are likely either the nasal portion of a prenasalized stop or the first consonant in a word-initial consonant cluster.

3.6 DISCUSSION

3.6.1 *Summary of Results*

The results indicate several properties of purported minor syllables in Bunong. First, Bunong C_ΔCVC vowels are distinct from underlap. In this way, Bunong minor syllables differ from Khmer “minor syllables”, which, as I have established previously, should not actually be considered syllables. When compared to nuclei of CVC monosyllables, Bunong C_ΔCVC vowels are yet again distinct. Not only are they shorter than monosyllabic nuclei, they often have a greater distribution in their acoustic realizations, which is likely due to their shorter duration, which in turn correlates with their lack of stress. These results, considered together, suggest that Bunong minor syllables are indeed syllables that are unstressed.

This conclusion supports the hypothesis that word types which are commonly called sesquisyllables are in fact disyllabic iambs. Under this interpretation, the reduced segmental inventory as well as the weight restrictions on minor syllables are expected properties of the

weak initial syllables that are characteristic of iambs (Hayes 1995). Therefore, I conclude that minor syllables in Bunong are indeed well-formed mora-bearing syllables and that sesquisyllables in Bunong should be reinterpreted as ordinary disyllables.

In addition, I have suggested that, at least for the speakers in this sample, Bunong does not contain minor syllables whose nuclei are syllabic nasals. Nasals that were previously interpreted as such, i.e. nasals which are heterorganic with the following consonant, are neither longer nor more intense than other nasals in Bunong, suggesting that these nasals are in fact part of word-initial consonant clusters.

3.6.2 *Minor Syllables as Gestures*

In Section 3.6, I presented an analysis of what “minor syllables” might look like as transitions between gestures. I also suggested that a gestural analysis of minor syllables is superior to a segmental analysis because it provides the means to explain and concretize the differences between excrescent and epenthetic or underlying vowels, i.e. whether or not a gesture is present. In this same light, having established that Bunong minor syllables are syllables with vocalic nuclei, I now turn to the question of how to represent these syllables in a gestural framework, which is crucially different than the representation of the schwa-like vocalic material resulting from the mistiming of gestures as in Khmer. Although over time, excrescence can be re-interpreted as epenthesis (and vice versa), I suggest that an epenthetic vowel is crucially phonological and has an associated gesture, whereas an excrescent “vowel” consists merely of a transition state. Certainly, these two may be difficult to distinguish in the acoustics, especially if, for example, a vowel gesture is almost entirely obscured by neighboring gestures. Still in some cases, like in Bunong underlap versus Bunong minor syllable vowels, the two have quite distinct acoustic realizations.

Because minor syllables are noted for their supposed instability and their tendency to change over time (Sidwell 2000, Brunelle and Pittayaporn 2012, *inter alia*), it is useful to understand how their representations as gestures can change over time. Based on data from related South Bahnaric languages, it is fairly well accepted that at least some Bunong minor syllables were formed as part of a reduction of prefixed derivational morphology (Sidwell 2000). This may explain in part the somewhat skewed inventory of consonants that occur in the onset position of minor syllables, i.e. /p, p^h, c^h, k, r, l/. If these are remnants of morphemes, the lack of other consonants in this position, e.g. /t/, is less surprising than if motivated purely on phonological grounds.

If the change from disyllables to monosyllables (with an assumed intermediate step of sesquisyllables) is due to a process of deletion, it is important to understand what that process looks like in terms of articulation. Browman and Goldstein (1990, 360) suggest that what are often thought of as categorical deletions in fast speech are actually not deletions at all: “In faster, casual speech, we expect gestures to show decreased magnitudes (in both space and time) and to show increasing temporal overlap. We hypothesize that the types of casual speech alternations observed (segment insertions, deletions, assimilations and weakenings) are consequences of these two kinds of variation in the gestural score.” In other words, it is quite often the case that “deletions” are not actually deletions but are just a matter of a particular gesture being obscured by gestures on other tiers. For example, Browman and Goldstein (1989, 1990) show that the /t/ in ‘perfect memory’ can be deleted in the sense that it is neither visible on a waveform nor audible in the speech stream, while remaining present in the articulation.

More relevant to the deletion of minor syllables, Davidson (2006b) suggests that pre-tonic wedge, which is often thought to be deleted in fast speech in English, is actually not deleted

entirely. Although her experiment is not articulatory, the results suggest that “acoustic residue,” e.g. longer consonant durations or aspiration of stops in fricative-stop clusters, indicates that the wedge gesture is not deleted entirely. Instead, some effects of the wedge gesture are still observable. Davidson concludes that this type of “deletion” is more consistent with an overlap account than an account in which rules only apply in fast speech. Nonetheless, Browman and Goldstein (1989) do acknowledge that it is possible for gestures to be deleted entirely. Indeed, we assume this is often the case when “sesquisyllables” become monosyllables. In particular, one generation of speakers may hide a gesture to such an extent that subsequent generations re-interpret the utterance omitting the hidden gesture altogether.

However, minor syllables in Bunong are not deleted entirely. Instead, vowels in the initial prefixes were reduced so that the only permitted vowel in the minor syllable is wedge. Such reduction is likely due to what Browman and Goldstein (1990) refer to as a “decreased magnitude” of the gesture, likely due to the shorter duration of the unstressed syllables. At any rate, minor syllable vowels in Bunong are highly restricted, and the set of consonants in initial position in the minor syllable is quite small relative to the consonant inventory of the language.

It is at this point that we can imagine two different synchronic grammars for speakers of Bunong, both of which are in accordance with the fact that minor syllable vowels have associated gestures and targets. First, the gesture for the minor syllable vowel might be (and likely is) underlying, as in Figure 3.28. This is the simplest scenario, which assumes no epenthesis or deletion rules are operative on the underlying representation. Each arc in the figure represents a set of gestures and their timed relationship with other sets of gestures. While there is some amount of overlap, each gesture is not only present but also fully attains its target.

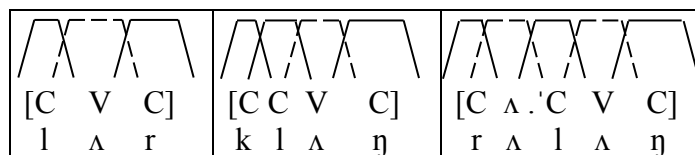


Figure 3.28: Gestural representation of Bunong monosyllables and sesquisyllable

Second, however, because the target in which the minor syllable vowel occurs is entirely predictable, in theory it could be epenthetic instead (Figure 3.29). In order to determine which of these is correct, some morpho-phonological evidence is needed; however, because of the extremely minimal morphology of Bunong, this evidence does not exist.

Underlying Representation			
	/C V C/ l ʌ r	/C C V C/ k l ʌ ŋ r l ʌ ŋ	
Rules	$\left[\begin{array}{l} p, p^h \\ c^h, k' \\ r, l \end{array} \right] - \left[\begin{array}{l} \text{not} \\ r, l \\ j, w \end{array} \right] \rightarrow \text{---}$		
	Stress assignment and vowel reduction		
Surface Form	 [C V C] l ʌ r	 [C C V C] k l ʌ ŋ [C ɛ C V C] k ɛ l ʌ ŋ	 [C ə . C V C] r ə l ʌ ŋ

Figure 3.29: Derivation of epenthetic minor syllable vowels in Bunong

Despite the lack of evidence, a rule-based type of epenthesis is in and of itself a somewhat controversial claim to make within an AP framework. As Warner et al. (2001) note, within a standard view of AP, a bundle of gestures cannot be inserted between the underlying representation (or gestural score) and the surface representations (or phonetic output). Indeed, one of the core properties of AP is the inherent relationship between the phonology and the phonetics, so to conceptualize gestural insertion as something that happens between levels of

representation, as in Figure 3.29, is problematic. However, some insertion phenomena have been accounted for in terms of gestural insertion. For example, Warner et al. (2001) argue that schwa insertion in Dutch must be due to epenthesis because there is no way for changes in either gestural magnitude or overlap to account for the changes associated with schwa insertion. In addition, McMahon et al. (1994) claim that English /r/-insertion (e.g. /waf/ → [warf]) must actually be due to a process of epenthesis and consequently AP must be integrated into a more stratal phonological model like Lexical Phonology.

However, more conservative and widely accepted AP models have accounted for a number of phenomena that are traditionally considered insertion by making use of the same processes invoked to account for deletion, i.e. changes in gestural magnitude and overlap. For example, /r/-insertion or /t/-insertion (e.g. /prɪns/ → [prɪnts]) can be accounted for in this way (Gick 1999). Indeed, Gick (1999, 44) suggests that, as McMahon et al. (1994) have stated about previous insertion accounts, “the arbitrary insertion of a gesture requires as much extra theoretical apparatus as the insertion of any other phonological unit.”

Although the correct model for synchronic epenthesis is still under debate, the concept of epenthesis as a diachronic sound change is much more widely accepted. For example, McMahon et al. (1994, 289) suggest that epenthesis which occurred between Latin and Spanish can be supported by work outlined in Browman and Goldstein (1992): “Epenthetic schwa vowels, especially when they break up consonant clusters, might then result from speakers slightly increasing the distance between the cluster consonants, until they no longer overlap. At that point, a schwa-like vowel would be perceived.” In this way, epenthesis as intergenerational language change is fairly uncontroversial.

In conclusion, Articulatory Phonology easily accommodates intergenerational language change whereby epenthesis of minor syllables occurs in some acoustically relevant environments. However, if this type of change were to take place within all word-initial consonant clusters or even within a subset of onsets defined by phonological characteristics, like sonority for example, a conservative AP account, i.e. one which is not integrated into a more derivational phonological framework, forces us to conclude that these non-phonetically motivated instances of epenthesis are the result of large-scale phonological constraints on the language. In other words, speakers must generalize from misinterpretations of epenthesis in certain environments to other environments as well. Crucially, however, the lexical representations/gestural scores of all affected environments are set during planning to include this “additional” gesture so that the need for gestural epenthesis between the gestural score and the phonetic output is unnecessary.

3.6.3 Conclusion

Minor syllables in Bunong are phonologically real. They have an associated gesture and are the weak first syllable of disyllabic iambs. In this way, they differ from minor syllable “vowels” in Khmer, which are actually transition states between consonants. Yet sesquisyllables in Bunong also differ from less canonical forms of sesquisyllables in several Southeast Asian languages, which are characterized by allowing multiple minor syllables. While an iambic analysis works for languages like Bunong, what can be said about languages like Burmese, whose maximal word shape does not fit cleanly into a disyllabic mold? I investigate this question in the following chapter.

CHAPTER 4

MULTIPLE MINOR SYLLABLES IN BURMESE

4.1 INTRODUCTION

Burmese is a Sino-Tibetan language spoken in Myanmar and has about 32 million speakers (Lewis et al. 2013). Like Khmer and Bunong, Burmese is also claimed to have minor syllables (Okell 1969, Bradley 1980, Green 1995, 2005, *inter alia*). Aside from its voiceless sonorants, which have been the subject of a number of studies (Dantsuji 1987, Bhaskararao and Ladefoged 2009, *inter alia*), it has a fairly standard consonant inventory (Table 4.1). A vowel inventory is given in Figure 4.1.

Table 4.1: Burmese stops (adapted from Green 2005)

	Labial		Dental		Alveolar		Palatal		Velar		Glottal	
Plain Stops	p	b			t	d	c	ɟ	k	g		ʔ
Aspirated Stops	p ^h				t ^h		c ^h		k ^h			
Plain Fricatives			θ	(ð)	s	z	ʃ					
Aspirated Fricatives					s ^h							
Trill					(r)							
Nasals	m	ᵹ			n	ᵈ	ɲ	ɲ̚	ŋ	ŋ̚		
Approximants					l	l̥						
Glides	w	(w̥)					j					

i	u
e	o
	ʌ
ɛ	ɔ
	a

Figure 4.1: Burmese vowels (adapted from Green 2005)

Burmese is also considered a tonal language, with four tones – high, low, creaky and checked (Green 1995) or level, heavy, creaky and stop (Okell 1969). However, Green (2005) notes that Bradley (1982) refers to these different categories as registers instead of tones since they also exhibit differences in phonation, intensity, duration and vowel quality. In addition, Bradley (1982, 120) suggests that there is a fifth possible tone, i.e. a “reduced” tone which is only possible on minor syllables and is not contrastive. It is described as “very low” and “very short”. In contrast, Okell (1969) explicitly states that minor syllables, or as he calls them “weak syllables,” do not have tone.

Burmese tones are represented in different ways orthographically across different sources. Because this is not an investigation into Burmese tone, I do not wish to make strong claims about which tonal analysis is correct. But for the sake of consistency, tones will be represented as in Table 4.2. In addition, when tones differ between sources, I use the tonal specifications presented by Okell (1969).

Table 4.2: Representation of Burmese tone

Tone	Symbol
high/level	á
low/heavy	à
creaky	ǎ
checked/stop	a?
reduced/minor syllable	a

Additionally, Burmese differs from the other languages investigated so far in that it allows multiple minor syllables within a single word. This is not claimed to be unique within the Tibeto-Burman languages, although it is still relatively rare. Matisoff (2003) cites a number of languages which are suggested to have multiple minor syllables due to historical prefixation and subsequent “reprefixation”. Some examples are listed in (4.1) below.

(4.1)

- | | | | | |
|----|-------------------|---------------|-----------------|---------------------------|
| a) | khΛ.mΛ.lek | ‘lick’ | Tangkhol Naga | |
| b) | sbrul < *s-b-ru:l | ‘snake’ | Written Tibetan | |
| c) | pă.să.wi | ‘plaid cloth’ | Jingpho | Hanson (1906) (contested) |

Given my claim that sesquisyllables should be considered non-iterative disyllabic iambs, these trisyllabic word forms (or *extended sesquisyllables*) are potentially problematic for my analysis. In this chapter, I present a pilot study of minor syllables in Burmese. The results, while suggestive of the structure of the prosodic word, are ambiguous in terms of foot structure, confirming that the role of phonetics in assessing prosodic structure is limited.

4.2 MINOR SYLLABLES IN BURMESE

Minor syllables in Burmese are different from those in Bunong and Khmer in that they often result from synchronic processes of compounding and reduction. Although similar processes were likely the origin of minor syllables in Bunong and Khmer (Sidwell 2008), such processes are no longer productive in those languages, aside from a limited amount of morphology in Khmer. Green (2005) provides his own characterization of minor and major syllables in Burmese. This is summarized in Table 4.3. As noted in the list of properties, minor syllables are always light, and major syllables are always heavy.

and reducing compounds. In reducing compounds, like those seen in (4.3) above, the first of the two words is phonologically reduced. On the other hand, non-reducing compounds (4.4) are claimed to comprise multiple prosodic words and have no phonological reduction. Although these types of words can include stems that already have minor syllables, no further reduction takes place during the compounding (4.4c).

(4.4)

- | | | | | |
|---------------------|------------|---|----------------------|-----------------------------|
| a) [cɛʔ.shíN] | ‘turkey’ | < | [cɛʔ] + [shíN] | ‘fowl’ + ‘elephant’ |
| b) [jáuN.wé] | ‘trade’ | < | [jáuN] + [wé] | ‘buy’ + ‘sell’ |
| c) [ʔʌ.jé:.ʔʌ.chìN] | ‘standard’ | < | [ʔʌ.jé:] + [ʔʌ.chìN] | ‘qualification’ + ‘quality’ |

(Okell 1969, Green 2005)

Note that while this distinction between compound types is not made by Okell (1969), he does draw a distinction between “tightly” and “loosely” linked forms. He includes vowel reduction, which is the criterion that Green (2005) uses to establish the category of reducing compounds, as one of the criteria for tightly linked compounding. Other criteria include the non-reversability of the order of the stems, the voicing of the onset of the second member of the compound if it is voiceless when independent and the impossibility of the intrusion of other words within the compound. These properties all seem to hold for Green’s (2005) reducing compounds, as well.

In addition to the disyllabic words given in (4.3), when reducing compounds include words that already have supposed sesquisyllables, they can result in trisyllabic extended sesquisyllables, which have multiple adjacent minor syllables (4.5). The current chapter explores these types of words.

(4.5)

- a) [thΛ.mΛ.jé:] ‘rice-water’ < [thΛ.mìN] + [jé:] ‘rice’ + ‘water’
b) [kΛ.lΛ.bjé:] ‘India’ < [kΛ.là:] + [bjé:] ‘Indian’ + ‘country’

(Green 1995)

There are at least two competing proposals for the potential foot structure of Burmese reducing compounds like those in (4.5). In Chapter 1, I noted how iterativity is necessary for determining whether a language is iambic or trochaic and that in maximally di- or monosyllabic languages, this property is not available. I also discussed how the distinction between these two language types may be overstated such that systems may be mixed or – in the case of monosyllabic languages – simply ambiguous. Burmese presents an interesting case study in this regard.

First, Green (1995, 2005) proposes that major syllables are trochaic feet, thereby suggesting a one-to-one correspondence between syllables and feet in Burmese. This leads to the metrical structures in (4.6). Feet are given in parentheses, and prosodic word boundaries are indicated by curly brackets.

(4.6)

- a) {(pàN)} ‘flower’
b) {zΛ.(bwé)} ‘table’
c) {{{(cε?)}.{(shìN)}}} ‘turkey’
d) {{{?Λ.(jé:)}.{?Λ.(chìN)}}} ‘standard’
e) {thΛ.mΛ.(jé:)} ‘rice-water’
f) {kΛ.lΛ.(bjé:)} ‘India’

(Green 2005)

Green's (1995) analysis relies on the fact that Burmese lacks words composed of two light syllables, e.g. *z Λ ba Λ . He suggests that if the word structure were actually iambic, such (LL) feet would be permitted. However, given that many iambic languages allow heavy monosyllables as well-formed words, Green's (2005) criterion may not be sufficient. In contrast, if we assume the sesquisyllabic foot structure that I have suggested, then these words should be parsed as in (4.7).

(4.7)

- a) {(pàN)} 'flower'
- b) {(z Λ .bwé)}
- c) {{{(cɛʔ)}.{(shíN)}}}
- d) {{{(? Λ .jé:)}.{(? Λ .chìN)}}}
- e) {th Λ .(m Λ .jé:)}
- f) {k Λ .(l Λ .bjé:)}

The differences in these two analyses are represented in Figure 4.2.

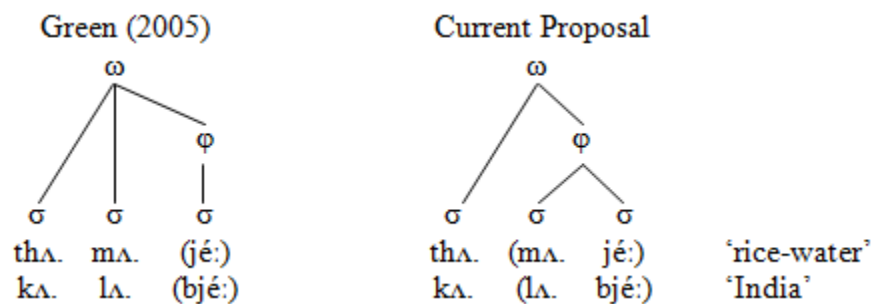


Figure 4.2: Possible footing of extended sesquisyllables in Burmese

In determining which analysis is correct, we turn to a language with potentially similar structural properties – San Martín Itunyoso (SMI) Trique, an Oto-Manguean language spoken in Oaxaca, Mexico. According to DiCanio (2008), morphological words in SMI Trique are minimally one syllable and maximally three, just as Burmese reducing compounds, although the

most common word type is disyllabic. The rightmost syllable in all words is necessarily heavy and stressed. In Trique, as in the other sesquisyllabic languages examined thus far, while all phonemes are licensed on the final syllable, only a subset are allowed on non-final syllables. SMI Trique stands apart, however, in that the permissible segment inventory is not identical in all non-final syllables, unlike in Burmese; instead, the antepenultimate syllable allows still fewer phonemes than the penult.

DiCanio (2008, 52) states, “The general pattern here is one of increasing phonological contrast toward right edge of the morphological word,” and indeed this is the pattern we find in vowels, whereby final syllables show five contrastive vowel qualities and contrastive vowel nasalization. In penultimate syllables, only four vowels are licensed while both /o/ and vowel nasalization are restricted. In antepenultimate syllables, only three vowels are licensed and of them /e/ possibly has restrictions on its distribution. The consonants pattern in the same way: of the 39 consonant phonemes in the language, only 15 occur in the penultimate syllable, while the antepenultimate syllable allows only 10. Although the antepenult does not permit any consonants that the penult disallows, /cn/ is only allowed in penultimate syllables, not in final syllables. One exception to this tendency is the distribution of consonant clusters, which are usually found only word initially. Nonetheless, the vowel and consonant distributions are quite striking in that they display a consistent relationship between markedness and proximity to stress.

I propose that this three-way difference in markedness is a top-down effect of prosodic structure. That is, footed syllables license more marked material than unfooted syllables, and stressed syllables license more marked material than unstressed syllables, as demonstrated in (4.8). If this footing is correct, Trique would appear to be a sesquisyllabic language, whereby

both unstressed and unfooted syllables might be considered minor syllables, albeit with different degrees of markedness.

(4.8)

- a) [ra.(ru.'βa)] 'breakfast'
- b) [ru.(ni.'ʔja)] 'tejocote fruit'

DiCanio (2008)⁸

But what do the SMI Trique data indicate about Burmese? If the markedness restrictions are suggestive of the phonological structure, then although my analysis is a priori possible for a language, the markedness restrictions Green (1995, 2005) reports for Burmese – that both minor syllables are subject to the same markedness restrictions – actually support his monosyllabic foot structure, not my proposal.

However, although not impossible, it is quite rare that syllable sequences which can be parsed remain unfooted in the output as Green (1995, 2005) suggests is the case for extended sesquisyllables in Burmese. In other words, unfooted footable syllable sequences are usually prohibited, so that multiple adjacent unfooted syllables are not allowed in the output of a language. The concept of forbidding unfooted syllables is not new. It was first argued for in Prince's (1980, 535) account of Estonian as a "general condition on [the] metrical composition of [a] word," and it simply stated that words must be exhaustively divided into feet. In fact, this prohibition is very widely enforced, as the various formalizations of the LAPSE constraint Table 4.4 reveal.

⁸ Tones omitted; stress and footing mine.

Table 4.4: Various prohibitions against sequences of unfooted syllables

<u>Syllable adjacency</u>	<u>Foot adjacency</u>
Gordon (2002), Alber (2005)	Kager (1994), Ishii (1996)
Rhythm is alternating; no two adjacent unstressed syllables.	Of every two stress units one must be parsed into a foot.
Elenbaas and Kager (1999)	Green & Kenstowicz (1995), de Lacy (2004)
Every weak beat must be adjacent to a strong beat or the word edge.	Adjacent unstressed moras or syllables must be separated by a foot boundary.
Selkirk (1984)	
Any weak position on a metrical level n may be preceded by at most one weak position on that level.	

In sum, if – as markedness patterns suggest – Green’s analysis is correct, this would place Burmese among the small number of languages that allow adjacent unfooted footable syllables. Given these facts, in what follows I investigate the phonetic properties of extended monosyllables to shed some light on these ambiguous word types. This in turn leads us back to the issue of what, if anything, phonetics can tell us about phonological structure.

4.3 HYPOTHESES

Before addressing the issue of the extended sesquisyllable in Burmese, I attempt to lay out the phonetic properties of vowels in minor syllables in Burmese in disyllabic words with [ʌ]

nuclei. Minor syllable vowels should be shorter than major syllable vowels and less peripheral in terms of formant structure due to lack of prominence. In addition, when a vowel occurs in a minor syllable as a result of compounding, like the underlined vowels in (4.9), as opposed to being the nucleus of a major syllable, it should again be shorter and more centralized.

(4.9)

/càN/ + /pò:/ = [cà.bò:]

floor + insect = bug

/θwà:/ + /jé:/ = [θà.jé:]

tooth + juice = saliva

Next, how do additional, i.e. antepenultimate, minor syllable vowels compare to penultimate minor syllable vowels in extended sesquisyllables? Regarding this second question, in a study asking similar questions, Chitoran and Hualde (2007) show that distance from stress has a significant effect on segment duration for vowels in Romance languages. For Brazilian and European Portuguese, as well as Romanian and Spanish, stressed vowels are longer than pretonic and pre-pretonic vowels. Differences between pretonic and pre-pretonic vowels, however, vary by speaker. In some cases pretonic vowels are significantly longer than pre-pretonic vowels, but in some cases the durations are not distinct. Although Burmese is a tone language rather than a stress language, because there is only one tone per prosodic word, we may expect distance from the heavy tone-bearing syllable to have some effect on duration of the nuclei of the minor syllable(s). However, given the variability in Chitoran and Hualde's (2007) results, we may find that speakers vary in their production.

The present experiment is designed to evaluate the claim that minor syllable vowels in disyllabic words are phonologically reduced in terms of formant values and duration. In addition, regarding extended sesquisyllables, there are a number of possible outcomes: First, phonetically, there may or may not be any gradient effects in terms of duration and formant values in terms of

distance from the prosodically prominent final syllable. If such effects are found, the antepenultimate syllable will be shorter and more centralized than the penultimate syllable. Still, if gradient effects are present, the phonological interpretation of them still remains an open question. This may suggest the sesquisyllable footing, i.e. a footed disyllable preceded by one unfooted syllable; however, the effect may be purely phonetic and not indicative of this phonological structure. If no gradient effects are found, we may take this as evidence for Green's (1995, 2005) monosyllabic foot. In other words, we may find phonetic gradiency without phonological categorality, but we will likely not find differences in phonological categorality which are not also reflected in the phonetics.

4.4 METHODS

4.4.1 *Participants*

Five native speakers of Burmese (two female and three male) were recorded in the phonetics lab at Cornell University. Participant age ranged from 21 to 60 ($\mu = 39$), and all speakers were born and grew up in Burma and four also have some proficiency in Mandarin.

4.4.2 *Stimuli*

Stimuli were randomized and presented to the participants in the Burmese script one word at a time. Participants were instructed to read each aloud three times in the frame sentence [mʌnega ___ lu? cʌnobjore] ('Yesterday, I said ____'). Burmese has both a formal and an informal register, and the former is usually used when reading a text. However, in the formal register, there is far less vowel reduction in compounding. Therefore, participants were instructed to read the words informally, as if they were talking to a friend. Speakers are well aware of this distinction, and although most thought it odd to read aloud in the informal register,

they were successful at doing so. Tokens produced in a formal register were considered errors and were either omitted from the analysis or else noted in the results presented below.

Table 4.5 shows the stimuli that were recorded. For each pair of words connected by an arrow, the word on the left side shows the free form of the word, and the word on the right side of the arrow shows the compounded form. For the non-reduced/reduced comparison, the words *floor/bug*, *fish/fish spawn* and *tooth/saliva* were used. For comparisons across minor syllables, the relevant stimuli are *India* and *rice water*. Corresponding vowel phonemes are underlined in each pair.

Table 4.5: Burmese wordlist

Monosyllables	Sesquisyllables	Extended Sesquisyllables	Non-reducing Compounds
/càN/ <i>floor</i>	→ /c _Λ .bò:/ <i>bug</i>		/cɛʔ.shíN/ <i>turkey</i> /ʔ _Λ .jé:.ʔ _Λ .chìN/ <i>standard</i>
/ηà:/ <i>fish</i>	→ /η _Λ .ʔú:/ <i>fish spawn</i>		
/θwà:/ <i>tooth</i>	→ /θ _Λ .jé:/ <i>saliva</i>		
/sà:/ <i>eat</i>	→ /z _Λ .bwè:/ <i>table</i>		
	/k _Λ .là:/ <i>Indian</i>	→ /k _Λ .l _Λ .bjé:/ <i>India</i>	
	/th _Λ .mìN/ <i>cooked rice</i>	→ /th _Λ .m _Λ .jé:/ <i>rice water</i>	
	/p _Λ .lú/ <i>ogre</i>		

4.4.3 Measurements

Formant and duration measurements were taken following the same procedure outlined in Section 2.4.3, which are repeated here. To obtain durations, the onset and offset of the second

formant were considered to be the beginning and ending of vowels, where possible. To obtain formant values, all sound files were downsampled to 10,000Hz and transformed into formant objects via a short-term spectral analysis with a window length of 25ms and a pre-emphasis of 50Hz. For males, the maximum value for three formants was set at 3,400Hz, and for females, 4,000Hz. Formant measurements were taken at vowel midpoints.

4.5 RESULTS

Results are presented in two parts. First, the experiment confirms that vowels in the minor syllables of disyllabic words are reduced in both quality and duration from their underlying, non-compounded forms. Second, in words with multiple minor syllables, vowel quality results are inconclusive and vowel duration results differ by stimulus, suggesting a reanalysis of prosodic word boundaries. In addition, results indicate that reduced vowels in both disyllabic words and extended sesquisyllables are phonological vowels rather than transitional vocoids. In this study, all three repetitions for this experiment were measured and included in the results. Box plots should be interpreted as explained in Section 2.5.

4.5.1 Reduction from Compounding

In general, vowels in the minor syllables of disyllabic words were found to be both shorter and more centralized than their counterparts in major syllables. Figure 4.3 shows the unreduced vowel /à/ as the nucleus of a monosyllable in *floor*, while Figure 4.4 shows the same vowel after being reduced and compounded to form the word *bug*.

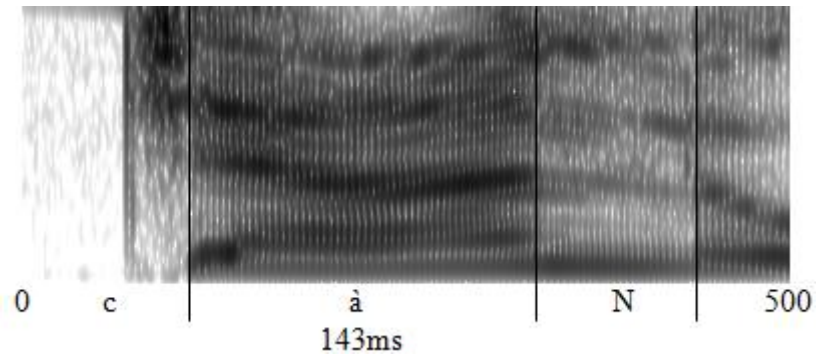


Figure 4.3: [Burmese] Spectrogram of [càN] 'floor'. Duration in ms.

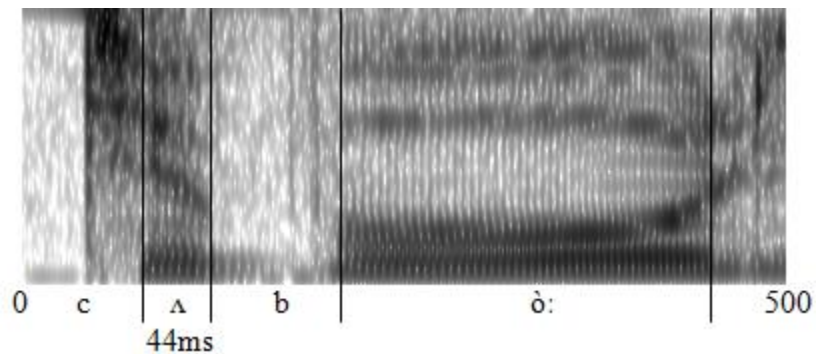


Figure 4.4: [Burmese] Spectrogram of [cΔ.bò:] 'bug'. Duration in ms.

Formant differences can be seen in Figure 4.5, which shows two vowels in two different conditions. First, the F1 of the vowel /à/ in *floor* is low when *floor* is compounded to create *bug*, as the vowel becomes much more schwa-like. This is also true of the vowel /à/ in *tooth*, which is reduced when compounded to create *saliva*. The /à/ in *floor* is more forward in the mouth than the /à/ in *tooth* presumably because the latter is preceded by /w/. Notably, there are five repetitions of the vowel in *bug* whose F1 is higher (>550Hz) and F2 is lower (<1500Hz) than the rest of the repetitions. These were all produced by the same speaker and presumably represent a dialectal or idiolectal difference. Nonetheless, they are still distinct from the realization of the same vowel in *floor*. Note that in Figure 4.5, the non-reduced vowels are represented by circles while the reduced vowels are represented by triangles.

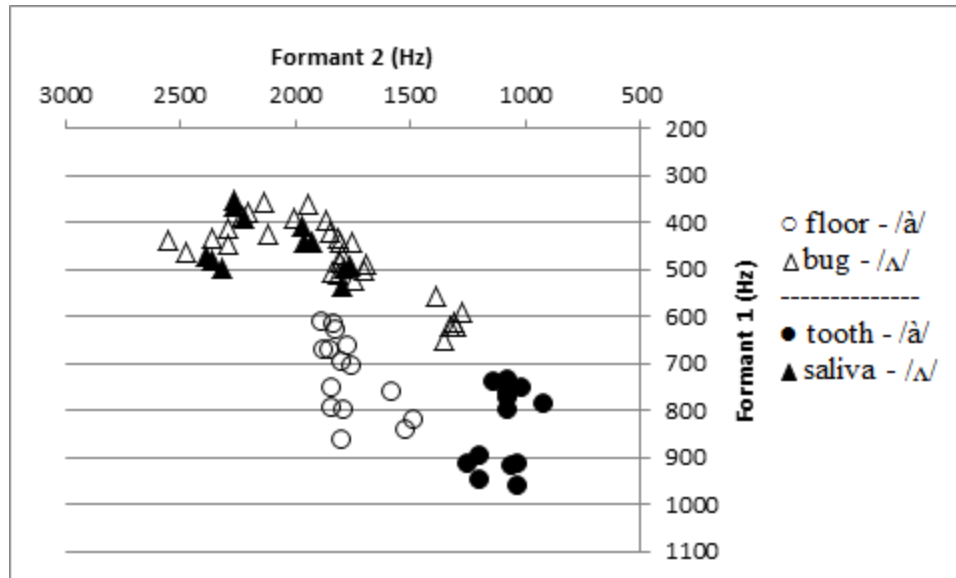


Figure 4.5: [Burmese] Vowel space for non-reduced [à] (circle) and reduced [Δ] (triangle)

As for duration, a linear regression shows that for the sesquisyllabic words in Table 4.5, minor syllable vowels are significantly shorter than major syllable vowels ($p < 0.0001$). As a control, durations were also measured for vowels in the disyllabic non-reducing compound, /cɛʔ.shiN/ ‘turkey’. This comparison is not ideal since /ɛ/ and /i/ are different vowels, and while we expect /i/ to be shorter than /ɛ/ due to height, /i/ should still be longer than MCVes in penultimate syllables of disyllabic words, given the propensity for MCVes to reduce. Unlike the sesquisyllabic forms, vowel durations for the two vowels in /cɛʔ.shiN/ were not found to be significantly different from each other ($p = 0.583$), likely because /i/ is subject here to word-final lengthening and is the locus of prosodic prominence while /ɛ/ is not.

In addition, as Figure 4.6 shows, /ɛ/ and /i/ are significantly longer than /Δ/ in purported sesquisyllabic words ($p = 0.0018$) and significantly shorter than word-final syllables in purported sesquisyllables, which here include /e, a, o, u/. Average durations for each of these are indicated in the column for “V”. Therefore, these durational differences may be due more to intrinsic differences between vowels than from metrical structure. More data are needed to make a strong conclusion.

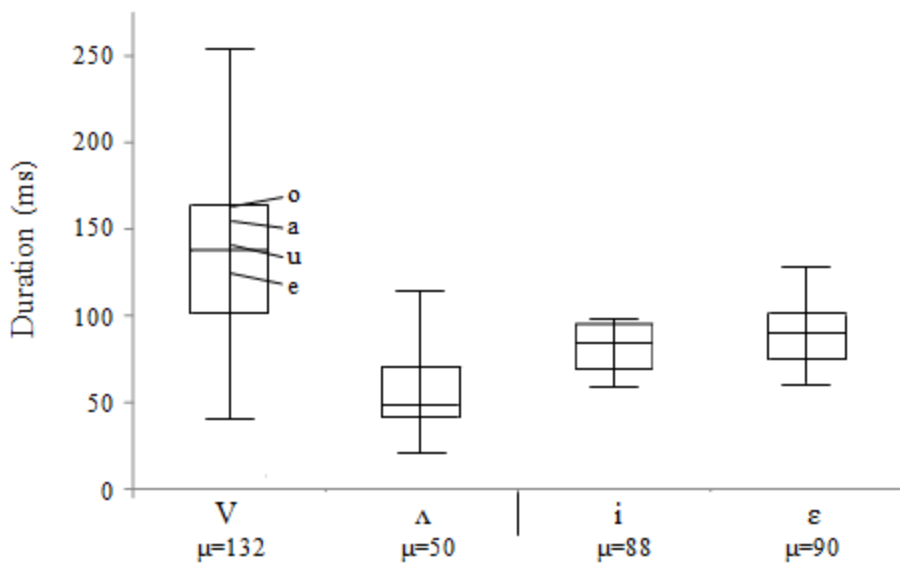


Figure 4.6: [Burmese] Durations of vowels in the final syllable of disyllables (V), the initial syllable of disyllables (Λ), the final syllable of /cεʔ.shiN/ (i) and the initial syllable of /cεʔ.shiN/ (ε)

4.5.2 Comparison across Minor Syllables

In comparisons between minor syllables in extended sesquisyllables, the results are much more variable. In general, the formant results are inconclusive across all speakers. The duration results, however, vary depending on speaker and stimulus. As such, the results for each stimuli are presented separately, beginning with [kΛ.lΛ.bjé:] (‘India’) and followed by [thΛ.mΛ.jé:] (‘rice water’). An example of [kΛ.lΛ.bjé:] is provided in Figure 4.7.

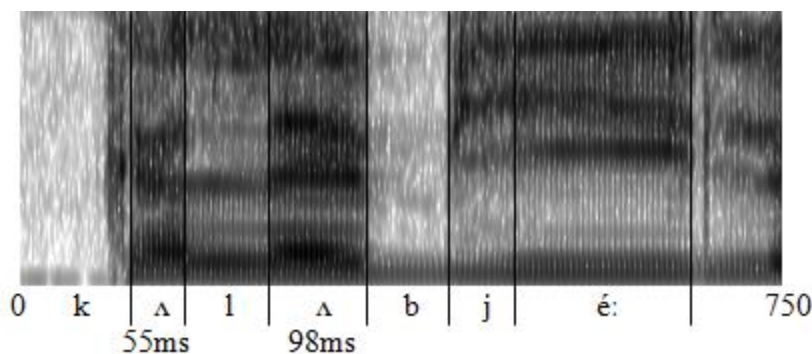


Figure 4.7: [Burmese] Purported multiple minor syllables in [kΛ.lΛ.bjé:] 'India'. Duration in ms.

In comparing minor syllable formant values against one another, post-hoc tests show that for [kΛ.lΛ.bjé:], there is a marginally significant difference in F1 between the two minor syllable

vowels ($p = 0.0451$), whereby the penultimate syllable is produced slightly higher in the mouth than the antepenultimate syllable, but there is no difference in F2 ($p = 0.805$), as can be seen in Figure 4.11. However, there is a greater distribution in terms of both F1 and F2 for the antepenultimate syllable than the penultimate syllable.

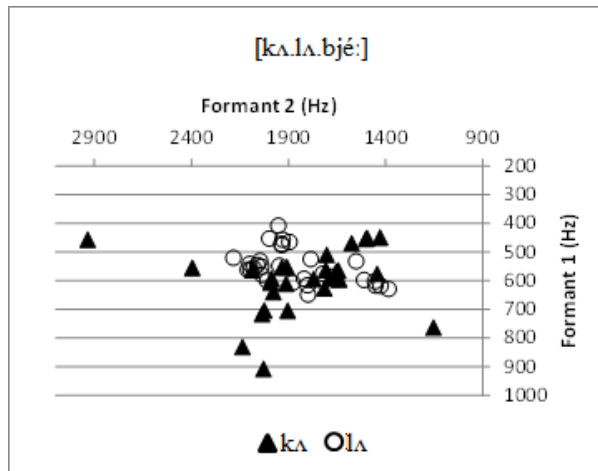


Figure 4.8: [Burmese] Vowel space for minor syllable nuclei in [kʌ.lʌ.bjé:]

Duration values are more telling but vary by speaker. When pooled across speakers, there is a significant three-way durational distinction by syllable ($p < 0.0001$), as shown in the box plot in Figure 4.9. When speakers are separated, as shown in Table 4.6, this pattern holds for only three of the five speakers. For speaker M2, only the antepenultimate syllable and the penultimate syllable are different, and for speaker F2, only the penult and the final syllable are different.

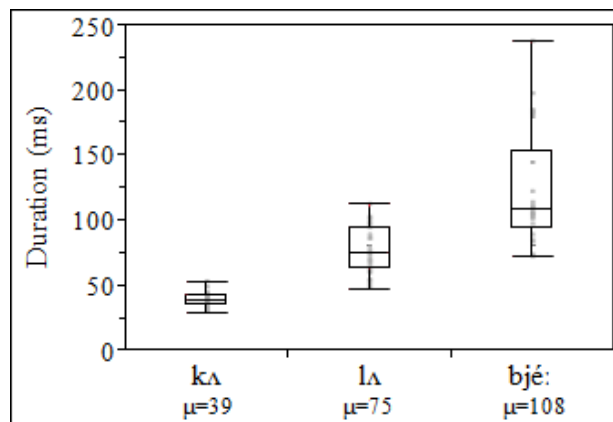


Figure 4.9: [Burmese] Durations (ms) for vowels in [kʌ.lʌ.bjé:]

Table 4.6: [Burmese] Mean durational differences (ms) between syllables by speaker for [kΛ.lΛ.bjé:]. Significant differences across the columns in each row are indicated by different shading.

Speaker	kΛ	lΛ	bjé:
M2	37	80	89
M1	44	85	111
F1	48	101	195
M3	37	63	97
F2	34	61	129

Results for [thΛ.mΛ.jé:] are somewhat different. As seen in the spectrogram in Figure 4.10, the formant values and duration for the penultimate syllable vowel are quite different than for the penult in [kΛ.lΛ.bjé:].

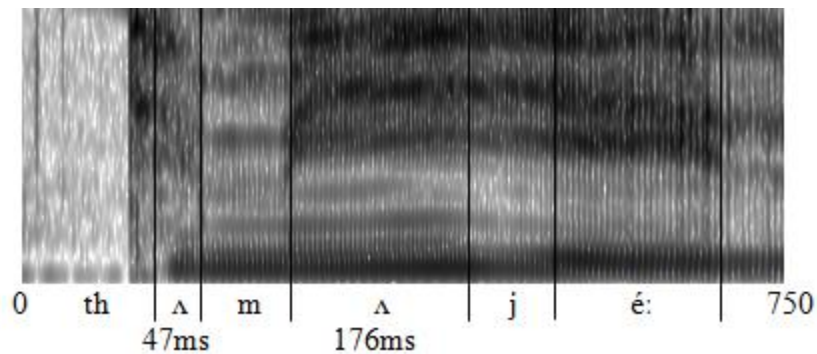


Figure 4.10: [Burmese] Purported multiple minor syllables in [thΛ.mΛ.jé:] 'rice water'. Duration in ms.

The antepenultimate syllable vowel has a significantly lower F1 ($p = 0.0007$) and higher F2 ($p < 0.0001$) than the penultimate syllable vowel. Although some amount of variation is expected in these data since the vowel plots represent tokens from both men and women, the vowels in the minor syllables of [thΛ.mΛ.jé:] are far more separated as a group than are the vowels in [kΛ.lΛ.bjé:]. I suggest that the penultimate syllable vowel in [thΛ.mΛ.jé:] is being pulled forward because of the following palatal glide. The differences between minor syllable

vowels in both words suggest that further experiments are needed in which phonetic environment is better controlled.

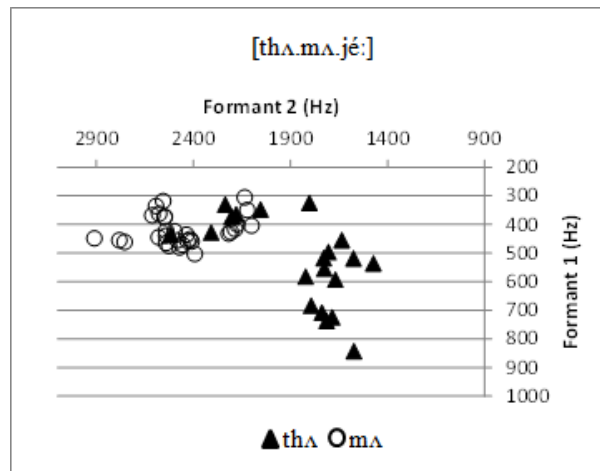


Figure 4.11: [Burmese] Vowel space for minor syllable nuclei in [thΛ.mΛ.jé:]

In terms of duration, Figure 4.12 shows that when speakers are pooled the antepenultimate syllable vowel is significantly shorter than the medial syllable ($p < 0.0001$), but the medial and final syllables are not durationally distinct ($p = 0.1469$). When speakers are separated, this pattern is maintained for three of the five speakers, but for speakers M3 and F2, there is a significant three-way distinction in vowel duration (Table 4.7). Note also that these duration measurements are potentially problematic due to difficulties in segmentation because of the glide in the /Λjé:/ sequence.

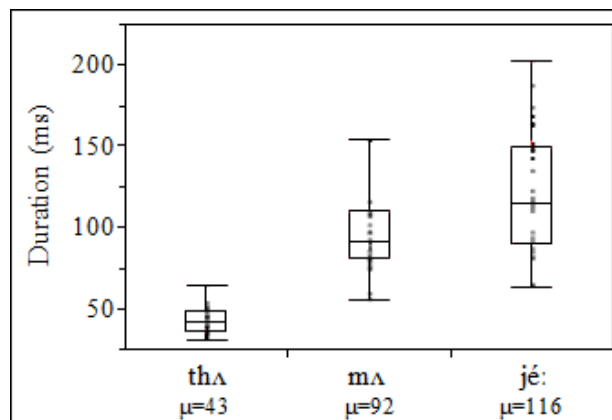


Figure 4.12: [Burmese] Durations (ms) for vowels in [thΛ.mΛ.jé:]

Table 4.7: [Burmese] Mean durational differences (ms) between syllables by speaker for [thΛ.mΛ.jé:]. Significant differences across the columns in each row are indicated by different shading.

Speaker	thΛ	mΛ	jé:
M2	44	87	93
M1	48	88	84
F1	45	174	176
M3	35	85	128
F2	52	97	125

4.5.3 *Summary of Results*

The acoustic data analyzed here support the claim that the vowel nuclei of minor syllables in disyllabic compound words are both shorter and more centralized than the vowel nuclei of non-compounded monosyllables. In addition, results suggest that in words with multiple minor syllables, the nuclei of those syllables differ by stimulus and speaker and that neighboring sounds may play an important role in their realization. Regarding duration, we do see some type of gradient effect in all cases although the nature of the gradiency, whether three-way or two-way, is variable.

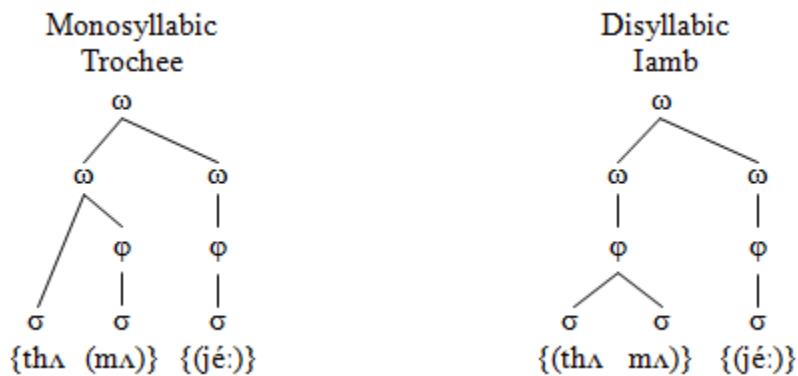
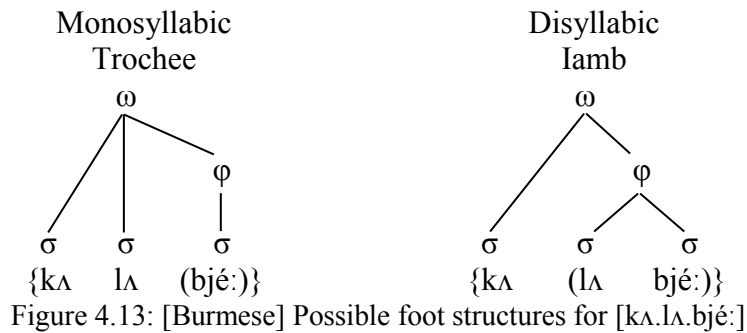
4.6 DISCUSSION

Given the data above, we have no reason to suspect that minor syllables in Burmese are akin to the Type A transitional vocoids found in Khmer. Indeed, vowel nuclei in minor syllables are more similar to Type B minor syllable nuclei in Bunong. However, Burmese is still distinct from languages like Bunong in that it allows multiple minor syllables. And while the presence of multiple minor syllables is not unique in mainland Southeast Asian languages, word forms in Burmese do differ from the canonical notion of the sesquisyllable and from the analysis proposed

in the preceding chapters by which sesquisyllables should be thought of as disyllabic iambs.

Therefore, the above results leave us with two questions: first, what is the metrical structure of Burmese extended sesquisyllables? Second, what does this say about sesquisyllables more generally?

First, given these results, the overall foot type of Burmese words is still ambiguous. The phonetics fully support neither the monosyllabic trochee analysis nor the disyllabic iamb analysis. However, the phonetic results do suggest that there are differences between the two purported extended sesquisyllables tested here. While [k Λ .l Λ .bjé:] may have either structure in Figure 4.13, I suggest that [th Λ .m Λ .jé:] actually comprises two prosodic words and, at least for some speakers, is not a reducing compound as Green (1995, 2005) suggests, but instead has one of the structures in Figure 4.14.



The present results suggest that this analysis is idiolectal and while it is probably relevant for speakers M2, M1 and F1, it may not be the case for speakers M3 and F2. A potentially influential factor is that reduction is representative of the less formal register, which, as stated above, is generally not used when reading. In other words, if the experimental design were different, speakers M2, M1 and F1 might have patterned more like the other speakers. Nonetheless, the duration values in Table 4.7 suggest that [th_Λ.m_Λ.jé:] comprises two prosodic words for at least some speakers. **MORE DATA NEEDED**

Regarding the second question – what do these results mean for sesquisyllables more generally? – the most important implication regards the status of the minor syllable. If the foot structure in Burmese really is ambiguous, and indeed if the distinction between trochaic and iambic languages in general is not so clear cut, then what is the phonological structure of the minor syllable? If sesquisyllables are disyllabic iambs, then minor syllables are definitively unstressed non-final syllables. But Burmese provides evidence that one or even both of the minor syllables might be unfooted, reminiscent of de Lacy's (2004) analysis of Maori discussed in Section 1.3.3.1. Indeed, from a historical perspective, sesquisyllables may reflect both of these structures at different points of change over time in the pathway from trochees to iambs, as suggested in Figure 4.15.

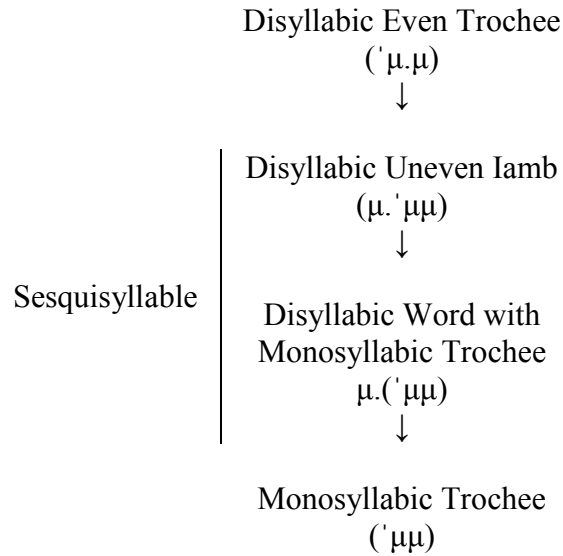


Figure 4.15: The sesquisyllable on the pathway of language change

Finally, although this chapter has offered a starting point for an analysis of extended sesquisyllables in Burmese, further studies should be carried out to support and/or clarify the phonetic results here. In addition to the need for more speakers, the set of extended sesquisyllables tested should be expanded in order to control for potentially confounding effects of formant differences due to coarticulation with surrounding consonants. However, given the relative rarity of reducing compounds in Burmese which actually result in extended sesquisyllables, as opposed to non-reducing compounds, this is methodologically difficult. Still, no matter what the outcome of future studies, the Burmese data in conjunction with the Khmer and Bunong results demonstrate that previous characterizations and descriptions of sesquisyllables and minor syllables in particular must be reconsidered in terms of their prosodic constituency.

CHAPTER 5

CONCLUSION

5.1 SUMMARY OF THEMES AND STRUCTURE

I have had two main goals in this dissertation. First, I sought to provide an analysis of the word type known as the *sesquisyllable*, based on both descriptive phonetic properties and a structural account of prosodic constituents. I argued that much of the disagreement among different accounts of sesquisyllables and minor syllables was due to an over-emphasis on one or the other of these approaches. By understanding the sesquisyllable as a maximally disyllabic iambic word whose word-initial syllable nucleus is a reduced yet phonological vowel, i.e. not just a phonetic transition, I was able to tease apart previous synchronic and diachronic accounts of the sesquisyllable. Using acoustic data from three purportedly sesquisyllabic languages in Chapters 2 – 4, I showed that the cross-linguistic variation in so-called sesquisyllables can be accounted for when word structure is viewed in light of articulation.

Second, I addressed the issue of what sorts of phonological conclusions can be drawn based on phonetic data. Taking an articulatory perspective, I suggested that the phonetic correlates of the phonological categories I use to describe sesquisyllables are quite strong, if not conclusive. This is apparent in Chapters 2 and 3, in which I demonstrated that purported sesquisyllables in Khmer are in fact monosyllables with word-initial consonant clusters with intervening transitional states, while purported sesquisyllables in Bunong are maximally disyllabic words with reduced phonological vowels. In other cases, as reflected in the Burmese data in Chapter 4, phonetic experiments, while suggestive, cannot definitively provide a model of prosodic structure.

In conclusion, I now outline a number of contributions that this dissertation makes to the study of sesquisyllables and to linguistic theory and practice more generally. Although they overlap, these contributions are divided into two types: descriptive and theoretical. In Section 5.2, I discuss the benefits of focusing on mainland Southeast Asian languages. I review the different word types that have historically been considered sesquisyllables and note that the voicing specification of transitional states has been a confounding factor in narrowing the definition of the sesquisyllable. In Section 5.3, I review the prosodic structure of the sesquisyllable and discuss how this structure can be viewed in light of a gestural framework. With this integrated approach in mind, I then explore how articulatory timing can be understood in terms of dynamical oscillators, which I suggest underlies the ubiquity of disyllabic word types and provides evidence against the rarity of the sesquisyllable.

5.2 DESCRIPTIVE CONTRIBUTIONS

A simple yet important contribution of this dissertation is the addition of phonetic data and careful description of a number of understudied languages to the literature. While Khmer and Burmese have the official status of being national languages, they are still insufficiently documented. A descriptive account is most important for Bunong because it is a minority language and has far fewer speakers than either Khmer or Burmese.

Along these same lines, I have indirectly advocated for a simple methodological approach to studying articulation in less accessible languages. In particular, drawing conclusions about articulation based on more easily measureable acoustic properties allows for analysis of a wider range of languages. Although studying acoustic correlates of articulators is not a replacement for direct observation of the articulators themselves, given the difficulty (though not impossibility) of conducting articulatory experiments outside of a laboratory setting, interpreting

acoustic data in light of articulation helps us begin to understand gestural configurations in a wider variety of languages. This knowledge can in turn be used to gain a better understanding of phonology more generally.

Throughout the dissertation, I sought to clarify ambiguity surrounding the term *sesquisyllable* which has emerged from the range of cross-linguistic descriptions of it. I maintain that the term has traditionally conflated a number of word types that are better understood using more standard notions of prosodic phonology. In particular, these include both (i) monosyllables with word-initial consonant clusters containing an intervening (voiced) excrescent vocoid and (ii) disyllabic iambs whose weak syllables contain either phonological schwa, wedge or a restricted set of vowels. We can clarify the meaning of the term *sesquisyllable* by defining it to include only the latter word type. The schwa-like material in the word-initial consonant clusters of monosyllables is qualitatively different from the phonological vowels found in the weak syllable of disyllabic iambs. These word types are labeled Type A and Type B, respectively, and each is represented in the model in Figure 5.1. This figure shows sesquisyllables as further defined by having one disyllabic foot per prosodic word. Understanding sesquisyllables as iambs allows us to easily capture the markedness differences between the penultimate and final syllables, which are characteristic of all descriptions of sesquisyllables.

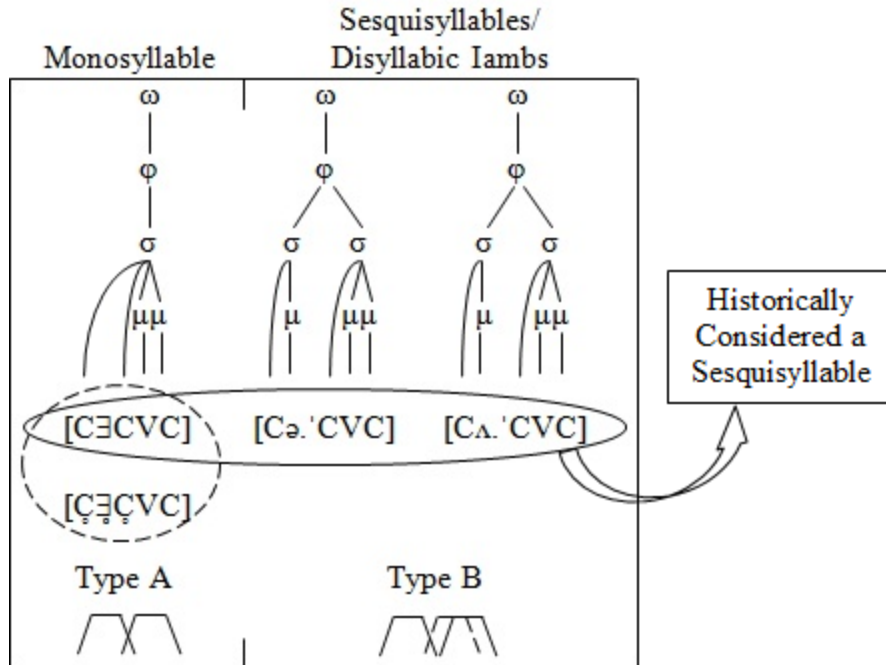


Figure 5.1: Traditional differences in sesquisyllables

This model also helps us understand the two kinds of sources contributing to the confusion of the term *sesquisyllable*. First, all the word types located in the solid circle have each at some point been called sesquisyllables. However, as outlined in Section 3.6.2, it is possible for the phonological status of a vowel to change over time due to intergenerational misinterpretation (cf. McMahon et al. 1994). Language learners may interpret Type B MCVEs as Type A MCVEs and vice versa. Therefore, what might be a sesquisyllable for one generation of speakers may not be for the next. The acoustic similarity between MCVE types may also explain difficulties field linguists have in differentiating the phonological nature of the two.

Second, the word types located in the dashed circle in Figure 5.1, which differ only in voicing, underlie another descriptive problem. Although I argue that neither should be considered a sesquisyllable, monosyllables with intervening voiced Type A MCVEs quite often have been while those with voiceless transitions have not been. This may be motivated by a desire to capture the difference in voicing between these structurally identical word types since

the voicing difference is likely more impressionistically salient than is the structural similarity. I suggest this has led to the over-use of the term *sesquisyllable* to label only the voiced types. Instead, I offer a unified treatment for these monosyllabic transitions whether voiced or voiceless and show that durational similarities of the transitional periods as well as the predictability of voicing attests to the similarity of these two word types, and that neither should be considered a sesquisyllable.

5.3 THEORETICAL CONTRIBUTIONS

In addition to its descriptive contributions, this dissertation offers a number of theoretical insights. In Section 5.3.1, I provide an overview of these contributions, the most salient of which is the proposal that an integrated model of prosody and articulation is necessary to account for the sesquisyllable and minor syllable. In Section 5.3.2, I expand on this idea by exploring the role of prosodic constituents as oscillators, which can account for the propensity of binary branching prosodic constituents. I summarize the claims in Section 5.3.3.

5.3.1 Overview of Contributions

. While phonetics and phonology have been traditionally seen as separate modules that interact with one another via some type of mapping (Cohn and Chitoran 2009), Articulatory Phonology differs from this view in regarding them as one entity. Because of this difference, metrical formalisms and articulatory formalisms are not usually integrated. However, I have argued that both are useful and necessary in understanding the sesquisyllable. Defining the sesquisyllable in terms of prosodic structure – and in particular as a maximally disyllabic iamb – not only provides a way to distinguish it from monosyllables with word-initial consonant clusters but also provides a model on which to test the structure of extended sesquisyllables in languages like Burmese. Understanding the mechanisms which underlie the minor syllable in terms of

articulation also provides a more explicit definition of the differences between epenthesis and excrescence.

I present here in Figure 5.2 a comprehensive version of the schema on which my sesquisyllable analysis is based, in order to highlight the differences between the excrescent vocoid (Type A) and unstressed schwa or wedge (Type B). Excrescent vocoids do not have an associated gesture while unstressed vowels do. There are certainly other indicators of the phonological reality of MCVEs that do not require a gestural approach – like determining if and how MCVEs participate in morphophonological processes. Yet in languages with extremely reduced morphology, which is often the case for sesquisyllabic languages, determining whether a sound has an associated gesture or not may be the only way to determine its cognitive reality.

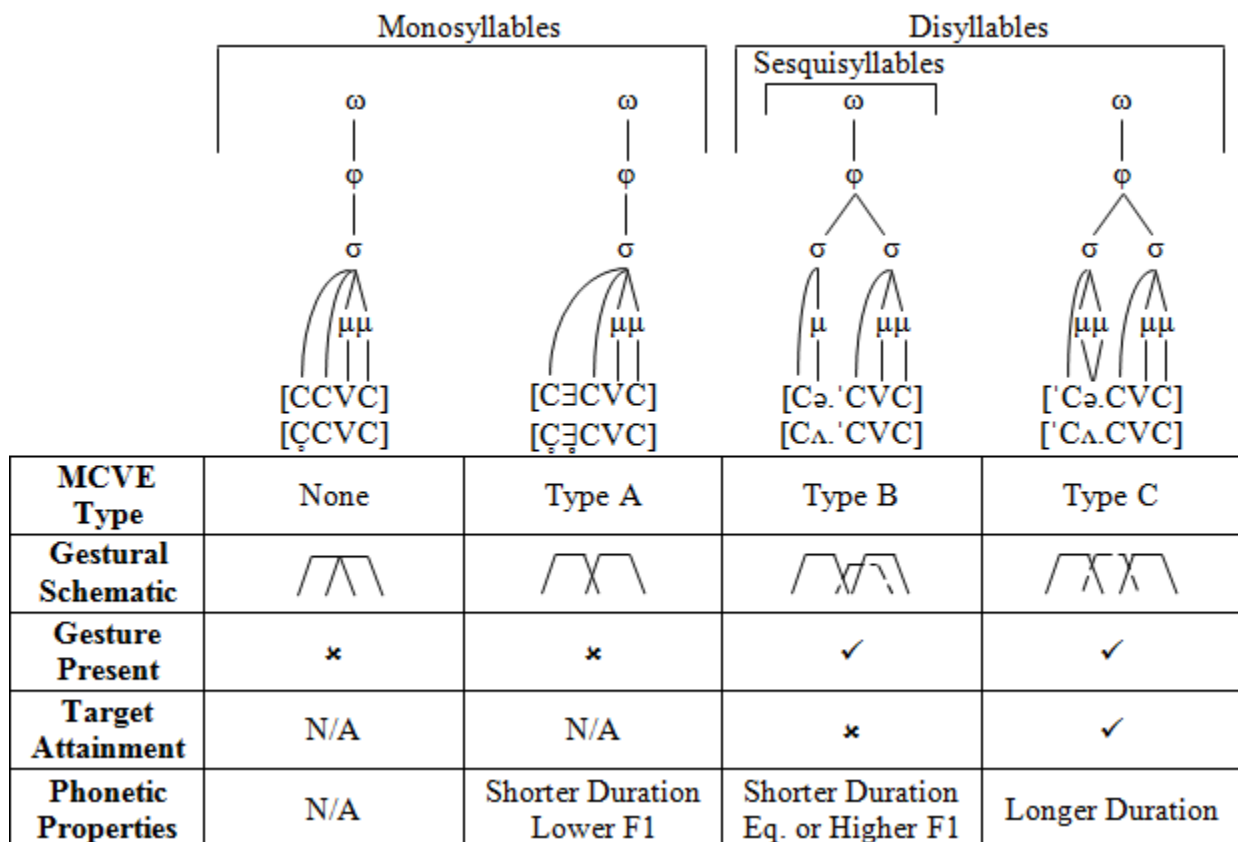


Figure 5.2: Integrated model of sesquisyllables

If an integrated approach like that represented in Figure 5.2 is indeed correct, then we must consider its implications. I propose that the integration of prosodic and articulatory models is particularly relevant to the study of sesquisyllables as it relates to binarity. Building upon Goldstein et al. (2007), O'Dell and Nieminen (2009), Tilsen (2009a,b), *inter alia*, I claim that prosodic constituents can be modeled as dynamical oscillators which interact with one another via frequency-locked relationships and that the simplest of these relationships which still maintains some amount of alternation is a 1-to-2 relationship. This can account for the ubiquity of binary branching constituents, hence disyllabic feet and sesquisyllables. The following section explores how this relationship functions.

5.3.2 *Prosody and Oscillation*

If formally motivated constraints are violable as discussed in Section 1.3.1, then why do we find such cross-linguistic regularity in metrical patterns and in rhythmic binarity in particular? This question is important here because I have defined the sesquisyllable as a word (or foot) which bifurcates into exactly two syllables. I suggest this pervasive binarity is due to other functionally motivated forces: alternation and unarity. While the principle of alternation has long been noted in grammars of language, unarity is discussed far less often. Yet both of these principles can be motivated by a theory of speech which considers prosodic constituents to be modeled as dynamic oscillators that are timed with one another. I argue that the pervasiveness of disyllabic word types both within and beyond mainland Southeast Asia is a result of the interaction of these two forces. In particular, alternation contributes to the weak-strong pattern that is characteristic of disyllabic iambs, while unarity ensures that there is a one-to-one correspondence between the prosodic word and the foot. Given this theory of speech timing, the

sesquisyllable should not be considered a unique word type but instead one type of response to competing forces on language.

In the following discussion, I establish that, as has long been noted, alternation is a fundamental property of speech (Section 5.3.2.1). Then, I explore the overwhelmingly binary relationship between prosodic constituents, first by reviewing how binarity has been stipulated in various grammatical formalisms (Section 5.3.2.2), then by grounding it in a theory of oscillation (Section 5.3.2.3).

5.3.2.1 Alternation as a Fundamental Property of Speech

Rhythmic alternation – or patterned changes in intensity, duration, loudness, etc. – aids in the perception of stress. That is, listeners perceive a sound to be stressed when it is more prominent than the sounds around it. Furthermore, through alternation, strings of sounds exhibit prominence relations such that they are perceived as being successively grouped, and this perception is contingent on the duration and loudness of their elements. One well known account of rhythmic alternation is Hayes' (1995) theory of metrical stress, motivated by the Iambic/Trochaic Law, which states that sounds contrasting in intensity form groupings with initial prominence (trochees) and sounds contrasting in duration form groupings with final prominence (iambs). Although this principle has been questioned (cf. Kager 1993), it provides a good basis for understanding the implications of rhythmic alternation. Indeed, Hayes' (1995) intention is to characterize feet as the effect of a “purely rhythmic principle” (79). As such, he cites literature from other metrically driven fields, like music and literature, which seem to support the Iambic/Trochaic Law. For example, Fraisse (1982) notes the prominent role of a periodic pulse in both perception and in physical activities. This suggests that basing a model of

rhythm on the psychological reality of periodicity, as in the case both of Hayes (1995) and, as we will see, of Coupled Oscillator Theory, is indeed a promising approach.

Alternations in stress have also been used as evidence for multiple prosodic constituents. As noted in Section 1.3.1, syllables are usually taken to be the domain of stress, and the existence of feet is motivated by grouping patterns found among syllables. Although in many cases, the prosodic word is defined by morphological (not necessarily phonological) properties (Downing 2006), it too can be seen as a stress domain. In previous accounts it was defined as such (Nespor and Vogel 1986) in the sense that there is often one primary stress per prosodic word, which is apparent relative to other unstressed or secondarily-stressed elements in the word. Additionally, partial evidence for phrasal groups is taken from the domain of pitch accent, whereby one word per phrase is marked with a primary accent, distinguished by its relative intensity, loudness and/or length relative to other words in the phrase. Indeed, this property of culminativity, or maximally one stress per stress domain, is often used to characterize grammatical units.

In addition to stress patterns, other prosodic constituents show alternation as well. For example, we find languages in which moras can alternate in terms of weight, although stress in such languages is computed over syllables. But because two moras may compose two separate light syllables or just one heavy syllable, alternations in stress may not be apparent at the syllable level. In addition to syllables and moras, Zec (2006) notes several other phonological phenomena that alternate at the word level. These include tonal accent in Creek (Hayes 1995), rhythmic lengthening in Chicasaw (Munro and Willmond 1994; Hayes 1995), and vowel quality in Old Church Slavonic (Isačenko 1970).

5.3.2.2 *Binarity as a Stipulation of Grammar*

Perhaps the simplest type of alternating pattern is the juxtaposition of two elements that differ in some property, in which one is more prominent than the other. This basic grouping of two elements, referred to as binarity, is extremely common in prosodic systems and in groupings of alternating strong and weak syllables, in particular. Yet, binarity has been formalized in grammars in a largely stipulative manner rather than as a phonetically or cognitively motivated principle relating to alternation. Indeed, Downing (2006) suggests that the concept of binarity is often taken to be inviolable as in Hayes (1995), Prince and Smolensky (2004) and McCarthy and Prince (1993a). Furthermore, Itô et al. (1996, 241) state that many grammars “enshrine strict binarity as a desideratum of all prosodic (or: linguistic) structure, by means of a specific constraint to that effect. While this line of attack is not without merit, its directness has a price: No further phonological explication of the unmarkedness of strictly binary structure is deemed possible, or necessary.” Before I present a possible grounding for alternation – and as a result binarity – via oscillation, I first lay out several ways in which binarity has been formalized in grammar, usually as rules or constraints.

Binarity has been formalized in numerous ways. Early accounts, like Kiparsky (1978), make use of binarity at the level of the syllable, suggesting that stress should be represented by binary branching trees. Subsequently, enforcement of binarity in feet was formalized by Prince (1980) in his work on quantity in Estonian. Prince (1980) follows Lehiste (1965) in suggesting that the foot is a “bisyllabic” unit in order to account for the domain of vowel lengthening in Estonian, i.e. foot-final vowels are lengthened if foot-initial vowels are short. Since then, one of the most prevalent formalizations of binarity constraints has been the FT-BIN constraint (McCarthy and Prince 1993a), which states that feet must be binarily parsed on the level of the

mora or the syllable. FT-BIN (5.1), as described in Prince and Smolensky (2004) is primarily understood as a constraint on word minimality, preventing degenerate word types.

(5.1)

FT-BIN: Feet are binary at some level of analysis (μ , σ).

A constraint on binarity is often formalized as two different constraints, one on binary minimality (that some unit i must dominate at least two units $i-1$) and another on binary maximality (that some unit i must dominate at most two units $i-1$). For example, in his analysis of modern Hebrew, Ussishkin (2005) enforces minimality requirements through a constraint PROSODICBRANCHING, which states that a prosodic category i must not be coextensive with a single member of prosodic category $i-1$. Ussishkin (2005) follows Itô et al. (1996) in formalizing the binary maximality constraint in terms of edge alignment. That is, if a prosodic constituent is binary branching, then every edge of it necessarily aligns with the edge of an immediately dominant constituent. This is formalized as Hierarchical Alignment: “Every prosodic constituent is aligned with some prosodic constituent, containing it” (Itô et al. 1996, 242). In this way, alignment is conceptualized in terms of prominence, reflecting again the pressure toward alternation within the linguistic system.

Making use of an apparatus already in the grammar, i.e. prosodic structure, eliminates some unnecessary stipulations. There are a number of other benefits to formulating binarity constraints in terms of prosodic categories and alignment, as well as to splitting one constraint into two. First, instead of stipulating binarity in terms of feet or syllables alone, these constraints predict that every prosodic constituent is binary branching. This not only allows for testable predictions, but also introduces more regularity into the grammar. Second, splitting binarity into two constraints versus one substantially clarifies the predictions of the model. If one of the

constraints, e.g. binary minimality, is a necessary property of the system, while binary maximality is not, the model makes drastically different predictions. Third, attempts at formulating binarity constraints using alignment properties seem to be somewhat less stipulative than a constraint like FT-BIN alone. Nonetheless, although such constraints can be accurate descriptors and predictors of a grammar, it is preferable that constraints be motivated in a cognitive or phonetic reality.

One such attempt to avoid stipulation by motivating binarity is proposed by Selkirk (1984). While she maintains that binarity is a fundamental part of stress patterns, she acknowledges the stipulativity of this claim. As a solution, she suggests motivating binary patterns via a rhythmic theory of stress, in which syllables align with some representational metrical structure. In this way, binary rhythmic alternation can be viewed as a derived outcome. Unfortunately, this assumes the alignment of syllables with the metrical tier is governed by grammatical rules. Therefore, although binarity may be motivated theoretically in this model, the theoretical structure is at best descriptive and fails to explain binary patterning.

Clearly both the evidence for alternation in speech as well as the range of approaches used to account for binarity within grammar suggest that alternation is a fundamental property of language. The issue at hand, then, is determining what sort of mechanism is motivating this propensity for alternation. I suggest that theories of articulation, and in particular of the role of oscillators in speech timing can shed light on this question.

5.3.2.3 Oscillation as Motivation for Binarity and the Sesquisyllable

Constraints on the prosodic hierarchy are best expressed as relations between prosodic constituents modeled as oscillators, which are roughly defined as any regularly repeating process (O'Dell and Nieminen 2009). For dynamical systems like speech, oscillators are thought to have

a natural phase, which regulate timing. In addition, oscillators can be coupled together, and this coupling is commonly expressed as a frequency-locking relation. It is precisely these relations between oscillators that give rise to the alternations so commonly seen in language. This alternation provides some explanation for the commonality of sesquisyllables/disyllables in mainland Southeast Asia as well as more generally.

5.3.2.3.1 Introduction to Oscillators

Although modeling speech as a system of oscillators was introduced into linguistics via the framework of Articulatory Phonology, the earliest research in AP was focused almost exclusively on gestural systems and did not address questions about higher level prosodic constituents. Apart from this, research in cognitive science has produced models of speech as an oscillatory system (Port 2003, O'Dell and Nieminen 2009). By now, in addition to the more broadly focused Coupled Oscillator Theory (COT) literature, prosodic constituents are conceived of as types of oscillators in the AP literature, as well. However, oscillators are motivated somewhat differently within the different frameworks. Indeed, several papers have questioned how these two models of speech – task dynamics and coupled oscillators – might interact with one another (Byrd and Saltzman 1998, Goldstein et al. 2007, Tilsen 2009a,b).

In an AP account, the oscillator is used primarily for the purpose of planning and activation of gestures. Once the planning has taken place, the gestural score is input to the task-dynamic model and the articulation is implemented. In a COT model, however, O'Dell and Nieminen (2009) suggest that “hierarchical synchronization does not necessarily imply planning at a certain level” (6) and that duration data are not sufficient to determine if and to what extent the phasing relations of oscillators are planned independently or are phased in *medias res*. As a result, the syllable “tier” is motivated differently in AP than it is in COT.

Using a gesture-based articulatory framework, Goldstein et al. (2007) seek to lay out the biological underpinnings of the syllable and explain the cross-linguistic ubiquity of CV syllable types. They assume that each gesture is associated with a planning oscillator that triggers the production of the gesture. These oscillators are coupled to each other in graphs that encode information about relative timing, and consonant and vowel gestures are either in-phase or anti-phase coupled with one another. The prevalence of the CV syllable shape is taken to be a consequence of the relative stability of the phasing relations between consonants and vowels.

In a more COT-based approach, O'Dell and Nieminen (2009) simply assume the presence of a mora oscillator as well as a relation between stress and the foot oscillator. Their data show that Finnish CVCVV words have a longer duration than CVVCV words (Figure 5.3). To capture this difference, it is necessary for them to have some way to vary parameters within a foot, thereby functionally motivating the syllable oscillator. In terms of computation, they require that the numerical value of the first mora is greater than that of the third, which is accomplished by setting a higher relative coupling strength of the syllable oscillator for the first mora than the third. Thus, the notion of the syllable oscillator is motivated by the need to introduce extra numeric variables to explain durational differences at the word level.

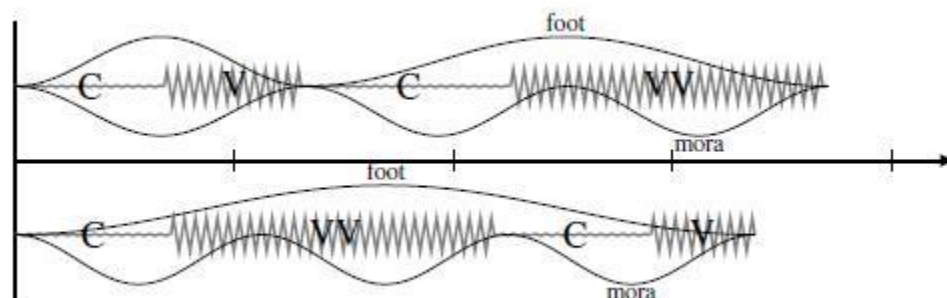


Figure 5.3: CVCVV and CVVCV words in Finnish (O'Dell and Nieminen, 2009, Figure 6)

Despite the fact that prosodic constituents are modeled as oscillators in both the AP and COT literature, the roles of the oscillators differ sharply between the two. For the present, I

simply assume that prosodic constituents *can* be modeled as a system of oscillators without considering their role (or lack thereof) in planning. The next step is to understand how conceptualizing prosodic constituents – and syllables in particular – as oscillators accounts for the pervasiveness of sesquisyllables – or iambs – in mainland Southeast Asian languages. I suggest this tendency is best explained through the mechanism of frequency locking.

5.3.2.3.2 Frequency Locking

The notion that frequency locking seeks to capture is not a new one. Early accounts of rhythm in language were largely concerned with the issue of isochrony, i.e. the notion that time is rhythmically divided into equal groups, and of whether these groups are based on stress patterns or on syllable patterns (Pike 1945, Abercrombie 1967). However, studies have shown that people perceive isochronous patterns when listening to speech, even if the timing is not actually isochronous (Lehiste 1977, Dauer 1983). Indeed, Benguerel and D'Arcy (1986) suggest that some rhythmic irregularity may be necessary for the perception of regularity. On the assumption that prosodic constituents can be thought of as oscillators in themselves, Cummins and Port (1998, 147) define rhythm with the following two parameters:

- (i) Rhythm is manifested as the temporal binding of events to specific and predictable phases of a superordinate cycle.
- (ii) Rhythm in speech is functionally conditioned.

Understanding prosodic constituents as coupled oscillators allows us to account for variations in these rhythmic patterns while maintaining the position that they are somehow coupled. Studies have shown that isochrony is more easily perceived at regular cyclic rhythms (Cummins and Port 1998, Tilsen 2009b, *inter alia*). In other words, when two oscillators are periodic and are related in such a way that the periodicities are related as simple multiples

(Figure 5.4), they are in a more stable rhythmic relationship. For example, a 1:2 frequency-locking relation between feet and syllables is more stable than a 5:6 relation. As Tilsen (2009b, 845) states, “Correlations between target phase and variability may arise due to the relatively lesser stability of 1:3 frequency-locking compared to 1:2 locking.”

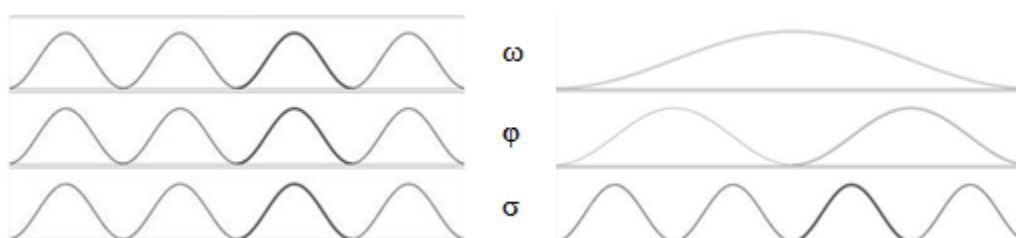


Figure 5.4: 1:1 and 1:2 frequency-locking relations

5.3.2.3.3 Unarity

It follows that the most stable frequency-locking relation of all is a 1:1 relation, where frequencies are not only simple multiples but are in fact isomorphic. Given that each tier of the prosodic hierarchy can be conceived of as an oscillator which may be coupled with any other adjacent oscillators, we may expect to find languages in which some prosodic constituents are in one-to-one relation with each other. Indeed, I suggest just such a relation holds for word and foot oscillators in the case of languages that are maximally sesqui- or disyllabic. Even in some languages with larger word types, e.g. Indonesian, there is a one-to-one relation between the word oscillator and certain morphological categories like roots (Cohn, p.c.). In fact, Downing (2006) argues that the lack of binary pressure on the prosodic word should be taken as evidence that it is not actually a prosodic constituent. Logically, we may also expect to find languages in which *all* of the prosodic constituents are in a one-to-one correlation with one another. In fact, as noted in Chapter 1, Vietnamese may be such a case (Figure 5.5).

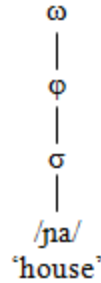


Figure 5.5: Redundant prosodic structure in Vietnamese

Using evidence from reduplication and compounding, Schiering et al. (2010) show that some prosodic constituents in Vietnamese are redundant. Because Vietnamese does not distinguish between polysyllabic words and strings of monosyllabic words for the purpose of stress, the syllable level is taken to be the domain for phonotactics, stress and tone, and the phrasal level to be the basis for rhythm. Based on perceptual experiments by Ingram and Nguyen (2006) and Nguyen and Ingram (2006), Schiering et al. (2010) also conclude that although stress alternation may exist at both the phrasal and word levels in Vietnamese, these alternation patterns are entirely redundant. In their own experiments, they find that native speakers do not perceive differences between disyllabic compound words and phrasal units made of two monosyllabic words.

However, languages like Vietnamese, i.e. languages with consistent one-to-one frequency locking between levels, are rare indeed even in mainland Southeast Asia. Brunelle and Pittayaporn (2012) report that of 235 Austroasiatic languages listed in the Mon-Khmer Etymological Dictionary database (Sidwell and Cooper 2007-2011), only nine are maximally monosyllabic. Even in languages which are usually considered monosyllabic, like Mandarin, monosyllabic words often occur not in isolation but in compounds. For example, Brunelle and Pittayaporn (2012) note that *gú* ‘bone’ only occurs in combination with *tóu* ‘head’ as *gútou* ‘bone’ (Norman 1988; Duanmu 2000), which is lexicalized as a single phonological word.

5.3.2.3.4 Functionally Motivated Constraints

With a better understanding of frequency-locking relations, we are now able to consider what we take to be the functionally motivated restrictions on the prosodic hierarchy noted in Section 1.3.1. Three properties of frequency-locking relations, i.e. relations between prosodic constituents, are cross-linguistically consistent and as such require some motivation: First, frequency-locking relations involve only integers, not fractions. Second, while ternary, quaternary and larger relations are possible, we have no evidence for stress systems larger than quaternary, even while a priori, an infinite number of possible frequency-locking relations between oscillators exists. Third, along with the ubiquity of alternation, languages with 1:1 relations are also relatively rare, and no language has a 1:1 correspondence of prosodic constituents for every level of the prosodic hierarchy at the same time. Yet, given the cognitive simplicity of one-to-one frequency-locking relations, we may expect unary systems to be much more common than they actually are. Frequency-locking relations between oscillators provide an explanation for each of these issues.

With respect to non-integer frequency-locking relations, all fractional ratios can be converted into whole number ratios, e.g. $1:\frac{1}{2} = 2:1$, $3:\frac{7}{8} = 24:7$, etc.; therefore the representation of the values is a non-issue. Yet practically, it is unclear as to what means to say that a foot comprises 1.5 syllables. This harkens back to the descriptive but not particularly explanatory conception of the sesquisyllable as one-and-a-half syllables. In addition, because perception is categorical (Lehiste 1977), there is no reason to assume a listener would hear a rhythmic relation as $1:\frac{1}{2}$ instead of 2:1.

Regarding larger frequency-locking relations, experiments on the interpretation of rhythm in music provide evidence for the psychological preference for simpler rhythmic ratios, i.e. ratios

with small least common multiples (LCMs). Povel (1981) shows that when participants attempt to tap in synchrony with various beats, ratios near 1:2 tend to be assimilated to that pattern. In these cases, trained musicians and non-musicians perform about equally. With more complex patterns, like 1:3 and 1:4, trained musicians perform better than non-musicians, and these patterns in general are more difficult. Krumhansl (2000, 5) notes that experiments by Friasse (1982) “suggest a psychological limit of about two duration categories, most often related by a 1:2 ratio, in rhythmic patterns”, although Gabrielsson (1988) argues that there may be as many as four duration categories; however, we see no evidence for categories greater than four. Indeed, Selkirk (1984) suggests that quaternary groups seem to be “felt” as two binary groups instead.

Finally, given a model of coupled oscillators, there is a functional preference for unarity, yet unary systems are relatively rare. This is because to achieve the perception of contrast, there must also be alternation. If these two factors are conceptualized as competing pressures on a grammar, there are several possible results. First, if the pressure for unarity is greater than the pressure for alternation, unary maximality is achieved, as in languages like Vietnamese. If, however, the pressure for alternation is greater than that of unarity, there can be various results, which all lead to non-equivalent ratios of frequency-locking relations. These could manifest as 1:2, 1:3, and so on although they are constrained by the psychological pressures mentioned above. The more similar the ratio, the more stable the frequency-locking relation. Therefore, next to a 1:1 frequency-locking relation, a 1:2 frequency-locking relation is the most stable. Indeed, binarity is identical with 1:2 frequency locking, and therefore, the cross-linguistic ubiquity of binarity in language is not at all surprising. This is a harmonic way to resolve the pressure for alternation with frequency locking.

Nonetheless, although binarity is remarkably common in the languages of mainland Southeast Asia and beyond, unarity still has a role. For example, a number of languages have syllabic maxima at the word or foot levels. In particular, many languages claimed to have sesquisyllables are maximally sesquisyllabic/disyllabic, e.g. Bunong, Kammu, Pacoh, etc. In these cases, there is a unary relationship between the word and foot oscillators. If redundant tiers are eliminated, maximally sesquisyllabic languages may be better represented by a structure which lacks feet altogether.

In general there is strong cognitive pressure for abstract linguistic units, i.e. prosodic constituents, to be timed with one another via simple integer frequency-locking relations. The pervasiveness of binary relationships, and hence disyllables, results from the competing pressures of rhythmic alternation and maximally simple 1:1 frequency-locking relations. These tendencies were well documented in phonological theory before the introduction of Articulatory Phonology or the integration of Coupled Oscillator Theory as a mechanism of speech timing. However, a theory of oscillatory relationships through frequency locking allows us to motivate these pressures on speech which have long been used to motivate structures within metrical and prosodic theories like stress placement, foot types, etc. The characterization of the sesquisyllable as a syllable-and-a-half is inherently problematic in the context of binary relationships, and it is better understood as a binary branching word or foot, i.e. as a disyllable.

5.3.3 Summary

An articulatory approach to understanding the linguistic tendencies behind the sesquisyllable is a useful way to understand the pervasiveness of the sesquisyllable in mainland Southeast Asia. This integration of prosodic phonology with oscillation demonstrates how the ideas presented in previous, more segment-based accounts can be successfully integrated into an

articulatory framework. In this theoretical context the sesquisyllable is best understood to be a maximally disyllabic word resulting from one-to-one frequency-locking between the word and foot oscillators, with prominence on the final syllable and phonological and/or phonetic reduction on the initial – or minor – syllable.

5.4 FUTURE DIRECTIONS

There are a number of avenues for further research prompted by the work in this thesis. Here I present just two directions, although there are likely many more. First, I address the potential for articulatory experiments based on my phonetic findings. Second, I discuss additional word types that might be investigated in light of our revised definition of the sesquisyllable. These include *super extended sesquisyllables* with more than two minor syllables, sesquisyllabic words with more than two degrees of markedness and words whose foot types are ambiguous.

5.4.1 *Articulatory Experiments*

A major limitation of the dissertation is the lack of direct articulatory evidence. While the phonetic studies presented in Chapters 2 – 4 provide informative acoustic results, the ways in which these results are mapped onto articulatory parameters are a matter of interpretation. Nonetheless, as stated in Chapter 1, there is a precedent for making articulatory claims based on acoustic correlates (cf. Davidson 2006b). Still, as a future avenue of research, the results presented in this dissertation could be used as hypotheses for more direct articulatory studies of the status of minor syllable vowels. The starting point should be a confirmation of the claims about minor syllable nuclei in languages like Khmer and Bunong based on the acoustic results presented in this thesis.

First, word-initial consonant clusters in Khmer should be evaluated much like clusters in Ridouane and Fougeron's (2011) study on Tashlhiyt Berber. I have concluded that there are not phonological vowels intervening in these clusters in the subset I investigated. Articulatory experiments could confirm or disprove this claim and could shed much more light on sibilant-initial clusters, which are difficult to interpret based solely on spectrographic evidence. These experiments emerge from the following hypothesis:

H1: Word-initial consonant clusters in Khmer are composed of consonant gestures, and what is often interpreted as a minor syllable nucleus results from an intergestural transition with no identifiable target.

Second, I have proposed that while Bunong minor syllable nuclei do indeed have targets, their targets are usually not achieved because of their shorter duration due to effects of stress and word length. On the other hand, target attainment can be obscured by overlap with other gestures. Only an articulatory study can detail the movement of the articulators, allowing us to tell whether or not a target has in fact been reached as well as the duration of that target attainment. Type B MCVEs in Bunong can be compared both with Type A MCVEs and Type C MCVEs in monosyllables to determine relative durational differences between MCVE types. These predictions are formalized in the following hypothesis:

H2: Purported minor syllable nuclei in Bunong have an associated gesture and target, yet because their duration is shorter than stressed vowels, that target will not always be reached; therefore the realization of minor syllable vowels will be more strongly affected by neighboring consonants than will the realization of stressed vowels.

5.4.2 *Other Sesquisyllable Types*

Besides the canonical cases of sesquisyllables mentioned in this dissertation and even those languages just slightly beyond the most conservative definition of the sesquisyllable, i.e. those with extended sesquisyllables, there are a number of other factors that might bear on our understanding of the sesquisyllable. I mention three of them here. First is a word type which, in addition to the extra minor syllable such as those permitted in Burmese, has an even larger number of minor syllables. Second, some sesquisyllabic-like word types, like those found in SMI Trique, also potentially have multiple minor syllables albeit with differing markedness restrictions. Finally, in addition to types of particular words, a more comprehensive study of sesquisyllables must revisit languages which blur the line between iambicity and trochaicity and consider how sesquisyllables can be understood in light of both foot types.

5.4.2.1 *Super Extended Sesquisyllables*

In addition to the extended sesquisyllables found in Burmese, there is evidence for what we might call *super extended sesquisyllables*, or words with more than two minor syllables. Although rare, such word shapes are found in Hatam, a language isolate of Irian Jaya, for example. Among other word shapes, Hatam does have a class of monomorphemic words that look like what are traditionally described as sesquisyllables (5.2).

(5.2) Hatam

a) [tʌ.'bɔr] 'arrow'

b) [nʌ.'ɲay] 'three' (Reesink 1999)

Like other sesquisyllabic languages, Hatam's stress pattern is largely iambic, except when personal pronouns are suffixed, in which case the pronoun receives stress. According to Reesink (1999), just as in monosyllables, stressed syllables in disyllabic words allow all possible

phonemes and combinations of phonemes. Pre-tonic syllables, however, only allow a subset of segments and combinations of segments. Apparent minor syllables are consonantly restricted in that only the nasals [m] and [n] are allowed in the presence of a minor syllable vowel, to the exclusion of [ŋ] and [ɲ]. Interestingly, however, if nasals are syllabic, all four nasals are allowed. Geminates are not allowed in minor syllables, and there does not appear to be a voicing contrast in the onsets of minor syllables. Consonant clusters, i.e. stop+/r/ or /sr/, are allowed in stressed syllables but not in pre-tonic syllables. Stressed syllables may have nasals and voiceless stops in coda position, but pre-tonic syllables allow nasals only.

Like the consonant inventory, the Hatam vowel inventory is also smaller in pre-tonic syllables than in stressed syllables. In stressed syllables, Hatam allows [i, e, a, o, u]; however, in unstressed syllables, only [i, a, u] are allowed if there is no onset (5.3a), and if an onset is present, only an MCVE is possible, although Reesink (1999) suggests that there is some variation in the possible realization of this vowel, but gives only one example of a token other than a MCVE. The choice of [i, a, u] is conditioned by the consonants in the unstressed syllable. If the onset of the major syllable is [h], the minor syllable nucleus is identical to the nucleus of the stressed syllable, and [h] is voiced (5.3b). In contrast, in polymorphemic words, i.e. those with prefixes, although the pre-tonic vowel is still reduced, its realization is much more predictable according to its environment.

(5.3) Hatam

- | | | |
|----|------------|------------|
| a) | [a.'kow] | ‘ant’ |
| | [un.'doy] | ‘hornbill’ |
| b) | [da.'fiat] | ‘to break’ |
| | [bu.'fiun] | ‘heavy’ |

(Reesink 1999)

Hatam differs from canonical sesquisyllabic languages and even from languages like Burmese in that it allows more than two pretonic syllables. Polymorphemic words consisting of pronominal clitics and verbal inflection may contain up to at least four minor syllables (5.4).

(5.4) Hatam

a)	/dV.go/	[də.'go]	‘cut up’
b)	/nV - dV.go/	[nə.də.'go]	‘1EXC - cut up’
c)	/nV - bV - dV.go/	[nə.bə.də.'go]	‘1EXC - INC - cut up’
d)	/bV - dV - bV - dV.go/	[bə.də.bə.də.'go]	‘PUR - 1SG - INS - cut up’

(Reesink 1999)

The word shapes found in Hatam raise a number of interesting questions and challenges for my definition of the sesquisyllable and metrical theory in general. Markedness patterns suggest it is possible that feet in Hatam are maximally one syllable. In other words, they might not actually be sesquisyllabic as I have defined the sesquisyllable but may resemble more closely the footing that Green (1995, 2005) proposed for Burmese. If this is the case, the Burmese data may also need to be reanalyzed, but this is a matter of future research.

5.4.2.2 *Sesquisyllables with Gradient Markedness*

In addition to words with more than two minor syllables, potential sesquisyllables with more than two levels of markedness in languages like SMI Trique, as described in Chapter 4, should also be investigated further. These words in Trique differ from those in Burmese and Hatam in that they display gradient markedness effects. While phonetic studies on languages like Burmese are limited in what they can tell us about phonological structure, an investigation of a language like Trique may be more enlightening. In addition to phonological gradience, does Trique have phonetic gradience? While we expect that phonetic gradience might be present

without phonological gradience, as appears to be the case for some speakers of Burmese, it would be surprising if phonological gradience were present in the absence of phonetic gradience.

The markedness differences in Trique trisyllabic words seem to indicate something about prosodic structure. Those differences should have an effect on the phonetics. Therefore, we may expect antepenultimate syllables to be shorter than penultimate syllables and for their vowels to be more reduced. Whatever the phonetic results, this raises interesting questions about the nature of minor syllables – Do they actually have to be parsed? Is there more than one type of minor syllable? Should languages with more than one minor syllable still be considered sesquisyllabic? These questions are all open for future research.

5.4.2.3 *Trochaic Sesquisyllables*

Finally, broadening the scope of the study of sesquisyllables to include trochees will no doubt provide a more comprehensive and accurate understanding of this word type. For one, it raises questions regarding language contact and structural change. Cham, for example, has become a sesquisyllabic/monosyllabic language (depending on register) while closely related Acehnese remains trochaic (Durie 1985). While sesquisyllables have been traditionally understood as quintessentially Southeast Asian, I have indicated that a number of languages outside of Southeast Asia proper may also be sesquisyllabic, suggesting that universal language tendencies may play just as important a role in the evolution of language as do genetic relatedness between languages or even language contact situations.

In addition, as noted in Chapter 1, disyllabic words in a number of otherwise iambic languages seem to be trochaic. Languages like these raise a number of questions for models of language change. Particularly in Southeast Asia, maximally disyllabic languages overwhelmingly tend to move from trochees to iambs (Brunelle and Pittayaporn 2012). If this is

the case, then we must ask why disyllables in a number of languages seem to resist iambicity (e.g. Hixkaryana, Lake Iroquoian, Sierra Miwok, Kashaya, etc.). What, if anything, does this indicate about the structure of maximally monosyllabic languages that were at one point maximally disyllabic? Perhaps languages whose words are maximally disyllabic and iambic are in a state of reduction. If, as Brunelle and Pittayaporn (2012) claim, trochaic languages always pass through a phase of iambicity on their way to monosyllabicity, maximally disyllabic iambs might actually be conceived of as a type of reduced trochee, as a sort of snapshot in time of an unstable state (cf. van de Vijver 1998 for the lack of iambs as phonological primitives). And in this historical sense, perhaps sesquisyllables are actually unique after all.

REFERENCES

- Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine Publishing Company.
- Alber, B. (2005). Clash, lapse and directionality. *Natural Language & Linguistic Theory* 23(3), 485–542.
- Allen, S. (1953). *Phonetics in Ancient India*. London: Oxford University Press.
- Auger, J. (2001). Phonological variation and Optimality Theory: Evidence from word-initial vowel epenthesis in Vimeu Picard. *Language Variation and Change* 13, 253–303.
- Beckman, M., & Kingston, J. (1990). Introduction. In M. Beckman & J. Kingston (Eds.), *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech* (pp. 1–16). Cambridge: Cambridge University Press.
- Benguerel, A. P., & D’Arcy, J. (1986). Time-warping and the perception of rhythm in speech. *Journal of Phonetics* 14(2), 231–246.
- Benjamin, G. (1976). An outline of Temiar grammar. *Oceanic Linguistics Special Publications* 13, 129–187.
- Bennett, J. F. (1995). *Metrical Foot Structure in Thai and Kayah Li: Optimality-Theoretic Studies in the Prosody of Two Southeast Asian Languages* (Ph.D. Dissertation). University of Illinois, Urbana-Champaign.
- Bequette, R. (2008). *Participant Reference, Deixis, and Anaphora in Bunong Narrative Discourse* (M.A. Thesis). Graduate Institute of Applied Linguistics, Duncanville, TX.
- Bhaskararao, P., & Ladefoged, P. (2009). Two types of voiceless nasals. *Journal of the International Phonetic Association* 21(2), 80–88.

- Bishop, N. (1996). A preliminary description of Kensiw (Maniq) phonology. *Mon-Khmer Studies* 25, 227–253.
- Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 206–244). Cambridge: Blackwell.
- Blood, H. (1966). *A Reconstruction of Proto-Mnong* (M.A. Thesis). Summer Institute of Linguistics, University of North Dakota.
- Boersma, P., & Weenink, D. (2012). *Praat: doing phonetics by computer*. Retrieved from www.praat.org.
- Bradley, D. (1980). Phonological convergence between languages in contact: Mon-Khmer structural borrowing in Burmese. In *Proceedings of the 6th Annual Meeting of the Berkeley Linguistics Society* (pp. 259–267). Berkeley: Berkeley Linguistics Society.
- Bradley, D. (1982). Register in Burmese. (D. Bradley, Ed.) *Papers in South-East Asian Linguistics No. 8: Tonation*, 117–132.
- Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook* 3, 219–252.
- Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201–251.
- Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press.
- Browman, C., & Goldstein, L. (1992a). Articulatory phonology: An overview. *Haskins Laboratories Status Report on Speech Research*.

- Browman, C., & Goldstein, L. (1992b). “Targetless” schwa: An articulatory analysis. *Haskins Laboratories Status Report on Speech Research*, 194–219.
- Brunelle, M. (2004). Monosyllabicization in Eastern Cham. In W. Khanittanan & P. Sidwell (Eds.), *Papers from the 14th Annual Meeting of the Southeast Asian Linguistics Society* (Vol. XIV, pp. 43–58).
- Brunelle, M., & Pittayaporn, P. (2012). Phonologically-constrained change: The role of the foot in monosyllabization and rhythmic shifts in Mainland Southeast Asia. *Diachronica* 29(4), 411–433.
- Burenhult, N. (2001). Jahai phonology: A preliminary survey. *Mon-Khmer Studies* 31, 29–45.
- Butler, B. (Accepted). Registrogenesis in Bunong: The interaction of phonetics and phonology. *Journal of the Southeast Asian Linguistics Society*.
- Butler, B. (In Progress). Bunong. In M. Jenny & P. Sidwell (Eds.), *The Austroasiatic Languages*. Brill.
- Butler, B. (2012). A gestural deconstruction of the minor syllable. Presented at the Mainland Southeast Asian Languages: The State of the Art in 2012, Leipzig.
- Butskhrikidze, M. (2002). *The Consonant Phonotactics of Georgian*. Utrecht: LOT.
- Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics* 83, 1–19.
- Byrd, D., & Saltzman, E. (1998). Intra-gestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26(2), 173–199.
- Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. (C. Gussenhoven & N. Warner, Eds.) *Laboratory Phonology* 7, 419–447.

- Chitoran, I., & Hualde, J. I. (2007). From hiatus to diphthong: the evolution of vowel sequences in Romance. *Phonology* 24(1), 37.
- Cho, T. (2011). Laboratory phonology. In N. Kula, B. Botma, & K. Nasukawa (Eds.), *The Continuum Companion to Phonology* (pp. 343–368). London: Continuum International Publishing Group.
- Cho, Y. Y., & King, T. H. (2003). Semisyllables and universal syllabification. In C. Fery & R. van de Vijver (Eds.), *The Syllable in Optimality Theory* (pp. 183–212). Cambridge: Cambridge University Press.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology 1: Between the Grammar and the Physics of Speech* (pp. 283–333). Cambridge: Cambridge University Press.
- Clements, G. N., & Keyser, S. J. (1981). *A three-tiered theory of the syllable* (Occasional Paper, Center for Cognitive Science No. 19).
- Clements, G. N., & Keyser, S. J. (1983). *CV Phonology: A Generative Theory of the Syllable*. Cambridge: MIT Press.
- Cohn, A. C. (2005). Truncation in Indonesian: Evidence for violable minimal words and AnchorRight. In *Proceedings of NELS 34* (Vol. 1, pp. 175–189).
- Cohn, A. C., & Chitoran, I. (2009). Complexity in phonetics, phonology, gradience, categorality, and naturalness. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupe (Eds.), *Approaches to Phonological Complexity* (pp. 21–46). Berlin: Mouton de Guyer.
- Cohn, A. C., & McCarthy, J. (1998). Alignment and parallelism in Indonesian phonology. *Working Papers of the Cornell Phonetics Laboratory* 12, 53–137.

- Cohn, A. C., & Riehl, A. (2012). The structure of nasal-stop sequences: Evidence from Austronesian. (B. Butler & M. Renwick, Eds.) *Cornell Working Papers in Phonetics and Phonology*, 21–37.
- Cooper, J., & Cooper, N. (1966). Halang phonemes. *Mon-Khmer Studies* 2, 87–98.
- Crowhurst, M. (1991). *Minimality and Foot Structure in Metrical Phonology and Prosodic Morphology* (Ph.D. Dissertation). University of Arizona.
- Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26, 145–172.
- Dantsuji, M. (1987). An acoustic study on the distinction of place of articulation for voiceless nasals in Burmese. *Folia Linguistica* 21, 281–291.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 52–62.
- Davidson, L. (2006a). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34, 104–137.
- Davidson, L. (2006b). Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 63, 79–112.
- Davidson, L., & Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production. In M. Tsujimura & G. Garding (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics*.
- de Lacy, P. (2004). Maximal words and the Maori passive. In J. McCarthy (Ed.), *Optimality Theory in Phonology: A Reader* (pp. 495–512). Blackwell.
- de Lacy, P. (2011). Evaluating the quality of stress data. Presented at the 2nd UConn Workshop on Stress and Accent.

- DiCanio, C. (2005). Expressive alliteration in Mon and Khmer. *UC Berkeley Phonology Lab Annual Report*, 337–393.
- DiCanio, C. (2008). *The Phonetics and Phonology of San Martín Itunyoso Trique* (Ph.D. Dissertation). University of California, Berkeley.
- Diffloth, G. (1976a). Minor-syllable vocalism in Senoic languages. *Oceanic Linguistics Special Publications* 13, 229–247.
- Diffloth, G. (1976b). Expressives in Semai. *Oceanic Linguistics Special Publications* 13, 249–264.
- Diffloth, G., & Zide, N. (1992). Austro-Asiatic Languages. In (W. Bright, Ed.) *International Encyclopedia of Linguistics*. New York.
- Downing, L. (2006). *Canonical Forms in Prosodic Morphology*. Oxford University Press.
- Duanmu, S. (2000). *The Phonology of Standard Chinese*. New York: Oxford University Press.
- Durie, M. (1985). *A Grammar of Acehnese on the Basis of a Dialect of North Aceh*. Dordrecht: Foris Publications. [Revised version of 1984 ANU Ph.D. Dissertation].
- Elenbaas, N., & Kager, R. (1999). Ternary rhythm and the lapse constraint. *Phonology* 16(3), 273–329.
- Enfield, N. (2005). Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34, 181–206.
- Ferlus, M. (1982). Spirantisation des obstruantes mediales et formation du system consonantique du vietnamien. *Cahiers de Linguistique - Asie Orientale* 11(1), 83–106.
- Ferlus, M. (1992). Essai de phonétique historique du khmer (Du milieu du premier millénaire de notre ère à l'époque actuelle). *Mon-Khmer Studies* 21, 57–89.

- Ferlus, M. (2009). What were the four divisions of Middle Chinese? *Diachronica* 26(2), 184–213.
- Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. (A. Albright & T. Cho, Eds.) *UCLA Working Papers in Linguistics* 7, 71–116.
- Flemming, E., & Johnson, S. (2007). Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association*, 37(1), 83–96.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music*. New York: Academic Press.
- Gabrielsson, A. (1988). Timing in music performance and its relation to expression. In J. Sloboda (Ed.), *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition* (pp. 27–51). Oxford: Oxford University Press.
- Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20(2), 269–337.
- Gafos, D. (1998). A-templatic reduplication. *Linguistic Inquiry* 29(3), 515–527.
- Gaudes, R. (1978). Zur phonologischen Relevanz der sogenannten Register im modernen Kambodschanischen. *Mon-Khmer Studies* 7, 39–59.
- Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology* 16(1), 29–54.
- Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. *ICPhS XVI*, 241–244.
- Gordon, M. (2002). A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20, 491–552.

- Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32, 141–174.
- Green, A. (1995). The prosodic structure of Burmese: A constraint-based approach. *Working Papers of the Cornell Phonetics Laboratory* 10, 67–96.
- Green, A. (2005). Word, foot, and syllable structure in Burmese. In J. Watkins (Ed.), *Studies in Burmese Linguistics*. Canberra: Pacific Linguistics.
- Green, T., & Kenstowicz, M. (1995). The lapse constraint. In *Proceedings of the 6th Annual Meeting of the Formal Linguistics Society of the Midwest* (pp. 1–15).
- Grice, M., Röttger, T., Ridouane, R., & Fougeron, C. (2011). The association of tones in Tashlhiyt Berber (pp. 775–778). Presented at the ICPhS XVII, Hong Kong.
- Hall, N. (2004). *Implications of vowel intrusion for a gestural grammar*. Unpublished ms. University of Haifa.
- Hall, N. (2011). Vowel epenthesis. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1576–1596). West Sussex: Wiley-Blackwell.
- Hanson, O. (1906). *A Dictionary of the Kachin Language*. Rangoon: Baptist Board of Publications (1954).
- Harms, R. (1976). The segmentalization of Finnish “nonrules”. *Texas Linguistics Forum* 5, 73–88.
- Harris, J. (1983). *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*. Cambridge: MIT Press.
- Haupers, R. (1969). Stiang phonemes. *Mon-Khmer Studies* 3, 131–137.

- Hayes, B. (1985). *A Metrical Theory of Stress Rules*. New York: Garland Press. [Revised version of 1980 MIT Ph.D. Dissertation].
- Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Rhythm and Meter*. Orlando: Academic Press.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Henderson, E. J. A. (1952). The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14(1), 149–174.
- Hendricks, S. (2001). Bare-consonant reduplication without prosodic templates: Expressive reduplication in Semai. *Journal of East Asian Linguistics* 10, 287–306.
- Hermes, A., Grice, M., Mücke, D., & Niemann, H. (2008). Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. *8th International Seminar on Speech Production*, 433–436.
- Hsin, T. H. (2000). Consonant clusters in Tsou and their theoretical implications. In *Proceedings of the 18th West Coast Conference on Formal Linguistics*. Cascadilla Press.
- Huffman, F. (1972). The boundary between the monosyllable and the disyllable in Cambodian. *Lingua* 29, 54–66.
- Huffman, F. (1976). The register problem in fifteen Mon-Khmer languages. *Oceanic Linguistics Special Publications* 13, 575–589.
- Hyman, L. (1985). *A Theory of Phonological Weight*. Dordrecht: Foris Publications.
- Hyman, L. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Sciences* 31, 213–238.

- Ingram, J., & Nguyen, T. A. T. (2006). Stress, tone and word prosody in Vietnamese compounds. In P. Warren & C. Watson (Eds.), *Proceedings of the 11th Australasian International Conference on Speech and Science Technology* (pp. 193–198).
- Inkelas, S. (1989). *Prosodic Constituency in the Lexicon* (Ph.D. Dissertation). Stanford University.
- Isačenko, A. V. (1970). East Slavic morphophonemics and the treatment of jers in Russian: A revision of Havlik's law. *International Journal of Slavic Linguistics and Poetics* 13, 73–124.
- Ishii, T. (1996). An optimality theoretic approach to ternary stress systems. In B. Agbayani & N. Harada (Eds.), *Proceedings of the South Western Optimality Theory Workshop* (pp. 95–111).
- Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language & Linguistic Theory* 7(2), 217–260.
- Itô, J., Kitagawa, Y., & Mester, A. (1996). Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5(3), 217–294.
- Itô, J., & Mester, A. (2004). The phonological lexicon. In J. McCarthy (Ed.), *Optimality Theory in Phonology: A Reader*. Malden, MA: Blackwell.
- Jacob, J. (1968). *Introduction to Cambodian*. Oxford: Oxford University Press.
- Jannedy, S. (1994). Rate effects on German unstressed syllables. *OSU Working Papers in Linguistics* 44, 105–124.
- Kager, R. (1993). Alternatives to the iambic-trochaic law. *Natural Language & Linguistic Theory* 11(3), 381–432.
- Kager, R. (1994). *Ternary rhythm in alignment theory*. Unpublished ms. Utrecht University.

- Kahn, D. (1976). *Syllable-based Generalizations in English Phonology* (Ph.D. Dissertation). Massachusetts Institute of Technology.
- Keyser, S. J., & Kiparsky, P. (1984). Syllable structure in Finnish phonology. In M. Aronoff & R. Oehrle (Eds.), *Language Sound Structures* (pp. 7–31). Cambridge: MIT Press.
- Kiparsky, P. (1978). Recent developments in generative phonology. In *Proceedings of the Third International Conference of Nordic and General Linguistics* (pp. 42–49). University of Texas at Austin: University of Texas at Austin.
- Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry* 10, 421–441.
- Kirby, J. (To appear). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*.
- Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54(4), 1102–1104.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5), 1208–1221.
- Kreitman, R. (2008). *The Phonology and Phonetics of Onset Clusters: The Case of Modern Hebrew* (Ph.D. Dissertation). Cornell University.
- Krumhansl, C. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin* 126(1), 159–179.
- Kühnert, B., Hoole, P., & Moonshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. *Sciences de l'Homme et de la Societe*.
- Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Malden, MA: Blackwell.

- Larish, M. (1999). *The Position of Moken and Moklen within the Austronesian Language Family* (Ph.D. Dissertation). University of Hawaii, Manoa.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics* 5, 253–263.
- Lehiste, I. (1965). The function of quantity in Finnish and Estonian. *Language* 41, 447–456.
- Lehiste, I., & Peterson, G. E. (1959). Vowel amplitudes and phonemic stress in American English. *Journal of the Acoustical Society of America* 31(4), 428–435.
- Levin, J. (1987). Between epenthetic and excrescent vowels. In *Proceedings of the 6th West Coast Conference on Formal Linguistics* (pp. 187–201).
- Lewis, M. P., Simons, G., & Fennig, C. (Eds.). (2013). *Ethnologue: Languages of the World* (17th ed.). Dallas, TX: SIL International. Retrieved from <http://www.ethnologue.com/>.
- Liberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America* 32(4), 451–454.
- Liberman, P., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2), 249–336.
- Lin, Y.-H. (1998). On minor syllables. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of LP* (pp. 163–183).
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773–1781.
- Maier, J. (1969). Cua phonemes. *Mon-Khmer Studies* 3, 9–19.
- Marin, S., & Pouplier, M. (2008). Organization of complex onsets and codas in American English: Evidence for a competitive coupling model. In *Proceedings of the 8th International Seminar of Speech Production* (pp. 437–440).

- Mascaró, J. (2004). External allomorphy as emergence of the unmarked. In J. McCarthy (Ed.), *Optimality Theory in Phonology: A Reader*. Malden, MA: Blackwell.
- Matisoff, J. (1973). Tonogenesis in Southeast Asia. (L. Hyman, Ed.) *Consonant Types and Tone* 1.
- Matisoff, J. (1990). Bulging monosyllables: Areal tendencies in Southeast Asian diachrony. In K. Hall (Ed.), *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society* (pp. 543–559).
- Matisoff, J. (2003). *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. University of California Publications in Linguistics.
- Matteson, E., & Pike, K. (1958). Non-phonemic transition vocoids in Piro (Arawak). *Miscellanea Phonetica* 3, 22–30.
- McCarthy, J. (2007). Derivations and levels of representation. In Paul de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 99–117). Cambridge: Cambridge University Press.
- McCarthy, J., & Prince, A. (1993a). Generalized Alignment. In *Yearbook of Morphology* (pp. 79–153).
- McCarthy, J., & Prince, A. (1993b). *Prosodic Morphology I: Constraint Interaction and Satisfaction* (Rutgers University Center for Cognitive Science Technical Report No. 3).
- McMahon, A., Foulkes, P., & Tollfree, L. (1994). Gestural representation and lexical phonology. *Phonology* 11(2), 277–316.
- Michaud, A. (2012). Monosyllabicization: Patterns of evolution in Asian languages. In N. Nau, T. Stolz, & C. Stroh (Eds.), *Monosyllables: From Phonology to Typology* (pp. 115–130). Berlin: Akademie Verlag.

- Migliazza, B. (2003). So stories: A preliminary analysis of texts in a Mon-Khmer language. *Mon-Khmer Studies* 33, 67–158.
- Morey, S. (2005). The tone and syllable structure of Turung. In P. Sidwell (Ed.), *Papers from the 15th Annual Meeting of the Southeast Asian Linguistics Society* (Vol. XV, pp. 149–168).
- Munro, P., & Wilmond, C. (1994). *Chickasaw: An Analytical Dictionary*. Norman, OK: University of Oklahoma Press.
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.
- Nguyen, T. A. T., & Ingram, J. (2006). Reduplication and word stress in Vietnamese. In P. Warren & C. Watson (Eds.), *Proceedings of the 11th Australasian International Conference on Speech and Science Technology* (pp. 187–192).
- Niebuhr, O., Clayards, M., Meunier, C., & Lancia, L. (2011). On place assimilation in sibilant sequences -- Comparing French and English. *Journal of Phonetics* 39, 429–451.
- Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.
- O'Dell, M., & Nieminen, T. (2009). Coupled oscillator model for speech timing: Overview and examples. In M. Vainio, R. Aulanko, & O. Aaltonen (Eds.), *Nordic Prosody: Proceedings of the 10th Conference*. Helsinki.
- Okell, J. (1969). *A Reference Grammar of Colloquial Burmese*. London: Oxford University Press.
- Parker, S. (2002). *Quantifying the Sonority Hierarchy* (Ph.D. Dissertation). University of Massachusetts, Amherst.
- Phaen, T., Sok, M., & Bequette, T. (2012). *Bunong-Khmer Bilingual Dictionary*. Mondulkiri Province, Cambodia: International Cooperation Cambodia, READ Project.
- Phillips, R. (1973). Mnong vowel variations with initial stops. *Mon-Khmer Studies* 4, 119–127.

- Pike, K. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Pinnow, H. (1979). Remarks on the structure of the Khmer syllable and word. *Mon-Khmer Studies* 8, 131–137.
- Pittayaporn, P. (2005). Moken as a mainland Southeast Asian language. (A. Grant & P. Sidwell, Eds.) *Chamic and Beyond*, 189–209.
- Port, R. (2003). Meter and speech. *Journal of Phonetics* 31, 599–611.
- Poupier, M., & Benus, S. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. (J. Cole, Ed.) *Laboratory Phonology*, 2(2), 243–273.
- Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance* 7, 3–18.
- Preecha, S. (1988). *Phonological Interference between Kuay and Northeastern Thai in Surin* (M.A. Thesis). Mahidol University.
- Prince, A. (1976). *Applying stress*. Unpublished ms. University of Massachusetts, Amherst.
- Prince, A. (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry* 11, 511–562.
- Prince, A. (1984). Phonology with tiers. In M. Aronoff & R. Oehrle (Eds.), *Language Sound Structures* (pp. 234–244). Cambridge: MIT Press.
- Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar* (Rutgers University Center for Cognitive Science Technical Report No. 2).
- Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell.
- Reesink, G. (1999). *A Grammar of Hatam*. Canberra: Pacific Linguistics.
- Ridouane, R., & Fougeron, C. (2011). Schwa elements in Tashlhiyt word-initial clusters. *Laboratory Phonology* 2(2), 275–300.

- Riehl, A. (2008). *The Phonology and Phonetics of Nasal-Obstruent Sequences* (Ph.D. Dissertation). Cornell University.
- Rubach, J., & Booij, G. (1990). Syllabic structure assignment in Polish. *Phonology* 7, 121–158.
- Sagart, L. (1999). *The Roots of Old Chinese* (Vol. 184). Amsterdam: John Benjamins.
- Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review* 94, 84–106.
- Schiering, R., Bickel, B., & Hildebrandt, K. A. (2010). The prosodic word is not universal, but emergent. *Journal of Linguistics* 46, 657–709.
- Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry* 11, 563–605.
- Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.
- Selkirk, E. (1995). The prosodic structure of function words. In J. McCarthy (Ed.), *Optimality Theory in Phonology: A Reader* (pp. 464–482).
- Shaw, P. (1994). The prosodic constituency of minor syllables. In E. Duncan, D. Farkas, & P. Spaelti (Eds.), *Proceedings of the 12th West Coast Conference on Formal Linguistics* (pp. 117–132).
- Shorto, H. L. (1960). Word and syllable patterns in Palaung. *Bulletin of the School of Oriental and African Studies* 23(3), 544–557.
- Sidwell, P. (2000). *Proto South Bahnaric: A Reconstruction of a Mon-Khmer Language of Indo-China*. Canberra: Pacific Linguistics.

- Sidwell, P. (2008). Issues in the morphological reconstruction of Proto-Mon-Khmer. In H. J. Koch, C. Bowern, B. Evans, & L. Miceli (Eds.), *Morphology and Language History: In Honour of Harold Koch* (pp. 251–268). Philadelphia: John Benjamins.
- Sidwell, P., & Cooper, D. (2007). *Mon-Khmer Etymological Dictionary*. Retrieved from <http://sealang.net/monkhmer/dictionary/>.
- Silverman, D. (2011). Schwa. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 628–642). Malden, MA: Blackwell.
- Sloan, K. (1988). Bare-consonant reduplication: Implications for a prosodic theory of redup. In H. Borer (Ed.), *Proceedings of the 7th West Coast Conference on Formal Linguistics*.
- Smorodinsky, I. (2002). *Schwas with and without Active Gestural Control* (Ph.D. Dissertation). Yale University.
- Svantesson, J. O. (1983). *Kammu Phonology and Morphology*. Lund: CWK Gleerup.
- Svantesson, J. O., & Karlsson, A. M. (2004). Minor syllable tones in Kammu. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages* (pp. 177–180).
- Thach, N. M. (1999). Monosyllabization in Kiengiang Khmer. *Mon-Khmer Studies* 29, 81–95.
- Thomas, D. (1971). *Chrau Grammar*. Honolulu: University of Hawaii Press.
- Thomas, D. (1992). On sesquisyllabic structure. *Mon-Khmer Studies* 21, 206–210.
- Thurgood, G. (1996). Language contact and the directionality of internal drift. *Language* 71(1), 1–31.
- Thurgood, G. (1999). *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*. Honolulu: University of Hawaii Press.

- Tilsen, S. (2009a). Toward a dynamical interpretation of hierarchical linguistic structure. *UC Berkeley Phonology Lab Annual Report*, 462–512.
- Tilsen, S. (2009b). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33, 839–879.
- Tilsen, S., Zec, D., Bjorndahl, C., Butler, B., L'Esperance, M.-J., Fisher, A., Heimisdottir, L., Renwick, M., & Sanker, C. (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology*, 51–81.
- Tsuchida, A. (1994). Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence. *Working Papers of the Cornell Phonetics Laboratory* 9, 183–222.
- Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language & Linguistic Theory* 23(1), 169–218.
- Van de Vijver, R. (1998). *The Iambic Issue: Iambs as a Result of Constraint Interaction* (Ph.D. Dissertation). Vrije Universiteit, Netherlands.
- Vogel, S. (2006). *Introduction a la langue et aux dits traditionnels des Phnong de Mondulkiri*. Phnom Pehn: Editions Funan.
- Vogel, S., & Filippi, J. M. (2006). *Elements de langue Phnong*. Phnom Pehn: Editions Funan.
- Warner, N., Jongman, A., Cutler, A., & Mucke, D. (2001). The phonological status of Dutch epenthetic schwa. *Phonology* 18(3), 387–420.
- Watson, R. (1964). Pacoh phonemes. *Mon-Khmer Studies* 1, 135–148.
- Wayland, R. (2002). Registrogenesis in Khmer: A phonetic account. *Mon-Khmer Studies* 32, 101–115.

- Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics* 31, 181–201.
- White, L., & Turk, A. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, 38(3), 459–471.
- Yap, N. T. (2009). Nonexhaustive syllabification in Temiar. *Journal of the Southeast Asian Linguistics Society* 1, 267–282.
- Zec, D. (1988). *Sonority Constraints on Prosodic Structure* (Ph.D. Dissertation). Stanford University.
- Zec, D. (1995). Sonority constraints on syllable structure. *Phonology*, 12, 85–129.
- Zec, D. (2006). Rhythmic alternations. In *Encyclopedia of Linguistics* (pp. 619–623). Elsevier.
- Zec, D. (2007). The syllable. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 161–194). Cambridge: Cambridge University Press.