An investigation of tone in Walungge

Cathryn Virginia Bartram

Thesis submitted for the degree of PhD in Linguistics

2011

Department of Linguistics

School of Oriental and African Studies

University of London

I have read and understood regulation 17.9 of the Regulations for students of the School of Oriental and African Studies concerning plagiarism. I undertake that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person. I also undertake that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for examination.

Signed: Date:

Abstract

This dissertation is an investigation of tone in Walungge from an acoustic, a phonological, and a perceptual perspective.

The acoustic analysis includes an investigation of pitch, duration, phonation, and voicing, all of which are correlates of tone. Not only does pitch vary according to the underlying tone of the word, but also according to the syllable shape. This is investigated and accounted for. Phonation and duration are shown to vary as the pitch rises and falls. The underlying tone of the word affects the VOT of some obstruents but not others. Stress also has an effect, with pitch and duration being higher on stressed syllables.

From the phonological perspective of tone, there are two main considerations in this dissertation. The first is the interaction between voicing and tone. Because of their correlation, there is a question of whether or not voicing is phonologically contrastive. Voicing is analysed as non-contrastive, following which the phonological nature of the consonant-tone interaction is discussed. The second main consideration is the derivation of the surface rises and falls when the language is analysed as having just one underlying tone per word. The concept of a tonal foot is proposed. This, along with boundary tone insertion, accounts for all surface patterns.

Tone is also considered from a perceptual perspective. Because VOT, pitch and phonation are all cues as to the tone of a word, experiments were carried out to investigate their relative salience. Similarly, because the surface tone pattern and the phonological length of a vowel correlate, perception experiments were carried out to investigate the relative salience of the pitch melody versus the vowel duration.

This dissertation contains a body of language data, acoustic measurements, phonological analysis and perceptual considerations, all of which contribute to the understanding of tone in Tibetan languages.

Acknowledgements

There are many people to whom I owe a great deal of thanks. My thanks must first go to the wonderful people of Lungthung, Olangchung Gola, Yangma, Gunsa and Kyabla, who welcomed me into their villages, showed me hospitality and friendship as we sat round the fireplace drinking salt tea, and willingly answered my many questions about their language. In particular I would like to thank Kippa and Yangi who gave me a room in their house, fed me, and made me feel part of their family. I am also extremely grateful to Phomo, Pasang Kippa, Kippa, Yangi, both Tashis and Phuttik who were willing for their voices to be recorded, and who showed incredible patience in answering all my questions and repeating and re-repeating language data. The forty people who participated in the perception tests also deserve my thanks.

This dissertation would not have been possible without the help, support and encouragement of my supervising committee: Justin Watkins, Monik Charette, and Peter Austin. Justin has been a wonderful supervisor, giving me extremely helpful advice particularly with all the acoustic and perception test analysis, and putting up with reading and re-reading all my draft chapters. No matter how much I piled on him, he read it and returned it with very helpful comments. I am extremely grateful for Monik's input particularly on the phonology sections of this dissertation, and for Peter's wealth and breadth of experience in so many aspects of linguistics and fieldwork. I am also grateful to Nathan Hill for his insights and help with the comparison of Walungge to Written Tibetan, to Bernard Howard for his help with recording equipment, and to Sophie Salffner and Anja Choon for their friendship and input.

There are a number of friends in Nepal I also wish to thank. Pemba, Peter, Khawang, Rose and Dikki have all on different occasions given me companionship on the trails in northern Taplejung. Pemba and Peter have shown incredible willingness and patience as they helped me to gather the data and to make recordings. Lobden and Doma are the best friends that I could ever ask for.

This PhD was made possible by postgraduate studentship from the AHRC.

Tim, Matthew, Josh and Samuel all deserve my thanks. They have been four wonderful sons who have put up with their mother trying to juggle being both a student and a mother. Tim and Matthew need a lot of thanks for the way they have been willing to help with their younger brothers even when they have had their own studies and commitments to attend to. Josh and Samuel have been willing to let me go to Nepal whilst they have been looked after by their grandparents or at school.

I am grateful to my parents, Russell and Gillian Feist. Mum has willingly travelled halfway around the world to look after grandsons so as I could make trips to Nepal, and Dad has been prepared to let her go. This far exceeds the duties of grandparents. My parents-in-law, Bryan and Olwen Bartram, have also far exceeded the duties of grandparents; they also have willingly looked after grandsons so as I could get to Nepal. They have been parents to me, not parents-in-law.

Finally, there is one person to whom I owe a great deal, but who will not be reading this. Without his encouragement, I would never have applied to do a PhD. When I started my MA at SOAS as a precursor to doing the PhD, he was willing to juggle his work with being a house-husband, looking after the boys, and cooking meals, to allow me to commute into SOAS and to work on my MA in the evenings. When I applied for the PhD I never envisaged that he wouldn't be there beside me. The road has been very hard, but with the love and encouragement of friends and family I have managed to complete this dissertation. This PhD, then, is in loving memory of Steve.

Contents

Abstract 3
Acknowledgements 4
Abbreviations and terminology 18
1 Introduction 22
1.1 Introduction and outline 22
1.2 Introduction to Walungge 25
1.3 Classification 31
2 Tone in Tibetan languages 36
2.1 Definition of tone 36
2.2 The Chao system of tone letters 36
2.3 Tonogenesis 38
2.4 Register 39
2.5 Examples of register and tone in Tibetan languages 41
2.5.1 Initial consonant clusters and register 41
2.5.2 Obstruent voicing and register 43
2.5.3 Summary 46
2.6 Pitch and tone contours 48
2.7 Controversy in the analysis of tone in Tibetan languages 52
2.7.1 Obstruent voicing or register contrast 52
2.7.2 Tone Contours and/or vowel length 54
2.7.3 Underlying forms 58
2.8 Conclusion on tone and register 59
3 Phonemic summary 61
3.1 Consonants 61
3.1.1 Consonant phonemes 61
3.1.2 Plosives 61
3.1.3 Affricates 67
3.1.4 Fricatives 68
3.1.5 Nasals 70
3.1.6 Lateral and vibrant 71
3.1.7 Approximants 73
3.2 Vowels 74
3.2.1 Vowel phonemes 74
3.2.2 Length contrasts 75
3.2.3 Front and back rounded vowel contrasts 77
3.2.4 Vowel height contrasts 78

3.3 Underlying tones 78
3.4 Syllable structure 79
3.4.1 Syllable patterns 79
3.4.2 Glides 80
3.5 Diachronic phonology 88
3.5.1 Introduction to Written Tibetan 88
3.5.2 Syllable onsets word initially 91
3.5.2.1 WT Voiceless unaspirated plosive and affricate radicals 91
3.5.2.2 WT aspirated plosive and affricate radicals 92
3.5.2.3 WT Voiced plosive and affricate radicals 93
3.5.2.4 WT fricative radicals 94
3.5.2.5 WT sonorant radicals 95
3.5.2.6 WT subscripts 97
3.5.2.7 The radical < ' > 99
3.5.3 Syllable onsets word medially 100
3.5.4 Word final syllable coda 102
3.5.5 Word medial syllable coda 104
3.5.6 Vowels 107
3.5.7 Other changes 109
4 Acoustic analysis of pitch, phonation and duration 110
4.1 Introduction 110
4.2 Data 112
4.3 Pitch 114
4.3.1 Description of monosyllabic pitch melodies 114
4.3.2 Statistical analysis of monosyllabic melodies 122
4.3.3 Description of disyllabic pitch melodies 126
4.3.4 Statistical analysis of disyllabic pitch melodies 131
4.3.5 Polysyllabic pitch melodies 140
4.3.6 Monosyllabic pitch melodies in phrases 141
4.3.6.1 Acoustic analysis of the monosyllabic fall 142
4.3.6.2 Simple noun phrases 147
4.3.6.3 Further evidence for monosyllabic falling tone pattern 153
4.3.6.4 Conclusion of pitch analysis for phrases 159
4.3.7 Phonetic representation of pitch 159
4.4 Phonation 161
4.4.1 Introduction to phonation 161
4.4.2 Acoustics of phonation 162
4.4.3 Description of closed quotient patterns 168
4.4.4 Analysis of closed quotient 171

4.4.5 Analysis of spectral tilt 173
4.4.6 Further discussion 175
4.5 Duration 176
4.5.1 Duration and tone 176
4.5.2 Duration for monosyllabic words 176
4.5.3 Duration for first syllable of disyllabic words 178
4.5.4 Duration for second syllable of disyllabic words 181
4.5.5 Summary of duration 182
4.6 Conclusion for acoustic analysis of tone 183
5 Tone, stress, and tone associations 185
5.1 Tone and stress 186
5.1.1 Introduction to tone and stress 186
5.1.2 Metrical structure and tone 186
5.1.3 Refugee Standard Tibetan 188
5.1.4 Tone and stress in Tibetan languages 191
5.2 Acoustic analysis of stress in Walungge 192
5.2.1 Introduction 192
5.2.2 Duration of vowels in disyllabic non-verbs 193
5.2.3 Pitch on disyllabic high tone words 198
5.2.4 Intensity 203
5.2.5 Stress patterns on verbs 205
5.2.6 Conclusion of acoustic analysis 209
5.3 Surface and underlying tone patterns in Walungge 209
5.3.1 Introduction 209
5.3.2 Stress or tone 211
5.3.3 Underlying and non underlying tones 212
5.3.4 Underlying L and LH 213
5.3.5 Tonal Feet 215
5.3.6 Tonal foot formation in Walungge 217
5.3.6.1 Summary 217
5.3.6.2 Basic foot formation in polysyllabic words 218
5.3.6.3 Moras and syllables 219
5.3.6.4 Syllables of differing weight 223
5.4 Tone patterns across morpheme boundaries 227
5.4.1 Prosodic structure 227
5.4.2 Compound nouns 232
5.4.3 Derivational morphemes 234
5.4.4 Inflectional morphemes on nouns 236
5.4.4.1 Introduction 236

5.4.4.2 The extra-tonality of suffixes 239
5.4.4.3 Polysyllabic words with the possessive 242
5.4.4.4 Monosyllabic nouns with open syllables 245
5.5 Tone patterns on verbs 247
5.5.1 Different types of suffix 248
5.5.2 Suffixes which are extra-tonal 249
5.5.3 Suffixes included in the tonal foot 251
5.5.4 The negative morpheme 252
5.5.5 Verb forms with auxiliaries 254
5.6 Conclusion to chapter 255
6 Consonants and tone 258
6.1 Introduction 258
6.2 Word medial obstruents 260
6.2.1 Introduction to word medial obstruents 260
6.2.2 Intervocalic plosive durations 261
6.2.2.1 Plosive duration and voicing in Hindi 261
6.2.2.2 Intervocalic obstruent duration in Walungge 262
6.2.3 Proportion of voicing 266
6.2.4 Effect of medial obstruents on F0 269
6.2.5 Distribution of other word medial consonants 270
6.2.6 Morpheme boundaries 273
6.2.6.1 Plosives between voiced segments 274
6.2.6.2 Sequences of identical plosives at morpheme boundaries 278
6.2.7 Summary 279
6.3 Word initial obstruents 279
6.3.1 Introduction 279
6.3.2 VOT of word initial plosives 280
6.3.2.1 Introduction to VOT 280
6.3.2.2 Description of Walungge plosives 281
6.3.2.3 Statistical analysis of VOT of Walungge plosives 284
6.3.3 Affricates and fricatives 287
6.3.3.1 Affricates 287
6.3.3.2 Statistical analysis of affricates 290
6.3.3.3 Fricatives 292
6.3.4 F0 and intervocalic obstruents 294
6.3.5 Summary of acoustic analysis 296
6.4 Word initial sonorants 296
6.5 Laryngeal features for Walungge segments 298
6.5.1 Laryngeal features and voicing 298

6.5.2 Voiceless sonorants and aspirated obstruents 299
6.5.3 Non word initial voicing 302
6.6 Tone and consonants in Walungge 303
6.6.1 Tone and voicing proposals 303
6.6.2 Tone and voicing in Walungge 306
6.6.2.1 Different proposals 306
6.6.2.2 Spontaneous voicing? 307
6.6.2.3 Non-spontaneous voicing 310
6.6.2.4 The case of high tone 312
6.7 Conclusion of chapter 315
7 Perception 317
7.1 Introduction 317
7.1.1 Tone and voicing 317
7.1.2 Tone and length 321
7.2 Experimental design 323
7.3 Tone and voicing experiments 324
7.3.1 Pitch melodies for words starting with sonorants/fricatives 325
7.3.1.1 Experimental design 325
7.3.1.2 Results for monosyllabic heavy syllables – words in frames 327
7.3.1.3 Results for disyllabic H versus L; stimuli in isolation 330
7.3.1.4 Effect of phonation 332
7.3.2 Obstruent voicing 335
7.3.2.1 Experimental design 335
7.3.2.2 Results for stimuli in frames 336
7.3.2.3 Exposure to Nepali 339
7.3.2.4 Results for disyllabic words in isolation 343
7.3.3 Voicing and pitch for monosyllabic light syllable words 346
7.3.3.1 Experimental design 346
7.3.3.2 Results of perception of monomoraic words 347
7.3.4 Discussion on voicing and tone 354
7.4 Length versus tone 355
7.4.1 Experimental design 355
7.4.2 Results of experiment 356
7.4.3 Discussion of length and tone 362
8 Conclusion 364
Bibliography 367

List of Figures

Figure 1-1: Map of Nepal 26
Figure 1-2: Walungge speaking area 27
Figure 1-3: Benedict's classification diagram 33
Figure 1-4: A summary of Bradley's classification of Tibetan languages 34
Figure 2-1: Lhasa Tibetan contrastive contours (from Hari (1979)) 56
Figure 3-1: spectrograms comparing [wa] and [ja] 84
Figure 4-1: mean pitch melodies for monosyllabic words 115
Figure 4-2: mean and standard deviation for pitch melodies 116
Figure 4-3: -V high examples 117
Figure 4-4: -VN high examples 118
Figure 4-5: VP high examples 119
Figure 4-6: V-low examples 120
Figure 4-7: VN-low examples 121
Figure 4-8: VP-low examples 122
Figure 4-9: Boxplot of starting pitch 123
Figure 4-10: Average pitch melodies with CV as second syllable 127
Figure 4-11: Average pitch melodies for disyllabic with CV first syllable 130
Figure 4-12: mean pitch melodies for low tone disyllabic words 131
Figure 4-13: Boxplot of starting pitch of 2nd syllable for tone and syllable weight 139
Figure 4-14: mean pitch melodies for monosyllabic words in isolation 141
Figure 4-15: mean pitch melodies for monosyllabic words in frames 143
Figure 4-16: amount of fall for low tone following 144
Figure 4-17: amount of fall for high tone following 145
Figure 4-18: continuum of phonation types 162
Figure 4-19: calculation of closed quotient from glottal waveform 164
Figure 4-20: spectrum of [a]; male speaker 165
Figure 4-21: spectrum of [i]; low tone; male speaker 166
Figure 4-22: spectrum of [i]; high tone; male speaker 166
Figure 4-23: spectrum of [i]; high tone; female speaker 167
Figure 4-24: spectrum of [i]; low tone; female speaker 167
Figure 4-25: Monosyllabic mean pitch melodies 168
Figure 4-26: Monosyllabic mean CQ 168
Figure 4-27: Mean disyllabic pitch melodies for CVCV words 169
Figure 4-28: Mean CQ for CVCV words 170
Figure 4-29: boxplots of CQ at start of 1st vowel, for each speaker 171
Figure 4-30: boxplot of H2-H1 for each speaker, consonant type, and tone 174

Figure 4-31: Boxplot of duration of rhyme for monosyllabic words 177
Figure 4-32: boxplot for duration of first syllable rhyme 179
Figure 4-33: Boxplot of rhyme duration for second syllable......................................181
Figure 5-1: scatter plot showing durations of 1st and 2nd syllable rhymes 195
Figure 5-2: mean normalized pitch melodies for each high tone word pattern...........198
Figure 5-3: Scatter plot showing max pitch (semitones) of 1st and 2nd syllables.........200
Figure 5-4: Scatter plot showing pitch difference and duration difference 202
Figure 5-5: scatter plot of intensity; words with 2nd vowel height higher 203
Figure 5-6: scatter plot of intensity; words with 2nd vowel height the same 204
Figure 5-7: scatter plot of intensity; words with 2nd vowel height lower 204
Figure 5-8: Scatter plot showing durations of syllable rhymes; heavy 1st syllable 207
Figure 5-9: Boxplot showing difference in syllable rhyme duration 208
Figure 5-10: relationships between gwords and pwords in Kyirong Tibetan.............229
Figure 5-11: relationships between gwords, pwords and feet in Walungge 231
Figure 6-1: intervocalic [g] 266
Figure 6-2: intervocalic [k:] 267
Figure 6-3: Boxplot of percentage of closure voiced 268
Figure 6-4: waveform and spectrogram for *[d]*, with voicing before the release........282
Figure 6-5: waveform and spectrogram for *[t]*..282
Figure 6-6: waveform and spectrogram for *[tʰ]* before high tone 283
Figure 6-7: waveform and spectrogram for *[tʰ]* before low tone................................283
Figure 6-8: Boxplot of VOT in ms 284
Figure 6-9: spectrogram of *tɕʰá* ‘fertilizer’ 288
Figure 6-10: spectrogram of *tɕáː* ‘iron’ 288
Figure 6-11: Boxplot of VOT for affricates 290
Figure 6-12: VOT for fricatives...292
Figure 7-1: percentage choosing H tone; stimuli in frames 328
Figure 7-2: percentage choosing H tone; disyllabic stimuli in isolation 330
Figure 7-3: Bar chart of percentages for phonation; monosyllabic frames 333
Figure 7-4: Bar chart of percentages for phonation; disyllabic isolation 334
Figure 7-5: percentage of responses choosing H tone, for pitch melody and VOT 337
Figure 7-6: percentages choosing H; education inside area..341
Figure 7-7: percentages choosing H; education outside area......................................341
Figure 7-8: percentage of responses choosing H; disyllabic stimuli in isolation 344
Figure 7-9: percentage of responses choosing H for pitch levels and VOT................349
Figure 7-10: chart of responses for VOT; people educated inside the area 350
Figure 7-11: chart of responses for VOT; people educated outside the area 351
Figure 7-12: bar chart showing responses for stimuli with differing phonation 353
Figure 7-13: percentage choosing H-L% & short vowel; grouped by duration 358

Figure 7-14: percentage choosing short vowel & H-L%; grouped by pitch melody ..358

List of Tables

Table 1-1: Population of “Sherpa” speakers in Taplejung district 28
Table 1-2: Lexical similarity between Walungge and other languages 31
Table 1-3: Tibeto-Burman languages by number of speakers 32
Table 2-1: Chao letters specified in IPA 37
Table 2-2: laryngeal configurations in Southeast Asian languages 39
Table 2-3: initial sonorants, Shigatse and Temchen..42
Table 2-4: initial obstruents with pre-scripted consonant, Shigatse and Temchen 42
Table 2-5: initial obstruents without pre-scripted consonant, Shigatse and Temchen ..42
Table 2-6: Daofu Tibetan examples..43
Table 2-7: Qiuji Tibetan initial sonorants...43
Table 2-8: Qiuji Tibetan examples for initial obstruents 44
Table 2-9: Kyirong Tibetan examples for initial obstruents 45
Table 2-10: Dolpo Tibetan examples for initial obstruents 46
Table 2-11: Summary of correlations with WT consonant clusters and voicing 47
Table 2-12: Aba Tibetan pitch melodies 48
Table 2-13: Daofu Tibetan pitch melodies 49
Table 2-14: Pitch melodies for 3 Tibetan languages...49
Table 2-15: Kyirong tones 51
Table 2-16: Obstruent initial words in Lhomi...52
Table 2-17: Lhasa Tibetan monosyllabic pitch patterns 55
Table 2-18: Dege Tibetan 56
Table 2-19: Lhasa Tibetan 6-way lexical contrast for tone (Dawson (1980))...............57
Table 2-20: Lhasa Tibetan underlying tone features (Dawson (1980)) 57
Table 3-1: Consonant phonemes..61
Table 3-2: Vowel phonemes of Walungge 74
Table 3-3: WT consonants with Wylie transcription and IPA.......................................89
Table 3-4: WT voiceless unaspirated radicals with Walungge correspondences 91
Table 3-5 WT voiceless unaspirated radicals with superscripted <l> 92
Table 3-6: Aspirated word initial radicals with Walungge correspondences 93
Table 3-7: WT voiced radicals without pre-script or superscript 93
Table 3-8: Voiced radicals with pre-script or superscript..94
Table 3-9: Voiceless WT fricative radicals 95
Table 3-10: Voiced fricative radicals 95
Table 3-11: Examples of subscript <w> 97
Table 3-12: Subscripted <y> with velar plosive radical 97
Table 3-13: Subscripted <y> with bilabial plosive radical 98

Table 3-14: Subscript <l> 98
Table 3-15: Plosive radical plus subscript <r> 99
Table 3-16: Fricative radical plus subscript <r> 99
Table 3-17: the radical <'> 100
Table 3-18: word medial plosive and affricate radicals 101
Table 3-19: intervocalic (including subscript) bilabial radicals 101
Table 3-20: Examples of alveolar place of articulation assimilation 102
Table 3-21: Word final syllable codas with nasals and plosives 103
Table 3-22: Word final <l>, <s>, <r> 104
Table 3-23: Word medial plosive syllable codas 105
Table 3-24: Word medial WT nasal syllable codas 106
Table 3-25: <r>, <l>, <s> in word medial syllable coda 106
Table 3-26: changes for WT vowels and following consonants 108
Table 4-1: Minimal pairs for tone in Walungge 110
Table 4-2: Correlates of tone in Walungge 111
Table 4-3: ANOVA for effects on starting pitch 124
Table 4-4: Mean and standard deviation of starting pitch 125
Table 4-5: Scheffé post-hoc test for syllable shape 125
Table 4-6: summary of disyllabic pitch melodies in isolation 126
Table 4-7: ANOVA for effects on starting pitch for disyllabic words 133
Table 4-8: table of mean relative starting pitch for vowel height and tone 133
Table 4-9: Scheffé test for effect of vowel height on starting pitch; high tone 134
Table 4-10: Scheffé test for effect of vowel height on starting pitch; low tone 134
Table 4-11: Mean starting pitch for different word initial consonants 135
Table 4-12: Scheffé test for effect of initial consonant on starting pitch; high tone 136
Table 4-13: Scheffé test for effect of initial consonant on starting pitch; low tone 136
Table 4-14: ANOVA for effects on starting pitch of second syllable 138
Table 4-15: Table of mean and s.d. for starting pitch of second syllable 139
Table 4-16: ANOVA for effects on amount of fall 144
Table 4-17: Mean and s.d. of fall for low tone following 145
Table 4-18: Mean and s.d. of fall for high tone following 145
Table 4-19: Scheffé post-hoc test for amount of fall for low tone following 146
Table 4-20: Scheffé post-hoc test for amount of fall for high tone following 146
Table 4-21: monosyllabic pitch melodies described using Chao system 160
Table 4-22: monosyllabic pitch melodies described using Chao system 161
Table 4-23: ANOVA for effects on closed quotient 172
Table 4-24: ANOVA for effects on spectral tilt of [a] 175
Table 4-25: Mean and s.d. for duration of syllable rhymes; monosyllabic words 177
Table 4-26: ANOVA for effects on duration of monosyllabic rhyme 178

Table 4-27: Mean and standard deviation of duration of first syllable rhyme 179
Table 4-28: ANOVA for effects upon duration of first syllable rhyme 179
Table 4-29: monosyllabic pitch melodies 180
Table 4-30: Table of mean and s.d. for rhyme duration of second syllable 181
Table 4-31: ANOVA for effect of tone upon duration of second syllable rhyme.......182
Table 5-1: Mean and standard deviation of rhyme durations 195
Table 5-2: paired T-test results comparing durations of 1st and 2nd syllable 196
Table 5-3: mean duration (ms) of syllables; differing syllable shapes; high tone.......197
Table 5-4: mean duration (ms) of syllables; differing syllable shapes; low tone........197
Table 5-5: paired T-test comparing 1st and 2nd syllable pitch; high tone words 201
Table 5-6: paired T-test results comparing intensity of 1st and 2nd syllable 205
Table 5-7:Walungge pitch melodies 210
Table 6-1: Durations of Hindi plosives 262
Table 6-2: mean and standard deviation for duration of medial plosives....................263
Table 6-3: ANOVA for effects on duration of intervocalic plosives 265
Table 6-4: mean and standard deviation for proportion voiced 268
Table 6-5: ANOVA for effects on starting F0 of second syllable 270
Table 6-6: mean and s.d. for 2nd syllable F0; intervocalic plosives 270
Table 6-7: Lhomi and Walungge medial consonants 271
Table 6-8: Types of plosives in Tibetan languages 281
Table 6-9: Word initial plosives of Walungge 281
Table 6-10: mean and standard deviation of VOT for different plosive types............284
Table 6-11: ANOVA for effects on VOT...286
Table 6-12: Scheffé post-hoc test for VOT; aspirated plosives 286
Table 6-13: Scheffé post-hoc test for VOT; unaspirated plosives 286
Table 6-14: Examples of aspirated plosives and affricates 289
Table 6-15: Examples of unaspirated plosives and affricates 290
Table 6-16: mean and standard deviation of VOT for different affricate types 291
Table 6-17: ANOVA for effects on affricate VOT 291
Table 6-18: mean and standard deviation of VOT for fricatives 293
Table 6-19: ANOVA for effects on fricative VOT 293
Table 6-20: voicing in plosives and fricatives..294
Table 6-21: ANOVA for effect of tone upon VOT for intervocalic plosives 295
Table 6-22: mean and standard deviation for VOT for intervocalic plosives 295
Table 6-23: laryngeal features proposed in Halle & Stevens (1971)...........................304
Table 7-1: manipulated pitch melodies..326
Table 7-2: percentages choosing H or L; monosyllabic heavy syllable stimuli 327
Table 7-3: significance statistics for each pair of pitch melodies 329
Table 7-4: percentages of responses choosing H or L; stimuli in isolation................330

Table 7-5: significance statistics for each pair of pitch melodies................................332
Table 7-6: significance statistics for pairs of stimuli with differing phonation...........333
Table 7-7: percentages choosing H or L; stimuli in frames..336
Table 7-8: significance statistics for pairs of positive and negative VOT words........338
Table 7-9: percentages of responses choosing H for differing education....................340
Table 7-10: significance statistics for VOT; education inside language area..............342
Table 7-11: significance statistics for VOT; education outside language area............342
Table 7-12: percentages choosing H or LH; disyllabic stimuli in isolation. 343
Table 7-13: significance statistics for VOT...344
Table 7-14: percentages of responses for pitch levels – sonorant onset......................348
Table 7-15: percentages of responses for pitch levels – plosive onset........................348
Table 7-16: percentages of responses for light syllabled words with level pitch........350
Table 7-17: significance statistics for VOT; education inside language area..............352
Table 7-18: significance statistics for VOT; education outside language area............352
Table 7-19: significance statistics for pairs of stimuli with differing phonation.........353
Table 7-20: manipulated pitch melodies..356
Table 7-21: percentages of responses for duration and pitch melody 357
Table 7-22: McNemar test p values for 4-1 versus 3-3 360
Table 7-23: McNemar test p values for 3-3 versus 4-4 360
Table 7-24: McNemar test p values for different pitch melodies and durations.........360
Table 7-25: McNemar test p values for phonation...362

Abbreviations and terminology

List of abbreviations used

1PL	1st person plural
1SG	1st person singular
2PL	2nd person plural
2SG	2nd person singular
3PL	3rd person plural
3SG	3rd person singular
ABL	ablative
C	consonant
CONJT	conjunct
CQ	closed quotient
DAT	dative
DISJT	disjunct
ERG	ergative
F	feature
F0	fundamental frequency
G	glide
gword	grammatical word
GEN	genitive
H	high tone
H1	1st harmonic
H2	2nd harmonic
IF0	intrinsic F0
IMP	imperative
L	low tone
LAR	laryngeal node
LOC	locative

M	mid tone
NEG	negative
NMLZ	nominalizer
OCP	obligatory contour principle
OQ	open quotient
PL	plural
POSS	possessive
pphrase	phonological phrase
pword	phonological word
PST	past
T	tone
TBU	tone bearing unit
V	vowel
Vː	long vowel
VN	either a long vowel, or vowel followed by sonorant
VOT	voice onset time
VP	vowel followed by plosive
WT	Written Tibetan
%	boundary tone
σ	syllable
μ	mora
+	morpheme boundary
#	word boundary
*	ungrammatical form

Terminology to do with tone

The terms pitch and tone are not used consistently throughout the literature. Therefore it is important to define how these are being used in this dissertation.

Tone: the word “tone” is being used to refer to a phonological unit of contrast whose primary phonetic realisation is pitch.

Underlying tone: this dissertation assumes an autosegmental framework where tones are autosegments; that is to say, they are unattached to the segments underlyingly. Walungge has two underlying tones H (high) and L (low) which are being transcribed as accents on the first vowel of each word, e.g. é for high tone and è for low tone.

Surface tone: in an autosegmental framework the tones attach to the segments according to language specific rules. The pattern of H’s and L’s across the segments is being referred to as surface tone.

F0: The surface tone pattern has to be realised by the articulators and in particular by the frequency at which the vocal folds vibrate. This frequency is F0 and is an absolute acoustic measurement.

Pitch: The height and contour of F0 may be perceived as high, mid, low, rising, falling, etc. Perception is relative; for example the same value of F0 might be perceived as high if the speaker is male or low if the speaker is female. The perceived height of F0 is pitch. In this dissertation pitch is being transcribed using numbers from 1 (lowest pitch) to 5 (highest pitch).

Register: in this dissertation “register” refers to a phonological contrast involving the configuration of the larynx. The acoustic correlates of register include pitch, voicing, and phonation.

Levels of representation

The following should be noted about the levels of representation used in this dissertation for the language data. Following convention, the phonological form of the segmental data is represented using //, and data which is phonetic is represented using []. However, in some instances it is helpful to be able to refer to an intermediate representation. In particular there are instances where it is more helpful overall to have a phonemic rather than phonetic transcription, but it is very necessary to transcribe the phonetic voicing of obstruents. For this intermediate level of representation italics are being used. The following are examples showing the different levels of representation:

/ʈùpse/	*ɖùpse*	[duɸse $^{11\ 44}$]	‘grain’
/ʈùk/	*ɖùk*	[ɖuʔ ^{24}]	‘dragon’

1 Introduction

1.1 Introduction and outline

Tone in Tibetan languages is a fascinating topic of study. Part of the interest is the great tonal variety that exists within Tibetan languages, both in terms of the phonological contrasts, and in the acoustic realisation of these contrasts. Not only that, but the analysis is often not clear-cut, and alternative analyses can be proposed for the same set of data. This dissertation investigates tone in the Walungge language, which is an endangered Tibetan language spoken in the far north-east of Nepal. To date there has been no published description or documentation of Walungge. Because of the lack of previous description, this dissertation focuses on investigating and describing the language data. It is hoped that the body of language data contained in this dissertation, the acoustic measurements that have been carried out, the phonological analysis which has been proposed, and the perception experiments, will all contribute to a greater understanding of tone in Tibetan languages in general, as well as tone in Walungge.

The rest of this introductory chapter gives an overview of the Walungge language, including the language area, number of speakers and classification. It also includes information about the Walung people and their culture and way of life.

Chapter 2 is also introductory in nature, being an overview of tone in Tibetan languages. For a number of Tibetan languages, the contrast is not only realised through a change in pitch, but also phonation, voicing, duration, etc. This “package” of correlates is often referred to as “register”. Chapter 2 considers how a register or tone contrast has come about in Tibetan languages. It gives an overview of the great variety of tonal/register systems that exist in Tibetan languages. It also gives an overview of some of the controversy that exists in the analysis of tonal Tibetan languages.

Because this is the first description to be done of the Walungge language, Chapter 3 is a phonemic analysis, giving a chart of the consonant and vowel phonemes, data to illustrate the contrasts and the allophonic variation, and a description of the syllable structure. Following this, it gives an overview of the diachronic phonology, comparing Walungge with Written Tibetan.

Chapter 4 contains the main body of the acoustic analysis. Tone in Walungge has the phonetic correlates of pitch, duration, phonation, and voicing. With the exception of voicing (which is considered in chapter 6), these correlates are investigated acoustically, with statistical analysis carried out to determine what effects are significant. The investigation shows that of these correlates pitch is the primary factor, with phonation and duration varying as the pitch rises and falls. The investigation also shows that pitch varies not only according to the underlying tone of the word, but also according to the syllable shape. This variation cannot be explained as a phonetic effect; it is a realisation of different tone patterns.

Using an autosegmental framework, Chapter 5 considers the question of what are the underlying tones and how can the surface tone patterns be accounted for. Interaction between stress and tone is investigated, and the concept of a tonal foot is proposed. With this proposal Walungge can be analysed as having one underlying tone per word (either H or L), and all surface tone patterns can be accounted for. Using this proposal, the chapter gives example derivations of surface patterns for nouns and verbs, both monomorphemic words and words incorporating morpheme boundaries. It also shows how the segments and the tone may not necessarily be operating at the same level in the prosodic hierarchy.

Chapter 6 discusses the relationship between tone and consonants. Occurring in conjunction with tone is a voicing distinction for plosives, thus an acoustic investigation of all obstruents is carried out. The question of whether voicing is phonologically contrastive, or is caused by the tone, is considered. To consider this question, the voicing of word medial plosives is investigated because their voicing correlates with factors other than tone. Having analysed voicing as non-contrastive, the phonological nature of the consonant-tone interaction is discussed.

Chapter 7 considers tone from a perceptual perspective. Because the VOT, pitch and phonation are all cues as to the tone of a word, experiments were carried out to investigate their relative salience. The results of the experiments show that pitch is the most salient cue. It is only when the pitch of a word is ambiguous that VOT becomes significant in determining the tone of a word. Similarly, because the surface tone pattern and the phonological length of a vowel correlate, perception experiments were carried out to investigate the relative salience of the pitch melody versus the vowel duration. In this situation it is the vowel duration which is of greater salience than the pitch melody.

Chapter 8 is a summing up of the preceding chapters and the conclusion of this dissertation.

1.2 Introduction to Walungge

Walungge Language

The name *Walungge* comes from *Walung*, one of the names of a principal village in the language area, and *skad* ([keʔ]), the Tibetan for 'language'. Other names for the language that may be found in the literature include *Halung*, which is the local name for the same village, and *Olangchung Gola*, which is the Nepali name for that village. *Walungge* is the name under which the language is listed in the Ethnologue (Gordon (2005)); the Ethnologue code is [ola].

The name *Walung* is used by the Walung Diaspora population in Kathmandu and further afield to refer to their people group and language. It is also used by the Nepal Federation of Indigenous Nationalities to refer to the people group and language. However, it is generally not used by the language community living in the actual language area. These people do not have a specific name for their language. When asked, they generally refer to their language as *Sherpa*, whilst recognising that it is different from the Sherpa of the Solu Khumbu district of Nepal. Other people groups living in Taplejung district will refer to the language as *Bhotia*, which is a word used across Nepal to refer to the Tibetan people groups of Nepal, and comes from the word *bod* which is Tibetan for 'Tibet'. The people from Olangchung Gola village will accept the words *Walungge* or *Halungge* as possible names for their language; however the people from other villages in the language area will generally say that *Walungge* or *Halungge* specifically refer to the variety of language spoken in Olangchung Gola village, but not to the language as a whole.

This makes it difficult to decide what name to use when referring to the language. It would be nice to refer to it using a local name, but calling it *Sherpa* would create confusion, as would calling it *Bhotia.* However, given that the Diaspora population has chosen the word *Walung* to refer to themselves, and given that the word *Walungge* is already in use in the literature to refer to the language, I have chosen to use *Walungge* to refer to the language, *Walung* to refer to the people group and *Olangchung Gola* to refer to the village.

Language Area

The Walungge speaking area is primarily the upper Tamur River and the Gunsa River in the north east of the Taplejung district of Nepal, and includes the villages of Olangchung Gola, Yangma, Gunsa and Lungthung. Figure 1-1 and Figure 1-2 below are maps of the area.

Figure 1-1: Map of Nepal

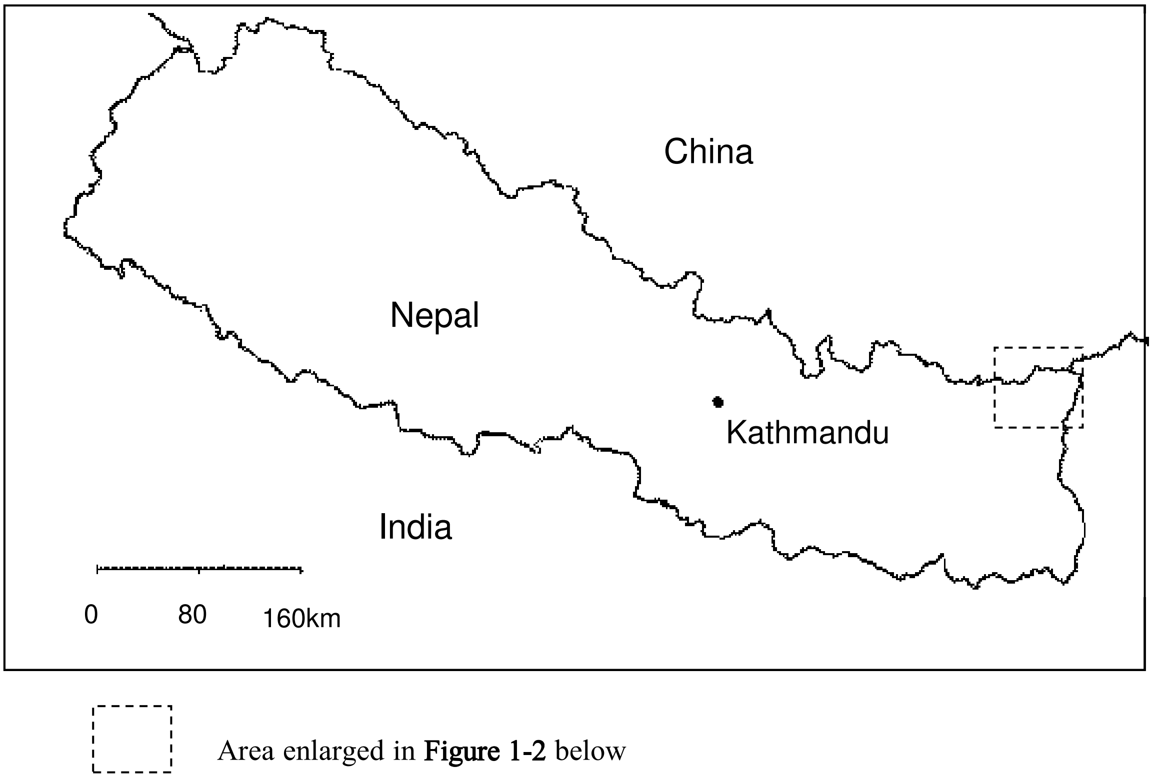
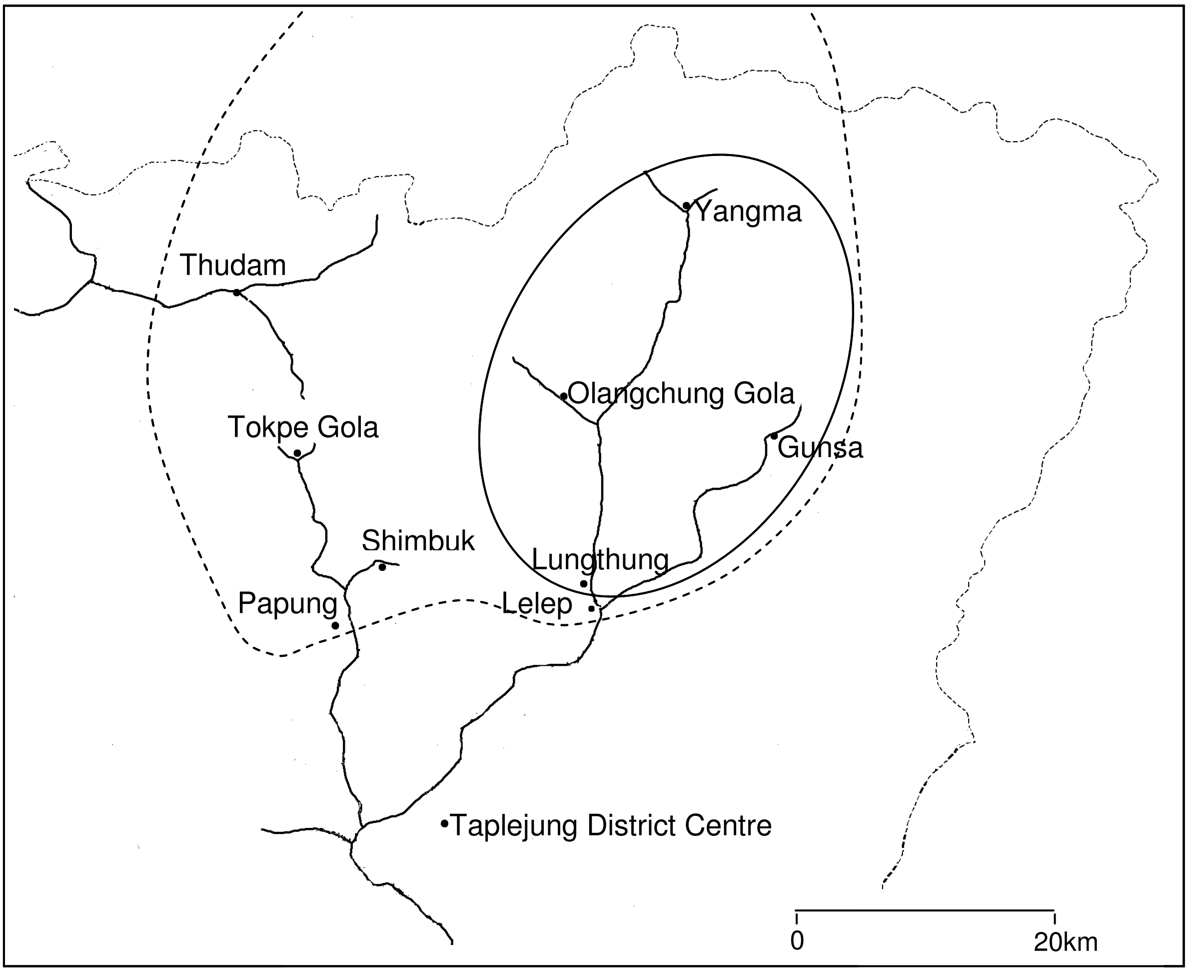


Figure 1-2: Walungge speaking area



River

Country border

Walungge area

closely related language area

As well as showing the area which is Walungge-speaking, Figure 1-2 also shows areas where the language variety appears to be very close to Walungge. These areas are discussed briefly below.

Caplow (2007) gives the language area of Tokpe Gola Tibetan as primarily the villages along the upper Mewa River, which include Tokpe Gola village, Shimbuk and Papung, plus the village of Thudam. From my personal communication with people from Thudam, Tokpe Gola, and Olangchung Gola, it would appear that the people from these areas can communicate with each other with ease using their mother tongue. The question of whether these language varieties are dialects of the same language or whether they are different languages is not addressed here.

The village of Lelep is only about an hour's walk away from Lungthung. However, from my personal communication with people from Lelep, Lungthung and Olangchung Gola it would appear that the variety of language spoken in Lelep, whilst being a Tibetan language, is different from that spoken in the villages north of Lelep.

North of Olangchung Gola village is a pass through the mountains to China. Trans-border trade is carried out on a regular basis. From my personal communication with people from across the border whilst they have been carrying out business in Olangchung Gola, it would appear that the Tibetan villages immediately across the border speak a very similar variety of language to that spoken in Olangchung Gola.

Number of speakers

There are approximately one thousand speakers of Walungge in the area given in Figure 1-2. However, the total number of speakers of the language is very hard to determine. Part of the difficulty in determining the number of speakers comes from the fact that the people of Taplejung whose mother tongue is a Tibetan language all tend to refer to their language as "Sherpa". The 2001 Nepal Census lists no speakers of Walungge, however in the whole of Taplejung district it lists approximately 12,600 mother tongue speakers of "Sherpa". These are roughly distributed as follows:

Table 1-1: Population of "Sherpa" speakers in Taplejung district

Area	*approx. number of "Sherpa" speakers*
Upper Tamur/Gunsa River	1100
Upper Mewa River	2100
Other areas	9400

It can be assumed that in the Upper Tamur River and Upper Mewa River areas the mother tongue "Sherpa" speakers are for the most part Walungge speakers and Tokpe Gola Tibetan speakers rather than Solu Khumbu Sherpa speakers. The villages in the "other" category are villages where the majority population is mother tongue

Limbu (a Kiranti language) or Nepali, and "Sherpa" is in the minority. In these villages some speakers of "Sherpa" are people from villages such as Olangchung Gola or Gunsa who have chosen to settle near the district centre or in a village where the climate is not as harsh as their own. In such cases "Sherpa" will refer to Walungge. However in some villages there are Tibetan peoples from other districts of Nepal, and also from China, who are now living in Taplejung district, possibly including Sherpas from Solu Khumbu.

Another difficulty with knowing the number of Walungge speakers is that ethnicity and mother tongue are so closely linked that people of "Sherpa" ethnicity will state their mother tongue as "Sherpa" even if it is only their parents or grandparents who speak it. The 2001 census give a population of approximately 13,100 for "Sherpa" ethnicity in Taplejung. This means that the proportion of those of "Sherpa" ethnicity stating "Sherpa" as their mother tongue is 96%. Yet from my own observations it would appear that in many villages, especially those where "Sherpa" is the minority, it is only the old people who are able to speak the language, and day to day communication is conducted in Nepali. Thus, the number of speakers is likely to be considerably less than what is reported in the Nepal Census.

As well as Walungge speakers living in the original area, there is also a Diaspora population. Gupha Pokhari village, south-west of Taplejung district centre and north of Basantapur, has a population of approximately 300 people, the majority of whom are from Thudam, Olangchung Gola and Papung. These people appear to have retained their mother tongue and not lost it in favour of Nepali. Much of the Walung population in Kathmandu, on the other hand, appears to have lost their mother tongue in favour of the variety of Central Tibetan spoken in Kathmandu by the wider Diaspora Tibetan population.

Surrounding languages, culture, etc.

To the south of the Walungge area, the primary languages spoken are Limbu a Kiranti language, and Nepali. To the north and east of the Walungge speaking area are the Himalayas with the Kanchenjunga mountain range. There are passes north to China and east to India. West of the area is the Arun valley where Lhomi, another Tibetan language, is spoken.

Olangchung Gola used to be a prosperous trading village on the main trading route through Eastern Nepal from India to Tibet. However once Tibet became part of China, trade was restricted. Subsequent to that, a landslide destroyed many of the houses in Olangchung Gola. This combination of events meant that by 1972 the trade had dried up and many people, particularly the wealthier ones, had left Olangchung Gola for places such as Darjeeling or Kathmandu (Fürer-Haimendorf (1975)). Today, whilst cross-border trade in and out of Tibet has resumed, many people in Olangchung Gola depend on carpet-making (by hand) for their livelihood. The villages of Yangma, Gunsa and Lungthung are agricultural villages. Gunsa is on the trekking route to Kanchenjunga base camp, and so derives much of its income from tourism.

Culturally Walungs are Tibetans and follow Tibetan-Buddhism. Thus classical literary Tibetan is used by the *lamas* in the monasteries; however from personal observation it would appear that it is only the *lamas* who are literate in Tibetan. There are Nepali medium primary schools in each of the villages, and the vast majority of children attend school. A handful of children are sent to school in Darjeeling. Adult literacy rates in Nepali are approximately 50%, according to the 2001 Census. Walungge is an unwritten language.

Previous research on Walungge

Very little research appears to have been done on the Walungge language.

The Ethnologue gives lexical-similarity statistics between Walungge and other Tibetan languages (Gordon (2005)). These are as follows:

Table 1-2: Lexical similarity between Walungge and other languages

Language	*percentage lexical similarity*
Lhasa Tibetan	71
Dolpo	68
Lowa	68
Kyirong	68
Lhomi	66
Yolmo	66
Nubri	64
Jirel	57
Sherpa	55

There is no indication of the source of these figures.

Van Driem (2001) mentions the existence of Walungge as a language and Bradley (1997) lists it in his classification of Tibetan languages. In Matisoff (1996) Walungge is listed along with the following information: “listed by Ethnologue as Bodic, and could be a Lhomi or Tibetan dialect”.

Fürer-Haimendorf (1975) outlines Walung culture as it was in 1957. He mentions that his informants in Thudam told him that the identical language is spoken in Thudam, Tokpe Gola, Olangchung Gola, and the Tibetan villages of Kudo and Sar.

1.3 Classification

From Tibeto-Burman to Tibetan

Tibeto-Burman is a family of languages stretching across the Himalayas and surrounding regions of India, Nepal, China, Bhutan, Bangladesh, Burma, Thailand, Laos and Vietnam. The exact number of Tibeto-Burman languages is unknown but is estimated at approximately 250 languages (Matisoff (2003)). Of these, whilst some have well over a million speakers (for example Burmese), approximately half have less than 10,000 speakers. The following table is taken from Matisoff (2003):

Table 1-3: Tibeto-Burman languages by number of speakers

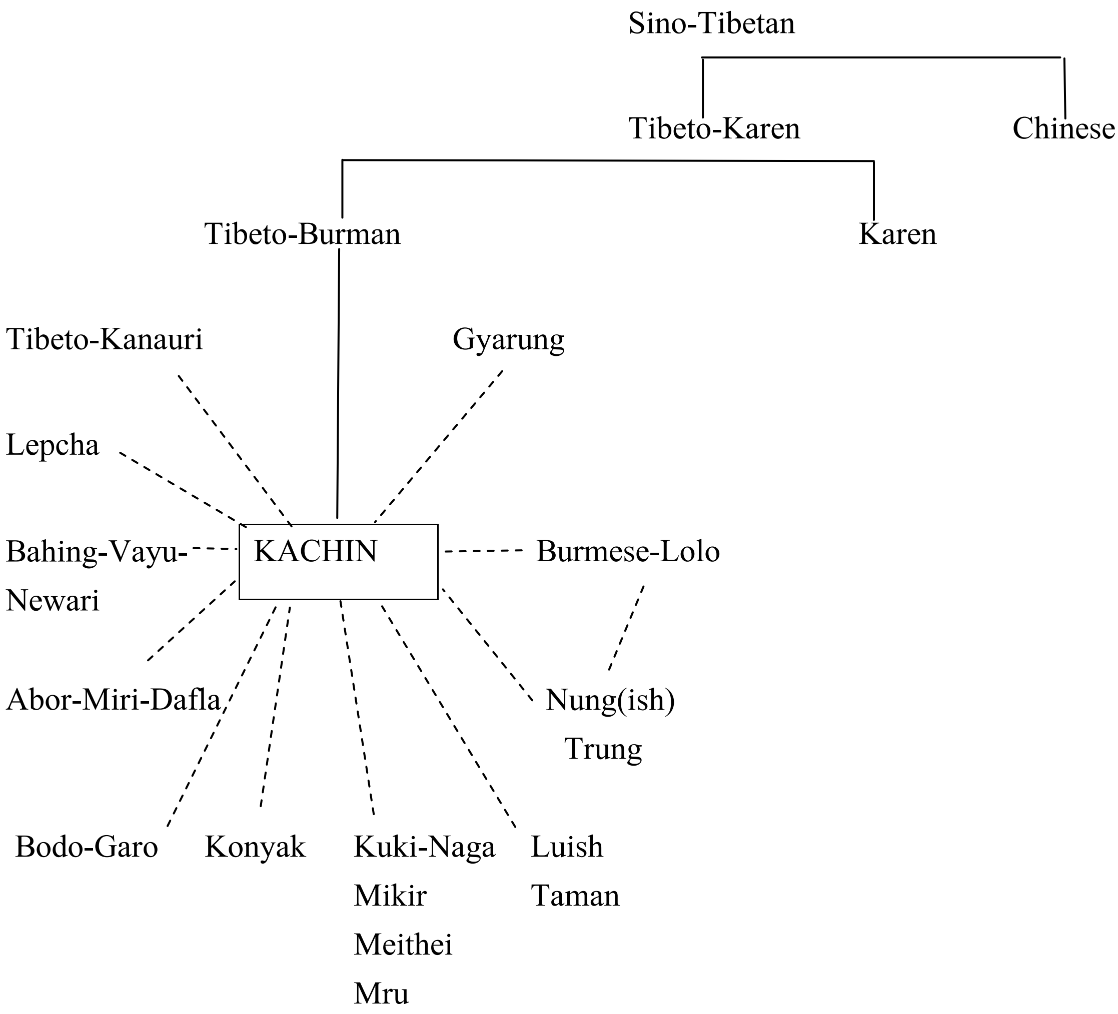
Number of Speakers	*Number of Languages*
more than 1,000,000	9
500,000 - 999,000	12
250,000 - 499,000	11
100,000 - 249,000	16
50,000 - 99,000	16
25,000 - 49,000	27
10,000 - 24,000	44
fewer than 10,000	123

There have been various classifications of Tibeto-Burman languages proposed, including Shafer (1966; 1974), Benedict (1972) and Bradley (1997). Although these differ, all include a group of languages known as Tibetan, or Bodish (from *Bod*, the Tibetan word for 'Tibet').

Shafer does not have a Tibeto-Burman division as such. Instead, he breaks Sino-Tibetan into six main divisions: Sinitic, Daic, Bodic, Burmic, Baric and Karenic. The latter four of these together correspond to what the other classifications term Tibeto-Burman. His Bodic division includes Himalayish sections and a Bodish section.

Benedict, using primarily the same body of lexical material as Shafer, presented a diagram of Tibeto-Burman with Kachin as the centre:

Figure 1-3: Benedict's classification diagram



Benedict's Tibeto-Kanauri division, which is one of the Kachin divisions, has two main branches: Bodish and Himalayish.

Bradley divides Tibeto-Burman into North-eastern India, Western or Bodic, South-eastern and North-eastern. His Western or Bodic branch is divided into Bodish and Himalayan.

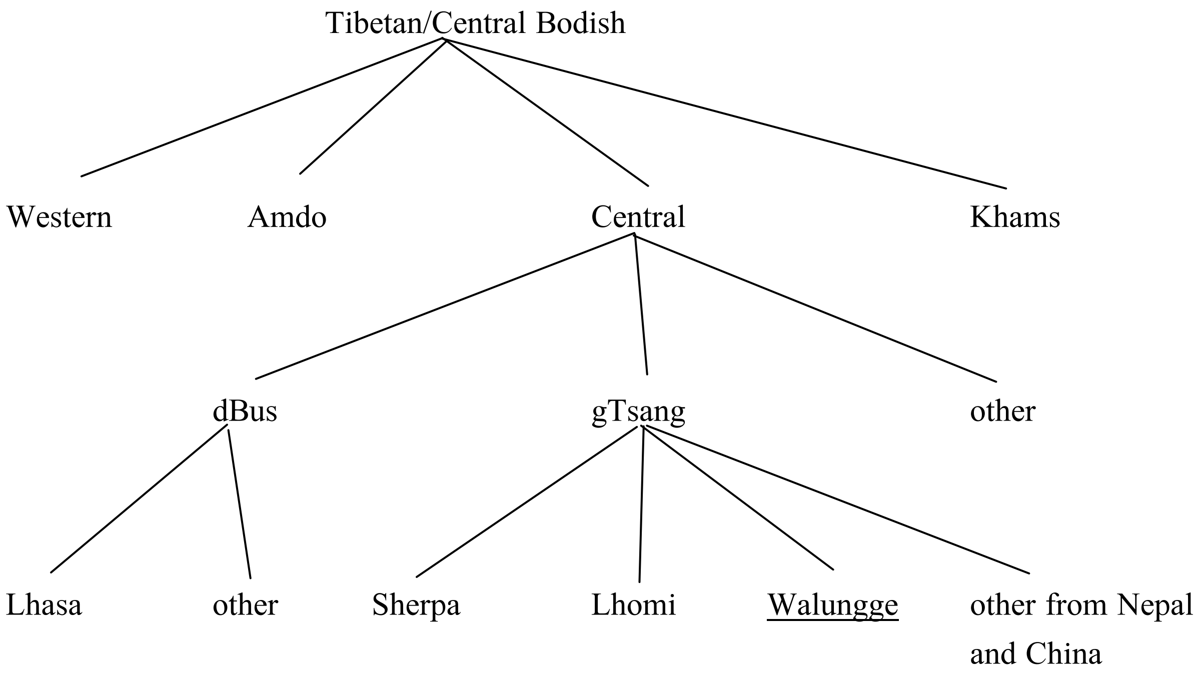
For all of the above classifications, within Bodish is a group of languages which includes the majority of languages of Tibet and the bordering areas of India, Nepal and Bhutan. For Shafer this group is Central Bodish, for Benedict Tibetan, and for Bradley Central Bodish/Tibetan. Bradley estimates the total number of speakers of Tibetan languages as about 5 million.

Walungge Classification

Very little has been done to classify individual Tibetan languages, and this is particularly true of the Tibetan languages of Nepal, many of which have little or no documentation. Walungge appears in very few lists of Tibetan or Tibeto-Burman languages. Hale (1982) gives a comparison of Tibeto-Burman classifications to that date along with a list of Tibeto-Burman languages, and Walungge does not appear. It is, however, listed in Matisoff (1996) which is a list of all known Tibeto-Burman languages and dialects (listed as Olangchung Gola).

Bradley (1997) classifies Walungge (which he lists as Halung) as *gTsang*, which is a branch within his Central Bodish group. Figure 1-4 below illustrates this:

Figure 1-4: A summary of Bradley's classification of Tibetan languages



DBus and *gTsang* are the names of two regions in Tibet, with *dBus* referring to the central region and *gTsang* referring to the region to the south west of that. Thus Bradley's *gTsang* and *dBus* groups of languages are classified primarily by geographical region. *gTsang* includes most of the Tibetan languages of Nepal, plus the languages of the towns Shigatse and Gyantse in Tibet. Bradley estimates that the *gTsang* languages have over half a million speakers in total, with many of the *gTsang* languages in Nepal having less that 5,000 speakers per language, and those in Tibet totalling over 400,000 speakers. Lhasa Tibetan is in Bradley's *dBus* group of languages.

Classification systems

One of the difficulties with classifying Tibeto-Burman languages is that there is not yet any reliable system for determining the genetic relationships within the sub-branches (DeLancey (1990)). LaPolla (2000) challenges the validity of Tibeto-Burman subgrouping which is based on geographical location. Geographical subgrouping assumes that all the languages in a particular area originated from a single ancestral language which was either in that area or migrated to that area. Yet for Tibeto-Burman languages there are known to have been waves of migration (LaPolla (2000)).

There is general agreement within Tibeto-Burman linguistics (Thurgood (1985), Sun (2003)) that classification needs to be based on shared innovations. One shared innovation in many Tibeto-Burman languages that has attracted debate is *tonogenesis*, or, the birth or tone. This has attracted debate particularly concerning the classification of Central Bodish (Tibetan) languages. Sprigg (1972), Qu & Jin (1981) and Hu (1991) are among those presenting arguments in support of using tonal systems as a contributing factor in the classification of Tibetan languages. On the other hand, Matisoff (1973) and Sun (2003) caution against using tone in the classification of Tibetan languages. The debate is centred on whether tonogenesis is a shared innovation which has occurred in a common ancestor as it branched off from the atonal languages, or whether tone has emerged independently in different Tibetan languages and thus cannot be used to establish genetic subgrouping relationships.

2 Tone in Tibetan languages

2.1 Definition of tone

A definition of a tone language is found in Hyman (2001, p.1368): "A language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes." This definition includes languages where some of the morphemes have tone and some are toneless, and languages where the underlying form of roots may be toneless but the addition of tone gives a lexical contrast, e.g. Somali (Hyman (2001)). It includes those Tibetan languages discussed below where pitch is one of the phonetic correlates of register (see section 2.4 for a discussion of register). However, the definition does not include languages such as English, where pitch is added post-lexically for purposes such as stress and intonation. Very roughly, about 60-70 percent of the world's languages may be tonal (Yip (2002)).

At this point it is also worth distinguishing between the following terms: tone, pitch, register and fundamental frequency. Tone refers to a phonological contrast. Pitch refers to the perceptual height at which a sound is heard. Register (section 2.4 below) refers to a phonological contrast involving laryngeal configuration. The acoustic term referring to the number of cycles per second at which the vocal folds are vibrating, is the fundamental frequency.

2.2 The Chao system of tone letters

When dealing with a tone language, it is necessary to have a system for transcribing the tone and/or the pitch. The Chao system of tone letters (Chao (1930)) is a system that is commonly used when describing the tones/pitch of Asian languages. The Chao system divides a speaker's natural pitch range into 5 levels, with 1 being the lowest pitch and 5 the highest. Each syllable is then assigned pitch numbers, typically one digit for the starting pitch and one digit for the ending pitch. Thus *11* describes a low level pitch, *53* a high to mid fall, etc. If the pitch changes through the syllable, then 3 digits can be used. For example, *121* describes a low rising-falling pitch, *551* describes a pitch that sustains high then falls to low. Instead of using digits, Chao letters may also be represented diagrammatically by a line showing the contour of the pitch attached to a vertical stick, e.g. ˥˧ for *53,* ˩˧˩ for *131*, ˦ for *44*, etc. The Chao system in its diagrammatic form is part of the IPA (International Phonetic Alphabet), and is specified in the IPA as follows:

Table 2-1: Chao letters specified in IPA

˥	extra high	˩˥	rising
˦	high	˥˩	falling
˧	mid	˦˥	high rising
˨	low	˩˨	low rising
˩	extra low	˧˦˧	rising-falling

It is important to realise that given the variable nature of pitch, there is a need to interpret Chao letters with an element of flexibility. For example, a language might make use of two pitch levels: high and low. These might be labelled *55* (as the highest pitch in the speaker's range) and *11* (as the lowest pitch in the speaker's range). However, given that these levels are high and low, not extra high and extra low, they could equally be labelled *44* and *22,* which is how IPA specifies high and low.

Chao letters, particularly in their diagrammatic form, can also have a degree of flexibility simply for visual clarity. For example the Lhasa Tibetan high falling pattern *54* is often written *53* for convenience and visual clarity (Tan (1987), quoted in Duanmu (1992)). Diagrammatically, *53* ˥˧ is clearer than *54* ˥˦. What is significant is not whether the pitch ends on level 3 or level 4, but that it has a falling pattern and the fall is not a great fall.

A further point to note is that Chao letters can be used both for the transcription of the phonetic pitch, and for the description of a phonological tone.

2.3 Tonogenesis

Tonogenesis is the development of contrastive tones in a language.

In the articulation of some sounds in the world's languages, a difference in pitch also occurs. For example, in many languages there is a correlation between voiced obstruents and relatively lower pitch, at least at the onset of the following vowel. Similarly there may be a correlation between voiceless obstruents and relatively higher pitch. The correlation between pitch and obstruent voicing is so wide-spread, even in non-tonal languages, that it has been termed a language universal (Matisoff (1973)). Hombert (1978) details experimental evidence showing this correlation in English. Hombert (1978) also discusses the tendency for aspirated stops to have higher pitch at the onset of the following vowel, and for post-vocalic glottal stops to raise the pitch of the vowel. The height of the vowel itself can affect the pitch, with higher vowels being associated with higher pitch (Connell (2002)).

For some languages, the proto-language appears to have been atonal, with pitch differences associated with a contrastive segmental feature, for example obstruent voicing. Over time, the pitch difference has become more significant and the segmental acoustic cues less significant, to the point where pitch has become reanalysed as contrastive tone, and the segmental distinction has been lost altogether. The reanalysis of pitch as tone is discussed by Maran (1973), who uses the term "cognitivization" to mean the point in time at which the cognitive contrast is transferred from the segmental cues to the prosodic cues, and thus the perception and analysis of linguistic data is now carried out largely in terms of the prosodic features of tone.

One very common way in which tonogenesis has occurred is the loss of a contrast in obstruent voicing in favour of a contrast in tone (Matisoff (1973), Hombert (1978), Yip (2002)). An example of a language where obstruent voicing distinctions have been lost altogether in favour of tone is Vietnamese (Haudricourt (1954)). However, some languages which have become tonal still have a correlation between obstruent voicing and tone. Examples of this include Kera a Chadic language (Pearce (2005)) and Shanghai Chinese (Yip (2002)).

2.4 Register

Many Tibetan languages have a word-level contrast involving pitch and related correlates. For some Tibetan languages, because the contrast is mainly or solely carried by the pitch, the contrast is simply referred to as tone. But for other Tibetan languages the term "register" has been used to describe the contrast, because of presence of other correlates besides pitch.

The difference in register can be equated with a difference in configuration of the larynx. Whilst register is generally a binary phonological contrast, when considering the articulatory possibilities, in Southeast Asian languages three possible laryngeal configurations may be recognised: tense, unmarked, and breathy. The exact phonetic qualities vary from language to language but may include the following (Thurgood (2007)):

Table 2-2: laryngeal configurations in Southeast Asian languages

	Tense	*Unmarked*	*Breathy*
original initials:	proto-voiceless;		proto-voiced
voice quality:	tense (creaky);	modal (clear)	breathy
vowel quality:	lower (open);		higher (closed)
	more fronted vowels;		more backed vowels;
	tendency to diphthongization;		tendency to centralization;
	often shorter		often longer
pitch distinctions:	higher pitch;		lower pitch;
	association with -ʔ (?)		association with –h (?)
state of larynx:	larynx tense and/or raised (= reduced supraglottal cavity)		larynx lax and/or lowered (=increased supraglottal cavity)

Although 3-way register contrasts do exist (Thurgood (2007)), it is most common to have a two-way contrast. Where this is the case, a language will have one register which corresponds to "unmarked" laryngeal configuration in contrast with another register which corresponds to either “tense” or “breathy” laryngeal configuration. For many Tibetan languages there is a 2-way contrast which has been termed “high” versus “low” register, or “tense” versus “lax” register; with “low” or “lax” register corresponding to “breathy laryngeal configuration” in the table above, and “high” or “tense” register corresponding to “unmarked” in terms of laryngeal configuration (i.e. modal phonation). Clearly a 2-way contrast means that some of the features that are in opposition to those of “breathy laryngeal configuration” will be those listed under “tense laryngeal configuration” (e.g. lower pitch versus higher pitch).

The configuration of the larynx which can account for the phonetic correlations in the different registers, appears to be a combination of vocal cord tension (Ladefoged (1971)) and height of the larynx (Hombert, Ohala et al. (1979)). One of the correlations in Table 2-2 above is between voicing and pitch, with voicelessness and higher pitch co-occurring, and voicing and lower pitch co-occurring. There are a number of explanations for the correlation between pitch and obstruent voicing. Halle and Stevens (1971) suggest that this is due to the stiffness of the vocal folds. Voiceless obstruents are produced with stiff vocal folds, the result of which is that the pitch on the following vowel will be raised. In order to voice obstruents, the vocal folds need to be slackened, and this slackness lowers the pitch on the following vowel. Maran (1971) suggests that the correlation between obstruent voicing and pitch is due to larynx height, with the larynx being raised for both voiceless obstruents and higher pitch, and lowered for voiced obstruents and lower pitch. Hombert, Ohala et al. (1979) give a fuller discussion of the effects of vocal fold tension and height of larynx.

Breathy phonation is produced, at least in part, by the vocal folds being held further apart than for normal phonation, and having less tension (Gordon & Ladefoged (2001)). It is often accompanied by a lowering of the larynx (Thurgood (2007)). Because of this laryngeal configuration, breathy phonation is often accompanied by lower pitch, whereas tenseness often correlates with raised pitch (Thurgood (2007)).

For an in-depth description of how all the phonetic correlates of the above registers can be accounted for by laryngeal configuration, see Thurgood (2007).

2.5 Examples of register and tone in Tibetan languages

A register contrast in Tibetan languages has typically arisen from the loss of word initial consonant clusters and word initial voicing contrasts. These consonant clusters and voicing contrasts can be found in Written Tibetan (WT), which dates back to the 7th century AD. However there is variety in Tibetan languages in terms of what contrasts and clusters have been lost, and hence the synchronic forms of the languages can be different both in terms of what segmental contrasts they have, and in terms of the correlates of register. Of the phonetic correlates of register listed in Table 2-2, for some Tibetan languages the most significant is pitch. Such languages are often referred to as having contrastive tone rather than contrastive register, although there is sometimes disagreement on the terminology (see section 2.7.2 below on Lhasa Tibetan).

The following gives an overview of the variety there is in Tibetan languages. All Written Tibetan (WT) data is given using Wylie transliteration of the Tibetan script (Wylie (1959)). For consistency in comparing the different languages, the term "register" will be used throughout this particular section, even for those languages which would normally be referred to as having a tone contrast rather than a register contrast.

2.5.1 Initial consonant clusters and register

Temchen (Haller (1999)) is an archaic Amdo Tibetan language. It has initial consonant clusters, but does not have a register/tone contrast. This is compared with Shigatse Tibetan (Haller (1999)) which has lost the consonant clusters in favour of a register contrast.

The following data illustrates the difference between Temchen and Shigatse for initial sonorants with and without a WT pre-scripted consonant (a consonant before the sonorant). Temchen has retained the WT consonant clusters whereas for Shigatse the WT consonant clusters correspond to high register (´ denoting high register and ` denoting low register).

Table 2-3: initial sonorants, Shigatse and Temchen

WT	Temchen	Shigatse	Gloss
nga	ŋa	ŋà	‘I’
rnga	rŋa	ŋá	‘drum’
lug	ləç	lùː	‘sheep’
glu	ɣlə	lú	‘song’

The next set of data shows WT initial obstruents with pre-scripted consonants. In Shigatse the pre-scripted consonants have been lost, and voiced obstruents have become voiceless with low register. Temchen has retained the consonant clusters.

Table 2-4: initial obstruents with pre-scripted consonant, Shigatse and Temchen

WT	Temchen	Shigatse	Gloss
spu	ʂ̥pə	pú	‘hair’
'bu	mbə	pù	‘insect’
rta	ʂ̥ta	tá	‘horse’
rdo	rdo	tò	‘stone’

In the case of WT voiced obstruents without any pre-scripted consonants, both Shigatse and Temchen have lost the voicing. For Shigatse, but not Temchen, this loss corresponds to a register contrast.

Table 2-5: initial obstruents without pre-scripted consonant, Shigatse and Temchen

WT	Temchen	Shigatse	Gloss
ja	tɕa	tɕʰà	‘tea’
gong	kuŋ	kʰòŋ	‘price’
kha	kʰa	kʰá	‘mouth’

In order for initial consonant clusters to give rise to a tone/register contrast, the presence or absence of the clusters needs to be associated with a difference in pitch or other laryngeal features. This can be seen in Daofu Tibetan, which is an Amdo Tibetan language (Huang (1997)). Daofu has consonant clusters and not contrastive register. However, Daofu has predictable pitch contours, as illustrated by the following data.

Table 2-6: Daofu Tibetan examples

WT	Daofu	Gloss
snabs	ʂnap^{53}	‘nasal mucus’
nas	neː^{24}	‘barley’
so	s^{h}o^{53}	‘tooth’
zam	zam^{24}	‘bridge’

In the above data the consonant cluster has a high falling pitch contour, whereas the nasal without pre-scripted consonant has a low rising contour. The voiceless obstruent has a high falling contour whereas the voiced obstruent has a low rising contour. If over time the pitch contours become more significant and the pre-scripted consonants and voicing less significant until the point whereby the burden of contrast has shifted from the segments to the pitch, the voicing and pre-scripted consonants can be lost altogether, which is what has happened with Shigatse.

Whilst in most Tibetan languages the WT pre-script + sonorant consonant clusters correspond to high register as in Shigatse above, occasionally the reverse seems to have happened. In Qiuji Tibetan (Sun & Lin (2002)) WT pre-script + sonorant consonant clusters correspond to breathy voice whereas WT sonorants without pre-script correspond to modal voice.

Table 2-7: Qiuji Tibetan initial sonorants

WT	Qiuji Tibetan	Gloss
nga	ŋɐ	‘I’
lnga	ŋa̤	‘five’
mar	maː	‘butter’
sngo	ŋo̤	‘blue/green’

2.5.2 Obstruent voicing and register

When considering the correlation between obstruent voicing and register in Tibetan languages there are several different situations: a) there are languages which have completely lost WT voicing in favour of a register contrast, b) there are languages which have both voicing and register but in correlation with each other, and c) there are languages where contrastive register and voicing are both present and independent of each other.

Qiuji Tibetan (Sun & Lin (2002)) has both phonologically contrastive obstruent voicing and breathy voice (Sun and Lin choose to call this breathy voice rather than register because in many Tibetan languages "register" has become synonymous with "tone", however in Qiuji Tibetan breathiness is not consistently correlated with low pitch). The breathy voice has come from the loss of pre-script + voiced obstruent clusters. In the case of the pre-script being a nasal, the loss of the pre-script has resulted in prenasalization of the voiced obstruent. Otherwise, the obstruent has become voiceless. In the case of no WT pre-script, obstruent voicing has been retained. This is illustrated by the following data.

Table 2-8: Qiuji Tibetan examples for initial obstruents

WT	Qiuji	Gloss
rkos	ko	'to dig'
go	go	'to hear'
sgo	ko̤	'door'
mgo	^{n}go̤	'head'
stong	toŋ	'thousand'
dung	doŋ	'white conch'
gdong	to̤ŋ	'face'
mdung	^{n}do̤ŋ	'spear'

The above data illustrates how in the synchronic form of Qiuji Tibetan, breathy voice and obstruent voicing are independent of each other. That is to say, both voiceless and voiced obstruents occur with both breathy voice and modal voice.

Kyirong Tibetan (Huber (2003)) is another example of a language which has both phonologically contrastive obstruent voicing and register. Unlike Qiuji Tibetan, in Kyirong a register contrast for initial obstruents has arisen from the loss of voicing for WT voiced obstruents without any pre-scripted consonant. WT voiced fricatives plus pre-scripted consonant have lost the pre-script, but the voicing has been retained. In the case of WT voiced plosives plus pre-scripted consonant, the pre-script has been lost, but the resulting voicing and register is determined by the WT pre-script. This means that for words with initial plosives there is a 3-way contrast, which does not appear to have been documented for any other Tibetan language. Huber refers to Kyirong as having three phonemically distinctive register tones. High register tone has modal voice and a high level pitch contour. Mid and low register tones have a slightly rising pitch contour. In addition, low register tone typically has concomitant lax or breathy voice. For consistency with Huber's convention of marking the register tones in Kyirong, in the data below [¯] denotes high, [_] denotes mid, and [_ ʱ] denotes low.

Table 2-9: Kyirong Tibetan examples for initial obstruents

WT	Kyirong	Gloss
tho	tʰō	'list'
lto	tō	'food'
do	to̱	'two'
rdo	to̱ʱ	'stone'
mdo	do̱	'lower valley'
bsangs	sāŋ	'incense'
zangs	sa̱ŋ	'copper'
bzang	za̱ŋ	'wise'

As can be seen in the above table, Kyirong Tibetan has both contrastive register and contrastive voicing. However aspirated plosives only occur in conjunction with high register, and voiced obstruents only occur in conjunction with mid register.

Dolpo Tibetan (own data) is similar to Kyirong Tibetan in that it has both contrastive register and contrastive voicing, with restrictions on co-occurrence. In Dolpo, like Kyirong, WT voiced plosives without a pre-script have become voiceless, and WT voiced plosives with a pre-script have lost their pre-script and either lost or retained their voicing depending on the pre-script. However, unlike Kyirong, Dolpo has a two-way register contrast. And unlike Kyirong, the contrastive voicing of fricatives has been lost completely.

Lhasa Tibetan (Hari (1979)), is an example of a Tibetan language which has lost all WT obstruent voicing contrastively in favour of register contrast.

The following data gives examples of Dolpo and Lhasa Tibetan for initial obstruents.

Table 2-10: Dolpo Tibetan examples for initial obstruents

WT	Dolpo	Lhasa Tibetan	Gloss
so	só	só	'tooth'
bzas	sò	sò	'eat!'
mthong	t^hóŋ	t^hóŋ	'see'
stong	tóŋ	tóŋ	'thousand'
dom	tòm	t^hòm	'bear'
rdo	tò	tò	'stone'
mdo	dò	tò	'lower valley'

Comparing Dolpo with Lhasa Tibetan, one of the differences is in the distribution of aspirated plosives. For Lhasa Tibetan, WT voiced plosives without a pre-script have become voiceless aspirated plosives with low register. This means that aspirated plosives in Lhasa Tibetan occur with both high and low register. However, for Dolpo WT voiced plosives without a pre-script have lost their voicing but have not become aspirated. This means that in Dolpo, aspirated plosives only occur in conjunction with high register. The other difference between Dolpo and Lhasa Tibetan is that whilst in Lhasa Tibetan all phonological voicing has been lost, in Dolpo the voicing has been retained for certain WT pre-scripts. This means that in Dolpo voiced plosives only occur in conjunction with low register.

2.5.3 Summary

The following table gives a summary of the language data discussed above.

Table 2-11: Summary of correlations with WT consonant clusters and voicing

WT pre-script + sonorant			WT N + B			WT R + B			WT B				WT Z			
ret.	> high	> low	ret.	> B low	> P low	ret.	> B low	> P low	> P	> B high	> Ph low	> P low	> S	> Z low	> S low	
x			x			x			x				x			Temchen
	x				x			x			x				x	Shigatse
		x		x				x		x						Quiji
	x			x (m)				x				x (m)		x (m)		Kyirong
	x			x				x				x			x	Dolpo
	x				x			x			x				x	Lhasa

B = voiced plosive; N = nasal; P = voiceless plosive; Ph = aspirated plosive; R = non-nasal pre-script; S = voiceless fricative; Z = voiced fricative; low = low register/breathy; high = high register/modal; m = mid register; ret = retained.

2.6 Pitch and tone contours

In addition to the pitch difference associated with register, for some Tibetan languages the pitch rises or falls as well as, or instead of, remaining level. The pattern of the pitch across the word will be referred to as the pitch melody. For some languages the pitch melody is predictable from the syllable shape, and for other languages it is contrastive in which case the language can be said to have contrastive tone as well as contrastive register. There is a lot of variety between the pitch melodies that occur in different Tibetan languages.

Aba Tibetan (Huang (1997)), an Amdo Tibetan dialect, does not have contrastive register. There is a high-falling pitch melody (53) for all syllable shapes. This can be seen as an embryonic form of the development of contrastive pitch melodies.

Table 2-12: Aba Tibetan pitch melodies

WT	Aba	Gloss
tshwa	$\text{ts}^{\text{h}}\text{æ}^{53}$	‘salt’
stag	ʂ̥tɐx^{53}	‘tiger’
dngul	ʂ̥ŋu^{53}	‘silver’
sgo	ɣgo^{53}	‘door’
na	na^{53}	‘ill’
'don pa	$^{\text{n}}\text{don}^{53}$	‘read’

Daofu Tibetan (Huang (1997)) does not have a register contrast, but has pitch melodies which are predictable from the syllable shape. For WT voiced initial consonants without any pre-script, the Daofu pitch melody is [24]. For other WT initials, the Daofu pitch melody depends upon the syllable rhyme. If the WT rhyme has no coda or a final stop then the Daofu pitch melody is high-falling [53]. If the WT rhyme has a continuant coda or it is an open syllable plus the suffix -*ba*, then the pitch melody is high-level [55].

Table 2-13: Daofu Tibetan pitch melodies

WT	Daofu	Gloss
gcin	χtɕin^{55}	‘urine’
pho ba	haː^{55}	‘belly’
rjes	rdʑiː^{55}	‘trace’
so	s^{h}o^{53}	‘tooth’
snabs	s̥nap^{53}	‘nasal mucus’
zam	zam^{24}	‘bridge’
nas	neː^{24}	‘barley’

In the light of the pitch melodies of Aba and Daofu, neither of which have contrastive register, attention is now turned to three languages which have contrastive register and have predictable pitch melodies according to the register and the syllable rhyme. Below is data from Lhomi (Vesalainen & Vesalainen (1976) and Watters (2003)), Lhasa Tibetan (Chang & Shefts (1964) and Duanmu (1992)) and Dolpo (own data) illustrating different pitch melodies for monosyllabic words. Register is indicated by h (high) and l (low).

Table 2-14: Pitch melodies for 3 Tibetan languages

WT	Dolpo		Lhomi		Lhasa Tibetan		Gloss
kha	k^{h}a^{51}	h	k^{h}a^{51}	h	k^{h}a^{54}	h	‘mouth’
kha ba	k^{h}aː^{551}	h	k^{h}aː^{551}	h	k^{h}aː^{55}	h	‘snow’
dngul	ŋul^{551}	h	ŋuː^{551}	h	ŋuː^{55}	h	‘silver’
shing	ɕiŋ^{551}	h	ɕiŋ^{551}	h	ɕiŋ^{55}	h	‘wood’
khab	k^{h}ap^{54}	h	k^{h}ap^{51}	h	k^{h}ap^{52}	h	‘needle’
me	me^{353}	l	me^{31}	l	me^{12}	l	‘fire’
nas	neː^{253}	l	neː^{131}	l	nɛː^{113}	l	‘barley’
dom	tom^{253}	l	t^{h}om^{131}	l	t^{h}om^{113}	l	‘bear’
lug	luk^{35}	l	luk^{31}	l	luː^{132}	l	‘sheep’

Comparing the words for ‘mouth’ with the words for ‘wood’, all three languages have retained the WT segments, and all three languages show a difference in pitch melody between ‘mouth’ (open syllable, short vowel), and ‘wood’ (sonorant coda). Now consider the word ‘silver’. Dolpo has retained the WT coda whereas the other two languages have lost the coda and gained compensatory vowel lengthening. For ‘snow’ all three languages have lost -*ba* and gained compensatory vowel lengthening. For all three languages, the pitch melodies are the same for words which have lost WT post-vocalic segmental material compared with words which have retained post-vocalic segmental material. This is the same for low register pitch melodies, comparing ‘barley’ and ‘bear’. Thus the pitch melodies over the synchronic long vowels can be analysed as having come from the pitch melody that was there because of segmental material that was originally there. The segmental material deleted in favour of compensatory lengthening, but the pitch melody was retained.

Although the pitch melodies of Dolpo, Lhomi, and Lhasa Tibetan are different from each other, they all share similarities with each other and with Aba and Daofu. Aba has a high-falling melody on all syllable shapes. It is this high-falling pattern which can be seen for all Dolpo and Lhomi high register melodies. In both Dolpo and Lhomi the high register melody across light open syllables is high-falling, and across heavy syllables (long vowel, or sonorant coda) with high register the melody is sustained high and then falling. The falling pattern can also be seen in the Lhomi low register patterns. In Lhasa Tibetan, the sustained high is the same as for Daofu, as is the low register rising pattern.

For the three Tibetan languages above, there is no consistency in the effect that a word final consonant has on the pitch. Consider in the above data the words for ‘mouth’ and ‘needle’. In Dolpo, the presence of the word final plosive in ‘needle’ causes the fall on the preceding vowel to be less than it would otherwise be (pitch pattern [54] instead of [51]). On the other hand, the presence of the word final plosive in Lhasa Tibetan causes the fall on the preceding vowel to be greater (pitch pattern [52] instead of [54]). In Lhomi the fall is the same for both words.

All three languages above have pitch melodies which differ according to syllable shape and register. In languages such as this, particularly in the case of Lhasa Tibetan, there is debate as to whether these are realisations of separate phonological tones, or whether there are simply two registers (or tones, if 'tone' is being used synonymously for 'register'), and within each register the pitch melodies are not phonologically contrastive (see section 2.7.2 below). As well as languages such as these, there are also Tibetan languages where tone, as distinct from register, is indisputably contrastive.

In Kyirong Tibetan (Huber (2005)) in addition to three register tones (see section 2.5.2 above) there are also two contour tones: level and falling. The following data uses Huber's conventions for marking Kyirong register and tone as follows: [¯] denotes high register level tone, [`] denotes high register falling tone, [̠] denotes mid register level tone and [̖] denotes mid register falling tone. Low register is transcribed the same as mid register, but followed by [ʱ]. Falling tone only occurs with long vowels, and thus the contrast between level and falling only occurs for words with long vowels.

Table 2-15: Kyirong tones

WT	Kyirong	Gloss
tshur	tsʰūː	'hither'
stag	tàː	'tiger'
gseb	sēp	'male horse'
ngal	ŋɛ̠ː	'rest, break'
nas	nɛ̖ː	'barley'
rdul	ty̠ʱː	'dust'
bdud	ty̖ʱː	'evil spirit'

The Kyirong falling tone corresponds to WT obstruent codas which have been deleted. However, the WT coda <b> has not been deleted, and the tone on such words is level. WT sonorant codas which have been deleted correspond to level tone.

2.7 Controversy in the analysis of tone in Tibetan languages

There are certain issues of controversy in the analysis of tone in Tibetan languages. These issues include whether it is the voicing of obstruents or the register/tone which is contrastive, whether it is the contour or the vowel length which is contrastive, and what are the underlying tonal contrasts. Each of these issues will be discussed below.

2.7.1 Obstruent voicing or register contrast

Consider the following data from Lhomi, a Tibetan language from Nepal, with the segments written phonetically with regard to voicing of obstruents (h denoting high register, and l denoting low register):

Table 2-16: Obstruent initial words in Lhomi

WT	Lhomi		Gloss	WT	Lhomi		Gloss
so	so	h	'tooth'	zo	so	l	'eat!'
shing	ɕiŋ	h	'firewood'	zhing	ɕiŋ	l	'field'
spags	pak	h	'dough'	'bag	bak	l	'mask'
spu	pu	h	'fur'	'bu	bu	l	'worm'
phag pa	p^hak	h	'pig'	ba	p^ha	l	'cow'

Lhomi is an example of a Tibetan language which has a correlation between phonetically voiced plosives and low register. In analysing the synchronic form of languages such as this, one must address the question of whether there is phonologically contrastive voicing which is determining the register, or whether it is the register which is causing the voicing.

Consider the data for words with initial fricative in Table 2-16 above. The significance of this data is that in Lhomi there are minimal pairs for register. In this case register is independent of the initial consonant. Diachronically the voicing contrast for fricatives has become a register contrast, and the voicing has been lost.

Now consider the data in Table 2-16 above for word initial plosives. Whilst aspirated plosives occur with both high and low register, it can be seen that voiceless unaspirated plosives occur in conjunction with high register, and voiced plosives occur in conjunction with low register.

Given that register has already been shown to be contrastive, one analysis for the plosives is that the voicing of plosives is not a phonological contrast. There are only voiceless plosive phonemes, with the register determining the voicing. This analysis favours economy in the number of consonant phonemes. Vesalainen & Vesalainen (1976) analyse voiced plosives as allophones of voiceless plosives.

An alternative analysis is that voicing on plosives is phonologically contrastive, and is determining the low register of the word. This analysis retains something of the consonant contrasts found in Written Tibetan. Watters (2002) and (2003) analyses Lhomi in this manner.

Voicing or tone in other languages

The correlation between voicing and low tone, and voicelessness and high tone is found in many languages worldwide (Bradshaw (1999)).

The following data is from the Songjiang dialect of Shanghai Chinese (Chen (2000), cited in Yip (2002)):

ti	[53]	'low'	di	[31]	'lift'
ti	[44]	'bottom'	di	[22]	'younger brother'
ti	[35]	'emperor'	di	[13]	'field'

Yip analyses this language as having three contrastive tone patterns: H, HL, and LH, and a feature [$\pm$Upper] the value of which is determined by the voicing on initial obstruents. Obstruent voicing is taken to be phonologically contrastive. However, an alternative analysis of Shanghai Chinese is found in Zee & Xu (2003). They analyse the voiced obstruents as allophones of the voiceless phonemes, with the voicing caused by breathiness on the vowel which is a concomitant of low tone.

Kera, a Chadic language, also has a correlation between voicing and tone. The following data is from Pearce (2005):

[H]	káasáw	‘millet’	[L]	dàagà	‘mat’
[M]	pāatāl	‘needle’	[LH]	dàktə́láw	‘bird’
[MH]	tāatá	‘big jar’	[HL]	táabùl	‘table
[HM]	táasā	‘cup’			

The right hand column of words shows the correlation between voiced obstruents and low tone. Previously it was claimed that Kera voiced obstruents cause the low tone (Ebert (1979), Pearce (1998)). However, Pearce (2005) and (2007) argues that it is the low tone which is causing the voicing.

2.7.2 Tone Contours and/or vowel length

In the synchronic analysis of tone in Tibetan languages, there can be debate as to whether the pitch melodies are realisations of underlying tone contours (be they sequences of high and low tone, or unit contour tones) or whether there is phonological vowel length which is determining the pitch melodies (or, indeed, both). Much of this debate has centred on the tones of Lhasa Tibetan, although similar arguments can be applied to other Tibetan languages. Mazaudon (2005) summarises this debate as follows “Depending on the features that each analyst chooses to include as “tonal”, this makes from 2 (only HIGH and LOW) to 6 “tones” (if length is considered an attribute of tone).”

Consider the following Lhasa Tibetan data (Hu, Qu et al. (1982) cited in Duanmu (1992)):

Table 2-17: Lhasa Tibetan monosyllabic pitch patterns

Lhasa Tibetan			Gloss
	pitch	register	
ka:	55	h	‘column’
kaʔ	52	h	‘prevent’
ka	53	h	‘order’
ka:	13	l	‘install’
kaʔ	121	l	‘clog’
ka	12	l	‘saddle’

Firstly there is a question of terminology. As described in section 2.4 above, register is a combination of phonetic correlates, one of which is pitch. Because in Lhasa Tibetan pitch is salient amongst the phonetic correlates of register, Lhasa Tibetan has usually been referred to as having 2 or more tones, rather than having 2 registers with possible tonal contrasts in addition to that of register.

If vowel length is taken to be phonologically contrastive, then within each register the pitch melodies in Table 2-17 above are predictable according to the syllable shape. Duanmu (1992) provides an explanation for each of the different pitch melodies within each register. He also considers the pitch melodies on polysyllabic words. He concludes that Lhasa Tibetan has 2 phonological tones with underlying forms H (high) and LH (low-high rise). Although Sprigg’s analysis of Lhasa Tibetan is different from Duanmu’s, he also concludes that Lhasa Tibetan has a 2-way tonal contrast: “...Lhasa Tibetan is a register tone language in which two terms are lexically distinguished; contour pitch distinctions have no lexically distinctive role to play.” (Sprigg (1993))

In some dialects of Lhasa Tibetan word final plosives are not pronounced, but the pitch melody caused by the plosive is retained, and there is compensatory lengthening. For example, from the above data *kaʔ 52* ‘prevent’ would be pronounced *ka: 52.* The loss of word final plosives and gaining of compensatory lengthening is reported by Dawson (1980) and Denwood (1999). In dialects where this occurs a pitch contrast exists between *55* and *52*, and between *13* and *121*. Thus Lhasa Tibetan can be analysed as having 4 tones: H (high), L (low), LH (rising) and HL (falling).

Geziben (1996) is an example of 4-tone analysis using the data given in Table 2-17 above (that is, with the word final plosive still present in the data). In this analysis, different Tibetan languages are considered. For some Tibetan languages, there is a fall in pitch without any final plosive either in the synchronic language or the Written Tibetan. The following is an example of this in Dege Tibetan:

Table 2-18: Dege Tibetan

WT	Dege Tibetan	Gloss
rko	ko^{51}	‘to dig’
bkog	koʔ ^{55}	‘to tear’

Geziben argues that given languages such a Dege Tibetan have a fall without the presence of a final plosive, and in addition the presence of a word final plosive does not cause a fall, then it is not the final plosive which is causing the fall in Lhasa Tibetan. He also points out that a drop of 4 or 5 semitones can be considered an indication of falling tone, and that a *51* fall can drop by up to 8 semitones. He argues that it is purely coincidence that in Lhasa Tibetan falling contours and final plosives occur together, and concludes that the falling pitch in Lhasa Tibetan is phonologically contrastive, and thus Lhasa Tibetan has 4 tones not 2 tones.

Hari (1979) is another 4-tone analysis of Lhasa Tibetan. (Her data is slightly different from the data given in Table 2-17 above.) The tone system of Lhasa Tibetan is presented as the intersection of two features: register and contour. These two features combined give rise to 4 contrastive pitch contours, as follows:

Figure 2-1: Lhasa Tibetan contrastive contours (from Hari (1979))

	moving pitch contour	basically level contour
low register	1	2
high register	3	4

Sprigg (1993) gives a discussion of several 4-tone analyses of Tibetan, specifically Chang & Shefts (1964), Hari (1979), and Hu (1982).

In addition to the contrast caused by the loss of word final plosives, if length is taken to be a property of tone, then Lhasa Tibetan can be analysed as having 6 contrastive tones. Dawson (1980) is an example of a 6 tone analysis of Lhasa Tibetan. Dawson's 6 lexically contrastive tones are as follows:

Table 2-19: Lhasa Tibetan 6-way lexical contrast for tone (Dawson (1980))

Tone	description	example		
		segments	pitch	gloss
Tone 1:	(short) high level	ka	5	'first letter of the alphabet'
Tone 2:	(long) high level	kaa	55	'pillar'
Tone 3:	(long) high falling	kaa	52 or 51	'to stop'
Tone 4:	(short) low rising	ka	23 or 2	'saddle'
Tone 5:	(long) low rising	kaa	23 or 24	'to fasten'
Tone 6:	(long) low rising falling	kaa	231	'to be stuck'

Having described the pitch contours of each of the tones as above, Dawson then uses the features proposed by Yip (1980) for the underlying forms of the tones. These features are a register feature [±Upper] and a tone feature [±high] (with [+high] being referred to as H and [-high] as L). The combination of these two features gives rise to the pitch contours as follows:

Table 2-20: Lhasa Tibetan underlying tone features (Dawson (1980))

Tone	Register	Tonality	Predicted pitch:	Citation pitch
Tone 1:	+Upper	H	5	5
Tone 2:	+Upper	H-H	55	55
Tone 3:	+Upper	H-L	53	52 (or 51)
Tone 4:	-Upper	H	2	23 or 2
Tone 5:	-Upper	H-H	22 or 33	23 or 24 (or 22)
Tone 6:	-Upper	H-L	31	231

Dawson notes that whilst the usual pitch contour of Tone 5 is a 23 or 24 rise, 22 is an acceptable variant for the citation form. He also proposes a word initial rise insertion rule to account for the word initial rise of tones 4, 5, and 6.

2.7.3 Underlying forms

Even when there is agreement on the number of tonal contrasts, there can be disagreement as to their underlying form. Again consider Lhasa Tibetan. Both Duanmu (1992) and Sun (1997) analyse Lhasa Tibetan as having a 2-way tonal contrast. For Sun this contrast is between tone H and tone L. But for Duanmu this contrast is between H and LH. As already shown in Table 2-17 above, Lhasa Tibetan pitch melodies are dependent upon the syllable shape. For polysyllabic words, the first syllable may be either high or low pitch. All medial syllables are high, level pitch. The final syllable starts high, with a fall according to the syllable shape.

Duanmu's analysis starts by recognising that it is not sufficient to merely say that as the pitch melodies are predictable they are not phonologically contrastive. It is desirable to provide a phonetic or phonological explanation for the melody with which they surface. He notes that the falling pitch melody quoted as [53] is really [54], but written as [53] for "convenience and visual clarity". (When using diagrammatic form, [53] is ˥˧ and [54] is ˥˦. Indeed, Sun quotes the same data, giving the pitch melody as [54]). This makes it only a very slight drop in pitch, which can be explained as a domain final intonation effect. The greater drop of [52] for a glottal coda can be explained by the sudden closure of the glottis which reduces the vocal cord vibration to zero. This also explains the pattern [121]. The difference in rise between [12] and [13] can be explained by the fact that [12] occurs on a short vowel and so doesn't have time to rise. Thus Duanmu concludes that the pitch melodies above are all realisations of just two phonological tones with underlying forms H (high) and LH (low-high rise). Using an autosegmental framework, Duanmu then gives rules to account for how the tones H and LH spread across a polysyllabic word.

Now consider the analysis of Sun (1997). Like Duanmu, Sun explains the falling pitch melodies on syllables ending with a glottal plosive as being due to the glottal closure. And like Duanmu he explains the slight [54] fall as being a domain final effect. However, whereas Duanmu accounts for the slight rising pitch melody of a light syllable as being a reduced rise due to the short duration, Sun accounts for all monosyllabic rises as being domain final effects. He argues that it is more consistent if both the rises and the falls are considered to be domain final effects than if only the fall is analysed as domain final. He also argues that it is simpler to have tones H and L than to have tones H and LH. Thus he says that the monosyllabic rise is a realisation of tone L not tone LH.

Because all non-initial syllables have high pitch, Sun analyses high tone as being the unmarked default tone, and as such does not necessarily need to be specified in the underlying form. As well as considering the synchronic data, he also argues from the diachronic perspective that high tone corresponds to the unchanged state; it was low tone, not high tone, that was introduced in the tonogenic process. Thus the underlying tones of Lhasa Tibetan are analysed to be H and L, not H and LH. The polysyllabic pitch patterns can be accounted for by saying that the underlying lexical tone (L or H) attaches to the first syllable of a word only. There is then a default tone rule which assigns non-lexical high tone to all subsequent syllables.

2.8 Conclusion on tone and register

Despite the similarities that exist between Tibetan languages at the segmental level, there are differences at the suprasegmental level, particularly regarding tone and register.

Many Tibetan languages have a register contrast, which has arisen through the loss of word initial consonant clusters and obstruent voicing. However, the degree to which obstruent voicing has been lost differs between languages. For some Tibetan languages, both obstruent voicing and register are phonologically contrastive. For other Tibetan languages, the obstruent voicing contrast has been lost altogether. And for yet other Tibetan languages, obstruent voicing is there phonetically, but there is debate as to whether it is phonologically contrastive. In addition, the way in which register surfaces is different for different languages. It is a contrast involving the state of the larynx, which typically involves differences in pitch, phonation, voicing and related concomitants. However the exact mix of phonetic correlates of register is dependent on the language.

As well as having lost initial consonant clusters and obstruent voicing, many Tibetan languages have lost a number of word final consonants, and this loss has given rise to vowel length and to pitch contours. The pitch contours vary between languages, and may or may not be phonologically contrastive, depending on language. And the same data may be analysed in different ways to give very different conclusions as to what is phonologically contrastive, and what is not.

It is these issues which provide the springboard from which to analyse tone in Walungge.

3 Phonemic summary

This chapter is a description of the consonant and vowel phonemes of Walungge, both from a synchronic and a diachronic perspective. The synchronic analysis describes the consonant and vowel phonemes, giving examples of contrast, and a description of the realisation of the phonemes including any allophonic variation. It also includes a description of the syllable structures found in Walungge. The diachronic analysis compares Walungge with Written Tibetan.

3.1 Consonants

3.1.1 Consonant phonemes

The following is a chart of the consonant phonemes of Walungge.

Table 3-1: Consonant phonemes

	labial	alveolar	retroflex	palatal	velar	glottal
plosive	/p/, /p^h/	/t/, /t^h/	/ʈ/, /ʈ^h/	/c/, /c^h/	/k/, /k^h/	
affricate		/ts/, /ts^h/		/tɕ/, /tɕ^h/		
fricative		/s/		/ɕ/		/h/
nasal	/m/	/n/		/ɲ/	/ŋ/	
vibrant		/r/, /r̥/				
lateral		/l/, /l̥/				
approximant	/w/			/j/		

3.1.2 Plosives

Unaspirated plosives

Walungge has the following unaspirated plosive phonemes: /p/, /t/, /ʈ/, /c/ and /k/. These are realised as voiced or voiceless depending on the context. For a full discussion as to why voicing of plosives has been analysed as not phonologically contrastive, see chapter 6.

Word initially, the voicing of plosives depends upon the tone of the word, with plosives becoming voiced before low tone. This is illustrated by the following examples.

/pák/	[paʔ ^{53}]	'pastry'	/péːtsa/	[peːza $^{44\ 44}$]	'lettuce'
/pàk/	[baʔ ^{24}]	'bamboo variety'	/pèːla/	[beːla $^{11\ 44}$]	'cement'
/túŋ/	[tuŋ ^{44}]	'drink'	/tá/	[ta ^{52}]	'horse'
/tùŋ/	[duŋ ^{14}]	'hit'	/tà/	[da ^{24}]	'bow' (and arrow)
/ʈá/	[ʈa ^{52}]	'hair'	/ʈíŋ/	[ʈiŋ ^{44}]	'high mist'
/ʈà/	[ɖa ^{24}]	'enemy'	/ʈì/	[ɖi ^{24}]	'female yak'
/cákpa/	[cakpa $^{44\ 44}$]	'excrement'	/cóː/	[coː ^{44}]	'ladle'
/càkre/	[ɟagre $^{11\ 44}$]	'fight'	/còwo/	[ɟowo $^{11\ 44}$]	'fast'
/kú/	[ku ^{52}]	'statue'	/kánti/	[kandi $^{44\ 44}$]	'sole of foot'
/kù/	[gu ^{24}]	'nine'	/kànte/	[gande $^{11\ 44}$]	'good'

The following are examples of word medial plosives following a nasal. All positions of articulation can occur, and all are realised as voiced regardless of tone.

/sámpa/	[samba $^{44\ 44}$]	'new'
/rìŋpo/	[riŋbo $^{11\ 44}$]	'long'
/kònta/	[gonda $^{11\ 44}$]	'monastery'
/síŋtaŋ/	[siŋdaŋ $^{44\ 44}$]	'grass'
/k^{h}ànʈa/	[k^{h}aɳɖa $^{11\ 44}$]	'what'
/námʈu/	[namɖu $^{44\ 44}$]	'aeroplane'
/sìŋca/	[siŋɟa $^{11\ 44}$]	'pink'

/káŋca/	[kaŋɟa $^{44\ 44}$]	‘stem’
/lǿŋkoŋ/	[løŋgoŋ $^{44\ 44}$]	‘wet’
/p^{h}àŋkok/	[p^{h}aŋgoʔ $^{11\ 44}$]	‘chest’

In a word medial syllable coda, the only plosives which can occur are /p/ /t/, and /k/. In this position /t/ has a very limited distribution, only occurring before another coronal plosive or affricate.

/ʈỳttok/	[ɖytːoʔ $^{11\ 44}$]	‘knot’
/sáttiŋ/	[satːiŋ $^{44\ 44}$]	‘floor’
/háttʈa/	[haʈːa $^{44\ 44}$]	‘bazaar’
/ŋáttʈo/	[ŋaʈːo $^{44\ 44}$]	‘morning’
/máttsi/	[matsi $^{44\ 44}$]	‘chilli’
/kéttɕa/	[ketɕa $^{44\ 44}$]	‘language’

When /p/ and /k/ occur before a voiced consonant, they are voiced, regardless of the tone of the word:

/ɕàpro/	[ɕabro $^{11\ 44}$]	‘dance’
/káple/	[kable $^{11\ 44}$]	‘skull’
/càkre/	[ɟagre $^{11\ 44}$]	‘fight’
/áklawa/	[aglawa $^{44\ 44\ 21}$]	‘crow’

/p/ and /k/ can occur before any other plosive or affricate, including another /p/ or /k/. The following are examples.

/làkpa/	[lakpa $^{11\ 44}$]	‘hand’
/sákti/	[sakti $^{44\ 44}$]	‘secret’
/tsákʈa/	[tsakʈa $^{44\ 44}$]	‘matches’
/lópʈa/	[lopʈa $^{44\ 44}$]	‘school’
/kùpca/	[gupca $^{11\ 44}$]	‘chair’
/t^{h}íptɕu/	[t^{h}iptɕu $^{44\ 44}$]	‘button’
/tɕàktɕa/	[dʑaktɕa $^{11\ 44}$]	‘broom’

/séppo/	[sepːo $^{44\ 44}$]	‘yellow’
/kàppo/	[gapːo $^{11\ 44}$]	‘all’
/tàkca/	[dacːa $^{11\ 44}$]	‘moon’

The following are examples of /p/ and /k/ in a word medial syllable coda, occurring before a fricative. They are realised as the fricatives [ɸ] and [x] respectively.

/ʈùpse/	[duɸse $^{11\ 44}$]	‘grain’
/tókse/	[toxse $^{44\ 44}$]	‘adze’

The plosives /t/, /c/ and /k/ may all occur intervocalically, where they are realised as voiced. /p/ does not occur intervocalically; there is intervocalic neutralisation of contrast between /p/ and /w/ (see section 3.1.7). So far no instances of intervocalic /ʈ/ have been found.

/ɕùtok/	[ɕudoʔ $^{11\ 44}$]	‘spindle’
/kútaŋ/	[kudaŋ $^{44\ 44}$]	‘religious painting’
/ràco/	[raɟo $^{11\ 44}$]	‘horn’
/jìkok/	[jigoʔ $^{11\ 44}$]	‘soap’
/tɕʰúcoŋ/	[tɕʰuɟoŋ $^{44\ 44}$]	‘stream’
/kíken/	[kigẽː $^{44\ 44}$]	‘dog’

Word final plosives

The only plosives which can occur word finally are /p/, /t/, and /k/. /p/ is unreleased ([p̚]), and at a pause boundary both /t/ and /k/ are realised as the glottal plosive [ʔ].

/kʰáp/	[kʰap̚ ^{53}]	‘needle’
/càp/	[ɟap̚ ^{24}]	‘nation’
/két/	[keʔ ^{53}]	‘voice’
/tỳt/	[dyʔ ^{24}]	‘demon’
/mík/	[miʔ ^{53}]	‘eye’
/lùk/	[luʔ ^{24}]	‘sheep’

Although word finally both /t/ and /k/ are realised as [ʔ] before a pause, in other contexts they behave differently from each other. If they occur word finally within a phrase /k/ is realised as [k], however there is free variation between realising /t/ as the glottal plosive [ʔ], and deleting the final consonant altogether but gaining compensatory lengthening on the preceding vowel. In general, the word final /t/ is more likely to be realised as [ʔ] in careful speech, but deleted otherwise. The pitch melodies rise or fall regardless of whether or not there is consonant deletion. The following are examples of this:

/tỳt/	[dyʔ ^{24}]	or	[dyː ^{14}]	'demon'
/pǿt/	[pøʔ ^{53}]	or	[pøː ^{51}]	'story'

The plosives /t/ and /k/ also behave differently from each other when suffixes are added. The following is an example of the locative suffix /-la/, which starts with a sonorant. In this case /k/ is realised as a voiced velar plosive whereas /t/ is deleted and there is compensatory lengthening on the vowel.

/lùk/	[luʔ ^{24}]	'sheep'	/lùk/ /-la/	[lugla $^{24\ 21}$]	'to the sheep'
/tỳt/	[dyʔ ^{24}]	'demon'	/tỳt/ /-la/	[dyːla $^{14\ 21}$]	'to the demon'

The following is an example of the nominalizing suffix /-pa/, which starts with a plosive. In this instance /k/ is realised as [k] whereas /t/ completely assimilates to the following plosive.

/ʈʰùk/	[ʈʰuʔ ^{24}]	'six'	/ʈʰùk/ /-pa/	[ʈʰukpa $^{11\ 44}$]	'6 o clock'
/cèt/	[ɟeʔ ^{24}]	'eight'	/cèt/ /-pa/	[ɟepːa $^{11\ 44}$]	'8 o clock'

There is no overlap in distribution between word final /k/ and /t/, with /t/ following the vowels /y/, /ø/, /e/, and /k/ following the vowels /i/, /a/, /o/, /u/. For information as to how this corresponds to Written Tibetan, see sections 3.5.4 and 3.5.6 below.

Aspiration

All places of articulation for plosives show a contrast between aspirated and unaspirated plosives word initially, which is the only environment in which aspiration occurs. Aspirated plosives can occur with both high tone and low tone. The following are examples of the contrast between aspirated and unaspirated plosives.

/p^{h}áŋ/	[p^{h}aŋ ^{44}]	‘spindle’	/p^{h}àː/	[p^{h}aː ^{14}]	‘wool’
/páŋ/	[paŋ ^{44}]	‘lap’	/pàː/	[baː ^{14}]	‘goitre’
/t^{h}ókme/	[t^{h}ogme $^{44\ 44}$]	‘lightning’	/t^{h}òm/	[t^{h}om ^{14}]	‘bear’
/tókse/	[toxse $^{44\ 44}$]	‘hoe’	/tòŋ/	[doŋ ^{14}]	‘front’
/ʈ^{h}ák/	[ʈ^{h}aʔ ^{53}]	‘cliff’	/ʈ^{h}èː/	[ʈ^{h}eː ^{14}]	‘mule’
/ʈá/	[ʈa ^{52}]	‘hair’	/ʈèː/	[ɖeː ^{14}]	‘rice’
/c^{h}ówa/	[c^{h}owa $^{44\ 44}$]	‘husband’	/c^{h}ùkpak/	[c^{h}ukpaʔ $^{11\ 44}$]	‘log’
/cóː/	[coː ^{44}]	‘ladle’	/càrok/	[ɟaroʔ $^{11\ 44}$]	‘beard’
/k^{h}áŋpa/	[k^{h}aŋba $^{44\ 44}$]	‘full’	/k^{h}ànʈa/	[k^{h}anɖa $^{11\ 44}$]	‘what’
/káŋpa/	[kaŋba $^{44\ 44}$]	‘leg’	/kànte/	[gande $^{11\ 44}$]	‘good’

Contrast between alveolar and retroflex

The following pairs of words demonstrate the contrast between alveolar and retroflex positions of articulation:

/tá/	[ta ^{52}]	‘horse’	/t^{h}ákpa/	[t^{h}akpa $^{44\ 44}$]	‘rope’
/ʈá/	[ʈa ^{52}]	‘hair’	/ʈ^{h}ák/	[ʈ^{h}aʔ ^{24}]	‘cliff’
/tà/	[da ^{24}]	‘bow’	/t^{h}áː/	[t^{h}aː ^{44}]	‘loom’
/ʈà/	[ɖa ^{24}]	‘enemy’	/ʈ^{h}áː/	[ʈ^{h}aː ^{44}]	‘tax’

Contrast between palatal and velar

The contrast between /c/ and /k/ and between /c^{h}/ and /k^{h}/ occurs before all vowels except /i/. Before /i/ only /k/ and / k^{h}/ occur; the place of articulation is slightly further forward than velar but is not as far forward as /c/ and /c^{h}/. The following examples show the contrast between palatal and velar positions before vowels other than /i/. Note in particular that there is a contrast between palatal and velar before other front vowels.

/céwa/	[cewa $^{44\ 44}$]	'birth'	/c^{h}ówa/	[c^{h}owa $^{44\ 44}$]	'husband'
/kéttɕa/	[ketɕa $^{44\ 44}$]	'language'	/k^{h}ólo/	[k^{h}olo $^{44\ 44}$]	'wheel'
/cýppo/	[cypːo $^{44\ 44}$]	'sour'	/càkre/	[ɟagre $^{11\ 44}$]	'fight'
/kýttok/	[kytːoʔ $^{44\ 44}$]	'thread'	/kàkca/	[gacːa $^{11\ 44}$]	'ginger'

3.1.3 Affricates

There are two places of articulation for affricates: alveolar and alveolo-palatal. As with plosives, there is a contrast between aspirated and unaspirated in word initial position, which is the only position in which aspiration occurs. Aspirated affricates occur with both high tone and low tone. Before low tone, unaspirated affricates have a significantly reduced VOT compared with unaspirated affricates before high tone (see section 6.3.3.2). For low tone words in isolation, the affricates are not truly voiced in that the VOT does not generally become negative. However, for low tone words in phrases the closure of the affricate does have voicing. For this reason they have been transcribed in the phonetic form as voiced. The following pairs are examples of the contrast between aspirated and unaspirated affricates.

/tɕ^{h}ú/	[tɕ^{h}u ^{52}]	'water'	/tɕ^{h}à/	[tɕ^{h}a ^{24}]	'tea'
/tɕú/	[tɕu ^{52}]	'ten'	/tɕà/	[dʑa ^{24}]	'rainbow'
/ts^{h}á/	[ts^{h}a ^{52}]	'salt'	/ts^{h}émak/	[ts^{h}emaʔ $^{44\ 44}$]	'thorn'
/tsá/	[tsa ^{52}]	'grass'	/tsémo/	[tsemo $^{44\ 44}$]	'game'

Whilst aspirated affricates only occur word initially, both /ts/ and /tɕ/ can occur word medially in a syllable onset. Intervocalically they occur with both high and low tone, and are weakened to be realised as the fricatives [z] and [ʑ] respectively. The following are examples of intervocalic affricates.

/tsátsik/	[tsaziʔ $^{44\ 44}$]	'meaning'
/t^{h}òtse/	[t^{h}oze $^{11\ 44}$]	'orphan'
/sátɕa/	[saʑa $^{44\ 44}$]	'earth, soil'
/látɕa/	[laʑa $^{44\ 44}$]	'wage'

Following a voiced consonant, affricates are always voiced:

/ts^{h}èmtsi/	[ts^{h}emdzi $^{11\ 44}$]	'scissors'
/lúŋtsup/	[luŋdzup˺ $^{44\ 44}$]	'storm'
/ámtɕok/	[amdʑoʔ $^{44\ 44}$]	'ear'
/màŋtɕak/	[maŋdʑaʔ $^{11\ 44}$]	'red millet'

Following a voiceless consonant, affricates are always voiceless:

/t^{h}íptɕu/	[t^{h}iptɕu $^{44\ 44}$]	'button'
/tùptɕa/	[duptɕa $^{11\ 44}$]	'bad'

If the preceding consonant is /t/, the sequences /tts/ and /ttɕ/ are realised as [ts] and [tɕ]. For more on this, see section 6.2.3.

/máttsi/	[matsi $^{44\ 44}$]	'chilli'
/ŋòttse/	[ŋotse $^{11\ 44}$]	'shy'
/kéttɕa/	[ketɕa $^{44\ 44}$]	'language'
/áttɕe/	[atɕe $^{44\ 44}$]	'elder sister'

3.1.4 Fricatives

Walungge has the three fricative phonemes /s/, /ɕ/ and /h/. /h/ only occurs word initially, whereas /s/ and /ɕ/ can occur in any syllable initial position. All occur with both high and low tone.

In words with low tone, /h/ starts voiceless but often it very rapidly becomes voiced. The voicing starts well before the transition from the fricative into the following vowel. It has thus been transcribed as [ɦ]. There is no apparent phonetic difference in the articulation of /s/ and /ɕ/ with low tone compared to high tone.

The following are examples of word initial fricatives with both high and low tone:

/háttʈa/	[haʈːa $^{44\ 44}$]	'bazaar'
/hàluŋ/	[ɦaluŋ $^{11\ 344}$]	'Halung village'
/sì/	[si ^{24}]	'zi stone'
/sá/	[sa ^{52}]	'earth'
/ɕì/	[ɕi ^{24}]	'four'
/ɕá/	[ɕa ^{52}]	'meat'

The fricatives /s/ and /ɕ/ occur following both voiceless and voiced consonants, and are always voiceless. The following are examples of fricatives following consonants:

/ʈùpse/	[duɸse $^{11\ 44}$]	'grain'
/tókse/	[toxse $^{44\ 44}$]	'hoe'
/háŋse/	[haŋse $^{44\ 44}$]	'duck'
/tɕʰáŋsa/	[tɕʰaŋsa $^{44\ 44}$]	'marriage'
/pákɕok/	[paxɕoʔ $^{44\ 44}$]	'bread'
/sáŋciŋ/	[saŋɕiŋ $^{44\ 44}$]	'pine tree'

Similarly intervocalic fricative phonemes are always voiceless. Because affricates become voiced fricatives intervocalically, this gives a surface contrast between voiced and voiceless fricatives.

/léɕa/	[leɕa $^{44\ 44}$]	'thigh'	/mèsøː/	[mesøː $^{11\ 44}$]	'tinder'
/látɕa/	[laʑa $^{44\ 44}$]	'wage'	/mètsak/	[mezaʔ $^{11\ 44}$]	'spark'

A possible reason why an affricate will become a voiced fricative whereas a phonemic fricative will remain voiceless is the difference in duration. The fricative portion of the affricate is much shorter in duration than a phonemic fricative. This makes it more likely to become voiced in a voiced environment (see section 6.5.3).

3.1.5 Nasals

The four nasal phonemes, /m/, /n/, /ɲ/ and /ŋ/ all occur word initially with both high and low tone, and are in contrast with each other:

/màː/	[maː ^{14}]	'butter'
/nàŋ/	[naŋ ^{14}]	'house'
/ɲà/	[ɲa ^{24}]	'fish'
/ŋà/	[ŋa ^{24}]	'I'
/mén/	[mẽː ^{44}]	'medicine'
/náː/	[naː ^{44}]	'blue-sheep'
/ɲáː/	[ɲaː ^{44}]	'put to sleep'
/ŋáː/	[ŋaː ^{44}]	'harvest'

Nasals can occur word medially following a non-homorganic nasal, or a plosive:

/námnuk/	[namnuʔ $^{44\ 44}$]	'snot'
/çùŋɲi/	[çuŋɲi $^{11\ 44}$]	'sun'
/çíŋmu/	[çiŋmu $^{44\ 44}$]	'louse'
/t^{h}ókme/	[t^{h}ogme $^{44\ 44}$]	'lightning'

All can occur intervocalically:

/tìmik/	[dimiʔ $^{11\ 44}$]	'key'
/séno/	[seno $^{44\ 44}$]	'fingernail'
/tàɲuk/	[daɲuʔ $^{11\ 44}$]	'arrow'
/ʈ^{h}áŋa/	[ʈ^{h}aŋa $^{44\ 44}$]	'rosary'

The palatal nasal /ɲ/ only occurs as a syllable onset, whereas the other nasals can also occur as syllable codas.

The following are examples of /m/, /n/, and /ŋ/ word finally. In this position, /n/ is realised as nasalisation and duration on the vowel.

/nám/	[nam ^{44}]	'sky'
/kàm/	[gam ^{14}]	'trunk'

/mén/	[mẽː ^{44}]	‘medicine’
/tỳn/	[dỹː ^{14}]	‘seven’
/sáŋ/	[saŋ ^{44}]	‘pine’
/ràŋ/	[raŋ ^{14}]	‘2SG’

The following are examples of nasals in a word medial syllable coda. Both /m/ and /ŋ/ can occur before consonants of all places of articulation:

/sámpa/	[samba $^{44\ 44}$]	‘new’
/c^{h}émra/	[c^{h}emra $^{44\ 44}$]	‘gardening fork’
/námʈu/	[namɖu $^{44\ 44}$]	‘aeroplane’
/ts^{h}ómcuk/	[ts^{h}omɟuʔ $^{44\ 44}$]	‘stone crusher’
/tsáŋpu/	[tsaŋbu $^{44\ 44}$]	‘river’
/kánti/	[kandi $^{44\ 44}$]	‘sole of foot’
/háŋse/	[haŋse $^{44\ 44}$]	‘duck’
/çìŋka/	[çiŋga $^{11\ 44}$]	‘field’
/sìŋca/	[siŋɟa $^{11\ 44}$]	‘pink’

Word internally, /n/ only occurs before another coronal consonant, assimilating in place of articulation to that consonant. The following are examples of word medial /n/:

/kànte/	[gande $^{11\ 44}$]	‘good’
/k^{h}ànʈa/	[k^{h}aɳɖa $^{11\ 44}$]	‘what’

3.1.6 Lateral and vibrant

Walungge has the lateral and vibrant phonemes /l/, /l̥/, /r/ and /r̥/. The voiceless phonemes /l̥/ and /r̥/ only occur word initially, and only with high tone.

The following words illustrate the contrast between /l/ and /l̥/.

/l̥ámʈo/	[l̥amɖo $^{44\ 44}$]	‘boot tie’	/l̥énta/	[l̥enda $^{44\ 44}$]	‘patch’
/láŋpa/	[laŋba $^{44\ 44}$]	‘steam’	/létta/	[letːa $^{44\ 44}$]	‘brain’

The phoneme /l/ occurs word initially with both high and low tone. The following words illustrate this.

/láŋ/	[laŋ ^{44}]	‘bull’
/làŋ/	[laŋ ^{14}]	‘stand’
/lúk/	[luʔ ^{53}]	‘pour’
/lùk/	[luʔ ^{24}]	‘sheep’

Word initially [r] only occurs with low tone and [r̥] with high tone. Because of this complementary distribution, the question is raised as to whether [r] and [r̥] are allophones of the one phoneme /r/. For a full discussion of this see section 6.4, the conclusion of which is that /r/ and /r̥/ are separate phonemes. The following are examples of word initial /r̥/ and /r/.

/r̥éna/	[r̥ena $^{44\ 44}$]	‘beans’
/r̥ýː/	[r̥yː ^{44}]	‘corn’
/rà/	[ra ^{24}]	‘goat’
/rỳt/	[ryʔ ^{24}]	‘landslide’

Both /l/ and /r/ can occur word medially as a syllable onset, with both high and low tone. This is illustrated by the following words:

/jàŋlaː/	[jaŋlaː $^{11\ 344}$]	‘candle’
/ts^{h}ílu/	[ts^{h}ilu $^{44\ 44}$]	‘fat’
/ŋòloŋ/	[ŋoloŋ $^{11\ 44}$]	‘face’
/tíŋri/	[tiŋri $^{44\ 44}$]	‘wooden crusher’
/tòri/	[dori $^{11\ 44}$]	‘stone’
/l̥ári/	[l̥ari $^{44\ 44}$]	‘artist’

/r/ also occurs word finally:

/k^{h}ùr/	[k^{h}ur ^{14}]	‘tent’
/sér/	[ser ^{44}]	‘gold’

3.1.7 Approximants

/w/

Word initially the occurrence of the approximant /w/ is extremely rare. With the exception of proper names, the only words found so far to have a word initial /w/ are:

/wàŋ/	[waŋ ^{14}]	‘religious observance; power’
/wà/	[wa ^{24}]	‘a type of fox’

The occurrence of /w/ intervocalically is considerably more common. The following are examples of intervocalic /w/:

/ʈʰàwe/	[ʈʰawe $^{11\ 44}$]	‘buckwheat’
/rìwoŋ/	[riwoŋ $^{11\ 344}$]	‘hare’
/tʰéwo/	[tʰewo $^{44\ 44}$]	‘thumb’
/cʰéwak/	[cʰewaʔ $^{44\ 44}$]	‘ice’

Intervocalically there is neutralisation between /w/ and /p/, with both /p/ and /w/ being realised as [w] intervocalically regardless of the tone of the word. This next example illustrates the contrast and also the neutralisation between /w/ and a bilabial plosive. The comparison is between the plural and the nominalised forms of numbers. The plural morpheme is /-wa/ with optional insertion of a velar plosive occurring following a non-velar consonant; the nominalizer morpheme is /-pa/, with /p/ being realised as [b] next to a voiced consonant.

/súm/	/-pa/	→	[sumba $^{44\ 44}$]	/súm/	/-wa/	→	[sumgwa $^{44\ 21}$]~[sumwa $^{44\ 21}$]
‘three’	NMLZ		‘third one’	‘three’	PL		‘three.PL’
/ʈʰùk/	/-pa/	→	[ʈʰukpa $^{11\ 44}$]	/ʈʰùk/	/-wa/	→	[ʈʰugwa $^{24\ 21}$]
‘six’	NMLZ		‘sixth one’	‘six’	PL		‘six.PL’
/kù/	/-pa/	→	[guwa $^{11\ 44}$]	/kù/	/-wa/	→	[guwa $^{24\ 21}$]
‘nine’	NMLZ		‘ninth one’	‘nine’	PL		‘nine.PL’

The above data shows that in a non-intervocalic environment the contrast between /-pa/ and /-wa/ is retained. However, intervocalically both are realised as [-wa].

/j/

The approximant /j/ can occur both word initially and intervocalically. It occurs with both high and low tone:

/jú/	[ju ^{52}]	‘turquoise’
/jù:/	[ju: ^{14}]	‘village’
/kájo:/	[kajo: $^{44\ 44}$]	‘teacup’
/màjen/	[majẽ: $^{11\ 44}$]	‘buffalo’

3.2 Vowels

3.2.1 Vowel phonemes

The following is a chart of the vowel phonemes that occur in Walungge:

Table 3-2: Vowel phonemes of Walungge

	front		central	back
	unrounded	rounded		
close	/i/	/y/		/u/
mid	/e/	/ø/		/o/
open			/a/	

Additionally there is phonological vowel length, giving the phonemes /i:/, /e:/, /y:/, /ø:/, /a:/, /o:/, /u:/.

3.2.2 Length contrasts

The five vowels /i/, /e/, /a/, /o/, /u/ contrast with their long counterparts /iː/,/eː/, /aː/, /oː/, /uː/ in both monosyllabic and polysyllabic words. The monosyllabic contrast is illustrated by the following examples:

/ʈʰì/	[ʈʰi ^{24}]	‘knife’	/mè/	[me ^{24}]	‘fire’
/ʈʰìː/	[ʈʰiː ^{14}]	‘write’	/nèː/	[neː ^{14}]	‘barley’
/ɕá/	[ɕa ^{52}]	‘meat’	/lò/	[lo ^{24}]	‘year’
/ɕáː/	[ɕaː ^{44}]	‘deer’	/lòː/	[loː ^{14}]	‘wall’
/kù	[gu ^{24}]	‘nine’			
/kùː/	[guː ^{14}]	‘wait’			

These five vowels and their long counterparts may all occur in disyllabic words. The following examples show the length contrast in the first syllable of a disyllabic word:

/sìma/	[sima $^{11\ 44}$]	‘eyelash’	/mèrok/	[meroʔ $^{11\ 44}$]	‘burning coal’
/sìːwa/	[siːwa $^{11\ 44}$]	‘dew’	/rèːlok/	[reːloʔ $^{11\ 44}$]	‘kid’
/ɲùŋa/	[ɲuŋa $^{11\ 44}$]	‘less’	/kòmo/	[gomo $^{11\ 44}$]	‘money’
/ŋúːwa/	[ŋuːwa $^{44\ 44}$]	‘cloud’	/kòːmo/	[goːmo $^{11\ 44}$]	‘evening’
/káma/	[kama $^{44\ 44}$]	‘star’			
/káːma/	[kaːma $^{44\ 44}$]	‘white rhododendron’			

The following are examples of long vowels in the second syllable of a disyllabic word:

/kòreː/	[goreː $^{11\ 344}$]	‘shawl’
/ʈʰòmaː/	[ʈʰomaː $^{11\ 344}$]	‘potato’
/kájoː/	[kajoː $^{44\ 44}$]	‘teacup’

/ɲémpuː/ [ɲembuː $^{44\ 44}$] ‘pillow’

Although the front vowels /y/ and /ø/ are separate phonemes from /yː/ and /øː/, the length is predictable. The evidence for short and long vowels being separate phonemes is the pitch melodies. A word such as [pøː ^{44}] ‘incense’ has a [44] pitch melody, which occurs on monosyllabic bimoraic nouns. A word such as [pøʔ ^{53}] ‘story’ has a [53] pitch melody, the pattern which occurs on all monosyllabic monomoraic nouns. A word such as [pøʔ ^{53}] may optionally drop the coda consonant, in which case it is pronounced [pøː ^{51}]. Although in the phonetic form there is a vowel which has additional duration, the prosodic structure still recognises the word as underlying monomoraic, so it surfaces with a falling pitch melody. Thus, the prosodic structure must be recognising [pøː ^{44}] as underlyingly bimoraic (for a full discussion of prosodic structure and tone, see chapter 5).

In monosyllabic words, the short vowels /y/ and /ø/ only occur when followed by the coronal consonants /t/ or /n/. This restricted distribution is a result of the diachronic phonology (see section 3.5.6).

/tỳt/	[dyʔ ^{24}]	‘demon’
/tỳn/	[dỹː ^{14}]	‘seven’
/ts^{h}ǿn/	[ts^{h}ø̃ː ^{44}]	‘paint’
/tø̀t/	[døʔ ^{24}]	‘sit’

The long vowels /yː/ and /øː/ occur in monosyllabic open syllables, e.g.

/pǿː/	[pøː ^{44}]	‘incense’
/ŋỳː/	[ŋyː ^{14}]	‘cry’

Vowels in polysyllabic words follow the same patterns as described above. Front rounded vowels in closed syllables are short, whereas front rounded vowels in open syllables are long. However, because the first syllable of a disyllabic word is normally unstressed, the long vowels /yː/ and /øː/ often lose an element of duration in this position. The following are examples of front rounded vowels in polysyllabic words:

/cýppo/	[cypːo $^{44\ 44}$]	‘sour’
/hýːra/	[hyːra $^{44\ 44}$]	‘whistle’
/jàŋkyt/	[jaŋgyʔ $^{11\ 44}$]	‘pumpkin’
/sǿːlok/	[søːloʔ $^{44\ 44}$]	‘charcoal’
/tɕʰýːtsøt/	[tɕʰyːzøʔ $^{44\ 44}$]	‘hour’
/pʰø̀tta/	[pʰøtːa $^{11\ 44}$]	‘Tibetan’

3.2.3 Front and back rounded vowel contrasts

Walungge contrasts front and back rounded vowels. The following are examples of front and back rounded vowel contrasts for monosyllabic words with long vowels.

/ʈýː/	[ʈyː ^{51}]	‘wash’	/r̥ýː/	[r̥yː ^{44}]	‘corn’
/ʈùː/	[ɖuː ^{14}]	‘snake’	/r̥úː/	[r̥uː ^{44}]	‘aunt’
/pǿː/	[pøː ^{44}]	‘incense’	/mǿː/	[møː ^{44}]	‘ploughing’
/póː/	[poː ^{44}]	‘old man’	/móː/	[moː ^{44}]	‘old woman’

Because the vowels /y/ and /ø/ have originated from the vowels /u/ and /o/ when followed by certain coronal codas (see section 3.5.6 below for a summary of the diachronic phonology of vowels) short rounded vowels do not contrast for front and backness when in monosyllabic words. In closed syllable monosyllabic words, /y/ and /ø/ occur before the coronal codas /t/ and /n/, whereas /u/ and /o/ occur before the coronal coda /r/, and all non coronal codas. The following data illustrates this.

/tỳt/	[dyʔ ^{24}]	‘demon’	/ʈùk/	[ɖuʔ ^{24}]	‘dragon’
/tỳn/	[dỹː ^{14}]	‘seven’	/tùŋ/	[duŋ ^{14}]	‘hit’
/tsʰǿn/	[tsʰø̃ː ^{44}]	‘colour’	/tsʰóm/	[tsʰom ^{44}]	‘stone crusher’
/pǿt/	[pøʔ ^{53}]	‘story’	/pór/	[por ^{51}]	‘spill’

However in the first syllable of a polysyllabic word, both front and back rounded vowels can occur in a CVC syllable before both coronal and non coronal places of articulation:

/tɕʰýttøn/	[tɕʰytːø̃ː $^{44\ 44}$]	‘shrine’	/tɕʰútto/	[tɕʰutːo $^{44\ 44}$]	‘beak’
/tỳmpa/	[dymba $^{11\ 44}$]	‘7 o clock’	/súmpa/	[sumba $^{44\ 44}$]	‘3 o clock’
/tɕø̀ntaŋ/	[dʑøndaŋ $^{11\ 344}$]	‘brave’	/sònta/	[sonda $^{11\ 44}$]	‘boot’
/pʰø̀tta/	[pʰøtːa $^{11\ 44}$]	‘Tibetan’	/ʈʰòtta/	[ʈʰotːa $^{11\ 44}$]	‘stomach’
/lǿŋkoŋ/	[løŋgoŋ $^{44\ 44}$]	‘wet’	/tsóŋkaŋ/	[tsoŋgaŋ $^{44\ 44}$]	‘shop’

3.2.4 Vowel height contrasts

The following pairs of words illustrate the contrast between the close vowels /i/, /y/, /u/ and the mid vowels /e/, /ø/ and /o/. The contrast exists for both long and short vowels.

/mì/	[mi ^{24}]	‘person’	/ʈʰìː/	[ʈʰiː ^{14}]	‘write’
/mè/	[me ^{24}]	‘fire’	/ʈʰèː/	[ʈʰeː ^{14}]	‘mule’
/tỳt/	[dyʔ ^{24}]	‘demon’	/tɕỳː/	[dʑyː ^{14}]	‘hold’
/tø̀t/	[døʔ ^{24}]	‘sit	/tsøː/	[tsøː ^{51}]	‘cook’
/ʈʰù/	[ʈʰu ^{24}]	‘boat’	/múː/	[muː ^{51}]	‘bite!’
/ʈʰò/	[ʈʰo ^{24}]	‘wheat’	/móː/	[moː ^{44}]	‘old woman’

3.3 Underlying tones

Underlyingly, Walungge has two phonologically contrastive tones, high (H) and low (L). The presence of a boundary tone, and the formation of a tonal foot for the purposes of attaching the tones to the tone bearing units, gives the rising, falling, and level pitch melodies that are transcribed above in the phonetic forms of words (see Chapter 5 for tonal analysis).

The following are examples of monosyllabic and disyllabic words with an underlying H versus L tonal contrast.

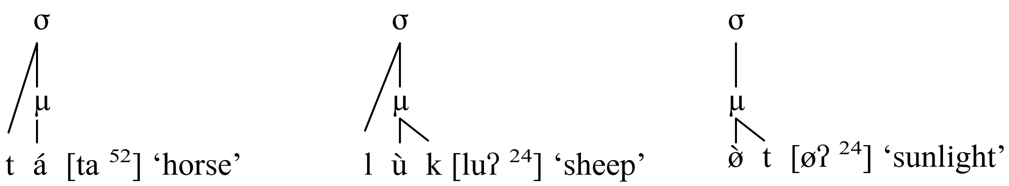
/ŋá/	[ŋa ^{52}]	‘drum’	/sáŋ/	[saŋ ^{44}]	‘pine’
/ŋà/	[ŋa ^{24}]	‘1SG’	/sàŋ/	[saŋ ^{14}]	‘screw’
/ɲíma/	[ɲima $^{44\ 44}$]	‘cereal head’	/sámpa/	[samba $^{44\ 44}$]	‘new’
/ɲìma/	[ɲima $^{11\ 44}$]	‘sun’	/sàmpa/	[samba $^{11\ 44}$]	‘bridge’

3.4 Syllable structure

3.4.1 Syllable patterns

The acoustic work of chapter 4 and the tonal analysis of chapter 5 show that syllable rhymes with a short vowel only and rhymes with a plosive coda have the same pitch melodies and can be analysed as monomoraic, whereas rhymes with either a long vowel or a short vowel plus a sonorant coda have the same pitch melodies and can be analysed as bimoraic.

Monosyllabic monomoraic words have been found with syllable onset but no coda, coda but no onset, or both coda and onset, e.g:



Apart from suffixes (e.g. /-i/ ‘POSS’, which are not of themselves phonological words (see 5.4.1)), no words have be found with a short vowel nucleus and neither onset nor coda.

Monosyllabic bimoraic words include those with a long vowel nucleus, and those with a short vowel nucleus plus sonorant coda. These may occur both with and without a syllable onset, e.g.

σ μ μ t í: [ti: ^{44}] ‘foal’ σ μ μ s á ŋ [saŋ ^{44}] ‘pine’ σ μ μ ú: [u: ^{44}] ‘breath’

Syllables with neither onset nor coda may occur as the first syllable of a polysyllabic word, e.g.

σ σ μ μ μ ó r a ŋ [oraŋ $^{11\ 344}$] ‘1PL’

As well as being the syllable nucleus of a monosyllabic word, a long vowel may occur in either the first or the second syllable of a polysyllabic word, or indeed in both syllables. The following are examples of words with long vowels.

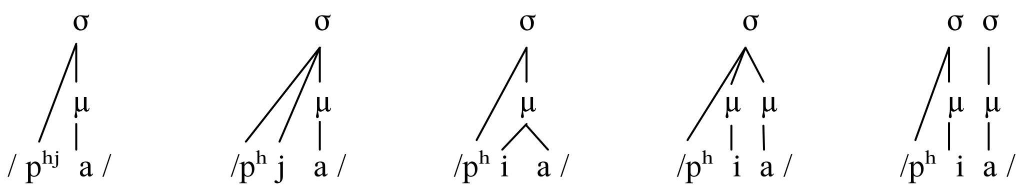
/k^há:/	[k^ha: ^{44}]	‘snow’
/ní:lam/	[ni:lam $^{44\ 44}$]	‘dream’
/$ʈ^h$òma:/	[$ʈ^h$oma: $^{11\ 344}$]	‘potato’
/ʈé:ma:/	[ʈe:ma: $^{44\ 44}$]	‘red monkey’

Consonant clusters are not permitted in the syllable coda, and there are restrictions on what consonants may occur in the coda. The consonants that are permitted are the plosives /p/, /t/, /k/, the nasals /m/, /n/, ŋ/, and the vibrant /r/. If the syllable has a coda, then the vowel nucleus is a short vowel.

All consonants are permitted as word initial syllable onsets. With the exception of aspirated plosives and affricates, and the consonants /l̥/, /r̥/ and /h/, all are also permitted in a word medial syllable onset. In the sections above there are examples of all consonants and the environments in which they occur.

3.4.2 Glides

Phonetically, Walungge has CGV (consonant, glide, vowel) and CVG glides involving [j] and [w], e.g. [p^hja ^{44}] ‘marmot’ and [daw ^{14}] ‘friend’. The question arises as to their underlying structure. By the very nature of [j] and [w], underlyingly they could be either consonants or vowels; if they are vowels they could be in the same syllable as the adjacent vowel, or a different syllable; and if the same syllable it could be monomoraic or bimoraic. The following diagrams illustrate various possibilities with the word [p^{h}ja ^{44}]:



The different CGV and CVG glides which are found in Walungge are discussed below.

glide [jV]

Examples of phonetic CGV glides involving [j] are:

[p^{h}ja ^{44}]	‘marmot’
[dju ^{14}]	‘bracelet’
[poːdja $^{44\ 41}$]	‘navel’

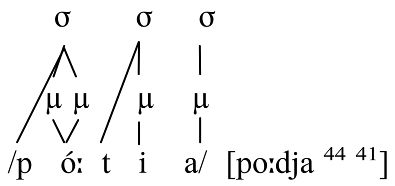
For a word such as [poːdja $^{44\ 41}$] ‘navel’, the underlying structure can be analysed by taking into account the realised pitch melody. A typical disyllabic word with underlying high tone has a surface pitch melody which is high across the word regardless of whether the second vowel is short or long e.g.

[ɲima $^{44\ 44}$]	‘cereal head’
[k^{h}adaː $^{44\ 44}$]	‘silk scarf’

A trisyllabic word with underlying high tone has a pitch melody which is high across the first two syllables of the word, then falling to low on the third syllable, e.g.

[orokpa $^{44\ 44\ 21}$]	‘toad’

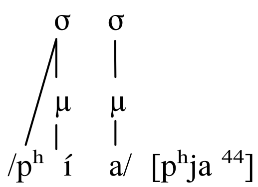
If a word such as [poːdja $^{44\ 41}$] ‘navel’ is underlyingly two syllables, then it would be expected to be realised with a high level pitch melody [44 44]. However, the realised pitch melody is in fact [44 41], suggesting that there are three syllables underlyingly, with the phonetic sequence [ja] being two underlying syllables. This would give the structure:



A high level pitch melody, as in [pʰja 44] 'marmot', occurs with both monosyllabic bimoraic words and disyllabic words, but not monosyllabic monomoraic words:

[ɕa 52]	'meat'
[ɕa: 44]	'deer'
[ɲima 44 44]	'cereal head'

This suggests that [pʰja 44] 'marmot' is bimoraic, but does not give any further indication as to whether the two moras are part of the same syllable or different syllables. If the two moras are in the same syllable, this would be a heavy rising diphthong /ia/. However, cross-linguistically rising diphthongs are typically light (Kehoe, Hilaire-Debove et al. (2008)), suggesting a disyllabic structure for [pʰja 44] 'marmot', which is consistent with analysing [po:dja 44 41] 'navel' as trisyllabic. Similarly the GV segment in a low tone word such as [dju 14] 'bracelet' will be analysed as two syllables. The following diagram illustrates the disyllabic structure:



Vowel plus glide [Vw] and [Vj]

The next examples are of glides [w] and [j] following a vowel:

[haj ^{44}] ‘3SG (far)’

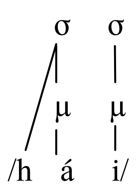
[daw ^{14}] ‘friend’

As with the glides discussed above, the pitch melodies [44] and [14] suggest the underlying structure is bimoraic, which means that structurally they could each be either one heavy syllable or two syllables. And if they are structurally monosyllabic, the segments [aj] and [aw] could be underlying heavy diphthongs /ai/ and /au/ or vowel plus sonorant coda /aj/ and /aw/.

The data which would be helpful to have would be data with [aj] or [aw] in the second phonetic syllable, e.g. CVCaj or CVCaw. If these have pitch melodies [44 41] or [11 41], that would suggest the segments [aj] and [aw] are structurally two syllables, whereas pitch melodies [44 44] or [11 44] would suggest one syllable. However, to date the only data found with such glides is when the glide is the result of an affix, e.g.

tóptçe -i → [toptçej $^{44\ 41}$] ‘food.POSS’

In these instances the VG segment is structurally two syllables. Although the disyllabic structure of the VG segment in [toptçej $^{44\ 41}$] does not necessarily imply a disyllabic structure of the VG segment in [haj ^{44}], in the absence of any other data, all VG segments will be treated as disyllabic, e.g.



[haj ^{44}]

Segments [wV]

The CGV glide [wV] presents a more complex situation than the glides discussed above. The following are examples of words with [wV] following a word initial consonant:

[bwaʔ ^{24}]	‘bubble’

[k^{h}wa ^{52}]	‘curry soup’

[t^{h}wa ^{52}]	‘hammer’

[k^{h}woŋ ^{44}]	‘3.SG’ (hon.)

Comparing these with [jV] in a word such as [p^{h}ja ^{44}] ‘marmot’ there are several differences. Firstly, in the above data it can be seen that it is possible to have a falling [52] pitch melody across [wV], whereas the only high tone pitch melody across [jV] is [44]. Elsewhere in the language, a falling [52] melody is associated with a monomoraic word and a level [44] melody is associated with a bimoraic word, e.g.

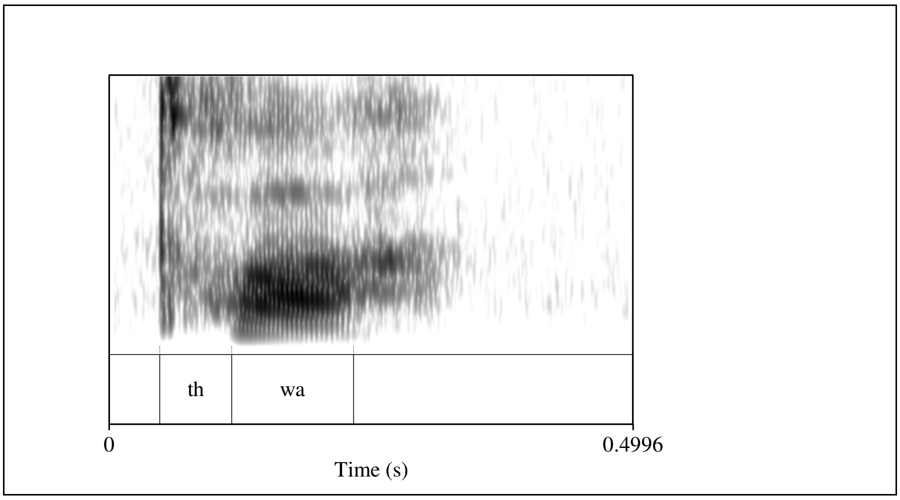
[ça ^{52}]	‘meat’

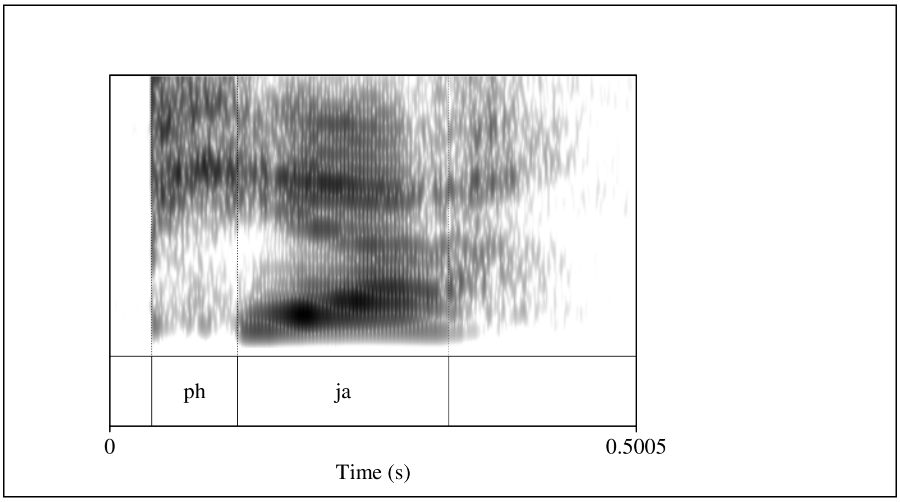
[çaː ^{44}]	‘deer’

This difference suggests that a word such as [t^{h}wa ^{52}] ‘hammer’ is monomoraic whereas a word such as [p^{h}ja ^{44}] ‘marmot’ is bimoraic.

Secondly, there is a difference in duration between [wV] and [jV]. Below are spectrograms which illustrate this difference.

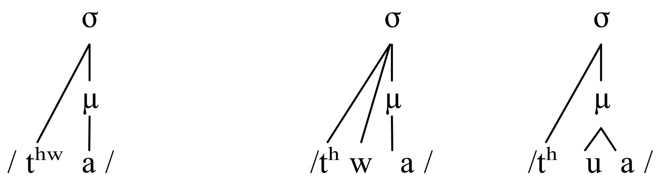
Figure 3-1: spectrograms comparing [wa] and [ja]





In both words there is a period of voicelessness at the end of the vowel [a]. The segmentation of [ja]/[wa] is from the onset of voicing following the release of the consonant until the cessation of voicing. The duration of this portion in the above spectrograms is 116ms for [wa] and 198ms for [ja]. This is a further indication that [wa] is monomoraic and [ja] is bimoraic.

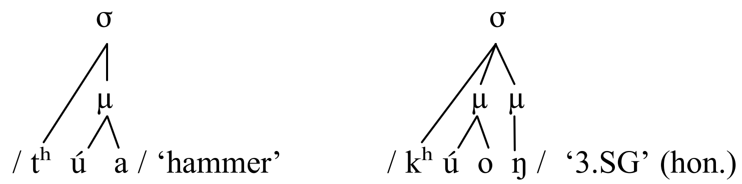
If a word such as [t^{h}wa ^{52}] 'hammer' is underlyingly monomoraic, there are three possibilities for underlying structure:



The possibility of a single phoneme /t^{hw}/ is ruled out because of the number of different consonants that can occur with the glide [w], e.g. [bwaʔ ^{24}] 'bubble', [lwa ^{52}] 'lung', [kwa ^{52}] 'leather', [gwa ^{24}] 'egg', [nwi ^{24}] '3SG.POSS', [$\underline{t}^{h}$wa ^{24}] 'taste.

The next possibility to be considered is that the phonetic sequence [CwV] is underlyingly /CwV/, i.e. a branching onset. The problem with this analysis is that there is no evidence for branching onsets. Apart from the glides already mentioned, there are no consonant clusters word initially.

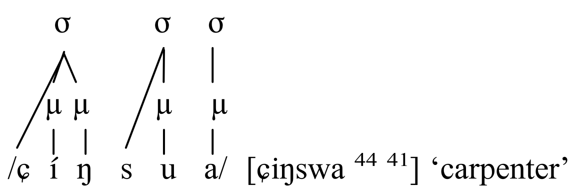
In the light of the above, it might appear that the [wV] in words such as [t^{h}wa ^{52}] 'hammer' should be analysed as a light diphthong /uV/. Further, if light diphthongs occur in the language, they might be expected to occur in both heavy and light syllables. Indeed, this appears to be the case, e.g.



However, there are occurrences of [wV] in the second phonetic syllable of a word, e.g.

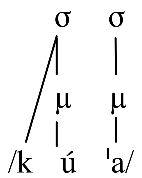
[ɕiŋswa $^{44\ 41}$]	'carpenter'
[ts^{h}émbwa $^{44\ 41}$]	'tailor'

These pattern with [jV]. Unlike the above examples where there is a duration difference between [wV] and [jV] in words such as [t^{h}wa ^{52}] 'hammer' and [p^{h}ja ^{44}] 'marmot', in words such as [ɕiŋswa $^{44\ 41}$] 'carpenter' and [poːdja $^{44\ 41}$] 'navel' there is no such duration difference. In addition to this, both have the same [44 41] pitch melody. If [wa] in [ɕiŋswa $^{44\ 41}$] was monosyllabic, then a [44 44] pitch melody would be expected. The [44 41] pitch melody suggests that underlyingly [ɕiŋswa $^{44\ 41}$] is three syllables, which is the same as [poːdja $^{44\ 41}$]. This would give an underlying structure:

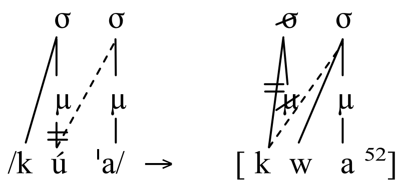


Whilst analysing [wa] in words such as [kwa ^{52}] 'leather' as a light diphthong and analysing [wa] in words such as [ɕiŋswa $^{44\ 41}$] 'carpenter' as two separate syllables may account for the data, it is problematic on several accounts. Firstly, if the language does indeed have light diphthongs /uV/ and also sequences /uV/ as separate syllables, it might be expected that both these structures would occur in both the first and the second syllable of a word. Yet the only occurrences of light diphthongs are in monosyllabic words, and the only occurrences of /uV/ as separate syllables are with the vowel /u/ in the second syllable of polysyllabic words. In addition, if light diphthongs /uV/ exist in the language, light diphthongs /iV/ might be expected as well. Yet words which could potentially have been analysed as having this structure (e.g. [pja ^{44}]) have in fact been analysed as being disyllabic due to the [44] pitch melody. Thus an alternative analysis will be proposed that seeks to address these problems.

It is proposed that [wV] as in [kwa ^{52}] 'leather' and [wV] as in [çiŋswa $^{44\ 41}$] 'carpenter' both have the same underlying structure of being two syllables. The difference in the surface realisation can be attributed to stress, which is typically on the second syllable (see section 5.2 below). If a word such as [kwa ^{52}] 'leather' has underlying form



its surface realisation [kwa ^{52}] can be explained by saying that a high vowel /u/ in an unstressed syllable becomes a glide:



This glide formation process must occur before the assignment of tone to TBUs in order to account for the monomoraic pitch melody [52]. On the other hand, whilst the vowel /u/ in /çíŋsua/ is phonetically realised as [w], structurally there is no glide formation because it is the nucleus of the stressed second syllable.

This analysis is not without problems; the question that this analysis raises is why /kúa/ is realised as [kwa ^{52}] whereas /pía/ is realised as [pja ^{44}]. If an unstressed /u/ in a /uV/ sequence undergoes glide formation, then it might be expected that an unstressed /i/ would also undergo glide formation. Yet the surface realisation [pja ^{44}] suggests that /i/ does not become a glide. Further research is needed in order to fully resolve the question of the underlying structure of glides. For this dissertation the analysis that will be adopted is that all occurrences of glide plus vowel are underlyingly two syllables.

3.5 Diachronic phonology

3.5.1 Introduction to Written Tibetan

The following introduction to Written Tibetan is taken from Beyer (1992), Denwood (1999), and Tournadre & Dorje (2003). The Tibetan alphabet was developed in the 7th century. Following this, there was a period of time when the writing system underwent various standardizations, but from approximately the 11th century until the present day, Written Tibetan has remained virtually unchanged. It is generally assumed that the writing reflects the phonology of the language at the time it was written down (but see Hill (2010) for an in-depth discussion of the relationship between Old Tibetan phonology and Written Tibetan, including arguments suggesting that in places of the alphabet there is not a one to one mapping between the letters of Written Tibetan and the phonemes of Old Tibetan).

Written Tibetan (WT) has thirty letters, each of which has an inherent ‘a’ vowel sound. In addition to this, there are also four (Tournadre & Dorje (2003)) or five (Hill (2010)) vowel diacritics. Wylie (1959) proposed a standard system for transcribing WT in Roman script. The following table gives the thirty WT letters, listed in their traditional order, along with the Wylie transcription, and IPA equivalents of that transcription.

Table 3-3: WT consonants with Wylie transcription and IPA

ཀ	ka	[ka]	ཁ	kha	[k^{h}a]	ག	ga	[ga]	ང	nga	[ŋa]
ཅ	ca	[tɕa]	ཆ	cha	[tɕ^{h}a]	ཇ	ja	[dʑa]	ཉ	nya	[ɲa]
ཏ	ta	[ta]	ཐ	tha	[t^{h}a]	ད	da	[da]	ན	na	[na]
པ	pa	[pa]	ཕ	pha	[p^{h}a]	བ	ba	[ba]	མ	ma	[ma]
ཙ	tsa	[tsa]	ཚ	tsha	[ts^{h}a]	ཛ	dza	[dza]	ཝ	wa	[wa] [1]
ཞ	zha	[ʑa]	ཟ	za	[za]	འ	'a	[a/h/ w/ɣ][2]	ཡ	ya	[ja]
ར	ra	[ɾa]	ལ	la	[la]	ཤ	sha	[ɕa]	ས	sa	[sa]
ཧ	ha	[ha]	ཨ	a	[a]						

[1]This did not exist as an independent letter in Old Tibetan (Hill (2010)).

[2]There has been debate over the phonetic value of this letter (Hill (2005)).

The four diacritic vowels below comprise the letter <a> plus a diacritic either above or below the letter, as follows:

ཨི	i	ཨུ	u	ཨེ	e	ཨོ	o

Syllable boundaries are marked by a dot.

The structure of the WT syllable is diagrammatically represented as follows:

	5			
	2			
1	3	6	7	8
	4			
	5			

1: pre-scripted letter
2: superscripted letter
3. radical (base letter) - obligatory
4. subscripted letter
5. vowel diacritic
6. final letter
7. additional final letter
8. syllable marker - obligatory

The numeric order above is the order of the letters when transcribing them completely linearly as in the Wylie transcription.

The smallest WT syllable consists of a single consonant radical (with an inherent vowel ‘a’) plus syllable marker, for example ས་ <sa> ‘earth’. There may be consonants pre-scripted before, superscripted above, or subscripted below the radical. For any vowel other than [a], there is a vowel diacritic. The syllable coda may have up to two consonants. An example of a syllable containing all these elements is བསྒྲིབས་ <bsgribs> ‘arrange.PST’.

As is to be expected, Walungge has some lexical items which are the same as in WT or have a clear parallel in WT, and other words which are possibly unique to Walungge. For the words that are common to Walungge and WT, what follows is an outline of how Walungge corresponds to WT. It should be noted that these are only generalisations, for which there are exceptions.

3.5.2 Syllable onsets word initially

3.5.2.1 WT Voiceless unaspirated plosive and affricate radicals

A WT voiceless unaspirated plosive or affricate radical remains voiceless unaspirated in Walungge and corresponds to high tone. All WT pre-scripted and superscripted consonants have disappeared with no obvious influence. The exception to this is a superscripted <l>, which is discussed below.

The following table gives WT segments comprising of a plosive or affricate radical plus pre-scripted or superscripted consonants, along with their Walungge equivalent and examples of each. The WT is in Wylie transliteration.

Table 3-4: WT voiceless unaspirated radicals with Walungge correspondences

WT	Walungge		Examples		
	phoneme	tone			
<k>, <pre-script or superscript + k>	> /k/	H	<ka ba>	/káː/	‘column’
			<skar ma>	/káma/	‘star’
			<dkar po>	/káppo/	‘white’
<c>, <pre-script or superscript + c>	> /tɕ/	H	<ca lag>	/tɕálak/	‘things’
			<gcig>	/tɕík/	‘one’
			<bcu>	/tɕú/	‘ten’
<t>, <pre-script or superscript + t>	> /t/	H	<tog tse>	/tókse/	‘hoe’
			<sta re>	/tári/	‘axe’
			<rta>	/tá/	‘horse’
<p>, <pre-script or superscript + p>	> /p/	H	<pus mo>	/píːmo/	‘knee’
			<dpung pa>	/púŋpa/	‘shoulder’
			<spun nye>	/pýn/	‘relative’
<ts>, <pre-script or superscript + ts>	> /ts/	H	<tsag sgra>	/tsákʈa/	‘matches’
			<gtsang chu>	/tsáŋpu/	‘river’
			<btsos>	/tsǿː/	‘cook’

WT plosives and affricates with a superscripted <l> sometimes equate to high tone and sometimes to low tone.

Table 3-5 WT voiceless unaspirated radicals with superscripted <l>

WT	Walungge		Examples		
<lk>	> /k/	H	<lkugs pa>	/kúkpa/	'mute'
<lc>	> /tɕ/	L	<lci ba>	/tɕàː/	'dung'
	> /tɕ/	H	<lce>	/tɕéle/	'tongue'
<lt>	> /t/	L	<lte mig>	/tìmik/	'key'
	> /t/	H	<lto>	/tó/	'food'
<lp>	> /p/	H	<lpags pa>	/pákpa/	'skin'

The combination of a plosive or affricate with a subscript is discussed in section 3.5.2.6 below.

3.5.2.2 WT aspirated plosive and affricate radicals

Aspirated plosive and affricate radicals in WT have remained phonologically voiceless in Walungge, and word initially they have remained aspirated and correspond to high tone. Pre-scripted and superscripted consonants have disappeared without apparent influence.

Table 3-6: Aspirated word initial radicals with Walungge correspondences

WT	Walungge		Examples		
<kh>, <pre-scripted consonant + kh>	> /k^{h}/	H	<kha ba> <mkhan pa> <'kher>	/k^{h}á:/ /k^{h}é:ma/ /k^{h}úr/	‘snow’ ‘fern’ ‘carry’
<ch>, <pre-scripted consonant + ch>	>/tɕ^{h}/	H	<chang> <mchongs> <'cham>	/tɕ^{h}áŋ/ /tɕ^{h}óŋ/ /tɕ^{h}ám/	‘beer’ ‘jump’ ‘religious dance’
<th>, <pre-scripted consonant + th>	> /t^{h}/	H	<thag pa> <mthe po> <'thams>	/t^{h}ákpa/ /t^{h}éwo/ /t^{h}ám/	‘rope’ ‘thumb’ ‘hug’
<ph>, <pre-scripted consonant + ph>	> /p^{h}/	H	<phag pa> <'phur>	/p^{h}ákpa/ /p^{h}úr/	‘pig’ ‘fly’ (v.)
<tsh>, <pre-scripted consonant + tsh>	> /ts^{h}/	H	<tshi lu> <mtsho> <'tsher ma>	/ts^{h}ílu/ /ts^{h}ó/ /ts^{h}émak/	‘fat’ ‘lake’ ‘thorn’

Aspirated radicals with subscripts are discussed in section 3.5.2.6 below.

3.5.2.3 WT Voiced plosive and affricate radicals

A word initial voiced plosive or affricate radical without any pre-scripting or superscripting has become an aspirated phoneme and low tone.

Table 3-7: WT voiced radicals without pre-script or superscript

WT	Walungge		Examples		
<g>	> /k^{h}/	L	<gur>	/k^{h}ùr/	‘tent’
<j>	> /tɕ^{h}/	L	<ja>	/tɕ^{h}à/	‘tea’
<d>	> /t^{h}/	L	<dom>	/t^{h}òm/	‘bear’
<b>	> /p^{h}/	L	<ba>	/p^{h}à/	‘cow’

Word initially, a voiced plosive with pre-scripting or superscripting has become an unaspirated plosive with low tone.

Table 3-8: Voiced radicals with pre-script or superscript

WT	Walungge	Examples		
<pre-script or superscript + g>	> /k/ L	<dgun ga>	/kỳnto/	'winter'
		<sgam>	/kàm/	'trunk'
		<mgo>	/kòkki/	'head'
<pre-script or superscript + j>	>/tɕ/ L	<rje po>	/tɕèwo/	'beautiful'
		<ljang khu>	/tɕàŋku/	'green'
		<'ja>	/tɕà/	'rainbow'
<pre-script or superscript + d>	> /t/ L	<sdon po>	/tòŋpo/	'tree'
		<rdo>	/tòri/	'stone'
		<mdangs>	/tàŋ/	'yesterday'
<pre-script or superscript + b>	> /p/ L	<sbal pa>	/pàwak/	'frog'
		<lbu ba>	/pùak/	'bubble'
		<'bu>	/pù/	'insect'
<pre-script or superscript + dz>	> /ts/ L	<rdz ma>	/tsàma/	'pottery'
		<mdzo>	/tsò/	'yak cross'
		<'dzegs>	/tsàk/	'climb'

3.5.2.4 WT fricative radicals

Voiceless fricative radicals in WT correspond to voiceless fricatives along with high tone in Walungge. All pre-scripts and superscripts have disappeared without influence. The exception to this is the WT segment <lh>, which is the phoneme /l̥/ in Walungge.

Table 3-9: Voiceless WT fricative radicals

WT	Walungge	Examples		
<s>, <pre-script + s>	> /s/ H	<sa>	/sá/	'earth'
		<gsum>	/súm/	'three'
<sh>, <pre-script + sh>	> /ɕ/ H	<sha>	/ɕá/	'meat'
		<gshog pa>	/ɕókpa/	'wing'
<h>	> /h/ H	<ha go>	/há k^hò/	'understand'
<lh>	> /l̥/ H	<lhan pa>	/l̥énta/	'patch'

WT voiced fricatives have become voiceless fricatives in Walungge. Word initially they correspond to low tone. All pre-scripts and superscripts have disappeared without apparent influence.

Table 3-10: Voiced fricative radicals

WT	Walungge	Examples		
<z>, <pre-script + z>	> /s/ L	<zo ra>	/sòra/	'sickle'
		<bza ba>	/sàtɕa/	'food'
		<gzig>	/sìk/	'snow leopard'
<zh>, <pre-script + zh>	> /ɕ/ L	<zhing>	/ɕìŋka/	'field'
		<bzhi>	/ɕì/	'four'

3.5.2.5 WT sonorant radicals

WT sonorant radicals <m>, <n>, <ny>, <ng>, <l>, <r> and <y> have remained sonorants in Walungge. The WT radicals without pre-scripting or suffixing word initially have given rise to low tone in Walungge. Any WT pre-script or suffix has been lost and word initially it has given rise to high tone in Walungge.

WT	Walungge		Examples		
<ng>	> /ŋ/	L	<nga>	/ŋà/	‘I’
<pre-script or superscript + ng>	> /ŋ/	H	<sngon po> <rnga> <lnga>	/ŋóno/ /ŋá/ /ŋá/	‘blue’ ‘drum’ ‘five’
<ny>	> /ɲ/	L	<nya>	/ɲà/	‘fish’
<pre-script or superscript + ny>	> /ɲ/	H	<gnyis> <snying > <rnying pa>	/ɲíː/ /ɲíŋ/ /ɲíŋpa/	‘two’ ‘heart’ ‘old’
<n>	> /n/	L	<nag po>	/nàppo/	‘black’
<pre-script or superscript + n>	> /n/	H	<gnam> <sna khug> <snum>	/nám/ /nákkik/ /núm/	‘sky’ ‘nose’ ‘oil’
<m>	> /m/	L	<mar>	/màː/	‘butter’
<pre-script or superscript + m>	> /m/	H	<dmar po> <rmar tsa> <sman>	/máppo/ /máttsi/ /mén/	‘red’ ‘chilli’ ‘medicine’
<l>	> /l/	L	<la>	/là/	‘mountain pass’
<r>	> /r/	L	<ra>	/rà/	‘goat’
<w>	> /w/	L	<wa>	/wà/	‘fox’
<y>	> /j/	L	<yi ge>	/jìː/	‘letter’
<gy>	> /j/	H	<gyag>	/ják/	‘yak’

It can be noted from the above table that there are no instances of the radicals <r> and <l> with a pre-script or superscript. Whilst <l> does not occur as a radical with pre-scripting or superscripting, it occurs as a subscript with a number of radicals. This is what has give rise to Walungge /l/ with high tone. Similarly <r> occurs as a subscript with a number of radicals. Subscript <l> and <r> are discussed in the section below.

3.5.2.6 WT subscripts

Written Tibetan has four subscripts <w>, <y>, <l>, and <r>.

Subscript <w>

The subscript <w> has disappeared without any influence on either the radical it was combined with, or the tone or the word. The WT radical (plus any pre-scripts or superscripts) follows the correspondences with Walungge which are outlined above. The following are examples of subscripted <w>:

Table 3-11: Examples of subscript <w>

WT	Walungge		Examples		
<tsw>	> /ts/	H	<rtswa>	/tsá/	‘grass’
<tshw>	>/ts^{h}/	H	<rtshwa>	/ts^{h}á/	‘salt’
<zhw>	> /ɕ/	L	<zhwa mo>	/ɕàmo/	‘hat’
<rw>	> /r/	L	<rwa co>	/ràco/	‘horn’

Subscript <y>

A subscripted <y> with a velar plosive radical has given rise to a palatal plosive. The other aspects of the resulting phoneme along with the tone of the word are determined by the radical as described in the sections on WT radicals above.

Table 3-12: Subscripted <y> with velar plosive radical

WT	Walungge		Examples		
<pre-script or superscript + ky>	> /c/	H	<skyog> <skyur mo>	/cóː/ /cýppo/	‘ladle’ ‘sour’
<khy>,	>/c^{h}/	H	<khyo ga> <khyags pa>	/c^{h}ówa/ /c^{h}áwak/	‘husband’ ‘ice’
<gy>	> /c^{h}/	L	<gyong>	/c^{h}ỳn/	‘loss’
<pre-script or superscript + gy>	> /c/	L	<rgya> <brgyad>	/cà/ /cèt/	‘beard’ ‘eight’

For subscripted <y> with an initial bilabial plosive radical, a variety of different patterns exist. Sometimes the <y> has deleted, leaving the bilabial plosive. Sometimes the combination of bilabial plosive plus subscript <y> has resulted in an alveolo-palatal affricate. Other times the radical and pre-scripted consonants have deleted, leaving the approximant <y> as the phoneme /j/. The following are examples of subscripted <y> following a bilabial plosive radical.

Table 3-13: Subscripted <y> with bilabial plosive radical

<dpy>, <spy>	> /tɕ/	H	<spyan>	/tɕén/	'eye (hon)'
<phy>	> /p^h/	H	<phyi ba>	/p^{h}íja/	'marmot'
			<phyis>	/p^{h}íː/	'blow nose'
	> /tɕ^h/	H	<phye ma leb>	/tɕ^{h}émalatɕok/	'butterfly'
			<phyur ba>	/tɕ^{h}úppe/	'cheese'
<by>	> /p^h/	L	<byi ba >	/p^{h}ìtsik/	'mouse'
	> /tɕ^h/	L	<byi 'u>	/tɕ^{h}ìpʈuk/	'chick'
			<bye ma>	/tɕ^{h}èːma/	'sand'
<dby>, <sby>	> /j/	H	<dbyar ga>	/jákka/	'summer'

Subscript <l>

A subscripted <l> has become the phoneme /l/ in Walungge, with radicals having deleted and given rise to high tone. The exception to this is the combination <zl> which has become /t/ with low tone.

Table 3-14: Subscript <l>

WT	Walungge		Examples		
<radical with subscript l>	> /l/	H	<bla ma>	/láma/	'monk'
			<rlon pa>	/lǿŋkoŋ/	'wet'
			<slob grwa>	/lópʈa/	'school'
<zl>	> /t/	L	<zla ba>	/tàwa/	'month'

Subscript <r>

In general, the combination of a plosive radical plus subscripted <r> has resulted in a retroflex plosive phoneme. The tone of the word and any aspiration of the resulting phoneme determined by the radical.

Table 3-15: Plosive radical plus subscript <r>

WT	Walungge	Examples		
<pre-script or superscript + kr>, <superscript + pr>	> /ʈ/ H	<bkrus> <skra> <spre >	/ʈýː/ /ʈá/ /ʈéːmaː/	‘wash’ ‘hair’ ‘monkey’
<khr>, <pre-script + khr>, <phr>, <pre-script + phr>,	> /ʈʰ/ H	<'phreng ba> <phru gu> <khron pa>	/ʈʰáŋa/ /ʈʰúku/ /ʈʰóŋka/	‘rosary’ ‘offspring’ ‘well (n)’
<gr>, <dr>	> /ʈʰ/ L	<grog ma> <drug>	/ʈʰòmakpa/ /ʈʰùk/	‘ant’ ‘six’
<pre-script or superscript +br>, <pre-script or superscript + gr>	> /ʈ/ L	<'bri> <sgra snyan>	/ʈì/ /ʈàmɲen/	‘female yak’ ‘Tibetan lute’

The combination <sr> has resulted in the phoneme /r̥/ plus high tone.

Table 3-16: Fricative radical plus subscript <r>

WT	Walungge	Examples		
<sr>	> /r̥/ H	<sran ma> <srab>	/r̥éna/ /r̥áp/	‘bean’ ‘horse’s bit’

3.5.2.7 The radical < ' >

The WT radical <'> has sometimes become the phoneme /h/ in Walungge. Other times it has deleted, but has given rise to low tone. The following are examples of this.

Table 3-17: the radical <'>

WT	Walungge	Examples		
<'>	> /h/ L	<'o ma>	/hòma/	‘milk’
		<'o cag>	/hòtɕaŋ/	‘we’
		<'ug pa>	/hùla/	‘owl’
	> Ø L	<'od>	/ø̀t/	‘sunlight’
		<'og>	/òk/	‘under’

3.5.3 Syllable onsets word medially

Word medial segments do not contribute to the tone of the word. In general, the following has happened to word medial syllable onsets:

- Pre-scripts and superscripts have disappeared without influence. The subscripts <r> and <y> have given rise to retroflex and palatal phonemes as described in section 3.5.2.6 above, but otherwise they have disappeared without influence.
- Fricative radicals have all become voiceless phonemes in the same way as they have word initially.
- Sonorant radicals have remained unchanged.
- Plosive and affricate radicals have generally become the equivalent voiceless unaspirated plosive phonemes, as illustrated by the following examples.

Table 3-18: word medial plosive and affricate radicals

WT radical	Walungge	Examples		
<k>, <kh>, <g>	> /k/	<las ka>	/lè:ka/	‘work’
		<lam khag>	/làŋkak/	‘path’
		<rgya gar>	/càka:/	‘India’
<c>, <ch>, <j>	> /tɕ/	<sgo lcags>	/kø̀:tɕaː/	‘lock’
		<am chi>	/ámtɕi/	‘doctor’
		<theb 'ju>	/t^{h}íptɕu/	‘button’
<t>, <th>, <d>	> /t/	<kha btags>	/k^{h}átaː/	‘silk scarf’
		<gur thag>	/k^{h}ùttak/	‘guy rope’
		<chod sdong>	/tɕ^{h}ýttøn/	‘shrine’
<p>, <b>	> /p/	<lag pa>	/làkpa/	‘hand’
		<phyur ba>	/tɕ^{h}úppe/	‘cheese’
<ts>, <tsh>	> /ts/	<jem tse>	/ts^{h}èmtsi/	‘scissors’
		<ja tshags>	/tɕ^{h}àtsaː/	‘tea strainer’

Both <ph> and <dz> should fit with the above patterns, but no examples were found.

There are a number of regular exceptions to the generalisation that plosive radicals all become the equivalent voiceless phoneme.

Intervocalic bilabial radicals (including those with subscripts) have generally become the phoneme /w/. This is illustrated by the following examples:

Table 3-19: intervocalic (including subscript) bilabial radicals

WT radical	Walungge	Examples		
<p>, <ph>, <b>	> /w/	<bra phye>	/ʈ^{h}àwe/	‘buckwheat’
		<mthe po>	/t^{h}éwo/	‘thumb’
		<zla ba>	/tàwa/	‘month’

Intervocalic velar radicals have sometimes become the phoneme /w/, for example <khyo ga> /cʰówa/ ‘husband’. Other times, the syllable starting with an intervocalic velar radical has deleted altogether and given rise to compensatory lengthening, for example <yi ge> /jìː/ ‘letter’. Intervocalic bilabial radicals have sometimes become a nasal plus alveolar plosive, for example <khe po> /kʰénte/ ‘cheap’.

When medial bilabial and velar plosive radicals occur following either <d> or <n>, generally it has been the onset plosive assimilating to the alveolar place of articulation rather than the coda consonant assimilating to the onset plosive’s place of articulation. This is illustrated by the following examples.

Table 3-20: Examples of alveolar place of articulation assimilation

WT	Walungge	Examples		
bilabial radical	> /t/	<pad pa>	/pétta/	‘leech’
		<nggon pa>	/kònta/	‘monastery’
		<mchin pa>	/tɕʰínta/	‘liver’
velar radical	> /t/	<grod khog>	/ʈʰòtta/	‘stomach’
		<dgun ga>	/kỳnto/	‘winter’

3.5.4 Word final syllable coda

Of the word final syllable codas found in WT, the ones that remain in Walungge are the plosives <b>, <d>, <g>, the nasals <m>, <n>, <ng>, and sometimes the vibrant <r>. The coronal codas <d> and <n> have caused the vowels <u>, <o> and <a> to become the phonemes /y/, /ø/, and /e/.

The following are examples of word final syllable codas with plosives and nasals. The WT plosives have become voiceless phonemes in Walungge.

Table 3-21: Word final syllable codas with nasals and plosives

WT	Walungge	Examples		
<g>	> /k/	<gzig>	/sìk/	‘snow leopard’
<ng>	> /ŋ/	<shing>	/çíŋ/	‘firewood’
<d>	> /t/	<bdud>	/tỳt/	‘demon’
<n>	> /n/	<sman>	/mén/	‘medicine’
<b>	> /p/	<khab>	/kʰáp/	‘needle’
<m>	> /m/	<dom>	/tʰòm/	‘bear’

Word final <l> has deleted, giving rise to compensatory lengthening on the preceding vowel. Word final <s> has also deleted. In cases where <s> is the sole consonant word finally, the deletion of <s> has given rise to compensatory vowel lengthening, and caused the vowels <u>, <o> and <a> to become the phonemes /y/, /ø/ and /e/. Similarly, <gs> has deleted giving compensatory vowel lengthening but not a change in the vowel phoneme. However for <ms>, <ngs> and <bs> only <s> has deleted and there is no vowel lengthening. Word final <r> has been deleted in some instances and been retained in others. In the instances where it has been deleted, there has been compensatory lengthening on the vowel. There appears to be no way of predicting which <r>'s have been deleted and which have been retained. Word final <'> has deleted without any apparent effect. The following examples illustrate <l>, <s>, <'> and <r> word finally:

Table 3-22: Word final <l>, <s>, <r>

WT	Walungge	Examples		
<l>	deletion; vowel length	<bkol>	/kóː/	‘boil’
		<bal>	/pʰàː/	‘wool’
<s>	deletion; vowel length	<yos>	/jøːː/	‘popped barley’
		<'bras>	/ʈèː/	‘rice’
	deletion following <m>, <ng> or <b>	<mdangs>	/tàŋ/	‘yesterday’
		<btsems>	/tsém/	‘sew’
<gs>	deletion; vowel length	<ja tshags>	/tɕʰàtsaː/	‘tea strainer’
		<lcags>	/tɕáː/	‘iron’
<r>	> /r/	<gur>	/kʰùr/	‘tent’
		<par>	/pár/	‘picture’
	> /Vː/	<mar>	/màː/	‘butter’
		<yar>	/jàː/	‘up’
<'>	deletion	<mda'>	/tà/	‘bow’ (‘arrow’ WT)

3.5.5 Word medial syllable coda

Some word medial syllable codas in WT have a variety of correspondences in Walungge. Plosives have generally become voiceless plosive phonemes; however there are also instances of deletion with compensatory vowel lengthening, assimilation to the place of articulation of the following phoneme, or coalescence with the following phoneme. The following examples illustrate these correlations:

Table 3-23: Word medial plosive syllable codas

WT	Walungge	Phoneme		
<g>	> /k/	<phug pa>	/pʰúkpa/	‘cave’
	assimilation to place of following phoneme	<nag po>	/nàppo/	‘black’
	deletion	<grog ma>	/ʈʰòmakpa/	‘ant’
	deletion plus vowel length	<sog pa>	/sóːwa/	‘shoulder blade’
<d>	> /t/	<pad pa>	/pétta/	‘leech’
	assimilation to place of following phoneme	<phyed ka>	/tɕʰékka/	‘half’
	deletion and vowel length	<gnyid lam>	/níːlam/	‘dream’
<b>	> /p/	<theb 'ju>	/tʰíptɕu/	‘button’
	coalescence and vowel length	<grib nag>	/ʈʰìːmak/	‘shadow’

Nasals have either remained unchanged, assimilated to the following place of articulation, or deleted. <ng> is the only nasal which appears to be unchanged in most, if not all instances. Both <m> and <n> have assimilated to a velar place of articulation before a velar plosive radical, but as a general rule <m> has not assimilated to alveolar before an alveolar plosive radical, and neither has <n> assimilated to bilabial before a bilabial plosive radical. Sometimes deletion of a nasal has occurred before another nasal, but not consistently. The following are examples of WT word medial nasal syllable codas.

Table 3-24: Word medial WT nasal syllable codas

WT	Walungge	Examples		
<ng>	> /ŋ/	<rkang pa>	/káŋpa/	'leg'
<m>	> /m/	<jem tse>	/tsʰèmtsi/	'scissors'
	place assimilation	<lam khag>	/làŋkak/	'path'
<n>	> /n/	<sen mo>	/séno/	'fingernail'
	deletion	<nyin ma>	/ɲìma/	'day'
	place assimilation	<pin khab>	/píŋkap/	'safety pin'

A word medial <l> in a syllable coda has deleted, leaving compensatory vowel lengthening. Before a plosive, <r> has assimilated in place and manner of articulation, becoming a voiceless plosive phoneme. Before an affricate, <r> has become the phoneme /t/. Otherwise <r> has deleted, but without any compensatory lengthening of the vowel. <s> without any other consonant in the syllable coda has deleted leaving compensatory length. <s> following <b>, <ng>, or <m> has deleted without any influence. The combination <gs> has deleted, sometimes leaving compensatory length and sometimes not. The following are examples of word medial syllable codas with <r>, <l> and <s>.

Table 3-25: <r>, <l>, <s> in word medial syllable coda

WT	Walungge	Examples		
<r>	assimilation to a voiceless plosive	<phyur ba>	/tɕʰúppe/	'cheese'
		<mar tsa>	/máttsi/	'chilli'
	deletion	<skar ma>	/káma/	'star'
<l>	deletion; vowel length	<sbal pa>	/pàːwak/	'frog'
<s>	deletion; vowel length	<pus mo>	/píːmo/	'knee'
		<rgyags pa>	/càːwo/	'fat'
	deletion; no length	<khyags pa>	/cʰéwak/	'ice'
		<rlangs pa>	/láŋpa/	'steam'

3.5.6 Vowels

Compensatory lengthening on vowels which has resulted from the deletion of syllable coda consonants has already been described above. The other major change which has happened to vowels is the change of vowel position from back vowels to front vowels.

In general the vowels <a>, <o>, and <u> have become the front vowels <e>, <ø> and <y> respectively, when followed by the WT coronal syllable codas <d>, <s> and <n> (but not <l> or <r>). This holds without exception for monosyllabic words. There are, however, disyllabic exceptions to this, for example <dgon pa> /kònta/ 'monastery'.

The vowel <e> has generally become <a> when followed by a velar syllable coda.

Whilst <i> has not undergone any regular change, when it is followed by <d> or <n> these consonants have changed to /k/ and /ŋ/.

The following examples illustrate the changes described above:

Table 3-26: changes for WT vowels and following consonants

WT	Walungge	Examples		
<i>	> /i/	<gnyid>	/ɲík/	‘sleep (n)’
		<gcin>	/tɕíŋ/	‘urine’
<e>	> /a/	<'phreng ba>	/ʈʰáŋa/	‘rosary’
		<dreg pa>	/ʈʰàkpa/	‘grime’
<a>	> /a/	<mar>	/màː/	‘butter’
		<bal>	/pʰàː/	‘wool’
	> /e/	<brgyad>	/cèt/	‘eight’
		<'bras>	/ʈèː/	‘rice’
		<sman>	/mén/	‘medicine’
<o>	> /o/	<bkol>	/kóː/	‘boil’
	> /ø/	<'od>	/ø̀t/	‘sunlight’
		<yos>	/jø̀ː/	‘popped barley’
		<tshon>	/tsʰǿn/	‘paint’
<u>	> /u/	<sbrul>	/ʈùː/	‘snake’
		<gur>	/kʰùr/	‘tent’
	> /y/	<bkrus>	/ʈýː/	‘wash’
		<bdud>	/tỳt/	‘demon’
		<bdun>	/tỳn/	‘seven’

The word initial radical <a> (with optional vowel diacritic) corresponds to a Walungge word initial vowel with high tone. The following are examples of this: <a zhang> /áɕaŋ/ ‘maternal uncle’, and <ig bu> /íkpa/ ‘hiccough’. Other Walungge word initial vowels which have a correspondence in WT have come from the deletion of a word initial radical, typically <'>, for example <'od> /ø̀t/ ‘sunlight’.

3.5.7 Other changes

The above sections outline the main regular correspondences between WT and Walungge. There are other correspondences which occur with sufficient frequency to be worthy of mention. These include resyllabification and deletion.

In some instances whole syllables have deleted, giving rise to compensatory lengthening on the preceding vowel. Where this has occurred, it has typically been a syllable starting with a bilabial plosive with an inherent vowel <a> on either side. However, there are also instances of deletion of syllables starting with a velar plosive or with the radical <'>. The following are examples of syllable deletion.

<gna' ba>	/náː/	'blue sheep'
<kha ba>	/kʰáː/	'snow'
<yi ge>	/jìː/	'letter'
<rte 'u>	/tíː/	'foal'

There are also instances of resyllabification of word medial WT consonants. In some instances a word medial nasal syllable coda has become the onset to the following syllable due to the deletion of the original syllable onset, e.g.

<'phreng ba>	/ʈʰáŋa/	'rosary'
<sngon po>	/ŋóno/	'blue'

In other instances, a pre-script consonant to the onset of the second syllable in WT has become the coda of the first syllable in Walungge:

<sa mtshams>	/sántsam/	'border'
<rnga bcu>	/ŋóptɕu/	'fifty'

Vowel sequences have occurred in instances where the WT has a back rounded vowel <o> or <u> followed by a bilabial plosive.

WT	*Walungge*	*gloss*
<ko ba>	/kúa/	'leather'
<dbu ba>	/pùak/	'bubble'

4 Acoustic analysis of pitch, phonation and duration

4.1 Introduction

As is the case with many Tibetan languages, Walungge has a 2-way phonological contrast which can be termed either “tone” or “register” (see section 2.4 above). The question of terminology is to do with the role of pitch, and its salience compared to the other concomitant factors. Throughout this dissertation Walungge is referred to as having two underlying tones (rather than two registers), because the contrast primarily involves pitch. Part of the purpose of this chapter is to demonstrate this through acoustic analysis.

The following are examples of minimal pairs for tone in Walungge, with the Written Tibetan given in Wylie Transliteration (Wylie (1959)):

Table 4-1: Minimal pairs for tone in Walungge

WT	*Walungge*	*Gloss*
gser	*sér*	‘gold’
gzer	*sèr*	‘screw’
blugs	*lúk*	‘pour’
lug	*lùk*	‘sheep’
khur	*k^húr*	‘carry’
gur	*k^hùr*	'tent’

Section 2.5 above gives a discussion of tone and register in Tibetan languages, and how the register/tone contrast has arisen from the loss of initial consonant clusters and the loss of contrastive obstruent voicing. Section 2.4 discusses the commonly found phonetic correlates of register.

The acoustic correlates of tone in Walungge include differences in pitch, VOT, phonation, and vowel duration, as listed in the table below:

Table 4-2: Correlates of tone in Walungge

high tone	*low tone*
higher starting pitch	lower starting pitch
modal phonation	breathier phonation
later onset of voicing	earlier onset of voicing
shorter duration of rhyme	longer duration of rhyme

In addition to the difference in starting pitch associated with the high-low underlying tonal contrast, Walungge also has different pitch melodies associated with different syllable shapes.

This chapter starts with an acoustic analysis of the pitch, and in particular the way that Walungge pitch melodies vary with syllable shape. The acoustic analysis looks at what differences in the pitch melodies are significant. Among other things, it shows that the starting pitch for monosyllabic words does not only depend upon the underlying tone of the word, but for each tone there is a significant difference in starting pitch correlating with a difference in syllable shape. The section on pitch analysis then considers how the differences in pitch change when the words are put into phrases. The conclusion is that whilst some aspects of the pitch melodies can be explained as phonetic effects, the difference between level and falling pitch for underlying high tone words is a phonological difference caused by a difference in the surface tone pattern. Similarly the rising pitch melody is the result of surface tone rather than a phonetic effect.

Following the analysis of the pitch melodies is an acoustic analysis of phonation and duration. The analysis of the phonation shows that the phonation across a vowel very closely follows the pitch melody across the vowel, becoming less breathy or more breathy as the pitch rises or falls. The analysis of duration suggests that duration correlates with the pitch melody rather than the underlying tone, in that it is longer if the pitch rises but not if the pitch is low and level. Because of the close way that both phonation and duration follow the pitch melody, the conclusion is pitch is the primary acoustic factor through which the contrast in tone is being realised.

The acoustic analysis below primarily analyses the pitch, phonation, and duration, for the isolation form of words. However, the shape of the pitch contour across a word can depend on whether the word is in isolation or in a phrase, the pitch of the surrounding words, the intonation over the phrase as a whole, etc. Thus for the analysis of pitch melodies, having considered the pitch melody across the isolation form of a word, the analysis then considers how this melody varies for words in phrases.

4.2 Data

The method used to collect the data was chosen in order to get as naturalistic speech as possible. A wordlist of approximately 500 words was chosen, with all items on the wordlist easily recognisable by photograph or picture. The words were chosen so as to include all syllable patterns, all phonemes, and both underlying tones. Because of the constraints of using pictures and including all syllable patterns and phonemes, the wordlist was slightly asymmetric in terms the number of high tone words versus the number of low tone words.

The speakers were shown a picture and had to respond to the question "What is this?" In the initial stages of deciding how to collect the data, there was an element of experimentation to determine what worked best. Techniques such as having a speaker repeat the word 3 times did not work. Repeating every word 3 times was not a natural thing to do, which meant that in order to concentrate on saying a word 3 times speakers would end up chanting the word rather than saying it naturally. When asked "what is this" the most natural response was for a speaker to either say the word, or to say the word in the phrase *nò .... dè* "this is a ...". Both responses were so natural that there was a tendency for a speaker to switch between the word in isolation and the word in a phrase without being aware that they had switched. Also very natural was to combine the responses: "...., it is a ...." i.e. to say the word, pause and then say it in the phrase. Giving this combined response was both naturalistic, and easy for speakers to maintain for every word in the wordlist without switching the format of their response. Thus it was decided that speakers should give the response [..., nò ... dè] "..., this is (a) ..." in answer to the question "What is this?" This meant that the words chosen were mostly nouns, but adjectives were also included as they could be said with a very similar format of response. Verbs were excluded as a) the format of response would need to be changed, and b) the phrase final environment can affect the pitch melody.

The wordlist was recorded for 4 speakers: 2 male and 2 female, and checked with other speakers of the language for accuracy. All were mother tongue speakers living within the Walungge language area. None were literate in Written Tibetan, or had received any education in Tibetan. This was an important consideration when choosing speakers. An educated knowledge of a more prestigious Tibetan language could influence the pronunciation of words which are the same in both Walungge and the more prestigious language (Indeed, one potential language consultant admitted that he would not be a suitable person to use for recording because of the number of years he had spent in a Tibetan medium school, which had influenced the way he now spoke Walungge.) Sprigg (1993) comments on "spelling style pronunciation of Written Tibetan" as a reason for conflicting data for the pitch contours of Lhasa Tibetan. Two of the four speakers were literate in Nepali. Whilst a knowledge of Nepali has the potential to influence pronunciation of Walungge (for example, Nepali has contrastive voicing of plosives but not contrastive tone, and this could influence the voicing and pitch correlations for Walungge plosives), care was taken to select speakers who were known by the community to speak the language well, and without any apparent influence from Nepali.

The recordings were made in domestic settings in the Walungge area. Due to the fact that Walungge houses generally have very few and very small windows, and minimal electricity, and there was a need for people to be able to see the pictures, the recordings were made in the quietest outside location available. Generally this was the speaker's balcony. This made for difficult recording conditions, due to background noise from animals or small children, interruptions by other people, etc. For each speaker the recording was completed in a single session. From each recorded wordlist, items were discarded for reasons such as background noise, or the speaker giving a different response from that which was required. This left between 300 and 400 words per speaker.

A subsection of the wordlist was then recorded again using a laryngograph (see 4.4.2 below for the use of laryngograph measurements in acoustic analysis of phonation).

The recordings were made using a Marantz PMD660 digital recorder, Sony ECM957 microphone, and were recorded in uncompressed WAV format. Acoustic measurements were taken using Praat computer software.

4.3 Pitch

This section describes the pitch melodies found for Walungge non-verbal content words, and investigates the differences in pitch which correlate not only with the tone of the word but also with the weight of the syllables. The analysis of monosyllabic words considers the starting pitch of the rhyme and investigates what factors in addition to tone are having an effect upon the starting pitch. For disyllabic words, the starting pitch of both syllables is considered. The patterns on words of three or more syllables are an extension of what is found for disyllabic words. Having considered the pitch melodies for words in isolation, words are then considered in phrases in order to investigate the rises and falls which occur in the pitch melody of words in isolation.

4.3.1 Description of monosyllabic pitch melodies

The monosyllabic pitch melodies found in Walungge vary according to two parameters: the underlying tone of the word and the syllable rhyme. Neither the vowel nor the initial consonant affects the overall contour of the pitch melody (i.e. whether it is rising, falling, or level). The syllable rhymes which influence the pitch melody are the following:

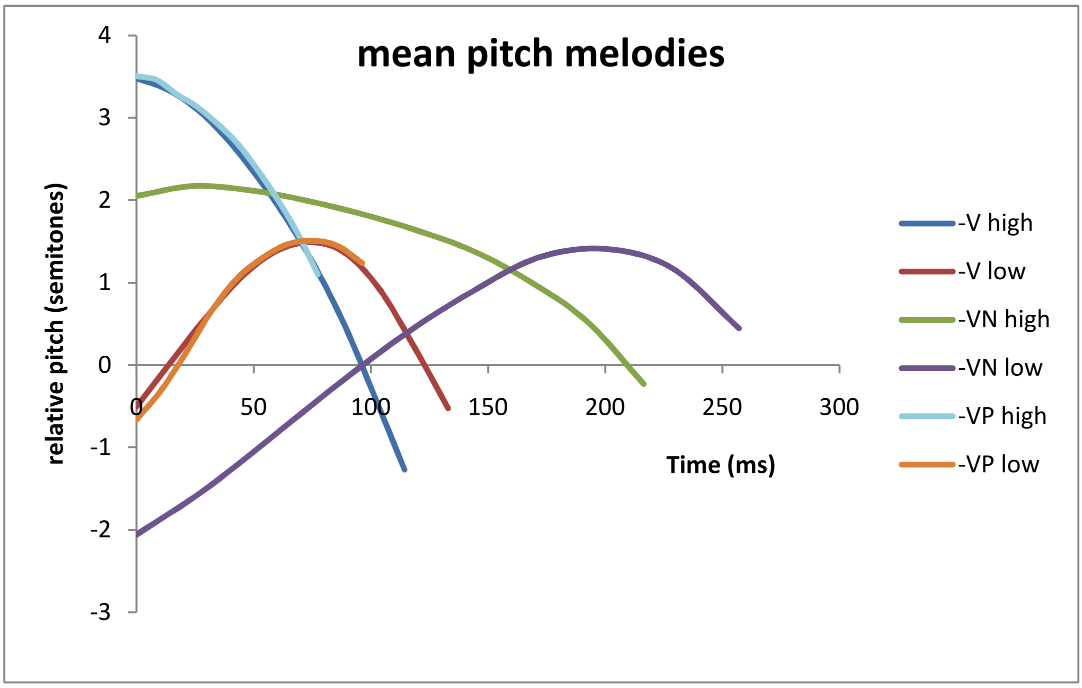
V – a short vowel nucleus and no syllable coda

VP – a short vowel nucleus followed by a voiceless plosive

VN – either a long vowel, or a short vowel plus sonorant coda. The pitch melody extends across the whole of the syllable rhyme, with both types of rhyme having the same pitch melody.

Figure 4-1 below displays the mean pitch melody for each syllable rhyme and tone, for words in isolation.

Figure 4-1: mean pitch melodies for monosyllabic words

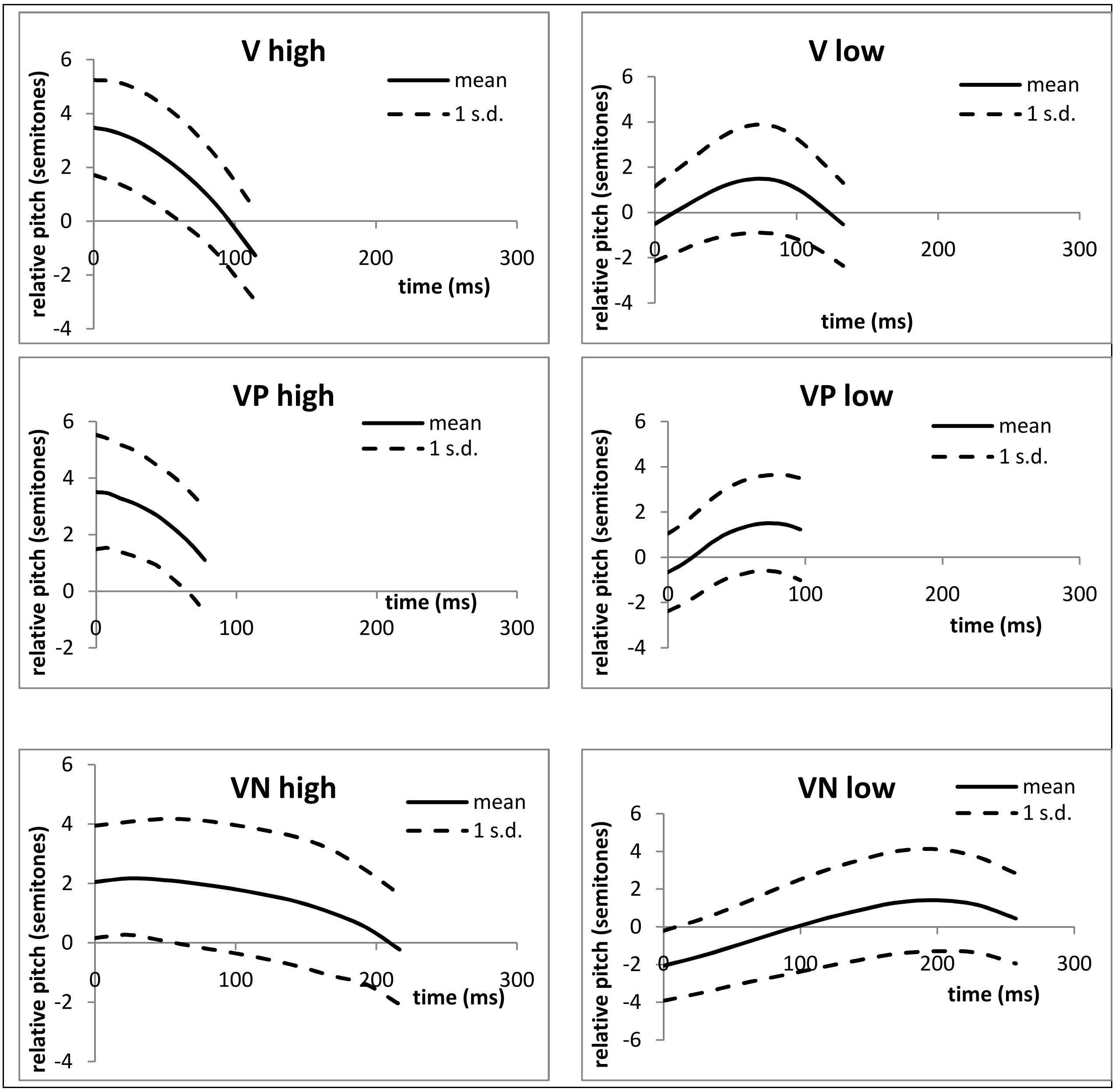


In order to calculate the mean pitch melodies displayed above, measurements were time normalized and averaged as follows:

The fundamental frequency (F0) in Hertz was measured at ten equidistant points across the syllable rhyme, and then converted to semitones in order to normalize between the pitch ranges of different speakers. Semitones were calculated relative to the mean F0 for each speaker. The conversion to semitones was also because semitones are a scale of measurement that closely corresponds to pitch perception (Nolan (2003)). The duration of the syllable rhyme was also measured. For each syllable shape and tone the average duration was calculated. For each word the 10 pitch measurements were positioned equidistantly across the average duration, and for each syllable shape and tone the average pitch at each measurement point was then calculated. This method of time normalization is taken from the method used in Xu (1997).

Figure 4-2 below separates out these pitch melodies, showing the mean pitch and one standard deviation on either side of each melody. These show the amount of variation that there is in the pitch of the melody.

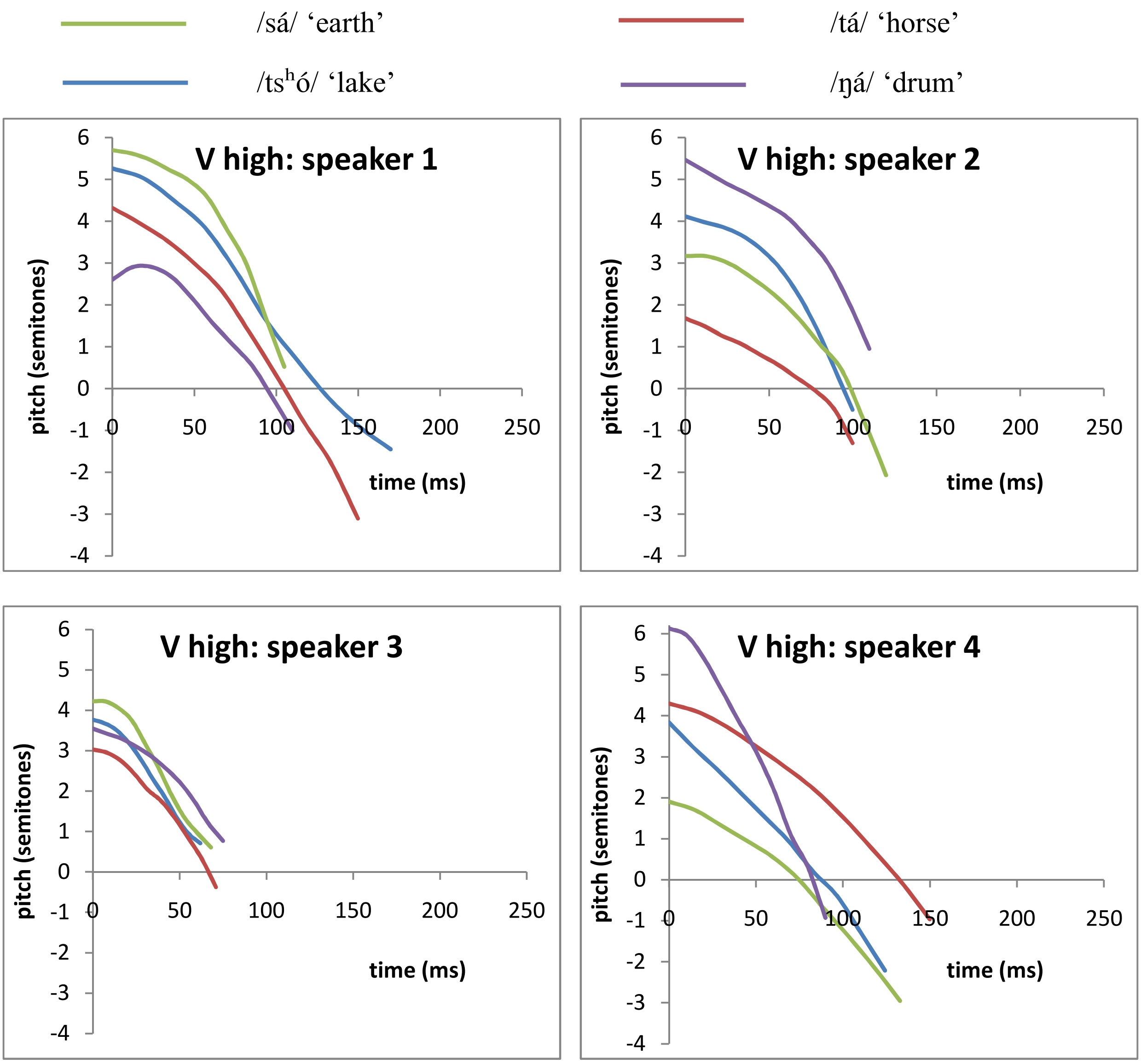
Figure 4-2: mean and standard deviation for pitch melodies



V high

The pitch melody for monosyllabic V words with high tone starts high and falls steeply. As can be seen from Figure 4-1 above, whilst all the pitch melodies for the different rhymes and tones fall towards the end of the syllable rhyme, it is the V high tone which has the steepest and greatest fall. It has the highest starting pitch, and falls below both the starting and ending pitch of the V low tone melody. Figure 4-3 below gives examples of this pitch melody.

Figure 4-3: -V high examples



The above examples illustrate the variation that there is in the starting pitch, the amount of fall and the duration of the syllable rhyme. However, despite this variation, the overall shape of the melody (starting high and falling) is always maintained regardless of speaker and regardless of whether the initial consonant is a plosive, affricate, fricative, or sonorant.

VN high

The rhyme VN comprises words that have a long vowel and no coda, and words with a short vowel and sonorant coda. The pitch melody is same regardless, and is a sustained high pitch which has a slight fall primarily in the latter part of the rhyme. The starting pitch, whilst in the upper half of the speaker's ranger, is generally not as high as the starting pitch of V and VP high. And the fall is considerably less than the fall of V high. Figure 4-4 gives examples of VN high tone words.

Figure 4-4: -VN high examples

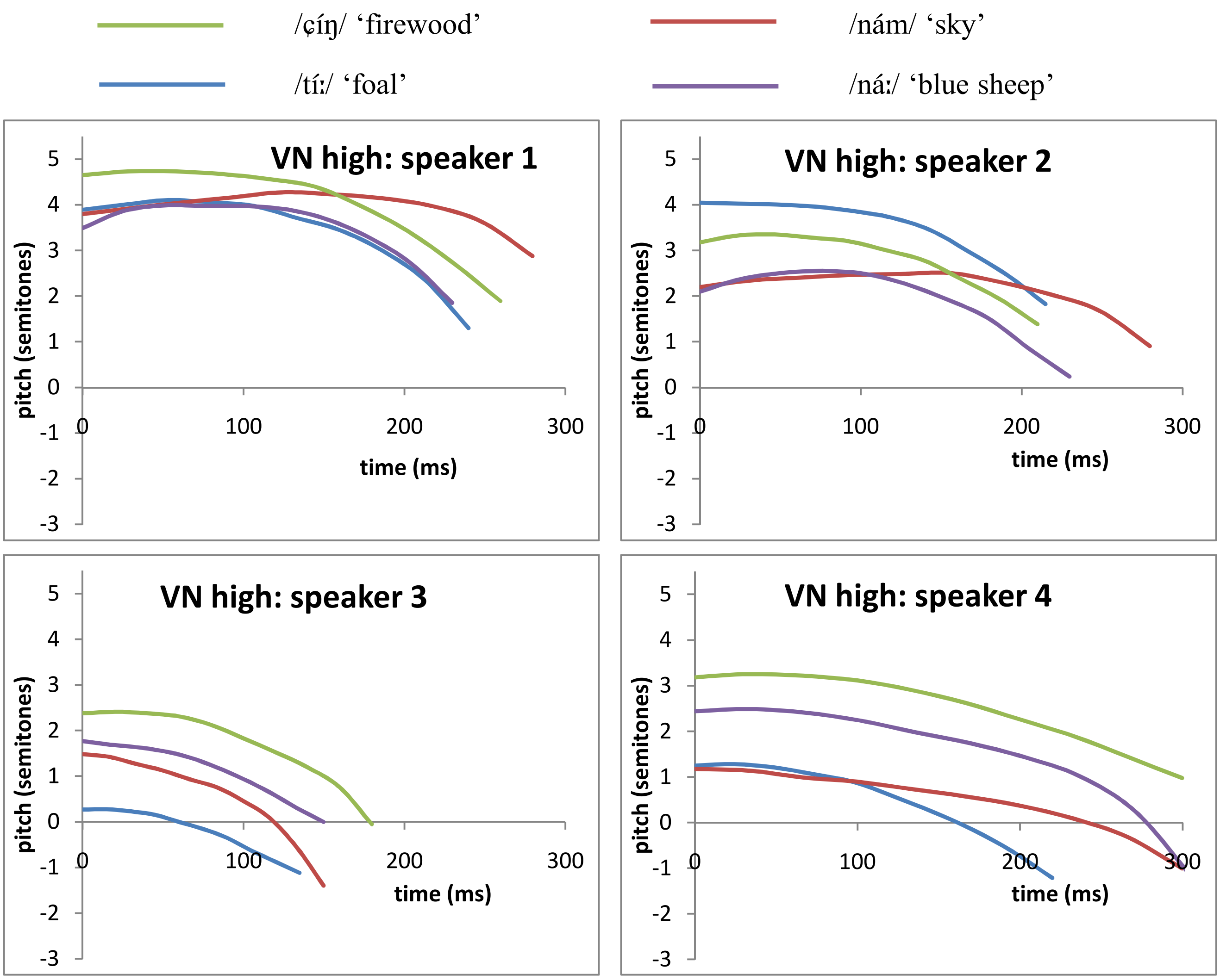
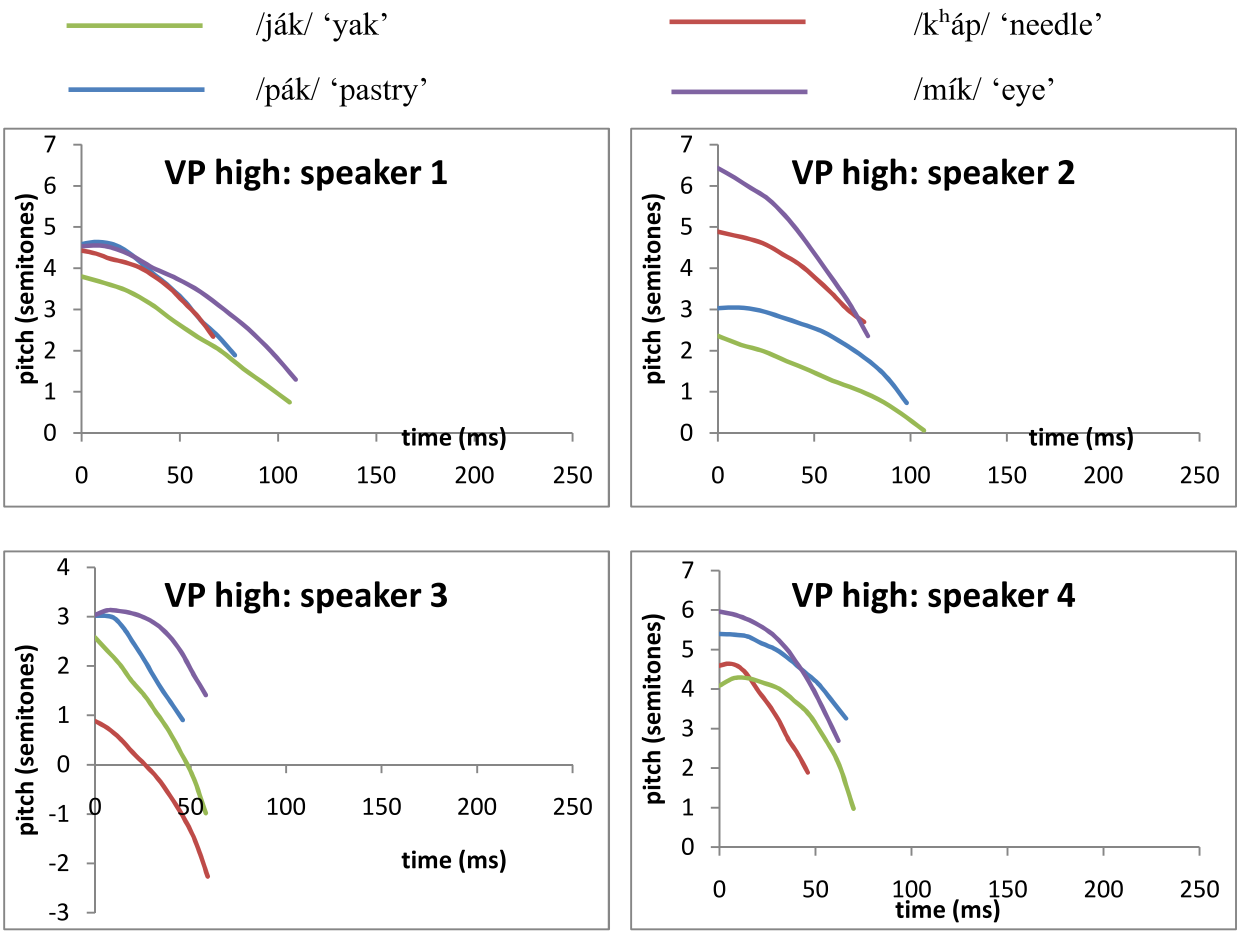


Figure 4-4 illustrates the variation that there is, both in the pitch and in the duration of the rhyme. It also illustrates that despite the variation, the basic shape of the pitch melody is the same in every case.

VP high

As with V-high, the VP pitch melody starts high and immediately starts to fall. Because the duration of the vowel is shorter for VP, the amount of fall is not as great as for V. However, from the graph of the mean pitch melodies (Figure 4-1) it can be seen that the trajectory of the fall is similar for both V and VP. This is different from some other Tibetan languages (e.g. Lhasa Tibetan (Duanmu 1992)) where the presence of a final plosive changes the trajectory of the fall. The following are examples of VP words:

Figure 4-5: VP high examples

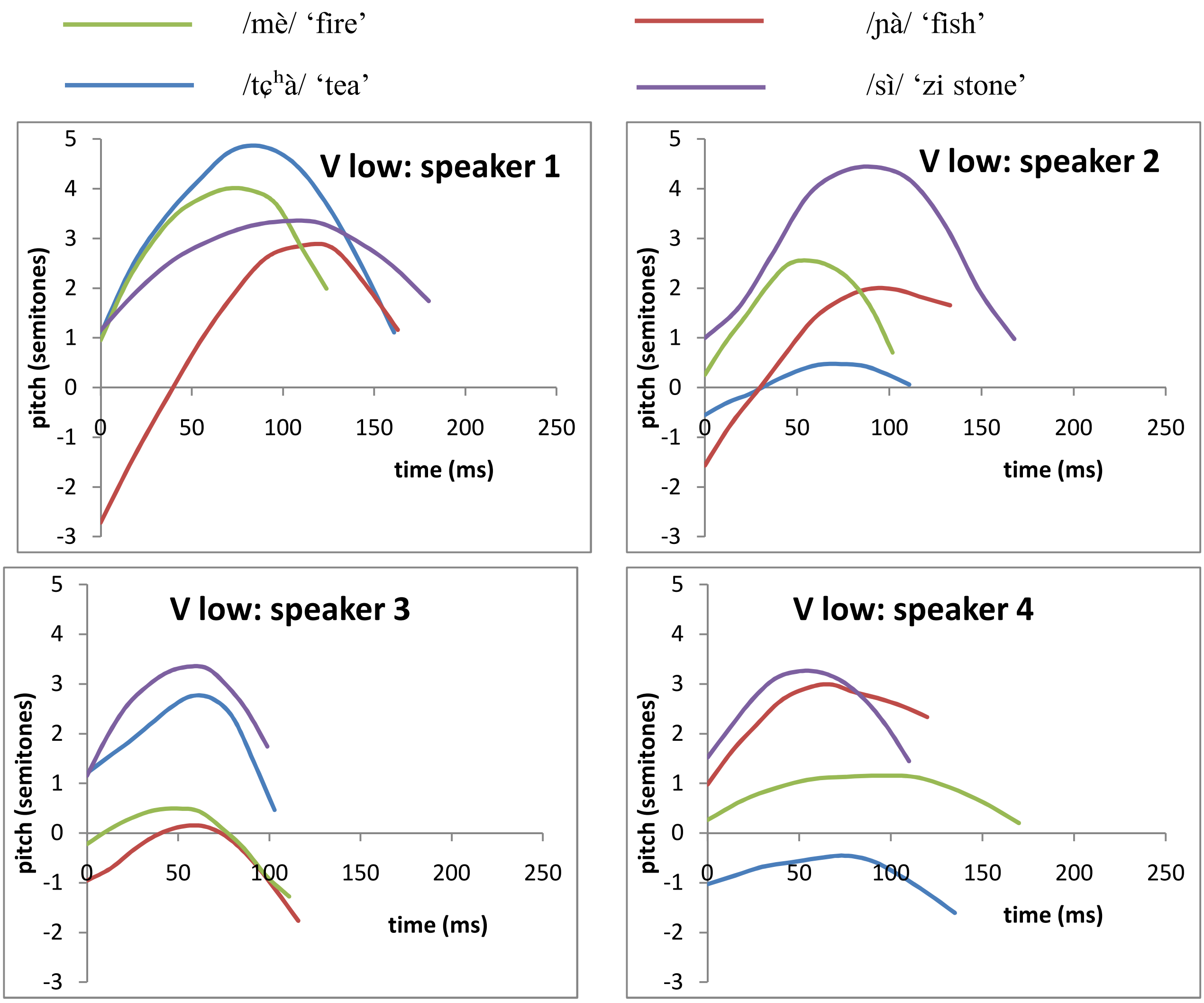


As with other high tone word rhymes, there is variation in the pitch and in the duration, but the basic shape of the pitch melody is maintained.

V low

The mean melody for low tone words with V rhyme rises then falls. Of all the monosyllabic pitch melodies, V-low is the one which displays the most variation between words when spoken. Whilst all the words recorded basically adhered to the rising-falling pattern, there were those which rose more than they fell, those which fell more than they rose, and those which rose and fell only a small amount. Below are examples of V-low words:

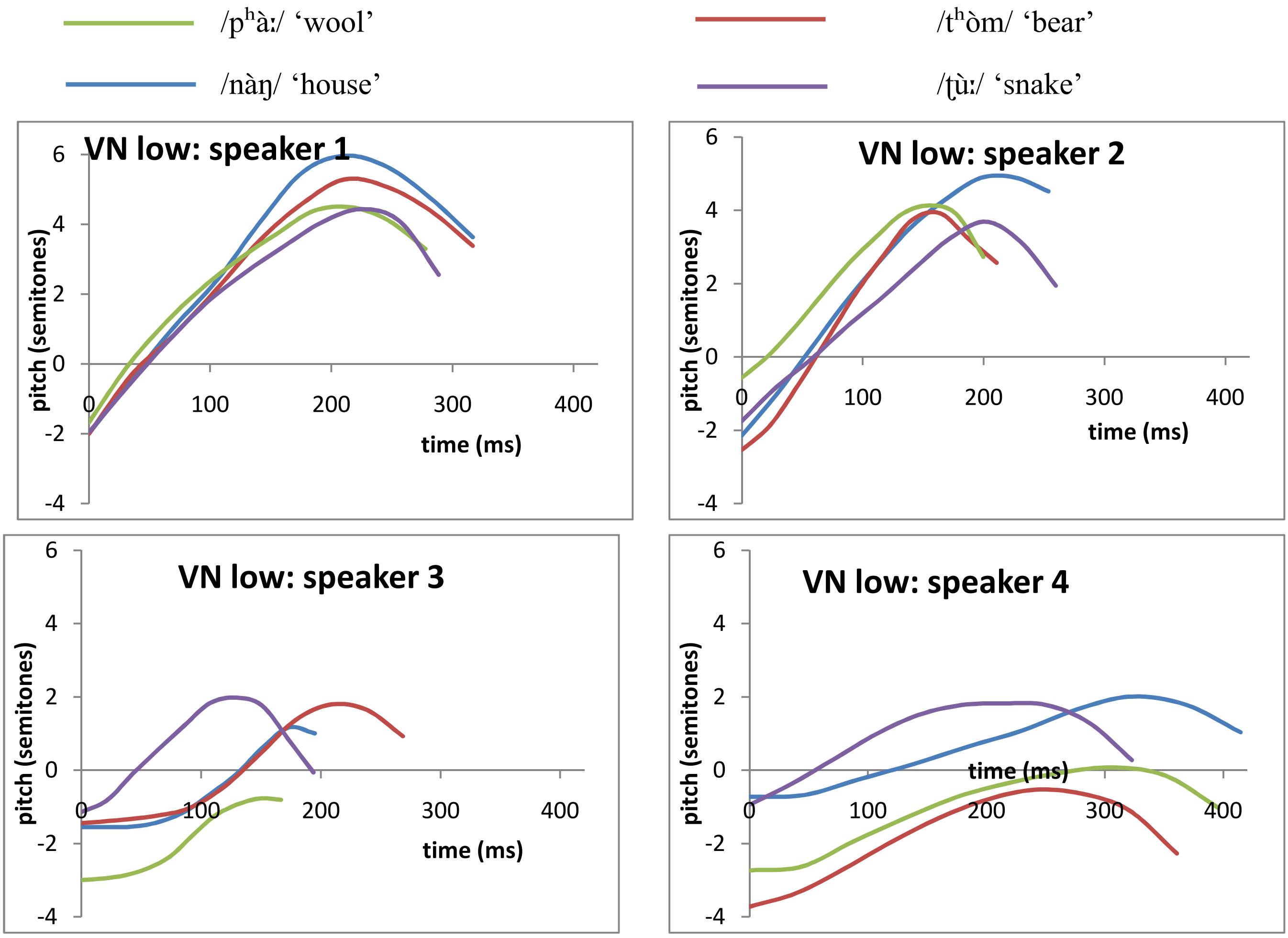
Figure 4-6: V-low examples



VN low

The VN-low pitch melody starts low, rises through most of the syllable rhyme and falls towards the end. This melody is the same for words which have a long vowel as the syllable rhyme, and those which have a sonorant coda. The starting pitch of VN low is the lowest pitch for all monosyllabic pitch melodies.

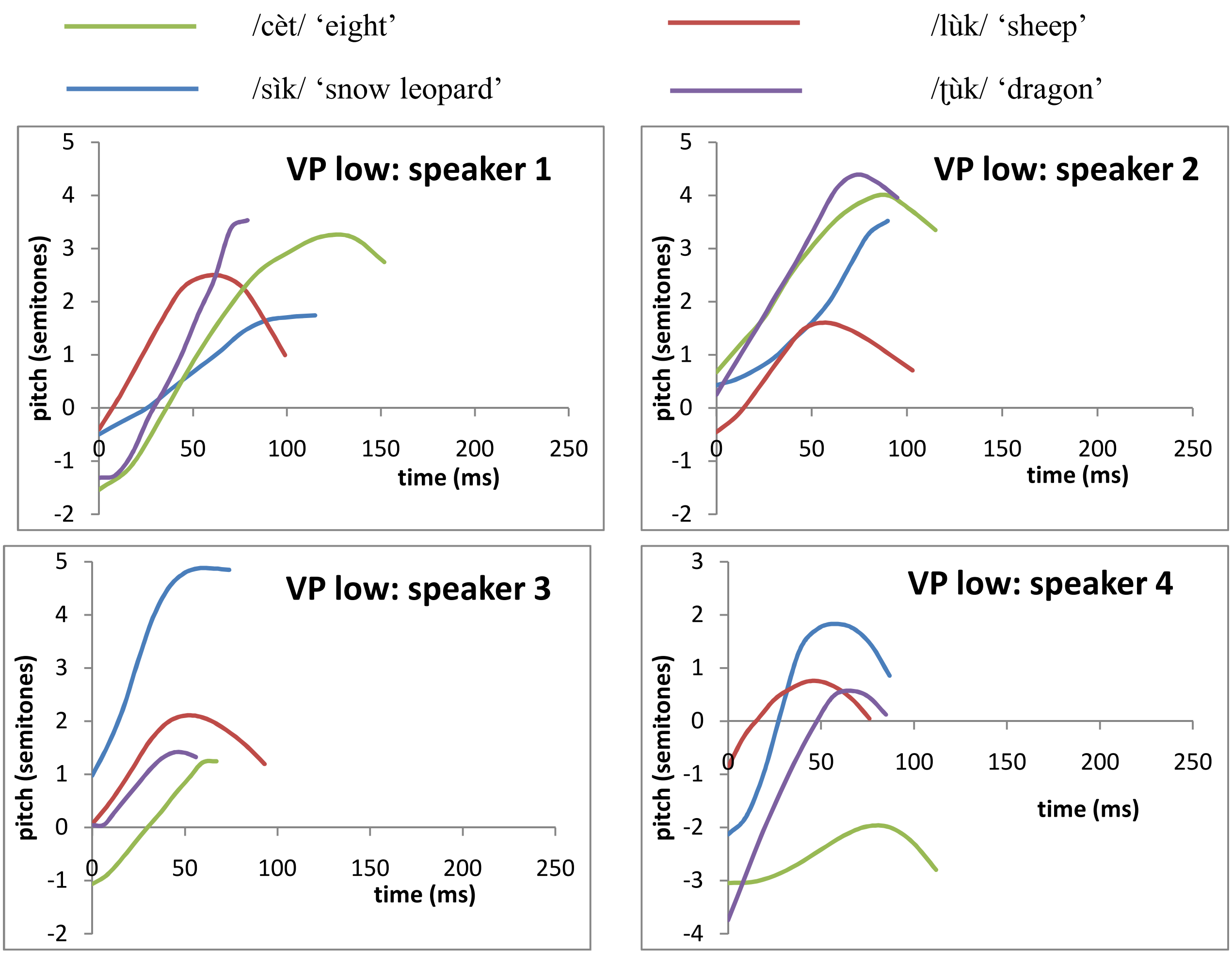
Figure 4-7: VN-low examples



VP low

The pitch melody of VP-low is rising pitch. The starting pitch and gradient of the rise are similar to that of V-low. However, the vowel duration for VP is shorter than for V, which means that VP reaches its maximum pitch and only just starts to fall before the end of the syllable rhyme.

Figure 4-8: VP-low examples



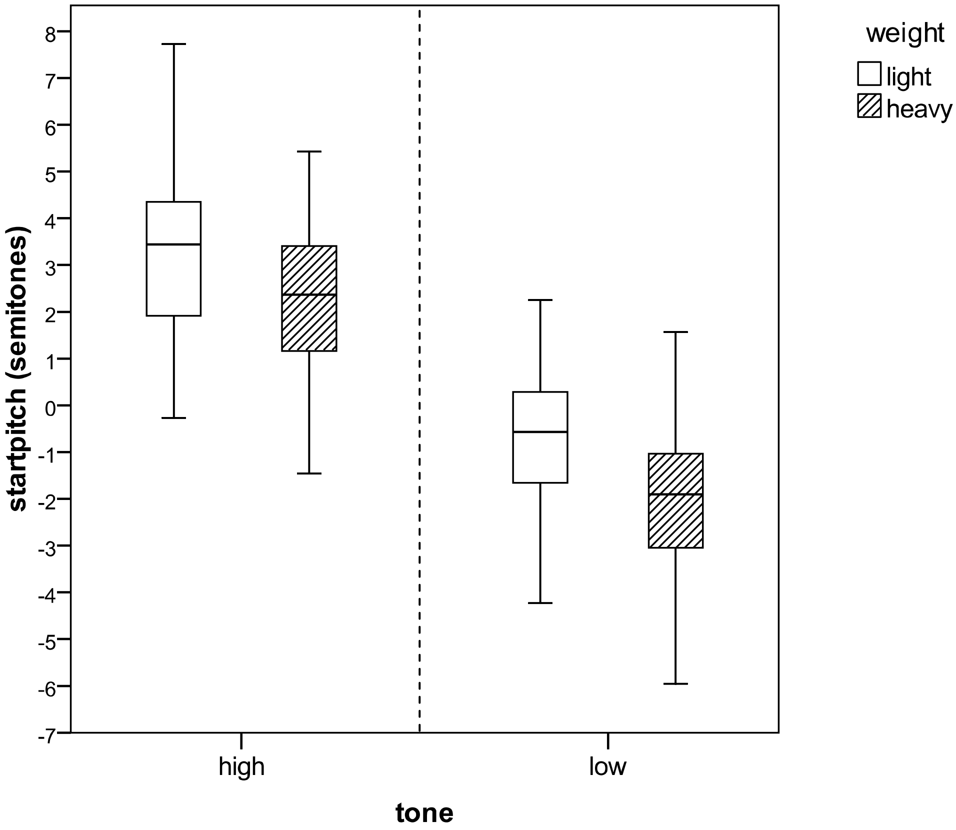
4.3.2 Statistical analysis of monosyllabic melodies

Relative starting pitch

One of the main quantitative ways in which the monosyllabic pitch melodies differ is in the relative starting pitch of the melody. As can be seen from Figure 4-1 above, there are 4 distinct relative starting pitches. The starting pitch differs not simply according to tone, but also according to syllable weight, with heavy syllables (VN) having a lower starting pitch than light syllables (V and VP). Light syllables with high tone have the highest starting pitch, then heavy syllables with high tone, then light syllables with low tone. The lowest starting pitch is for heavy syllables with low tone.

Figure 4-9 is a boxplot of starting pitch for each tone and syllable weight, showing median, inter-quartile range, and whiskers to 1.5 x inter-quartile range.

Figure 4-9: Boxplot of starting pitch



It can be seen that although the average starting pitch of heavy rhymes might be lower than the average starting pitch of light rhymes, there is considerable overlap in the starting pitches of light and heavy rhymes for each tone. There is also overlap in the starting pitches for high and low tone. Of particular interest is not only whether there is a significant difference in starting pitch correlating with tone (which there clearly is), but also whether the difference in starting pitch correlating with the weight of the syllable is significant.

The significance of a difference in starting pitch correlating with a difference in syllable weight as well as a difference in tone was statistically investigated by means of an ANOVA. The ANOVA was a 4 way ANOVA, which included the following independent variables:

- Tone: high and low.
- Syllable weight: heavy and light.

- Initial consonant: initial consonants were divided into four types: sonorants, aspirated plosives & affricates, unaspirated plosives & affricates, and fricatives. Consonants are well known for affecting the pitch of the surrounding segments. Voiced obstruents are generally associated with lower pitch. Aspiration, also, can affect the pitch, but there is no universal way in which it affects it (Tang (2008)). This separates sonorants from obstruents, and aspirated obstruents from unaspirated obstruents. Further, in Walungge unaspirated obstruents can be divided into two groups. Plosives and affricates form one group as a difference in VOT correlates with a difference in tone. Fricatives form a separate group as there is no difference in VOT correlating with a difference in tone (see section 6.3).

- Vowel height: vowels were divided into groups according to vowel height: high vowels /i, y, u/, mid vowels /e, ø, o/ and low vowel /a/. There is a known cross-linguistic correlation between vowel height and fundamental frequency, with high vowels having and intrinsically higher F0 and low vowels having an intrinsically lower F0. (Whalen & Levitt (1995)).

The significance level for the ANOVA is taken to be $p < 0.05$. That is to say, there is less than 0.05 probability that the difference in measurements is being caused purely by chance rather than by the independent variable being considered. The results of the ANOVA are as follows:

Table 4-3: ANOVA for effects on starting pitch

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
initial consonant	425	3	2.02	0.11	
vowel height	425	2	2.12	0.12	
tone	425	1	557	<0.0005	•
syllable weight	425	1	15.2	<0.0005	•

Neither the initial consonant nor the vowel height has a significant effect on the starting pitch of the syllable rhyme. This is interesting given the way that both consonants and vowels can affect the pitch. It is possible that the effects of tone and syllable weight are overriding any slight variation in pitch due to consonant or vowel. ((Connell (2002)) discusses the question of how much the intrinsic F0 of vowel height is present in tone languages.)

Both tone and syllable weight do have a significant effect. The combination of tone and syllable weight gives the four combinations: high light, high heavy, low light, and low heavy, with the following means and standard deviations:

Table 4-4: Mean and standard deviation of starting pitch

type	*mean (semitones)*	*s.d.*	*N*
high light	3.28	1.76	108
high heavy	2.16	1.58	83
low light	-0.80	1.83	123
low heavy	-2.01	1.79	111

In order to confirm these four different combinations have significantly different starting pitches from each other, a post-hoc Scheffé test was carried out, with the following results:

Table 4-5: Scheffé post-hoc test for syllable shape

tone & weight	high heavy		low light		low heavy	
	P	*sig.*	*P*	*sig.*	*P.*	*sig.*
high light	< 0.0005	•	< 0.0005	•	< 0.0005	•
low heavy	< 0.0005	•	< 0.0005	•		
low light	< 0.0005	•				

Each combination of weight and tone has a significantly different starting pitch from each other combination. That is to say, there are four significantly distinct starting pitches, with high light being the highest pitch, followed by high heavy, then low light, and low heavy having the lowest starting pitch. It is particularly interesting that the starting pitch difference which is associated with syllable weight is a significant difference. This raises the question of what is causing this difference. Although a difference in syllable weight might correlate with a difference in pitch, this does not necessarily mean that it is the syllable weight itself which is causing the difference in pitch. Rather, the differences in starting pitch could be an indication that the surface pitch melodies are realisations of separate tone patterns. The possibility of the pitch melodies being realisations of separate tone patterns is returned to in section 4.3.6.

4.3.3 Description of disyllabic pitch melodies

Summary

The disyllabic pitch melodies in isolation form can be summarized as follows. The pitch on the first syllable of a disyllabic word is high or low according to the underlying tone of the word. Unlike monosyllabic words, there is no significant difference in starting pitch correlating with the weight of the first syllable. The second syllable always has high pitch which falls. The following is a table outlining the disyllabic pitch melodies for different combinations of syllable shape and tone.

Table 4-6: summary of disyllabic pitch melodies in isolation

tone	*first syllable rhyme*	*second syllable rhyme*	*pitch melody*	
			1st syllable	2nd syllable
high	-V, -VN, -VP	-V, -VP		starts high and falls
	-V, -VN, -VP	-VN	high	level then falling
low	-V, -VN, -VP	-V, -VP	low	starts high and falls
	-V, -VN, -VP	-VN	low	rises to high, then falling

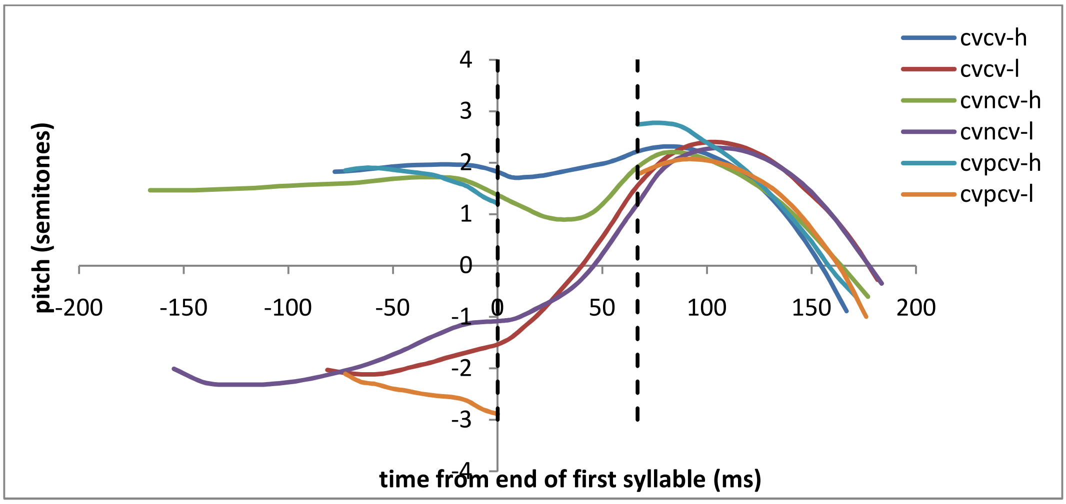
These pitch melodies will be discussed in detail below.

Comparison of disyllabic words with differing first syllables

To display all the disyllabic patterns graphically at the same time makes for an unmanageable graph, and is unnecessary. The value of graphically displaying pitch melodies is for ease of visual comparison. One way of doing this is to divide the disyllabic patterns up into groups which keep one syllable constant and the other syllable differing. This first section compares different first syllable patterns for disyllabic words. In order to keep the second syllable constant, disyllabic words were chosen which have CV as the pattern of the second syllable.

Figure 4-10 below compares the pitch melodies for disyllabic words with CV as the second syllable. The first syllable can have the rhyme V, VN or VP, and the tone can be either high or low.

Figure 4-10: Average pitch melodies with CV as second syllable



The above average pitch melodies are time normalized according to the method outlined for monosyllabic pitch melodies. However there are a few important points about the disyllabic pitch melodies that need to be noted. In order to simultaneously compare the contour of the pitch both for the first syllable and for the second syllable, it was decided to fix the time at two time points. The first of these points is the end of the voicing of the first syllable. For first syllable rhymes V and VN this gives time as zero at the onset of the second syllable. For first syllable rhyme VP this gives time as zero at the closure of the syllable final plosive. Thus the pitch melodies over the first syllable can be visually compared on the above graph. In order to compare across the rhyme of the second syllable, the pitch melodies were also lined up at the start of the second syllable rhyme. However, in order to line up all the pitch melodies in these two places, the duration of the consonantal segments between these two points has been completely time normalized across all pitch melodies. These two points are shown on the graph by dashed lines.

Most words with the first syllable rhyme as V or VN have the vocal folds vibrating through the onset of the second syllable. (The exception to this is when the onset of the second syllable is /s/ or /ç/, neither of which becomes voiced between voiced phones.) Thus in general for words with the first syllable rhyme as V or VN, there is continuous pitch across the whole word. The pitch across the second syllable onset for these patterns has been included on the above graph. For words with the first syllable rhyme as VP, however, the pitch stops at the closure of the voiceless plosive. (Whilst there is a combination of voiced plosive followed by sonorant, this pattern is not a common pattern; the predominant pattern is to have a voiceless plosive coda followed by an obstruent onset). Thus, no pitch has been graphed between the tone bearing units for CVPCV.

The following generalizations about the above disyllabic patterns can be made.

Unlike monosyllabic melodies which have different starting pitch according to the weight of the syllable rhyme as well as tone, disyllabic melodies only differ in starting pitch according to tone.

A difference can be observed between those melodies which have reached their target height at the start of the second syllable (low tone melodies with word medial voicelessness, and all high tone melodies), and those melodies which do not reach their target height until partway through the second syllable rhyme (low tone melodies which are voiced word medially). This difference can be explained by the well-known phenomenon of peak delay (see Xu (1999) for more on peak delay). If there is a low tone syllable followed by a high tone syllable, there is a tendency for the pitch to remain low throughout the first syllable, and only then to start rising. This means that the target high pitch is not reached until at least partway through the second syllable.

Apart from the effect of peak delay, the pitch melody on the second syllable is unaffected by either the shape of the first syllable, or by the underlying tone of the word. In each case the second syllable pitch is high and falling.

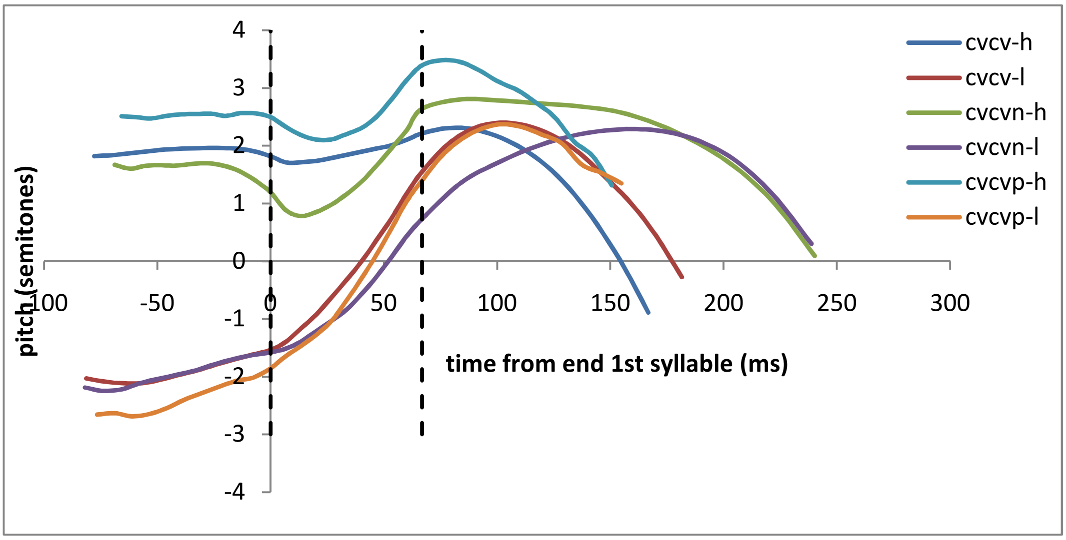
For high tone words, the starting pitch on the first syllable is high and is sustained throughout the first syllable rhyme. For first syllable rhyme VP, however, the pitch drops at the end of the first syllable rhyme, which is the effect of the voiceless plosive. (This is a similar effect as described for syllable final voiceless plosives in Lhasa Tibetan (Duanmu (1992))). Comparing first syllable rhymes V and VP, the duration of the first syllable is unaltered by the presence of a syllable final plosive. This is a different effect from syllable final plosives for monosyllabic words. For monosyllabic words, the effect of the plosive is to shorten the duration of the vowel, but keep the gradient of the pitch melody. This difference in effect can possibly be explained by the different phonetic realization that the voiceless plosive has in each of these environments. Word finally, /t/ and /k/ are both realised as [ʔ], whereas in a word medial syllable coda they are generally realised as [t] and [k]. The glottal closure of [ʔ] could be what is giving rise to the shorted vowel duration.

In the case of low tone words, the starting pitch is low. For first syllable rhymes V and VN, both rise slightly through the rhyme of the first syllable, in anticipation of the high pitch of the second syllable. However, most of the rise between the low of the first syllable and high of the second syllable occurs during the consonantal onset to the second syllable. For low tone with first syllable rhyme VP, the pitch does not rise at all during the first syllable. Rather, it falls, which is the same effect as for VP with high tone. The word medial break in the voicing of the segments enables the second syllable rhyme to start high.

Comparison of disyllabic words with differing second syllable

Figure 4-11 below shows the average pitch melodies for disyllabic words with the first syllable held constant as CV, and varying the second syllable pattern. This gives the word patterns CVCV, CVCVN and CVCVP. The CVCV patterns have already been described above, but are included here for comparison with CVCVN and CVCVP. As was done for the pitch melodies in Figure 4-10 above, the pitch melodies in Figure 4-11 have been time normalized, and have been lined up both at the end of the voicing of the first syllable rhyme, and at the start of the second syllable rhyme.

Figure 4-11: Average pitch melodies for disyllabic with CV first syllable



There are similarities between these pitch melodies, and those in Figure 4-10 above, in that the first syllable shows a pitch distinction which correlates with tone, and the second syllable is essentially high and falling.

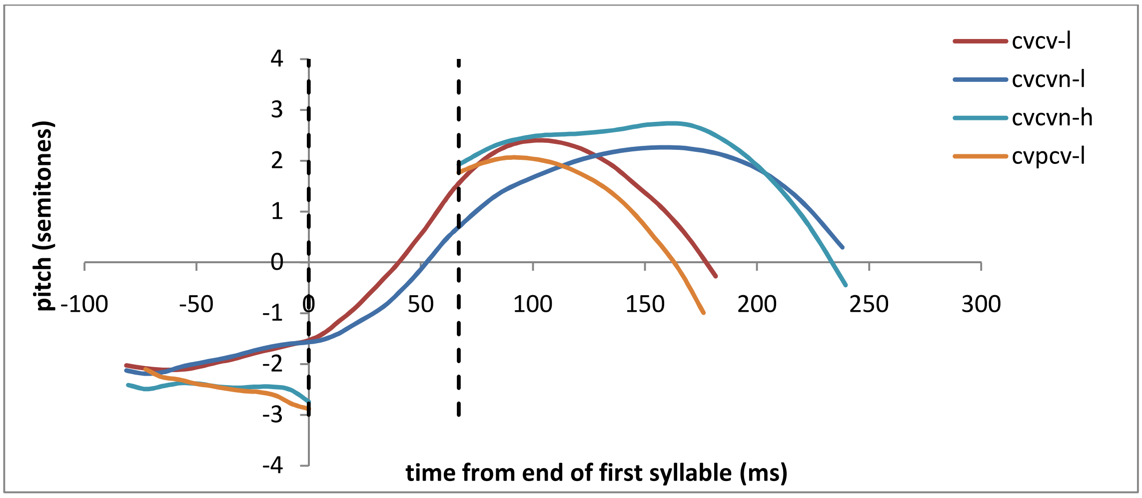
The rhyme of the second syllable has an effect on the pitch melody as follows.

As is the case with monosyllabic words, the effect of a word final plosive is to shorten the duration of the syllable rhyme, but to keep the gradient of the pitch melody roughly the same.

Comparing second syllables CV and CVN for words with underlying high tone, it can be seen that the effect of the heavy syllable is to sustain the high tone before falling. This is comparable to the difference between CV and CVN for monosyllabic high tone words, though with a notable difference. For monosyllabic CV-high words, the starting pitch is significantly higher than that of CVN-high words, whereas in the second syllable of a disyllabic word the starting pitches of CVN and CV are not significantly different (see statistical analysis below). Also for monosyllabic words the mean fall for CV words is considerably greater than that of CVN words, whereas for disyllabic words the mean fall for CV words is approximately the same as the mean fall for CVN words.

Comparing second syllables CV and CVN for words with low underlying tone, it can be seen that there is a difference in the positioning of the second syllable high pitch. Whilst for both patterns the rise from low to high occurs primarily during the consonantal onset to the second syllable, for CVN the rise continues well into the second syllable and the peak is not reached until more than halfway through the rhyme. In order to further consider the effect of second syllable CVN for words with low tone, the following graph compares the mean low tone pitch melodies for CVCV (voiced medial consonant), CVCVN , CVPCV (voiceless medial consonants) and CVPCVN.

Figure 4-12: mean pitch melodies for low tone disyllabic words



The difference between CVCV and CVPCV has already been discussed; the effect of the voiceless compared with voiced medial consonant is for the high pitch of the second syllable to be positioned at the start of the syllable rhyme rather than partway through. However, this is not the case when the second syllable is CVN. For both CVPCVN and CVCVN the peak is reached more than midway through the second syllable. Part of the statistical analysis below considers whether or not this difference is significant.

4.3.4 Statistical analysis of disyllabic pitch melodies

Starting pitch of first syllable

Looking at the above graphs of the disyllabic pitch melodies, whilst there is a clear difference in starting pitch correlating with tone, there is very little obvious difference correlating with the weight of the first syllable. Given the significant difference in starting pitch correlating with weight for monosyllabic words, it is important to determine whether there is any similar significant difference for disyllabic words. If there is no significant difference correlating with weight for disyllabic words, this is one further indication that the differences in monosyllabic words are caused by tone rather than by the weight of the syllable.

Because all disyllabic words have high pitch on the second syllable, it was decided to use this as a reference pitch, and thus consider the starting pitch relative to the maximum pitch of the second syllable for each word. This reduces the variation that there is in the starting pitch due to speaker variation, without reducing the variation caused by the segments or the tone.

An ANOVA was carried out for influences on the starting pitch of the first syllable rhyme for disyllabic words, with the following independent variables:

- tone
- Initial consonant: the possible effect on pitch due to initial consonant has been discussed in section 4.3.2. In summary, initial consonants can be divided into 4 sets: sonorants, aspirated plosives & affricates, unaspirated plosives & affricates, and fricatives. For disyllabic words, in the data there are words starting with a vowel. However, there are only very few of these, and they have been excluded from the ANOVA.
- Vowel height: the possible effect on pitch due to vowel height has been discussed in section 4.3.2. In summary, there are 3 vowel heights: high, mid and low.
- Weight of first syllable: for monosyllabic words, a difference in syllable weight correlates with a difference in starting pitch. Thus it is important to include syllable weight as an independent variable for an ANOVA for disyllabic words.

Table 4-7 below gives the results of the ANOVA.

Table 4-7: ANOVA for effects on starting pitch for disyllabic words

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
initial consonant	868	3	4.17	0.055	(•)
vowel height	888	2	6.27	0.005	•
tone	888	1	1429	<0.0005	•
weight of 1st syllable	888	1	1.14	0.286	

The ANOVA clearly shows that tone has a significant effect on the starting pitch, which is the expected result. However, unlike monosyllabic words, a difference in the weight of the first syllable does not correlate with a difference in pitch, suggesting that in the monosyllabic case it is not the weight of the syllable which is causing the difference there.

A further difference between monosyllabic and disyllabic words is that in the disyllabic case initial consonant and vowel height do have a significant effect on the starting pitch (although initial consonant is on the border of being significant).

Below is a table showing the mean and standard deviation of the starting pitch for different vowel heights, with pitch in semitones relative to the high pitch of the second syllable:

Table 4-8: table of mean relative starting pitch for vowel height and tone

tone	*vowel height*	*mean starting pitch*	*s.d.*	*N*
high	high	-0.87	1.27	142
	mid	-0.92	1.28	175
	low	-1.05	1.25	155
low	high	-4.55	2.29	112
	mid	-4.56	2.48	123
	low	-4.72	2.45	181

From Table 4-8 it can be seen that whilst the differences in mean starting pitch for different vowel heights are extremely slight and there is a large amount of variance, for each tone vowel height produces a gradient for mean starting pitch, with high vowels having slightly higher pitch, low vowels having slightly lower pitch, and mid vowels in the middle. This fits with the findings of Connell (2002) where in general vowels have an intrinsic fundamental frequency (IF0), and this is a) gradient with respect to vowel height (i.e. higher vowels have higher IF0) and b) gradient with respect to tone (i.e. higher tones have higher IF0).

For Walungge, the differences in mean pitch for different vowel heights are very slight, with only about a quarter of a semitone between high vowel pitch and low vowel pitch for each tone. A post-hoc Scheffé test on mean starting pitch was carried out for vowel height and tone, with the following results:

Table 4-9: Scheffé test for effect of vowel height on starting pitch; high tone

vowel height	mid		low	
	P	*sig.*	*P*	*sig*
high	0.941		0.443	
low	0.614			

Table 4-10: Scheffé test for effect of vowel height on starting pitch; low tone

vowel height	mid		low	
	P	*sig.*	*P*	*sig*
high	0.995		0.776	
low	0.846			

What can be seen from the post-hoc Scheffé test is that as soon as the results are separated by tone and each vowel height is compared, there are no significant differences. The differences only become significant when all the results are pooled together in the ANOVA.

Connell notes that not all tone languages show IF0. Whether or not a language has IF0 appears to depend on the tone inventory and the modulation used in producing the tonal contrasts. This could explain the difference in Walungge between monosyllabic words and disyllabic words. For disyllabic words, tone gives a 2-way pitch distinction, whereas for monosyllabic words there is a 4-way distinction correlating with both tone and syllable weight. The effect of both tone and syllable weight suppresses any IF0 effect for monosyllabic words, whereas for disyllabic words, whilst IF0 is too slight to be picked up in the post-hoc Scheffé test, it is sufficient to be picked up in the ANOVA.

Below is a table of means and standard deviation of the starting pitch for different initial consonants.

Table 4-11: Mean starting pitch for different word initial consonants

tone	*initial consonant*	*mean starting pitch*	*s.d.*	*N*
high	unaspirated	-0.91	1.32	143
	aspirated	-0.93	1.27	139
	fricative	-1.10	1.31	55
	sonorant	-0.84	1.08	104
low	unaspirated	-3.95	2.28	141
	aspirated	-4.66	2.35	122
	fricative	-4.70	2.29	54
	sonorant	-5.35	2.45	99

From the table above, it can be seen that:

- for each tone the differences in pitch associated with the type of consonant are very slight in comparison with the differences in pitch associated with tone
- the differences in pitch due to initial consonant are primarily with low tone.

For each tone, the post-hoc Scheffé test below considers the difference in effect on starting pitch for different consonant types:

Table 4-12: Scheffé test for effect of initial consonant on starting pitch; high tone

initial cons.	unasp.		fricative		sonorant	
	P	*sig.*	*P*	*sig*	*P.*	*sig.*
asp.	0.852		0.999		0.718	
sonorant	0.997		0.963			
fricative	0.992					

Table 4-13: Scheffé test for effect of initial consonant on starting pitch; low tone

initial cons.	unasp.		fricative		sonorant	
	P	*sig.*	*P*	*sig*	*P.*	*sig.*
asp.	0.299		0.383		0.317	
sonorant	0.307		0.396			
fricative	0.375					

From Table 4-12 and Table 4-13 above, it can be seen that as soon as the measurements are split by tone and by initial consonant type, there are no significant differences between the measurements. As is the case for the effect of vowels upon pitch, the differences for consonant type are only very slight, and are too slight to register as significant for the post-hoc Scheffé test.

Starting pitch of second syllable

For both high and low tone, the pitch melody across the rhyme of the second syllable is high and falling. There are several factors of relevance when considering the starting pitch of the second syllable.

- Even though the underlying tone contrast is realised as high versus low pitch on the first syllable only, it is important to consider whether there is any effect from the tone on the second syllable. From the pitch graphs above (Figure 4-10 and Figure 4-11) it can be seen that the pitch melody on low tone words is still rising from low to high at the start of the second syllable rhyme. The result is that the mean value of the second syllable starting pitch for low tone words is slightly lower than that of high tone words. A statistical analysis must consider whether or not this difference is significant.

- For monosyllabic words a difference in syllable weight correlates with a difference in starting pitch; however the first syllable of disyllabic words has no significant difference in pitch correlating with syllable weight. Given the similarity of pitch melodies on the second syllable with monosyllabic high melodies (falling for a light syllable; sustained high then falling for a heavy syllable) the question is whether there is a difference in starting pitch of the second syllable correlating with syllable weight. In section 4.3.3 above it was noted that for low tone words with heavy second syllable, the high peak of the second syllable did not occur until the latter part of the syllable, making the starting pitch lower for heavy syllables than for light syllables.

- Whether or not word medial consonants affect the starting pitch is also an issue to be investigated. However, as chapter 6 is specifically about the interaction between consonants and tone, the effect of word medial consonants is considered in that chapter.

- Vowel height of second syllable: The vowel height of the first syllable was found to cause a very slight difference in the starting pitch of the first syllable. For completeness, this is included as a variable in the statistical analysis that follows.

An ANOVA was carried out to consider the effects of tone, vowel height, and syllable weight upon the starting pitch of the second syllable. Just as in the analysis of first syllable starting pitches, the starting pitch of the second syllable was taken as a measurement relative to the maximum pitch of the second syllable. Because of the differences in pitch melody associated with tone described above, when considering the effects of syllable weight the data was separated according to tone.

Table 4-14: ANOVA for effects on starting pitch of second syllable

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	888	1	49.9	<0.0005	•
vowel height	888	2	1.73	0.178	
syllable weight for high tone words	472	1	0.13	0.720	
syllable weight for low tone words	416	1	15.19	<0.0005	•

The vowel height does not have a significant effect upon the starting pitch of the second syllable. This matches with the ANOVA results for monosyllabic words. What is interesting is that vowel height has a significant (though very slight) effect upon the pitch of the first syllable but not on the pitch of the second syllable. However, as Connell points out (Connell (2002)), it is not just the tone inventory size that can have an effect on IF0, but also the degree of modulation permitted for each tone. Given the significant difference in starting pitch of second syllable for both high and low tone words and for syllable weight, that could possibly be overriding any effect of IF0.

The starting pitch of the second syllable is significantly affected by the tone, with low tone words having a lower starting pitch. This is attributed to peak delay, which has already been discussed above. For low tone words the second syllable starting pitch is also significantly affected by the weight of the second syllable; whereas for high tone words, the syllable weight does not significantly affect the starting pitch. Below is a boxplot and table of means and standard deviation for the starting pitch of the second syllable by weight and tone.

Figure 4-13: Boxplot of starting pitch of 2nd syllable for tone and syllable weight

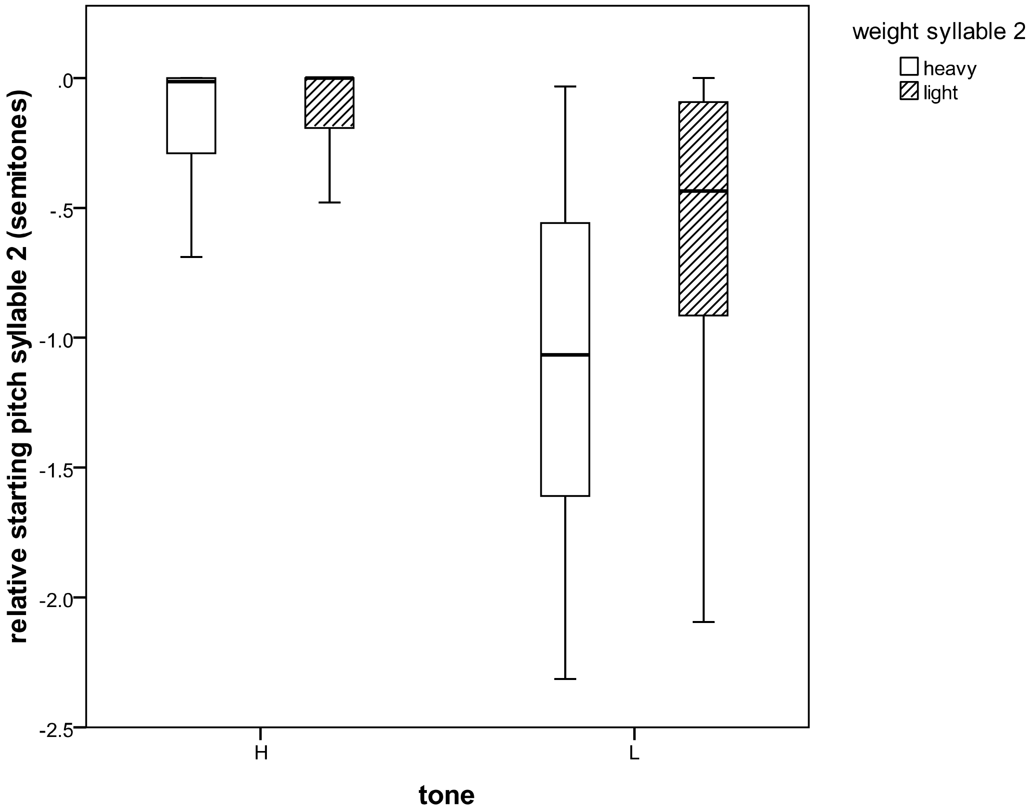


Table 4-15: Table of mean and s.d. for starting pitch of second syllable

tone	*syllable weight*	*mean pitch 2nd syll.*	*s.d.*	*N*
high	heavy	-0.15	0.21	84
	light	-0.13	0.22	388
low	heavy	-1.08	0.69	70
	light	-0.59	0.59	374

As can be seen from the above boxplot and table of means, both heavy and light second syllables have a starting pitch which is lower for low tone words. And low tone words have a second syllable starting pitch which is lower for heavy syllables.

It is of interest that for high tone words there is no significant difference in starting pitch correlating with syllable weight, whereas for monosyllabic words there is a significant difference. For both monosyllabic and the 2nd syllable of disyllabic high tone words, light syllables start high and fall whereas heavy syllables sustain the high before falling. If for monosyllabic words the difference in starting pitch was simply a phonetic effect acting as an enhancement to the difference between long and short vowels, one might expect the same phonetic enhancement to be there on the 2nd syllable of disyllabic words. But it is not. The difference can be explained if the fall for monosyllabic words is taken to be phonological, but the (smaller) fall for disyllabic words is taken to be a boundary effect caused by the isolation environment (investigated further in section 4.3.6.3 below). The difference in starting pitch for monosyllabic high tone words can then be explained as a phonetic enhancement to the difference between a falling tone and a level tone (with the higher starting pitch giving the speaker a greater range for the fall). The lack of difference in starting pitch for the 2nd syllable of disyllabic words would then be explained by saying that as the fall at the end of disyllabic words is a boundary effect rather than phonological tone the speaker has no need to enhance the fall by starting at a higher pitch.

It is also of interest that for low tone words there is a significant difference in the starting pitch of the second syllable correlating with syllable weight. This difference will be returned to in chapter 5, in the assigning of tones to moras. In summary, the explanation is that there is a non-lexical high tone which attaches to the final mora of the second syllable of a word. A heavy syllable is bimoraic, which means that the high pitch associated with the high tone placement is not achieved until partway through the syllable.

4.3.5 Polysyllabic pitch melodies

The pitch melodies of polysyllabic words can be summarized as follows:

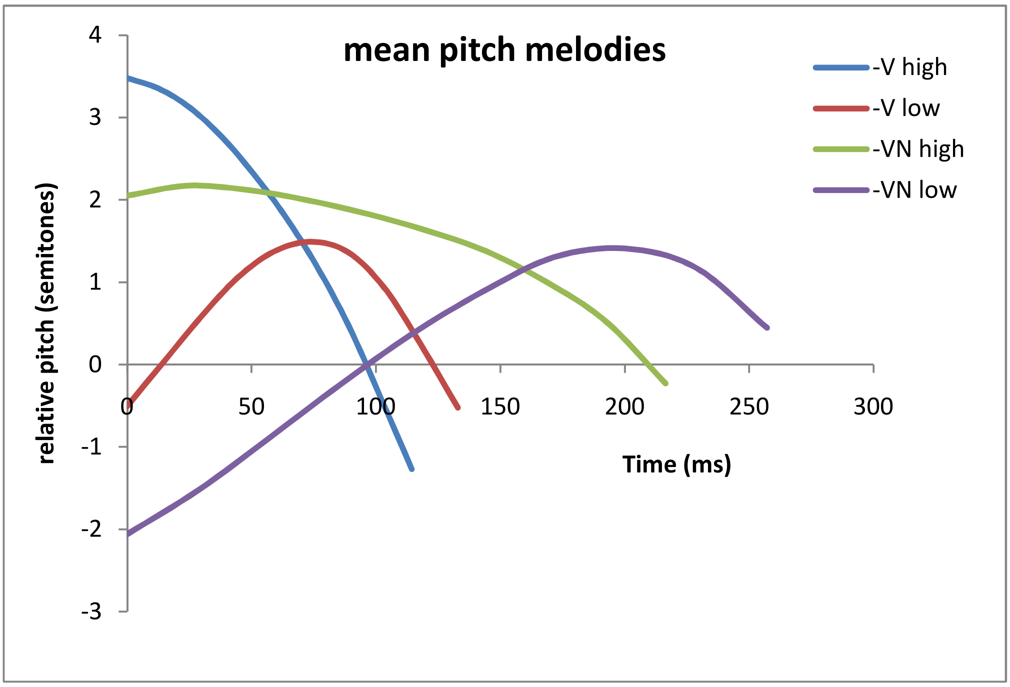
- The first syllable has high or low pitch correlating with the tone of the word.
- The second syllable has high pitch, regardless of the underlying tone of the word. If the second syllable rhyme is VN and the tone is low, then the high pitch is not reached until midway through the rhyme.
- The pitch starts falling across the onset to the third syllable. It continues to fall across the rhyme of the third syllable, reaching low at the end of the syllable.
- All subsequent syllables are low.

As the pitch melodies for words of 3 or more syllables are essentially an extension of the disyllabic pitch melodies, no further acoustic analysis was carried out on them.

4.3.6 Monosyllabic pitch melodies in phrases

Below shows again the graph of the mean pitch melodies for monosyllabic nouns in isolation (see 4.3.1).

Figure 4-14: mean pitch melodies for monosyllabic words in isolation



This section considers the question of whether the above four pitch melodies are all realisations of separate tone patterns, or whether there is only a high-low tone distinction and any further differences in the pitch melodies are not phonological. Only four of the six monosyllabic pitch melodies are considered because the pitch melodies for VP rhymes have already been accounted for by analysing them as variants of the pitch melodies for V rhymes, with the final plosive shortening the vowel and thus "chopping off" part of the pitch melody.

In isolation form, all four pitch melodies have a fall. This section argues that whilst for three of the pitch melodies (V-low, VN-high, VN-low) the fall can be explained as a boundary effect, the fall associated with V-high cannot be explained in this way and has to be attributed to tone. The rise of both low tone melodies is also argued to be attributed to tone. The conclusion of the analysis of monosyllabic pitch melodies is that the above four pitch melodies are realisations of three tone patterns, which will be referred to as HL (falling), H (high level), and LH (rising), The analysis of these as surface tone patterns rather than underlying tone patterns is discussed in Chapter 5. By extension, it is also argued that the falls that occur for disyllabic words are caused by the isolation environment, and thus a disyllabic word with underlying tone H has surface tone H, and a disyllabic word with underlying tone L has surface tone pattern LH. It should be noted that these are the patterns on nouns. Verbs have different patterns, which are discussed in Chapter 5.

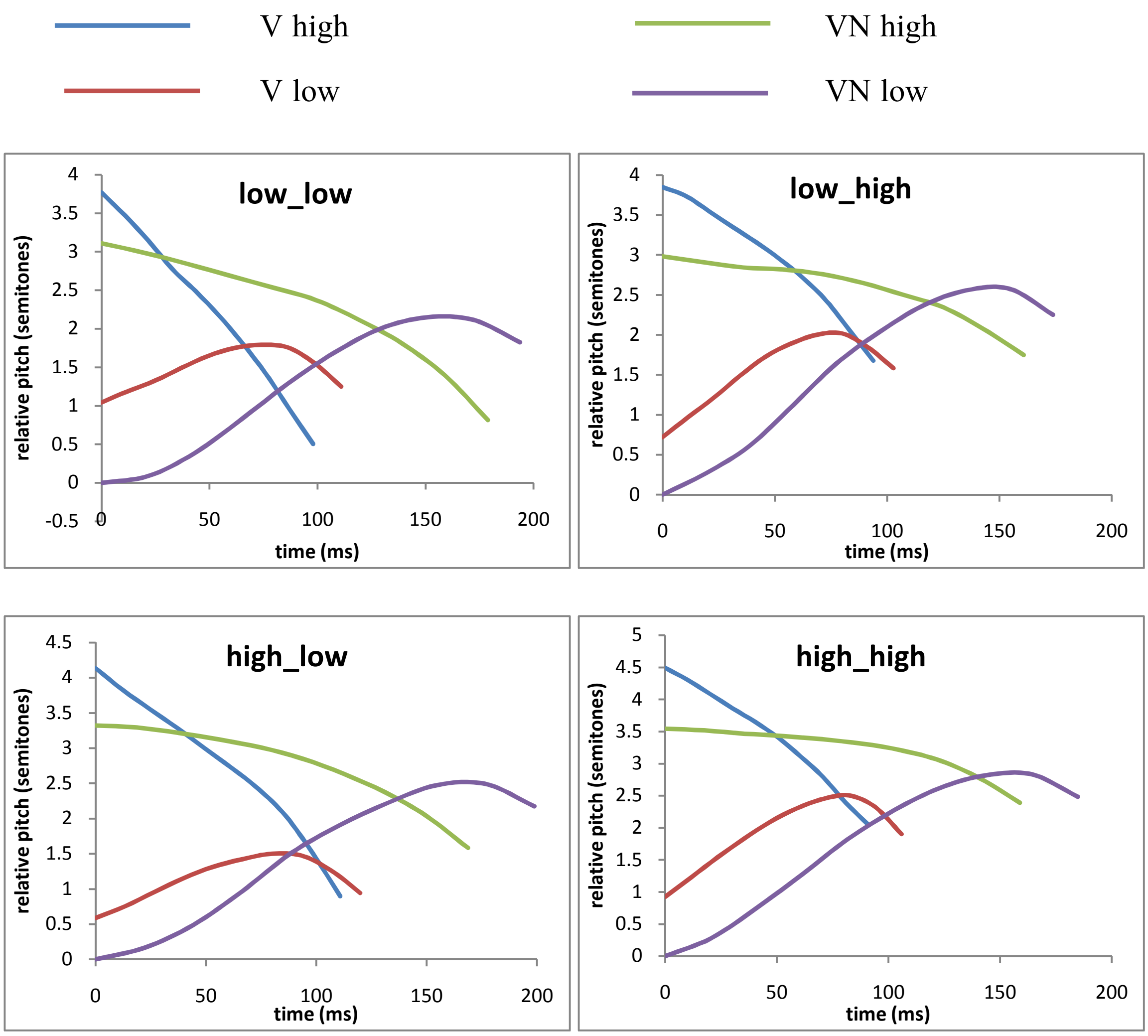
4.3.6.1 Acoustic analysis of the monosyllabic fall

When words are spoken in isolation, the utterance final environment contributes to the amount of fall on a word. In order to remove this effect, and using the same methodology as outlined in section 4.2 above, words were elicited in four different frames in order to get the combinations of both low and high tone words both before the substitution word and after the substitution word. The frames used were:

- *nò ... dè* 'this is a ...'
- *hái ... dè* 'that is a ...'
- *nỳ: ... t^hóŋgik* 'he (near) sees a ...'
- *hái ... t^hóŋgik* 'he (far) sees a ...'

Methodology for creating mean pitch melody graphs normalised for speaker and duration was the same as outlined in section 4.3.1. However, instead of calculating semitones with respect to a speaker's mean F0, for each utterance semitones were calculated relative to the starting F0 of the utterance. The first word of each frame acts as a reference point against which the pitch on the substitution word can be compared. This normalises the measurements of each frame. For comparison between the different frames, zero semitones was taken to be the lowest mean pitch measurement for each frame. Below are graphs showing the mean pitch melodies for each frame:

Figure 4-15: mean pitch melodies for monosyllabic words in frames



The following observations about the mean pitch melodies can be noted. Regardless of the frame, the pitch melodies retain their basic shape for each underlying tone and syllable rhyme. The four-way distinction of the starting pitch is maintained regardless of the tone of the preceding word, and there is still a slight fall at the end of the frame word regardless of whether the following word has high tone or low tone. However, the amount of fall for high tone words is less when followed by a high tone word than when followed by a low tone word.

One indication that the fall for V-high is of a different nature from the other falls is a significant statistical difference in the amount of fall compared with the falls for other pitch melodies. An ANOVA was carried out to investigate this, with independent variables

- word type (V-high, V-low, VN-high, VN-low)
- underlying tone of following word (high or low)
- underlying tone of preceding word (high or low)

The following are the results of the ANOVA (significance level $p < 0.05$), with boxplots (showing median, inter-quartile range, and whiskers to 1.5xinter-quartile range), and table of means and standard deviations.

Table 4-16: ANOVA for effects on amount of fall

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
word type	935	3	340	<0.0005	•
tone following	935	1	18.8	<0.0005	•
tone preceding	935	1	0.435	0.510	

Figure 4-16: amount of fall for low tone following

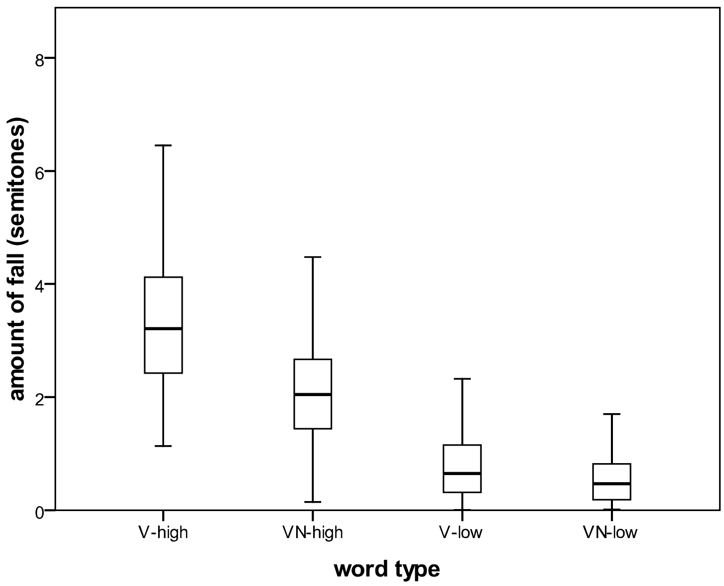


Table 4-17: Mean and s.d. of fall for low tone following

type	*mean (semitones)*	*s.d.*	*N*
V-high	3.31	1.13	110
VN-high	2.21	1.10	142
V-low	0.80	0.70	106
VN-low	0.58	0.51	119

Figure 4-17: amount of fall for high tone following

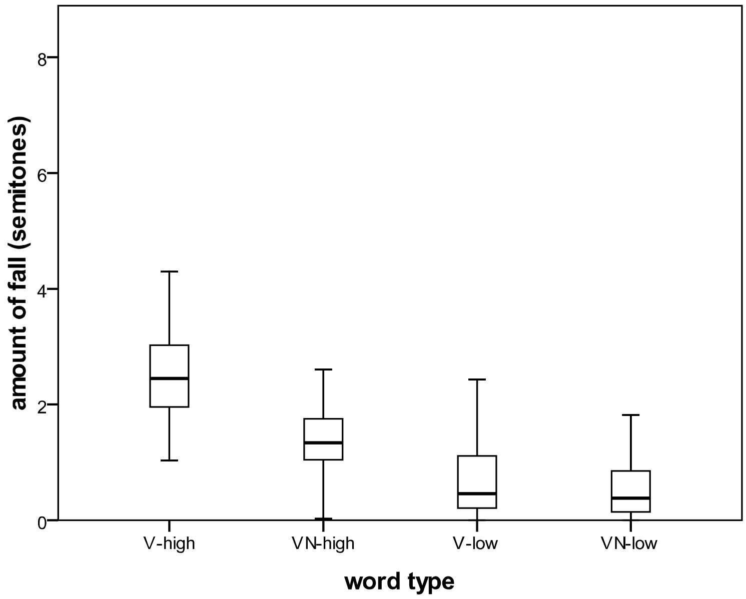


Table 4-18: Mean and s.d. of fall for high tone following

type	*mean (semitones)*	*s.d.*	*N*
V-high	2.58	1.10	122
VN-high	1.41	0.61	129
V-low	0.75	0.70	101
VN-low	0.59	0.61	106

From the above ANOVA results, it can be seen that the tone of the preceding word does not have a significant effect upon the amount of fall, however the tone of the following word does, with the falls for words with underlying high tone being less when followed by a high tone word. Because the difference in the amount of fall is significant both for word type and for tone of following word, the measurements were separated according to the tone of the following word and for each group a post-hoc Scheffé test was carried out to compare the amount of fall for each word type (significance $p < 0.05$).

Table 4-19: Scheffé post-hoc test for amount of fall for low tone following

type	VN-high		V-low		VN-low	
	P	*sig.*	*P*	*sig.*	*P.*	*sig.*
V-high	< 0.0005	•	< 0.0005	•	< 0.0005	•
VN-low	< 0.0005	•	0.325			
V-low	< 0.0005	•				

Table 4-20: Scheffé post-hoc test for amount of fall for high tone following

type	VN-high		V-low		VN-low	
	P	*sig.*	*P*	*sig.*	*P.*	*sig.*
V-high	< 0.0005	•	< 0.0005	•	< 0.0005	•
VN-low	< 0.0005	•	0.492			
V-low	< 0.0005	•				

Regardless of whether the tone of the following word is high or low, there is no significant difference between the amount of fall for V-low and the amount of fall for VN-low. For both of these word types the fall is small (on average less than a semitone) and sometimes there is no fall; the pitch melody for these word types can thus be taken as a rising pitch melody. The fall for V-high is significantly greater than any of the other falls regardless of the tone of the following word. This is one indication of the different nature of the fall for V-high compared with the other falls. The fall for VN-high, whilst less than for V-high, is greater than for V-low and VN-low. However it is noticeable that when there is a high tone word following, the VN-high fall only comes right at the end of what is a predominantly level pitch melody and that there is sometimes no fall at all. This is different from V-high where the fall is always present. Maddieson (1978) defines a level tone as ‘one for which a level pitch is an acceptable variant'. By this definition, the VN-high pitch melody is a realisation of a level tone, but the V-high pitch melody is not.

4.3.6.2 Simple noun phrases

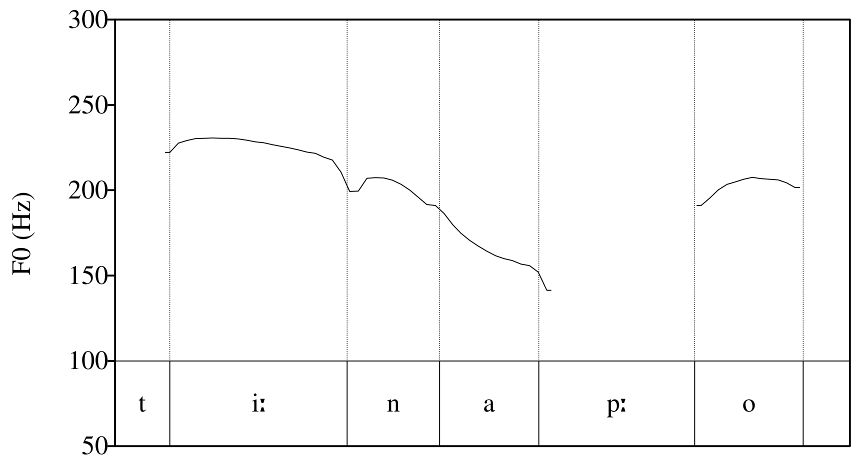
This section describes the pitch melody of a monosyllabic noun when it is in a simple noun phrase, giving further evidence of the lack of fall on VN-high, VN-low and V-low, but the presence of a fall on V-high. Because a noun phrase is left headed, any effect from a phrase boundary can be eliminated by following a noun with an adjective.

Each of the noun phrases described below was elicited in the sentence: *dàŋ ŋàj ... tʰóŋdẓu* "yesterday I saw a ...". The phrases were elicited from five different speakers and checked for accuracy. All show the same patterns described below.

VN-high noun followed by adjective

The following are two examples of a high tone VN noun followed by a low tone adjective:

tíː nàppo 'black foal'



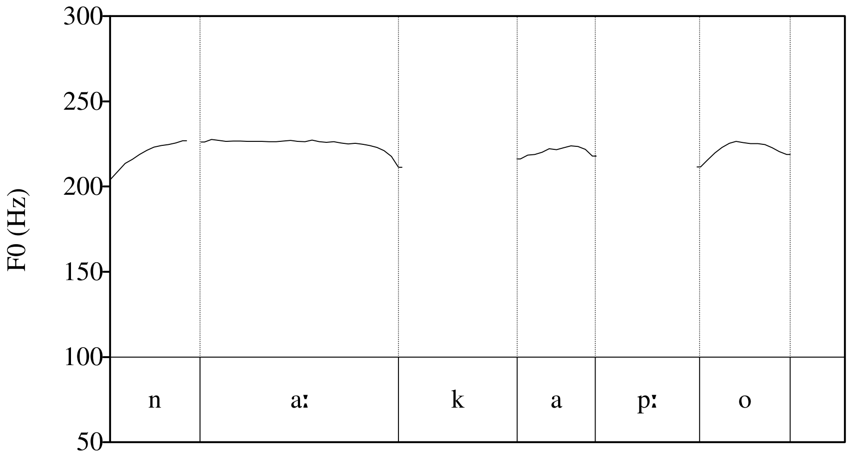
ná: nàppo ‘black blue-sheep’

In both examples the pitch remains high and level across the syllable rhyme of the noun, dropping only as the articulators transition between the vowel of the rhyme and the consonant which follows. The pitch drops during the onset to the first syllable of the adjective and only reaches the target low at the end of the first syllable rhyme. The second syllable of the adjective is high, which is the case for all disyllabic nouns and adjectives regardless of tone.

The sustained high pitch of a VN-high noun illustrates that the fall associated with VN-high in the frames of the section above is a phrase boundary fall. Phrase internally the fall is not present, and the pitch melody is high-level.

The following is an example of a high tone VN noun followed by a high tone adjective. Here the pitch stays level and high throughout the phrase, which is a further illustration that the pitch melody of VN-high is high and level.

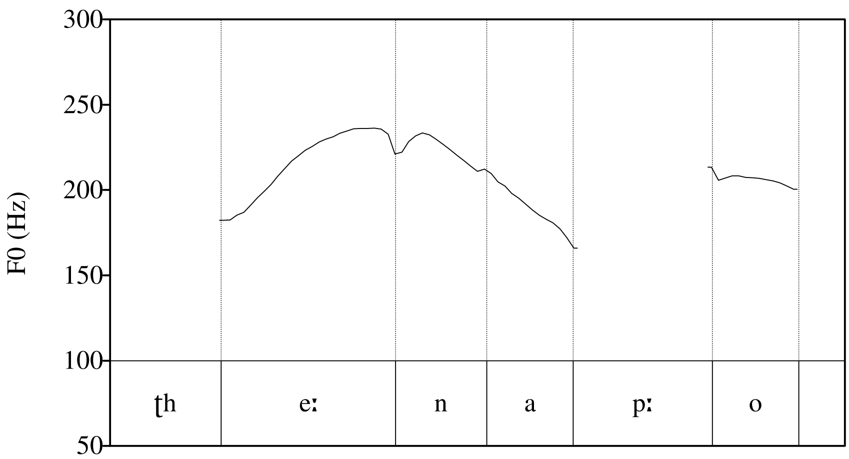
ná: káppo 'white blue-sheep'



VN-low noun followed by adjective

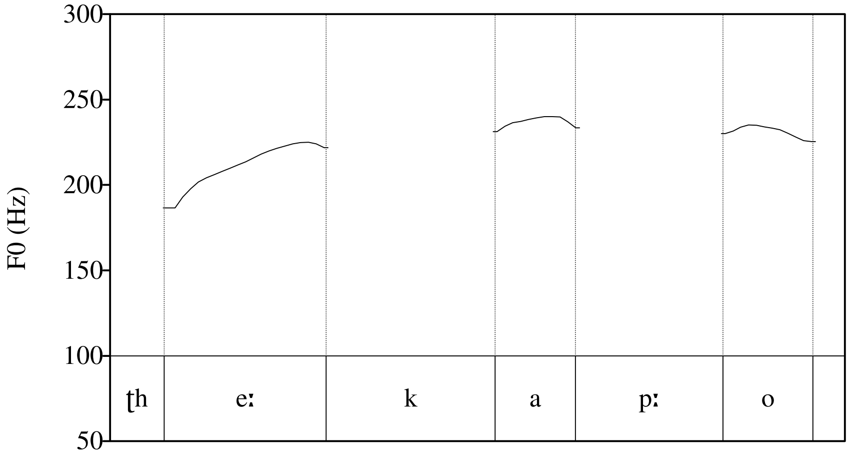
A low tone VN noun followed by a low tone adjective is illustrated in the next example. The pitch melody across VN-low rises throughout the syllable rhyme. The pitch starts falling during the onset to the adjective, only reaching its target low at the end of the first syllable. The second syllable of the adjective is high.

ʈʰè: nàppo 'black mule'



The following is an example of a low tone noun with VN shape, followed by a high tone adjective. Again the pitch melody for VN-low rises throughout the syllable rhyme. The pitch melody on the adjective is high throughout.

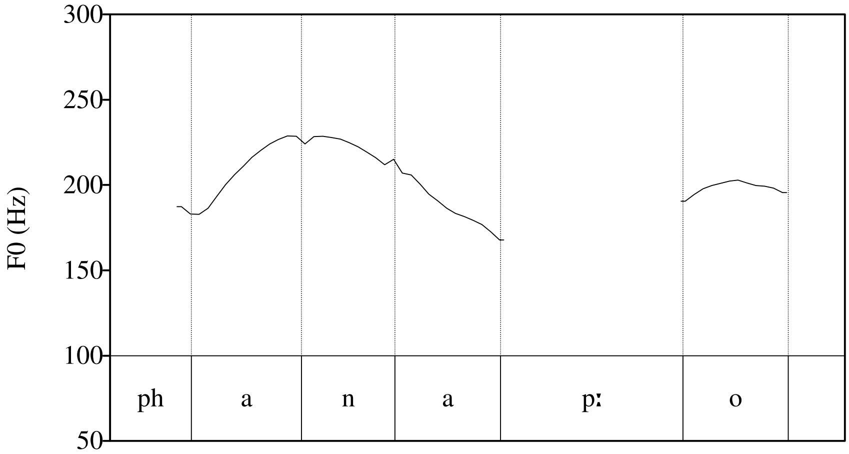
t̪ʰèː káppo ‘white mule’



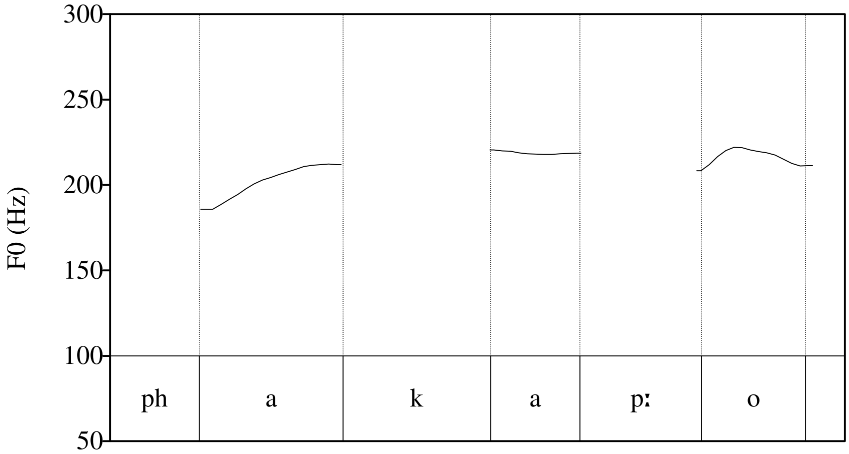
V-low noun followed by adjective

As with VN-low, the pitch melody for V-low within a phrase is a rising melody, regardless of the tone of the following word. The following are examples of V-low words followed by low tone and high tone adjectives. The patterns are the same as for VN-low.

pʰàː nàppo ‘black cow’



pʰà: káppo ‘white cow’



V-high noun followed by adjective

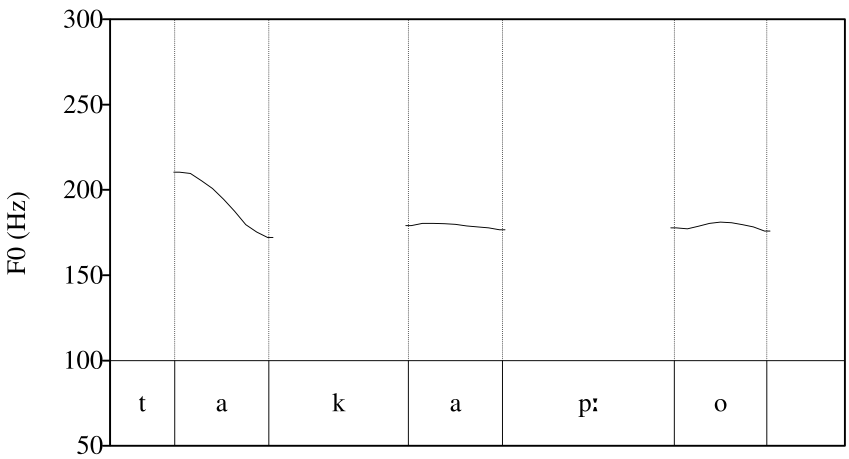
The next examples show V-high nouns followed by low tone adjectives. The pitch is falling across the noun. This is different from the VN-high examples above, where the pitch is level, and is evidence that the fall for V-high is of a different nature from the falls that may occur on other types of monosyllabic nouns. Whereas the falls on VN-high, VN-low and V-low all disappear as soon as the word is in a phrase, the fall on V-high remains.

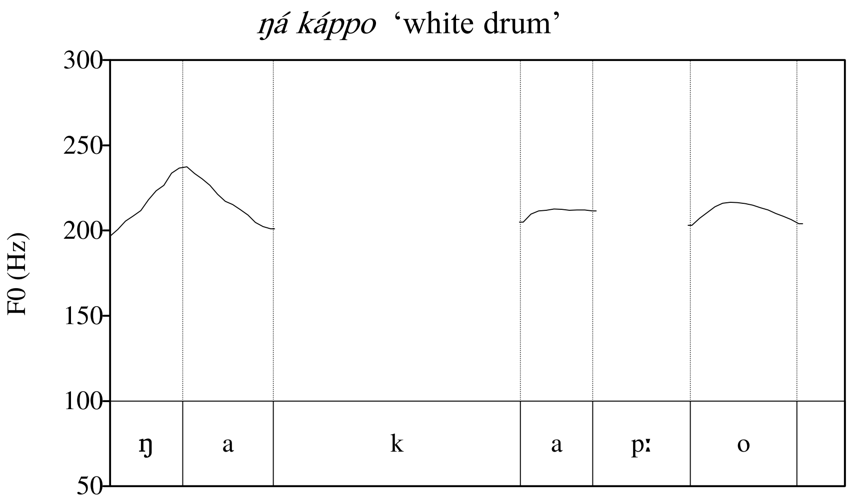
ŋá nàppo ‘black drum’

tá nàppo ‘black horse’

Even more interesting is the pitch melody across a V-high noun followed by a high tone adjective. Not only is there a fall across V-high, but the pitch for the high tone adjective is closer to the finishing pitch of the V-high fall than it is to the high pitch of V-high. For VN-high, which is illustrated above as having a high-level pitch melody, the examples showed the pitch at a constant level across the noun plus high tone adjective. However for V-high, the pitch falls across the noun then levels out for the high tone adjective.

tá káppo ‘white horse’





The above acoustic analysis and examples of noun plus adjective are evidence that the falling pitch melody of a high tone light syllable word is different from the falls which can occur on other monosyllabic words. The other falls can be explained as being a boundary effect. They are there for words in isolation and they are there to a lesser degree between a noun phrase and a verb phrase. However, as soon as the words are within a phrase the fall disappears. The fall for a high tone light syllable word is there at all times, regardless of whether the word is in isolation, at the end of a phrase, or within a phrase. There are two possible explanations for this. The first is that it is a phonetic enhancement to contrastive vowel length, e.g.

ɕá	'meat'
ɕáː	'deer'

The second is that the fall is a realisation of a falling tone pattern. The sections that follow show that the fall is a realisation of a falling tone pattern.

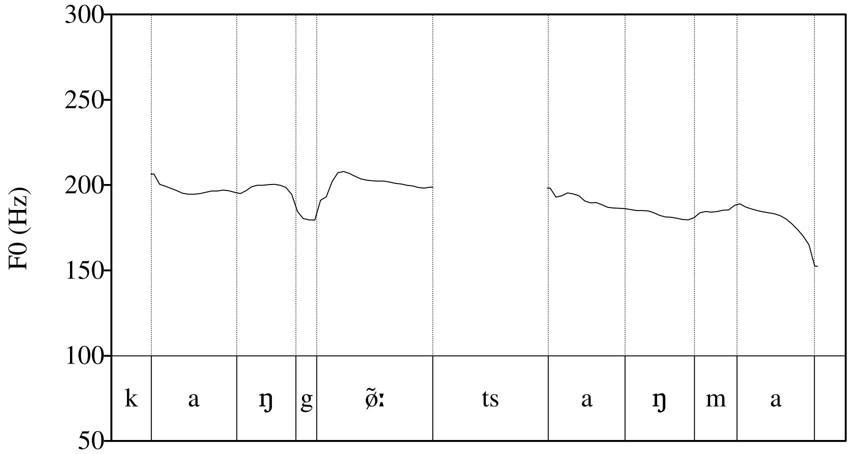
4.3.6.3 Further evidence for monosyllabic falling tone pattern

Disyllabic falls

Section 4.3.3 above describes the disyllabic pitch melodies. As with monosyllabic pitch melodies, in isolation all disyllabic pitch melodies end with a fall. However when they are in a phrase such as the noun phrases above, the fall disappears. The following are examples of the pitch melodies for disyllabic nouns followed by an adjective. The adjectives are both high and low tone, and the nouns have both heavy and light second syllables. In every case there is no fall on the second syllable of the noun.

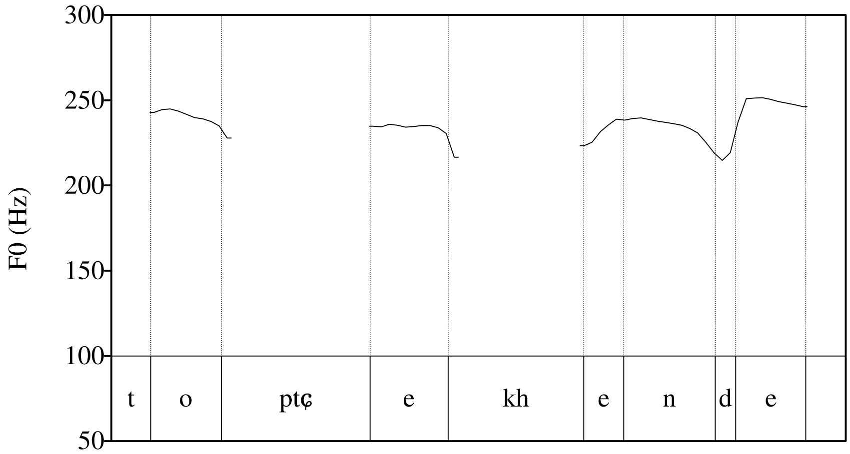
- A high tone noun with heavy second syllable followed by a high tone adjective

káŋgøn tsáŋma ‘clean clothes’



- A high tone noun with light second syllable followed by a high tone adjective

tóptɕe kʰénde ‘cheap food’



- A high tone noun with light second syllable followed by a low tone adjective *kʰá:lo dʑìtto* ‘heavy load’

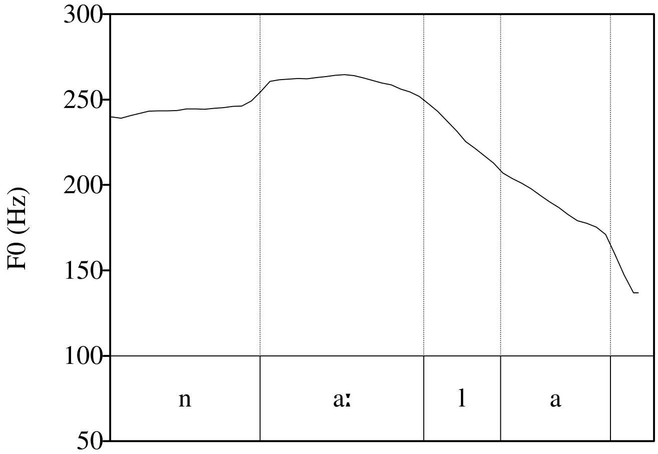
Comparing the patterns above with the case of monosyllabic words, if the difference between level and falling pitch for monosyllabic high tone words was a phonetic enhancement to the vowel length, then it might be expected that for disyllabic words the fall would be retained for words with second syllable short vowel. However, for disyllabic words the fall is lost in all cases. Thus the lack of disyllabic fall is evidence in support of the fall on monosyllabic V-high words being the realisation of a falling tone pattern.

Morpheme boundaries

Further evidence that the monosyllabic V-high fall is the realisation of a falling tone pattern rather than a phonetic enhancement to the length distinction is found at morpheme boundaries. The examples below show the pitch melody when the dative/locative morpheme /-la/ or the plural morpheme /-wa/ is added to monosyllabic high tone open syllable words. Both morphemes are underlyingly toneless and are realised with low pitch.

In the case of words with long vowels, high level pitch is sustained throughout the root word and then falls through the dative/locative or plural morpheme, as the following examples show:

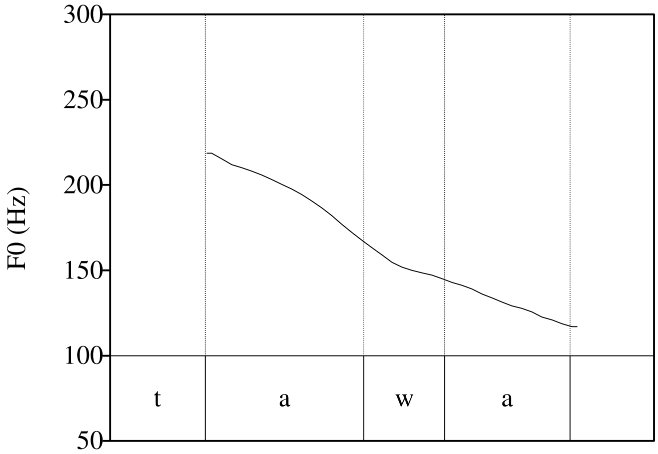
ná:la 'blue-sheep.LOC'



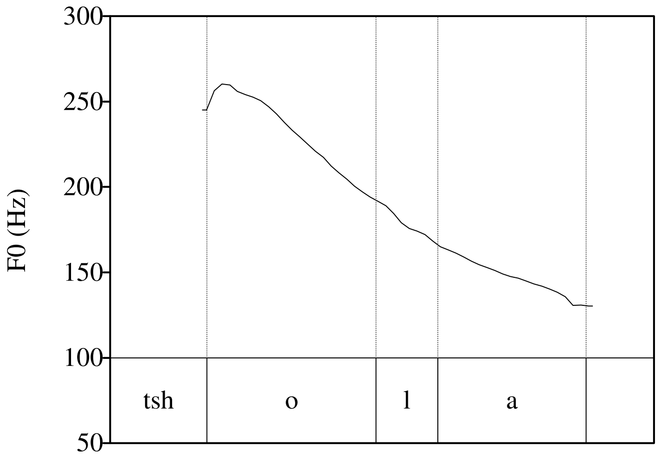
mó:wa 'old woman.PL'

For words with short vowels, the pitch falls during the root word.

táwa 'horse.PL'

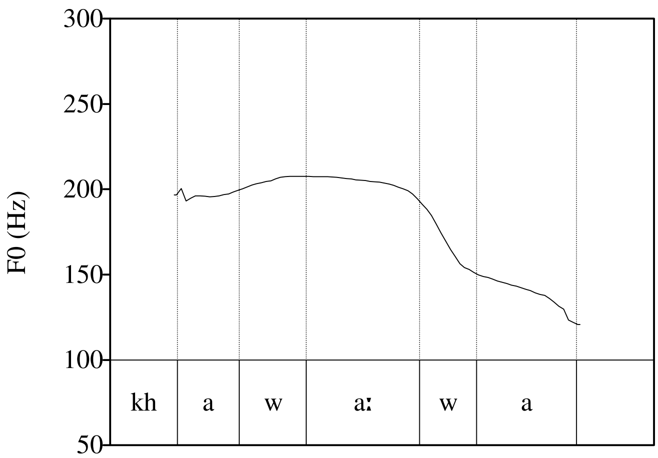


tsʰóla ‘lake.LOC’

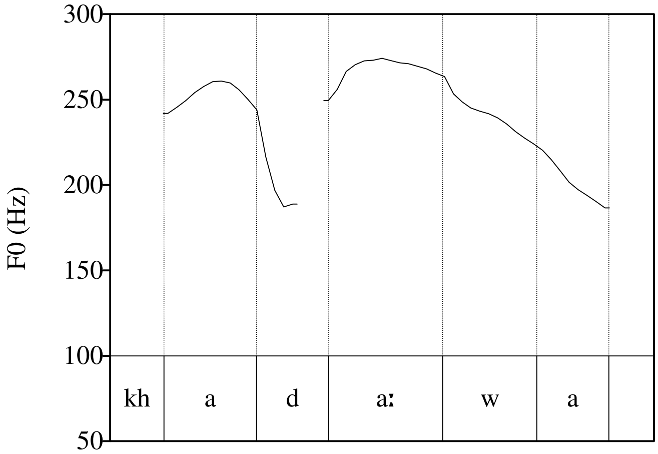


If the difference between level and falling pitch was an enhancement to the length distinction, then it might be expected that disyllabic words would show the same difference in pitch when the plural morpheme is added. But they don’t. The following examples show that regardless of whether the second syllable of a disyllabic word has a long or a short vowel, the pitch across that syllable remains high and level, with falling pitch on the plural morpheme.

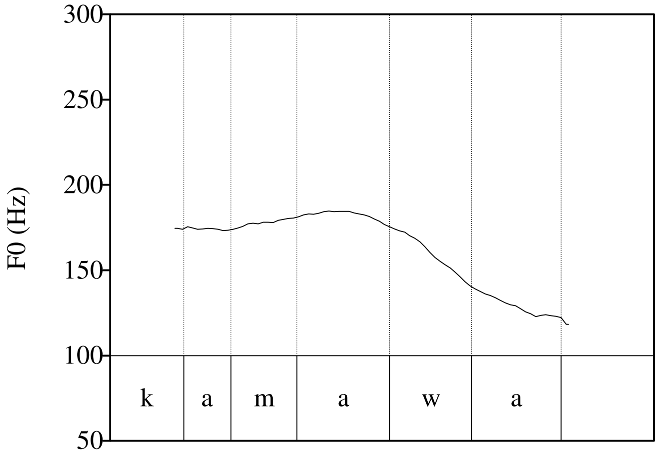
kʰáwaːwa ‘telephone.PL’



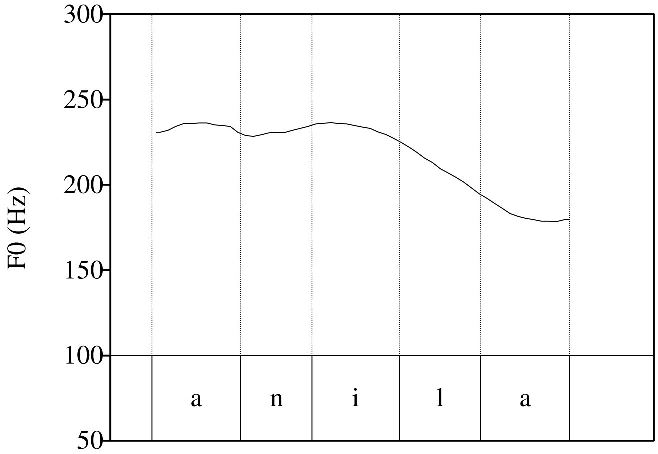
kʰáda:wa ‘silk.scarf.PL’



káma:wa ‘star.scarf.PL’



ánila ‘aunt.DAT’



This is further evidence that the monosyllabic V-high fall is a realisation of a falling tone pattern rather than an enhancement to the length distinction.

4.3.6.4 Conclusion of pitch analysis for phrases

The conclusion of the above analysis of the pitch across nouns, is that the difference in falling versus level pitch melody for monosyllabic high tone light syllable versus heavy syllable words cannot be explained by saying that it is a phonetic enhancement to the length distinction. The fall is best analysed as being a realisation of a falling tone pattern, which for now will be referred to as HL. Just as the fall for light syllables does not disappear in phrases or when morphemes are added, neither does the rise disappear. This suggests that the rise is a realisation of an LH tone pattern. Thus for monosyllabic nouns, the possible tone patterns are H, HL, and LH.

These three tone patterns are realised with four different starting pitches. Tone pattern HL has a higher starting pitch than H. This higher starting pitch can be explained by saying that starting at a higher pitch gives greater phonetic space for the fall. The LH rise for monosyllabic words ends at approximately the same pitch for both heavy and light syllables. However for light syllables the starting pitch is higher. This can be explained by saying that the shorter duration means the starting pitch must be higher in order to reach high pitch by the end of the word. Interestingly it appears that reaching high by the end of the word is more important than starting low at the beginning of the word.

4.3.7 Phonetic representation of pitch

The pitch melodies of words in Tibetan languages have often been represented using Chao letters (see section 2.2 for a description of the Chao system). Any system of transcription, be it for the segments or the prosodic features, is an abstraction from what is actually uttered, thus there is a question of how much phonetic detail to include. In terms of using Chao letters to describe the pitch melodies of words in Walungge, the following principles have been adopted.

- The slight differences in pitch such as the very slight difference caused by the height of the vowel (see section 4.3.4), or the very slight differences caused by the preceding consonant will not be included.

- For disyllabic low tone words, the high pitch of the second syllable is not always reached at the start of the second syllable. If the second syllable is light, the high pitch is reached at the start if the preceding consonant was voiceless and slightly after the start if it was voiced, as a result of peak delay. This slight delay will not be transcribed. However, for heavy syllables the pitch of the second syllable starts significantly lower and the peak is not reached until the latter half of the rhyme. This will be transcribed.

- The falls that are there for words in isolation but which disappear once the words are put into phrases will not be transcribed.

- In general other differences will be transcribed.

On the basis of the mean pitch melodies for each tone and syllable rhyme, and the statistical analysis indicating which differences in pitch are significant, the monosyllabic pitch melodies can be described using the Chao system, as follows:

Table 4-21: monosyllabic pitch melodies described using Chao system

underlying tone	*syllable rhyme*	*pitch melody*
high	V	52
high	VN	44
high	VP	53
low	V	24
low	VN	14
low	VP	24

Similarly, the pitch melodies across polysyllabic words can be described:

Table 4-22: monosyllabic pitch melodies described using Chao system

number of syllables	*tone*	*2nd syllable weight*	*pitch melody*
2	high	light/heavy	44 44
2	low	light	11 44
2	low	heavy	11 344
3	high	light/heavy	44 44 21
3	low	light	11 44 21
3	low	heavy	11 344 21
4	high	light/heavy	44 44 21 11
4	low	light	11 44 21 11
4	low	heavy	11 344 21 11

By the nature of the system, Chao letters have an element of flexibility, and this flexibility needs to be used when describing the pitch across morpheme boundaries. For example, in [tawa ^{53} ^{21}] 'horse.PL' no significance should be attached to the fact that the pitch of *tá* 'horse' has been described as [53] rather than [52]. It has been described as [53] simply because the pitch on the following word falls even further. It should also be noted that Chao letters are describing the pitch across syllable rhymes. The pitch across an intervocalic sonorant, for example, is not shown.

4.4 Phonation

4.4.1 Introduction to phonation

Before considering differences in phonation, it is first necessary to briefly describe how the vocal folds vibrate. The vocal folds can be held closer together or wider apart, thus opening or closing the glottis (the space between the vocal folds). For a voiced sound, as the vocal folds are brought together air pressure builds up behind the closure until the pressure difference on either side of the opening is sufficient for the vocal folds to be forced apart as the air bursts through. The change in pressure results in the vocal folds being pulled back together, then pressure builds up and the cycle starts again.

A difference in phonation depends, at least in part, upon the way in which the vocal folds are held, including the tension and the openness. Laver (1980) proposes that one of the main characteristics of different phonation types is tension, with creaky phonation having greater laryngeal tension and breathy phonation having less laryngeal tension. Ladefoged (1971) and Gordon & Ladefoged (2001) propose a continuum based on openness of the vocal folds, which is depicted by the following diagram:

Figure 4-18: continuum of phonation types

	most open ⟵			⟶	most closed
phonation type:	voiceless	breathy	modal	creaky	glottal closure

If the vocal folds are held sufficiently wide apart, then the air can move through the glottis freely, without causing the vocal folds to vibrate open and closed. The resulting sound is voiceless. At the other extreme, the vocal folds are held sufficiently closed that they cannot vibrate open and closed. Normal voicing occurs midway between these two extremes, i.e. modal phonation. Breathy voice occurs, at least in part, by the vocal folds being held further apart than normal but still close enough for vibrations to occur, so more air can escape through the glottis. Similarly, creaky voice occurs, at least in part, by the vocal folds being held more closely together than normal.

As phonation is a continuum, different speakers may be at different points on this continuum. Thus certain speakers' normal phonation may be breathier than that of other speakers. Females tend to have breathier phonation than males. For a language that contrasts breathy and modal phonation, it is the relative breathiness that is significant.

4.4.2 Acoustics of phonation

There are potentially a number of acoustic correlates of phonation (Gordon & Ladefoged (2001)). The two that will be considered here are closed quotient (the percentage of time that the vocal folds are closed per cycle) and spectral tilt (how much the intensity decreases as the frequency increases).

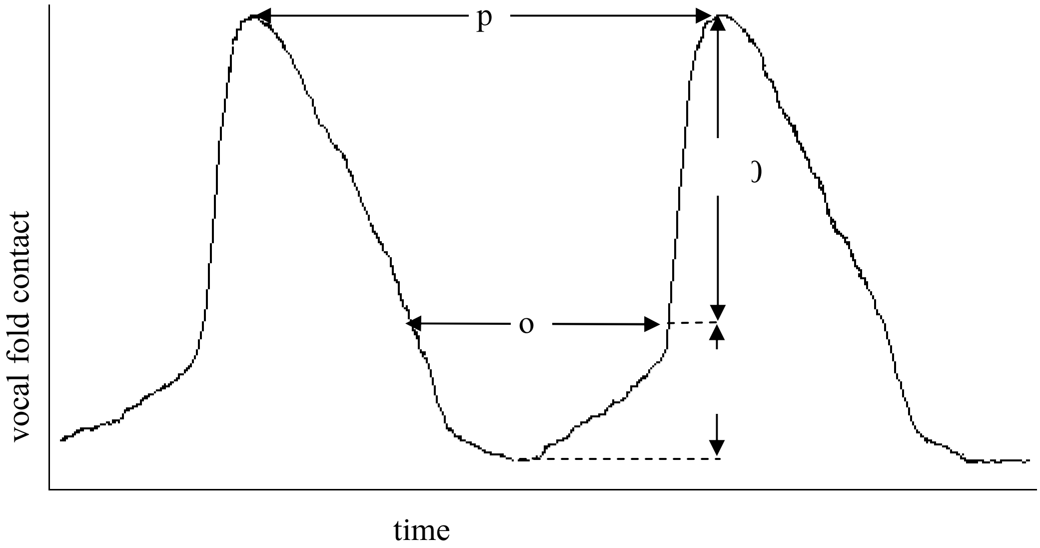
Closed quotient

As discussed above, breathy phonation, at least in part, correlates with the vocal folds being held further apart than normal. Although for breathy phonation the vocal folds might never completely close and air is still able to escape through the glottis, the nature of the vibrating vocal folds means it is still possible to refer to an open phase and a closed phase in the glottal cycle. One of the correlates of breathy phonation, is that for every cycle of the vocal folds opening and closing they are closed for a lower percentage of the time. This is known as the closed quotient (CQ). A lower closed quotient correlates with a greater degree of breathiness.

The closed quotient can be measured by a technique known as *electroglottography* or *laryngography.* Electrodes are positioned on the neck, on either side of the larynx, and a very weak current is passed between them. The amount of current that passes between them depends upon the surface area of contact of the vocal folds, which can be taken as an indication of how open or closed the vocal folds are. As the current flow increases and decreases, the resistance can be measured. The piece of equipment which has the electrodes and the circuitry for measuring the signal is a *laryngograph.* The laryngograph can be connected to one channel of a stereo recorder and a microphone to the other channel. Thus the recorder can record both the sound wave and the glottal wave simultaneously. Both the sound wave and the glottal wave can be analysed using Praat.

Figure 4-19 below shows how the closed quotient can be calculated from the glottal waveform.

Figure 4-19: calculation of closed quotient from glottal waveform



p = duration of one period

o = duration of open phase

closed quotient (CQ) = 100 x (1 - o/p) %

In the above waveform, the peaks correspond to the most closed position in the vocal fold cycle, and the valleys correspond to the most open position in the cycle. The downward slopes correspond to the vocal folds opening, and the upward slopes correspond to the vocal folds closing. Because the amount of vocal fold contact is a continuum, there is a question of where to take the boundary between the closed phase and the open phase. The method of determining the boundary that will be used here is to take the boundary at 30% of the difference between the maximum and minimum vocal fold contact. Whilst 30% is an arbitrary cut-off point, previous studies have used this method as a reliable means of differentiating between phonation types (Davies, Fourcin et al. (1986), Lindsey, Hayward et al. (1992), Watkins (2002)).

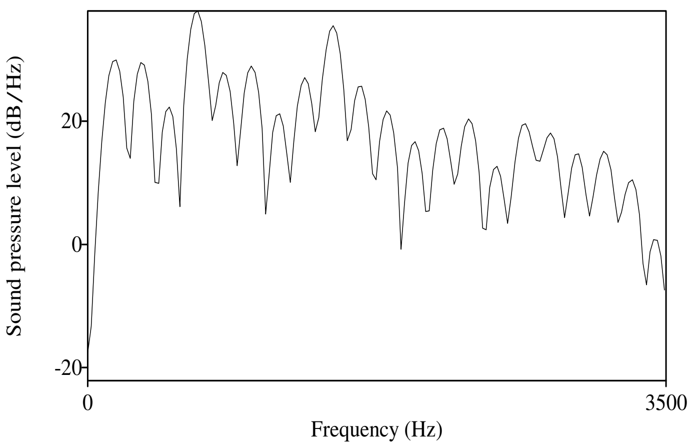
Phonation is relative, and absolute values of CQ can vary considerably from speaker to speaker. However, as extremely approximate measurements, a CQ of 50% might indicate modal phonation, a CQ of 35% might indicate breathy phonation, and a CQ of 65% might indicate creaky phonation (Johnson (2003)).

Spectral tilt

Spectral tilt compares the intensities of the different harmonics in a voiced sound. For breathy phonation, where the vocal folds are held only loosely together, as the frequency increases the intensity decreases. This can be quantified by measuring the amplitude of the fundamental frequency (F0 = H1) and the amplitude of the second harmonic (H2), and then subtracting the amplitude of H1 from the amplitude of H2. This difference in amplitudes will be termed H2-H1. The breathier the phonation, the smaller H2-H1 will be; the creakier the phonation, the greater H2-H1 will be. (It has been chosen to calculate H2-H1 rather than H1-H2 because then a lower value will correlate with a lower CQ, and a higher value of H2-H1 with higher CQ, rather than the opposite correlation.) Gordon & Ladefoged (2001) describe spectral tilt in greater detail, and give a list of languages where this method of quantifying spectral tilt has been successfully used to differentiate between phonation types. Watters (2002) uses this method to investigate the phonation of the five Tibetan languages in his study.

One of the difficulties of using the difference in amplitude between H2 and H1 is the interference of vowel formants. Consider the spectrum below of the vowel [a], said by a male speaker of Walungge:

Figure 4-20: spectrum of [a]; male speaker



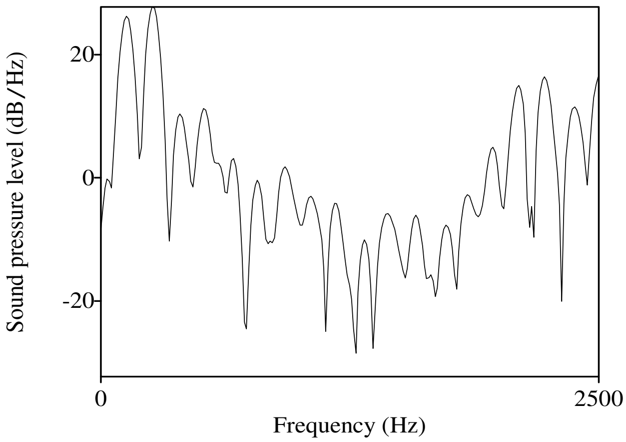
In the above spectrum, the frequencies for H1, H2, and F1 are:

H1: 165 Hz
H2: 330 Hz
F1: 664 Hz

The frequency of F1 is sufficiently high not to interfere with the intensities of H1 and H2. In the case of the vowel [a] this holds for both male and female speakers, and for both high and low tone.

Now consider the spectrum below of [i], for low tone and for a male speaker:

Figure 4-21: spectrum of [i]; low tone; male speaker



H1: 132 Hz

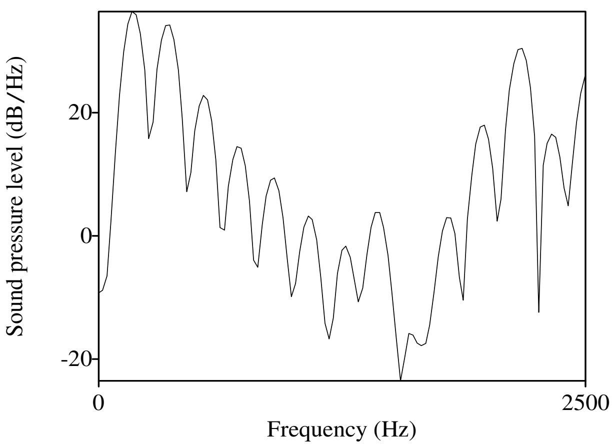
H2: 264 Hz

F1: 274 Hz

In this instance, the frequency of H2 is close to the frequency of F1. The amplitude of H2 has been amplified due to its proximity to F1, which means that H2-H1 is higher than what would otherwise be expected.

The following spectrum is of [i] for the same speaker, but for high tone:

Figure 4-22: spectrum of [i]; high tone; male speaker



H1: 181 Hz

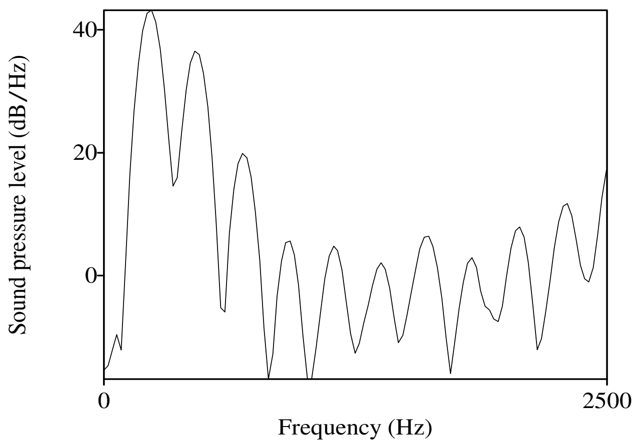
H2: 362 Hz

F1: 296 Hz

In the case of high tone H2 is further away from F1 than it is for low tone, and so the intensity of H2 is not as amplified. Thus for high tone H2-H1 is lower than it is for low tone despite low tone having breathier phonation than low tone.

Similarly F1 of [i] affects H2-H1 for female speakers.

Figure 4-23: spectrum of [i]; high tone; female speaker

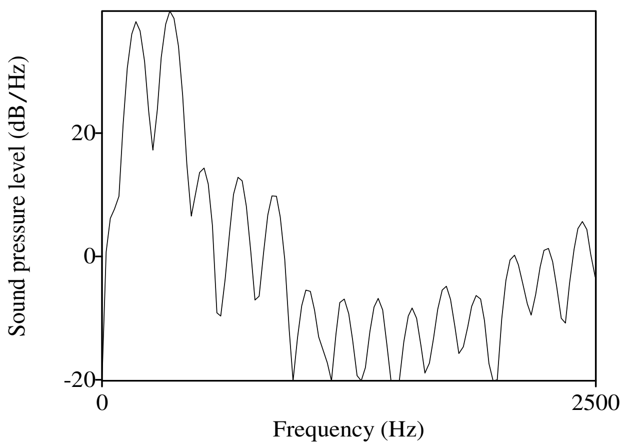


H1: 231 Hz

H2: 342 Hz

F1: 288 Hz

Figure 4-24: spectrum of [i]; low tone; female speaker



H1: 174 Hz

H2: 348 Hz

F1: 298 Hz

For high tone the pitch is higher and H1 is closer to F1. Thus the intensity of H1 is amplified, making H2-H1 lower. For low tone, because the pitch is lower, H1 is further away from F1 and so the intensity of H1 is not amplified. On the other hand, for low tone H2 is close enough to F1 to be amplified. Thus for low tone H2-H1 is higher than for high tone despite the breathier phonation.

In the Walungge data used for this analysis, the only vowel for which there is no influence from F1 upon H2-H1 is [a]. Thus spectral tilt will only be considered for the vowel [a].

4.4.3 Description of closed quotient patterns

Section 4.3.1 above describes how the pitch melodies for monosyllabic words were time normalized and averaged. Similarly, the closed quotient (CQ) was measured at 10 equidistant points across each vowel, time normalized, and averaged. However, because phonation in Walungge is taken to be a property of vowels, rhymes with codas were excluded from the average measurements for CQ. Below are graphs showing the mean pitch melodies for monosyllabic words and the mean CQ across the vowel for monosyllabic words.

Figure 4-25: Monosyllabic mean pitch melodies

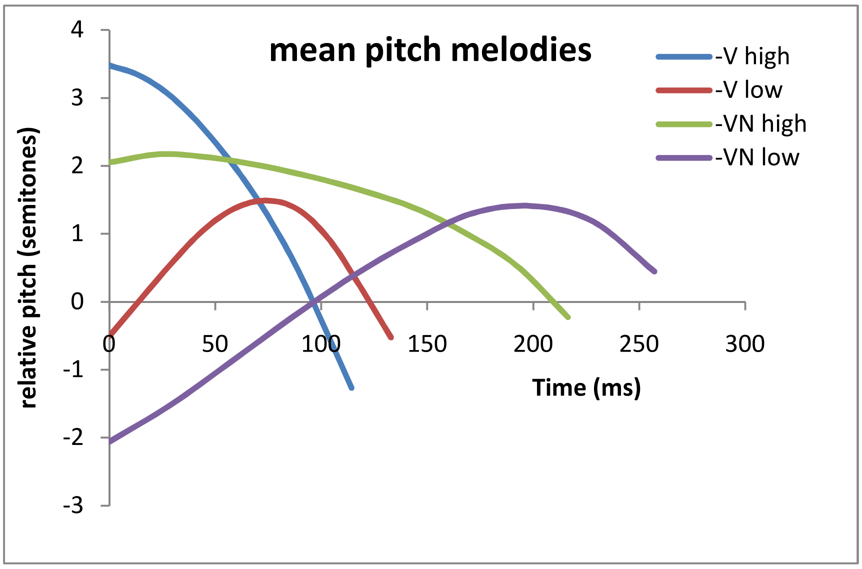
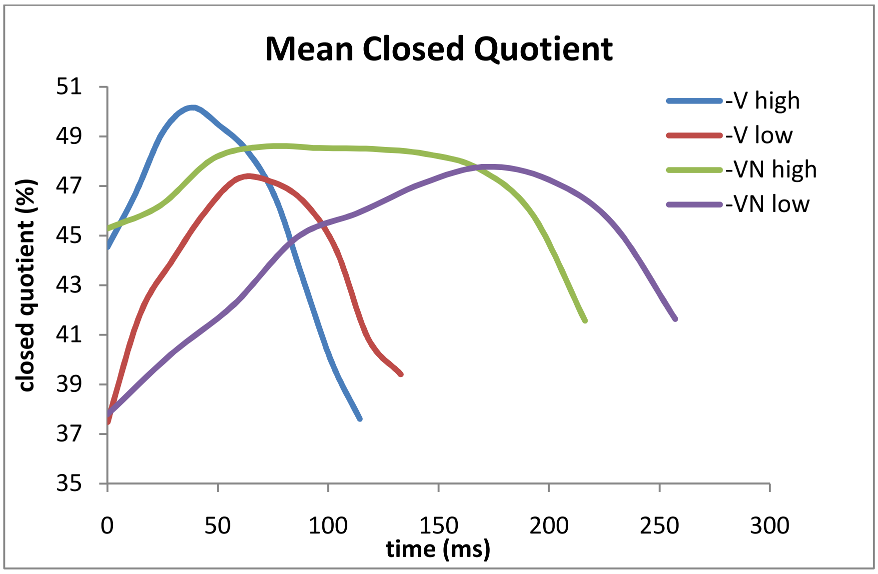


Figure 4-26: Monosyllabic mean CQ



Comparing the above graphs of mean pitch melodies and mean CQ, there are some striking similarities and differences. The main difference is that for the pitch melodies there are four different starting values for pitch, and the differences between these values are statistically significant (see section 4.3.2). For CQ there are only two different starting values, one for high tone and one for low tone. What is particularly striking, however, is that as the vowel progresses all the CQ graphs match in overall shape with the pitch melody graphs. From approximately 50ms into the vowel, V-high has a sharply falling CQ which matches the sharply falling pitch melody. VN-high has a CQ which sustains and then falls, as does its pitch melody. Both V-low and VN-low have a rising-falling CQ, matching the rise and fall of the pitch melodies. For the first 50ms for both high and low tone the CQ is rising, which could in part be due to the effect of the initial consonant.

A similar correlation between pitch melody and CQ is seen with disyllabic words. Below are graphs comparing disyllabic pitch melodies and disyllabic CQ for CVCV words. Because phonation is a property of vowels, the graph for CVCV CQ shows the CQ across the first vowel and the CQ across the second vowel, with a break in the graph for the intervocalic consonant.

Figure 4-27: Mean disyllabic pitch melodies for CVCV words

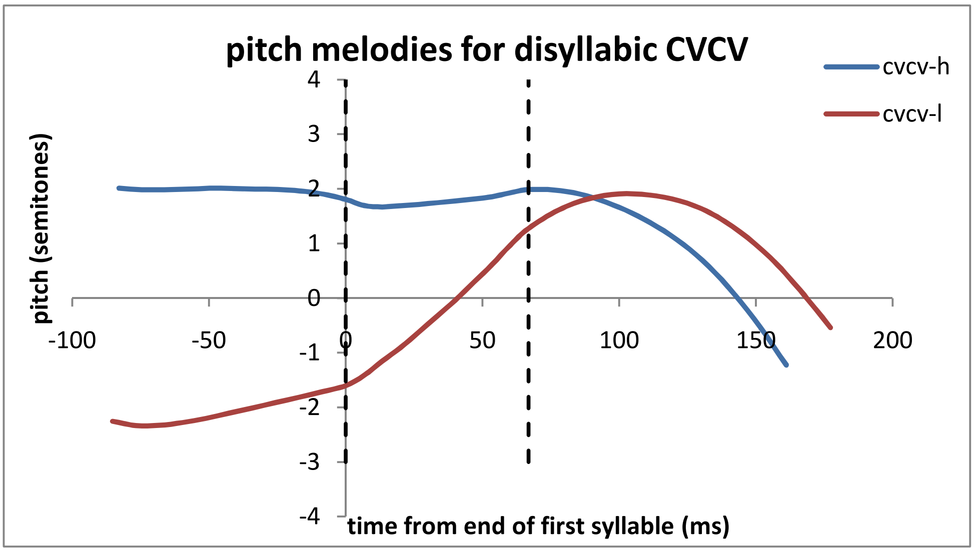
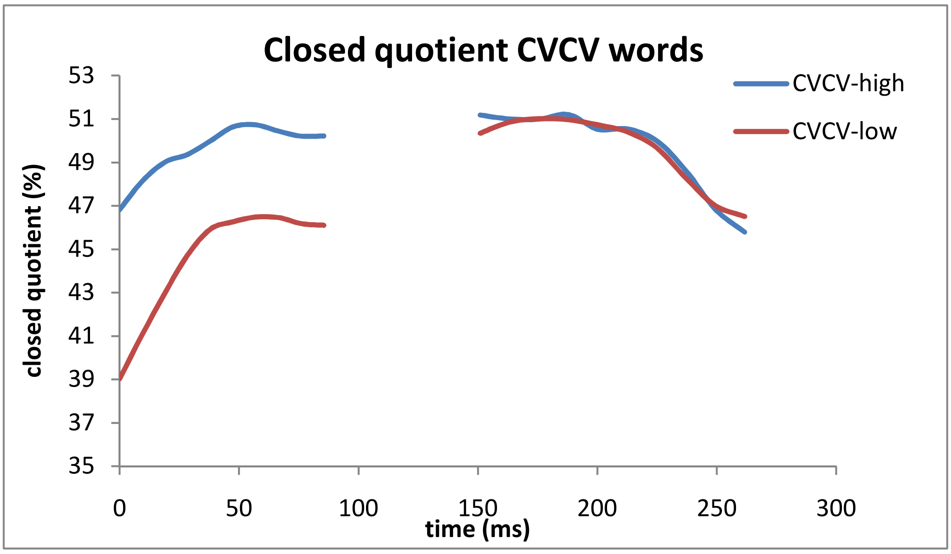


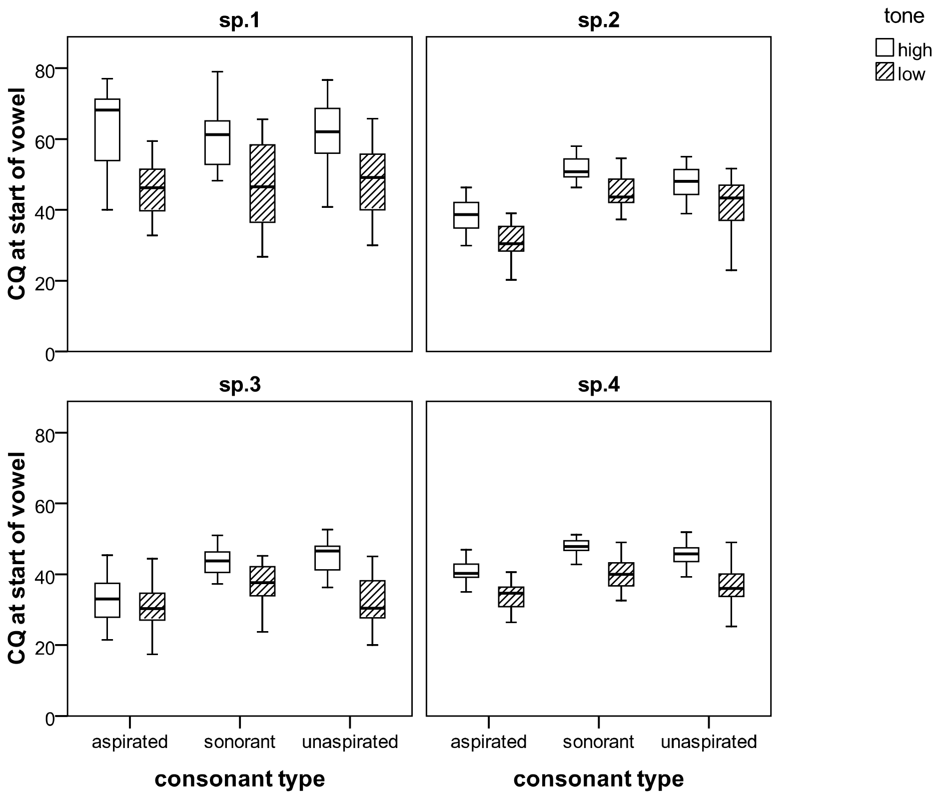
Figure 4-28: Mean CQ for CVCV words



Comparing the graphs of mean pitch melody and closed quotient for disyllabic words, the same similarities and differences exist as for monosyllabic words. From approximately 50ms onwards, the CQ follows the pitch melody. For CVCV high tone words both the pitch melody and the CQ are high for the first syllable and drop during the second. In the case of CVCV low tone words, both CQ and pitch are lower for the first syllable, and high and falling for the second. For the first 50ms the CQ is rising.

Breathiness, and hence CQ, is relative according to speaker, with female speakers generally having breathier phonation than male speakers. Also the initial consonant may have an effect on the breathiness of the following vowel. In particular, aspiration is known to correlate with breathier phonation at the start of the following vowel (Blankenship (1997)). Hence, there is a large degree of variation in the CQ which cannot be seen in the above graphs. Below is a boxplot showing CQ at the start of each first syllable vowel (monosyllabic and disyllabic words all included), separated by speaker and by initial consonant type (aspirated obstruents, unaspirated obstruents and sonorants). The boxplot gives median, interquartile range, and whiskers of 1.5x interquartile range.

Figure 4-29: boxplots of CQ at start of 1st vowel, for each speaker



From the above boxplots it can be seen that whilst there is considerable variation in CQ, for each speaker and consonant type there is a difference in CQ correlating with tone.

4.4.4 Analysis of closed quotient

An ANOVA was carried out in order to consider which factors are affecting the CQ at the start of the vowel. Once the vowel is underway the CQ follows the pitch melody of the word.

The independent variables included in the ANOVA are:

- Tone: it can be seen from the above graphs and boxplots that tone and starting CQ are correlated, with lower CQ correlating with low tone and higher CQ with high tone.
- Initial consonant type: as mentioned above, different consonant types affect the breathiness of the following vowel. In particular aspirated consonants have

spread vocal folds and so give rise to breathier phonation. Thus initial consonant type is divided into aspirated obstruents, unaspirated obstruents, and sonorants.

- Syllable weight: it has been argued in the section on pitch above, that the four monosyllabic pitch melodies are realisations of three surface tone patterns H, HL, and LH. However, there are four significantly different starting pitches, with a difference in pitch not only correlating with the first tone of the surface tone pattern, but also correlating with a difference in syllable weight. Closed quotient is a correlate of breathiness which is a correlate of tone. So clearly closed quotient and pitch are closely linked, and this is seen in the above graphs where from 50ms onwards the CQ follows the pitch. If CQ is completely correlated with pitch, then it might be expected that a difference in syllable weight correlates with a difference in CQ. This is only relevant for monosyllabic words.

- vowel height: higher vowels correlate with slightly higher pitch and lower vowels with slightly lower pitch. Thus vowel height is included as a possible effect upon CQ.

The significance level for the ANOVA is taken to be $p < 0.05$

Table 4-23: ANOVA for effects on closed quotient

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	588	1	115	<0.0005	•
consonant type	588	2	18.4	<0.0005	•
syllable weight	208	1	0.883	0.348	
vowel height	588	6	0.594	0.552	

Both tone and consonant type have a significant effect upon the closed quotient, and hence breathiness. This effect has already be seen in Figure 4-29 above.

The variables which do not have a significant effect on the closed quotient are syllable weight and vowel height. For vowel height, it should be pointed out that although vowel height correlates with pitch, the differences in pitch for the differing vowel heights are only very slight, and not all differences are statistically significant (see section 4.3.4). Thus is it not surprising that vowel height does not cause a significant difference in closed quotient.

From the ANOVA, the result which is interesting to note is that tone has a significant effect on the closed quotient, and hence breathiness, but syllable weight does not. As has already be stated, in Walungge both pitch and breathiness are concomitants of tone and thus there is a close link between pitch and breathiness. If they were completely linked together, then both would always vary together, and in particular syllable weight might be expected to show on the ANOVA as having a significant effect upon the closed quotient. However, it does not. At the very start of a word pitch and breathiness only correlate in so far as they are both concomitants of tone, but there is no further correlation between them and thus pitch further varies without breathiness varying along with it. The most likely explanation for this is that right at the beginning of the vowel the effect of the consonant upon phonation is greater than the effect of the pitch. However, one further explanation is the inter-relationship between pitch and tone at the start of the word. The pitch melodies [52] and [44] are realisations of surface tone patterns HL and H; the pitch melodies [14] and [24] are realisations of surface tone pattern LH; the four separate starting pitches are phonetic effects. Because there is a phonological distinction between H and L, but not between [1] and [2] or between [4] and [5], it could be that the vowel phonation at the start of the word is acting as an enhancement to the phonological distinction rather than the phonetic pitch. Thus a two way phonation distinction is maintained at the start of the vowel even though there are 4 separate starting pitches.

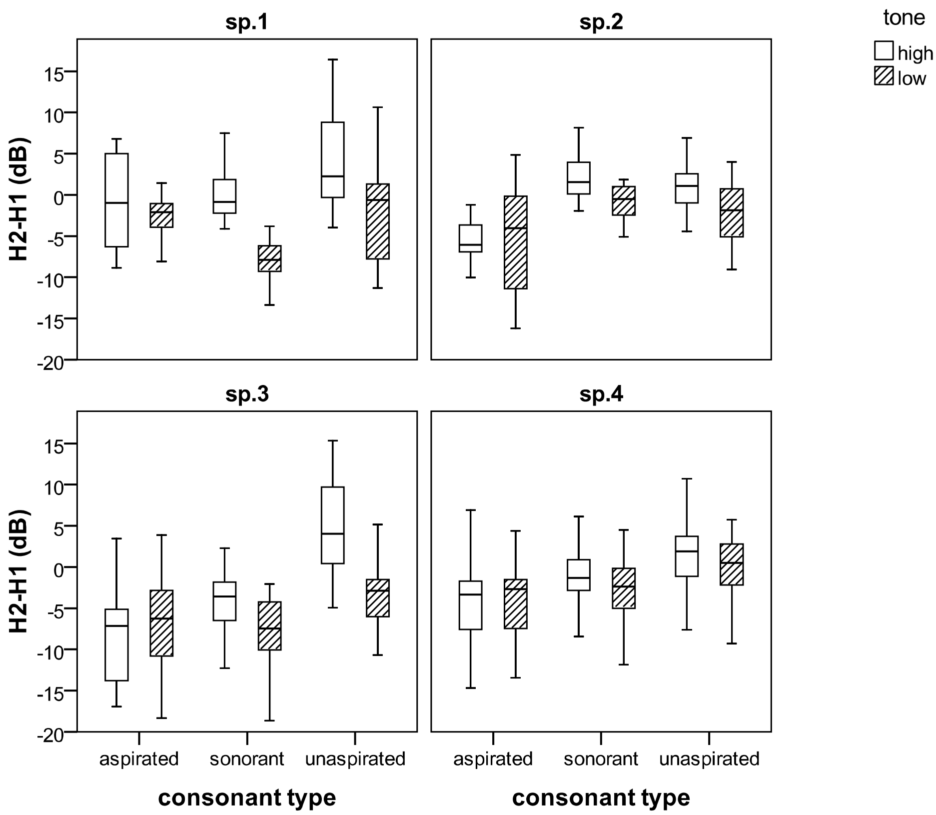
4.4.5 Analysis of spectral tilt

Because of the interference of first formants of vowels upon H2-H1 (the difference in amplitude between H2 and F0) for all vowels except for [a] (see section 4.4.2), a spectral tilt analysis was only carried out for [a].

From the closed quotient analysis above, it can be seen that the difference in phonation correlating with the difference in underlying tone occurs at the start of a vowel. By 50ms into the vowel the closed quotient closely follows the rises and falls of the pitch melody. In order to consider the effect of the underlying tone upon H2-H1, an average power spectrum across a time window needs to be calculated. This time window is taken from the onset of the vowel, and needs to be as small as possible, and in any case less than 50ms because by that time the phonation is following the pitch melody. On the other hand, the time window needs to be great enough to cover at least one full cycle of vocal fold vibration. Thus a spectrum was taken across the first 20ms of the vowel. From the spectrum, the amplitudes of H1 and H2 were measured, and H2-H1 was calculated.

Below is a boxplot showing H2-H1 for the four different speakers, with H2-H1 separated by consonant type and tone.

Figure 4-30: boxplot of H2-H1 for each speaker, consonant type, and tone



Although there is considerable overlap in H2-H1 measurements for high tone and low tone words, in general for all of the speakers for sonorants and unaspirated obstruents H2-H1 is lower for low tone. However, the effect of aspiration is that H2-H1 does not appear to be different for high and low tones. With the exception of aspirated consonants, Figure 4-30 above for H2-H1 is very similar to Figure 4-29 above for closed quotient.

With tone and consonant type as independent variables, an ANOVA was carried out for spectral tilt, with the following results.

Table 4-24: ANOVA for effects on spectral tilt of [a]

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	633	1	21.5	< 0.0005	•
consonant type	633	2	54.7	< 0.0005	•

Both consonant type and tone have a significant effect on spectral tilt, with low tone correlating with a lower value of H2-H1.

Both H2-H1 and closed quotient are considered to be correlates of phonation, with lower H2-H1 correlating with breathier phonation and also lower closed quotient correlating with breathier phonation. In Walungge, tone has a significant effect on both of these measurements, with low tone significantly lowering both H2-H1 and closed quotient at the start of the vowel. The above analysis of H2-H1 reinforces the results that were found from the analysis of closed quotient.

4.4.6 Further discussion

There are a number of studies looking at the phonation across a vowel, for languages where both phonation and tone are separately phonologically contrastive, and for languages where phonation is not contrastive. These include Löfqvist & McGowan (1992), Silverman, Blankenship et al. (1995), and Blankenship (1997). In those languages for which phonation is separately contrastive, the contrast in phonation only lasts for a proportion of the vowel duration. For example, Silverman, Blankenship et al. (1995) found that in Mazatec, contrastive breathiness only lasts for 43% of the vowel duration, with modal phonation for the rest of the duration. The possible reason for contrastive phonation only lasting for a proportion of the vowel is the need to distinguish phonation and tone independently of each other. This is different from Walungge in that phonation and tone are not separately contrastive. However there is a parallel with these studies, in that it is the first portion of the vowel where phonation can be argued to be enhancing the underlying tone distinction rather than simply varying according to the pitch.

4.5 Duration

4.5.1 Duration and tone

One of the effects that has been observed in various languages with contrastive register is for low register to lengthen the duration of the vowel. This has been discussed in section 2.4 above, and is separate from any phonological length that the language might have. If the language in question has phonological vowel length, and in addition register has an effect on vowel duration, then phonetically there might be four distinct vowel durations. Lhomi (Watters (2003)) is an example of a Tibetan language where one of the phonetic correlates of register is longer vowel duration for low register.

4.5.2 Duration for monosyllabic words

Walungge has three basic syllable rhyme shapes, each with a difference in duration: V (single phonologically short vowel), VN (phonologically long vowel, or short vowel plus sonorant coda) and VP (short vowel plus obstruent coda). The difference in duration between V and VP is caused by the presence of the final obstruent, which shortens the vowel duration. Vowels which are phonologically long have a duration similar to that of rhymes comprising of a short vowel plus sonorant coda, and the duration is approximately double that of a short vowel without any coda.

In monosyllabic words, for each type of syllable rhyme (V, VN, VP) there is a slight difference in duration for high and low tone, with low tone words being slightly longer in duration than high tone words. This can be seen in the boxplot (showing median, interquartile range, and whiskers of 1.5x interquartile range) and table of means below, for monosyllabic words in isolation.

Figure 4-31: Boxplot of duration of rhyme for monosyllabic words

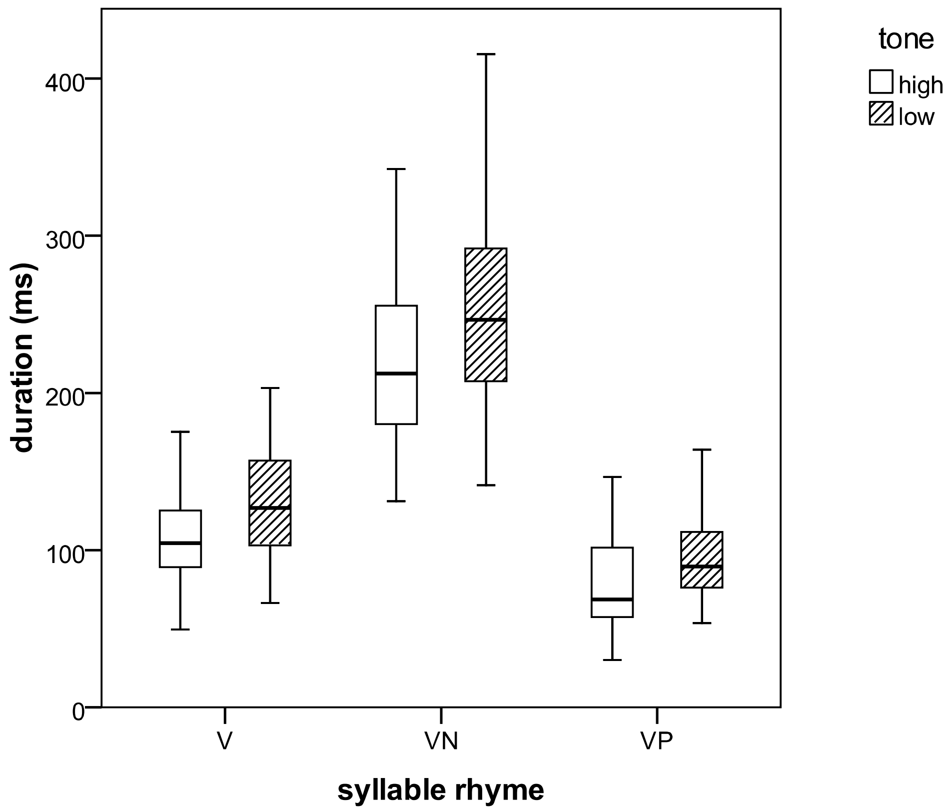


Table 4-25: Mean and s.d. for duration of syllable rhymes; monosyllabic words

tone	*syllable rhyme*	*mean duration (ms)*	*std. dev.*	*N*
high	V	108.7	29.4	80
	VN	220.1	54.5	83
	VP	75.9	28.7	28
low	V	129.6	32.7	88
	VN	254.3	65.1	111
	VP	99.9	35.5	35

As well as the difference in rhyme duration between high and low tone, the above boxplot and table also show the variation that there is in duration for each type of rhyme. In particular, the amount of variation causes an overlap in duration measurements for V and VN. The overlap is interesting to note because vowel length (included in the rhyme shape VN) is considered to be a phonological contrast. This suggests that some of the acoustic cue for the contrast is carried by the pitch melody rather than the vowel length (see chapter 7 for perception experiments investigating this), and is further evidence that the pitch melodies *[52]* and *[44]* are realisations of two separate tone patterns H and HL rather than being phonetic variants according to syllable weight.

An ANOVA was carried out for effects upon rhyme duration, with independent variables:

- tone
- rhyme shape
- vowel: aside from contrastive length, there are 7 vowels in Walungge (/i/, /e/, /a/, /o/, /u/, /y/, /ø/).

The following table gives the ANOVA results, with significance level as $p < 0.05$.

Table 4-26: ANOVA for effects on duration of monosyllabic rhyme

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	425	1	17.1	<0.0005	•
vowel	425	6	1.96	0.070	

The vowel does not have a significant effect upon the duration of the rhyme, however tone does have a significant effect, with low tone words having significantly longer duration than high tone words.

4.5.3 Duration for first syllable of disyllabic words

Below is a boxplot and table of means for the duration of first syllable rhymes in disyllabic words, along with the results of an ANOVA investigating effects upon the duration. What is particularly interesting is that the underlying tone attaches to the first syllable of disyllabic words, yet the first syllable rhyme duration is not significantly affected by the tone.

Figure 4-32: boxplot for duration of first syllable rhyme

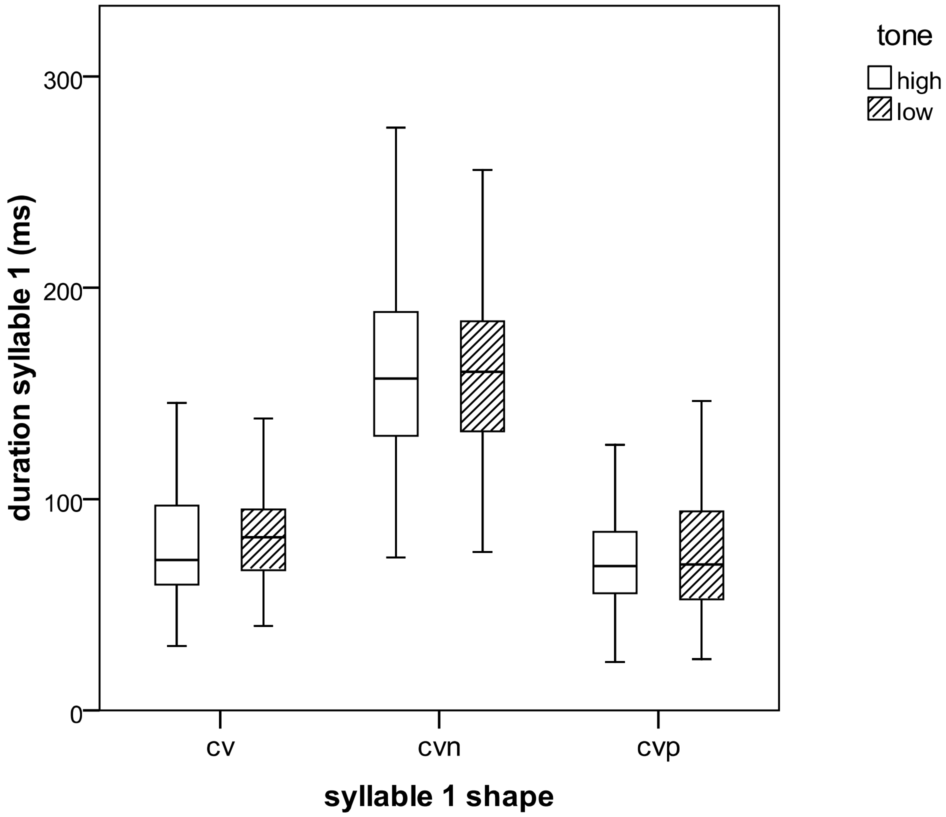


Table 4-27: Mean and standard deviation of duration of first syllable rhyme

tone	*1st syllable shape*	*mean duration (ms)*	*std. dev.*	*N*
high	CV	73.1	22.8	133
	CVN	160.4	41.7	194
	CVP	69.7	18.4	145
low	CV	79.5	21.0	156
	CVN	157.1	38.2	171
	CVP	73.0	27.5	89

Table 4-28: ANOVA for effects upon duration of first syllable rhyme

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	888	2	0.35	0.556	

The underlying tone attaches to the first syllable of polysyllabic words, and thus the expectation is that the acoustic cues for tone surface on the first syllable. Indeed the acoustic cues of a difference in pitch, VOT and breathiness are all there on the first syllable. And for monosyllabic words a difference in tone correlates with a difference in syllable rhyme duration. Yet for disyllabic words, tone has no significant effect upon the duration of the first syllable.

A possible explanation for this is that the difference in duration for monosyllabic words is not caused directly by the underlying tone but rather it is caused by the rising or falling pitch melody. The pitch melodies for monosyllabic words are as follows:

Table 4-29: monosyllabic pitch melodies

tone	*syllable rhyme*	*pitch melody*
high	V	52
high	VN	44
high	VP	53
low	V	24
low	VN	14
low	VP	24

For each shape of syllable rhyme, underlying high tone is associated with a fall (surface tone HL) or with level pitch (surface tone H), and low underlying tone is associated with a rise (surface tone LH). Xu & Sun (2002) consider the speed at which a change of pitch can take place, with the conclusion that the speed of pitch change is different for rises and falls. Unless the amount of change is very small, a fall can occur faster than a comparable rise. This would explain why for monosyllabic words the high tone melodies have a slightly shorter duration than the low tone melodies.

In the case of disyllabic words, the pitch melody on the first syllable is essentially level, with level high for high tone and level low for low tone. If the duration is being affected by a rise or fall in the pitch, then this would explain why a difference in tone does not correlate with a difference in duration for the first syllable of disyllabic words.

4.5.4 Duration for second syllable of disyllabic words

Below is a boxplot and table of means for the duration of the syllable rhyme for the second syllable of disyllabic words, along with the results of an ANOVA for effects upon the duration. The boxplot and table show a slight difference in duration between high and low tone for each shape of syllable rhyme. This difference is statistically significant, as can be seen from the ANOVA (significance level $p < 0.05$).

Figure 4-33: Boxplot of rhyme duration for second syllable

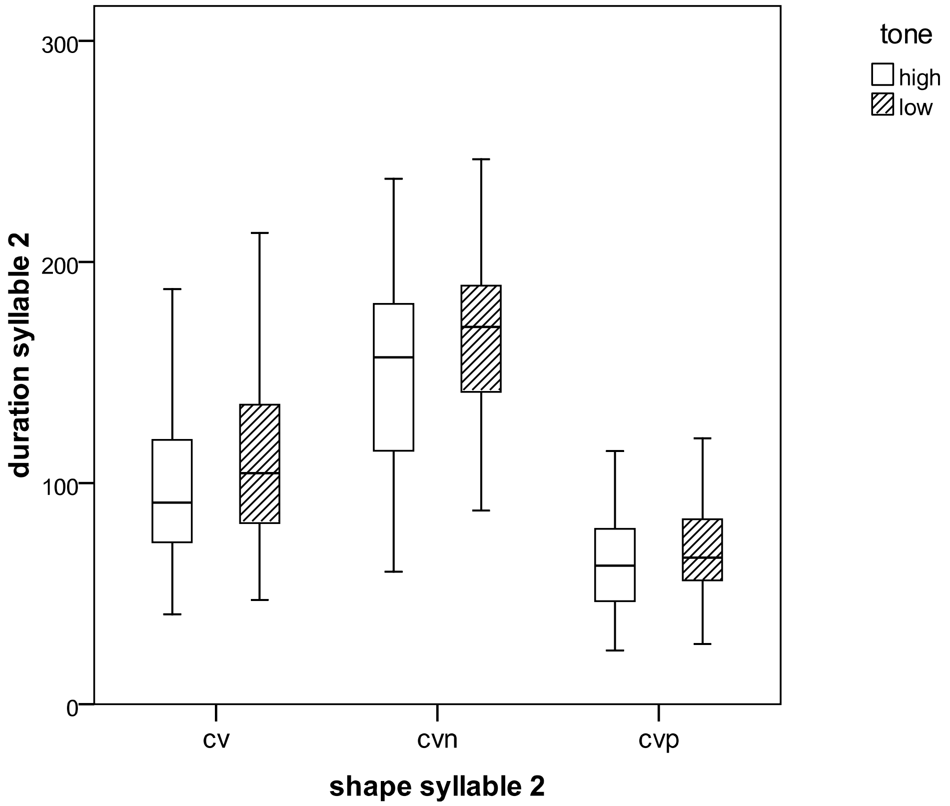


Table 4-30: Table of mean and s.d. for rhyme duration of second syllable

tone	*2nd syllable shape*	*mean duration (ms)*	*std. dev.*	*N*
high	CV	104.9	28.4	313
	CVN	170.4	30.9	85
	CVP	78.2	16.6	74
low	CV	114.2	32.8	208
	CVN	177.9	33.0	103
	CVP	86.5	16.8	105

Table 4-31: ANOVA for effect of tone upon duration of second syllable rhyme

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	888	1	7.96	0.005	•

The ANOVA result is interesting because the primary pitch difference associated with a difference in tone is only on the first syllable of polysyllabic words, yet a difference in tone correlates with a difference in the duration of the second syllable for disyllabic words. In order to give a possible explanation for this difference in duration, the disyllabic pitch melodies for high and low tone need to be considered. These are described in detail in section 4.3.3 above. In summary, for both high and low underlying tone the pitch on the second syllable is high. However for low underlying tone, the pitch is low on the first syllable and rises predominantly through the onset of the second syllable. The pitch has not reached its target height by the start of the second syllable rhyme, and so there is a peak delay effect where the target height is not reached until partway through the second syllable rhyme. This delay in reaching the high peak is a possible explanation for why the overall duration of the syllable rhyme is longer.

4.5.5 Summary of duration

Whilst a difference in duration correlates with a difference in tone for monosyllabic words, with low tone words having longer duration, it does not correlate with tone for the first syllable of disyllabic words. The difference between low tone monosyllabic words and the first syllable of low tone disyllabic words is that the pitch melody is rising for monosyllabic and level for the first syllable of disyllabic. Thus the difference in duration can be attributed to the rising pitch. The duration difference for the second syllable of disyllabic words can also be attributed to pitch, namely the peak delay caused by the rising pitch between the first and second syllables.

4.6 Conclusion for acoustic analysis of tone

The pitch melodies on Walungge nouns do not only vary according to the underlying tone of the word. Monosyllabic melodies also vary according to syllable weight. The acoustic analysis carried out sought to provide an explanation for the differences in the pitch melodies. All pitch melodies for words in isolation end with a fall. In most cases this fall disappears when words are phrase internal. However, the high-falling pitch melody for high tone light syllable monosyllabic words is retained and may also affect the pitch of the following word. This suggests that the high-falling pitch melody of light syllables and the high-level pitch melody of heavy syllables are realisations of separate tone patterns HL and H. Similarly, the rising pitch melodies of low tone monosyllabic words are retained, and taken to be realisations of a rising tone pattern LH. The analysis of the pitch also shows four distinct starting pitches for monosyllabic words according to tone and weight. In the case of high tone words, the difference in starting pitch can be accounted for if the pitch melodies are analysed as realisations of two separate tone patterns H and HL, with the higher starting pitch of HL giving greater phonetic space for the fall. The difference in starting pitch for LH words can be accounted for by the duration of the rhyme.

Vowel phonation closely follows the pitch melody. As the pitch rises the phonation becomes less breathy and as the pitch falls the phonation becomes more breathy (taking closed quotient as an indication of breathiness). However right at the start of a word, for both monosyllabic and polysyllabic words, there is only a two way distinction for starting phonation rather than the 4-way distinction of the starting pitch. This is analysed to be in part due to the effect of the initial consonant, and in part due to the phonation enhancing the H versus L tonal contrast. Because of the way that phonation closely follows the pitch melody throughout the duration of all vowels in the word, in general phonation can be analysed as being a phonetic enhancement to the pitch melody, and hence tone, of the word.

Although a difference in vowel duration correlates with a difference in tone for monosyllabic words, it would appear that vowel duration is being influenced by the rise and fall of the pitch melody rather than by the underlying tone of the word. A rise takes longer than a fall, and hence the duration is longer for rising melodies than for falling melodies. If the pitch is level, as in the first syllable of polysyllabic words, then there is no significant difference in duration between the two tones.

Of the correlates pitch, phonation, and duration, pitch is the predominant factor, with phonation and duration being dependent upon the pitch melody as a whole rather than purely being dependent upon the underlying tone of the word. The pitch melodies themselves are best analysed as realisations of different surface tone patterns. For these reasons, Walungge is best analysed as being a tone language rather than a register language.

5 Tone, stress, and tone associations

This chapter is concerned with issues surrounding the attachment of tone to tone bearing units in Walungge, and in particular how the surface patterns can be accounted for if Walungge is analysed as having one underlying tone per word: H or L.

The chapter starts with an introduction to stress and tone, and languages which have an interaction between stress and tone. Also outlined is the theory that the tone patterns of disyllabic words in tonal Tibetan languages have arisen from stress. Following this introductory section, the question of stress in Walungge is considered from an acoustic perspective. For disyllabic nouns the duration of vowels, the pitch, and the intensity are analysed. Stress patterns on verbs are also briefly considered.

Using an autosegmental framework, attention is then turned to the issue of underlying tones, surface tone patterns, and pitch melodies. In particular the concept of a tonal foot is introduced. This can be seen as a parallel structure to a metrical foot, which is formed for the purpose of assigning tones to tone bearing units. It is proposed that Walungge a) forms a tonal foot for the purpose of attaching tones to tone bearing units and b) has boundary tone insertion. This enables Walungge to be analysed as having one underlying tone per word, either H or L. Using this proposal, the derivation of surface tone pattern from underlying tone is explained using examples from a wide variety of different words.

Finally attention is turned to tone patterns across morpheme boundaries. Morpheme boundaries in Walungge are particularly interesting because of the different relationships between stem and affix. The boundary between stem and affix is at a different level in the prosodic hierarchy for different affixes; and in addition, affixes may be tone-bearing or extra-tonal. This affects both the segmental alternations and the tone patterns.

5.1 Tone and stress

5.1.1 Introduction to tone and stress

Hyman (2007) summarises definitions of a tone language and a stress language as:

stress system: a language with word-level metrical structure, e.g. English

tone system: a language with word-level pitch features, e.g. Mandarin

From these definitions he continues "Stated this way, stress and tone have virtually nothing inherently in common: Stress is a structural property in which syllables are metrically hierarchized as relatively strong vs. weak, while tone is a featural property referring to contrastive relative pitch." (p. 656). However, as Hyman points out, whilst there are some languages which only have the properties of a stress language and there are other languages which only have the properties of a tone language, there are also languages which have some of the properties of both. In this situation there is no single way in which languages combine some of the properties of a tone system and some of the properties of a stress system. Even amongst languages which have been termed "pitch-accent languages" there are a variety of different systems. Hyman gives a number of examples of languages which combine different properties of tone systems and stress systems.

The following subsections outline in more detail some of the interactions between stress and tone which are of particular relevance to the analysis of tone in Walungge.

5.1.2 Metrical structure and tone

Metrical structure and tone have been found to interact in a number of the world's languages. De Lacy (2002) considers this interaction, making the following generalisations. Some languages have tone-driven stress, and others have stress driven tone. For languages with an interaction between stress and tone, de Lacy proposes that there is a hierarchy of tone types, analogous to the sonority hierarchy, with higher tone being more prominent than lower tone (pp. 3-4). This tone hierarchy may interact with a metrical stress system in that

(a) Foot heads and higher tone have an affinity for each other.

(b) Foot non-heads and lower tone have an affinity for each other.

(c) (a) and (b) can motivate

i. attraction of (non-)heads to tone;

ii. attraction of tone to (non-)heads;

iii. neutralisation of tone on (non-)heads.

To illustrate these relationships in detail, de Lacy considers Ayutla Mixtec which is an example of tone-driven stress, and Lamba (a Bantu language) which is an example of stress-driven tone.

Stress placement in Ayutla Mixtec (De Lacy (2002:5-7)) is determined by the formation a single trochaic foot per prosodic word. The head of the foot aligns with the highest tone in the word if that tone is followed by a lower tone. If all tones are the same, the foot aligns with the leftmost edge of the word. The following are examples:

LMˈHL	lùlū(ˈúrà)	‘he is not small’
MˈML	lā(ˈʃārà)	‘his orange’
ˈHHH	(ˈʃíní)rá	‘he understands’

Lamba (De Lacy (2002:14-15)) has predictable high tone placement. The stress domain is parsed into trochaic feet, and high tones may only occur on heads. The following are examples:

u-ku-mu-pám-a	→	u(ˈkúmu)pama	‘to beat him/her’
u-ku-pám-a	→	u(ˈkúpa)ma	‘to beat’
tá-tu-luku-mu-kom-a	→	tatu(ˈlúku)(ˈmúko)ma	‘we are not hurting him’

Kera (a Chadic language; Pearce (2007)) is another example of the placement of tone being affected by the metrical structure of the language. Kera forms iambic feet. If a word has only one foot, then each syllable may carry a tone. If a word has more than one foot, then each foot can only have one tone. The following are examples:

1 foot		*more than 1 foot*	
(gèdɛ̀l)	‘mud’	(gɔ̀r)(nɔ̀y)	‘hyena’
(gùɓúː)	‘stand for pots’	(dàk)(tə́láw)	‘bird’
(kítīr)	‘moon’	(kúɓúr)sī	‘coal’

Hyman (2007) gives the example of Seneca, an Iroquoian language. Seneca marks the first syllable of a word as extrametrical and then builds disyllabic trochees left-to-right. Tone H is assigned to the first syllable of a trochee if and only if either syllable is closed. This gives structures such as <σ>(CáC.Ca)(Ca.Ca), <σ>(Cá.CaC)(Ca.Ca), <σ>(Cá.CaC)(CáC.Ca), <σ>(Ca.Ca)(Ca.Ca).

Mandarin Chinese (Yip (2002)) is an example of an Asian language showing an interaction between metrical structure and tone. Although most syllables in Mandarin take their own lexical tone, some syllables have no tone of their own but surface with a variety of tones depending on the tone of the preceding syllable. These underlyingly toneless syllables are unstressed, and are short, monomoraic syllables. The analysis that Yip presents is that Mandarin forms trochaic feet which are minimally bimoraic. Heavy syllables must be stressed, and stressed syllables must be heavy. The tone of the head of the foot spreads to non-head syllables.

Drawing on Duanmu’s analysis of Lhasa Tibetan (Duanmu (1992)), Yip (2002) analyses Lhasa Tibetan as having an initial head syllable which keeps its tone and spreads that tone to all non-head syllables. All the non-head syllables lose their tone. Shanghai Chinese (Yip (2002)) shows a similar pattern of tone retention and deletion.

5.1.3 Refugee Standard Tibetan

Meredith (1990) uses a footing process to describe the attachment of tone in the variety of Tibetan which he refers to as “Refugee Standard Tibetan (RST)” as spoken in the Tibetan refugee community in Nepal and India. This is a particularly interesting analysis, briefly outlined below, because of the similarities between RST and Walungge.

RST is described as having two tone registers: low and high. For monosyllabic nouns, high register can either have a level or a falling tone pattern; low register can either have a falling or a rising tone pattern, e.g.

caa	high falling	‘iron’
chẽẽ	high level	‘mother’ (hon.)
ree	low falling	‘cotton’
phöö	low rising	‘Tibet’

Meredith gives the geometry of the tones first in terms of features [stiff] and [slack], and then for convenience simplifies them to the following, with t = tonal node, r = register node, c = contour node:

Low register rise

```
t
|
r
|
L
```

Low register fall

```
   t
  / \
 r   c
 |   |
 L   fall
```

High register level

```
t
|
r
|
H
```

High register fall

```
   t
  / \
 r   c
 |   |
 H   fall
```

The rise in low register is analysed as being an inserted rise, and not part of the underlying phonological form.

Disyllabic nouns have H on their second syllable. In addition the second syllable may have a fall, but first syllables have either L or H and never a rise or a fall. Words with three or more syllables have pitch which falls to low on third and subsequent syllables (they are analysed as being toneless).

Using the framework of a metrical grid, Meredith analyses RST as having separate metrical processes for stress and tone, with the metrical process for tone preceding the process for stress. For tone, binary right-headed feet are formed from left to right, with conflation to leave one metrical head per word. This head will be the second syllable of the word. H is inserted onto the metrical head. Any unmetrified syllable has its tonal root node delinked. Any metrified syllable that is not the head has its contour node delinked.

The following is an example of an RST trisyllabic compound in which *pu* 'child' + *su* 'care' + *qhan* 'house' becomes 'nursery'. The underlying tones of each word are shown below in diagrammatic form. In the grid above the word, line 0 shows the formation of right-headed feet with the heads of these feet projected onto line 1. Line 1 shows the formation of a left-headed foot from the line 0 heads, the head of which is in line 2.

```
line 2          *
line 1         (*     *)
line 0     (*    *)   (*)
word       pu   su   qhan
            |    |     |
            t    t     t
            |    |\    |
            r    r c   r
            |    | |   |
            L    H fall H
```

Continuing with the derivation, line 1 is deleted to leave only one head for the whole word and line 2 becomes the new line 1 (conflation). The second foot of line 0 no longer has a head, so becomes unfooted. The third word loses its tonal node because it becomes unmetrified. The second word keeps its fall as it becomes the head.

```
line 2          *
line 1         (*     *)                *
line 0     (*    *)   (*)          (*    *)   *
word       pu   su   qhan    →     pu   su   qhan     'nursery'
            |    |     |            |    |
            t    t     t            t    t
            |    |\    |            |    |\
            r    r c   r            r    r  c
            |    | |   |            |    |  |
            L    H fall H           L    H  fall
```

The next example of a disyllabic compound word in RST: ‘cotton’ + ‘robe’ becomes ‘cotton robe’. The derivation is similar to the above example. A right headed foot is formed on line 0 with its head on line 1. A left headed foot is formed on line 1 with its head on line 2, following which is conflation. The L of the second word becomes H as the second word becomes the head of the compound. The fall on the first word is lost because it is not a head.

```
line 2          *
line 1         (*)                 *
line 0      (*    *)          (*    *)
word       ree    see    →    ree  see     ‘cotton robe’
            |      |           |    |
            t      t           t    t
            |\     |           |    |
            r c    r           r    r
            | |    |           |    |
            L fall L           L    H
```

The stress on nouns is determined by tone, syllable weight, and syllable position. Only syllables which are metrified by the tonal metrical process may be stressed. For stress allocation heavy syllables take priority over light syllables and syllables with H take priority over syllables with L. Because the allocation of stress is dependent upon the placement of the tones, the metrical process for stress takes place after the process for tone assignment.

5.1.4 Tone and stress in Tibetan languages

As well as tone and stress interacting in the synchronic form of a language, there might also have been a diachronic interaction, with tone patterns originating at least in part from stress.

Caplow (2009) considers the second-syllable-high tone patterns of Tibetan languages in the light of the stress patterns found in two Tibetan languages which have been described as being atonal. In particular she looks at stress in Balti and in Rebkong Amdo Tibetan, which are archaic Tibetan languages. In both these languages, disyllabic nouns, adjectives, and numerals have second syllable stress. In Balti the primary acoustic correlate of stress is pitch, which is higher on the second syllable than the first. For Rebkong Amdo it is both the height and the slope of the pitch. Using what she terms "historical comparative acoustics", not only is stress reconstructed for Proto-Tibetan, but also the acoustic correlates of stress, and in particular the prominence of pitch. Thus she proposes that Proto-Tibetan had default pitch patterns of low pitch on unstressed syllables and high pitch on the stressed syllable. The prominence of pitch on the second syllable of Proto-Tibetan nouns can account for the second syllable high tone in tonal Tibetan languages.

5.2 Acoustic analysis of stress in Walungge

5.2.1 Introduction

Section 5.1 above gives an overview of a number of languages which have an interaction between tone and stress. Attention is now turned to Walungge polysyllabic words, all of which have high pitch on the second syllable, as illustrated below:

disyllabic	*sámba* 'new'	[44 44]
	sàmba 'bridge'	[11 44]
three syllables	*órokpa* 'toad'	[44 44 21]
	bàrakpa 'big'	[11 44 21]
four syllables	*tɕémaladʑok* 'butterfly'	[44 44 21 11]
	dʑàŋbagadze 'lizard'	[11 44 21 11]

The pattern of high pitch on second syllables is observed throughout the language in all content words (with the exception of those affixes which are outside of the tonal domain, see section 5.4.4 below). Whilst the second syllable high pitch can be analysed as a realisation of high tone, the aim of this section is to demonstrate from an acoustic perspective that as a general pattern Walungge has second syllable stress, and not just second syllable high tone. Section 5.3 below draws on this acoustic analysis when looking in detail at the placement of tone in Walungge.

The acoustic correlates of a stressed syllable may include longer duration, higher pitch, and greater acoustic intensity (Hayward (2000), Ladefoged (2003)). However, the mix of these correlates may vary from language to language, and each of these correlates is influenced by a number of factors, which may make it difficult to measure stress acoustically. For example the duration of a vowel might not only be affected by stress, but also by the vowel height, by the number and shape of the syllables in the word, and whether there is any phonological length. This means that not all stressed syllables will necessarily have longer duration. However, all other things being equal, namely if both syllables are the same phonological shape, both vowels the same height and the same phonological length, etc, then an increased duration on one of the syllables would be an indication of stress on that syllable, particularly if it was accompanied by higher pitch and/or greater intensity. Ladefoged (2003) illustrates the interplay of these correlates with examples from English, showing that in general English stress is more to do with higher pitch and longer duration than greater intensity, with pitch generally being the most important correlate, but he shows some acoustic measurements where the stressed syllable had lower pitch but greater intensity and longer duration. Therefore any acoustic analysis on stress has to be done controlling for variables that might influence the various correlates, and it is important to consider all three correlates and not just any one of them.

The acoustic analysis which follows considers duration, pitch, and intensity for Walungge disyllabic words. In particular it considers disyllabic non-verbs, but following this it briefly considers verbs. The data used for the acoustic analysis below is the same data that was used for the acoustic analysis in chapter 4 above.

5.2.2 Duration of vowels in disyllabic non-verbs

This section considers whether vowels on the second syllable of a disyllabic word are typically longer in duration than first syllable vowels. There are two factors which are already known to affect the duration of vowels in Walungge. The first is tone, with low tone words having slightly longer second syllable vowels than high tone words (section 4.5 above). The second is the phonological length contrast of vowels, which can occur on both the first and the second syllable of a disyllabic word, e.g.

t^hòmaː	'potato'	*hòma*	'milk'
gòmo	'money'	*gòːmo*	'evening'

Because of these influences on duration, when considering whether second syllables are longer than first syllables, words need to be separated according to whether they have long or short vowels, and also according to whether they are high or low tone words.

In order to account for the different syllable shapes of either syllable, two different statistical analyses were performed, which are discussed below. The first considers words which have the same syllable shape for both first and second syllable, and looks at the question of whether the second syllable is significantly longer than the first. The second analysis looks at words of all shapes and considers whether vowels in a syllable of a particular shape are significantly longer when that syllable is the second syllable.

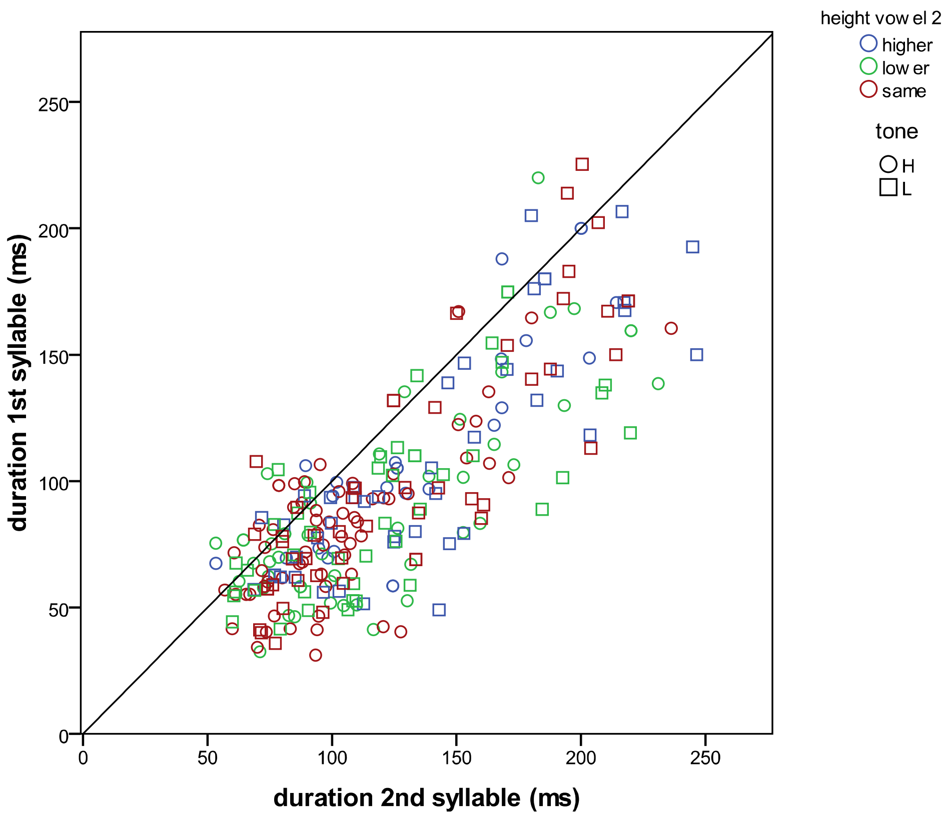
Same syllable shape on both syllables

The following word patterns have similarly shaped first and second syllables: CVCV, CVPCVP (plosive coda), and CVNCVN, where VN represents either a long vowel or a short vowel plus a sonorant. Rhymes with vowel + sonorant or with long vowels were grouped as one pattern because they are alike in duration and pitch melody. Below is a table of mean durations of first and second syllable rhymes for different word patterns and tones, which shows that for each word pattern and tone the mean duration of the second syllable is longer than the mean duration of the first syllable. This is followed by a scatter-plot of first and second rhyme durations. Because vowel height can affect duration (in particular high vowels tend to have shorter durations), the scatter plot indicates whether the second vowel is higher, lower, or the same height as the first vowel. The diagonal line on the scatter plot indicates duration of syllable 1 equal to duration of syllable 2.

Table 5-1: Mean and standard deviation of rhyme durations

pattern	*tone*	*syllable*	*mean duration (ms)*	*s.d.*	*N*
CVCV	H	1	77.0	22.3	85
		2	99.0	24.5	
	L	1	82.6	22.0	78
		2	114.0	32.1	
CVPCVP	H	1	66.2	13.4	25
		2	80.9	15.3	
	L	1	64.4	18.7	23
		2	81.7	14.8	
CVNCVN	H	1	153.4	44.5	32
		2	176.1	29.9	
	L	1	160.1	36.3	41
		2	183.5	32.9	

Figure 5-1: scatter plot showing durations of 1st and 2nd syllable rhymes



From the above scatter plot it can be seen that with only a few exceptions the duration of the second syllable rhyme is longer than the duration of the first syllable rhyme. This holds regardless of the height of the second vowel in relation to the height of the first vowel, regardless of word shape, and regardless of tone.

In order to confirm statistically that 2nd syllables are significantly longer in duration than 1st syllables, for each word pattern and tone the durations of first and second rhymes were paired for each word, and a paired T-test was carried out. The following table is a summary of the results.

Table 5-2: paired T-test results comparing durations of 1st and 2nd syllable

word pattern	*tone*	*N*	*d.f.*	*t*	*P*	*Sig*
CVCV	H	85	84	7.133	<0.0005	•
	L	78	77	8.386	<0.0005	•
CVPCVP	H	25	24	4.335	<0.0005	•
	L	23	22	3.772	0.001	•
CVNCVN	H	32	31	2.864	0.007	•
	L	41	40	3.583	0.001	•

For each word pattern and for both H and L tone, the duration of the second syllable is significantly longer than the duration of the first syllable.

All word shapes

The syllable shapes CV, CVP and CVN can each occur in either first or second syllable position and can each occur with any other syllable shape, which gives rise to 9 different disyllabic word shapes. For a word shape such as CVCVN the second syllable rhyme will be longer in duration than the first simply because of the shape of the second syllable. Similarly CVNCV will have a first syllable rhyme which is longer in duration than the second. In order to take into account all word shapes, the question is whether a particular syllable rhyme shape is shorter in duration when it is in first syllable position than when it is in second syllable position regardless of word shape.

The following table gives the mean duration of syllable rhymes for each combination of first and second syllables.

Table 5-3: mean duration (ms) of syllables; differing syllable shapes; high tone

	2^{nd} = CV		2^{nd} = CVP		2^{nd} = CVN	
	1^{st}	2^{nd}	1^{st}	2^{nd}	1^{st}	2^{nd}
1^{st} = CV	77	99	69	77	68	164
1^{st} = CVP	72	102	66	81	64	170
1^{st} = CVN	161	111	132	76	153	176

Table 5-4: mean duration (ms) of syllables; differing syllable shapes; low tone

	2^{nd} = CV		2^{nd} = CVP		2^{nd} = CVN	
	1^{st}	2^{nd}	1^{st}	2^{nd}	1^{st}	2^{nd}
1^{st} = CV	83	114	75	87	78	171
1^{st} = CVP	73	109	64	82	81	180
1^{st} = CVN	159	117	151	89	160	183

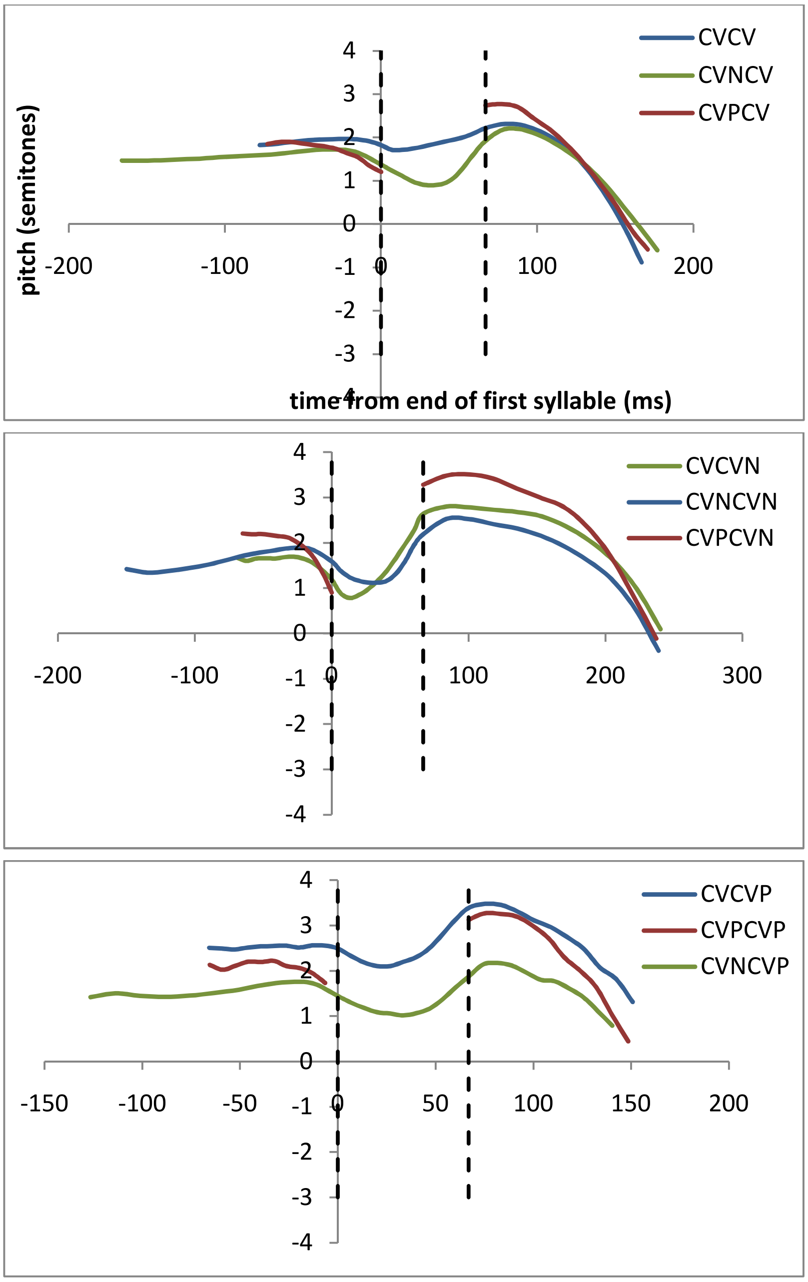
Notice from the above tables that no matter what combination of first and second syllables and no matter whether the tone is high or low, each syllable shape (CV, CVP or CVN) has shorter mean duration when it is in first syllable position than when it is in second syllable position. In particular notice the case of when the first syllable is a heavy syllable (CVN), which because of its weight might be expected to attract stress. If the second syllable is a light syllable, whilst it does not have as long duration as the first syllable, it has longer duration than when the equivalent shape is a first syllable.

Thus, if duration is to be taken as an indication of stress, then Walungge has a general pattern of second syllable stress for all disyllabic non-verbs. As has already been discussed, duration is only one possible correlate of stress; the following sections consider the correlates of pitch and intensity.

5.2.3 Pitch on disyllabic high tone words

Data has already been presented to show that Walungge has high pitch on the second syllable regardless of the tone of the word. This section considers the pitch across a high tone word. The graphs below are from the data and graphs in section 4.3.3 above and show the mean pitch melodies (in semitones) for disyllabic high tone words. For ease of display, these are drawn three pitch melodies to the graph. Notice that for each word pattern the highest pitch on the second syllable is higher than the highest pitch on the first syllable.

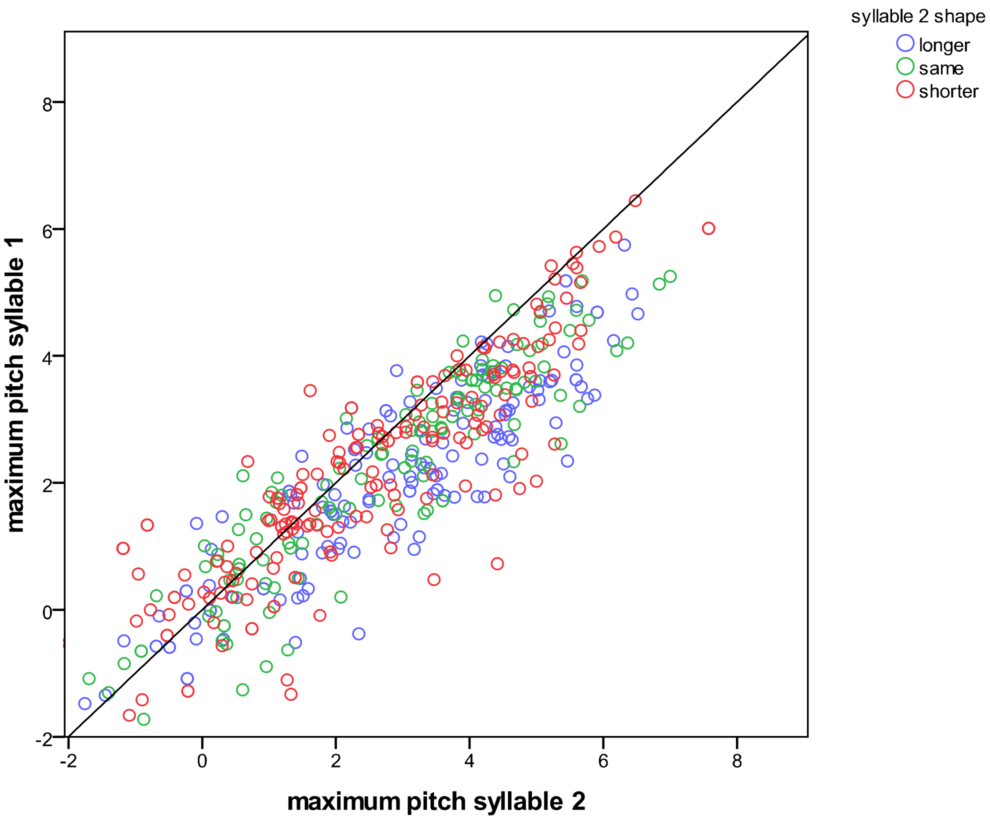
Figure 5-2: mean normalized pitch melodies for each high tone word pattern



In order to investigate whether the observed difference in pitch between first and second syllables is significant, for each word the maximum value of F0 for each syllable was measured. The maximum value of F0, rather than the mean value of F0, was chosen because of the fall on the second syllable which occurs for words in isolation. In order to adjust for the effect of a consonant upon the value of F0 immediately following the consonant, the first 10ms of a vowel were excluded when looking for the maximum value of F0. The maximum values of F0 were then converted to pitch in semitones relative to the mean value of F0 for each speaker, in order to normalize for all speakers (see section 4.3 above for detailed description of the normalisation process).

The following is a scatter plot showing the maximum pitch values for syllable 1 and syllable 2 for each recorded high tone word. For ease of display, rather than coding the scatter plot for all nine word patterns, the word patterns have been grouped according to whether the shape of the 2nd syllable is longer (CVPCVN, CVCVN, CVPCV), the same (CVPCVP, CVCV, CVNCVN) or shorter (CVCVP, CVNCVP, CVNCV) than the shape of the 1st syllable. The diagonal reference line shows the pitch on syllable 1 equal to the pitch on syllable 2.

Figure 5-3: Scatter plot showing max pitch (semitones) of 1st and 2nd syllables



From the above scatter plot, whilst there are some data tokens with the pitch of syllable 1 higher than the pitch of syllable 2, the vast majority of tokens shown a higher 2nd syllable pitch than 1st syllable pitch. This pattern of second syllable pitch being higher than first syllable pitch can be statistically investigated by means of a paired T-test.

First and second syllable measurements were paired, and a paired T-test was carried out for the maximum value of the normalised pitch on each syllable, for each word pattern. The following are the results of the T-test:

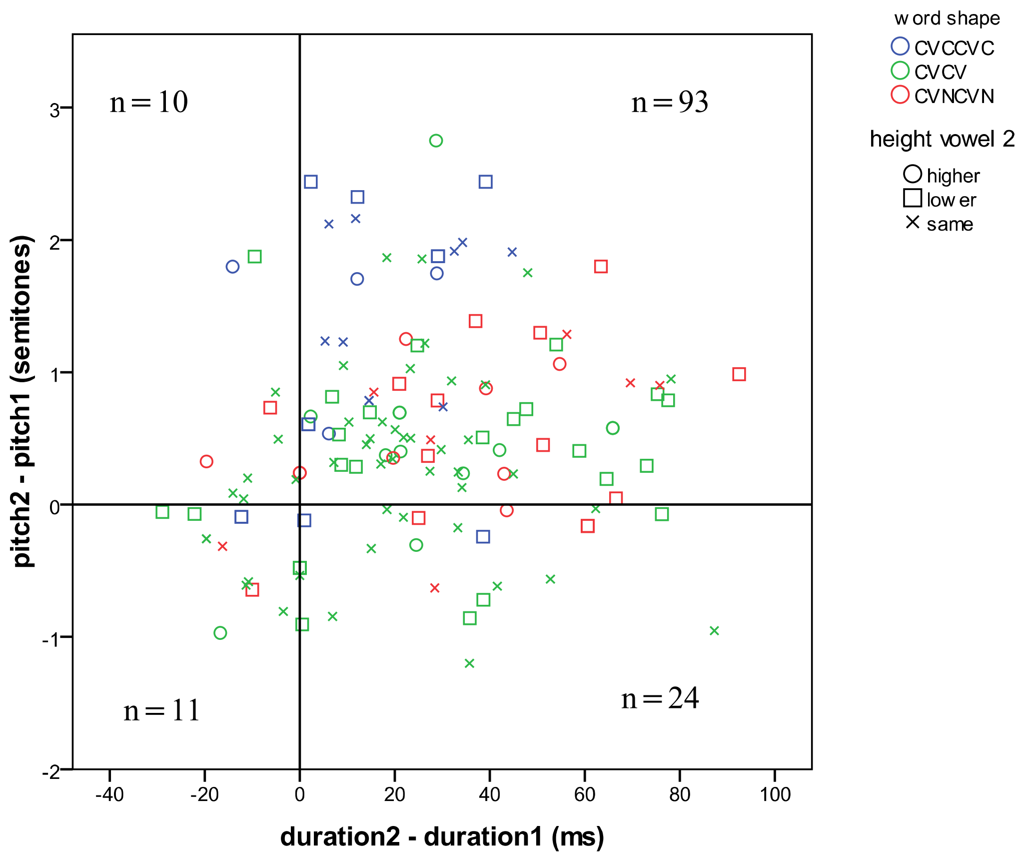
Table 5-5: paired T-test comparing 1st and 2nd syllable pitch; high tone words

word pattern	*N*	*d.f.*	*t*	*P*	*Sig*
CVPCV	94	93	5.590	<0.0005	•
CVPCVP	25	24	8.567	<0.0005	•
CVPCVN	26	25	6.662	<0.0005	•
CVCV	85	84	4.181	<0.0005	•
CVCVP	21	20	3.841	0.001	•
CVCVN	26	25	5.704	<0.0005	•
CVNCV	134	133	6.224	<0.0005	•
CVNCVP	28	27	2.975	0.006	•
CVNCVN	32	31	5.660	<0.0005	•

The T-test results show that for every word pattern the maximum value of normalized pitch on the second syllable is significantly higher than the maximum value of normalized pitch on the first syllable.

The following scatter plot combines pitch with duration. For each high tone word with the same shape on both syllables (CVPCVP, CVCV, CVNCV) the duration difference between syllables (duration2 minus duration1) was calculated, and the pitch difference (pitch2 minus pitch1) was calculated. The scatter plot shows the pitch difference and duration difference for each word. The number of tokens in each quadrant of the scatter plot is also shown.

Figure 5-4: Scatter plot showing pitch difference and duration difference



n = the number of tokens in each quadrant

The majority of high tone words with same shaped syllables have both a longer duration and a higher pitch on the second syllable. This is the top right quadrant of the above scatter plot. Of the remainder, most have either higher pitch or longer duration on the second syllable. There are only very few words which had neither longer duration nor higher pitch.

This is a further indication that the general stress pattern in Walungge is second syllable stress. It should be pointed out, though, that whilst this is the general pattern there are some exceptions. From the above acoustic analysis it can be noted that there was a small handful of high tone nouns which had neither higher pitch nor longer duration on the second syllable (which can be seen in the bottom left quadrant of the above scatter plot).

5.2.4 Intensity

The third possible correlate of stress which is investigated here is intensity, measured in decibels. In order to minimize any effect from the consonants on either side of the vowel, intensity was measured as the mean intensity across the mid 50% of the vowel, which is the method used in Caplow (2009). Intensity varies with vowel height, with a lower vowel in general having greater intensity than a higher vowel. Thus, in comparing intensity of first and second syllables, the data was split according to whether the second vowel height is higher, the same, or lower than the height of the first vowel. The following are three scatter plots showing the intensity measurements for the first and second syllable vowels. The diagonal line is where the intensity of both syllables is equal.

Figure 5-5: scatter plot of intensity; words with 2nd vowel height higher

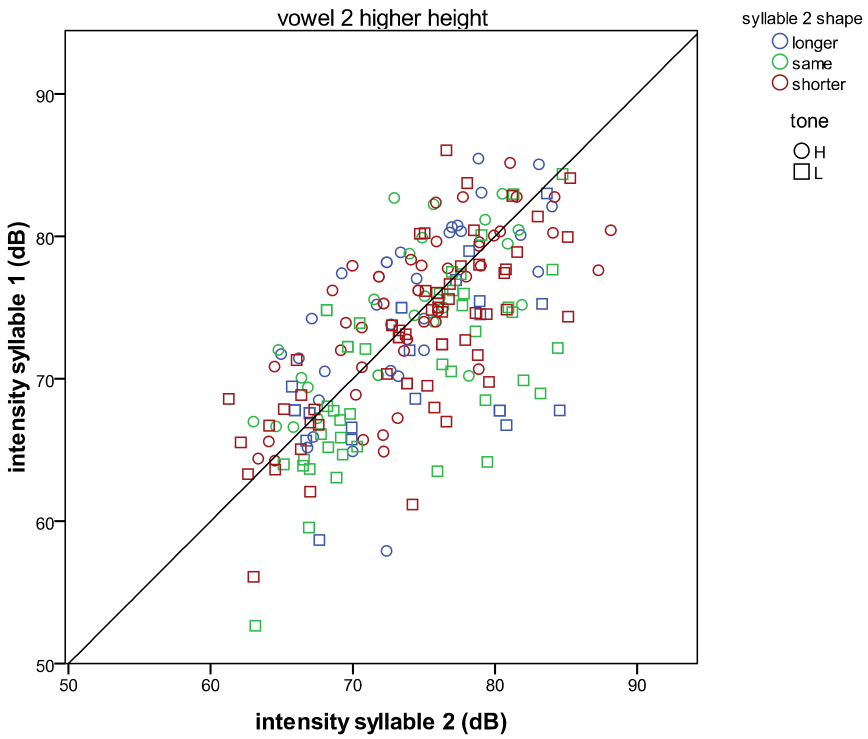


Figure 5-6: scatter plot of intensity; words with 2nd vowel height the same

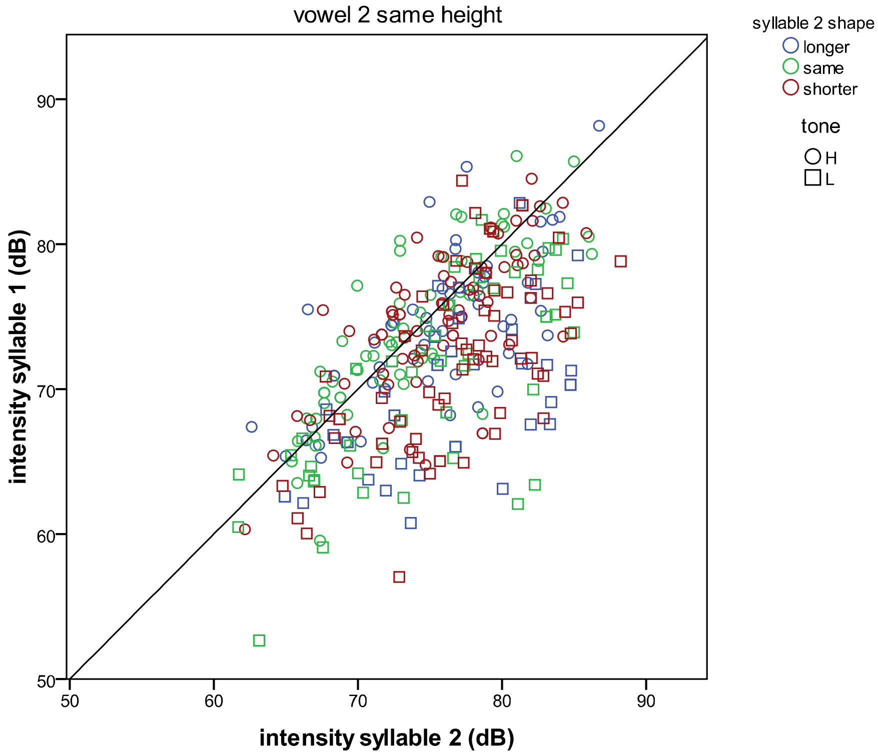
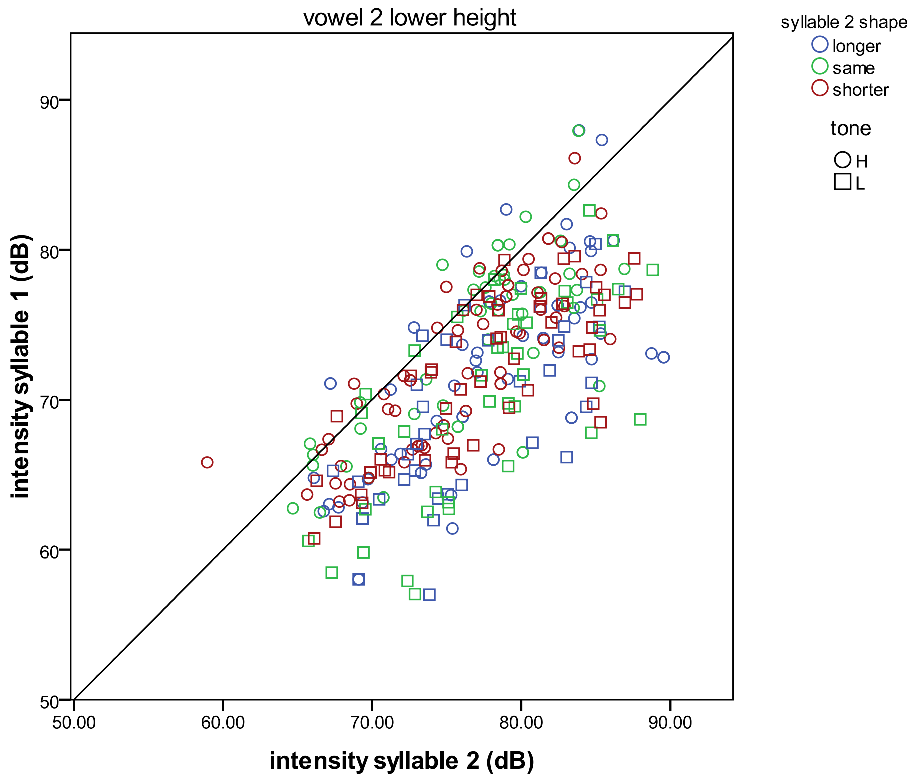


Figure 5-7: scatter plot of intensity; words with 2nd vowel height lower



It can be seen from the above scatter plots, when the vowel height is lower for the second vowel than the first vowel, in almost all instances the intensity is greater on the second syllable, which is to be expected. When the vowel heights are the same, it is still the case that the vast majority of words have greater intensity on the second syllable than the first syllable. When the vowel height of the second syllable is higher than the height for the first syllable the majority of words still have greater intensity on the second syllable, however the effect isn't as pronounced as when the vowel height is the same or lower. This is not surprising given that higher vowels generally have less intensity than lower vowels, and the fact that the majority of words with higher 2nd syllable vowels still have greater intensity on the second syllable is an indication of stress.

In order to investigate the significance of these trends, paired T-tests were carried out for intensity for first and second vowels, with the following results:

Table 5-6: paired T-test results comparing intensity of 1st and 2nd syllable

2nd vowel height	*tone*	*N*	*d.f.*	*t*	*P*	*Sig*
higher	H	101	100	2.085	0.040	•
	L	124	123	5.638	< 0.0005	•
same	H	194	193	2.663	0.008	•
	L	158	157	13.096	< 0.0005	•
lower	H	177	176	12.359	< 0.0005	•
	L	134	133	17.928	< 0.0005	•

In all cases, even when the second syllable vowel height is higher than the first, intensity is significantly affected by syllable position, with second syllable vowels in general having greater intensity than first syllable vowels.

5.2.5 Stress patterns on verbs

The work above has focused on non-verbs. Attention is now briefly turned to verbs. Most verb roots are monosyllabic. It was thus decided to consider a particular verb form where the same segmental representation can occur syntactically as both a noun and a verb. This is the verb root plus an affix with the segmental form *-en.* A verb plus affix *-en* can occur syntactically as a noun, meaning "someone doing ..." e.g.

dàŋ *ŋàj* *tɕ*^{h}*à* *túŋen* *t*^{h}*óŋdʑu*

yesterday 1SG.ERG tea drink.NMLZ see.PST.CONJT

'Yesterday I saw someone drinking tea.' lit: 'yesterday I saw a tea-drinker'

In this context it can also take nominal affixes, e.g. the plural affix, as in

dàŋ *ŋàj* *tɕ*^{h}*à* *túŋengwa* *t*^{h}*óŋdʑu*

yesterday 1SG.ERG tea drink.NMLZ.PL see.PST.CONJT

'Yesterday I saw some people drinking tea.'

The same segmental form can also occur as a verb[1]:

ŋà *tɕ*^{h}*à* *túŋen*

1SG tea drink.IPFV.CONJT

'I am drinking tea.'

In order to consider any difference in stress between nouns and verbs, a series of verbs with morpheme *-en* was recorded when a) it was functioning syntactically as a verb (e.g. "I am drinking tea") and b) when it was functioning syntactically as a noun (e.g. "I saw someone drinking tea"). Recordings were made for two speakers of the language (two of the four used for the main acoustic analysis in chapter 4).

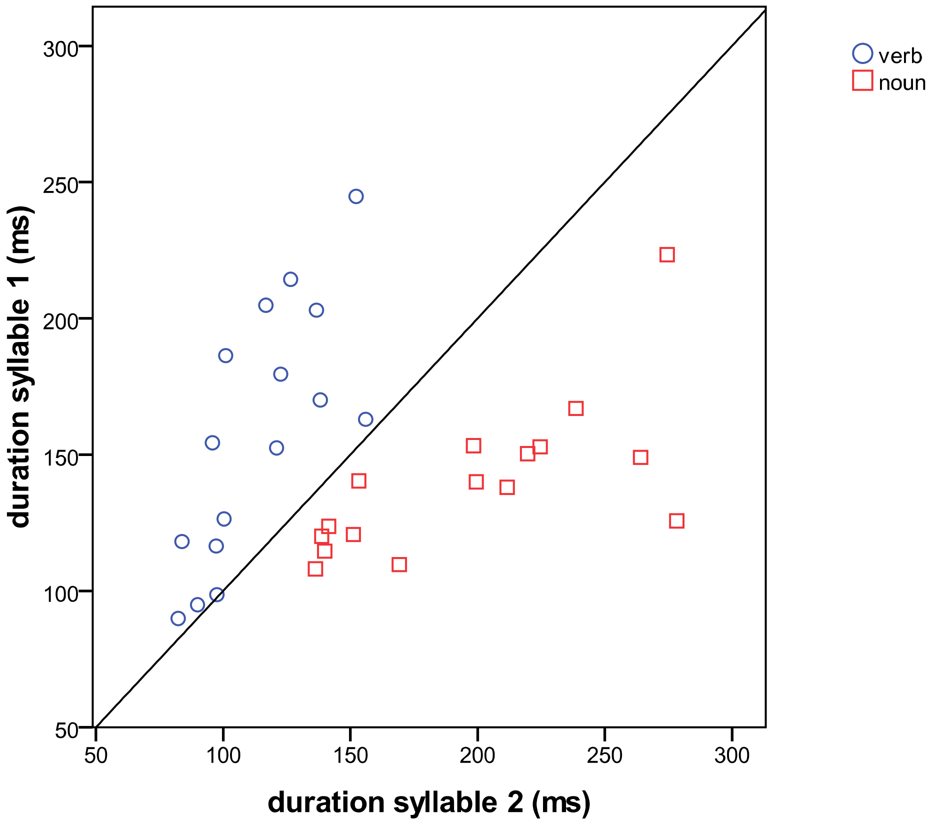
The duration of the syllable rhymes was considered, as that was one of the main indications of stress from the above analysis.

Words were separated according to whether the weight of the first syllable was light or heavy (the second syllable is the affix *-én* which is heavy). The following scatter plot is for words with a heavy first syllable. It shows durations of first and second syllable rhymes in the case of words functioning as verbs and the case of the

[1] It is a question for further research whether the segmental form /-en/ is the same morpheme in each case, or whether it is a different morpheme with the same segmental form.

same segmental forms functioning as nouns. The diagonal line is where the duration of syllable 1 is equal to the duration of syllable 2.

Figure 5-8: Scatter plot showing durations of syllable rhymes; heavy 1st syllable



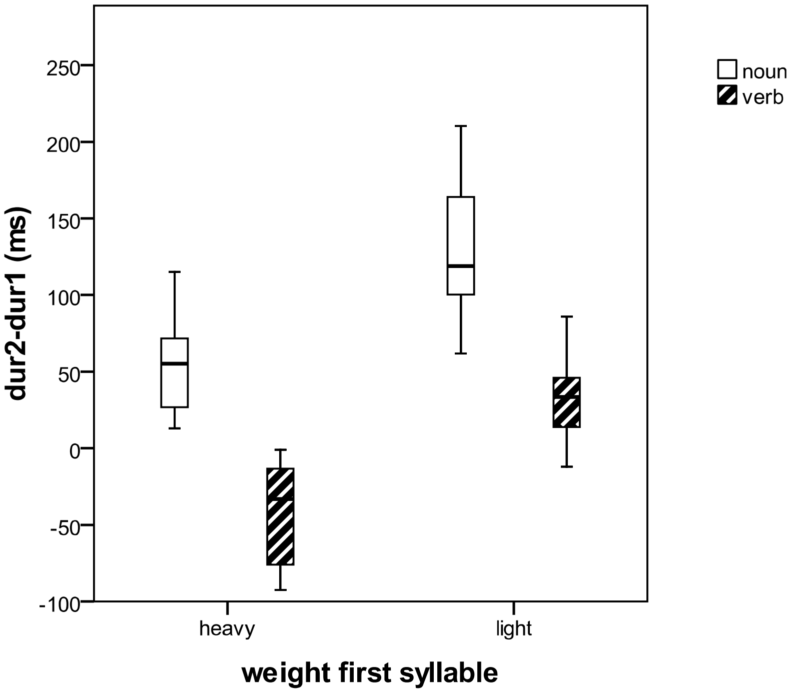
In every case the duration of the second syllable rhyme is greater than the first syllable rhyme when the word is functioning as a noun, and less than the first syllable rhyme when it is functioning as a verb. This is a strong indication that the stress is different for verbs and nouns. For nouns it is on the second syllable, and for verbs it is on the first syllable.

For words with light first syllable, clearly the second syllable rhyme is going to be longer in duration than the first. However, if there is a different stress pattern for verbs and nouns, the difference in duration between the syllables is likely to be different. If there is stress on the second syllable then the difference in duration will be greater than if there is stress on the first syllable.

For each word the difference in duration between syllable rhymes was calculated by subtracting the duration of the first syllable rhyme from the second syllable rhyme.

The following is a boxplot showing the duration difference for words functioning as verbs or nouns for different weight of the first syllable.

Figure 5-9: Boxplot showing difference in syllable rhyme duration



The above boxplot clearly shows that the difference in duration between the syllable rhymes is greater when the word is functioning as a noun than when it is functioning as a verb, indicating that verbs have first syllable stress whereas nouns have second syllable stress.

The acoustic analysis of stress on nouns also considered pitch and intensity. Intensity varied considerably according to the height of the vowel, so it was decided not to consider intensity here. In terms of pitch, the pitch of the final syllable of a sentence always falls to low regardless of the underlying tone, so comparing pitch between words functioning as verbs and nouns would not be a fair comparison. Thus the acoustic analysis of verbs only involved duration. However, the duration measurements show a very clear difference between words functioning as verbs and nouns. For verbs to have first syllable stress is not surprising, given that the first syllable is generally the verb root. The one case when the verb root is not the first syllable is with the negative morpheme, which is a prefix. In this case the stress is on the second syllable, which is the verb root.

5.2.6 Conclusion of acoustic analysis

In general, all polysyllabic non-verbs on their second syllable have a combination of greater duration, higher pitch, and greater intensity. For disyllabic words the differences between first and second syllable measurements for each of these acoustic parameters have been analysed statistically and are significant. These differences cannot simply be explained as correlates of high tone because they occur on the second syllable even when both the first and second syllables have high tone. The conclusion is that as a general pattern Walungge has second syllable stress for words which are not verbs. Verbs have stress on their root, which is in general the first syllable.

This predominant pattern of second syllable stress is important when considering the attachment of tones to TBUs in Walungge. The section which immediately follows draws on this as the basis for Walungge forming a disyllabic right-headed tonal foot with subsequent syllables excluded from the tonal domain. Second syllable stress is also important when considering the relationship between tone and pitch melody in low tone words with heavy second syllables (section 5.3.6.4). Not only can second syllable stress be analysed as the basis for the tone patterns, but it is also a relevant perceptual cue in word identification. Chapter 7 below details a series of perceptual experiments for tone. In disyllabic situations with a somewhat rising pitch melody (see section 7.3.1.3), the rise appears to be perceived as second syllable stress rather than the rise of an LH tone pattern, causing the majority of subjects to identify stimuli as tone H words. A further area which draws upon stress patterns is the analysis of glides (section 3.4.2 above).

5.3 Surface and underlying tone patterns in Walungge

5.3.1 Introduction

Chapter 4 above gives an acoustic analysis of the pitch melodies of Walungge non-verbal content words. In the case of monosyllabic monomorphemic words, for each syllable weight there is a two-way contrast in the pitch melody, but the pitch melodies also vary depending upon syllable weight. The following are examples.

Table 5-7:Walungge pitch melodies

shape of word	*example*	
monosyllabic; light syllable	[ŋa ^{52}]	'drum'
	[ŋa ^{24}]	'1SG'
monosyllabic; heavy syllable	[naː ^{44}]	'blue sheep'
	[neː ^{14}]	'barley'

From the acoustic analysis in chapter 4, the conclusion was drawn that the steeply falling [52] pitch melody is a realisation of tone pattern HL. The fall could not be explained as an utterance final or phrase final effect because it is there even in non final contexts when the slight falls which are on other isolation words disappear. Neither could it be explained as phonetic enhancement to the distinction between long and short vowels. Similarly the [24] and [14] rises occur in non phrase final contexts. The difference in these two melodies was explained as a phonetic effect due to the length of the vowel. The acoustic analysis gave an initial conclusion that the pitch melodies on monosyllabic words are realisations of tone patterns H, HL, and LH, but without touching on the question of whether these are surface or underlying tone patterns.

The acoustic analysis of polysyllabic words was also considered in chapter 4. In summary, the first syllable has high or low pitch, the second has high pitch, and subsequent syllables have low pitch, e.g.

disyllabic	[samba $^{44\ 44}$]	'new'
	[samba $^{11\ 44}$]	'bridge'
three syllables	[orokpa $^{44\ 44\ 21}$]	'toad'
	[barakpa $^{11\ 44\ 21}$]	'big'
four syllables	[tɕemalaʑoʔ $^{44\ 44\ 21\ 11}$]	'butterfly'
	[dʑaŋbagaze $^{11\ 44\ 21\ 11}$]	'lizard'

The purpose of this section is to account for the surface pitch melodies of Walungge from a phonological perspective. Firstly several possible alternatives are presented all of which go some way to explaining the data, but are ruled out because they fail to completely account for all the data. Following this, the concept of a tonal foot is introduced and an analysis is presented using a tonal foot, which accounts for all the data.

5.3.2 Stress or tone

Section 5.2 above considers disyllabic words, and demonstrates through acoustic analysis that Walungge typically has second syllable stress on nouns. Given this, there is a question of how much of the pitch melody on disyllabic words is the result of underlying lexical tone, and how much is the result of stress. For disyllabic nouns, one alternative is to say that the underlying tone attaches to the first syllable of the word only, and the second syllable high pitch is the result of stress rather than tone. This would mean that the underlying tone pattern for disyllabic words starting with low tone is L rather than LH, e.g.

```
L            L
             |
σ ˈσ         σ  ˈσ
sàmba   →   [samba ¹¹ ⁴⁴]  'bridge'
```

Words of more than two syllables would similarly be accounted for by saying that the tone attaches to the first syllable only, the high pitch on the second syllable is due to stress, and the low pitch on subsequent syllables is due to lack of stress, e.g.

```
H                    H
                     |
σ ˈσ σ σ             σ ˈσ σ σ
tɕémaladʑok    →    [tɕemalaʑoʔ ⁴⁴ ⁴⁴ ²¹ ¹¹ ]  'butterfly'
```

However, although the typical pattern on nouns is second syllable stress, section 5.2 also shows that there are a minority of nouns which have first syllable stress, e.g.

[ˈtɕʰuwa $^{44\ 44}$] 'tap'

[ˈʈʰukpa $^{11\ 44}$] 'sixth one'

In these instances there is high pitch on the second syllable even when there is stress on the first syllable. Thus the second syllable high pitch cannot be the result of stress. In addition to this, section 5.2 above considers the stress patterns of verbs as well as nouns, showing that verbs have stress on their root which is typically the first syllable. The following data shows verbs with stress on the first syllable and high pitch on the second syllable. In these instances, also, the high pitch on the second syllable cannot be the result of stress.

[ˈduŋgi $^{11\ 44}$] ‘hit.IPFV’

[ˈtuŋgi $^{44\ 44}$] ‘drink.IPFV’

Thus it is not possible to account for the polysyllabic pitch melodies simply by stress.

5.3.3 Underlying and non underlying tones

Having determined that stress alone cannot account for the surface pitch melodies, another alternative is a “template” tone system such as proposed by Sun (1997) for the analysis of Lhasa Tibetan. Sun’s analysis of Lhasa Tibetan is outlined in section 2.7.3 above. In summary, he proposes a word template where all non-initial syllables take a non-underlying default H and the first syllable of the word takes the underlying tone. Clearly this would have to be modified for Walungge, by saying that the second syllable takes H and third and subsequent syllables take L (or are toneless, with the low pitch the result of them being toneless rather than the result of L). Examples of polysyllabic words in Walungge would then be as follows:

$$\begin{array}{lcl}
\text{L} & & \text{L H} \\
 & & | \ \vdots \\
\sigma\ \sigma & & \sigma\ \sigma \\
\textit{sàmba} & \rightarrow & \text{[samba}\ ^{11\ 44}\text{] ‘bridge’}
\end{array}$$

$$\begin{array}{lcl}
\text{H} & & \text{H H L L} \\
 & & | \ \vdots\ \vdots\ \vdots \\
\sigma\ \sigma\ \sigma\ \sigma & & \sigma\ \sigma\ \sigma\ \sigma \\
\textit{tɕémaladʑok} & \rightarrow & \text{[tɕemalaʑoʔ}\ ^{44\ 44\ 21\ 11}\text{] ‘butterfly’}
\end{array}$$

However, one of the main differences between Lhasa Tibetan and Walungge regarding tone is that whilst Sun analyses Lhasa Tibetan as having level H and L for monosyllabic words, Walungge has monosyllabic rises and falls which cannot be explained away as phrase final effects or phonetic enhancements. In order to use the concept of a word template for Walungge, the monosyllabic rises and falls need to be accounted for. One option would be to say that Walungge polysyllabic words have underlying tones H or L but monosyllabic words have underlying tones H, HL, and LH. Whilst this would explain all the pitch melodies it has a problem in the lack of underlying tone L for monosyllabic words. The acoustic analysis of chapter 4 shows that monosyllabic pitch melodies [14] and [24] are best analysed as realisations of the same tone pattern rather than as realisations of LH and L respectively. The [24] rise is always present in non phrase final situations just as the [14] rise is always present. Alternatively, if the monosyllabic rise was taken to be the result of interaction between underlying L and stress (rather than underlying LH), monosyllabic words could be analysed as having tones H, HL, and L. However, this creates a discrepancy with disyllabic words which may have both L and stress on the first syllable and yet the first syllable pitch is low and level rather than rising.

5.3.4 Underlying L and LH

Section 2.7.3 outlines two alternative analyses of Lhasa Tibetan: Duanmu (1992) and Sun (1997). Having considered an application of Sun's analysis to Walungge in the section above, this section considers an application of Duanmu's analysis. Whereas Sun analyses Lhasa Tibetan as having underlying tones, H and L, Duanmu analyses Lhasa Tibetan as having underlying H and LH. If Walungge were to be analysed as having H and LH, polysyllabic pitch melodies could be accounted for, e.g.

```
H                      H
                      /\
σ σ σ σ              σ σ σ σ
tɕémaladʑok    →   [tɕemalaʑoʔ 44 44 21 11 ]  'butterfly'

LH                     L H
                       | |
σ σ σ                  σ σ σ σ
bàrakpa        →   [barakpa 11 44 21 ]  'big'
```

Similarly [44] and [14]/[24] monosyllabic words can be accounted for, e.g.

H   H
σ   σ
sáŋ   →   [saŋ ^{44}] ‘pine’

L H   L H
σ   σ
sàŋ   →   [saŋ ^{14}] ‘saucepan’

However, [52] would have to be accounted for by proposing an underlying HL:

H L   H L
σ   σ
tá   →   [ta ^{52}] ‘horse’

Whilst this would account for the data, it still remains that HL and H are in complementary distribution, with HL occurring on monosyllabic light syllable words and H occurring on monosyllabic heavy syllable words, e.g.

[sa ^{52}]   ‘earth’

[saŋ ^{44}]   ‘pine’

Because of this complementary distribution an analysis is sought which accounts for HL and H as two different surface patterns of the one underlying tone.

The sections that follow present an analysis based on the concept of a tonal foot that argues for Walungge having two underlying tones H and L, with HL and LH being surface tone patterns. The formation of a tonal foot, plus the insertion of boundary tones, accounts for all the tone patterns found in Walungge, both in monomorphemic words and across morpheme boundaries.

5.3.5 Tonal Feet

Some languages with a stress and tone interaction form standard metrical feet in the same way that a language without any tone might form metrical feet; for example a language such as Kera (Pearce (2007)), outlined in section 5.1 above, shows the parsing of syllables in a word into standard metrical feet. In Kera the syllables are parsed into iambic feet which, typically for iambs, require the head of the foot to be a heavy syllable. The placement of tone is determined by this metrical structure. However, some languages show a placement or grouping of tones which in some respects can be accounted for by the formation of feet, but in other respects these languages might not show the properties of a standard metrical system. This introduces the concept of a 'tonal foot'.

Leben (2001) uses the concept of a tonal foot to account for tone assignment in Hausa for loan words from English, and for tone distribution in Bambara. A tonal foot can be viewed as a structure parallel to a metrical foot, with a metrical foot assigning stress and a tonal foot assigning tone. However, whilst metrical feet are very well documented and by definition have one stress per foot, the concept of a tonal foot is an emerging concept and the makeup of tonal feet is not consistent across languages (Leben (2001 p. 1)).

Leben (p. 5) gives the following diagrams to illustrate that a tonal foot can refer to the prosodic structure on the segmental tier, on the tonal tier, or on both. The diagrams illustrate the tonal foot as a grouping of the segments and/or tones. Though, as Leben points out, segments belong to syllables and perhaps other units, and tones associate to TBUs, so the diagrams are illustrative only.

Footing on the segmental tier:

(gwamna) 'governor'
  | |
  H   L

Footing on the tonal tier

gwamna
 | |
(H   L)

Footing on both the segmental and tonal tiers

(gwamna)
| |
(H   L)

If the grouping is on the segmental tier, then syllables can be organised into metrical feet, into tonal feet, or into both.

For the assignment of tones to English borrowings in Hausa, Leben gives the following rules for tonal foot formation.

a. At the left edge of each English syllable perceived as stressed in Hausa, initiate a tonal foot:

(gwamna	garan(tii	(tan(kiifaa
‘governor’	‘guarantee’	‘timekeeper’

b. Locate the right edge of the foot so as to make it maximally binary:

(gwamna)	garan(tii)	(tan)(kiifaa)

c. A tonal foot has tones HL. Unfooted syllables have L:

(gwámnà)	gàràn(tîi)	(tân)(kíifàa)

Other languages have been analysed as assigning tone based on a foot structure that would fit the concept of a tonal foot. One of particular relevance to the analysis of Walungge is the analysis by Meredith (1990) of Refugee Standard Tibetan (RST), which is described in section 5.1.3 above. The analysis that he presents for the assignment of tone fits the concept of a tonal foot: the language is constructing feet for the purpose of assigning tones. In RST, the tones then influence the placement of stress.

In summary, the concept of a tonal foot can be very useful for describing the assignment of tone in a language. It is a grouping on the segmental tier and/or the prosodic tier for the purposes of assigning tones to TBUs. It has parallels with a metrical foot, although it can vary considerably from language to language. Because it is not as precisely defined as a metrical foot, there is considerable scope for a language specific definition. The analysis which follows uses the concept of a tonal foot in order to account for the assignment of tones in Walungge.

5.3.6 Tonal foot formation in Walungge

5.3.6.1 Summary

As discussed above, Walungge has rising, falling, and level pitch melodies. These can be accounted for by proposing that Walungge forms tonal feet in order to assign tones. The proposal not only accounts for pitch melodies across monomorphemic words, but can also account for pitch melodies across morpheme boundaries. The following gives a bare summary of how tonal feet are formed in Walungge, and how the tones attach to the TBUs. This summary is then followed by a discussion of the justification for each step in the foot formation process, along with detailed derivations of the surface forms for a wide variety of words, including those which incorporate morphological boundaries.

a. Form a single right-headed foot from the leftmost edge of the word. This foot is maximally binary on syllables and is the domain of the tone. (Equivalent to a metrical iambic foot.)

b. Attach the underlying tone to the first mora of the foot.

c. Attach a boundary tone to the final mora of the foot. The boundary tone is polar to the underlying tone if both the boundary and the underlying tone are on the same mora. Otherwise the boundary tone is H regardless of the underlying tone.

d. Spread the underlying tone to all free moras in the tonal foot.

e. Any unfooted moras are outside of the tonal domain. These surface with low pitch.

5.3.6.2 Basic foot formation in polysyllabic words

One of the main motivations for a tonal foot proposal for Walungge is the high pitch on the second syllable of polysyllabic words, which cannot completely be accounted for by stress. It is this second syllable high pitch which is typical of many Tibetan languages and has been a source of motivation for the word template proposal of Sun (1997) and the footing process proposed by Meredith (1990) for Refugee Standard Tibetan.

The proposed formation of a Walungge tonal foot is that it is maximally binary on syllables counted from the left-most edge, and is right-headed. The following examples illustrate this for words of two, three, and four syllables. The head of the foot is underscored.

(σ $\underline{\sigma}$)	(σ $\underline{\sigma}$) σ	(σ $\underline{\sigma}$) σ σ
làkpa 'hand'	*sàdukpa* 'nettle'	*tɕ^{h}émaladʑok* 'butterfly'

At this stage, and with these few words, this looks like a single metrical iambic foot. Indeed, it is likely that it is the predominant pattern of second syllable stress which has given rise to a tonal foot. The above acoustic analysis shows that stress on non-verbs typically occurs on the second syllable. The stress on verbs occurs on the root (which is usually the first syllable). However, stress never occurs on the third or subsequent syllables. Caplow (2009) argues that proto-Tibetan had second syllable stress for nouns, and this is the source of second syllable high tone in tonal Tibetan languages. Applying this to Walungge, in the light of the synchronic form of Walungge non-verbs having both stress and high tone on the second syllable it is probable that pre-tonogenesis Walungge had second syllable stress for non-verbs. As tone has entered the language, the second syllable stress has given rise to second syllable non-lexical high tone. This has become the tonal pattern for all words, both verbs and non-verbs, including words with stress on the first syllable. The lack of stress on third and subsequent syllables has resulted in these syllables being excluded from the tonal domain; the predominant pattern of stress on the second syllable has resulted in the first two syllables becoming the domain for tone.

The next stage in the process is stated as attaching the underlying tone to the first mora of the foot. The three words in the illustration immediately below have been deliberately chosen so as to have every syllable monomoraic; thus for illustrative purposes now, the tone will be attached to the syllable. The following section discusses why the mora rather than the syllable has been chosen as the tone bearing unit.

L	L	H
|	|	|
(σ <u>σ</u>)	(σ <u>σ</u>) σ	(σ <u>σ</u>) σ σ
làkpa 'hand'	*sàdukpa* 'nettle'	*tɕʰémaladʑok* 'butterfly'

A boundary tone is attached to the final mora of the foot. Because the final mora is the head of the foot it attracts high tone, written H%:

L H%	L H%	H H%
| |	| |	| |
(σ <u>σ</u>)	(σ <u>σ</u>) σ	(σ <u>σ</u>) σ σ
[lakpa $^{11\ 44}$]	[sadukpa $^{11\ 44\ 21}$]	[tɕʰemalaʑoʔ $^{44\ 44\ 21\ 11}$]
'hand'	'nettle'	'butterfly'

The third and subsequent syllables are outside of the tonal domain, which is the foot. They surface with low pitch. If the tonal domain is indeed the foot, which is what is being proposed here, then the low pitch on these syllables is the result of them being toneless rather than the attachment of any non-underlying low tone.

5.3.6.3 Moras and syllables

For simplicity in introducing the concept of a tonal foot, the previous section showed the tones attaching to syllables. However, the evidence from monosyllabic words is that the tone system is taking account of the number of moras in the syllable. The following are examples of pitch melodies on monosyllabic words with high tone. Short open syllables and syllables with a plosive coda pattern together, as do long vowels and syllables with a sonorant coda:

[ɕa ^{52}] ‘meat’

[k^{h}ap˺ ^{53}] ‘needle’

[ɕaː ^{44}] ‘deer’

[saŋ ^{44}] ‘pine’

The difference in level versus falling pitch melody can be accounted for by saying that the tone bearing unit is the mora. The words [ɕaː ^{44}] and [saŋ ^{44}] are bimoraic. A tonal foot is formed from the single syllable, the underlying tone is attached to the first mora, and a boundary tone H% is attached to the second mora. This results in the observed high level pitch melody[2].

H H H%

μ μ (μ μ)

σ σ

ɕáː → [ɕaː ^{44}] ‘deer’

For monomoraic words, however, the situation is slightly different. In the summary of the tonal footing process (section 5.3.6.1 above), it was stated that if the underlying tone and the boundary tone attach to the same mora, then the boundary tone is polar to the underlying tone. In all the examples seen so far there has been more than one mora in the word so the boundary tone has been H% regardless of the underlying tone. In the case of monomoraic words, the underlying tone and the boundary tone must share a mora. Thus, if the underlying tone is H, the boundary tone will be a polar L. This is possibly due to an OCP (obligatory contour principle) constraint preventing two identical tones on the one TBU. The following diagram illustrates a monomoraic word with underlying H:

[2] It can be noted in the monosyllabic diagrams that the footing has been shown at the level of the moras rather than the syllable. The reason for this is discussed in greater detail in section (5.3.6.4 below).

H $\qquad\qquad\qquad$ H L%

μ $\qquad\qquad\qquad$ ($\underline{\mu}$)

σ $\qquad\qquad\qquad$ σ

tá $\qquad$ → $\qquad$ [ta ^{52}] ‘horse’

The above example shows a foot, which, if it was a metrical foot, would be a degenerate foot, i.e. a foot of the logically smallest size possible. Hayes (1995:86-105) discusses in detail degenerate feet in metrical systems, noting that whilst in some languages they are permissible, there are many more languages in which they are not permissible, and for some languages there is a “repair” process which changes a degenerate foot into a permissible foot, e.g. by vowel lengthening. However, in Walungge these are tonal feet rather than metrical feet. Even in monomoraic words there is underlying tone which attaches; which means that if the tonal domain is the tonal foot then even a monomoraic word must form a foot.

The next example is of a monosyllabic bimoraic noun with underlying tone L. The derivation is the same as the other examples above of words with more than one mora. The underlying tone attaches to the first mora, and an H% boundary tone attaches to the second mora.

L $\qquad\qquad\qquad$ L H%

μ μ $\qquad\qquad\quad$ (μ $\underline{\mu}$)

σ $\qquad\qquad\qquad$ σ

sàŋ $\qquad$ → $\qquad$ [saŋ ^{14}] ‘saucepan’

The final monosyllabic structure is a monomoraic word with tone L. This shows nothing very different from what has already been described. Both the underlying tone and the boundary tone attach to the single mora of the foot. Because the underlying tone is L, the boundary tone is a polar H%.

```
L              L  H%
               \ /
μ              (μ̲)
|               |
σ               σ
rà      →      [ra ²⁴]  'goat'
```

Taking the mora rather than the syllable as the TBU gives the following diagrams for the polysyllabic words of the section above. Extra-tonal moras are marked in brackets < >.

```
L                 L   H%
                  |   ¦
μ  μ              (μ   μ̲)
|  |               |   |
σ  σ               σ   σ
làkpa       →     [lakpa ¹¹ ⁴⁴]  'hand'
```

```
L                 L   H%
                  |   ¦
μ  μ  μ           (μ   μ̲)  ‹μ›
|  |  |            |   |    |
σ  σ  σ            σ   σ    σ
sàdukpa     →     [sadukpa ¹¹ ⁴⁴ ²¹]  'nettle'
```

```
H                      H  H%
                       |  ¦
μ  μ  μ  μ             (μ   μ̲) ‹μ› ‹μ›
|  |  |  |              |   |   |   |
σ  σ  σ  σ              σ   σ   σ   σ
tɕʰémaladʑok     →     [tɕʰemalaʑoʔ ⁴⁴ ⁴⁴ ²¹ ¹¹]  'butterfly'
```

Further evidence that the TBU is the mora comes from polysyllabic words with heavy second syllable (see section immediately following) and patterns across morpheme boundaries (see 5.4.4).

5.3.6.4 Syllables of differing weight

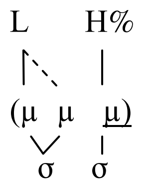
Having given an overview of the footing process for a variety of words, attention is now turned to the more complex case of polysyllabic words where at least one of the syllables is heavy.

Consider the pitch melodies on the following low tone words:

[ɲima $^{11\ 44}$]	‘sun’
[siːwa $^{11\ 44}$]	‘dew’
[ʈʰomaː $^{11\ 344}$]	‘potato’
[gøːʑaː $^{11\ 344}$]	‘lock’
[sadukpa $^{11\ 44\ 21}$]	‘nettle’
[biliŋma $^{11\ 344\ 21}$]	‘worm’

It is clear from the above words that the footing process is counting syllables rather than moras. Regardless of the number of moras in each syllable the pitch on the first syllable is low, the pitch on the second syllable is high (or rising to high) and the pitch on subsequent syllables is low. This raises some interesting issues relating to the difference between the proposed tonal feet and typical metrical iambic feet.

Under the proposal of a tonal foot, a word such as [siːwa $^{11\ 44}$] ‘dew’ can be illustrated by the following diagram:



[siːwa $^{11\ 44}$] ‘dew’

The underlying tone L attaches to the first mora of the foot and H% attaches to the final mora, as already described above; tone L then spreads to the spare mora. This results in low pitch on the first syllable and high pitch on the second syllable, which is a discrepancy between the tonal feet proposed in this chapter and typical metrical iambic feet. A fundamental principle of metrical systems is the “iambic-trochaic law”:

elements contrasting in intensity naturally form groupings with initial prominence whereas elements contrasting in duration naturally form groupings with final prominence (Hayes (1995:80)). Metrical systems will generally parse these elements such that the elements of greater intensity/duration are the heads of feet with the following inventories of feet (listed in Kager (2007)):

	licit feet	degenerate feet
Syllabic trochee:	($\underline{\sigma}$, σ)	(σ)
Moraic trochee:	($\underline{\sigma}_{\mu}$,σ_{μ}), ($\underline{\sigma}_{\mu\mu}$)	(σ_{μ})
Iamb:	(σ_{μ},$\underline{\sigma}_{\mu\mu}$), ($\underline{\sigma}_{\mu\mu}$), ($\sigma_{\mu}$,$\underline{\sigma}_{\mu}$)	(σ_{μ})

Applying this to Walungge and forgetting about tone for the minute, if Walungge was a typical iambic system then a word such as *si:wa* might be expected to be parsed into feet: *(si:)wa* (if degenerate feet were disallowed) or *(si:)(wa)* (if degenerate feet were allowed), but not *(si:.wa)*. Kager (1993) discusses the case of truly quantity-insensitive iambic systems, that is to say systems which have duration distinctions but form iambs without regard to these distinctions. Although such systems are rare, they do exist and Kager outlines several such languages. A quantity-insensitive iambic system would allow the parsing of *si:wa* into an iambic foot *(si:.wa).* However, it has already been demonstrated above that, in respect to tone, Walungge is not quantity-insensitive. This highlights one difference between typical metrical iambic feet and the proposed tonal feet for Walungge. Whereas a foot *(si:.wa)* might be problematic if it were an iambic foot given the rarity of quantity-insensitive iambic systems, the formation of tonal feet in the manner described above enables the empirical data to be accounted for.

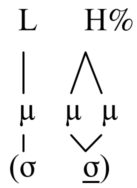
A further difference between the proposed tonal feet and typical iambic feet can be seen from the difference in second syllable pitch melodies of low tone words with light and heavy second syllables, e.g.

[ɲima $^{11\ 44}$] ‘sun’

[t̪ʰoma: $^{11\ 344}$] ‘potato’

In both instances the first syllable has low pitch due to the low tone. When the second syllable is light it has high pitch across it, whereas when the second syllable is heavy the pitch continues to rise through the first part of the syllable. This difference is transcribed as [44] versus [344]. An important point to note is that the acoustic analysis of section 4.3.4 showed that the difference in second syllable starting pitch is significant, and that it cannot be explained by peak delay; it has to be explained by the phonology[3].

If the feet formed were typical iambic feet, and H% was stress rather than tone, the foot formation and attachment of H% could be illustrated as:



A disyllabic foot is formed, with the second syllable (rather than the final mora) as the head. H% as stress would be realised as high pitch (and/or longer duration/greater intensity) across the whole of the second syllable. However, this would result in a [11 44] pitch melody such as in [ɲima $^{11\ 44}$] ‘sun’ rather than the observed [11 344] pitch melody.

On the other hand, if the tonal foot is formed from the TBUs with the final mora (rather than the final syllable) as the head of the foot, the structure with the initial attachment of L to the first mora and H% to the final mora would look like:

---

[3] As an aside, this rise for heavy second syllables is not unique to Walungge; Lhasa Tibetan also has a second syllable rise for low tone words with heavy second syllable (Duanmu (1992).

L H%

(μ μ $\underline{\mu}$)

σ σ

To account for L spreading to the middle mora, it can be said that in the attachment of tones to TBUs the underlying tone attaches to the first mora and spreads to all moras except the head mora:

L H%

(μ μ $\underline{\mu}$)

σ σ

[t̪^{h}oma: $^{11\ 344}$]

The final aspect of the pitch melody that needs to be accounted for is that the heavy syllables have a [34] rise. The attachment of L to the first mora of the second syllable and H% to the second mora might be expected to be realised with pitch melody [14], just as monosyllabic bimoraic words with L-H% are realised with [14] pitch (e.g. [ne: ^{14}] 'barley'). However, the actual pitch melody on the second syllable of the disyllabic word is [344][4]. This discrepancy in pitch melody can be accounted for by second syllable stress. Stress has not been mentioned up until this point because in the previous examples it has not had any effect upon the pitch melody; the pitch melody has been able to be accounted for purely by the underlying and the boundary tones. However, in this particular example the pitch melody can be accounted for by saying that stress raises the starting pitch of the syllable. This is in keeping with the acoustic properties of stress, which have already been discussed in section 5.2 above.

[4] The pitch levels 1 to 5 are abstractions of F0, so it is not so much a question of whether the starting pitch is level 1, 2, or 3, but rather that the second syllable of the disyllabic word has pitch which starts significantly higher (transcribed as level 3) than it does in the case of the monosyllabic word (transcribed as level 1), and it starts significantly lower than the starting pitch of a monomoraic second syllable (transcribed as level 4).

5.4 Tone patterns across morpheme boundaries

5.4.1 Prosodic structure

The prosodic hierarchy

Before considering in Walungge the ways in which morphemes attach to stems and the resulting tone patterns, it is helpful to first consider the prosodic hierarchy because this is influencing both the segmental patterns and the tonal patterns found at morpheme boundaries in Walungge. The foot and the syllable, which are the two lowest constituents of the prosodic hierarchy listed below, have already been seen in Walungge in the preceding sections, with feet being built from syllables. Just as feet are comprised of syllables, phonological words are comprised of feet, phonological phrases are comprised of words, etc, the largest constituent in the hierarchy below being the phonological utterance. The constituent which is of particular relevance to the analysis of tone in Walungge is the phonological word or pword.

The following information is taken from Hall (1999), who draws on authors including Selkirk (1980), Booij (1983), Selkirk (1984) and Nespor & Vogel (1986). The phonological, or prosodic, word (pword) is one constituent in a series of prosodic constituents arranged in a hierarchy:

phonological utterance
|
intonational phrase
|
phonological phrase
|
(clitic group)*
|
phonological word
|
foot
|
syllable

*Clitic group is bracketed because some authors, such as Nespor & Vogel (1986), propose a clitic group between the phonological word and the phonological phrase, however other authors (e.g. Hall (1999)) do not include it.

Importantly, the pword is distinct from the morphological or grammatical word (gword), which is defined in Nespor & Vogel (1986) as a unit that corresponds to the terminal node of a syntactic tree, and in Dixon & Aikhenvald (2002) as a number of grammatical elements which a) always occur together rather than scattered through the clause, b) occur in a fixed order and c) have a conventionalised coherence and meaning. Because a pword is distinct from a gword, this means that a single gword may be comprised of two or more pwords (e.g. each part of a compound word, or a stem plus affix), or alternatively two gwords may form a single pword (e.g. a lexical word plus a clitic). However, it is generally agreed that pword boundaries must align with morphosyntactic boundaries. This means that if a grammatical word (gword) is comprised of more than one pword, the pword boundary corresponds to a morpheme boundary. It also means that a monomorphemic gword will not consist of more than one pword.

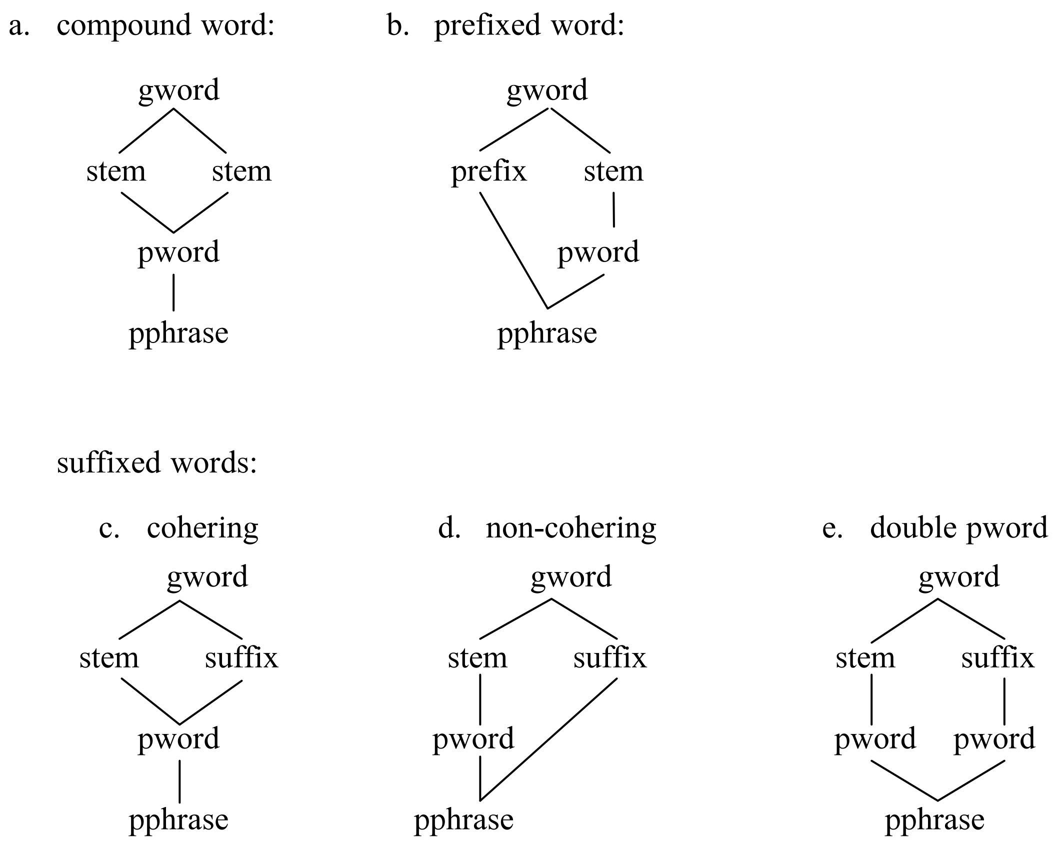
A debated principle of prosodic phonology is the "strict layer hypothesis" (Selkirk (1984)), where a prosodic constituent of rank n is immediately dominated by a single constituent of rank n + 1 (syllable being the lowest rank and phonological utterance being the highest, in the above diagram). This means, for example, a foot cannot be directly dominated by a phonological phrase (pphrase); it must be dominated by a pword. It also means, for example, a pword cannot be dominated by another pword (recursivity). However, more recent work (including Selkirk (1995), Booij (1996), Peperkamp (1996), Peperkamp (1997)) argue that both the skipping of levels and recursivity are allowed.

One of the ways that languages differ is the relationship between morphemes and pwords. In one language a stem plus affix might form only one pword, whereas in another language (or even in the same language) a stem plus affix might be two separate pwords. The following section gives a brief outline of the relationships between stems and affixes in Kyirong Tibetan, which is very helpful to consider before considering Walungge, because of similarities between the languages.

Pwords and gwords in Kyirong Tibetan

Hall & Hildebrandt (2008) outline different relationships which exist between affixes and stems involving the gword, the pword and the pphrase in Kyirong Tibetan. These relationships can be illustrated by the following diagrams:

Figure 5-10: relationships between gwords and pwords in Kyirong Tibetan



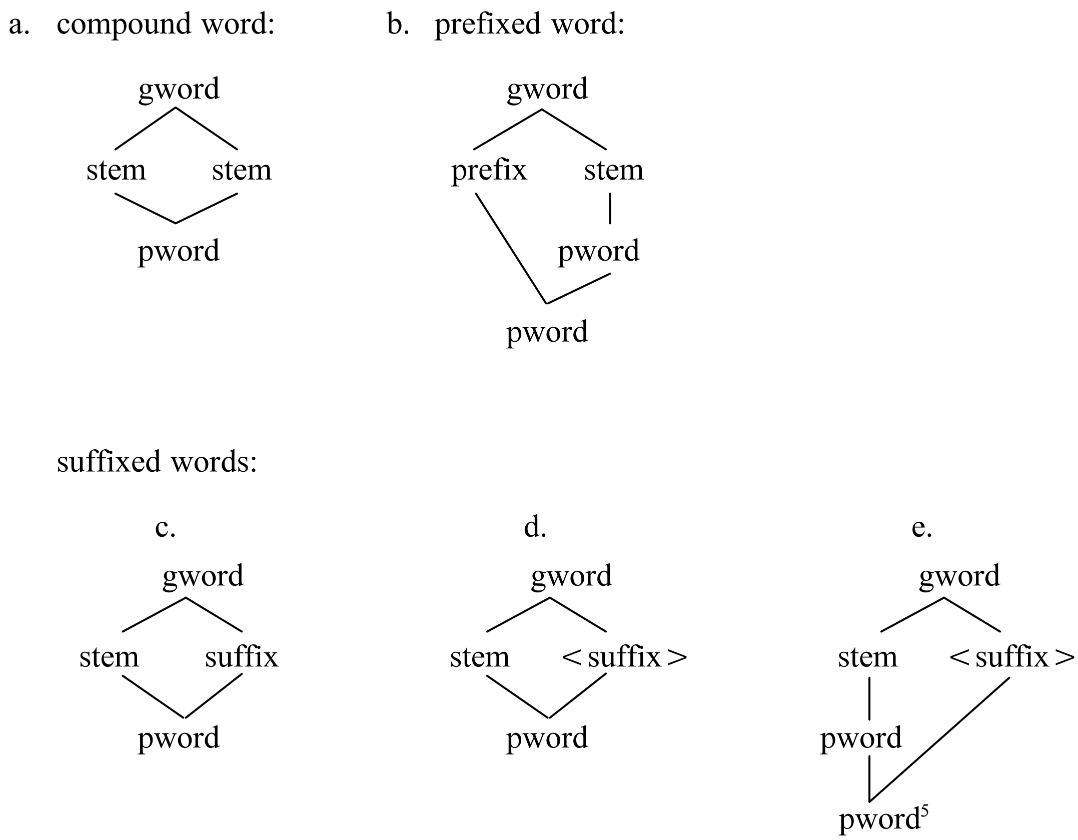
Diagrams a. and c. show a gword matching with a pword: the two stems or the stem plus suffix of the gword together form a pword. Hall & Hildebrandt argue that Kyirong Tibetan is typologically unusual due to the two parts of a compound word forming a single pword rather than separate pwords. Diagrams b. and d. show an affix which is outside of the pword, but is part of the pphrase. Hall & Hildebrandt note that these cases could alternatively be analysed by forming a clitic group with the stem rather than a pphrase, or by the formation of a recursive pword rather than a pphrase. Diagram e. shows the stem and suffix each an individual pword, but together forming a pphrase. Because stem and suffix relationships in Kyirong Tibetan can be either c., d., or e., the language requires suffixes to be lexically marked for which type of prosodic structure is required.

The considerations that Hall & Hildebrandt use to determine the pword boundaries are phonotactic constraints (e.g. aspirated phonemes only occur word initially), and tone (e.g. rising and falling contours only occur on monosyllabic words or on the word final syllable of polysyllabic words).

Prosodic constituents and gwords in Walungge

Walungge shows similar relationships between prosodic constituents and gwords to those of Kyirong Tibetan. The diagrams below summarize the different relationships between prosodic constituents and gwords that have been found in Walungge.

Figure 5-11: relationships between gwords, pwords and feet in Walungge



Whilst the above diagrams show similarities with Kyirong Tibetan, there are also a number of notable differences. Diagram b. shows an analysis of the prefix forming a recursive pword, rather than a pphrase, with the stem. This was one of the alternative analyses for the prefixes in Kyirong Tibetan. Diagrams c. and d. both show a suffix forming a pword with the stem. However, the difference between c. and d. is the tone on the suffix. The suffixes in d. are always extra-tonal (marked in brackets < >), and so are always outside of the tonal foot even though they form a single pword with the stem. On the other hand, the suffixes in c. do form a tonal foot with the stem, as do the prefixes in b. The suffixes in e. are also extra-tonal, but form a recursive pword with the stem.

The sections that follow discuss these relationships in more detail, giving particular consideration to the tone patterns across the morpheme boundary.

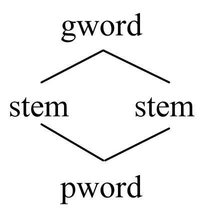
[5] An alternative for this is a clitic group.

5.4.2 Compound nouns

A compound word is described in Fabb (1998:66) as “a word which consists of two or more words ... the words in a compound retain a meaning similar to their meaning as isolated words”. In Walungge there are many words which fit this description of a compound. The following are examples of two monosyllabic nouns forming a disyllabic compound word.

ɖàŋ	*ts^{h}áŋ*	→	[ɖaŋdzaŋ $^{11\ 344}$]	*píŋ*	*k^{h}áp*	→	[piŋgap˺ $^{44\ 44}$]
‘honey’	‘nest’		‘honeycomb’	‘pin’	‘needle’		‘safety pin’
mík	*ɕé:*	→	[mixɕeː $^{44\ 44}$]	*k^{h}ùr*	*t^{h}ákpa*	→	[k^{h}utːaʔ $^{11\ 44}$]
‘eye’	‘glass’		‘glasses’	‘tent’	‘rope’		‘guy rope’
ʈá	*píŋ*	→	[ʈawiŋ $^{44\ 44}$]	*lók*	*mè*	→	[logme $^{44\ 44}$]
‘hair’	‘pin’		‘hairclip’	‘light’	‘fire’		‘battery’
ɖùk	*két*	→	[ɖukːeʔ $^{11\ 44}$]	*rà*	*lùk*	→	[raluʔ $^{11\ 44}$]
‘dragon’	‘voice’		‘thunder’	‘goat’	‘sheep’		‘flocks’

The following diagram illustrates the prosodic relationship between the two words of a compound word in Walungge:



What is interesting about this structure is that from a prosodic perspective Walungge treats both words of the compound as the one pword. This is the same as Kyirong Tibetan (KT), the structures of which have been outlined above. Hall & Hildebrandt (2008:216) note the following about KT compounds: “KT is also unusual typologically because it treats compound words as single pwords and not as two separate pwords, which is probably the default option cross-linguistically.” From the analysis of Refugee Standard Tibetan (RST) in Meredith (1990), it would appear that RST also treats compounds in this way. Lhasa Tibetan (Duanmu (1992)) appears to be similar.

In Walungge, part of the evidence for both parts of the compound word forming the one pword is the segmental changes, which include loss of aspiration, the labial weakening from /p/ to /w/, and the complete assimilation of /r/ to the following consonant.

Aspiration only occurs word initially (see section 3.1.2). The loss of aspiration in the above examples suggests that onset of the second word of the compound is not the start of a pword. Similarly /r/ only occurs in a syllable coda when it is word final. The total assimilation of /r/ to the following consonant suggests that there is no pword boundary at the end of the first word of the compound.

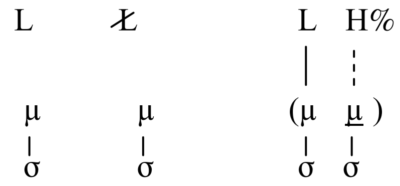
The labial plosive /p/ may occur word initially, finally, or next to another consonant, but never intervocalically. In the above data /p/ becomes /w/ intervocalically, e.g. [ʈawiŋ ^{44} ^{44}] 'hairclip'. Plosives and affricates become voiced next to a sonorant, e.g. [logme ^{44} ^{44}] 'battery'. Plosives become fricatives next to fricatives, e.g. [mixçeː ^{44} ^{44}] 'glasses'. None of these changes occur across the boundary of two separate pwords elsewhere in the language. Thus these changes are all indications that the two parts of the compound are not two separate pwords.

Further evidence that the two words of the compound are forming the one pword is the tone pattern across the compound. Regardless of the tone of either the first word or the second word, when two words come together to form a compound word, the resulting compound word has a pitch melody which is the same as any monomorphemic polysyllabic pitch melody: high or low pitch on the first syllable but always high pitch on the second syllable.

The derivation of tone patterns for compound words is simply an extension of what has already been proposed in section 5.3.6 above, which outlines the formation of a tonal foot. If two words form a tonal foot the tone from the second word is deleted. The underlying tone from the first word attaches to the first mora of the foot.

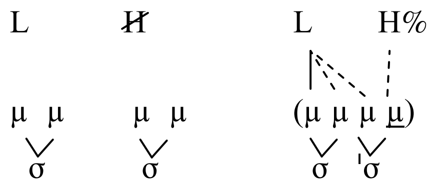
A boundary tone attaches to the final mora. The underlying tone spreads across any free moras.

The following are example derivations. The first example is when the underlying tone of both words is L.



rà *lùk* → [raluʔ 11 44] 'flocks'

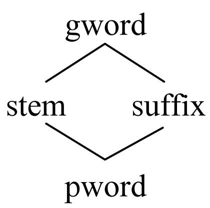
The next example shows a tone L word followed by a tone H word. Because the second word is bimoraic, the underlying tone of the first word spreads to the first mora of the second word. Second syllable stress causes the starting pitch of the syllable to be raised (see 5.3.6.4 above).



ɖàŋ *tsʰáŋ* → [ɖaŋdzaŋ 11 344] 'honeycomb'

5.4.3 Derivational morphemes

Derivational morphemes form a pword with the stem, and are suffixes. They will form a tonal foot with the stem unless they are the third or subsequent syllable of the word. This is the same as the compound words described above. The following diagram shows the relationship between the stem and suffix.



Segmental evidence that they form a pword with the stem is similar to that which is presented in the section above for compound words. The phonological process p → w / V_V holds pword internally throughout the language, and the nominalizer *-pa* is no exception to this, as illustrated by the following example:

ŋá	*-pa*	→	[ŋawa $^{44\ 44}$]
'five'	NMLZ		'fifth one'

A syllable coda /r/ only occurs word finally. The following example shows the deletion of a stem final /r/ when the nominalizer *-sa* is added:

k^{h}úr	*-sa*	→	[k^{h}usa $^{44\ 44}$]
'carry'	NMLZ		'place for carrying'

Prosodically, the resulting word follows the same pattern of foot formation and tone attachment as described in section 5.3.6 above. The following is an example of a monosyllabic, a disyllabic, and a trisyllabic stem, each with the nominalizer *-pa*. Because derivational affixes only occur affixed to a stem there is no evidence as to any underlying tone that they may have. They are therefore taken to be toneless.

```
L                  L  H%
                   |  ¦
μ    μ            (μ  μ̲ )
|    |             |  |
σ    σ             σ  σ
```

ʈ^{h}ùk -pa → [ʈ^{h}ukpa $^{11\ 44}$] 'the sixth one'

```
L                  L  H%
                   |  ¦
μ  μ   μ          (μ  μ̲ )‹μ›
|  |   |           |  |   |
σ  σ   σ           σ  σ   σ
```

nàppo -pa → [napːowa $^{11\ 44\ 21}$] 'the black one'

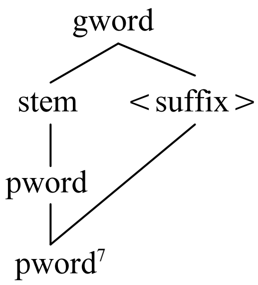
```
L                  L  H%
                   |  ¦
μ  μ  μ   μ       (μ  μ̲ )‹μ›‹μ›
|  |  |   |        |  |   |  |
σ  σ  σ   σ        σ  σ   σ  σ
```

bàrakpa -pa → [barakpawa $^{11\ 44\ 21\ 11}$] 'the big one'

5.4.4 Inflectional morphemes on nouns

5.4.4.1 Introduction

Inflectional morphemes associated with nouns include case markers and the plural marker. In Walungge they are all suffixes, and are all extra-tonal, (see below). These morphemes are interesting because of the relationship with the stem, which can be illustrated by the following diagram[6]:



Evidence for a pword boundary between the stem and the suffix is the segments /r/ and /n/. In monomorphemic words /r/ only occurs in syllable coda position when it is word final. Section 5.4.2 above on compound words showed a word final /r/ completely assimilating to a following consonant in order not to violate this restriction. However in the case of the morphemes here, a stem final /r/ remains, e.g.

k^h*ùr*	*-la*	→	[k^{h}urla $^{14\ 21}$]	'tent.LOC'
sér	*-i*	→	[sergi $^{44\ 21}$]	'gold.POSS'

In monomorphemic words, when the segment /n/ occurs in a word final position, it is realised as nasalisation and duration on the preceding vowel. However, when it is in a word internal syllable coda it is realised as [n], e.g.

mén	[mẽː ^{44}]	'medicine'
gànde	[gande $^{11\ 44}$]	'good'

With inflectional morphemes, /n/ is realised as nasalisation and duration:

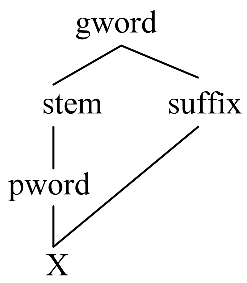
---

[6] In Walungge all inflectional morphemes on nouns appear to be accounted for by this diagram, whereas for verbs there are different types of inflectional morphemes having different relationships with the stem (see 5.5 below).

[7] An alternative structure would be a clitic group.

kígen -la → [kigẽːla $^{44\ 44\ 21}$] ‘dog.LOC’

A pword boundary between stem and suffix can be represented by the following structure:



where X is still to be determined. Under the strict layer hypothesis (Selkirk (1984)) X would be a pphrase, as that is what is generally considered to immediately dominate a pword[8]. However there is an epenthetic velar consonant, described below, which may occur between stem and suffix, but does not occur between two pwords in a pphrase. The epenthetic velar consonant occurs when the stem ends with a consonant and an affix starts with a vowel or an approximant. The following data shows the possessive morpheme /-i/ affixed to stems ending with a vowel:

tá -i → [tai ^{51}] ‘horse.POSS’

ʈʰò -i → [ʈʰoi ^{242}] ‘wheat.POSS’

mè -i → [mei ^{242}] ‘fire.POSS’

When the possessive morpheme follows a consonant there is an epenthetic velar plosive, e.g.

kʰáp -i → [kʰapki $^{53\ 21}$] ‘needle.POSS’

kʰùr -i → [kʰurgi $^{14\ 21}$] ‘tent.POSS’

tʰòm -i → [tʰomgi $^{14\ 21}$] ‘bear.POSS’

The velar plosive is voiced if it is next to a sonorant consonant. The plural morpheme /-wa/ shows a similar pattern of epenthesis. When the plural affix follows a stem ending with a vowel there is no epenthetic velar consonant. However when it

---

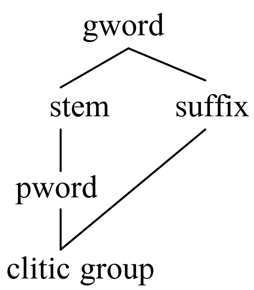
[8] Certain authors (e.g. Nespor & Vogel (1986)) allow for a clitic group between a pword and a pphrase.

follows a stem final consonant a velar plosive is optionally inserted. The voicing of the velar plosive and any stem final plosive follows the patterns already described:

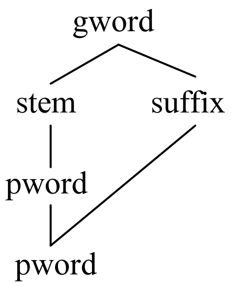
tá -wa	→	[tawa $^{53\ 21}$]	'horse.PL'
bù -wa	→	[buwa $^{24\ 21}$]	'insect.PL'
k^{h}áp -wa	→	[k^{h}apkwa $^{53\ 21}$] ~ [k^{h}abwa $^{53\ 21}$]	'needle.PL'
t^{h}òm -wa	→	[t^{h}omgwa $^{14\ 21}$] ~ [t^{h}omwa $^{14\ 21}$]	'bear.PL'

Although this epenthesis occurs between stem and affix if the stem ends with a consonant and the affix starts with a vowel or approximant, it never occurs between pwords within the one phrase. For example, in the noun phrase of a noun followed by an adjective *síttop áttɕuŋma* 'small ring' the word 'ring' ends with /p/ and the word 'small' starts with /a/, yet no epenthesis occurs. This suggests that in the case of stem plus affix, they are combining at a level lower than a pphrase.

As mentioned in 5.4.1 above, a clitic group is a debated structure between a pword and a pphrase. Under a model which allows for a clitic group, the relationship between stem and suffix can be represented as:



Alternatively, as discussed in 5.4.1 above, many models allow for recursivity. If this is the case, then the relationship between stem an suffix can be seen as a recursive pword. The stem forms its own pword, the boundary of which accounts for /r/ and /n/ described above. The suffix then attaches to the stem forming a further pword, which results in the epenthesis. This recursive structure can be represented as:



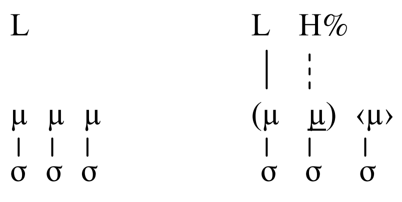
Either of the above structures would account for the segmental evidence presented above. The section immediately below discusses the tone across the stem and suffix, showing the extra-tonal nature of the suffixes.

5.4.4.2 The extra-tonality of suffixes

When an inflectional suffix attaches to a disyllabic noun to create a trisyllabic gword, the pitch melody across the stem plus suffix is the same as all the trisyllabic words described in the sections above: the pitch on the first syllable is high or low depending on the underlying tone, the pitch on the second syllable is high, and the pitch on the third syllable is low, e.g:

tɕʰàmo -wa	→	[tɕʰamowa $^{11\ 44\ 21}$]	‘chicken.PL’
lópʈa -la	→	[lopʈala $^{44\ 44\ 21}$]	‘school.LOC’

Simply considering these trisyllabic pitch melodies, a structure is implied which is no different from what has already been described for trisyllabic words, with tonal foot formation across the first two syllables, and an extra-tonal third syllable mora, which in this case is the suffix.

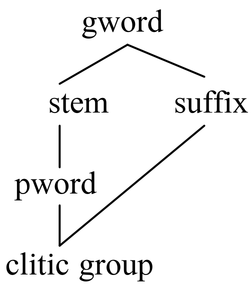


tɕʰàmo -wa → [tɕʰamowa $^{11\ 44\ 21}$] ‘chicken.PL’

However, if a suffix attaches to a monosyllabic stem to form a disyllabic gword, the realised pitch melody is different from the typical disyllabic pitch melody. Instead of the pattern of high or low level pitch on the first syllable and high pitch on the second syllable, the monosyllabic stem continues to have the pitch melody of a typical monosyllabic word and the suffix has low pitch:

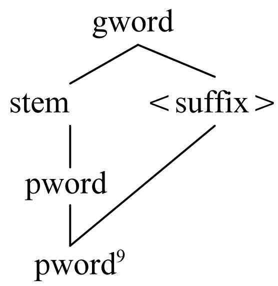
p^hà -wa	→	[p^hawa $^{24\ 21}$]	‘cow.PL’
nàŋ -la	→	[naŋla $^{14\ 21}$]	‘house.LOC’
nám -i	→	[namgi $^{44\ 21}$]	‘sky.POSS’
tá -wa	→	[tawa $^{53\ 21}$]	‘horse.PL’

The pitch melody across the stem suggests that the suffix does not form a tonal foot with the stem. This is different from the derivational suffixes described above which will form a tonal foot with the stem unless they are the third or subsequent syllable of the resulting word. However, even though the inflectional suffix does not form a tonal foot with the stem, neither does it form its own tonal foot, as evidenced by its low pitch rather than a rising or falling pitch melody. The pitch melody across stem and suffix might be taken as evidence for a clitic group construction, with the lack of tone on the suffix the result of it forming a clitic group rather than a pword with the stem:

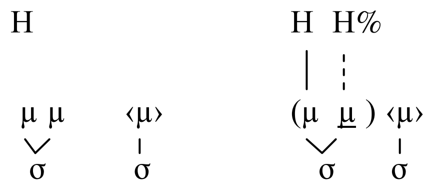


However, whilst this accounts for the pitch melodies across noun plus inflectional suffix, it does not account for pitch melodies across verb plus suffix (see section 5.5.2). In the case of verbs plus certain suffixes the segmental evidence suggests that the stem plus suffix form the one pword with no recursivity, yet the suffixes have the same lack of tone as the suffixes described above. If the lack of tone on the suffixes shown above was the result of them not forming a pword with the stem, that wouldn’t account for the lack of tone on the verb suffixes which do form a pword with the stem.

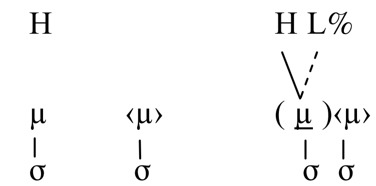
On the other hand, the lack of tone on the suffixes can be accounted for by describing these suffixes as extra-tonal in their structure, with the diagram below showing their extra-tonality by brackets < >:



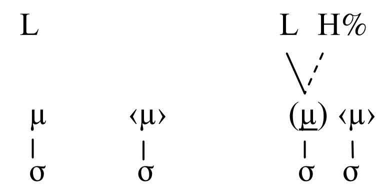
The following are illustrations of the tone attachment for an inflectional suffix and a monosyllabic stem. In each case a tonal foot is formed across the stem, with underlying tone and boundary tone attachment as already described. Because the suffix is extra-tonal it is not incorporated into the tonal foot, and surfaces with low pitch.



ná: -*wa* → *ná:wa* [na:wa $^{44\ 21}$] ‘blue-sheep.PL’



tá -*wa* → *táwa* [tawa $^{53\ 21}$] ‘horse.PL’



lùk -*la* → *lùgla* [lugla $^{24\ 21}$] ‘sheep.LOC’

[9] Could be a clitic group; but it would still need brackets < > to show the extra-tonality of the suffix

For stems with three or more syllables, the pattern is simply an extension of what has already been described. The stem forms a foot from the first two syllables and all remaining syllables of both the stem and the affix are toneless.

```
H                          H  H%
                           |  ¦
μ  μ  μ  μ  ‹μ›            (μ  μ̲) ‹μ›‹μ›‹μ›
|  |  |  |   |              |  |   |  |  |
σ  σ  σ  σ   σ              σ  σ   σ  σ  σ
```

tɕ^{h}émaladʑok -wa → [tɕ^{h}emalaʑogwa $^{44\ 44\ 21\ 11\ 11}$] ‘butterfly.PL’

Similarly, if more than one inflectional morpheme is added, there is no difference to what has already been described above. The formation of a tonal foot happens on the noun stem as normal, and all inflectional morphemes are realised with low pitch. The following is an example:

```
L                  L H%
                   |,'
μ ‹μ›‹μ›           (μ̲) ‹μ› ‹μ›
| |  |              |   |   |
σ σ  σ              σ   σ   σ
```

p^{h}à -wa -la → *p^{h}àwala* [p^{h}awala $^{24\ 21\ 11}$] ‘cow.PL.LOC’

5.4.4.3 Polysyllabic words with the possessive

This section considers the attachment of the possessive /-i/ to polysyllabic words ending with vowels. The first case to be considered is when the vowel of the stem is a short vowel /i/. The following are examples:

l^{h}ári -i →	[l^{h}ariː $^{44\ 41}$]	‘artist.POSS’
k^{h}àŋri -i →	[k^{h}aŋriː $^{11\ 41}$]	‘mountain.POSS’

Of particular note is the falling pitch melody across the second (phonetic) syllable of the resulting gword. A monomorphemic disyllabic word with long second vowel has a high level pitch melody across the second syllable, e.g. [k^{h}adaː $^{44\ 44}$] ‘silk scarf’. The falling pitch melody in the case of the possessive is due to the toneless nature of the possessive morpheme. Whereas the long vowel in [k^{h}adaː $^{44\ 44}$] ‘silk scarf’ is underlyingly one syllable, the long vowel in [gokːiː $^{11\ 41}$] ‘head.POSS’ is underlyingly two syllables. The final syllable is extra-tonal:

L   L H%

μ μ ‹μ›   (μ μ)‹μ›

σ σ σ   σ σ σ

gòkki -i   →   [gokːiː [11 41]] ‘head.POSS’

When the possessive /-i/ is attached to a polysyllabic noun ending with the short vowel /a/, the vowel /a/ deletes and there is compensatory lengthening. As in the case above, the pitch melody across the resulting [iː] is falling. The following are examples of this:

áppa -i	→	[apːiː [44 41]]	‘father.POSS’
gòkpa -i	→	[gokpiː [11 41]]	‘garlic.POSS’

The segmental process at the morpheme boundary can be illustrated as:

σ σ

μ μ

/ a i / [iː]

The following illustrates the formation of the tonal foot in the normal way. Because the long vowel in the phonetic form is structurally two syllables with the second syllable extra-tonal, it has falling pitch:

L   L   H%

μ μ ‹μ›   (μ   μ)‹μ›

σ σ σ   σ   σ σ

gòkpa -i   →   [gokpiː [11 41]] ‘garlic.POSS’

When the possessive /-i/ follows all other short vowels there is no deletion and compensatory lengthening as with /a/. The vowel plus affix is realised phonetically as vowel plus glide which is structurally two syllables:

H   H   H%

μ μ ‹μ›   (μ   μ)‹μ›

σ σ σ   σ   σ σ

tóptçe -i → [toptçej $^{44\ 41}$] ‘food.POSS’

The following are examples of the possessive affixing to a stem which ends with a long vowel, including the vowel /aː/:

t̪ʰomaː -i	→	[t̪ʰomaj $^{11\ 41}$]	‘potato.POSS’
míkçeː -i	→	[mikçej $^{44\ 41}$]	‘glasses.POSS’

The duration of the resulting vowel plus glide in a word such as [mikçej $^{44\ 41}$] is no different from the duration of the vowel plus glide in a word such as [toptçej $^{44\ 41}$], and neither is it any different from the duration of a long vowel [eː]. This suggests that even though the underlying form *míkçeː -i* has a bimoraic second syllable plus a further mora for the possessive, the surface [ej] in [mikçej $^{44\ 41}$] only has two moras. This segmental process can be illustrated:

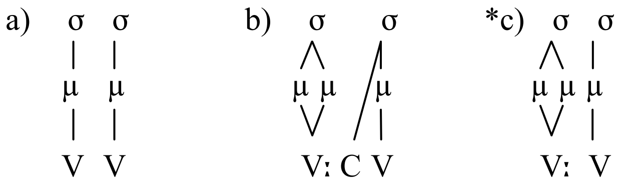
σ σ

μ μ μ

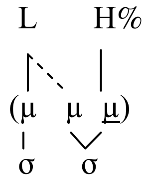
/ eː i / [ej]

A possible explanation for the deletion of a mora is that whilst Walungge permits two adjacent syllable nuclei if both are monomoraic, it does not permit a bimoraic syllable nucleus to be immediately adjacent to another nucleus without an intervening consonant; i.e. Walungge will permit the structures a) and b) but not c)[10]:

[10]An alternative explanation is that the vowel of the stem and the vowel of the affix become a heavy diphthong, which causes the stem vowel to shorten because of a restriction that heavy diphthongs are bimoraic.



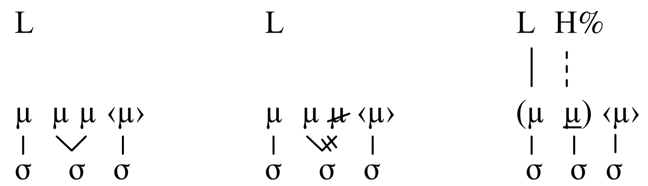
Further, a stem such as [ʈʰoma: $^{11\ 344}$] ‘potato’ has rising pitch at the start of the second syllable caused by the spreading of tone L to the first mora of the second syllable (see section 5.3.6.4 above):



[ʈʰoma: $^{11\ 344}$]

whereas the possessive [ʈʰomai $^{11\ 41}$] does not have this rise, which is further evidence that the surface vowel [a] is monomoraic.

The following is the suggested derivation of the surface form:



ʈʰòma: -i → *ʈʰòmai* → [ʈʰomaj $^{11\ 41}$] ‘potato.POSS’

5.4.4.4 Monosyllabic nouns with open syllables

Examples of open syllable monosyllabic stems combined with the possessive suffix are:

ɖù: -i	→	[ɖuj ^{242}]	‘snake.POSS’
ná: -i	→	[naj ^{51}]	‘blue-sheep.POSS’
tí: -i	→	[ti: ^{51}]	‘foal.POSS’
tá -i	→	[taj ^{51}]	‘horse.POSS’

p^{h}à -i → [p^{h}aj ^{242}] 'cow.POSS'

The derivation in these instances is similar to what has already been described above for polysyllabic nouns ending with vowels. In each case the resulting pword is a disyllabic word with two adjacent monomoraic syllable nuclei. The following are examples of a stem ending with a short vowel. With a low tone word the pitch melody across the resulting vowel plus glide is rising then falling; with a high tone word the pitch melody is falling:

L   L H%
μ   ‹μ›   (μ̲) ‹μ›
σ   σ   σ σ
p^{h}à   *-i*   →   [p^{h}aj ^{242}]   'cow.POSS'

H   H L%
μ   ‹μ›   (μ̲) ‹μ›
σ   σ   σ σ
tá   *-i*   →   [taj ^{51}]   'horse.POSS'

In the case of a stem ending with a long vowel, the resulting vowel plus glide when the possessive is added is no different in duration from the vowel plus glide which results from a stem with short vowel, and neither is the pitch melody any different; e.g. [naj ^{51}] and [taj ^{51}] have the same pitch melody and the same duration duration despite having a different number of moras in their roots (*ná:* versus *tá*). This is the same as the polysyllabic case described above, and can be illustrated:

H   H   H L%
μ μ   ‹μ›   μ ~~μ~~‹μ›   (μ̲) ‹μ›
σ   σ   σ σ   σ σ
ná: *-i*   →   *nái*   →   [naj ^{51}]   'blue-sheep.POSS'

When a bimoraic stem has the vowel /iː/ as its syllable rhyme and the possessive morpheme is affixed, the result is a pair of words that are phonetically the same at the segmental level but are contrastive at the prosodic level. e.g.

```
H            H   H%
             |   ¦
μ  μ        (μ   μ̲ )
 \/           \/
 σ            σ
tí:      →    [tiː 44]  'foal'
```

versus

```
H                 H              H L%
                                 |/
 μ μ     ‹μ›      μ μ‹μ›        (μ̲) ‹μ›
  \/      |        \/  |         |    |
  σ       σ        σ   σ         σ    σ
tí:      -i   →   tí-i      →   [tiː 51]   'foal.POSS'
```

5.5 Tone patterns on verbs

When describing the tone patterns on verbs, it is necessary to note that verbs are phrase final and all end with a fall. The following data compares the pitch melodies of nouns with the pitch melodies of the imperative form of verbs, the imperative form being a root without any affixes:

túŋ	[tuŋ 51]	'drink.IMP'	*sáŋ*	[saŋ 44]	'pine'
lúk	[luʔ 53]	'pour.IMP'	*ják*	[jaʔ 53]	'yak'
ŋù:	[ŋuː 241]	'cry.IMP'	*sù:*	[suː 14]	'body'
sò	[so 231]	'eat.IMP'	*ʈʰò*	[ʈʰo 24]	'wheat'

The pitch melodies on the verbs can be accounted for by saying that they receive a boundary tone L* in addition to their underlying tone L or H, and in addition to L% or H% which attaches to the head of the foot in the same way as all the other examples above. The following example illustrates this:

L L H% L*

μ μ (μ μ)

σ σ

dùŋ → [duŋ ^{241}] 'hit.IMP'

The level of the prosodic hierarchy with which L* associates is a subject for further research. The purpose of the sections which follow is to describe the prosodic relationships between verb roots and affixes, particularly noting whether or not the affix forms a tonal foot with the verb root. Whilst it is important to note the presence of the phrase final fall, the analysis of the prosodic relationship between root and affix is not affected by the analysis of this fall. Thus L* will be included in the descriptions of the following sections, whilst acknowledging that nothing further is known about L* at this stage.

5.5.1 Different types of suffix

Walungge is similar to other Tibetan languages which have been shown to have different types of suffixes for verbs.

Meredith (1990) describes Refugee Standard Tibetan (RST) as having 'strong' and 'weak' verbal suffixes. Weak suffixes are described as those outside the metrical structure, and do not take tone. Such suffixes include the infinitive and the past tense. Strong suffixes are described as being part of the metrical structure, and take tone. Strong suffixes include the imperfective.

Kyirong Tibetan shows a similar pattern. Hall & Hildebrandt (2008) analyse Kyirong verbal suffixes as being cohering (forming a pword with the stem) or non-cohering. The suffixes which they analyse as being non-cohering but not being a separate pword correspond to the suffixes in RST which Meredith analyses as being weak suffixes. The cohering suffixes of Kyirong Tibetan correspond to the strong suffixes of RST.

Walungge is very similar to these two languages. Some verbal suffixes correspond to the weak suffixes of RST or the non-cohering suffixes of Kyirong. These Walungge suffixes are extra-tonal. Other suffixes correspond to the strong/cohering suffixes, and form a tonal foot with the stem.

5.5.2 Suffixes which are extra-tonal

An example of a Walungge verbal suffix which is extra-tonal is the past tense suffix /-pa/. The relationship which such morphemes have with the stem is different from the extra-tonal suffixes of nouns which have been described above showing a pword boundary between the stem and the suffix. Whilst the verbal suffix is extra-tonal, the evidence suggests that there is no pword boundary between stem an suffix, as illustrated by the following diagram.

gword

stem <suffix>

pword

One difference between a suffix such as /-pa/ and inflectional suffixes of nouns is when it follows a coronal nasal /n/. In the case of nouns, the stem final nasal is realised as vowel nasalisation and duration (see 5.4.4.1) which was taken as evidence of a pword boundary between stem and suffix. However, in the case of /-pa/ there is assimilation of place of articulation of the plosive to the nasal and the nasal is realised as [n]. Given that the segment [n] does not occur word finally, this is evidence that there is no pword boundary between the verb root and suffix.

çø̀n -pa → [çønda $^{14\ 21}$] ‘ride.PST’

Further evidence of the lack of pword boundary between root and suffix is the phoneme /r/, which only occurs in a syllable coda position when it is word final. When inflectional suffixes are added to nouns, a word final /r/ remains in syllable coda position (see 5.4.4.1). However, if the past tense suffix is added to a verb stem with word final /r/, the initial consonant of the suffix is deleted allowing resyllabification of /r/ to the onset of the following syllable.

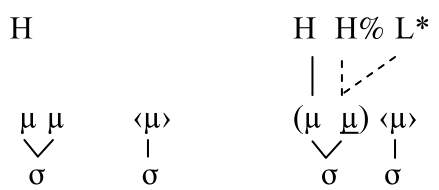
k^húr -pa → [k^hura $^{53\ 21}$] ‘carry.PST’

The affixation of other similar extra-tonal suffixes also shows the avoidance of syllable coda /r/ at the morpheme boundary. For example, before the suffix /-so/ ‘PST’[11] /r/ deletes:

k^húr -so → [k^huso $^{53\ 21}$] ‘carry.PST’

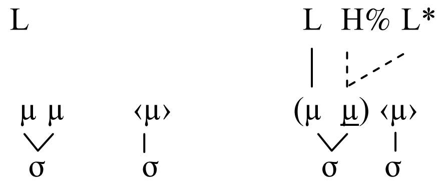
Whilst the segmental evidence is that a suffix such as /-pa/ forms a pword with the stem, the pitch melodies indicate that the suffix is extra-tonal. The suffix is always realised with low pitch, and the rises and falls on the root (as seen in the above data) indicate the boundary of the tonal foot.

The following two examples are of past tense verbs with heavy syllable stems. Both examples show the formation of a tonal foot with attachment of underlying tone and boundary H% as normal. In addition to this, a boundary tone L* also attaches. This gives a falling pitch melody on the stem of a verb with underlying tone H, which is of particular interest because heavy syllabled monosyllabic words with underlying H would normally have a high level pitch melody. This is evidence that the boundary tone L* is attaching to the tonal foot rather than the extra-tonal suffix. Because of peak delay, if the underlying tone is L the fall H%-L* does not occur until the end of the stem.



tsém -pa → [tsemba $^{53\ 21}$] ‘sew.PST’

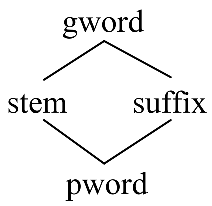
[11] both /-pa/ and /-so/ are used for verbs in the past. More research needs to be done in order to determine the difference between them.



çɵ̀n *-pa* → [çønda $^{14\ 21}$] ‘ride.PST’

5.5.3 Suffixes included in the tonal foot

Suffixes which form a tonal foot with the verb stem include the suffix /-ik/ ‘IPFV.DISJT’ and the suffix /-en/ ‘IPFV.CONJT’[12]. The relationship can be illustrated:



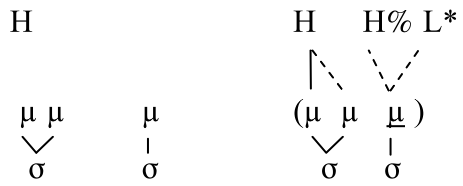
The segmental evidence that these suffixes form a pword with the stem is similar to the evidence that has already been given for other suffixes, including an epenthetic velar consonant (see 5.4.4.1), and avoidance of stem final /r/ (see 5.4.3). The following examples are of the suffix /-en/. There is an epenthetic velar consonant which inserts if the verb stem ends with a consonant. Because of the restriction that /r/ can only occur in a syllable coda if it is word final, /r/ deletes.

ŋù: -en	→	[ŋu:ẽ: $^{11\ 242}$]	‘cry.IPFV.CONJT’
p^{h}àp -en	→	[p^{h}apkẽ: $^{11\ 242}$]	‘dismount.IPFV.CONJT’
k^{h}úr -en	→	[k^{h}ugẽ: $^{44\ 41}$]	‘carry.IPFV.CONJT’

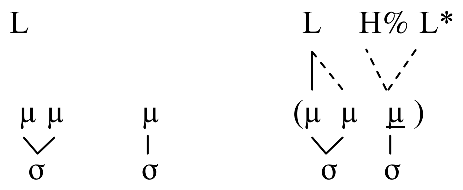
The evidence that the suffix is included in the tonal foot is the lack of rises or falls on the verb stem, and high pitch on the suffix. In each of the following examples, a tonal foot is formed from both the stem and the suffix. The underlying tone and boundary tone H% associate with the moras of the tonal foot in the normal manner.

[12] The terms ‘conjunct’ and ‘disjunct’ have been used in Tibeto-Burman languages to refer to a morpho-syntactic pattern where 1st person statements, 2nd person questions and co-referential relative clauses take one set of verbal suffixes (conjunct); otherwise a different set of verbal suffixes is used (disjunct).

Clause final position also means the attachment of boundary tone L* to the final mora of the tonal foot, which results in falling pitch on the final syllable.



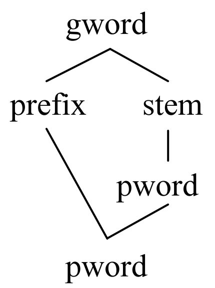
túŋ -*ik* → [tuŋgiʔ 44 42] ‘drink.IPFV.DISJT’



ɕòn -*ik* → [ɕøŋgiʔ 11 42] ‘ride.DISJT’

5.5.4 The negative morpheme

The negative morpheme /ma-/ is interesting because it is a prefix, not a suffix. Further, there is a pword boundary between the prefix and the stem yet the prefix is not extra-tonal so forms a tonal foot with the stem. This relationship between the stem and the prefix can be illustrated by the following diagram:



The evidence for the presence of a pword boundary between the prefix and the stem is stem initial plosives and affricates.

Firstly, in Walungge aspiration only occurs word initially. Section 5.4.2 above gives examples of the two parts of compound words forming a single pword, where any aspiration at the start of the second part of the compound is lost. The negative form of verbs, however, retains any stem-initial aspiration:

ma- kʰúr → [makʰur $^{44\ 41}$] 'carry.NEG'

ma- ʈʰì: → [maʈʰiː $^{11\ 342}$] 'write.NEG'

Secondly, unaspirated word initial plosives are voiced or voiceless according to whether the underlying tone is H or L whereas word internal plosives are voiced or voiceless according to the segmental environment (see 3.1.2 above). In the negative form of verbs, stem initial plosives remain voiced or voiceless according to the underlying tone:

ma- túŋ → [matuŋ $^{44\ 41}$] 'drink.NEG'

ma- dùŋ → [maduŋ $^{11\ 342}$] 'hit.NEG'

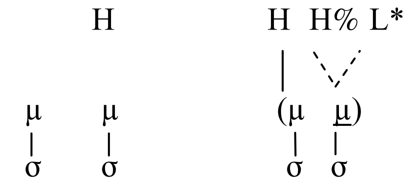
Thirdly, within a pword intervocalic affricates become fricatives whereas stem initial affricates in the negative verb forms do not become fricatives:

ma- tsém → [matsem $^{44\ 41}$] 'sew.NEG'

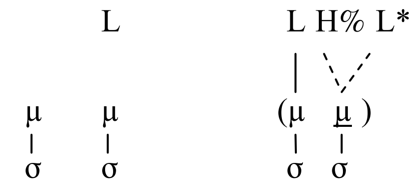
However, consider the underlying tone and the realised pitch melody in the above data, and in the data which follows. In each case the pitch on the negative morpheme is high or low depending on whether the underlying tone of the verb stem is high or low. Further, the pitch on the second syllable verb stem is always high (or rising to high) and falling.

[tuŋ ^{51}]	'drink'	[matuŋ $^{44\ 41}$]	'drink.NEG'
[luʔ ^{53}]	'pour'	[maluʔ $^{44\ 42}$]	'pour.NEG'
[duŋ ^{14}]	'hit'	[maduŋ $^{11\ 342}$]	'hit.NEG'
[loʔ ^{24}]	'return'	[maloʔ $^{11\ 42}$]	'return.NEG'

These pitch melodies can be accounted for by saying that the negative morpheme is underlyingly toneless and forms a tonal foot with the stem; the underlying tone of the verb stem attaches to the first mora of the foot which is the negative morpheme; the boundary tones H% and L* attaching to the final mora of the foot which is the stem. The following examples illustrate this.

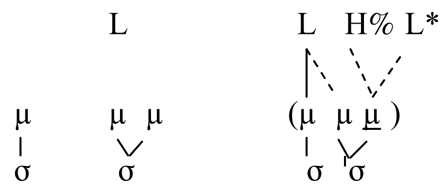


ma- *lúk* → [maluʔ [44 42]] ‘pour.NEG’



ma- *lòk* → [maloʔ [11 42]] ‘return.NEG’

Low tone bimoraic negative verbs have rising-falling pitch across the root. This rise is the same as the rise which has already been described for low tone nouns with heavy second syllables (see section 5.3.6.4). It is the result of a) the underlying tone L spreading to all non-final moras and b) second syllable stress which has the effect of raising the starting pitch of the second syllable.

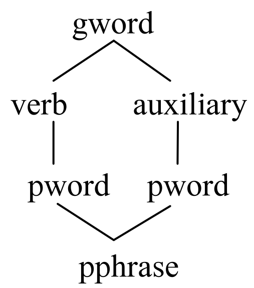


ma- *ʈʰìː* → [maʈʰiː [11 342]] ‘write.NEG’

Given a characteristic of all unambiguous pwords is the formation of a tonal foot from the first two syllables of the word, the formation of a tonal foot from the negative prefix and the verb stem suggests that they are together forming a pword. Thus the combination of the segmental evidence at the morpheme boundary and the tonal foot suggest an analysis of a recursive pword.

5.5.5 Verb forms with auxiliaries

A number of verb forms make use of a copula as an auxiliary verb. Unlike the suffixes described in the sections above which are underlyingly toneless, auxiliary verbs do have underlying tone, and form their own tonal foot independently from the tonal foot of the main verb. Relating this to the analysis of Kyirong in Hall & Hildebrandt (2008), the relationship between the auxiliary and the main verb can be illustrated as:



The following example shows a verb plus suffix forming one tonal foot, and the auxiliary as a separate tonal foot. The boundary tone L* attaches to the auxiliary tonal foot.

H		L		H H% L H% L*
μ	μ	μ		(μ μ <u>μ</u>)(<u>μ</u>)
σ	σ	σ		σ σ σ
túŋ	*-i*	*jø̀t*	→	[tuŋgijøʔ $^{44\ 44\ 121}$] ‘I am drinking’
drink	IPFV	AUX		

In a negative form of verb plus auxiliary, it is the auxiliary which is negated.

H			L		H H% L H% L*
μ	μ	μ	μ		(μ μ <u>μ</u>)(μ μ <u>μ</u>)
σ	σ	σ	σ		σ σ σ σ
túŋ	*-i*	*ma-*	*dè*	→	[tuŋgimende $^{44\ 44\ 11\ 41}$]
drink	IPFV	NEG	AUX		‘he is not drinking’

The formation of two separate tonal feet, and also the negative prefix on the auxiliary, is evidence that the verb and the auxiliary are separate pwords.

5.6 Conclusion to chapter

The focus of this chapter has been to provide an explanation for the surface pitch melodies which are found in Walungge, both in monomorphemic words and across morpheme boundaries. To provide such an explanation has involved an analysis of stress patterns, the proposition that Walungge forms a tonal foot to which is attached the underlying tone and a boundary tone, and a consideration of the prosodic hierarchy.

The acoustic analysis of stress patterns showed that in general Walungge has second syllable stress, and it is proposed that this general pattern of stress has been the origin of the polysyllabic tone patterns in Walungge. However, whilst diachronically stress might have caused the tone patterns, synchronically tone is independent of stress in as much as the second syllable of a polysyllabic word has high tone regardless of whether the word has the prevailing second syllable stress pattern or whether the word is one of the minority of words which have first syllable stress.

Following on from stress, and in order to account for the surface tone patterns, the main proposition of this chapter is that Walungge forms a tonal foot from the first two syllables of a word. This foot is the domain of the tone and any further syllables remain unfooted and are realised with low pitch. The underlying tone of the word attaches to the first mora of the tonal foot, with a boundary tone attaching to the final mora of the foot, which is the head of the foot. If the monomoraic, and thus the underlying tone and boundary tone attach to the same mora, the boundary tone is polar to the underlying tone. Otherwise the boundary tone is high. The underlying tone then spreads to any free moras in the tonal foot. It is proposed that this tonal foot is different from a typical metrical iambic foot in that the tonal foot is not sensitive to the syllabic structure; the formation of a tonal foot happens in this same way across the first two syllables of a word regardless of the weight of either of these syllables

In order to account for tone patterns and segmental realisations across morpheme boundaries it is necessary to consider the prosodic hierarchy, and in particular the formation of pwords. The tone pattern across the resulting polymorphemic word, and the segmental realisation at the morpheme boundary, is determined by whether or not the affix is extra-tonal, whether or not there is a pword boundary between stem and affix, and the tonal foot formation. Derivational affixes form a pword with the stem, and similarly the two parts of a compound word form the one pword. Tonal feet are formed across these words ignoring the morpheme boundary. Inflectional morphemes on nouns are extra tonal, and in addition to this they have been analysed as forming either a recursive pword or a clitic group. The extra-tonal analysis accounts for the low pitch with which they surface, and the recursive pword/clitic analysis accounts for the segmental realisation. Verbal affixes include suffixes which are extra-tonal, suffixes which form a tonal foot with the stem, and the negative prefix which can be analysed as forming a recursive pword with the stem and also forming a tonal foot with the stem. Verb forms with auxiliaries can be analysed as the stem and the auxiliary being separate pwords.

6 Consonants and tone

6.1 Introduction

The primary question that this chapter considers is whether Walungge has phonologically contrastive obstruent voicing. Phonetically Walungge has both voiced and voiceless plosives and affricates word initially, with voiced plosives/affricates occurring with low tone, and voiceless plosives/affricates occurring with high tone, for example:

pák	'pastry'
bàk	'a kind of bamboo'
tsá	'vein'
dzà	'crag'

However, in situations other than when there is a word initial unaspirated plosive or affricate, an underlying tone H contrasts with an underlying tone L independently of any of the segments:

k^{h}úr	'carry'	*k^{h}ùr*	'tent'
sér	'gold'	*sèr*	'screw'
náŋ	'day after tomorrow'	*nàŋ*	'house'
ŋá	'drum'	*ŋà*	'1SG'
órokpa	'toad'	*òraŋ*	'we'

Because there is contrastive tone independent of voicing it could be concluded that whilst diachronically the tone of *pák* versus *bàk* has arisen from contrastive obstruent voicing (see section 3.5 above and 7.1.1 below), synchronically it is the tone which is causing the voicing. Thus synchronic voicing in plosives would not be phonologically contrastive unless it could be shown to be contrastive in other situations in the language.

However, even in data such as the above where tone is independently contrastive but voicing always correlates with tone, it can be debatable whether it is the voicing determining the tone or the tone determining the voicing. Lhomi (see section 2.7.1 above) has tone and voicing very similar to Walungge. Vesalainen & Vesalainen (1976) analyse Lhomi plosives as being underlyingly voiceless with the tone determining the voicing; however Watters (2003) analyses Lhomi plosives as underlyingly having the features for voicing. Lhasa Tibetan is also very similar; voicing has been analysed as being non-contrastive (e.g. Hari (1979)) and as being contrastive (e.g. Denwood (1999). Kera is an example of an African language where there has been the same debate Pearce (2007).

In Walungge, the intervocalic voicing of plosives does not correlate with the tone; both voiced and voiceless plosives occur independently of the tone pattern (see 6.2.1 immediately below for examples). It is the word intervocalic consonants in Walungge which shed light on the issue of whether voicing is phonologically contrastive.

The chapter starts by considering intervocalic obstruents. Their acoustic properties are considered, with particular consideration given to duration and voicing. The question of whether intervocalic obstruents affect the F0 of the following syllable is also considered. The analysis of intervocalic obstruents draws on the acoustic work, and also considers the phonological behaviour of obstruents at morpheme boundaries. The conclusion is that word medial plosives and affricates do not contrast phonologically for voicing. Word initial obstruents are then considered, with a similar conclusion.

Having arrived at the conclusion that voicing of plosives and affricates is not a phonological contrast but is determined by the tone, the chapter considers the issue of laryngeal features for Walungge segments, and then considers the phonological nature of the consonant-tone interaction.

The data used for the acoustic analysis in this chapter is the same data as used in Chapter 4 for the acoustic analysis of pitch, phonation, and duration.

In looking at the question of voicing there is an issue of what level of representation to use for transcribing the data. Until the conclusion of section 6.2 of this chapter, no assumptions are being made as to whether or not voicing is phonological. Thus any phonetic voicing (or lack of) needs to be transcribed. However, much of the other phonetic detail (such as the unrelease [˺] of word final consonants) is unnecessary to the discussion, and will be omitted. This intermediate level of representation is written in italics. The phonetic form is written in brackets [], and the underlying phonological form is written in brackets //.

6.2 Word medial obstruents

6.2.1 Introduction to word medial obstruents

In addition to the word initial voicing of plosives and affricates which correlates with tone, there is also word medial voicing. Next to a sonorant consonant, plosives are always voiced, e.g.

gònda 'monastery'

lúŋbu 'wind'

Next to an obstruent, plosives are always voiceless, e.g.

dùptɕa 'bad'

tsákʈa 'matches'

Intervocalically, both voiced and voiceless plosives can occur with both high and low tone words, as illustrated by the following examples, specifically written in their phonetic form:

[kudaŋ $^{44\ 44}$] 'religious painting'

[tutːoŋ $^{44\ 44}$] 'jacket'

[radaʔ $^{11\ 44}$] 'root'

[ʈʰotːa $^{11\ 44}$] 'stomach'

[takːa $^{44\ 44}$] 'walnut'

[tʰaga $^{44\ 44}$] 'fireplace'

[napːo $^{11\ 44}$] 'black'

What is of particular interest in these words in their phonetic form is the intervocalic lengthened plosives.

If the lengthened plosives are analysed as sequences of two identical plosives (i.e. [tː] is analysed as /tt/, [pː] as /pp/, etc), then the above pattern of voicing can be analysed as being allophonic variation, with voiced plosives occurring between voiced segments, and voiceless plosives occurring next to a voiceless segment, e.g.

/kútaŋ/	[kudaŋ $^{44\ 44}$]	'religious painting'
/túttoŋ/	[tutːoŋ $^{44\ 44}$]	'jacket'

However, if the duration difference between voiced and voiceless plosives is treated as a phonetic effect caused by the presence or absence of voicing, then voicing of plosives can be analysed as being phonologically contrastive, e.g.

/kúdaŋ/	[kudaŋ $^{44\ 44}$]	'religious painting'
/útoŋ/	[tutːoŋ $^{44\ 44}$]	'jacket'

The sections that follow consider the issue of lengthened plosives from an acoustic perspective, a distribution perspective, and a morpheme boundary perspective. Whilst some of the evidence is not completely clear-cut, and alternative analyses have been presented, the body of evidence taken as a whole suggests that that the lengthened plosives are sequences of two identical plosives, and thus voicing is not contrastive.

6.2.2 Acoustic consideration of word medial obstruents

6.2.2.1 Intervocalic plosive durations

Plosive duration and voicing in Hindi

Before considering the duration of intervocalic obstruents in Walungge, attention is first turned to Hindi (Ohala (2007)), which has both voiced and voiceless intervocalic plosives both lengthened and non-lengthened (with the lengthened consonants traditionally analysed as sequences of two identical consonants with a syllable break in-between: VC.CV). Examples of Hindi words are:

[pət̪ɑ] ‘address’

[pət̪ːɑ] ‘leaf’

[gəd̪ɑ] ‘mace’

[gəd̪ːɑ] ‘mattress’

In Ohala’s study of Hindi intervocalic plosives, the following approximate mean durations were found:

Table 6-1: Durations of Hindi plosives

intervocalic segment	*mean duration (ms)*
voiced non-lengthened	~ 70
voiceless non-lengthened	~ 80
voiced lengthened	~ 130
voiceless lengthened	~ 150

It can be seen from the above figures that for both lengthened and non-lengthened plosives the voiced plosives are typically shorter in duration than the voiceless plosives. The above figures give a ratio voiceless:voiced of approximately 1.15:1. The correlation between the duration of a plosive closure and voicing can be accounted for by the pressure build up behind the closure of a voiced plosive. This pressure build up means that the closure cannot be held for as long as it can be if the plosive is voiceless. It can also be seen from the above figures that the lengthened consonants have nearly double the duration of the non-lengthened consonants (Ohala gives a ratio of 1.96:1).

Using the above study of Hindi as a basis for exploring plosive durations in Walungge, the ratio of durations between voiceless lengthened plosives and voiced non-lengthened plosives might be an indication of whether lengthened plosives are single consonants or sequences of two identical consonants.

Intervocalic obstruent duration in Walungge

As discussed above, one indication of the phonological nature of voiceless lengthened plosives might be their duration compared to voiced plosives. For this study the durations of three different types of plosives were measured: voiceless lengthened plosives (e.g. [tː] in [tutːoŋ $^{44\ 44}$] ‘jacket’), voiced intervocalic plosives (e.g. [d] in [kudaŋ $^{44\ 44}$] ‘religious painting’, and sequences of voiceless plosives (e.g. [kʈ] in [lukʈuʔ $^{11\ 44}$] ‘lamb’). Sequences of voiceless plosives were included because if lengthened plosives are phonologically sequences their duration might be expected to be similar to that of other sequences of plosives.

Durations were measured as follows. For intervocalic voiced plosives, two points which can easily be identified for measuring the duration are the point of closure and the point of release of the plosive. Thus the measurement taken was the duration of the closure. In the case of voiceless plosives, three points can easily be identified: the point of closure, the point of release, and the onset of voicing. However for consistency with voiced plosives, the duration of the closure was measured, and the duration between the release of the plosive and onset of voicing was ignored. When there is a sequence of two non-identical plosives (e.g. [kʈ]), the transition between them is close transition, that is to say, the release of the first plosive happens after the closure of the second plosive. They are similar to the voiceless lengthened consonants in that acoustically they have the one period of closure, which is the measurement that was taken.

The following is a table of means and standard deviation of the measurements:

Table 6-2: mean and standard deviation for duration of medial plosives

plosive type	*mean duration (ms)*	*std. dev.*	*N*
non-identical plosives	145.1	27.7	74
lengthened plosive	114.2	29.9	131
voiced plosive	59.1	15.4	39

From the above figures, the ratio of mean duration of intervocalic lengthened plosives to intervocalic voiced plosives is 1.93:1, which is very similar to the ratio that Ohala found between lengthened and non-lengthened consonants in Hindi. This is evidence suggesting that the lengthened plosives are sequences of two identical plosives. The duration ratio of non-identical plosive sequences to lengthened plosives is 1.27:1; the lengthened plosives are clearly considerably shorter. However, if lengthened plosives were single plosives, it might be expected that their duration would be approximately half the duration of non-identical sequences. An explanation for the longer duration of the non-identical plosives is the additional time it takes for the transition between the two places of articulation.

Whilst the duration of the lengthened plosives suggests that they are sequences, it is important to rule out other factors that could be influencing the measurements. These include the tone of the word, and the weight of the second syllable[13]. Chapter 5 above discusses the attachment of tone. In disyllabic words the underlying H or L tone attaches the first mora of the word and spreads to free moras, and in addition to this, a boundary tone H% is attached to the final mora of the word. This results in a [44 44] pitch melody across all disyllabic words with underlying tone H, a [11 44] pitch melody for words with underlying tone L and a light second syllable, and a [11 344] pitch melody for words with underlying tone L and a heavy second syllable. In the case of all low tone words there is a rise in F0 which occurs across the onset to the second syllable if the onset is voiced. This means that across word medial voiced consonants F0 depends upon the underlying tone of the word. Further, for both voiced and voiceless word medial consonants, if the underlying tone of the word is low the F0 at the start of the second syllable depends upon the weight of the second syllable. These differences in F0 caused by tone could be influencing the duration of plosives. As well as any difference in duration caused by the tone, there could also be compensatory lengthening of the medial consonant due to the weight of the second syllable.

These possible effects were investigated by means of an ANOVA, the results of which are as follows:

[13] The weight of the first syllable is not included because in the intervocalic situation the only data which may have both heavy and light first syllables is data with voiced intervocalic plosives, and there are too few words with heavy syllables for meaningful statistical analysis.

Table 6-3: ANOVA for effects on duration of intervocalic plosives

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
tone	170	1	2.081	0.151	
weight of second syllable	170	1	2.420	0.122	

From the above ANOVA, it can be seen that a difference in tone does not correspond with a difference in duration of the medial consonant. Neither does the weight of either the first syllable or the second syllable. These factors can thus be ruled out as factors which could be influencing duration measurements.

In order to add further weight to the duration evidence that lengthened intervocalic plosives are sequences of two identical plosives, the duration of word initial plosives was considered for comparison. The closure durations of word initial unaspirated plosives were measured for words in the frame *nò ... dè* "this is a ...".

type of plosive	*tone*	*mean duration (ms)*	*std. dev.*	*N*
voiceless	high	113.3	3.6	84
voiced	low	102.1	3.1	130

These measurements give a duration ratio voiceless:voiced of 1.11:1, which is very similar to the Hindi ratio of voiceless:voiced. In the Walungge word initial environment voiceless unaspirated plosives occur with high tone and voiced plosives with low tone. So it could be the tone instead of, or as well as, the voicing which is causing the slightly shortened duration of the voiced plosives. However, if it was the tone rather than the voicing causing the difference, this would be an even stronger argument that it is the number of phonological segments rather than the voicing which is causing the difference in intervocalic duration. Thus the duration figures seem to suggest that the lengthened Walungge plosives are sequences of two identical consonants.

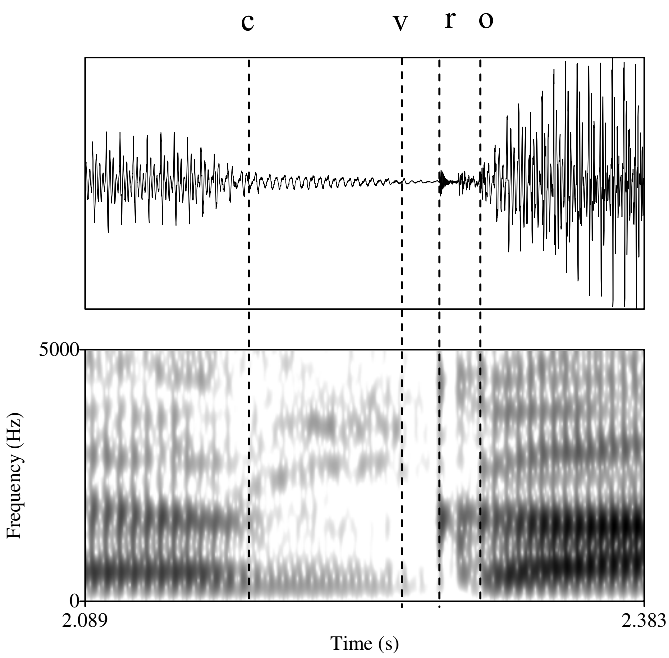
If the lengthened plosives are to be analysed as single voiceless plosives rather than two identical consonants, an explanation for these figures needs to be found. In Hindi both voiced and voiceless consonants can be either doubled or single. If the duration of a single consonant was increased, then it could be mistaken for a sequence of two consonants. In Walungge however, if the phonological distinction between [pː] and [b] is one of voicing rather than one of double/single, then the additional duration added to the voiceless plosive does not erode a contrast between single and double. Rather, it acts as an enhancement to the voicing distinction.

6.2.2.2 Proportion of voicing

Besides duration and voicing, one further acoustic difference between the lengthened voiceless plosives and the voiced plosives is the proportion of time that an obstruent is voiced. Following a sonorant consonant, plosives are voiced for 100% of their duration. For word initial plosives with low tone, once the voicing starts it continues for the rest of the duration of the plosive. However, intervocalic voiced plosives are not always voiced throughout their duration, and neither are lengthened voiceless plosives always completely voiceless throughout their duration.

The following waveform and spectrogram shows an intervocalic voiced plosive, with the voicing ceasing before the release of the closure, and resuming after the release of the closure.

Figure 6-1: intervocalic [g]



c = closure

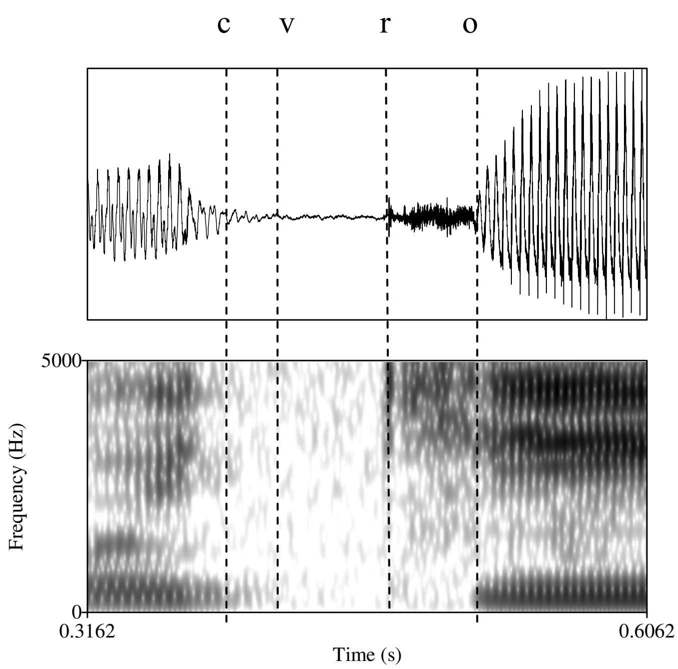
v = cessation of voicing

r = release of closure

o = onset of voicing

The following waveform and spectrogram show a voiceless plosive:

Figure 6-2: intervocalic [k:]



This pattern of voiced intervocalic plosives not being completely voiced, and voiceless plosives not being completely voiceless holds for both high tone and low tone words.

The following is a boxplot and table giving the mean percentage of the duration which is voiced.

Figure 6-3: Boxplot of percentage of closure voiced

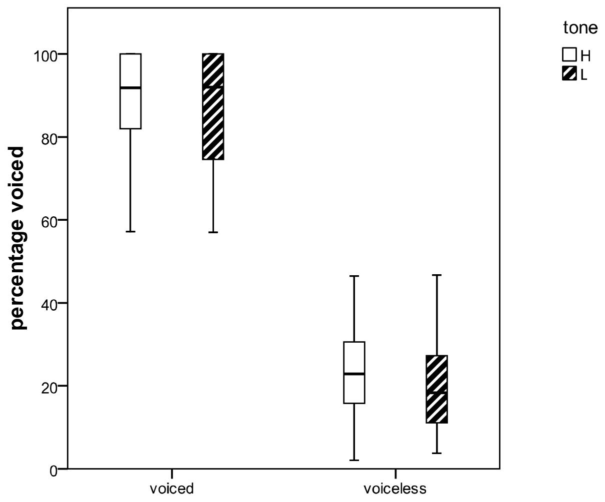


Table 6-4: mean and standard deviation for proportion voiced

voicing	*underlying tone*	*proportion voiced (%)*	*std. dev.*	*N*
voiced	H	88.0	13.3	17
	L	86.1	14.1	22
voiceless	H	24.3	11.6	83
	L	21.2	14.7	48

As can be seen from the above boxplot and table, voiced plosives have voicing in the closure for well over 50% of the duration, with a mean percentage of over 85%, whereas voiceless plosives have voicing in the closure for less than 50% of the time, with a mean percentage of less than 25%. Comparing the difference in measurements between tone H and tone L words, the tone of the word causes very little difference in the percentage of voicing for either voiced or voiceless plosives.

The proportion of voicing in the plosive can be explained if the medial plosives are all analysed as phonologically voiceless and the lengthened plosives are analysed as sequences of two identical plosives, with the period of voicing explained as spontaneous rather than non-spontaneous voicing. Halle & Stevens (1971) make a phonological distinction between plosives which are voiced because a deliberate configuration of the larynx is preventing them from becoming voiceless (non-spontaneous voicing), and plosives which may occur as voiced but will lose their voicing because of subglottal pressure (spontaneous voicing). In Walungge, the voicing during the first portion of the closure of the voiceless lengthened plosives is clearly spontaneous voicing. Similarly the voicing of the voiced plosives can be considered to be spontaneous voicing. The shortness in duration of these plosives combined with the intervocalic environment causes the vocal fold vibrations to continue across the most part of the plosive. However, the air pressure builds up behind the closure, which means that without a deliberate gesture of the vocal apparatus the voicing ceases before the plosive is released.

6.2.2.3 Effect of medial obstruents on F0

Whilst the voicing of word initial plosives correlates with the underlying tone of the word, medial plosive voicing and tone are independent. A word medial plosive can be either voiced or voiceless and the surface tone patterns can be either H-H% (underlying H) or L-H% (underlying L). In a disyllabic word the voicing of an intervocalic plosive does not affect the tone of the second syllable. Neither does the tone affect the voicing of the intervocalic plosive. However, even though the voicing of an intervocalic plosive doesn't affect the tone on the following vowel, given the word initial correlation between tone and voicing it is of particular interest to determine whether there is any difference in the starting F0 of the second syllable correlating with a difference in voicing of the preceding plosive. If the voicing of medial obstruents is shown to affect F0, this might be evidence that word initially it is the voicing of obstruents which is affecting F0, which in turn might suggest phonologically contrastive voicing.

For the question of whether medial obstruents affect F0, words with underlying tone H and L need to be considered separately. When there are voiceless consonants word medially, the high tone of the second syllable is able to be realised as high F0 right at the start of the second syllable rhyme, for both H-H% and L-H% patterns. However, for voiced medial consonants with L-H% tone pattern, the F0 increases from low to high during the voiced onset of the second syllable but the peak is not reached in time for the start of the second syllable. Thus for L tone words with voiced medial consonants, the starting F0 of the second syllable is lower than that of words with voiceless medial consonants, and it is not possible to determine to what degree the lowered F0 is due to peak delay and to what degree the lowered F0 is due to voicing.

However, because H tone words have F0 which is already high during the first syllable, it is possible to investigate whether the starting F0 of the second syllable is lower when the medial obstruent is voiced rather than voiceless.

An ANOVA was carried out to consider the above questions, with the following results ($p < 0.05$):

Table 6-5: ANOVA for effects on starting F0 of second syllable

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
effect of voicing of obstruents (tone H)	261	1	1.54	0.216	

Table 6-6: mean and s.d. for 2nd syllable F0; intervocalic plosives

voicing	*tone*	*pitch (semitones)*	*std. dev.*	*N*
voiced	H	1.92	2.24	93
voiceless	H	2.27	2.19	168

It is clear from the above results that the voicing of an intervocalic plosive does not significantly affect the starting F0 of the second syllable. This is a further indication that voicing is not contrastive.

6.2.3 Distribution of word medial consonants

6.2.3.1 Distribution of word medial plosives

The distribution of word medial plosives is evidence in favour of treating the lengthened plosives as sequences of identical plosives. Walungge has syllable rhymes V (short vowel), VP (short vowel with plosive coda), and VN (either a long vowel, or a short vowel plus sonorant coda). In particular, a rhyme with a long vowel never has a coda.

Voiced intervocalic plosives can occur following either a long vowel or a short vowel. However, voiceless lengthened plosives only occur following a short vowel:

[jiːguʔ $^{11\ 44}$] ‘envelope’

[t^{h}aga $^{44\ 44}$] ‘fireplace’

[takːa $^{44\ 44}$] ‘walnut’

If the voiceless lengthened plosives are underlyingly single phonemes, then they would be syllable onsets and thus it might be expected that they could occur following both long or short vowels. If, however, they are underlyingly sequences of two identical plosives with a syllable break in-between, this would explain why they never occur following a long vowel.

6.2.3.2 Distribution of other word medial consonants

One of the arguments in favour of treating the length of intervocalic voiceless plosives as additional duration caused by the difference in voicing is the lack of other intervocalic lengthened consonants in Walungge. In Walungge there are word medial sequences of nasals as well as word medial sequences of plosives, e.g.:

námnuk	‘snot’	*lópʈa*	‘school’
píŋmo	‘pigeon’	*t^{h}ákpa*	‘rope’

Given that the word medial distribution of nasals is much the same as plosives, then if sequences of two identical plosives are permissible intervocalically it follows that sequences of two identical nasals might be permissible as well. Consider the following data comparing Walungge with Lhomi (Vesalainen & Vesalainen (1976)) and with Written Tibetan.

Table 6-7: Lhomi and Walungge medial consonants

WT	Lhomi	Walungge	Gloss
<pad pa>	/píppɑ/ [pː]	[petːa $^{44\ 44}$]	‘leech’
<ser po>	/síppu/ [pː]	[sepːo $^{44\ 44}$]	‘yellow’
<mngar mo>	/ŋámmu/ [mː]	[ŋamo $^{44\ 44}$]	‘sweet’
<mtshan mo>	/ts^{h}émmu/ [mː]	[ts^{h}eno $^{44\ 44}$]	‘night’

Phonetically, both Lhomi and Walungge have lengthened plosives. These correspond in both languages to a plosive + plosive or a vibrant + plosive in WT. The lengthened nasals in Lhomi have a similar WT correspondence, whereas in Walungge there has been deletion of a word medial consonant leaving a single nasal. If the Walungge lengthened plosives are simply single consonants, then the above word medial consonants all show the same pattern of deletion leaving either a single plosive or a single nasal. In Lhomi, all lengthened consonants can be analysed synchronically as sequences of two identical consonants. Diachronically in Lhomi, there has been complete assimilation of the coda consonant to the onset consonant of the following syllable.

If the lengthened plosives in Walungge are to be analysed as sequence of two plosives, the following is a possible explanation for the above data. It could be argued that diachronically there was a similar assimilatory process in Walungge as in Lhomi. Following this, there was deletion of one of the identical nasals. Deletion of one of the identical plosives possibly did not happen because the voicing according to environment served to enhance the distinction between one or two plosives.

A further potential difficulty with treating Walungge lengthened plosives as sequences of two identical plosives is the intervocalic affricates. Intervocalically there are voiceless affricates and voiced fricatives (phonemically affricates) for example:

[matsi $^{44\ 44}$]	‘chilli’
[ŋotse $^{11\ 44}$]	‘shy’
[atɕe $^{44\ 44}$]	‘elder sister’
[gøːʑaː $^{11\ 344}$]	‘lock’
[ɟazo $^{11\ 44}$]	‘ocean’

One option would be to treat the segments [ts] and [tɕ] as sequences of /t/ followed by /s/ or /ɕ/. However, when other plosives are followed by /s/ or /ɕ/ they are realised as fricatives, e.g. *tókse* [toxse $^{44\ 44}$] ‘adze’. Further, at all derivational morpheme boundaries with a plosive followed by a fricative, a /t/ deletes, e.g. *dø̀t -sa* [døːsa $^{11\ 44}$] ‘sitting place’ whereas other plosives are realised as fricatives, e.g. *lúk -sa* [luxsa 44 44] ‘pouring place’. Section 5.4.3 above shows that a stem plus derivational morpheme forms the one pword, which suggests that the stem final /t/ is deleting at a derivational morpheme boundary because pword internally a sequence of /t/ plus fricative is not permitted. These are arguments that the segments [tɕ] and [ts] in the above data are not sequences of plosive plus fricative.

The simplest analysis of the affricates is to say that the voiceless affricates [ts] and [tɕ] are realisations of the phonemes /ts/ and /tɕ/, and the voiced fricatives are realisations of the phonemes /dz/ and /dʑ/. This would give an intervocalic contrast between voiced and voiceless affricates. Because of the phonological patterning of plosives and affricates in Walungge, if Walungge is analysed as having a voicing contrast for affricates, then it should also be analysed as having a voicing contrast for plosives.

If voicing is to be analysed as not being contrastive, the intervocalic segments [ts] and [tɕ] would need to be analysed as phonologically being sequences of a plosive followed by an affricate, i.e. /tts/ and /ttɕ/. This being the case, there needs to be an explanation as to why the phonetic realisation of these sequences is [ts] and [tɕ] rather than [tːs] and [tːɕ]. The explanation could be that because intervocalic affricates /ts/ and /tɕ/ are realised as fricatives [z] and [ʑ], there is no duration contrast between [ts]/[tɕ] and [tːs]/tːɕ] in the language. Without the need for closure duration as a phonetic cue to distinguish between /ts/ and /tts/ (or /tɕ/ and /ttɕ/), /tts/ and /ttɕ/ are realised as [ts] and [tɕ].

6.2.4 Morpheme boundaries

Part of the argument that voicing is not phonologically contrastive comes from the data at morpheme boundaries. In order to discuss the data at morpheme boundaries it is helpful present some of the phonological processes found in Walungge. However, as yet no discussion has been presented on the distinctive features of Walungge, whether they are privative or binary, or what the geometry of the features is. The discussion on laryngeal features is found in section 6.5 below. The following is a summary of the proposals of the later sections of this chapter that are relevant to this discussion:

- Aspirated plosives/affricates have the laryngeal feature [spread] but are unspecified for any other laryngeal features.

- Fricatives are unspecified for any laryngeal features, and always surface as voiceless. Similarly voiced sonorants are unspecified for any laryngeal features. Their voicing is the default phonetic state due to them being sonorants.

- If voicing of obstruents is not contrastive, then all unaspirated obstruents are unspecified for all laryngeal features.

- Plosives or affricates which have no laryngeal features (either because there are none underlyingly, or because they have been lost as a result of a phonological process) may have surface voicing which is due to the phonetics not the presence of any phonological feature (e.g. when they are next to a voiced consonant).

6.2.4.1 Plosives between voiced segments

When a plosive/affricate is next to a sonorant consonant, it is always voiced. This is always the case throughout the language, regardless of whether the plosive is morpheme internal or at a morpheme boundary. The following examples illustrate this at morpheme boundaries:

lók	*mè*	→	[logme $^{44\ 44}$]
'light'	'fire'		'battery'
ɖàŋ	*tsʰáŋ*	→	[ɖaŋdzaŋ $^{11\ 344}$]
'honey'	'nest'		'honeycomb'
píŋ	*kʰáp*	→	[piŋgap˺ $^{44\ 44}$]
'pin'	'needle'		'safety pin'

Next to an obstruent, an obstruent is always voiceless:

lùk	*ʈʰúk*	→	[lukʈuʔ $^{11\ 44}$]
‘sheep’	‘offspring’		‘lamb’

ŋá	*ɟà*	→	[ŋopca $^{44\ 44}$]
‘five’	‘hundred’		‘five hundred’

A single plosive or affricate at an intervocalic morpheme boundary is always voiced:

tɕʰà	*tʰàm*	→	[tɕʰadam $^{11\ 344}$]
‘tea’	‘fixed’		‘thermos’

mè	*tsʰák*	→	[mezaʔ $^{11\ 44}$]
‘fire’	‘fragment’		‘spark’

In the light of the above data, consider now the word internal data, e.g:

[tutːoŋ $^{44\ 44}$] ‘jacket’ [kudaŋ $^{44\ 44}$] ‘religious painting’

<u>Analysing lengthened plosives as sequences:</u>

If the lengthened plosives are analysed as sequences (e.g. [tː] is the realisation of /tt/), all the voicing in the above data is the result of the phonetics: if the plosive is intervocalic or between a sonorant and vowel, the plosive is realised as voiced. Otherwise it is realised as voiceless. The only laryngeal feature in Walungge, then, is the feature [spread] for aspiration (see 6.5. below for a discussion of laryngeal features). The loss of aspiration in the above data can be described as the delinking of the laryngeal node in all situations except word initially[14]:

X X

╪

LARYNGEAL

[14] See Lombardi (1995) for a discussion as to why laryngeal neutralisation should be described as the delinking of the laryngeal node rather than the delinking of the features attached to the laryngeal node.

The following is an example:

tɕʰà *tʰàm* → [tɕʰ a d a m 11 344] 'thermos'

C

LARYNGEAL

<u>Analysing lengthened plosives as single plosives:</u>

If lengthened plosives are analysed as single plosives with phonetic duration (e.g. [tː] is the realisation of /t/), then voicing is phonologically contrastive. One possibility for features for the voicing contrast is a single privative feature [voice], e.g.

/t ú t o ŋ/ 'jacket' /k ú d a ŋ/ 'religious painting'

C C

LARYNGEAL LARYNGEAL

[voice]

If the lack of a feature [voice] means that an intervocalic plosive is typically realised as voiceless, as in the above data, this creates a problem in accounting for all plosives being realised as voiced intervocalically at a morpheme boundary. The morpheme boundary voicing would have to be accounted for by the presence of [voice] on the intervocalic plosive. If [voice] is spreading to the plosive, e.g.

tɕʰà *tʰàm* → [tɕʰ a d a m 11 344] 'thermos'

V C

LAR LAR

[voice]

this means that vowels would have to have the feature [voice] even though there is not a voicing contrast for vowels. To avoid this problem, the other possibility is a binary feature [± voice], e.g.

/t ú t o ŋ/ ‘jacket’ /k ú d a ŋ/ ‘religious painting’

C | LARYNGEAL | [-voice]

C | LARYNGEAL | [+voice]

The presence of [+voice] means a plosive is always realised as voiced, the presence of [-voice] means a plosive is always realised as voiceless, and the absence of both [+voice] and [-voice] means that a plosive is realised as voiced or voiceless according to the segmental environment, e.g.

súm *tɕú* → [s u m dʑ u $^{44\ 44}$] ‘thirty’

‘three’ ‘ten’

C | LARYNGEAL

Under this proposal, all the above plosive data at morpheme boundaries can be accounted for by:

X + X | LARYNGEAL

X + X | LARYNGEAL

A morpheme boundary (+) needs to be specified to ensure that in the intervocalic case the delinking of the laryngeal node only happens at morpheme boundaries and not word internally. Whilst this accounts for all data at morpheme boundaries, it fails to acknowledge that, for example, plosives are voiced when next to sonorants and this is throughout the language and not just at morpheme boundaries. The alternative is to generalise the data by describing two separate processes similar to the above: one for the intervocalic voicing situation (in which case a morpheme boundary is specified) and one for all other situations (in which case a morpheme boundary is not specified).

This is an argument in favour of treating lengthened plosives as sequences of identical plosives and thus saying that voicing is not phonologically contrastive.

6.2.4.2 Sequences of identical plosives at morpheme boundaries

The following are examples which have lengthened plosives at morpheme boundaries (reduplication, compound word, and derivational morpheme) where the two morphemes together form a single phonological word (see 5.4 above). In each case the lengthened plosive has originated from two consonants.

ɖùk	*két*	→	[ɖukːeʔ $^{11\ 44}$]
'dragon'	'voice'		'thunder'
ɟèt	*-pa*	→	[ɟepːa $^{11\ 44}$]
'eight'	-NOM		'eight o'clock'
k^húr k^húr	*éːla*	→	[k^hukːur $^{44\ 44}$ eːla $^{44\ 21}$]
'carry'	'before'		'before carrying'
p^hòp p^hòp	*éːla*	→	[p^hopːop̚ $^{11\ 44}$ eːla $^{44\ 21}$]
'dismount'	'before'		'before dismounting'

The lengthened plosives are most easily analysed as sequences of identical consonants, with the first consonant, if it is a coronal consonant, having completely assimilated to the second consonant.

If the lengthened plosives are to be analysed as single consonants, there must be processes which delete a) a coronal consonant when it comes before another consonant and b) a consonant if it is next to an identical consonant. However, if the lengthened plosives are single consonants then they are in an intervocalic environment at a morpheme boundary. This is the very situation which has been described above, where all plosives are realised as voiced.

This situation is a further argument in favour of treating the lengthened consonants as sequences.

6.2.5 Summary

The following is a summary of the different types of data which have been considered in looking at the question of whether the lengthened voiceless plosives (e.g. [pː]) are underlyingly sequences of identical consonants (e.g. /pp/) or whether the length is phonetic duration and the plosives are underlyingly single voiceless plosives (e.g. /p/). The acoustic measurements for duration and proportion of voicing in the closure suggest that the lengthened plosives are sequences. Whilst the distribution of consonants suggests that the lengthened plosives are better analysed as single plosives, it is still possible to account for the data if the lengthened plosives are analysed as sequences. The segments which occur at morpheme boundaries also suggest that the lengthened plosives are sequences. The conclusion, therefore, is that the lengthened voiceless plosives are underlyingly sequences of two identical plosives, and thus voicing is not contrastive word medially.

6.3 Word initial obstruents

6.3.1 Introduction

The conclusion from looking at the word medial situation is that voicing of plosives and affricates is not phonologically contrastive. Given that voicing is not contrastive word medially, that the voicing of word medial obstruents does not affect the F0 immediately following them, and given that tone is contrastive but word initial voicing correlates with tone, it can be concluded that word initial obstruent voicing is not phonologically contrastive; all plosives/affricates (both word initial and word medial) can be analysed as being underlyingly voiceless. However, word medial and word initial plosives/affricates are realised as voiced through two different processes. Word medially the voicing is the result of the segmental environment. Word initially, the voicing is the result of the low tone. This raises a further issue for word initial obstruents: is the word initial voicing caused directly by the tone or is it caused by the F0? If the relationship is with the tone, then it might be expected that plosives are voiced with low tone and voiceless with high tone regardless of the actual values of F0. If, however, the relationship is with F0, then it might be expected that the degree of voicing (as measured by voice onset time (VOT)) varies as F0 varies. This next section investigates word initial obstruent voicing from an acoustic perspective in order to consider this issue.

6.3.2 VOT of word initial plosives

6.3.2.1 Introduction to VOT

VOT (voice onset time) is the duration between the release of the articulators and the onset of vocal fold vibrations. Measuring VOT is the primary method of distinguishing between different types of plosives, with voiceless aspirated plosives having a greater VOT than voiceless unaspirated plosives, and voiced plosives having the least VOT, which is a negative value if the vocal fold vibrations occur before the release of the articulators. VOT is a continuum, and the VOT values which distinguish between different types of plosives vary from language to language. For example, in English a word initial /p/ might have a VOT between 50 and 60ms, and a word initial /b/ might have a VOT of about 10ms. However, a French /p/ might have a VOT similar to that of an English /b/, and a French /b/ might have a VOT of -100ms and lower (all duration measurements from Ladefoged (2006)).

VOT of plosives in 5 Tibetan languages

Watters (2002) identifies four different types of plosive common to the five Tibetan languages of his study: Dzongkha (Bhutan), Lhomi (Nepal), Sherpa (Nepal), Dolpo Tibetan (Nepal), and Mugom Tibetan (Nepal). These are summarized as follows:

Table 6-8: Types of plosives in Tibetan languages

type	*description of plosive*
"prevoiced"	Begins with a brief period of voicing with little amplitude. Tapers off to almost complete silence, following which it is released as a voiceless sound.
voiceless	Silence, followed by the release of the plosive into the vowel.
"devoiced" [15]	Has a slightly longer VOT than voiceless unaspirated plosives, and is accompanied sometimes by a small degree of aspiration.
aspirated	Heavily aspirated.

There is a difference in VOT between voiceless, devoiced, and aspirated plosives, all of which have positive VOT. Although the voicing for the prevoiced plosive tapers off to silence during the closure of the plosive, it can still be seen as negative VOT.

6.3.2.2 Description of Walungge plosives

Phonetically, Walungge has four types of plosive word initially which are similar to the four types that Watters (2002) describes for other Tibetan languages. The following table summarizes the plosive types of Walungge:

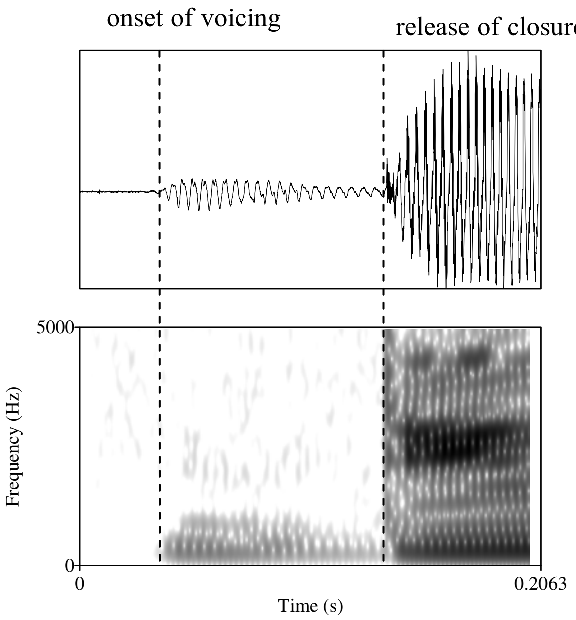
Table 6-9: Word initial plosives of Walungge

type	*tone*	*example*	*description of plosive*
voiceless	H	*kú* 'statue'	A short delay between the release of the plosive and the onset of voicing.
voiced	L	*gù* 'nine'	The amount of voicing varies. Typically the voicing occurs upon or before release of the plosive.
heavily aspirated	H	*k^{h}úr* 'carry'	There is a period of aspiration between the release of the plosive and the onset of voicing
lightly aspirated	L	*k^{h}ùr* 'tent'	The aspiration for a low tone word is both shorter in duration and less in intensity than with high tone.

[15] Referred to as "devoiced" in Watters (2002) because it derives historically from a voiced obstruent.

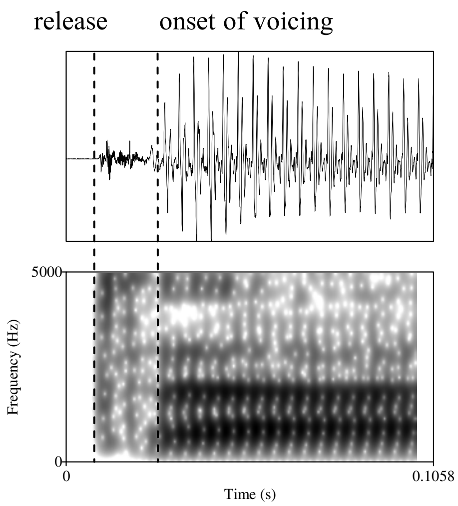
In general in Walungge, the onset of voicing for a voiced plosive occurs upon the release of the plosive, or before the release of the plosive. If the voicing occurs before the release of the plosive, it continues throughout the closure and into the release, which is different from the plosives that Watters describes. The waveform and spectrogram below show the release of the closure of the plosive, and the voicing which starts before the release.

Figure 6-4: waveform and spectrogram for *[d]*, with voicing before the release



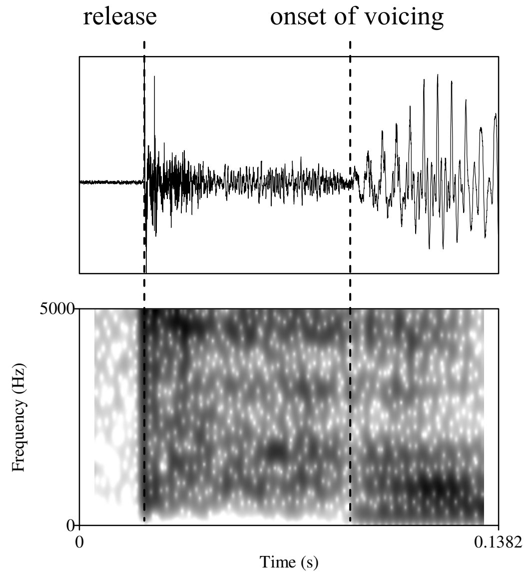
For a voiceless unaspirated plosive in Walungge, the voicing starts shortly after the release of the plosive. This corresponds to Watters' voiceless plosive. Below is a waveform and spectrogram to illustrate the voiceless plosive.

Figure 6-5: waveform and spectrogram for *[t]*



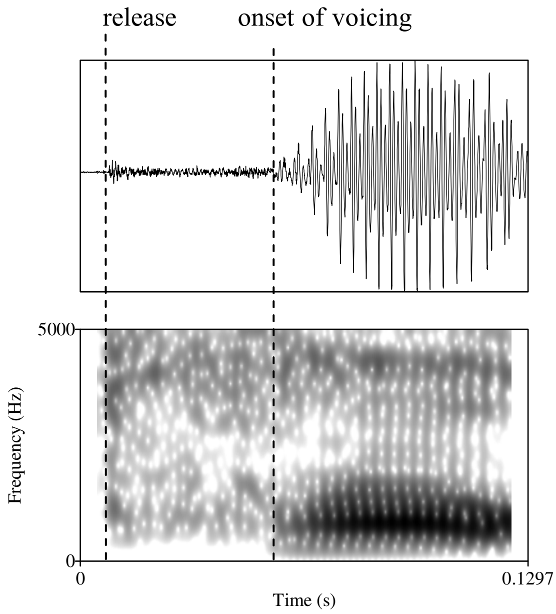
Heavily aspirated plosives in Walungge correspond to Watters' aspirated plosives, and acoustically appear to be very similar. The waveform and spectrogram below illustrates a Walungge aspirated plosive.

Figure 6-6: waveform and spectrogram for *[tʰ]* before high tone



The Walungge lightly aspirated plosives correspond to Watters' "devoiced" plosives, and are very similar acoustically to the plosives that Watters describes. The aspiration is generally less intense and shorter in duration than that of aspirated plosives before high tone. Below is the waveform and spectrogram of a Walungge lightly aspirated plosive.

Figure 6-7: waveform and spectrogram for *[tʰ]* before low tone



6.3.2.3 Statistical analysis of VOT of Walungge plosives

The four (phonetic) types of plosive described above for Walungge differ in terms of their average VOT. This is seen from the boxplot (giving median, interquartile range, and 1.5x interquartile range) and table of means below.

Figure 6-8: Boxplot of VOT in ms

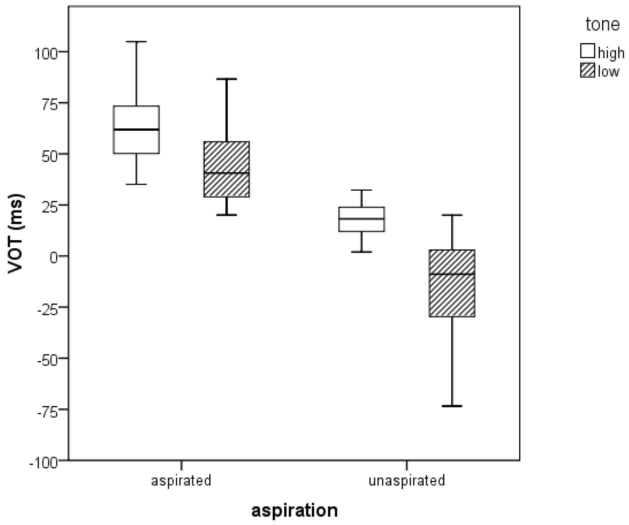


Table 6-10: mean and standard deviation of VOT for different plosive types

plosive type	*mean VOT (ms)*	*std. dev.*	*N*
heavily aspirated	66.0	18.9	120
lightly aspirated	48.5	18.7	130
voiceless	18.3	7.5	158
voiced	-13.2	32.7	202

As can be seen, there is considerable overlap in the VOT for different plosive types. Aspiration is a phonological distinction, and it can be seen from the boxplot that there is very little overlap between the VOTs for aspirated and unaspirated phonemes. For plosives before high tone, the cut-off in VOT is about 30ms, with aspirated plosives having a VOT of greater than 30ms, and unaspirated plosives having a VOT of less than 30ms. For plosives before low tone, the cut-off between aspirated and unaspirated is about 20ms. However, between the two types of aspirated plosive and between the two types of unaspirated plosive there is considerable overlap. For the unaspirated plosives, whilst low tone generally correlates with voicing there is not always voicing present, and this can be seen in the boxplot. For the aspirated plosives, whilst the aspiration is generally lighter before low tone than before high tone, and has a reduced mean VOT, there is overlap in the amount of aspiration there is before high or low tone.

An ANOVA was carried out to investigate the differences in VOT between the different types of plosives, and the effects upon the VOT. There is clearly a correlation between VOT and tone, but what is not clear is whether the correlation is with tone itself or with relative F0. The acoustic work in chapter 4 above showed that in monosyllabic words there are four significantly different starting pitches: two for high tone words and two for low tone words. Thus the starting value of F0 is not simply determined by the tone. If the VOT varies as the starting value of F0 varies, this is evidence in support of plosive voicing being a phonetic effect caused by F0 rather than a phonological one caused by the tone.

The independent variables included in the ANOVA are as follows:

- Aspiration: this clearly has an effect on VOT as VOT is the main acoustic cue that signals the difference between aspirated phonemes and unaspirated phonemes.

- Consonant place of articulation: different places of articulation are known to have a difference in VOT, with VOT generally greatest for velar plosives and least for bilabial plosives (Hayward (2000)). The places of articulation for Walungge are labial, alveolar, retroflex, palatal and velar.

- Tone: as has already been noted, plosives before low tone have on average a lower VOT than plosives before high tone.

- Syllable weight (monosyllabic words only): Syllable weight has been included in the ANOVA because for monosyllabic words a difference in syllable weight correlates with a difference in relative starting F0 of the syllable rhyme (see section 4.3.2 above). If VOT correlates with F0 (rather than underlying tone), then because relative F0 varies with syllable weight VOT also might be expected to vary with syllable weight.

Table 6-11 below gives the results of the ANOVA. The significance level is taken to be $p < 0.05$.

Table 6-11: ANOVA for effects on VOT

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
place of articulation	610	4	2.56	0.037	•
tone	610	1	93.7	<0.0005	•
aspiration	610	1	676	<0.0005	•
syllable weight	184	1	0.99	0.321	

As well as a difference in aspiration and a difference in tone correlating with a significant difference in VOT, place of articulation also has a significant effect. These results are unsurprising, and are included in the ANOVA for completeness. The important result is that the above figures show no significant correlation between syllable weight and VOT. These results include both aspirated and unaspirated data, thus a post-hoc Scheffé test was carried out to investigate further whether weight (and hence F0) has any effect, with the following results:

Table 6-12: Scheffé post-hoc test for VOT; aspirated plosives

tone; weight	high; heavy		low; light		low; heavy	
	P	*sig.*	*P*	*sig.*	*P.*	*sig.*
high; light	0.753		0.001	•	0.001	•
low; heavy	0.006	•	0.995			
low; light	0.005	•				

Table 6-13: Scheffé post-hoc test for VOT; unaspirated plosives

tone; weight	high; heavy		low; light		low; heavy	
	P	*sig.*	*P*	*sig.*	*P.*	*sig.*
high; light	0.970		<0.0005	•	<0.0005	•
low; heavy	<0.0005	•	0.916			
low; light	<0.0005	•				

For neither aspirated not unaspirated plosives, and for neither high nor low tone words, does weight have any significant effect upon VOT. In the case of aspirated plosives, Figure 6-8 above shows that there is considerable overlap in VOT between high tone and low tone words. Because of this overlap, any slight correlation of VOT with weight (and hence F0) could be masked. In the case of unaspirated plosives there is not the same overlap in VOT; further the VOT values have a large range, from approximately -75ms to approximately 30ms. Yet it would appear that weight (and hence F0) is not influencing these values. This suggests that for unaspirated plosives whilst a difference in VOT correlates with a difference in tone, a difference in VOT does not correlate with a difference in F0 any further than relative F0 is a concomitant of tone. The starting F0 of the syllable rhyme varies according to the syllable weight but VOT does not vary along with it.

6.3.3 Affricates and fricatives

In considering the question of whether word initial voicing is caused by tone or F0, attention is now turned to affricates and fricatives. In Walungge fricatives are a separate phonological class of sounds from plosives and affricates in that they do not become voiced with low tone. Affricates are of particular interest because phonologically they pattern with plosives yet phonetically they also have similarities with fricatives. In particular, both have a period of frication immediately before the vowel which is carrying the tone. If it is F0 not tone which is having an effect on word initial obstruents, then there might be changes to this period of frication for both affricates and fricatives. On the other hand, if the effect is from tone not F0 then affricates and fricatives might show different acoustic changes.

6.3.3.1 Affricates

Phonetically affricates are sequences of a plosive released into a homorganic fricative. Depending on language, an affricate may be one phoneme (as in Walungge) or it may be a sequence of two phonemes. In Walungge affricates pattern phonologically with plosives both diachronically and synchronically with regard to aspiration.

Synchronically, aspiration is phonologically contrastive for affricates just as it is for plosives. The following spectrograms show an aspirated and an unaspirated affricate:

Figure 6-9: spectrogram of *tɕʰá* 'fertilizer'

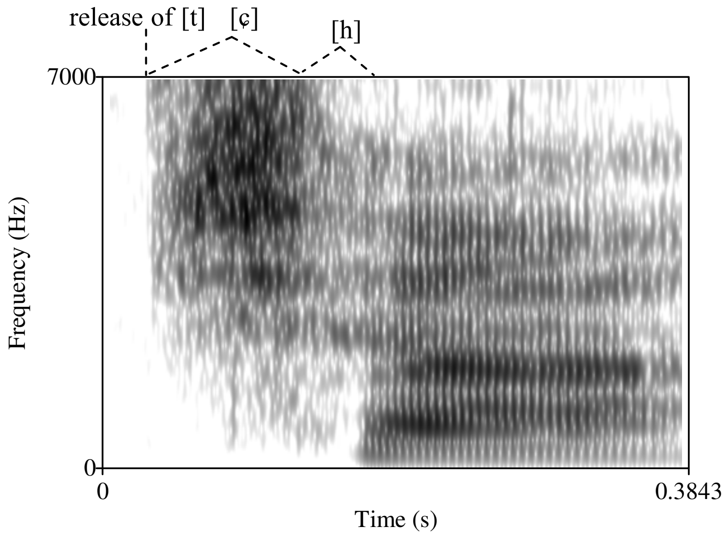
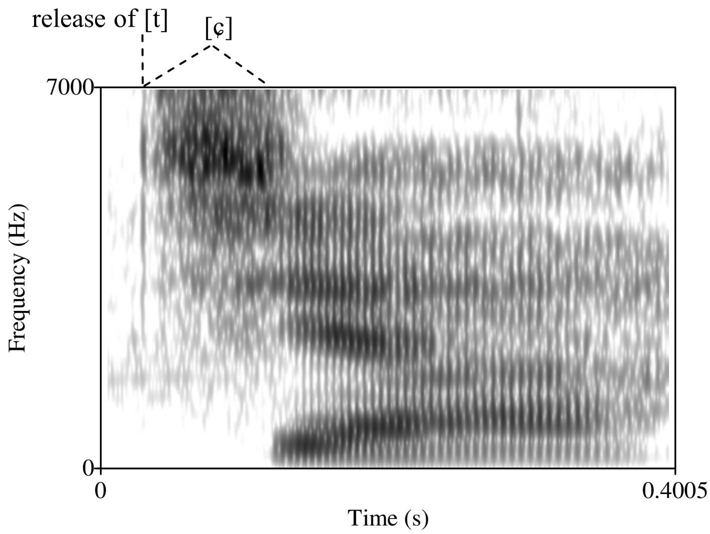


Figure 6-10: spectrogram of *tɕá:* 'iron'



Diachronically, comparing the Written Tibetan with Walungge the same correlations for tone and aspiration exist between Written Tibetan and affricates as between Written Tibetan and plosives. The following table gives examples of this:

Table 6-14: Examples of aspirated plosives and affricates

WT		Walungge		gloss
onset	*example*	*plosive/affricate*	*example*	
aspirated	<thag pa> <tshi lu>	aspirated; high	*t*^{h}*ákpa* *ts*^{h}*ílu*	‘rope’ ‘fat’
voiced	<dom> <ja>	aspirated; low	*t*^{h}*òm* *tɕ*^{h}*à*	‘bear’ ‘tea’

Walungge aspirated affricates are phonetically different for high and low tone words in a similar way to the difference in aspirated plosives for high and low tone words. Before high tone Walungge aspirated affricates have heavy aspiration, as do aspirated plosives. Before low tone the aspiration of affricates has a tendency to be lighter and shorter in duration. However, the aspiration difference for affricates is not as pronounced as for plosives.

Whilst affricates pattern with plosives in terms of aspiration, there is a difference between affricates and plosives for unaspirated segments. In particular, word initial unaspirated plosives before low tone are phonetically voiced, whilst unaspirated affricates before low tone have lowered VOT but are not voiced for words in isolation. However, if word initial unaspirated affricates are in a phrase, there is generally voicing through the closure of the affricate but not through the fricative portion of the affricate. In other respects unaspirated plosives and affricates pattern together both synchronically and diachronically. The following table gives examples of unaspirated plosives and affricates, along with the Written Tibetan. Note that because the VOT is lowered, because they become voiced in phrases, and because they pattern with plosives, unaspirated affricates with low tone are transcribed as voiced affricates.

Table 6-15: Examples of unaspirated plosives and affricates

WT		Walungge		gloss
onset	*example*	*plosive/affricate*	*example*	
voiceless unaspirated	<rta> <rtsa>	unaspirated high	*tá* *tsá*	‘horse’ ‘vein’
pre-scripted consonant + voiced	<mdangs> <mdzo>	unaspirated low	*dàŋ* *dzò*	‘yesterday’ ‘yak cross’

6.3.3.2 Statistical analysis of affricates

Word initial plosives, both aspirated and unaspirated, have a significant difference in VOT correlating with a difference in tone. In order to consider whether there is a comparable difference in VOT for affricates, the duration of word initial affricates was measured from the point of release of the closure until the onset of voicing. This essentially measures the duration of the frication for the unaspirated affricate, and duration of frication plus aspiration for aspirated affricates.

Below is a boxplot (with median, interquartile range, and whiskers of 1.5x interquartile range) and table of means of VOT for affricates.

Figure 6-11: Boxplot of VOT for affricates

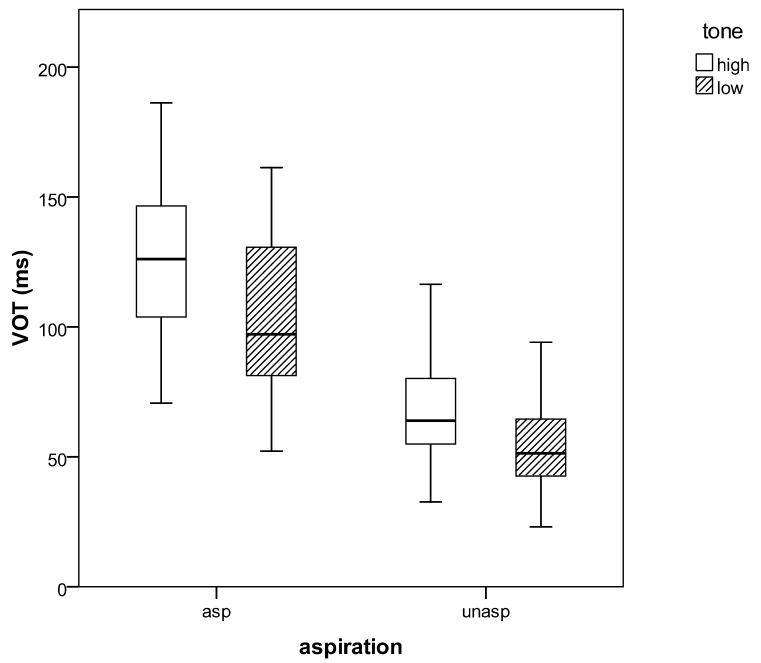


Table 6-16: mean and standard deviation of VOT for different affricate types

tone	*affricate type*	*mean VOT (ms)*	*std. dev.*	*N*
high	aspirated	127.7	29.5	69
high	unaspirated	68.6	19.2	60
low	aspirated	103.5	29.1	51
low	unaspirated	55.4	18.2	37

Both aspirated and unaspirated affricates, with both high and low tone, have considerable variation in the duration of the fricative, and hence in the VOT. This variation in the duration of the fricative means that there is overlap in VOT between aspirated and unaspirated affricates. Unlike plosives, where a difference in VOT corresponds to a difference in aspiration, for affricates it is the presence or absence of aspiration, not merely the duration of VOT, which gives the contrast[16]. Despite the variation in VOT, for both aspirated and unaspirated affricates the average VOT is less before low tone.

An ANOVA was carried out for VOT, with independent variables: aspiration, tone, and place of articulation. The significance level is taken to be $p < 0.05$. The following table gives the results of the ANOVA.

Table 6-17: ANOVA for effects on affricate VOT

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
place of articulation	217	1	0.37	0.544	
tone	217	1	10.8	0.001	•
aspiration	217	1	213.9	<0.0005	•

A difference in place of articulation (alveolar versus palatal) does not correlate with a significant difference in VOT however both tone and aspiration do, with low tone correlating with lower VOT. This means that a difference in tone correlates with

[16] .Measuring affricate VOT from the release of the sibilant rather than the release of the initial closure would give VOT measurements more comparable to those of plosives. However, in many cases there was not a visually clear point of release of the sibilant, which is why the measurements were taken from the point of release of the closure.

a difference in the VOT of affricates in the same way as it does for plosives, even though the fricative portion of the affricate does not become voiced before low tone.

6.3.3.3 Fricatives

Word initial fricatives in Walungge are always voiceless, regardless of tone. In this way they are similar to the fricative portion of affricates, which is always voiceless word initially. Even though word initial affricates in isolation words do not become voiced before low tone, the VOT (essentially a measurement of the duration of the frication) is significantly reduced when the tone is low. In order to consider whether a difference in tone correlates with a difference in word initial fricatives, the VOT can similarly be measured. The correlation between the duration of a fricative and tone can be found in other languages. For example, in Kera, a Chadic language, fricatives have a duration difference rather than a voicing difference, with shorter duration correlating with low tone (Pearce (2005)).

Statistical analysis of fricatives

The VOT of fricatives was measured as being the duration from the start of the fricative until the onset of vocal fold vibrations. The following boxplot shows the VOT measurements for high and low tone.

Figure 6-12: VOT for fricatives

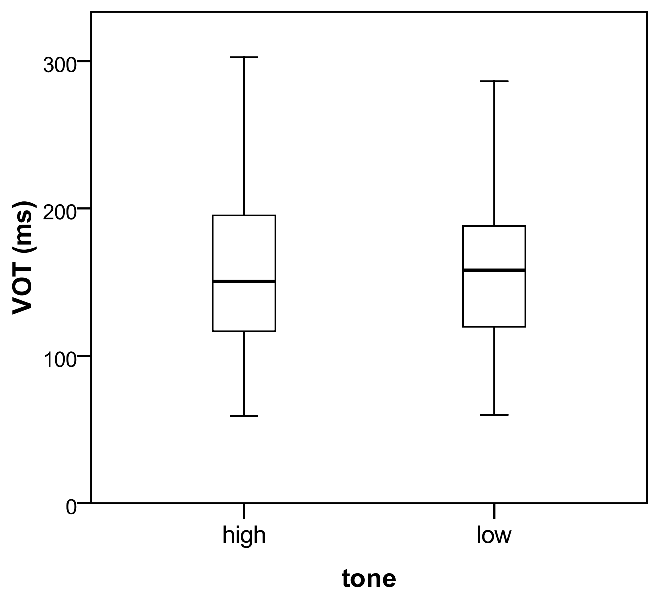


Table 6-18: mean and standard deviation of VOT for fricatives

tone	*mean VOT (ms)*	*std. dev.*	*N*
high	166.2	71.8	83
low	166.8	74.9	73

Two things are apparent from the above boxplot and table. Firstly, there is considerable variation in the VOT. Secondly, tone makes no apparent difference to the VOT.

An ANOVA was carried out for the effects on fricative VOT, with independent variables: place of articulation and tone.

Table 6-19: ANOVA for effects on fricative VOT

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
place of articulation	156	1	3.18	0.076	
tone	156	1	0.002	0.964	

Neither a difference in place of articulation nor tone correlates with a significant difference in fricative VOT (although place of articulation is only just above the significance level of 0.05). It is notable that a difference in tone correlates with a difference in VOT for affricates but not for fricatives. If it is simply the phonetics of articulating a fricative with lower F0 that is causing the reduced period of frication for affricates, then the same effect might be expected for fricatives. This is an indication that although phonetically, affricates might be similar to fricatives, phonologically they are different. Further, it is an indication that it is the low tone rather than the phonetics of articulating a fricative with lower F0 which is causing the lowered VOT of affricates.

Although the voicing of plosives and affricates is being analysed as being caused by the tone rather than being a phonological contrast, it is of relevance to note that cross linguistically a voicing contrast in plosives is far more common than a voicing contrast in fricatives. Maddieson (2008) gives the following data from a cross-linguistic survey of 567 languages:

Table 6-20: voicing in plosives and fricatives

contrast	number of languages
no voicing contrast	182
voicing contrast in plosives alone	189
voicing contrast in fricatives alone	38
voicing in both plosives and fricatives	158

Maddieson analyses these figures, showing that the presence of fricative voicing is significantly dependent upon the presence of plosive voicing.

The possible implication of this, when considering the diachronic shift away from an obstruent voicing contrast, is that languages are likely to lose their fricative voicing before their plosive voicing. This can be seen with Tibetan languages. There some Tibetan languages which have retained both plosive and fricative voicing contrastively (e.g. Kyirong (Huber (2005)), those which have retained contrastive voicing for plosives but lost it for fricatives (e.g. Dolpo (own data)), and those which have no voicing contrast for any obstruents (e.g. Lhasa Tibetan (Chang & Shefts (1964)). There are none which have contrastive voicing for fricatives but not for plosives.

6.3.4 F0 and intervocalic obstruents

The purpose of the above acoustic analysis has been to consider whether it is the tone or the F0 which is affecting the VOT of word initial obstruents. If it is F0 rather than tone which is affecting the VOT of word initial obstruents, there might be a similar effect upon word medial obstruents. It has already been discussed that even though the tone on the second syllable is always high, the starting F0 of the second syllable varies according to the underlying tone of the word due to peak delay. Also, if the tone is low and the second syllable is bimoraic, the low tone spreads to the first mora of the second syllable resulting in lowered F0 at the start of the second syllable. It has also been discussed that intervocalic plosives are not completely voiced. The voicing tapers off during the closure and starts upon release of the plosive. Thus the VOT of intervocalic plosives can be measured. If the VOT of intervocalic plosives varies as F0 varies, this could be evidence that it is F0 which is affecting the VOT of initial plosives.

The duration between the release of the plosive and the onset of voicing was measured for both voiced and voiceless intervocalic plosives. Because of the way that voicing typically tapers off during the closure, if a plosive was voiced throughout its closure its VOT was taken to be zero. In this respect, although this is being termed VOT it is a slightly different measurement from the normal way that VOT is calculated for word initial plosives; there is no negative VOT in these measurements. An ANOVA was carried out to check whether either the F0 caused by the underlying tone or the F0 caused by the spreading of tone L to the first mora of a heavy second syllable has a significant effect upon the VOT of intervocalic plosives. The following are the results of the ANOVA and a table of mean VOT measurements.

Table 6-21: ANOVA for effect of tone upon VOT for intervocalic plosives

Effect	*N*	*d.f.*	*F*	*P*	*Sig*
effect of underlying tone for voiced plosives	39	1	1.74	0.195	
effect of underlying tone for voiceless plosives	131	1	2.34	0.129	
effect of low tone spreading for voiced plosives	22	1	0.484	0.495	
effect of low tone spreading for voiceless plosives	48	1	0.242	0.625	

Table 6-22: mean and standard deviation for VOT for intervocalic plosives

voicing	*underlying tone*	*VOT (ms)*	*std. dev.*	*N*
voiced	H	3.8	5.4	17
	L	6.7	7.8	22
voiceless	H	17.1	6.3	83
	L	18.7	4.6	48

There is no significant correlation between intervocalic VOT and underlying tone or intervocalic VOT and low tone spreading. That is to say, the VOT of an intervocalic plosive is not significantly influenced by the F0 at the start of the second syllable.

6.3.5 Summary of acoustic analysis

The above acoustic analysis shows that the VOT of initial plosives varies as the tone varies, becoming higher for high tone and lower for low tone. For monosyllabic words the relative starting F0 varies not only according to the underlying tone but also according to the syllable shape of the word. However when the F0 varies according to syllable shape the VOT does not similarly vary. The above acoustic analysis also shows that the duration of the fricative portion of an affricate varies with tone, but the duration of a normal fricative does not vary with tone. If the fricative portion of an affricate was varying in duration because of F0 rather than tone, then it might be expected that a normal fricative would similarly vary. However it does not, suggesting that it is tone rather than F0 which is affecting the affricate. The final section of the above acoustic analysis shows that the VOT of intervocalic plosives does not vary as F0 varies. If word initial VOT was varying with F0 it might be expected that intervocalic VOT would also vary with F0. However, it doesn't, which is further evidence that the word initial variation is being caused by tone rather than F0. These results are drawn on in section 6.6 when considering the phonology of tone.

6.4 Word initial sonorants

Word initial r, r̥, l, l̥

In addition to voiced and voiceless unaspirated plosives and affricates, Walungge also has both voiced and voiceless liquids word initially. The following are examples of these:

r̥ý:	‘corn’	*rỳt*	‘landslide’
r̥ánde	‘solid’	*ràdak*	‘root’
l̥ári	‘artist’	*láʐa*	‘wage’
l̥énda	‘patch’	*létta*	‘brain’
		làŋa	‘upper arm’
		lùŋba	‘place’

The voiceless segments [r̥] and [l̥] only occur with high tone, and only word initially. The voiced segment [r] only occurs with low tone, but [l] occurs with both high and low tone. Word medially and word finally neither [r̥] nor [l̥] occurs. This is, to some extent, an “opposite” relationship with tone to the plosives. Whereas for plosives it is low tone and voicing which correlate, for liquids it is high tone and voicelessness. However, in the case of plosives the low tone can be analysed as causing the voicing, whereas with liquids high tone cannot be analysed as causing the voiceless, due to both [l] and [l̥] occurring with high tone.

The difference in distribution between [l] and [r] comes from the diachronic phonology. Whereas WT consonant clusters C + lateral have become lateral + high tone, e.g:

<klad pa> *létta* ‘brain’

consonant clusters C + vibrant have generally become a retroflex consonant, e.g:

<skra> *ʈá* ‘hair’

The WT consonant clusters of <sl> and <sr> both give rise to a high tone liquid. The cluster <sl> gives rise to [l] with high tone; the cluster <sr> gives rise to [r̥] with high tone, e.g.

<srap>	*r̥ap*	‘horses bit’
<slob grwa>	*lópʈa*	‘school’

The segment [l̥] has arisen from WT consonant cluster lh, e.g:

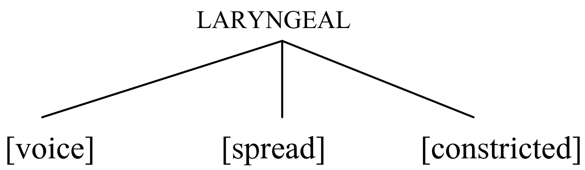
<lhan pa> *l̥énda* ‘patch’

Because of the complementary distribution between [r̥] and [r], these could be analysed as allophones. However this creates a dissymmetry in that [r̥] and [r] would be analysed as realisations of the one phoneme /r/, whereas [l] and [l̥] would be analysed as realisations of separate phonemes /l/ and /l̥/. The question of how to analyse [r̥] and [r] is not unique to Walungge. Other Tibetan languages show the same distribution of liquids. In some cases in these languages [r] and [r̥] (or its equivalent, e.g. [r^h]) have been analysed as realisations of separate phonemes, e.g. Lhasa Tibetan (Denwood (1999), Tournadre & Dorje (2003)). In other cases they have been analysed as allophones of the one phoneme /r/, e.g. Lhomi (Vesalainen & Vesalainen (1976)). There appears to be no clear answer for Walungge; however given that [r̥] patterns distributionally with [l̥] and [l̥] must be analysed as the realisation of /l̥/, it has been decided for this dissertation to analyse [r̥] as the realisation of the phoneme /r̥/.

6.5 Laryngeal features for Walungge segments

6.5.1 Laryngeal features and voicing

Before considering further the interaction between tone and consonants in Walungge, attention is first given to the laryngeal features which are operational in Walungge at a segmental level. Laryngeal features have generally been assumed to be attached to a laryngeal node (Clements (1985), Lombardi (1991)) as in the following diagram[17]:



However, whilst the features [spread] and [constricted] are generally taken to be privative features, there has been debate over the nature of the feature [voice], as to whether it is privative or binary, and even if it is binary whether there is

[17] Avery & Idsardi (2001) propose a different organisation of laryngeal features from what is described here.

underspecification. In a language where there is no voicing contrast in obstruents, it can be argued that the feature [voice] is not present in the phonology for obstruents, and they are articulated as voiceless by default. If there is a contrast for voicing of obstruents, one possibility is that it is the presence or absence of a privative feature [voice] which is causing the difference in voicing. Another possibility is that voiced obstruents are specified as [+voice] and voiceless obstruents are specified as [-voice]. In a language where there is a voicing contrast for plosives but not for fricatives, for example, even if binary features are adopted it is possible to have underspecification for the class of sounds for which there is no voicing contrast. Lombardi (1991) argues that all laryngeal features including [voice] are privative. They are also treated as privative in Clements & Hume (1995). Bradshaw (1999) treats [voice] as privative. Avery & Idsardi (2001) also treat all laryngeal features as being privative. On the other hand, Wetzels & Mascaro (2001) argue for binary rather than privative features for [voice] because [-voice] is sometimes active phonologically.

6.5.2 Voiceless sonorants and aspirated obstruents

Most languages do not have contrastive voicing for sonorants, in which case all sonorants are realised as voiced. Thus sonorants can be seen as being unspecified for voicing in their phonology. If this is the case, there are several possibilities that can account for them being realised as voiced. One is that the feature [voice] attaches to them post-lexically. Another is that they are underspecified even in their surface form, and their surface voicing is simply the result of them being sonorants.

Section 6.3.1 above came to the conclusion that voicing of plosives and affricates in Walungge is not phonologically contrastive. Neither do fricatives have contrastive voicing; they are always realised voiceless. Thus for obstruents the only laryngeal feature that is active is the feature for aspiration, which can be specified by the feature [spread]:

```
LARYNGEAL
    |
 [spread]
```

Aspiration only occurs word initially. The neutralisation of contrast between aspirated and unaspirated segments in all environments other than word initially can be represented by the following diagram which illustrates delinking of the laryngeal node when it follows another segment:

X X

LARYNGEAL

Lombardi (1995) argues that laryngeal neutralisation is the result of the delinking of the laryngeal node, and that this delinking happens because the language only licenses laryngeal features in particular configurations. She argues that languages which have laryngeal neutralisation have the constraint that “a laryngeal node is only licensed in a consonant if it immediately precedes a [+sonorant] segment in the same syllable”. Walungge can be analysed as having an even stronger version of this constraint, namely that the laryngeal node is only licensed in a consonant word initially.

In the case of sonorants, there is a word initial voicing contrast for liquids, e.g:

l̥énda	‘patch’	*létta*	‘brain’

In all other environments all sonorants are voiced, which is the default way in which sonorants are realised cross linguistically. If the contrast between voiced and voiceless sonorants is to be described using the feature [voice], then voice would need to be a binary feature with the negative value of the feature as the active value which is creating the contrast. Lombardi (1991) argues that the feature responsible for a voicing distinction in sonorants is [asp] not [voice]. This is the same feature which is responsible for aspiration of obstruents, and is equivalent to the feature [spread] used elsewhere in the literature. Lombardi presents data from a number of different languages which have a voicing distinction in sonorants, showing that voiceless sonorants and voiceless obstruents do not pattern together phonologically, and neither do voiced sonorants and voiced obstruents. Rather, the phonological patterning is between voiceless sonorants and aspirated obstruents.

Two of the languages Lombardi draws on are the Tibetan languages of Kagate and Lhomi, which have the aspirated sonorants [r^{h}] and [l^{h}]. These sonorants are equivalent to what Watters (2002) describes as the preaspirated sonorants [hr] and [hl] in the Tibetan languages that he discusses, including Lhomi and Lhasa Tibetan. This is also equivalent to the Kyirong lateral fricative /ɬ/ (Huber (2005)). They are also equivalent to the Walungge phonemes /r̥/ and /l̥/. From Tibetan language to language the phonetic realisation might be slightly different, but in each case [l̥], [ɬ], [hl] and [l^{h}] all equate to the Written Tibetan <lh>, and [r̥], [hr] and [r^{h}] equate to the Written Tibetan <sr>. Thus from a comparative perspective, the voicing distinction in Walungge sonorants is equivalent to an aspiration distinction, so using the feature [spread] for this distinction is appropriate. However, it is important to consider the phonological patterning in Walungge itself, and not just the comparative perspective.

In Walungge both aspirated obstruents and voiceless sonorants only occur word initially. So in this respect using the feature [spread] for both types of sounds does indeed pick up on this similarity between them. The following data shows reduplicated verbs which start with either an aspirated obstruent or voiceless sonorant. Non-initially the aspiration is lost and the voiceless sonorant becomes voiced.

r̥éː	'mix'	[r̥eːreː $^{44\ 44}$ eːla $^{44\ 21}$]	'before mixing'
k^{h}*úr*	'carry'	[k^{h}ukːur $^{44\ 44}$ eːla $^{44\ 21}$]	'before carrying'

However, whereas aspirated obstruents occur with both high and low tone:

tɕ^{h}*ú*	'water'	*tɕ*^{h}*à*	'tea'
ʈ^{h}*í*	'bed'	*ʈ*^{h}*ì*	'knife'

voiceless sonorants only occur with high tone.

If the voiceless sonorants were to be analysed as being the cause of the high tone of the word, then the feature that distinguishes them could not be [spread] because aspirated obstruents (and hence the feature [spread]) occur with both high and low tone. However, if the voiceless sonorants were analysed as having a distribution which is limited to high tone rather than them being the cause of the high tone, then the feature [spread] could be used for them. This is the analysis that is being opted for here. The explanation for a limited distribution is due to the diachronic phonology, with the link between Written Tibetan and voiceless sonorants discussed in section 6.4 above. This means that voiceless sonorants have the feature [spread] and voiced sonorants are unspecified for any laryngeal feature.

6.5.3 Non word initial voicing

In Walungge whilst the word initial voicing of plosives and affricates is caused by the tone, in other environments it is caused by the surrounding segments. Next to a sonorant consonant or intervocalically, plosives and affricates are voiced. If the privative feature [spread] is being used to distinguish between voiced and voiceless sonorants, with voiced sonorants unspecified for laryngeal features, there is no feature [voice] to spread from sonorant to obstruent. However, the voicing of obstruents next to sonorants may be treated as spontaneous rather than non-spontaneous voicing, and thus the result of the phonetics rather than the result of phonological assimilation (Halle & Stevens (1971), Lombardi (1991), Harris (2003), Harris (2009)). The voicing of sonorants happens spontaneously as a result of the lack of pressure in the oral cavity. In contrast, the build-up of oral pressure for an obstruent means that generally voicing cannot happen spontaneously; it requires a non-spontaneous compensatory gesture. However, when obstruents are next to sonorants it is possible for the spontaneous voicing that happens for the sonorants to be passed passively through the obstruent, particularly if the duration is short.

The acoustic properties of the Walungge intervocalic obstruents support the analysis that non word initial voicing is spontaneous, and thus a result of the phonetics rather than the phonology. Section 6.2.2.2 describes how intervocalic plosives are not completely voiced. At the point of closure of the plosive the vocal folds are vibrating because of the preceding vowel. The voicing gradually dies away during the closure, and starts again shortly after the release of the closure. Neither are sequences of two plosives completely voiceless. As is the case with voiced plosives, the voicing from the preceding vowel gradually dies away during the closure. However because of the longer duration of closure, the proportion of closure which is voiced is far less (approximately 25% compared with approximately 85%). All phonological fricatives remain voiceless, whereas intervocalic affricates are realised as voiced fricatives. This can also be attributed to the phonetics. Fricatives which are phonologically affricates have a considerably shorter duration than phonological fricatives; and spontaneous voicing is more likely to occur with shorted segments. Similarly, the voicing of plosives and affricates, but not fricatives, next to sonorant consonants can be accounted for as spontaneous voicing.

In summary then, the only laryngeal feature which is specified underlyingly for Walungge consonants is the privative feature [spread]. The voicing of sonorants and voicelessness of obstruents is not the result of any phonological feature [voice] (either privative or binary), but is simply the result of the phonetics. The voicing of plosives next to sonorants is also the result of the phonetics and not the result of feature spreading.

6.6 Tone and consonants in Walungge

This section considers the word initial correlation between tone and voicing. Firstly this section outlines two different models that have been put forward to capture the relationship between tone and voicing of obstruents, which have been chosen because of their relevance to Walungge. These are Halle & Stevens (1971) and Bradshaw (1999)[18]. Following this is a discussion of how best to capture the Walungge relationship between tone and voicing.

6.6.1 Tone and voicing proposals

Halle & Stevens (1971)

One of the earliest models to capture the link between tone and voicing was that proposed by Halle & Stevens (1971). The basis of this model is the tension in the

---

[18] Other models which are not discussed here include Bao (1999), Duanmu (2000) and Avery & Idsardi (2001). These all have a greater complexity than is needed for Walungge, e.g. more than two levels of tone and/or different combinations of voicing/tone/phonation.

vocal folds. Increased tension raises pitch and decreased tension lowers pitch. The tension in the vocal folds also affects the ability of the vocal folds to vibrate. Thus Halle & Stevens proposed two binary features [±stiff vocal folds] and [±slack vocal folds], as well as two further laryngeal features, [±spread glottis] and [±constricted glottis]. The following table shows the different combinations of these features (Halle & Stevens (1971) reprinted in Halle (2002:51)).

Table 6-23: laryngeal features proposed in Halle & Stevens (1971)

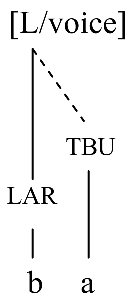
	1	2	3	4	5	6	7	8	9
obstruents	b_1	b	p	p_k	b^h	p^h	ɓ	ʔb	$p^{ʔ}$
vowels	V	V̀	V́	V̥	V̤			V̰	ʔV
spread	-	-	-	+	+	+	-	-	-
constricted	-	-	-	-	-	-	+	+	+
stiff	-	-	+	-	-	+	-	-	+
slack	-	+	-	-	+	-	-	+	-

Column 2 in the above table shows that when the feature [+slack] attaches to an obstruent (and all other features take their negative value), the result is a voiced obstruent. When it attaches to a vowel the result is low tone. Similarly column 3 shows that [+stiff] (with other features negative) gives rise to a voiceless obstruent and high tone. If all features are negative (column 1), the result on a vowel is mid tone (the unmarked tone in a 3 tone system); for an obstruent the result is contextual voicing. The difference between the obstruents in columns 1 and 2 can be seen as the difference between spontaneous and non-spontaneous voicing. The feature [+spread] is responsible for aspiration on obstruents (with the obstruent in column 4 being a moderately aspirated plosive; thus the difference between columns 4 and 6 is the degree of aspiration), and breathy vowels or voiceless vowels. The feature [constricted] is responsible for implosives or glottalised obstruents, and glottalised or creaky vowels.

Bradshaw (1999)

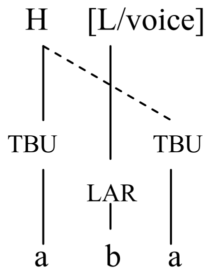
Bradshaw (1999) adopts a multiplanar approach in order to capture the relationship that exists in many languages between voiced obstruents and low tone. She notes that there is a wealth of cross-linguistic evidence to suggest that tone is suprasegmental, whereas voicing is subsegmental. Under an autosegmental model this means that TBUs and laryngeal nodes are on separate geometric planes. Given this consonant-tone interaction, Bradshaw proposes that there is the one phonological entity which on the one hand attaches to suprasegmental TBUs and on the other hand attaches to the laryngeal node of segments. This phonological entity she refers to as the privative feature [L/voice]; when attached to a TBU there is low tone, when attached to a laryngeal node there is voicing.

The following is a diagram showing [L/voice] spreading from the consonant [b] to the vowel [a], with voicing on the consonant and low tone on the vowel:



The feature [L/voice] is very similar to Halle & Stevens' [slack] in that it is responsible for both low tone and voicing. However, unlike Halle & Stevens who have the feature [stiff] for both high tone and voicelessness, Bradshaw does not have an equivalent feature for high tone; high tone is simply H. Her claim is that it is the presence not the absence of voicing which is interacting with tone.

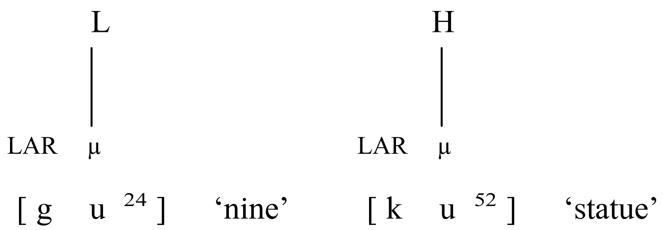
A further advantage of saying that TBUs and laryngeal nodes are on separate geometric planes and yet it the one phonological entity which can attach on both planes, is that tones can spread across words from TBU to TBU without having to cross an association line from a consonant. This means it is possible to have, for example in the following diagram, tone H spread across two vowels [a] with a voiced obstruent [b] in-between:



6.6.2 Tone and voicing in Walungge

6.6.2.1 Different proposals

When considering the correlation between tone and voicing[19] in Walungge it is important to note that tone in Walungge is suprasegmental in nature, with the TBU as the mora and the foot as the tonal domain, as discussed in Chapter 5 above. However, voicing is subsegmental. For this reason, the distinction that Bradshaw (1999) makes between the TBU and the laryngeal node is appropriate for Walungge. Even if this distinction between the TBU and the laryngeal node is adopted, there is more than one possibility for describing the tone and voicing correlation. The first is that neither of the above models apply. In the case of a low tone word starting with an unaspirated plosive/affricate, tone L attaches to the TBU but not to the laryngeal node of the initial consonant, and the voicing on the consonant is spontaneous rather than non-spontaneous; similarly H attaches only to a TBU[20]:



The other possibility is that that the voicing is non-spontaneous, in which case both Halle & Steven's model and Bradshaw's model need to be considered. Assuming a distinction between the TBU and the laryngeal node, the main difference between the

[19] i.e. surface voicing: voicing has already been argued to be not phonologically contrastive

[20] Because the current discussion is about the nature of the underlying tone and its interaction with the initial consonants, boundary tones H% and L% have been omitted from the diagrams below.

two models is the phonological nature of high tone. This is illustrated below, applying the two models to Walungge.

As outlined above, Halle & Stevens (1971) use the features [±stiff] and [±slack] for both tone and voicing. The combinations of these feature give 3 different tone levels ([+stiff][+slack] is not a possible combination). However as Walungge only has 2 tone levels, it is only necessary to use one of these features. Thus the feature [±slack] will be used for the contrast between H and L. In this model high and low tone are opposite values of the same type of phonological entity, e.g

[+slack]   [-slack]

LAR   μ   LAR   μ

[g   u ^{24}]   'nine'   [k   u ^{52}]   'statue'

Whereas in Bradshaw's model, it is low tone but not high tone which interacts with consonants, e.g.

[L/voice]   H

LAR   μ   LAR   μ

[g   u ^{24}]   'nine'   [k   u ^{52}]   'statue'

Each of the above proposals will be discussed in the sections which follow. What is important to note here is that it is not the terminology which is the issue but the structure. That is to say, the issue is not whether a phonological entity is termed L or [+slack] or [L/voice] but whether that phonological entity attaches to a TBU or to both a TBU and a laryngeal node.

6.6.2.2 Spontaneous voicing?

Consider Walungge data such as the following, where the initial plosive is voiced if there is low tone.

pák   'pastry'

bàk   'a kind of bamboo'

The voicing of plosives in Walungge has been argued to be not phonologically contrastive, and thus plosives are not specified for [voice] (or any equivalent feature such as [slack]) in their underlying form. The phonetic realisation of the low tone requires the F0 to be lowered. The changes in the laryngeal configuration which are needed to lower the F0 are the same configurations which affect voicing of plosives and vowel phonation (laryngeal configuration and the link between pitch, voicing and phonation has been discussed in section 2.4). Thus the lowering of F0 has the added effect of causing the vowel phonation to become breathier and causing the plosive to become voiced.

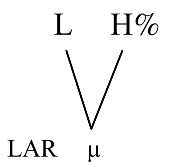
The analysis of section 6.3 above considers from an acoustic perspective whether it is the tone or the F0 which is responsible for the change in VOT, with the conclusion that VOT is affected by tone rather than by F0. If the VOT of word initial plosives was the result of F0 then VOT would be expected to vary as the relative value of F0 varies. However, the acoustic analysis in section 6.3.2 shows that a difference in VOT correlates with a difference in tone not a difference in relative F0. Similarly, it is the tone rather than F0 which can explain why affricates have reduced frication duration with low tone, but fricatives do not (section 6.3.3). If the reduced duration was the result of F0 then fricatives would also be expected to have reduced duration. However, if the reduced duration is the result of the low tone interacting with affricates but not fricatives, the data is accounted for.

Apart from the acoustic arguments, the primary reason for analysing the word initial voicing as non-spontaneous rather than spontaneous comes from the negative forms of verbs, such as:

root	*negative form*	*gloss*
k^{h}úr	[mak^{h}ur $^{44\ 41}$]	‘carry’
p^{h}àp	[map^{h}ap˺ $^{11\ 42}$]	‘dismount’
túŋ	[matuŋ $^{44\ 41}$]	‘drink’
dø̀t	[madøʔ $^{11\ 42}$]	‘sit’

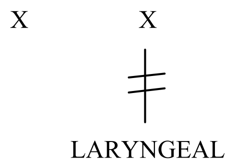
Negative forms of verbs are discussed in greater detail in section 5.5.4 above and 6.6.2 below. Section 5.5.4 presents arguments suggesting that a) there is a prosodic word (pword) boundary between a negative morpheme and verb root and b) the negative morpheme and the verb root together form a pword; the analysis is a recursive pword. The primary argument for a pword boundary between the negative morpheme and the root is the aspiration on the plosives; the primary argument for the negative morpheme and verb root together forming a pword is the attachment of the underlying tone of the verb root to the negative morpheme rather than the root.

Consider now the unaspirated plosives. The following diagram[21] illustrates the possibility that there is no attachment to the laryngeal node and thus voicing is spontaneous:



[d ø ʔ ^{24}] 'sit'

Section 6.5.2 argues that the laryngeal node is deleted in all environments other than word initial:



In the negative form of verbs, the laryngeal node of a root initial plosive is not deleted because it immediately follows a pword boundary. This accounts for the aspiration on the initial consonant of the verb root:

[21] For clarity, boundary tone L* (see 5.5) is omitted from the diagrams below.

L H%

[spread]

μ LAR μ

[m a # pʰ a p˺ ¹¹ ⁴²] 'dismount.NEG'

The following illustrates the situation for root initial unaspirated plosives, with the pword boundary marked #:

L H%

μ LAR μ

[m a # d ø ʔ ¹¹ ⁴²] 'sit.NEG'

H H%

μ LAR μ

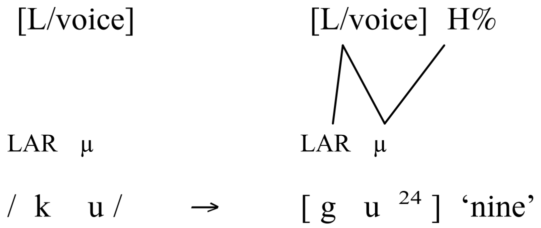
[m a # t a p˺ ⁴⁴ ⁴²] 'plant.NEG'

As described in section 6.5 above, at an intervocalic morpheme boundary plosives become voiced. However, for an underlyingly high tone verb in negative form the root initial plosive is voiceless and for a low tone verb it is voiced. This difference in voicing cannot be due to the intervocalic environment, and neither can it be due to the following pitch. It could not be due to the pitch of the preceding syllable because the acoustic analysis of section 6.2.2.2 above shows that voicing of an intervocalic plosive is not affected by the pitch of the preceding syllable. For the negative verbs, there is no purely phonetic explanation for the intervocalic voicing in low tone verbs and voicelessness in high tone verbs. This suggests that the voicing is not spontaneous voicing, but is being caused by the spreading of tone L (see below).

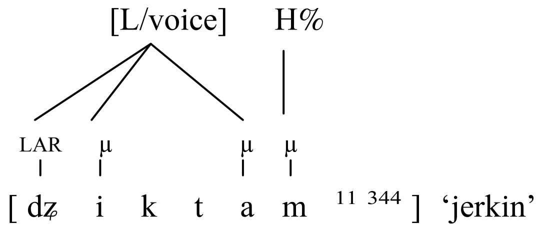
6.6.2.3 Non-spontaneous voicing

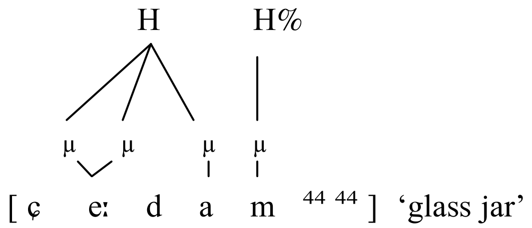
If word initial voicing is considered to be non-spontaneous, then it can be accounted for by the low tone attaching to the laryngeal node of the consonant. In this respect it can be modelled by either Halle & Stevens (1971) or Bradshaw (1999). The following diagrams in this section show different situations in Walungge where there is non-spontaneous voicing. In these diagrams the low tone is being referred to as [L/voice] and high tone as H, as in Bradshaw (1999) but the diagrams themselves are the same regardless of whether Bradshaw's model or Halle & Steven's model is used. In all the diagrams, because the boundary tone is a different phonological entity from the underlying tone, it will continue to be described as H% or L%[22]. The section following this considers situations where there is a difference in the application of the two models.

This first diagram shows the spreading of low tone to the laryngeal node of a word initial plosive, illustrating the non-spontaneous nature of the voicing:



In the case of word medial plosives/affricates the laryngeal node delinks because the environment is not word initial, [L/voice] or H spreads from TBU to TBU, and any word medial voicing is spontaneous:





---

[22] As the purpose of the diagrams below is to illustrate the interaction of tone with consonants, the diagrams show the surface tone patterns without the tonal foot structure, and the reader is referred to section 5.3 for this structure.

The first diagram above shows the spreading of [L/voice] to the second syllable. The intervening plosives remain voiceless because of their environment. Similarly the second diagram shows the spreading of H with the plosive surfacing as voiced because of the intervocalic environment[23].

6.6.2.4 The case of high tone

The main difference between Halle & Stevens (1971) and Bradshaw (1999) which is relevant to Walungge is the nature of high tone. If high tone is simply H which does not interact with consonants, a high tone word could be illustrated as:

H   H   L%

LAR   μ   LAR   μ

/ k   u /   →   [k   u ^{52}]   ‘statue’

The initial plosive is voiceless simply because that is the default realisation of a plosive which does not have any laryngeal features specified.

If, however, high tone is the opposite of low tone and is the same as low tone in its interaction with consonants, e.g. [-slack] for high tone and [+slack] low tone, then a high tone word would be illustrated as:

[-slack]   [-slack]   L%

LAR   μ   LAR   μ

/ k   u /   →   [k   u ^{52}]   ‘statue’

Because the default state of plosives and affricates is voiceless, the segment /k/ would surface as [k] even without the attachment of [-slack]. This is an argument in support of high tone being nothing other than a phonological entity H which only attaches to TBUs, and does not interact with the segments.

---

[23] It is assumed that any segment for which there are no laryngeal features does not have a laryngeal node in the structure (Lombardi (1991)), hence the lack of laryngeal node for word initial [ç] in the above diagram.

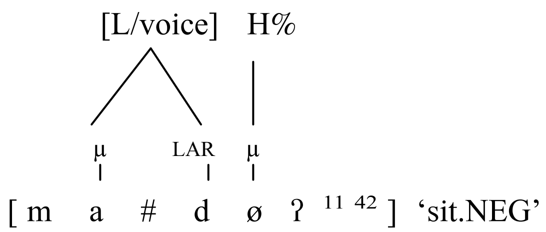
Consider now the case of negative verbs, e.g.

[madøʔ $^{11\ 42}$] ‘sit.NEG’

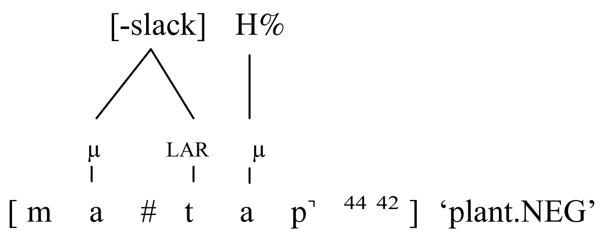
[matap˺ $^{44\ 42}$] ‘plant.NEG’

Data has already been presented above to show that within a pword intervocalic plosives gain spontaneous voicing. However, in the case of negative verbs there is a pword boundary between the negative morpheme and the root. The above data shows that spontaneous voicing does not occur across this boundary, just as spontaneous voicing does not occur in any other word initial environment.

The following diagram shows the attachment of [L/voice][24]. Because of the pword boundary between the negative morpheme and the root, the laryngeal node does not delink. [L/voice] attaches to this laryngeal node, meaning that the consonant surfaces as voiced because of the attachment of [L/voice]:



This next diagram shows a verb with high tone and a root initial plosive. If high tone is [-slack], then [-slack] attaches to both the first mora and the laryngeal node of the root initial plosive:



If, instead, high tone is simply H, it will not attach to the root initial plosive:

[24] It could equally be [+slack]; the diagram would be the same.

H H%

μ LAR μ

[m a # t a p˺ $^{44\ 42}$] ‘plant.NEG’

In this case, also, the voicing or voicelessness of the plosive can be accounted for by the spreading of [L/voice] and without the need for any high tone feature to spread to the laryngeal node. This is further evidence that high and low tone are not opposite values of the same binary feature, but that it is L rather than H which interacts with plosives.

However, whilst the above data suggests that spontaneous voicing does not occur across a pword boundary, consider now the following examples which are of a noun with the possessive suffix. In 5.4.4 above this construction was analysed as a recursive pword in a similar manner to the negative verbs shown above: the stem and affix together form the one pword and yet there is a pword boundary between stem and affix. In the following two examples the intervocalic plosive at a pword boundary is not at the start of a pword so does not have a laryngeal node. The voicing, therefore, must be spontaneous:

[L/v] H%

μ < μ >

[l u g # i $^{24\ 21}$] ‘sheep.POSS’

H L%

μ < μ >

[j a g # i $^{53\ 21}$] ‘yak.POSS’

The voicing can be attributed to a combination of factors. Although the voicing is not the result of the pitch immediately beforehand as the examples show both high and low pitch beforehand, it could be at least in part due to the low pitch which follows the plosive. Further, if voicing is spontaneous then the lack of voicing in a word such as [makoː $^{44\ 41}$] ‘boil.NEG’ and the voicing in a word such as [lugi $^{24\ 21}$] ‘sheep.POSS’ can be attributed to the difference in phonetic environment rather than the phonological presence of a pword boundary. In particular, it is important to note that the verb root takes stress, e.g. [maˈkoː $^{44\ 41}$]. Because the plosive is in the onset of a stressed syllable there are phonetic differences between its articulation and the articulation of the intervocalic plosive in [ˈlugi $^{24\ 21}$] ‘sheep.POSS’, for example. One difference is the duration of the plosive, with the duration being longer for the plosive in stressed position. As has already been discussed in the analysis of 6.2.2.2 and 6.5.3 above, the degree to which spontaneous voicing occurs in an intervocalic situation is dependent upon the duration. The longer duration in the stressed position could mean that spontaneous voicing does not occur, whereas the shorter duration in the unstressed position enables voicing to occur.

6.7 Conclusion of chapter

One of the main considerations of this chapter has been whether or not the voicing of plosives and affricates is phonologically contrastive. By looking at word medial plosives from both an acoustic perspective and a phonological perspective, it was concluded that voicing is not phonologically contrastive word medically. The acoustic analysis of medial plosives considered the duration of the plosives, the proportion of the closure which is voiced, and the possible effect of voicing on pitch. The phonological analysis considered the voicing of plosives at morpheme boundaries.

Having considered word medial plosives, attention was then turned to word initial obstruents, with the conclusion that voicing is not contrastive. VOT of plosives was considered, as was the duration of affricates and fricatives. From this acoustic analysis the conclusion was that is the tone rather than the pitch which has an effect on word initial obstruents.

The next consideration of this chapter was laryngeal features. The only laryngeal feature that is analysed as contrastive for Walungge segments is the feature [spread]. This accounts for the aspiration distinction in plosives and affricates. It also accounts for the voicing distinction in liquids. In all environments other than word initially, the laryngeal node delinks.

Low tone and high tone have been analysed as different types of phonological entity, as in Bradshaw (1999). High tone is simply H, which attaches to TBUs but does not interact with consonants in any way. Low tone, however, can be analysed as a multi-planar feature [L/voice]. It attaches to tone bearing units as low tone. It also attaches to the laryngeal node of segments where it is realised as voicing.

The voicing of plosives and affricates can be accounted for in two different ways. Non word initial plosives and affricates have spontaneous voicing when they are between sonorant segments (either sonorant consonants or vowels). However, the word initial voicing of plosives and affricates is best analysed as the spreading of [L/voice] to the laryngeal node of the consonant. Analysing the voicing on plosives in this way accounts for not only the voicing patterns in monomorphemic words, but also the voicing across morpheme boundaries.

7 Perception

7.1 Introduction

This chapter explores two aspects of tone in Walungge which are of particular interest because the phonetic distinction between words with opposing tones is not purely a difference in pitch. The first is the interaction between tone and obstruent voicing, and the second is the interaction between tone and vowel length.

7.1.1 Tone and voicing

Chapter 6 above discusses data such as the following minimal pair, which shows a surface correlation between tone and the voicing of the initial plosive:

WT	*Walungge*	*gloss*
<rta>	*tá*	'horse'
<mda'>	*dà*	'bow (Walungge)'; 'arrow (WT)'

Chapter 6 presents the arguments for and against an analysis of voicing in Walungge as phonologically contrastive, and concludes that voicing is not in itself phonologically contrastive but is determined by the tone. However, when a Walungge speaker hears a word, both the voicing (or lack of it) and the pitch are available as acoustic cues to the identification of the word. One of the issues that this chapter considers is the relative salience of these two cues.

Before considering Walungge speakers' perception of pitch versus voicing, it is helpful to explore the way in which the tone contrast in Walungge has developed from the WT voicing contrast.

Section 2.3 gives a general overview of tonogenesis, particularly the development of a tonal contrast from a voicing contrast, a process which has occurred in many languages worldwide. For many Tibetan languages, including Walungge, a tonal contrast has also arisen from the loss of word initial consonant + sonorant clusters, e.g.

WT	*Walungge*	*gloss*
<rnga>	*ŋá*	‘drum’
<nga>	*ŋà*	‘1SG’

Lhasa Tibetan also has the same loss of consonant clusters. Mazaudon (1977) claims that in Tibetan languages such as Lhasa Tibetan is it the loss of consonant + sonorant clusters rather than the loss of the obstruent voicing contrast which was the origin of the tone distinction. She proposes a scenario where “the reduction of initial clusters to simple consonants turned the prefixed nasals and resonants into a fortis series, probably voiceless. This fortis series merged with the plain series (WT plain unprefixed) triggering a compensatory tonal split where the fortis initials produced high-tone syllables, and the plain produced low-tone syllables. ... The stop initials went along with the split started by the continuants.”

In the Walungge context, it is possible that the tone contrast in Walungge arose in the way that Mazaudon proposes for Lhasa Tibetan. However, consider the following data from Daofu Tibetan (Huang (1997)):

WT	*Daofu*	*Gloss*
<snabs>	s̥nap^{53}	‘nasal mucus’
<nas>	neː^{24}	‘barley’

Daofu Tibetan does not have contrastive tone, but has predictable pitch melodies. For consonant + sonorant initial clusters the Daofu pitch melody is falling, which corresponds to the Walungge pitch melody [52] on high tone monomoraic words (which is a realisation of H underlying tone followed by L% boundary tone). For plain sonorants the Daofu pitch melody is rising, which corresponds to the Walungge pitch melodies [14] and [24] on low tone words (a realisation of L underlying tone followed by H% boundary tone). Thus it is possible that at an earlier stage of its development Walungge was similar to Daofu, with lower initial pitch associated with a simple sonorant, and higher initial pitch associated with a consonant + sonorant cluster. Over time these pitch perturbations became more pronounced, until the point of reanalysis where the burden of contrast shifted to the pitch melody from the consonant cluster.

The clusters were lost, leaving simple sonorant onsets plus contrastive tone. As Mazaudon proposes for Lhasa Tibetan, it is possible that this happened before the loss of voicing contrast.

Turning now to obstruents, it is possible that originally the voiced and voiceless obstruents of WT caused a slight raising or lowering of the pitch of the start of the following vowel, in much the same way as is described for English (Hombert, Ohala et al. (1979)). This mean that a word with a voiceless initial obstruent would have had a pitch melody which was slightly falling, and a word with a voiced obstruent would have had a pitch melody which was slightly rising. This is a similar scenario to what has been described above for sonorants. When contrastive tone was established for words with sonorant initials, the obstruents followed.

For words with initial fricatives, the shift from a voicing contrast to a tonal contrast is complete. The voicing has been lost phonetically as well as phonologically, and any initial consonant clusters have been lost without any knock-on effect:

WT	*Walungge*	*gloss*
<bzhi>	*ɕì*	‘four’
<zho>	*ɕò*	‘yoghurt’
<sha>	*ɕá*	‘meat’

In the case of plosives and affricates, not only has there been a shift from a voicing contrast to a tonal contrast, but the loss of word initial consonant clusters has also had an effect. Where there was no initial consonant cluster, a voiced plosive or affricate has been reanalysed as an aspirated consonant, while those with pre-scripted consonants have retained their voicing phonetically.

WT	*Walungge*	*gloss*
<ja>	*tɕʰà*	‘tea’
<'ja'>	*dʑà*	‘rainbow’
<dang>	*tʰà*	‘and’
<mda'>	*dà*	‘bow’

However, for voiceless plosives/affricates with pre-scripted consonants, the pre-scripted consonants have been lost without any further effect. The following shows voiceless plosives and affricates with and without pre-scripts and in all cases the resulting obstruent is voiceless with high tone.

WT	*Walungge*	*gloss*
<rta>	*tá*	'horse'
<mtsho>	*tsʰó*	'lake'
<chu>	*tɕʰú*	'water'

The diachronic reanalysis from word initial voicing contrast to tone contrast can be illustrated as follows using the feature [L/voice] (c.f. Bradshaw (1999)) as a possible feature for both tone and voicing (see section 6.6.1 above).

WT	*(1)*	*(2)*	*(3)*	*'gloss'*
[L/v] <md a ngs>	[L/v] <md a ngs>	[L/v] *d à ŋ*		'yesterday'
[L/v] <'j a'>	[L/v] <'j a'>	[L/v] *dʑ à*		'rainbow'
[L/v] <gz i g>	[L/v] <gz i g>	[L/v] *z ì k*	[L/v] *s ì k*	'snow leopard'

Taking WT as the historical form of the Walungge words, in each case the WT voiced obstruent has the feature [L/voice]. Column (1) shows the feature [L/voice] originating from the consonant and spreading to the vowel, which would have initially given rise to lowered pitch at the start of the vowel, and thus a rising pitch melody. This pitch melody has become "phonologised" as low tone followed by a boundary tone. Column (2) shows the feature [L/voice] underlyingly independent from the segments. It attaches to the TBU as low tone (the subsequent boundary tone is not shown in the above diagrams) which is realised on the vowel primarily as low pitch, and also spreads to the consonant as voicing. This is the synchronic situation for plosives and affricates. Column (3) shows the synchronic situation for fricatives. The feature [L/voice] no longer attaches to the initial consonant; the initial consonant is unaffected by the tone.

This chapter is concerned with words with initial plosives/affricates, because of the word initial voicing of these sounds. For a mother tongue speaker distinguishing between a pair of words such as *tá* 'horse' and *dà* 'bow', has the pitch melody become so salient that the voicing is insignificant, or does the voicing still play a role in the identification of words?

7.1.2 Tone and length

The second issue of interest is the relationship between the surface tone pattern and vowel length. This is particularly relevant for monomorphemic monosyllables which have surface tone patterns H-H% and H-L% (where H% and L% are boundary tones), realised as high-level and high-falling pitch.

In polysyllabic words, length is phonologically contrastive, for example:

gò:mo	'evening'
gòmo	'money'

In monosyllabic words, the high-level and high-falling pitch melodies are contrastive in certain situations, for example when high-falling pitch has originated from a word with underlying tone H and an affix which is extra-tonal:

H H%		H L%	
| |		|/	
μ μ		μ ‹μ›	
[ti: ^{44}]	'foal'	[ti: ^{51}]	'foal.POSS'

However, in monosyllabic monomorphemic nouns the surface tone patterns depend upon the vowel length, with tone pattern H-L% occurring with short vowels and tone pattern H-H% occurring with long vowels, for example:

H L%		H H%	
çá	'meat'	*çá:*	'deer'

The vowel length has arisen from the diachronic deletion of segments, either syllable codas or second syllables, for example:

WT	*Walungge*	*gloss*
<dngul>	*ŋú:*	'silver'
<mar>	*mà:*	'butter'
<kha ba>	*$k^há$:*	'snow'
<gzugs po>	*sù:*	'body'

In order to consider how the H-H% versus H-L% surface contrast has arisen, consider the following data from Dolpo (own data), compared with Walungge:

WT	*Dolpo*	*Walungge*	*gloss*
<dngul>	*ŋúl*	*ŋú:*	'silver'
<bkol>	*kól*	*kó:*	'boil'
<bsnyal>	*ɲál*	*ɲá:*	'put to sleep'

Given the existence of closely related languages such as Dolpo, which has lost initial consonant clusters but retained certain word final consonants which languages like Walungge have lost, it is reasonable to propose that in Walungge the loss of initial consonant clusters arose before the loss of word final segments. Now consider a word such as WT <dngul> 'silver' (Walungge *ŋú:*). If the outline of Walungge tonogenesis proposed above is correct, then <dngul> would have had slightly raised pitch at the start of the vowel, giving a falling pitch melody. When a tonal contrast arose as a result of the loss of initial consonant clusters, the word might have initially become *ŋúl* with the high tone realised as a falling pitch melody. The word final consonant was then lost, giving rise to compensatory vowel lengthening. One plausible scenario is that in order to enhance the length distinction, the falling pitch on short vowels became more pronounced, and the pitch on long vowels became a sustained high. Over time, this difference between falling pitch and sustained high pitch became phonologised, to give H-L% and H-H% tone patterns. However, the vowel length distinction has been retained.

Because pairs of words such as *ɕá* 'meat' and *ɕáː* 'deer' have both the surface tone (realised as pitch) and the quantity (realised as duration) acting together to distinguish between the words, perceptual experiments may be used to try to determine whether one of these two factors is more salient for word identification.

7.2 Experimental design

The sections that follow describe perception experiments that were carried out with mother tongue speakers of Walungge. The experiments investigate the relative salience between a) pitch and voicing and b) pitch and duration.

A series of binary choice word perception tests was devised. Subjects were played recordings and asked to identify the word they heard by choosing one of two pictures. For investigating the salience of pitch versus voicing, tonal minimal pairs were chosen, and recordings were made of a mother tongue female speaker saying the words. Using these recordings, F0 and VOT were artificially manipulated to create sets of stimuli with different combinations of F0 and VOT (see below for details). The vowel duration and F0 of minimal pairs with contrasting vowel length were similarly manipulated. All acoustic manipulations were carried out using the computer program Praat v. 5.1.07. A number of different manipulated values of F0, VOT and duration in the set of stimuli were chosen in order to have enough combinations for a thorough investigation, balanced against the need to prevent producing a set of stimuli so large that the experiment would take an unreasonably long time for subjects to complete. Each set of stimuli was presented to subjects in randomised order, and all stimuli were presented twice in the experiment.

In most of the experiments (see below) the words were placed in a frame sentence, so that subjects had a reference point for the pitch of each word, rather than the words themselves acting as reference points for the pitch of the following word, since a word with the same F0 might be perceived as having higher or lower pitch depending on what word it followed. The frame sentence was *nò ... dè* 'this is a .....'. The words *nò* and *dè* were the same recording in every case.

The sections 7.3 and 7.4 below describe the various experiments in detail, giving more information about the acoustic manipulations that were carried out for each set of stimuli, along with the results of the experiment. All experiments were carried out at the same time, giving each subject a total of 284 items to listen and respond to, typically completed within a half-hour session.

The experiments were carried out in the Walungge speaking area, and all subjects participating in the experiments were mother tongue Walungge speakers. All had been born and raised in the Walungge area, and although some had lived away from the area for education, work, etc, all were currently living in the area. A total of 40 subjects participated, and 38 of these gave usable results. The other 2 appeared to be listening to the first word in each group of words then alternating their answers regardless of what they heard, so their data were excluded from the analysis. Ages of subjects ranged from 13 to 61 years old, with the majority being between 20 and 40 years old. Approximately half were male and half were female. The subjects' level of education ranged from those who had never been to school to one person who had a BA degree. Most of the subjects had travelled outside the Walungge speaking area. For some, the extent of their travel was to Taplejung district centre. Others regularly made trips to Kathmandu, India, or Tibet. All spoke Nepali to some degree, and some spoke Lhasa Tibetan.

7.3 Tone and voicing experiments

When a Walungge speaker hears a pair of words, for example

tá [52]	'horse'
dà [24]	'bow'

there are a number of acoustic cues available to enable the two words to be perceived as distinct, namely whether the starting pitch is high or low, whether the overall pitch melody rises or falls, whether the initial consonant is voiced or voiceless, and how phonation changes through the vowel (the phonation correlates with the pitch, see section 4.4).

A series of perceptual experiments was designed to test the perceptual salience of each of the potential cues. Because pitch melodies [52] and [24] are arguably the pair of pitch melodies with the greatest difference, it was decided to look first at pairs of bimoraic words, where the pitch melody of a word with underlying tone H is level rather than falling. Pairs of words starting with fricatives or sonorants were chosen, so that the cues of starting pitch, overall pitch melody and phonation could be considered before additionally considering initial consonant voicing. Pairs of words starting with plosives were then added to the experiment. Finally words with [52] versus [24] pitch melodies were added to the experiment.

7.3.1 Pitch melodies for words starting with sonorants/fricatives

7.3.1.1 Experimental design

Monosyllabic monomorphemic heavy syllable words and disyllabic monomorphemic words have a level high pitch melody for underlying tone H and a rising pitch melody for underlying tone L.

In order to investigate the relative salience of the pitch versus the initial consonant voicing, it is first necessary to consider the pitch on its own without the interaction with voicing. In distinguishing between words which only differ in their underlying tone, e.g.

sáŋ [44] ‘pine’
sàŋ [14] ‘saucepan’

are Walungge speakers responding to the difference in high versus low initial pitch, or are they responding to the difference between a level versus a rising pitch melody?

In order to investigate this, a set of stimuli with three distinct values of F0 was prepared. The lowest value chosen was 185Hz, which is the approximate average starting value of F0 for words with low tone for the speaker used for the recordings. From the acoustic analysis described in section 4.3 above, and adjusting for the speaker in question, the approximate average rise of the rising pitch melody is 3.5 semitones, and it rises to approximately the same height as the average pitch of words with high level pitch melody. Thus two more values of F0 were chosen: 1.75 semitones above 185Hz (＝205Hz) and 3.5 semitones above 185Hz (＝226Hz)

The set of stimuli contrasted six rising or level pitch melodies based on these three values of F0:

Table 7-1: manipulated pitch melodies

shape of pitch melody						
pitch melody “name”	3-3	2-2	1-1	1-2	2-3	1-3
starting F0 (Hz)	226	205	185	185	205	185
ending F0 (Hz)	226	205	185	205	226	226

Starting with a recording of the word *sáŋ* ‘pine’, the pitch was manipulated to create 6 recordings, each with one of the above 6 pitch melodies.

It can be seen from the acoustic description in section 4.5, that words with low tone have a slightly longer average duration than words with high tone. It can also be seen that on average the breathiness across the word rises and falls as the pitch melody rises and falls. Whilst these acoustic differences are only very slight, they are available as cues for word identification. To investigate the effect of phonation on word identification, a recording of the word *sàŋ* ‘saucepan’ was also manipulated to create 6 recordings with each of the pitch melodies. All 12 recordings were then manipulated to have identical durations, set to their average duration. It was not easily possible to manipulate the breathiness across the word, so a pair of stimuli was produced for each pitch melody, identical except that one had originated from *sáŋ* and the other from *sàŋ*, preserving the slightly different phonation type of each. If subjects were responding in any way to the breathiness, this would be evident in the results of the experiment.

Because the level and rising pitch melodies occur on both monosyllabic and disyllabic words, recordings of the words *ɲìma* 'sun' and *ɲíma* 'cereal head' were manipulated in a similar way. The difference between monosyllabic and disyllabic words is that for monosyllabic words with underlying low tone the pitch melody is a rise over one syllable and for disyllabic words with low tone the pitch melody is low pitch on the first syllable, a rise through the medial consonant, and high pitch on the second syllable.

Whereas the monosyllabic words were placed in frames when presented to subjects, the disyllabic words were played to people in isolation. In isolation, because of the lack of a reference point for the pitch, it is possible that greater salience is attached to the contour of the pitch melody – whether it rises or is level.

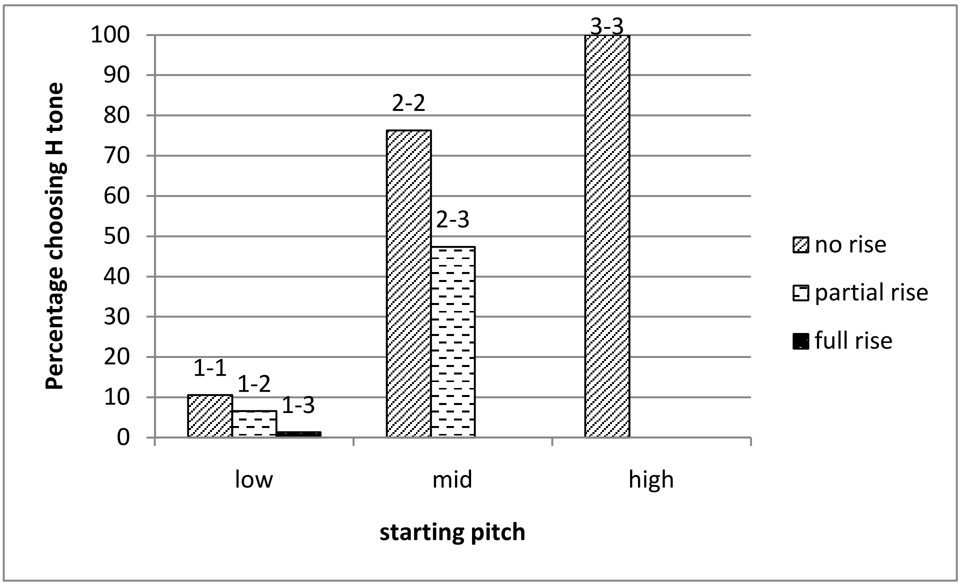
7.3.1.2 Results for monosyllabic heavy syllables – words in frames

The following chart and table gives for each (manipulated) pitch melody the percentage of responses which identified stimuli carrying that pitch melody as the low tone word and the percentage of responses which identified stimuli carrying that pitch melody as the high tone word. As explained above, for the one word [saŋ] each pitch melody appeared twice in the experiment and there were 38 usable subjects, giving a total of 76 responses for each pitch melody.

Table 7-2: percentages choosing H or L; monosyllabic heavy syllable stimuli

pitch melody	1-1	1-2	1-3	2-2	2-3	3-3
% responses choosing L	89	93	99	24	53	0
% responses choosing H	11	7	1	76	47	100
number tokens	76	76	76	76	76	76

Figure 7-1: percentage choosing H tone; stimuli in frames



The high level pitch melody 3-3 and the low rising pitch melody 1-3 are the unaltered pitch melodies, and as expected virtually all subjects identified stimuli with pitch melody 3-3 as the H tone word *sáŋ* 'pine', and the stimuli with pitch melody 1-3 as the L tone word *sàŋ* 'saucepan'.

When the pitch is low and level (1-1), subjects are presented with apparently conflicting acoustic cues. Whilst the low starting pitch is a cue that the item is an L tone word, a level pitch melody is normally associated with an H tone word. The responses to pitch melody 1-1 show that whilst the vast majority of subjects associated the low level melody with an L tone word, approximately 10% of responses identified the low level melody with an H tone word. If people have to choose between the starting pitch and the shape of the melody, the starting pitch is by far the more salient cue. However, the shape of the melody still has a slight effect upon perception.

A mid starting pitch is a potentially ambiguous cue as to the tone of the word. However, this ambiguous starting pitch was presented to subjects in a frame preceded by an L tone word with unambiguous rising pitch melody. The starting pitch for the ambiguous item is lower than the high pitch at the end of the word before, which could be a cue for low tone. However, the starting pitch for the ambiguous item is higher than the initial low pitch of the preceding word, which could be a cue for high tone. It is this situation where the shape of the pitch melody becomes a more salient acoustic cue. When the pitch melody was mid and level (2-2), the majority of responses (76%) associated this with an H tone word. However when the pitch melody was mid and rising (2-3), just over half the responses (53%) associated this with an L tone word.

For each starting pitch, subjects respond differently to the overall contour of the pitch melody, as described above. McNemar tests were applied in order to investigate the significance of these responses. This test was chosen because the responses for each pitch melody and each speaker can be paired, but the responses are neither normally distributed (T-test) nor symmetrical (Wilcoxon test). For each pair of pitch melodies with the same starting pitch, the responses for each subject and each word were paired; the null hypothesis is that there is no significant change in response according to pitch melody; the significance level is $p < 0.05$. The following are the *p* results.

Table 7-3: significance statistics for each pair of pitch melodies

	N	p	$p < 0.05$
1-1 vs. 1-2	76	0.549	
1-2 vs. 1-3	76	0.219	
1-1 vs. 1-3	76	0.039	•
2-2 vs. 2-3	76	<0.0005	•

When the starting pitch is low, the slight differences in response according to pitch melody are not significant when there is only a small change in pitch melody (1-1 to 1-2 or 1-2 to 1-3). Comparing 1-1 with 1-3, although the actual change in percentage is not large, with 89% associating 1-1 with low tone compared with 99% associating 1-3 with low tone, the change is significant. In other words, with over 95% certainty it is the overall contour of the pitch melody, rather than chance, which is causing subjects to respond differently to different stimuli. The statistical analysis also confirms that the difference in responses between 2-2 and 2-3 is significant; subjects are responding to the overall contour of the pitch melody and not just the starting pitch.

In summary, providing the starting pitch is not ambiguous (i.e. providing it is either high or low), when presented with apparently contradictory cues from the starting pitch and the shape of the pitch melody the starting pitch is of far greater salience than the shape of the melody. However when the starting pitch is ambiguous (mid), the contour of the melody becomes a salient cue in determining the tone of the word.

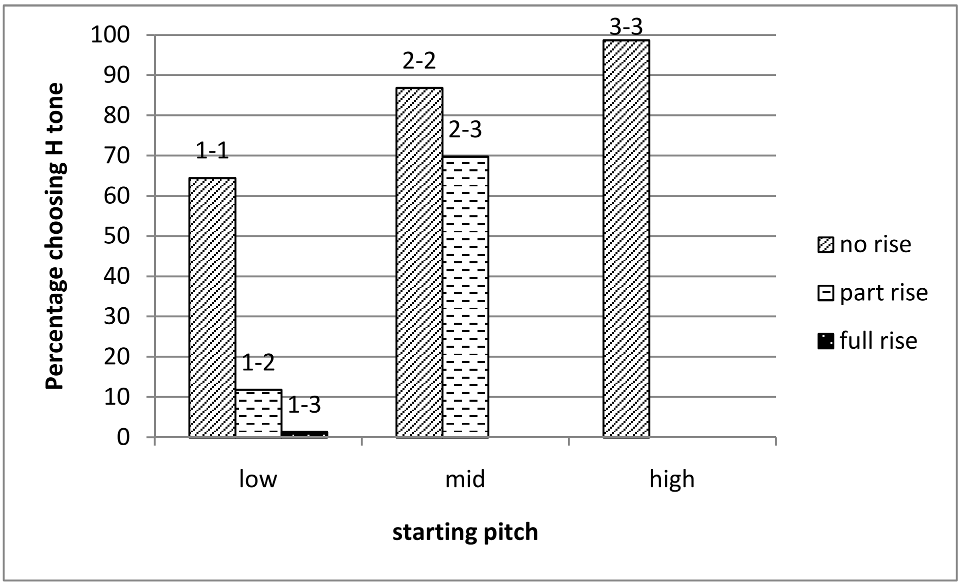
7.3.1.3 Results for disyllabic H versus L; stimuli in isolation

The following graph and table gives for each pitch melody the percentages of responses matching the stimuli, which are disyllabic and in isolation, with L and H tone words.

Table 7-4: percentages of responses choosing H or L; stimuli in isolation.

pitch melody	1-1	1-2	1-3	2-2	2-3	3-3
% responses choosing L	36	88	99	13	30	1
% responses choosing H	64	12	1	87	70	99
number tokens	76	76	76	76	76	76

Figure 7-2: percentage choosing H tone; disyllabic stimuli in isolation



The disyllabic stimuli were presented in isolation rather than in the frame *nò ... dé* "this is a ...". This meant that subjects did not have the additional acoustic cue of the pitch melody on the words of the frame as a reference pitch. Despite this, nearly 100% of subjects matched a 1-3 pitch melody with an L tone word and a 3-3 pitch melody with an H tone word.

When the starting pitch is low, the apparent effect on subjects' responses of the interaction between starting pitch and shape of pitch melody for the disyllabic isolation stimuli is different from the monosyllabic framed stimuli. In frames, because of the reference pitch beforehand, stimuli with low starting pitch were generally identified as L tone words regardless of whether the pitch rose or remained level. However when there was no reference pitch beforehand, as is the case for disyllabic isolation stimuli, the shape of the pitch melody became more salient. This can be seen in the above graph for disyllabic stimuli. When the pitch is low and level (1-1) approximately two-thirds of responses (64%) identify this as an H tone word. This is very different from the 11% which identified the same pitch melody as an H tone word when the word was in a frame. However when the pitch is rising, regardless of whether or not the word is in a frame, the majority of responses identified this as an L tone word.

In the case of stimuli presented in isolation with mid starting pitch, regardless of whether the pitch melody is level 2-2 or slightly rising 2-3, the majority of responses identified the stimuli as H tone words. This is different from monosyllabic framed stimuli, where the majority of responses for 2-2 were H tone, but not for 2-3. The possible reason why the rising 2-3 was associated with H tone for disyllabic stimuli is because of 2nd syllable stress (see section 5.2), which causes the pitch on the second syllable to be higher than the pitch on the first syllable even if the tone across both syllables is underlyingly H.

The responses for pitch melodies with the same starting pitch and different amount of rise were paired, and McNemar tests were carried out, with results below. Also included in the McNemar tests is the pairing of the pitch melodies 1-1, 2-2, and 3-3 because of the closeness of percentage responses for these.

Table 7-5: significance statistics for each pair of pitch melodies

	N	p	$p < 0.05$
1-1 vs. 1-2	76	<0.0005	•
1-2 vs. 1-3	76	0.021	•
2-2 vs. 2-3	76	0.021	•
1-1 vs. 2-2	76	0.003	•
2-2 vs. 3-3	76	0.012	•

All the differences in responses for the pairs of different pitch melodies are statistically significant. Again this is different from the monosyllabic framed stimuli where the differences between 1-1 and 1-2 and between 1-2 and 1-3 were not significant, and is a further indication of the salience of the shape of the pitch melody when words are in isolation.

In summary, when there are no additional words giving a reference pitch, the pitch melody across a word is of greater salience than the starting pitch of the word. The starting pitch of the word still has an effect (which is statistically significant), in that the higher the starting pitch the more likely it is for a listener to perceive that word as an H tone word. However, if it is a straight choice between the starting pitch and the overall pitch melody, then the shape of the pitch melody is more salient than the starting pitch.

7.3.1.4 Effect of phonation

In the above experiments (monosyllabic in frames and disyllabic in isolation) each pitch melody appeared twice. One occurrence was when the word that was manipulated to produce the pitch melody was originally an L word; the other occurrence was when the original word was an H word. Because the durations of the stimuli were manipulated to be identical, for each pitch melody the two stimuli were identical except for the phonation, which correlated with the original pitch melody of the word. Those stimuli which came from an L word had phonation which started breathier. Those stimuli which came from an H word had phonation which was less breathy.

In order to investigate whether the phonation of the original word was having any effect upon the perception of the pitch of the stimuli, McNemar tests were carried out by pairing responses for stimuli with differing phonation but identical pitch melodies. The following table gives the results of this test.

Table 7-6: significance statistics for pairs of stimuli with differing phonation

word environment	N	p	$p < 0.05$
frame	228	0.877	
isolation	228	<0.0005	•

Figure 7-3: Bar chart of percentages for phonation; monosyllabic frames

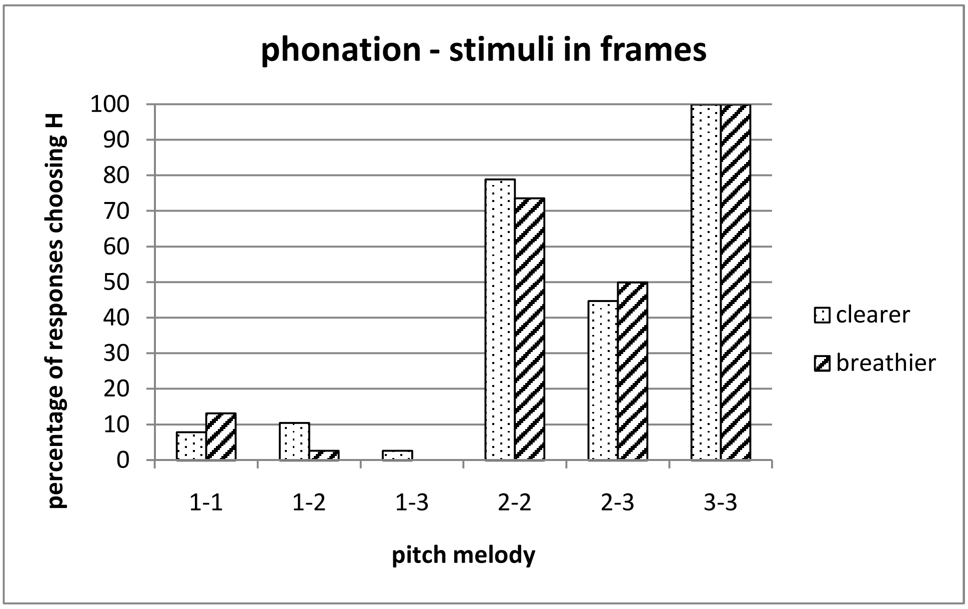
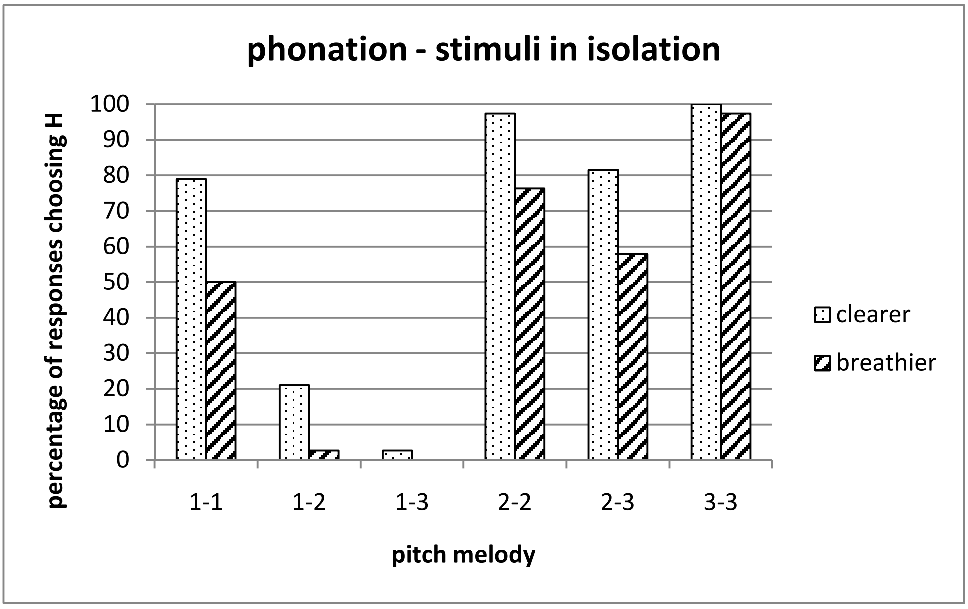


Figure 7-4: Bar chart of percentages for phonation; disyllabic isolation



For stimuli in frames, phonation appears to play no role in word identification. There is no significant difference in the responses for those stimuli which had clearer phonation from those which had breathier phonation. However for stimuli in isolation phonation does play a role in word identification. As seen in the above bar chart, for every pitch melody the percentage of responses identifying the stimuli as an H tone word was greater if the stimuli had clearer phonation than if they had breathier phonation. Further, from the McNemar test this difference is statistically significant. When the pitch melody is unambiguous (1-3 or 3-3) there is virtually no difference in perception for the different phonation; subjects were responding to the pitch melody not the phonation. However as the pitch melody becomes ambiguous the salience of the phonation increases. For example, in the case of a 1-1 pitch melody with clearer phonation almost 80% of responses identified this as an H tone word. However if the phonation of a 1-1 pitch melody was breathier (on the first vowel), only 50% of responses matched this with an H tone word.

For words in isolation, the lack of a reference pitch from the surrounding words means that not only does the shape of the pitch melody across the word become the salient cue, but phonation also has a bearing upon how the word is perceived.

7.3.2 Obstruent voicing

7.3.2.1 Experimental design

Using the pair of words *cáː* 'wooden spoon' and *ɟàː* 'window', stimuli were created by manipulating the pitch to correspond to the same pitch melodies as used for fricatives, in Table 7-1 above. Durations were also manipulated in order for all stimuli to have identical durations. Following this, for each item the VOT was then manipulated to be one of 3 values: 20ms, zero, and -20ms. (These values were chosen according to the acoustic analysis done in section 6.3) Because of the way that the VOT manipulations were carried out, it was possible to manipulate a word starting with a voiceless plosive to have either positive or zero VOT, and a word starting with a voiced plosive to have either negative or zero VOT. Thus all stimuli with positive VOT originated from the word *cáː* 'wooden spoon', and all stimuli with negative VOT originated from the word *ɟàː* 'window'. These stimuli were duplicated to give two identical recordings for each pitch melody and VOT. In the case of stimuli with zero VOT, each pitch melody appeared twice: once from the stimuli originating from *cáː* and once from the stimuli originating from *ɟàː*.

The following table gives a summary of the stimuli created from the above manipulations:

number of pitch melodies:	6
number of VOT values:	3
number of stimuli for each pitch melody and VOT:	2
total number of stimuli:	36

The same manipulations were also carried out for the pair of words *túŋ* 'drink' and *dùŋ* 'hit', giving an additional 36 stimuli. Although these are verbs not nouns, this pair of words was chosen because of the need to provide pictures to point to in response to the stimuli. When a verb is phrase final, it receives an additional boundary tone (section 5.5) . However, it is also possible to place the words *túŋ* and *dùŋ* in a non-final position such as *nò túŋ dè* 'this (the word) is "drink"'. In this case it has the same pitch melody as any other non-final content word, and mother tongue speakers gave assurance that a sentence such as the above was acceptable.

The disyllabic pair of words *tóŋba* 'empty' and *dòŋba* 'face' was chosen for stimuli in isolation rather than in frames. The same manipulations as above were carried out. (As an aside it should be noted that the more commonly used word for face is *ŋòloŋ*, but *dòŋba* is an acceptable alternative.)

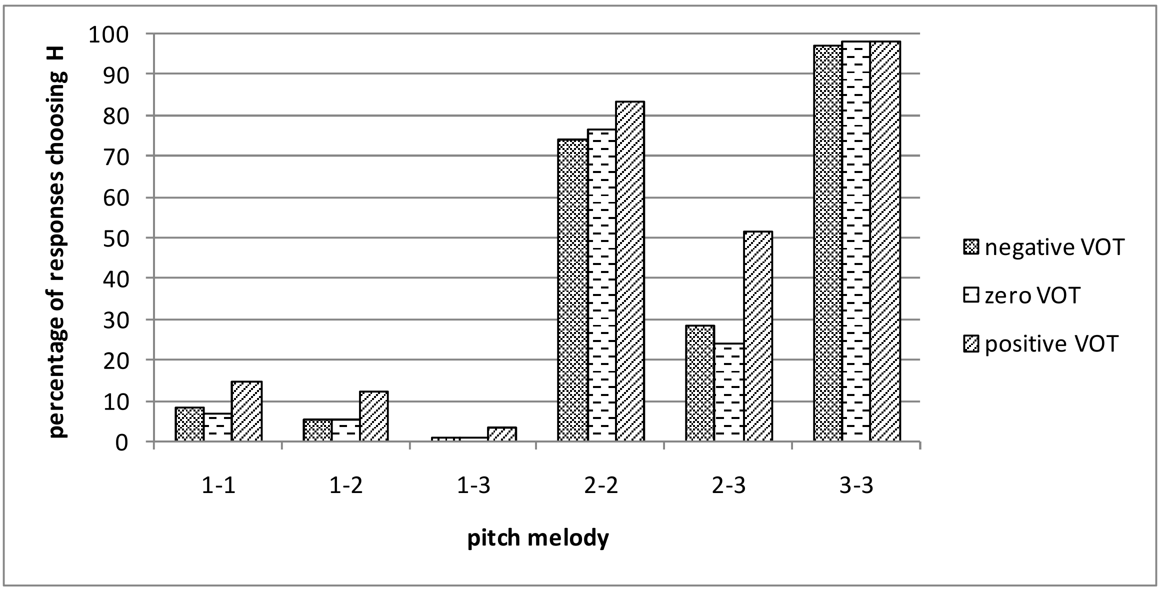
7.3.2.2 Results for stimuli in frames

The responses for all the stimuli in frames were collated and combined according to the (manipulated) pitch melody and the (manipulated) VOT. This gives 152 responses for each combination of VOT and pitch melody. The table and graph below give a summary of the percentages of responses identifying the stimuli as L and as H tone words.

Table 7-7: percentages choosing H or L; stimuli in frames.

pitch melody	VOT	% choosing L	% choosing H	number tokens
1-1	negative	92	8	152
	zero	93	7	152
	positive	86	14	152
1-2	negative	95	5	152
	zero	95	5	152
	positive	88	12	152
1-3	negative	99	1	152
	zero	97	3	152
	positive	97	3	152
2-2	negative	74	26	152
	zero	76	24	152
	positive	83	17	152
2-3	negative	72	28	152
	zero	76	24	152
	positive	49	51	152
3-3	negative	3	97	152
	zero	2	98	152
	positive	2	98	152

Figure 7-5: percentage of responses choosing H tone, for pitch melody and VOT



When the pitch melody is unambiguous (either 3-3 or 1-3) then VOT does not appear to play a role in word identification. For the high pitch melody 3-3 virtually 100% of responses equated this with an H tone word regardless of the VOT of the stimuli. Similarly virtually 100% of responses equated the 1-3 pitch melody with an L tone word regardless of the VOT. Interestingly this was the case regardless of other factors that could have been of influence, e.g. level of schooling in Nepali, amount of travelling outside the language area, etc. Exposure to Nepali, in particular, could have influenced the results due to the fact that Nepali is non-tonal and has a phonological voicing contrast for plosives. The experiment included subjects from a wide range of educational backgrounds, from those who had never been to school to those who had completed Nepali secondary school education and beyond by going to boarding school in a Nepali speaking town. Similarly there were subjects who had never travelled outside the Walungge speaking area and subjects who spent half of every year living and travelling in Nepali speaking areas of Nepal. Despite the large amount of contact with Nepali that some people had, the pitch rather than the voicing was still the salient feature.

For pitch melodies other than 1-3 and 3-3, there are some differences in responses due to the VOT of the stimuli. In general the difference in responses correlates with the difference between positive VOT on the one hand and zero or negative VOT on the other hand. To investigate these differences, for each pitch melody positive and negative VOT responses were paired and McNemar tests were carried out with the following results.

Table 7-8: significance statistics for pairs of positive and negative VOT words

pitch melody	N	p	$p < 0.05$
1-1	152	0.100	
1-2	152	0.064	
1-3	152	0.219	
2-2	152	0.045	(•)
2-3	152	<0.0005	•
3-3	152	0.727	

If the pitch starts low but does not do a complete rise (1-1 or 1-2), there is a very slight change in the responses as the VOT changes. But it has to be stressed that this is very slight, and the vast majority of responses matched a low starting pitch with L tone regardless of the VOT of the word. Further, this slight change in responses is not statistically significant.

In the case of a 2-2 pitch melody, the majority of responses identify this as H tone regardless of the VOT of the word. The 2-2 pitch is ambiguous because the level shape of the melody matches H tone but the height of the pitch is lower than it would normally be for H tone. This ambiguity results in there being a very slight VOT effect, with slightly more responses matching this to H tone when the VOT is positive. Testing this statistically, it is on the border of being statistically significant ($p=0.045$, with significance level $p < 0.05$).

The pitch melody where VOT does have a greater effect is 2-3. This is arguably the most ambiguous pitch melody: the starting pitch is neither high nor low, and the shape of the pitch melody neither rises its full amount nor remains level. This ambiguity has already been seen in the results of the perception test for stimuli which don't start with a plosive, in which case approximately half the responses chose L and half chose H (see section 7.3.1 above). For stimuli starting with plosives, if the initial consonant had positive VOT approximately half the responses identified this as H tone, which is the same as the responses for stimuli which didn't start with a plosive. However, if the VOT was zero or negative, approximately three quarters of responses chose L (76% and 72% respectively).

In summary, the salience of VOT is related to the ambiguity of the pitch melody. If the pitch melody is unambiguous, the VOT does not affect perception of words. Even if the pitch melody has some ambiguity associated with it, any VOT effect is generally only very slight. However, if the pitch melody is very ambiguous (as in the case of 2-3), the VOT does become a salient cue in the perception of words. It should be noted, though, that these results are the combined results for everybody who took part in the experiment. Given it is the ambiguous pitch melodies where VOT has any effect, it is these pitch melodies where exposure to Nepali could be influencing the results. The effect of exposure to Nepali upon perception is discussed in the section that immediately follows, with the result that for the ambiguous pitch melodies VOT only has a significant effect upon perception for people who are sufficiently exposed to Nepali.

7.3.2.3 Exposure to Nepali

One of the factors that is most likely to contribute to people's perception as the VOT changes is exposure to Nepali or another language (such as English or some other Tibetan languages) which has a phonological voicing distinction. People are exposed to Nepali a number of different ways, for example Nepali media, travel, education, proximity to Nepali speaking villages etc. But not all these factors are easy to measure. One factor that is both easy to measure and is most likely to be having an effect is if people have been educated outside of the Walungge speaking area. The schools in the Walungge speaking villages are Nepali medium schools but are only primary schools, and outside of the classroom Walungge is the language of village communication. For secondary school education and beyond, pupils have to board in a village or town outside of the language area; and all are sent to places where they have far greater contact with Nepali speakers (most commonly Lelep, Taplejung or Darjeeling).

Approximately half the people taking part in the experiment had gone to school outside the Walungge speaking area.

In order to consider whether exposure to Nepali has an effect, the results of the above experiment were split according to whether or not people had gone to school outside the Walungge area. The following table and chart gives the percentages of responses identifying the different pitch melodies as H tone words for both positive and negative VOT, and whether or not people have been educated outside the language area. If exposure to Nepali is having an effect on these percentages, the responses for both positive and negative VOT stimuli could be affected, with positive VOT resulting in a higher percentage choosing H, and negative VOT resulting in a lower percentage choosing H. Thus the difference in the percentages between positive VOT and negative VOT is also calculated.

Table 7-9: percentages of responses choosing H for differing education

pitch melody	education outside	% choosing H for +ve VOT	% choosing H for -ve VOT	difference in %	number tokens
1-1	y	17	4	13	72
	n	12	11	1	80
1-2	y	11	5	6	72
	n	13	5	8	80
1-3	y	4	1	3	72
	n	3	0	3	80
2-2	y	92	74	18	72
	n	75	74	1	80
2-3	y	61	28	33	72
	n	43	29	14	80
3-3	y	99	93	6	72
	n	98	100	2	80

Figure 7-6: percentages choosing H; education inside area

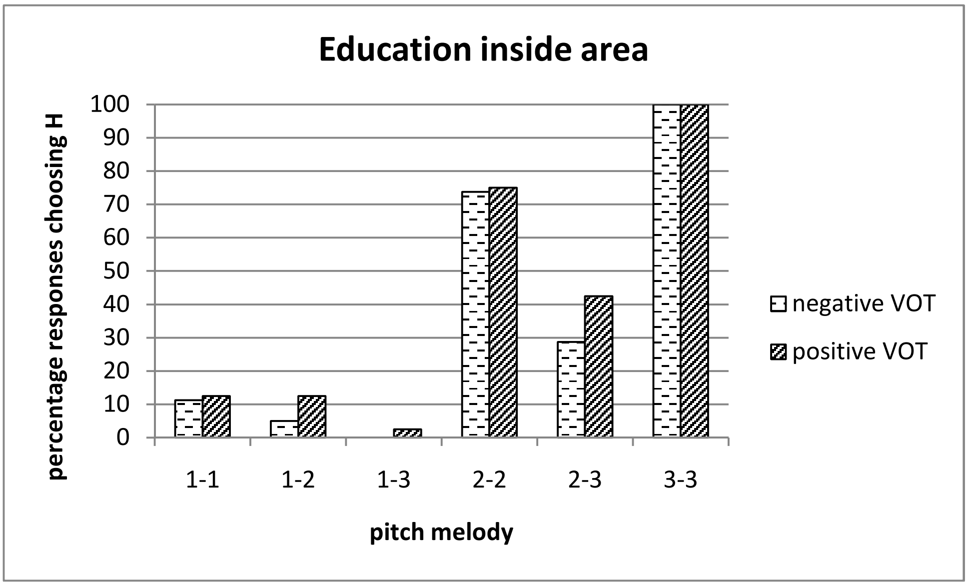
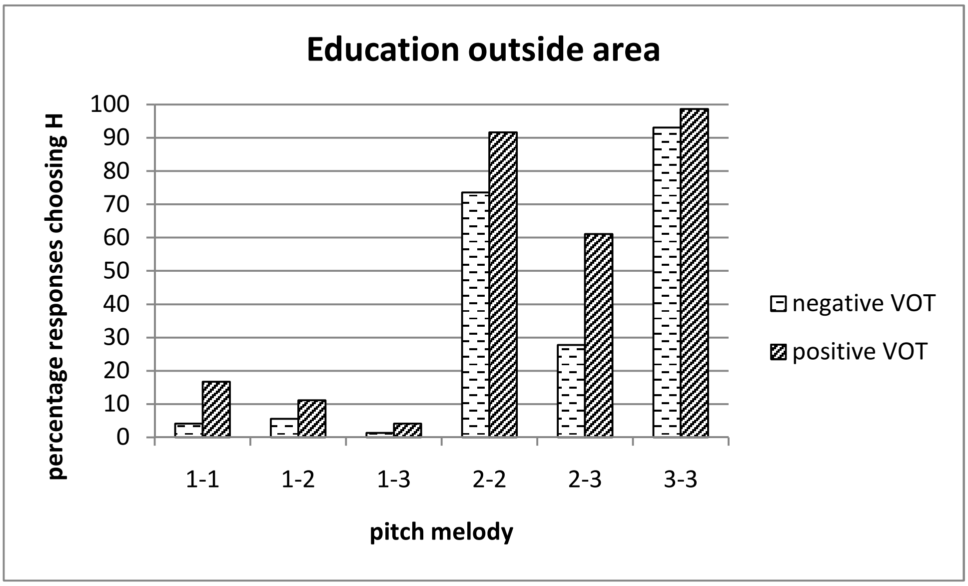


Figure 7-7: percentages choosing H; education outside area



Pitch melodies 1-3 and 3-3 are the unambiguous pitch melodies with virtually everybody identifying 1-3 as L and 3-3 as H, regardless of VOT and regardless of whether people were educated inside or outside of the language area. From the analysis above, the pitch melody where VOT made the greatest difference to the responses was 2-3. But notice from the above table and charts that the difference in responses according to VOT is greater for people who have been educated outside of the language area than inside the area. The analysis above also showed that VOT had an effect for pitch melody 2-2, but notice from the above table and charts that in the case of people educated inside the area VOT makes virtually no difference to the responses; it is only for people educated outside the area that VOT makes a difference. Also notice pitch melody 1-1. In the combined analysis above, VOT did not have a significant effect upon the responses. But notice in the above table and charts that whilst VOT makes virtually no difference to the response of people educated inside the language area, it makes a slight difference to the response of people educated outside the language area.

In the light of these differences, having split the responses according to education McNemar tests were again carried out to test whether or not VOT makes a difference to the responses. The p values for the McNemar tests are below.

Table 7-10: significance statistics for VOT; education inside language area

pitch melody	N	p	$p < 0.05$
1-1	80	1.000	
1-2	80	0.180	
1-3	80	0.500	
2-2	80	0.100	
2-3	80	0.072	
3-3	80	0.500	

Table 7-11: significance statistics for VOT; education outside language area

pitch melody	N	p	$p < 0.05$
1-1	72	0.035	•
1-2	72	0.344	
1-3	72	0.625	
2-2	72	0.001	•
2-3	72	<0.0005	•
3-3	72	0.219	

For people who have only been educated inside the language area, VOT makes no significant difference to the responses for any of the pitch melodies. Even when the pitch melody is ambiguous (such as pitch melody 2-3) VOT is not a significant acoustic cue in perception of the word. However, in the case of people who have been educated outside the language area VOT is making a significant difference to the responses not only for pitch melody 2-3 but also for pitch melodies 1-1 and 2-2, all of which are ambiguous pitch melodies.

7.3.2.4 Results for disyllabic words in isolation

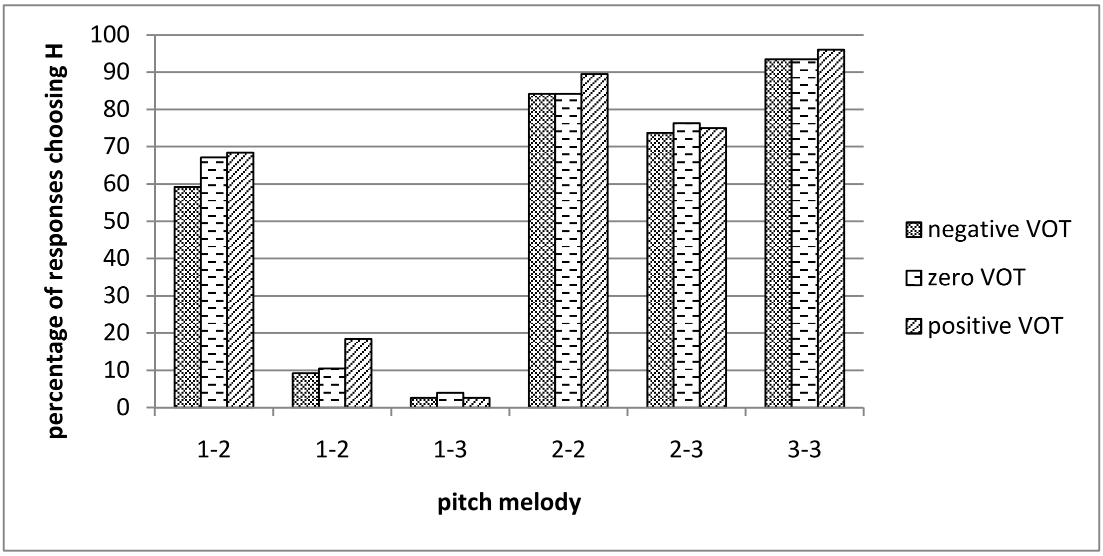
VOT and pitch

The table and graph below give a summary of the percentages of responses equating each pitch melody with L and with H tone words, for disyllabic stimuli in isolation.

Table 7-12: percentages choosing H or LH; disyllabic stimuli in isolation.

pitch melody	VOT	% choosing L	% choosing H	number tokens
1-1	negative	40	60	76
	zero	33	67	76
	positive	32	68	76
1-2	negative	91	9	76
	zero	89	11	76
	positive	82	18	76
1-3	negative	97	3	76
	zero	96	4	76
	positive	97	3	76
2-2	negative	16	84	76
	zero	16	84	76
	positive	11	89	76
2-3	negative	26	74	76
	zero	24	76	76
	positive	25	75	76
3-3	negative	7	93	76
	zero	7	93	76
	positive	4	96	76

Figure 7-8: percentage of responses choosing H; disyllabic stimuli in isolation



Unlike monosyllabic words, the disyllabic stimuli were in isolation rather than in a frame. As has been shown by the results for disyllabic stimuli without an initial plosive (see section 7.3.1.3 above), when words are in isolation the shape of the pitch melody has greater salience than it does when words are in frames. This difference in the salience of the shape of pitch melody changes the effect of VOT.

As is the case with stimuli in frames, the VOT does not have an effect upon perception of the unambiguous pitch melodies 1-3 and 3-3. Word identification was based upon the difference in pitch melody.

For the other pitch melodies there are very slight changes in response as VOT is changed, and these were investigated statistically with McNemar tests. The summary of the significance statistics is as follows.

Table 7-13: significance statistics for VOT

pitch melody	N	p	$p < 0.05$
1-1	76	0.248	
1-2	76	0.118	
1-3	76	1.000	
2-2	76	0.454	

2-3	76	1.000	
3-3	76	0.625	

None of the differences in response due to a change in VOT are statistically significant.

Pitch melodies 1-1 and 2-2 have a level pitch melody, which is the pitch melody shape of an H tone word. The lack of a frame giving a reference pitch means that despite the difference in starting pitch the majority of responses equated these pitch melodies with high tone. This was regardless of the VOT of the word.

Pitch melody 2-3 is the pitch melody where VOT made a difference to the responses if the stimuli were in frames. However, for disyllabic stimuli in isolation the ambiguity associated with 2-3 is largely removed. This is because disyllabic words have second syllable stress which raises the pitch of the second syllable in comparison to the first syllable even in high tone words. A 2-3 pitch melody thus associates with a high tone word, and this is regardless of VOT.

Looking at the above chart, VOT appears to have a slight effect upon perception of the 1-2 pitch melody, but this is not statistically significant, and the majority of responses associated the 1-2 pitch melody with L tone regardless of VOT. Both 2-3 and 1-2 have the same ambiguous amount of rise, which from a listener's perspective could be the result of underlying H tone with second syllable stress or could be the result of underlying L tone (with boundary H% on the second syllable). The higher starting pitch of 2-3 causes the majority to match it with H tone, whereas the lower starting pitch of 1-2 causes the majority to match it with L tone.

In summary, when words are in frames the starting pitch is of greater salience than the overall shape of the pitch melody. In the natural (unmanipulated) form of the language, VOT correlates with the tone, and hence the starting pitch. Because of the greater salience of the starting pitch as a cue to the meaning of the word, when the starting pitch becomes ambiguous, and in addition the shape of the pitch melody is ambiguous, VOT significantly affects the perception of the word, but generally only for people who have had sufficient exposure to Nepali. When words are in isolation the shape of the pitch melody across the word is of greater salience than the starting pitch. Because of the greater salience of the overall shape compared with the starting pitch, even when the pitch melody becomes ambiguous VOT has no significant effect upon the perception of the word.

Phonation

When disyllabic stimuli in isolation start with a sonorant, the phonation has a slight but significant effect upon perception. For this reason, the possible effect of phonation was also considered for disyllabic stimuli starting with obstruents. The particular responses considered were the responses for words with zero VOT. This is because of the way in which the manipulations to the VOT were carried out. In the case of stimuli with zero VOT every pitch melody appeared twice in the experiment – once originating from an L tone word (originally negative VOT and breathier phonation on the first syllable), and once originating from an H tone word (originally positive VOT and clearer phonation on the first syllable). The responses to zero VOT stimuli with identical pitch melody but differing first syllable phonation were paired, and a McNemar test was carried out. The test statistics are: $N = 228$ and $p = 0.358$, leading to the conclusion is that phonation is not a salient cue for word identification in the case of disyllabic words starting with obstruents.

7.3.3 Voicing and pitch for monosyllabic light syllable words

7.3.3.1 Experimental design

All monomoraic content words have a boundary tone which is polar to the underlying tone (see section 5.3.5 above). This gives surface tone patterns H-L% (realised with pitch melody [52] and L-H% (realised with pitch melody [24]), for example:

tá [52] ‘horse’
dà [24] ‘bow’

Because these two pitch melodies are so different, it was decided that manipulating the pitch in the same way as described in 7.3.1 above to produce different degrees of rises and falls starting and ending on different pitches was unlikely to give any further information than that which could be ascertained through the manipulations already carried out for bimoraic words. Thus it was decided to manipulate the pitch of monomoraic words to produce level pitch melodies both in the case of underlying H and in the case of underlying L tone. It is when the pitch melody is most ambiguous that VOT has any effect, and manipulating the pitch melodies to be level gives that extra degree of ambiguity.

In order to first determine to what degree level pitch melodies are ambiguous (without the added complexity of VOT), the following pair of words was chosen for manipulation:

là	'mountain pass'
lá	'wage'

Because the starting pitch for underlying H tone on monomoraic words is higher than the starting pitch for H tone on bimoraic words (see section 4.3) it was decided to take the 3 values of F0 used for the manipulations above, and to add one further value of F0 (250Hz), giving 4 levels for F0.

Two pairs of words starting with plosives were chosen for manipulation of F0 and VOT. These were:

pák	'pastry'	*bàk*	'type of bamboo'
kú	'statue'	*gù*	'nine'

For each word, F0 was manipulated to the four level values of F0 described above. Then for each F0 manipulation, the VOT was manipulated to the three values used above for bimoraic words, resulting in 12 different combinations of F0 and VOT for each original word.

7.3.3.2 Results of perception of monomoraic words

VOT and pitch

The following tables show the percentages of responses associating each pitch level with H and with L tone words. The first table contains the results for words starting with a sonorant, and the second table is for words starting with a plosive. This is following by a graph of the results.

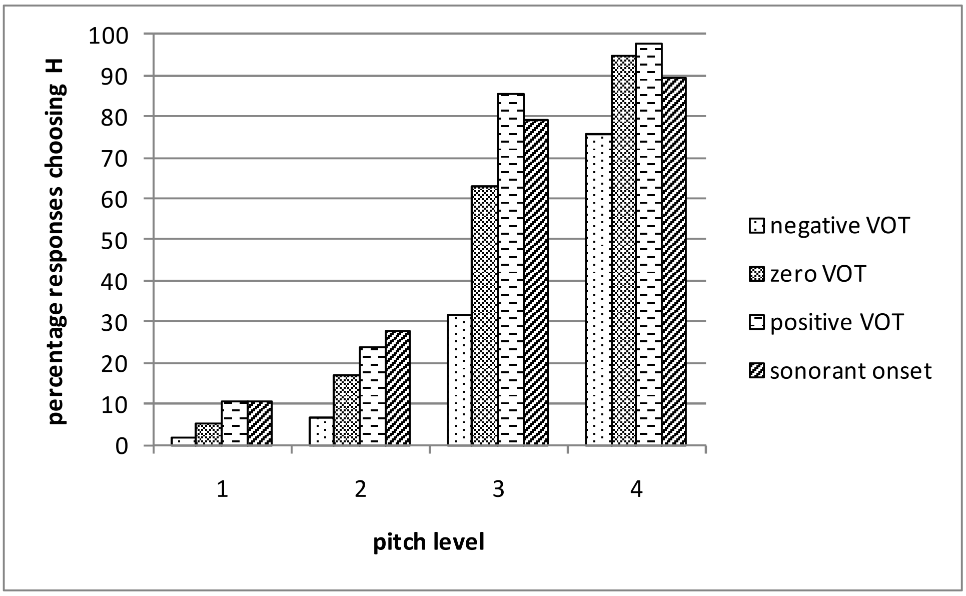
Table 7-14: percentages of responses for pitch levels – sonorant onset

pitch level	1	2	3	4
% responses choosing L	89	72	21	11
% responses choosing H	11	28	79	89
number tokens	76	76	76	76

Table 7-15: percentages of responses for pitch levels – plosive onset

pitch level	VOT	% responses choosing L	% responses choosing H	number tokens
1	-ve	98	2	152
	zero	95	5	152
	+ve	89	11	152
2	-ve	93	7	152
	zero	83	17	152
	+ve	76	24	152
3	-ve	68	32	152
	zero	37	63	152
	+ve	14	86	152
4	-ve	24	76	152
	zero	5	95	152
	+ve	2	98	152

Figure 7-9: percentage of responses choosing H for pitch levels and VOT



As well as showing the percentage responses for stimuli with differing VOT, the above chart also includes for comparison the responses when the stimuli started with a sonorant.

When the stimuli started with a sonorant, the majority of responses identified stimuli with pitch levels 1 and 2 as L tone words and stimuli with pitch levels 3 and 4 as H tone words. However, the ambiguity caused by the lack of a rise or fall in the pitch melody meant that even for the highest and lowest levels approximately 10% of responses associated low level with H and high level with L. The ambiguity increases for levels 2 and 3.

Because of this ambiguity, VOT does have an effect upon the perception of stimuli starting with plosives. From the above graph and tables of percentages, it can be seen that in general if the VOT is positive the percentages of responses are very similar to the percentages for stimuli starting with a sonorant. However, the percentage of responses choosing H decreases as the VOT decreases.

The results for bimoraic stimuli in the section above show that when the pitch melody was ambiguous it was in general the people educated outside the Walungge area who were responding to the VOT; in general VOT was not a salient cue for people educated inside the Walungge area. Because of this, the responses for monomoraic stimuli with level pitch melodies were split according to education. The charts and table below show the percentages of responses for each group of people.

Table 7-16: percentages of responses for light syllabled words with level pitch

pitch level	education outside	% choosing H for +ve VOT	% choosing H for -ve VOT	difference in %	number tokens
1	y	14	1	13	72
	n	8	3	5	80
2	y	24	6	18	72
	n	24	8	16	80
3	y	86	24	63	72
	n	85	39	46	80
4	y	97	68	29	72
	n	99	83	16	80

Figure 7-10: chart of responses for VOT; people educated inside the area

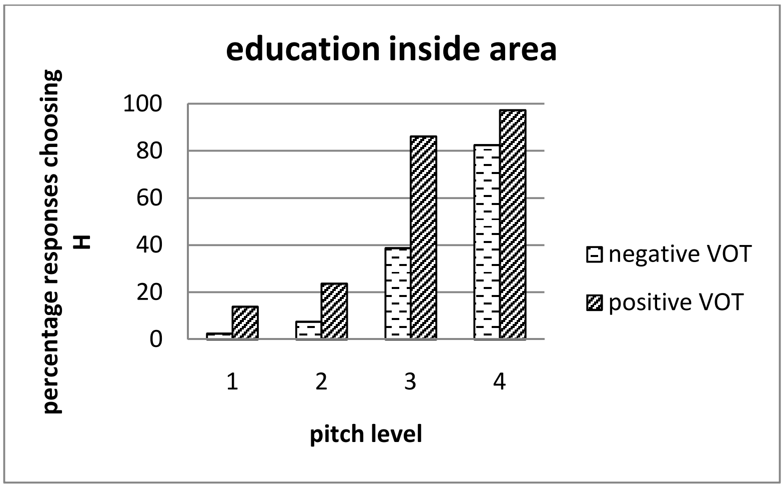
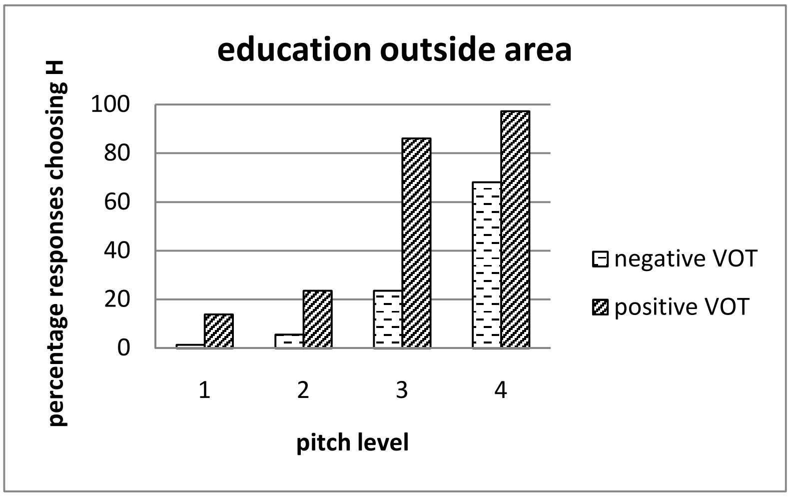


Figure 7-11: chart of responses for VOT; people educated outside the area



For pitch level 1 the change in responses associated with a change in VOT is only slight, with the vast majority of responses identifying stimuli with pitch level 1 as L tone words regardless of VOT. There is very little difference in the responses between those educated inside and those outside the Walungge speaking area. In the case of pitch level 2 the change in responses is greater, but still with the majority of responses identifying level 2 as L tone regardless of VOT. Again there is very little difference in the responses according to education.

When the pitch is at level 4, whilst there is a difference in responses correlating with a difference in VOT, the majority of responses identify level 4 as H tone regardless of VOT. However, education did make a difference to how people responded to stimuli with level 4 pitch but negative VOT. Where people have been educated inside the language area, 83% of responses equated negative VOT with H tone. However, for people who have been educated outside, this percentage drops to 68%.

In the case of pitch level 3, however, VOT does cause a big difference in response for both groups of people. When the VOT was positive approximately 85% of responses chose an H word regardless of where they were educated. When the VOT was negative, approximately 40% of responses of those educated inside the language area identified level 3 as H tone. But this percentage drops to approximately 25% for those educated outside the area. An explanation as to why this pitch level in particular results in VOT causing such a difference could be as follows. It is negative VOT rather than positive VOT which people are responding to (the positive VOT results are similar to the results for words starting with sonorants). The majority of people already identify pitch levels 1 and 2 as L tone regardless of VOT, so negative VOT simply causes a few more responses to be equated with L tone. For pitch level 4, the pitch is sufficiently high that the majority identify it as H tone regardless of VOT. This is not the case when the pitch is at level 3. The pitch is high enough that in the absence of VOT the majority identify it as H tone, but not sufficiently high that the majority ignore VOT. Thus when the VOT becomes negative, and in the absence of a rise or a fall to indicate the tone of the word, the negative VOT becomes a salient cue causing the majority to identify it as L tone.

For each population group, McNemar tests were carried out for paired responses of positive and negative VOT for each pitch level. Below is the summary of the significance statistics for these tests.

Table 7-17: significance statistics for VOT; education inside language area

pitch level	N	p	$p < 0.05$
1	80	0.219	
2	80	0.011	•
3	80	<0.0005	•
4	80	0.001	•

Table 7-18: significance statistics for VOT; education outside language area

pitch level	N	p	$p < 0.05$
1	72	0.012	•
2	72	0.007	•
3	72	<0.0005	•
4	72	<0.0005	•

Apart from pitch level 1 for people educated inside the area, the effect of VOT is statistically significant for all pitch levels for people educated both inside and outside of the language area. However, when the pitch is at levels 1, 2, or 4 it has to be stressed that even though the effect of VOT is statistically significant the actual change in the percentages is very slight. The vast majority of people choose L tone for pitch levels 1 and 2 and H tone for pitch level 4 regardless of VOT and regardless of where they have been educated. It is only for pitch level 3 that VOT makes a large difference to the responses.

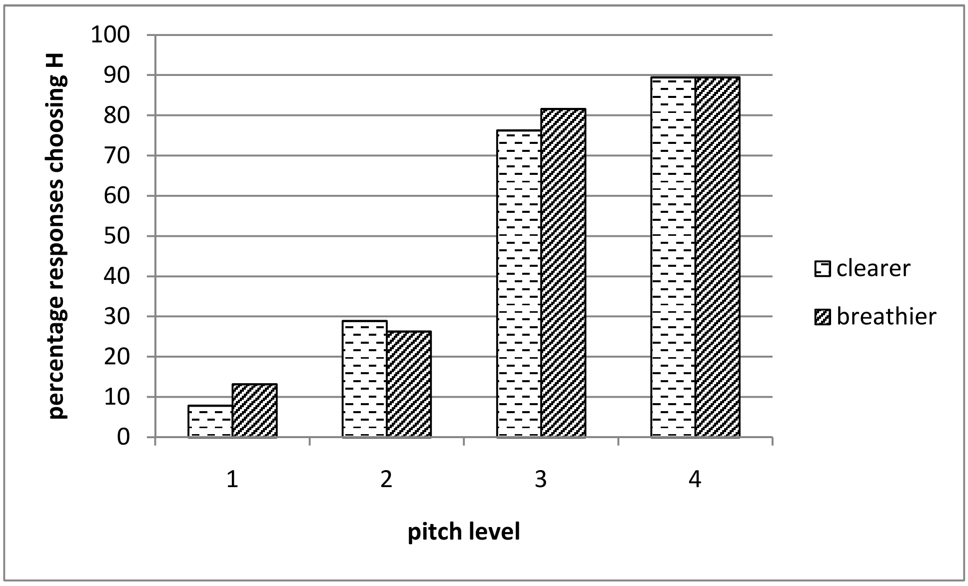
Phonation

In order to consider the effect of phonation, the responses for stimuli starting with sonorants were paired by phonation and a McNemar test was carried out. The results of this test are below, and as can be seen from them there is no significant effect due to phonation.

Table 7-19: significance statistics for pairs of stimuli with differing phonation

N	p	$p < 0.05$
152	1.000	

Figure 7-12: bar chart showing responses for stimuli with differing phonation



Because phonation is not a salient cue for identifying words starting with sonorants, it was not considered any further for words starting with obstruents.

7.3.4 Discussion on voicing and tone

One of the phonological issues that has already been discussed in chapter 6 of this dissertation is whether in a pair of words such as *tá* 'horse' and *dà* 'bow' voicing is a phonological contrast and is determining the tone, or whether voicing is not a phonological contrast but is being determined by the tone. The conclusion in section 6.3.1 is that voicing is not phonologically contrastive for obstruents. However, the arguments are not completely clear-cut, and arguments both for and against voicing being phonological were presented.

Now consider the diachronic process from a voicing contrast to a tone contrast. There is a point in this process for which Maran (1973) uses the term "cognitivization" whereby "the basic system of cognitive contrast is transferred from a segmental system of cues to a prosodic system. This event means, among other things, that it is now far simpler to conduct the cognitive process of perceiving and analysing linguistic data largely in terms of prosodic features of tone." In going from a voicing contrast to a tonal contrast there is a change in perception of the prosodic cues and the segmental cues. The results of the perception experiments above suggest that this change has taken place, and give additional weight to the arguments of chapter 6 that voicing is not a phonological contrast in Walungge.

Exposure to Nepali (as measured by education outside the language area) affects perception, with a greater salience given to voicing in the instances where the pitch is ambiguous. The experiments described above for Walungge are similar to those described in Pearce (2007) for Kera. In a similar manner to Walungge, Kera has voicing of obstruents correlating with tone. Manipulations were carried out to create a series of stimuli with different combinations of VOT and F0. Pearce found that whilst those who had less exposure to French (e.g. village women) responded to F0 as the salient cue, those with greater exposure to French (e.g. town men) responded to the VOT as the most salient cue. Pearce concluded that whilst diachronically the tone contrast in Kera has arisen from a voicing contrast, in the synchronic form of the language (as spoken by those without exposure to French) voicing is not contrastive. However, she also concluded that exposure to French is causing voicing to be reanalysed as contrastive in the town dialect of the language. It is possible that in Walungge a similar reanalysis could happen. However, from the perception experiments it appears that Walungge is not yet at that point. Even for those exposed to Nepali, it is only in the cases where the pitch is ambiguous that VOT becomes salient. In all other instances, pitch is still the most salient cue.

7.4 Length versus tone

7.4.1 Experimental design

As outlined above, monosyllabic monomorphemic nouns with underlying high tone have surface tone patterns, and hence pitch melodies, which differ according to the syllable weight. Those with light syllables have a surface tone pattern H-L% (realised as pitch [52]), and those with heavy syllables have a surface tone pattern H-H% (realised as pitch [44]) e.g:

ɕá [52] 'meat'

ɕá: [44] 'deer'

When distinguishing between these words there are two main acoustic cues, namely the duration of the vowel and the level versus falling pitch melody. Further, the acoustic work in section 4.3.2 above shows that the starting pitch of words with light syllables is significantly higher than the starting pitch of words with heavy syllables. Thus, when considering the salience of the pitch melody it is important to consider the effect of starting pitch height as well as whether the overall contour is level or falling.

In order to create a series of recordings with differing vowel durations and F0, manipulations were carried out in much the same way as the manipulations for the experiments detailed in the above sections. Using the same method of choosing values of F0 as in the previous sections of the experiment, the starting values of F0 were taken to be 250Hz for surface H-L% tone and 226Hz for surface H-H% tone. The value of F0 for the finish of the fall was taken to be 185Hz. In order to investigate the salience of the starting pitch as well as the salience of the shape of the melody, level melodies were created at both 250Hz and at 226Hz, and falls were created starting at both 250Hz and at 226Hz. In order to take into consideration the salience of the amount of fall, for each starting pitch a smaller fall was also created. This gives a total of 6 different manipulated pitch melodies, summarised below:

Table 7-20: manipulated pitch melodies

shape of pitch melody	╲	╲	—	╲	╲	—
name of pitch melody	4-1	4-2	4-4	3-1	3-2	3-3
starting F0 (Hz)	250	250	250	226	226	226
ending F0 (Hz)	185	215	250	185	203	226

Using the word *ɕá* ‘meat’ as a basis for the manipulations, F0 was manipulated to create each of the 6 melodies listed above.

For each of the 6 new recordings, the duration of the vowel was also manipulated to be 4 different durations (based on average durations from the acoustic analysis in section 4.5): 100ms; 130ms; 160ms; 190ms. This gives 24 recordings.

In order to have a duplicate set of recordings, but also to account for any residual effect from the original recording (e.g. the phonation across the vowel), the same manipulations were done for the word *ɕáː* ‘deer’. The same manipulations were also carried out for a further pair of words: ná ‘nose’ and *náː* ‘blue sheep’. All manipulated recordings were then put into the frame *nò ... dè* “this is a ...”.

7.4.2 Results of experiment

Effect of duration versus pitch

The following table shows the percentages of responses associating each combination of duration and pitch melody with a short vowel H-L% (surface) tone word or a long vowel H-H% tone word. The table is followed by two charts of the results. The first shows how the percentages change with pitch melody when the results are grouped by duration. The second shows how the percentages change with the duration when the results are grouped by pitch melody.

Table 7-21: percentages of responses for duration and pitch melody

duration (ms)	pitch melody	% responses choosing short vowel HL tone	% responses choosing long vowel H tone	number tokens
100	4-1	98	2	152
	4-2	89	11	152
	3-1	84	16	152
	3-2	82	18	152
	3-3	80	20	152
	4-4	84	16	152
130	4-1	88	12	152
	4-2	67	33	152
	3-1	66	34	152
	3-2	43	57	152
	3-3	41	59	152
	4-4	39	61	152
160	4-1	39	61	152
	4-2	32	68	152
	3-1	29	71	152
	3-2	16	84	152
	3-3	8	92	152
	4-4	11	89	152
190	4-1	19	81	152
	4-2	14	86	152
	3-1	13	87	152
	3-2	9	91	152
	3-3	3	97	152
	4-4	3	97	152

Figure 7-13: percentage choosing H-L% & short vowel; grouped by duration

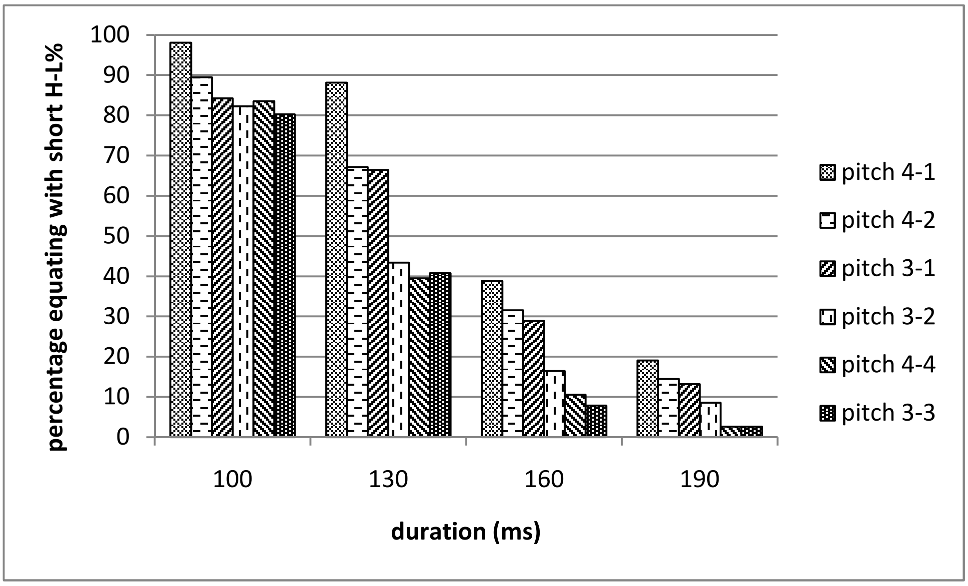
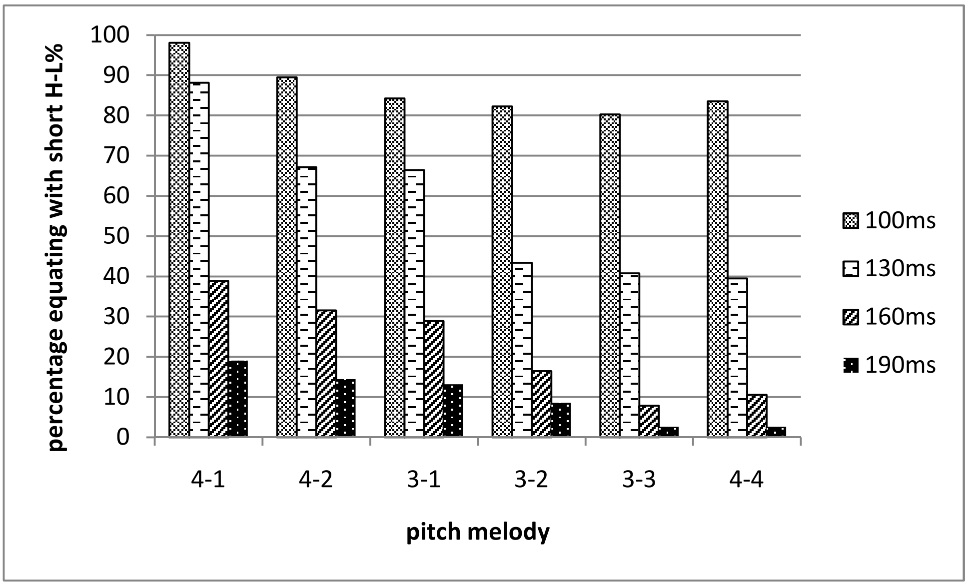


Figure 7-14: percentage choosing short vowel & H-L%; grouped by pitch melody



The first thing to notice from the above results is that when there is a simple choice between a pitch melody which matches the surface tone (either 4-1 or 3-3) and a duration which matches the phonological length (either 100ms or 190ms), then the majority choose vowel duration over pitch melody. That is, for the combination of 4-1 pitch melody with 190ms duration the majority identify this as a long vowel H-H% word; and for the combination of 3-3 pitch melody with 10ms duration the majority identify this as a short vowel H-L% word. However even though the majority chose duration over pitch, in each case approximately 20% of responses chose pitch over duration, identifying the 190ms 4-1 combination as a short vowel H-L% word, and the 100ms 3-3 combination as a long vowel H-H% word.

When the duration is neither at its longest or its shortest (i.e. 130ms and 160ms) the pitch has a greater effect upon the perception of the stimuli. From the acoustic analysis in chapter 4, a duration of 130ms is approximately the third quartile for durations of short vowels with H-L% tone. That is to say, approximately a quarter of short vowels with H-L% tone are longer than 130ms. So although 130ms is not as unambiguous as 100ms, it still has a duration comparable to that of a short vowel. Yet in the case of 130ms the pitch melody has a great effect upon perception. When the pitch melody was unambiguous for H-L% tone (i.e. 4-1), the vast majority (88%) identified this as a short vowel H-L% word. When the fall of the pitch melody was decreased, either by starting at pitch level 4 but not falling quite as much (4-2) or by starting slightly lower but falling to low (3-1), the fall was still sufficient for the majority to perceive this as a short H-L% word, but the percentage was reduced (67% and 66% respectively). A 3-2 fall does not fall enough to be clearly associated with H-L% tone. Indeed the responses for 3-2 fall, and 3-3 and 4-4 level pitches were all very similar. Pitch and duration were in conflict, with approximately 60% responding to pitch as the more salient cue (and thus identifying the stimuli as a long vowel H-H% word) and 40% responding duration (identifying the stimuli as a short vowel H-L% word).

In the case of 160ms vowel duration, pitch has a similar, but not quite as great, effect upon subjects' perception of the stimuli. From the acoustic analysis in chapter 4, 160ms is approximately the first quartile for long vowel durations. When the pitch of the stimuli is level (3-3) or (4-4), approximately 90% of responses identify this as a long vowel H-H% word. However, as a fall is introduced this percentage drops. By the time the fall is at its maximum (4-1), the percentage of responses identifying this as a long vowel H-H% word has dropped to 61%.

What is clear from the above graphs is the effect of duration upon the responses to the stimuli. What is not so clear is whether the effect of the pitch melody is significant. In order to statistically investigate the effect this, the responses for different pitch melodies of the same duration were paired and McNemar tests were carried out.

This first table of p values compares the two unambiguous pitch melodies 4-1 and 3-3, confirming that the difference in these pitch melodies significantly affects the responses for all vowel durations; the pitch melody is a salient cue in word identification.

Table 7-22: McNemar test p values for 4-1 versus 3-3

N = 152	100ms		130ms		160ms		190ms	
pitch melodies compared	p	sig.	p	sig.	p	sig.	p	sig.
4-1 vs. 3-3	<0.0005	•	<0.0005	•	<0.0005	•	<0.0005	•

When the pitch melody is level the starting pitch does not have any significant effect upon perception; it is the contour of the pitch melody rather than the starting pitch which is of salience.

Table 7-23: McNemar test p values for 3-3 versus 4-4

N = 152	100ms		130ms		160ms		190ms	
pitch melodies compared	p	sig.	p	sig.	p	sig.	p	sig.
3-3 vs. 4-4	0.441		0.906		0.523		1.000	

Now consider the p values when comparing the large fall 4-1 with a slightly smaller fall (4-2 or 3-1).

Table 7-24: McNemar test p values for different pitch melodies and durations

N = 152	100ms		130ms		160ms		190ms	
pitch melodies compared	p	sig.	p	sig.	p	sig.	p	sig.
4-1 vs. 3-1	<0.0005	•	<0.0005	•	0.073		0.200	
4-1 vs. 4-2	0.002	•	<0.0005	•	0.222		0.310	
4-2 vs. 3-1	0.170		1.000		0.688		0.850	

There is a significant difference between 4-1 and 3-1 for the shorter two vowel durations. But notice that there is also a significant difference between 4-1 and 4-2, and no significant difference between 4-2 and 3-1. This implies that the difference in perception between 4-1 and 3-1 is not because the starting pitch is different but because the amount of fall is different. This adds weight to the suggestion given in section 4.3.6.4 that the higher starting pitch for H-L% words compared with H-H% words is a phonetic effect which is there to create enough "phonetic space" for the fall.

In summary, then, the starting pitch does not significantly affect perception, but the amount of fall does. Even though the most salient factor in word identification is the duration (which is the result of an underlying phonological contrast), the shape of the pitch melody (which is the result of a difference in surface tone) still plays an important role in word identification.

Phonation

The effect of phonation upon the responses was also considered. Phonation varies across a word in correlation with its pitch melody, so even if the original pitch melody is manipulated the phonation across the resulting stimuli will still reflect the original pitch melody. Responses were paired according to whether the stimuli came from a word that was originally H-L% surface tone or a word that was originally H-H% surface tone. For each duration and (manipulated) pitch melody a McNemar test was carried out on "original H-L%" and "original H-H%" pairs to see whether there is any significant effect on responses from phonation. The table below gives the p values for each combination of duration and pitch melody.

Table 7-25: McNemar test p values for phonation

N = 76	100ms		130ms		160ms		190ms	
pitch melody	p	sig.	p	sig.	p	sig.	p	sig.
4-1	1.000		0.804		1.000		1.000	
4-2	0.754		0.571		0.377		1.000	
3-1	0.077		0.424		0.123		0.096	
3-2	1.000		0.391		0.332		0.267	
3-3	0.824		0.105		0.146		1.000	
4-4	0.383		0.405		0.774		0.125	

Taking a significance level of $p < 0.05$, phonation does not have a significant effect on perception for any duration or pitch melody. However, from the above table it can be seen that there is a wide range of "p" values, ranging from those which are only slightly above the 0.05 level, to those which have a value of $p = 1.000$. The values of $p = 0.077$ and $p = 0.096$ could be attributed to the fact that pitch melody 3-1 is one of the more ambiguous pitch melodies and so phonation is having a very slight effect upon perception. On the other hand, from the earlier analysis pitch melody 4-2 has similar ambiguity to 3-1. Yet the "p" values for 4-2 are mostly over 0.5, suggesting that any difference in perception is more likely to be the result of chance than phonation. The conclusion is that phonation is not a salient acoustic cue.

7.4.3 Discussion of length and tone

In some ways the perception test for length and tone is parallel to the one for voicing and tone. Both tests have one variable which is phonologically contrastive versus another variable which is dependent upon the first. Voicing is dependent upon the underlying tone; the distinction between H-H% and H-L% is determined by the number of moras. However, there is an important difference between the two types of perception test. There is no situation in the language where speakers are purely dependent upon voicing to determine the meaning of the word; pitch is always there as the main cue to for word identification. However, there are situations where the duration of the vowel is the same but the pitch melody is different, e.g.

[tiː ^{44}] ‘foal’

[tiː ^{51}] ‘foal.POSS’

Because in situations like this people have to use pitch as the main cue for word identification, it is conceivable that in a situation such as [ɕa ^{52}] versus [ɕaː ^{44}] they will be more aware of the pitch than they will of the voicing in [pak ^{53}] versus [bak ^{24}]. This is what the perception tests show. For a straight choice between pitch and voicing, virtually 100% of responses went for pitch rather than voicing. But for a straight choice between duration of vowel or shape of pitch melody, approximately 80% went for duration and 20% went for the shape of the pitch melody. In the case of voicing versus pitch, it was only in situations where the pitch was extremely ambiguous that voicing caused a significant difference to the responses. Whereas in the case of pitch melody versus duration, the pitch melody caused a significant difference to the responses for all vowel durations.

The salience of the level versus falling pitch melody raises an interesting issue to do with tonogenesis. Section 2.6 gives examples of a number of different Tibetan languages. These include languages which simply have a high/low tone contrast, languages where it is debatable whether or not the rises and falls in the pitch are realisations of phonologically contrastive tones, and languages which have rising and falling tone as well as high and low tone. In this final group of languages, the contrastive rises and falls correspond to the deletion of word final consonants which are there in the Written Tibetan. Returning to Walungge, it has already been noted (section 3.1.2) that a word final /t/, e.g. /pǿt/ ‘story’, may either be realised as a glottal stop or may be deleted with compensatory vowel lengthening, but in both cases the pitch is falling, e.g. [pøʔ ^{53}] or [pøː ^{51}]. The perception experiments above have already shown that the falling pitch is a salient cue for word identification. Thus it could be that the H-H% and H-L% surface tone patterns are one stage along the route towards the development of an H versus HL underlying tone contrast. If the optional deletion of word final glottals became permanent deletion, Walungge would be very similar to other Tibetan languages which have an underlying high versus falling contrast.

8 Conclusion

This dissertation has investigated a number of issues involving tone in Walungge, including acoustic analysis, phonological analysis, and perception experiments. Each of these different types of analysis and investigation has broadly revolved around two main facets of Walungge tone. One is the inter-relationship between the correlates of tone, which include pitch, voicing and phonation. The other is the relationship between tone, pitch melodies, and syllabic structure.

The correlates of tone were investigated acoustically in chapters 4 (pitch, phonation, and duration) and 6 (pitch and voicing). This acoustic work was then used as a basis for investigating the phonological relationship between tone and voicing (chapter 6) and the perception experiments (chapter 7). The conclusion that was reached from the phonological analysis is that the voicing of plosives/affricates in Walungge is not phonologically contrastive, but is determined by the tone. However, it was also concluded that voicing is not purely a phonetic effect caused by the lower pitch. Rather, low tone in Walungge can be analysed as being the feature [L/voice] which spreads to a word initial plosive/affricate resulting in voicing. The perception experiments in chapter 7 showed that in the identification of words which differ phonologically only by tone, it is pitch rather than voicing which is of primary salience.

The relationship between the syllabic structure and the realised pitch melodies was acoustically investigated in chapter 4, with the conclusion that the difference between falling and level pitch melodies, which correlates with a difference in syllable structure, could not be explained as a purely phonetic effect. On the basis of this conclusion, chapter 5 analyses Walungge as forming one tonal foot per word, to which is attached an underlying tone (H or L) and a boundary tone (H% or L%). The boundary tone is determined by the underlying tone and the number of moras in the tonal foot. This analysis can account for the realised pitch melodies both on monomorphemic words and across morpheme boundaries. The relationship between the syllabic structure and the surface tone patterns was returned to in chapter 7, where perception experiments were carried out to determine the relative salience of the cues of vowel duration and shape of pitch melody. Vowel duration was clearly the more salient cue, but the shape of the pitch melody still had a significant effect upon perception.

This dissertation is by no means a complete investigation into Walungge tone, and there are a number of ways in which the research could be extended. One very interesting area of further research would be to adapt the perception experiments of chapter 7 for other Tibetan languages. Chapter 2 gives an overview of the wide variety there is in Tibetan languages, both in terms of the relationship between tone and voicing, and in terms of the tones/pitch melodies. For example, Lhomi is similar to Walungge in the relationship between tone and voicing, but there is controversy in the way in which this is analysed (see section 2.7.1). Perception experiments could shed further light on what is happening in Lhomi. Dolpo (section 2.5.2) has both phonological voicing and phonological tone, but there is still a relationship between voiced plosives and low tone. It would be interesting to compare a Dolpo speaker's perception of the relative salience of the acoustic cues with a Walungge speaker's perception. Lhasa Tibetan is a language which has been analysed as having 2 tones, 4 tones, or even 6 tones, depending in part on whether vowel length is taken to be a separately contrastive feature or a property of tone (see section 2.7.2). The perception experiments that were carried out for Walungge tone versus length could be adapted for Lhasa Tibetan, and give further insight into the relationship between tone and length in Lhasa.

A central part of the analysis in chapter 5 is the use of a tonal foot to account for the realised patterns. A similar analysis has already been used for describing the patterns found in Refugee Standard Tibetan (see section 5.1.3). It is possible that tonal feet could also be used to account for the patterns found in other Tibetan languages.

In order to account for patterns found at different types of morpheme boundaries, the analysis of chapter 5 drew on the prosodic hierarchy. This is a whole area of research that could be extended. The prosodic hierarchy has a number of different levels, but for the most part chapter 5 was only looking at the level of the phonological word. It is clear that there are other things happening at other levels. For example, in order to account for the fall which happens on verbs, the analysis in section 5.5 proposed a boundary tone L*, mentioning that the prosodic level at which L* was operating was a topic for further research. Related to the issue of tone patterns across morpheme boundaries is the fact that different types of verbal suffixes have different tone patterns. This was accounted for by analysing some suffixes, but not others, as being extra-tonal. An area of further research would be to extend the study of tone patterns on verb forms to investigate whether there is any commonality (either phonologically or morphologically) between suffixes which is determining whether or not they are analysed as extra-tonal.

Although the focus of this dissertation has been tone in Walungge, it contains a body of language data which is hoped will be of benefit to those investigating other areas of Tibetan linguistics. It is also hoped that the tonal analysis contained in this dissertation will be of benefit not just to those researching Tibetan languages, but also to those working with tone in other areas of the world.

Bibliography

Population Census of Nepal 2001. Kathmandu, Central Bureau of Statistics, Nepal.

Avery, P. and W. Idsardi (2001). Laryngeal dimensions, completion and enhancement. *Distinctive Feature Theory*. T. Hall. Berlin, Mouton de Gruyer.

Bao, Z. (1999). *The structure of tone*. Oxford, Oxford University Press.

Benedict, P. (1972). *Sino-Tibetan: a conspectus*. Cambridge, Cambridge University Press.

Beyer, S. (1992). *The Classical Tibetan language*. Albany, State University of New York Press.

Blankenship, B. (1997). The time course of breathiness and laryngealization in vowels, UCLA. PhD.

Booij, G. (1983). "Principles and parameters in prosodic phonology." *Linguistics* 21: 249-280.

Booij, G. (1996). "Cliticization as prosodic integration: the case of Dutch." *The linguistic review* 13: 219-242.

Bradley, D., Ed. (1997). *Papers in Southeast Asian linguistics no. 14: Tibeto-Burman Languages of the Himalayas*. Canberra, Australian National University (Pacific Linguistics).

Bradshaw, M. (1999). A cross-linguistic study of consonant-tone interaction, Ohio State University. PhD.

Caplow (2007). Directionals in Tokpe Gola Tibetan discourse. *Linguistics of the Himalayas and Beyond*. R. Bielmeier and F. Haller. Berlin, Mouton.

Caplow, N. (2009). The role of stress in Tibetan tonogenesis: a study in historical comparative acoustics. Santa Barbara, University of California. PhD.

Chang, K. and B. Shefts (1964). *A manual of spoken Tibetan (Lhasa dialect)*. Seattle & London, University of Washington Press.

Chao, Y. (1930). "A system of tone letters." *Le Maitre Phonetique* 45: 24-27.

Chen, M. (2000). *Tone Sandhi*. Cambridge, Cambridge University Press.

Clements, G. N. (1985). "The geometry of phonological features." *Phonology yearbook* 2: 225-252.

Clements, G. N. and E. Hume (1995). The internal organization of speech sounds. *The handbook of phonological theory*. J. Goldsmith. Oxford, Blackwell.

Connell, B. (2002). "Tone languages and the universality of intrinsic F0: evidence from Africa." *Journal of Phonetics* 30: 101-129.

Davies, P., A. Fourcin, et al. (1986). "Variation in glottal open and closed phase for speakers of English." *Proceedings of the Institute of Acoustics* 8: 539-546.

Dawson, W. (1980). Tibetan Phonology, University of Washington. PhD.

De Lacy, P. (2002). "The interaction of tone and stress in optimality theory." *Phonology* 19(1): 1-32.

DeLancey, S. (1990). Sino-Tibetan languages. *The major languages of East and South-East Asia*. B. Comrie. London, Routledge: 69-82.

Denwood, P. (1999). *Tibetan*. Amsterdam

Philadelphia, John Benjamins.

Dixon, R. and A. Aikhenvald (2002). *Word: a cross-linguistic typology*. Cambridge, Cambridge University Press.

Duanmu, S. (1992). An autosegmental analysis of tone in four Tibetan languages. *Linguistics of the Tibeto-Burman Area*. Berkley, CA. 15.1: 65-91.

Duanmu, S. (2000). *Phonology of standard Chinese*. Oxford, Oxford University Press.

Ebert, K. (1979). Sprache und Tradition der Kera (Tschad). *Teil III: Grammatik*. Berlin, Reimer.

Fabb, N. (1998). Compounding. *The handbook of morphology*. A. Spencer and A. Zwicky. Oxford, Blackwell.

Fürer-Haimendorf, C. v. (1975). *Himalayan Traders: life in highland Nepal*. London, J. Murray.

Geziben, D.-S. (1996). Trochaic structure in Tibetan phonology: a metrical analysis of tone in Lhasa Tibetan. Arlington, University of Texas. MA.

Goldsmith, J. (1976). Autosegmental Phonology, MIT. PhD.

Goldstein, M. (1984). *English-Tibetan Dictionary of Modern Tibetan*. Berkeley, University of California Press.

Gordon, M. and P. Ladefoged (2001). "Phonation types: a cross-linguistic overview." *Journal of phonetics* 29: 383-406.

Gordon, R. G., Ed. (2005). *Ethnologue: Languages of the World, fifteenth edition*. Dallas, Tex, SIL International.

Hale, A. (1982). *Research on Tibeto-Burman Linguistics*. Berlin, Mouton.

Hall, T. (1999). The phonological word: a review. *Studies on the phonological word*. T. Hall and U. Kleinhenz. Amsterdam, John Benjamins.

Hall, T. and K. Hildebrandt (2008). "Phonological and morphological domains in Kyirong Tibetan." *Linguistics* 46(2): 215-248.

Halle, M. (2002). *From memory to speech and back*. Berlin, Mouton de Gruyter.

Halle, M. and K. Stevens (1971). A note on laryngeal features. *Quarterly Progress Report*. MIT. 101: 198-213.

Haller, F. (1999). "A brief comparison of register tone in Central Tibetan and Kham Tibetan." *Linguistics of the Tibeto-Burman Area* 22.2: 77-97.

Hari, A. M. (1979). Investigation of the Tones of Lhasa Tibetan. *Asian Pacific Series no.13*. Huntington Beach, SIL.

Harris, J. (2003). *Grammar-internal and grammar-external assimilation*. Proceedings of the 15th international congress of phonetic sciences, Barcelona, Casual Productions.

Harris, J. (2009). Why final obstruent devoicing is weakening. *Strength relations in phonology*. N. Nasukawa and P. Backley. Berlin, Mouton de Gruyter.

Haudricourt, A.-G. (1954). "De l'origine des tons en veitnamien." *Journal Asiatique* 242: 68-82.

Hayes, B. (1995). *Metrical stress theory: principles and case studies*. Chicago, University of Chicago press.

Hayward, K. (2000). *Experimental phonetics*. Harlow, Longman.

Hill, N. (2005). "Once more on the letter a-chung." *Linguistics of the Tibeto-Burman Area* 28(2): 107-137.

Hill, N. (2010). "An overview of old Tibetan synchronic phonology." *Transactions of the philological society* 108(2): 110-125.

Hombert, J. M. (1978). Consonant types, vowel quality and tone. *Tone: a linguistic survey*. V. Fromkin. New York, Academic Press.

Hombert, J. M., J. J. Ohala, et al. (1979). "Phonetic explanation for the development of tone." *Language* 55-1.

Hu, T. (1982). "Recherches sur let tons du Tibétain (dialecte de Lhasa)." *Cashiers de Linguistique Asie Orientale* XI(1): 11-46.

Hu, T. (1991). Zangyu (The Tibetan Language). *Han-Zangyu gailun [A general introduction to Sino-Tibetan languages)*. X. e. al. Beijing, Beijing University Press: 109-207.

Hu, T., J. Qu, et al. (1982). "Zangyu (Lasa hua) shengdiao shiyan (Experiments in Lhasa Tibetan)." *Yuyan Yanjiu* 1982.1: 18-38.

Huang, B. (1997). "Conditions for tonogenesis and tone split in Tibetan dialects." *Linguistics of the Tibeto-Burman Area* 18.1.

Huber, B. (2003). Tone in Kyirong Tibetan. *Themes in Himalayan languages and linguistics*. M. Turin. Heidelberg: South Asia Institute; Kathmandu:, Tribhuvan University.

Huber, B. (2005). *The Tibetan dialect of Lende (Kyirong): a grammatical description with historical annotations*. . Bonn, VGH Wissenschaftsverlag.

Hyman, L. (2001). Tone Systems. *Language Typology and Language Universals: an international handbook*. M. Haspelmath, E. König, W. Oesterreicher and W. Raible. Berlin & New York, Walter de Gruyter. 2.

Hyman, L. (2007). How (not) to do phonological typology: the case of pitch-accent. *UC Berkeley Phonology Lab annual report*. Berkeley, University of California.

Jäschke, H. (1881). *A Tibetan-English dictionary*. London, Routledge & Kegan Paul.

Johnson, K. (2003). *Acoustic and Auditory Phonetics (2nd Edition)*. Oxford, Blackwell.

Kager, R. (1993). "Alternatives to the iambic-trochaic law." *Natural language and linguistic theory* 11: 381-432.

Kager, R. (2007). Feet and metrical stress. *The Cambridge handbook of Phonology*. P. De Lacy. Cambridge, Cambridge University Press.

Kehoe, M., G. Hilaire-Debove, et al. (2008). "The structure of branching onsets and rising diphthongs: evidence from the acquisition of French and Spanish." *Language Acquisition* 15(1).

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago, University of Chicago.

Ladefoged, P. (2003). *Phonetic data analysis: an introduction to fieldwork and instrumental techniques*. Oxford, Blackwell publishing.

Ladefoged, P. (2006). *A course in phonetics (5th edition)*. Boston, Thomson Wadsworth.

LaPolla, R. (2000). Subgrouping in Tibeto-Burman *33rd International Conference on Sino-Tibetan Languages and Linguistics*. Bangkok.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge, Cambridge University Press.

Leben, W. (2001). Tonal Feet. *Typology of African Prosodic Systems Workshop*. Bielefeld University, Germany.

Lindsey, G., K. Hayward, et al. (1992). "Hausa glottalic consonants: a laryngographic study." *Bulletin of the School of Oriental & African Studies, University of London* 55(3): 511-527.

Löfqvist, A. and R. McGowan (1992). "Influence of consonantal envelope on voice source aerodynamics." *Journal of phonetics* 20: 93-110.

Lombardi, L. (1991). Laryngeal features and laryngeal neutralisation, University of Massachusetts at Amherst. PhD.

Lombardi, L. (1995). "Laryngeal neutralization and syllable wellformedness." *Natural language and linguistic theory* 13(1): 39-74.

Maddieson, I. (1978). Universals of tone. *Universals of human language, vol. 2: Phonology*. J. H. Greenberg, Stanford University Press: 335-66.

Maddieson, I. (2008). Voicing in plosives and fricatives. *The world atlas of language structures online*. M. Haspelmath, M. Dryer, D. Gil and B. Comrie. Munich, Max Planck Digital Library, chapter 4.

Maran, L. R. (1971). Burmese and Jinghpo: a study of tonal linguistic processes. *Occasional papers of the Wolfenden Society of Tibeto-Burman linguistics*. Urbana, Illinois. 4.

Maran, L. R. (1973). On becoming a lone language: A Tibeto-Burman model of tonogenesis. *Consonant types and tone*. L. Hyman, Southern California Occasional Papers in Linguistics, Linguistics Program USC. LA. no 1.

Matisoff, J. (1973). Tonogenesis in Southeast Asia. *Consonant Types and Tone*. L. Hyman, Southern California Occasional Papers in Linguistics 1.

Matisoff, J. (1996). *Languages and dialects of Tibeto-Burman*. Berkeley, Center for Southeast Asia Studies, University of California.

Matisoff, J. (2003). *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan Reconstruction*. Berkeley, University of California Press.

Mazaudon, M. (1977). "Tibeto-Burman tonogenetics." *Linguistics of the Tibeto-Burman Area* 3(2): 1-123.

Mazaudon, M. (2005). On tone in Tamang and neighbouring languages: synchrony and diachrony. *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena. Historical Development, Tone-Syntax Interface and Descriptive Studies*. Tokyo, ILCAA, Tokyo University of Foreign Studies: 79-96.

Meredith, S. (1990). Issues in the phonology of prominence, Massachusetts Institute of Technology. PhD.

Nespor, M. and I. Vogel (1986). *Prosodic phonology*. Dordrecht, Foris.

Nolan, F. (2003). "Intonational equivalence: an experimental evaluation of pitch scales." *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*: 771-774.

Ohala, M. (2007). Experimental methods in the study of Hindi geminate consonants. *Experimental Approaches to Phonology*. M.-J. Sole, P. Beddor and M. Ohala. Oxford, Oxford University Press.

Pearce, M. (1998). "Consonants and tone in Kera (Chadic)." *Journal of West African Languages* 27.1: 33-70.

Pearce, M. (2005). "Kera tone and voicing." *Working Papers in Linguistics, University College London.*

Pearce, M. (2007). The interaction of tone with voicing and foot structure: Evidence from Kera phonetics and phonology, University College London. PhD.

Peperkamp, S. (1996). On the prosodic representation of clitics. *Interfaces in phonology, Studia Grammatica 41.* U. Kleinhenz. Berlin, Akademic Verlag.

Peperkamp, S. (1997). *Prosodic words.* The Hague, Holland Academic Graphics.

Qu, A. and X. Jin (1981). "Zangu fangyan de yanjiu fanga (Methods in the study of Tibetan dialects)." *Xinan Minyuan Xuebao* 3: 76-83.

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, Massachusetts, MIT publishing.

Selkirk, E. (1995). The prosodic structure of function words. *Papers in optimality theory.* J. Beckman, Urbanczyk and Walsh. Cambridge, Massachusetts, University of Massachusetts Occasional papers.

Shafer, R. (1966). *Introduction to Sino-Tibetan, part 1.* Wiesbaden, Otto Harrassowitz.

Shafer, R. (1974). *Introduction to Sino-Tibetan.* Wiesbaden, Otto Harrassowitz.

Silverman, D., B. Blankenship, et al. (1995). "Phonetic structures in Jalapa Mazatec." *Anthropological Linguistics* 37: 70-88.

Sprigg, R. K. (1972). "A polysystemic approach, in Proto-Tibetan reconstruction, to tone and syllable-initial consonant clusters." *Bulletin of the School of Oriental & African Studies, University of London* 35(3): 546-587.

Sprigg, R. K. (1993). "Controversy in the tonal analysis of Tibetan." *Bulletin of the School of Oriental & African Studies, University of London* 56(3).

Sun, J. (1997). The typology of tone in Tibetan. *Chinese languages and linguistics IV: Typological studies of languages in China.* Taipei, Academia Sinica

(Symposium series of the Institute of History and Philology, number 2): 485-521.

Sun, J. (2003). Variegated tonal developments in Tibetan. *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff.* B. Bradley, R. LaPolla, B. Michailovsky and G. Thurgood. Canberra, Australian National University, Pacific Linguistics.

Sun, J. and Y.-J. Lin (2002). On breathy voice in Qiuji Tibetan. *The eighth international symposium on Chinese languages and linguistics*. Taiwan, Academia Sinica.

Tan, K. (1987). "Xiarba zangyu de shengdiao xitong (Tonology of Xiarba Tibetan)." *Minzu Yuwen* 87.2: 22-28.

Tang, K. (2008). The Phonology and Phonetics of Consonant-Tone Interaction. Los Angeles, University of California. PhD.

Thurgood, G. (1985). Pronouns, verb agreement systems, and the subgroupings of Tibeto-Burman. *Linguistics of the Sino-Tibetan area: The state of the art. Papers presented to Paul K. Benedict for his 71st birthday*. G. Thurgood, J. Matisoff and D. Bradley. Canberra, Pacific Linguistics: 376-400.

Thurgood, G. (2007). Tonogenesis revisited: revising the model and the analysis. *Studies in Tai and Southeast Asian Linguistics*. J. Harris. Bangkok, Ek Phim Thai Co: 263-291.

Tournadre, N. and S. Dorje (2003). *Manual of Standard Tibetan*. Ithaca, Snow Lion Publications.

van Driem, G. (2001). *Languages of the Himalayas*. Leiden, Brill.

Vesalainen, O. and M. Vesalainen (1976). *Lhomi Phonemic Summary*. Kathmandu, Tribhuvan University, SIL and Institute of Nepal and Asian Studies.

Watkins, J. (2002). *The Phonetics of Wa*. Canberra, Pacific Linguistics.

Watters, S. (2002). "The sounds and tones of five Tibetan languages of the Himalayan region." *Linguistics of the Tibeto-Burman Area* 25(1).

Watters, S. (2003). An acoustic look at pitch in Lhomi. *Themes in Himalayan languages and linguistics*. Heidelberg: South-Asia Institute; Kathmandu: Tribhuvan University.

Wetzels, W. and J. Mascaro (2001). "The typology of voicing and devoicing." *Language* 77(2): 207-244.

Whalen, D. H. and A. G. Levitt (1995). "The universality of intrinsic F0 of vowels." *Journal of Phonetics* 23: 349-366.

Wylie, T. (1959). "A standard system of Tibetan transcription." *Harvard Journal of Asiatic Studies* 22: 261-267.

Xu, Y. (1997). "Contextual tonal variations in Mandarin." *Journal of phonetics* 27: 55-105.

Xu, Y. (1999). F0 peak delay: when and why it occurs. *International Congress of Phonetic Sciences 1999*. J. Ohala. San Francisco: 1881-4.

Xu, Y. and X. Sun (2002). "Maximum speed of pitch change and how it may relate to speech." *Journal of the Acoustical Society of America* 111: 1399-1413.

Yip, M. (1980). The tonal phonology of Chinese, MIT. PhD.

Yip, M. (2002). *Tone*. Cambridge, Cambridge University Press.

Zee, E. and L. Xu (2003). Shanghai. *The Sino-Tibetan languages*. G. Thurgood and R. LaPolla. London & New York, Routledge.