The Intonational Phonology of Indian English:
An Autosegmental-Metrical Analysis Based on Bengali and Kannada English

Olga Maxwell

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy
March 2014
School of Languages and Linguistics
The University of Melbourne
Abstract

Indian English (IndE) is one of the varieties of world Englishes. Despite its phonological system being a subject of ongoing interest, only a limited number of studies have examined IndE intonational features. This study presents the first detailed description of the intonational and prosodic features of the speech of eight educated IndE speakers from Bengali (Indo-Aryan) and Kannada (Dravidian) L1 backgrounds. Through a series of four experiments, prominence, phrasing, tune-text alignment and pitch range were investigated within an Autosegmental-Metrical framework of intonation analysis.

The similarities found between Bengali English (BE) and Kannada English (KE) confirm previous claims that IndE has developed its own phonology and show that its intonation bears certain resemblances to other well-established Englishes. IndE exhibits similarities with other Englishes in its prosodic structure, boundary tones inventory, accentual and focal prominence levels, presence of rising accent/s and a wide range of nuclear tunes. The findings of the research also reveal that IndE has a somewhat ‘hybrid’ system of intonation, most likely as a result of its development and use, and the influence of indigenous languages upon it. It was found that duration, amplitude and f0 are reliable cues to post-lexical prominence while vowel quality plays only a marginal role. Similarly, deaccenting is used alongside other focus marking strategies that are not typical for English intonation.

The findings also support previous experimental research that IndE is not uniform. Greater accentual density, a smaller pitch accent inventory, frequent use of nuclear falls and narrower pitch range are characteristic of KE, while a wide use of nuclear and prenuclear rising accents and an L*+H pitch accent in narrow focus are part of BE intonation. The differences within IndE as well as between IndE and well-established Englishes are exhibited in the use of phonological categories, their phonetic realisation and function.
Declaration

This is to certify that:

i. the thesis comprises only my original work towards the PhD,

ii. due acknowledgement has been made in the text to all other material used,

iii. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed

Olga Maxwell

March 2014
Acknowledgements

I would like to thank my supervisors, Associate Professor Janet Fletcher, Dr Celia Thompson and Dr Judith Rochecouste, for their invaluable advice, generous support and encouragement. Particularly, I would like to thank my main supervisor Janet Fletcher, whose expertise in the field of prosody and intonation, and whose patience and guidance had such an invaluable impact on my evolvement as a researcher.

I also would like to express my gratitude to the School of Languages and Linguistics and the Faculty of Arts for providing me with the opportunity to carry out this research and help with the funding to attend a number of international conferences. I received a great deal of support from my senior colleagues at the School and was given a valuable opportunity to talk about my research in formal and informal settings. I am also grateful to the participants of this study for their generosity and commitment.

I would like to thank everyone in the Phonetics Laboratory for creating a wonderful and collegiate environment, and making the process very enjoyable. I am especially grateful to Amos Teo, Debbie Loakes, Colleen Holt, Hywel Stoakes and Simone Graetzer for their support and friendship. I also thank Hywel Stoakes and Simone Graetzer for sharing some of the R scripts, which made the data analysis process more efficient.

I am much obliged to Sandy Clarke for her advice on statistical analyses and Caitilin Punshon for adding the polishing touches to the final document.

I would like to express extreme gratitude to my family, especially my husband, Aaron Maxwell, for patience and care. I would not have made it without their support. My parents’ encouragement to grow and work hard led me to where I am now.

I also thank my friends Judith Rochecouste, Julie Murray and Dana Chahal for looking at earlier drafts, providing feedback, giving encouragement and expressing their faith in my abilities. Last, but not the least, I thank my friend Laurie Paine for ongoing support and time spent listening to my worries and self-doubts.
To my mother
Table of Contents

List of Figures ................................................................................................................. xi
List of Tables ...................................................................................................................... xviii
Abbreviations ................................................................................................................... xx

Chapter 1: Introduction .................................................................................................... 1
  1.1 Context ......................................................................................................................... 1
  1.2 Definition of key terms ............................................................................................... 3
  1.3 Indian English ............................................................................................................. 4
    1.3.1 Indian English and world Englishes framework .................................................. 5
    1.3.2 Complexity and variation within IndE ................................................................. 6
  1.4 Research aims and design ......................................................................................... 8
  1.5 Outline of the study .................................................................................................... 10

Chapter 2: Literature review ............................................................................................ 11
  2.1 Intonational phonology within the Autosegmental-Metrical framework .............. 11
    2.1.1 Overview of the AM framework and its main components ............................... 11
    2.1.2 English intonation within the AM framework .................................................. 16
    2.1.3 Variation across English language varieties ...................................................... 19
  2.2 L2 prosody and intonation ......................................................................................... 23
    2.2.1 World Englishes and Second Language Acquisition ....................................... 24
    2.2.2 Phonological and phonetic influences on L2 prosody ..................................... 25
  2.3 Previous research on IndE prosodic features ........................................................... 30
    2.3.1 Word-level prominence and rhythm .................................................................. 31
    2.3.2 Intonational phonology ...................................................................................... 36
  2.4 Previous studies on the intonation of relevant Indian languages ......................... 40
    2.4.1 Profile of L1s in this study ................................................................................. 40
    2.4.2 Intonational aspects of Kannada and Bengali ..................................................... 42
    2.4.3 Features shared across Indian languages ........................................................... 46
  2.5 Research questions ................................................................................................... 50

Chapter 3: Methodology .................................................................................................. 52
  3.1 Introduction ................................................................................................................ 52
  3.2 Laboratory phonology ............................................................................................... 52
  3.3 Experimental design ................................................................................................ 53
    3.3.1 Participants ......................................................................................................... 53
    3.3.2 Corpus ............................................................................................................... 55
    3.3.3 Data collection and recording procedure ........................................................... 56
  3.4 Experiment 1: Intonational realisation of information structure ......................... 57
    3.4.1 Materials ............................................................................................................ 57
    3.4.2 Annotation, acoustic measurements and analysis ............................................. 59
    3.4.3 Statistical analyses ............................................................................................ 65
  3.5 Experiment 2: Tonal alignment and scaling characteristics of rising accents ....... 66
    3.5.1 Materials ............................................................................................................ 66
    3.5.2 Annotation, acoustic measurements and analysis ............................................. 68
List of Figures

Figure 2.1 A map of the states of India with the corresponding main official language. http://www.corfizz.com/language.html .................................................. 41

Figure 3.1 Example of annotation in EMU with sound wave, labels and f0 track. The green panel in the top left corner shows the corresponding annotation tiers. Speaker K2............. 60

Figure 3.2 Example of the hierarchy in EMU, showing each label in a tier linked to the corresponding label in the tier below. ................................................................. 62

Figure 3.3 Waveform display, f0 contour and annotations in EMU corresponding to the word Lulu produced by speaker B1. The L target corresponds to the first elbow in the f0 contour, while the H target corresponds the f0 peak. .................................................. 69

Figure 3.4 Example of annotation in EMU with sound wave, labels and f0 track. The green panel in the top left corner shows the corresponding annotation tiers. QC on the type tier stands for a polar question produced as one intonational phrase. The tonal categories are shown on the tone tier................................................................. 77

Figure 4.1 Illustration of accentuation on subject and object in broad focus structure, with a rising pitch on the nuclear accented word. The f0 track illustrates suspension of declination. Speaker B1.................................................. 96

Figure 4.2 Illustration of accentuation on subject in broad focus structure, with a rising pitch on the prenuclear accented word. Speaker B2.................................................. 96

Figure 4.3 Illustration of accentuation on subject and object in broad focus structure, with a rising pitch on prenuclear and nuclear accented words. The f0 track illustrates suspension of declination on the nuclear accented word. Speaker B3.............................. 96

Figure 4.4 Illustration of accentuation on subject, verb and object in broad focus structure, with a rising pitch on prenuclear accented words. The f0 track illustrates declination throughout the utterance. Speaker B4.................................................. 97

Figure 4.5 Illustration of accentuation on subject, verb and object in broad focus structure. Speaker K2. ........................................................................................................ 98

Figure 4.6 Illustration of accentuation on subject, verb and object in broad focus structure. The f0 track illustrates declination throughout the utterance. Speaker K1.......................... 98

Figure 4.7 Illustration of accentuation on subject and verb in broad focus structure, with a sustained high pitch or a ‘hat pattern’ on prenuclear accented words, followed by a sharp drop in f0 after the nuclear accent. Speaker K3.................................................. 98

Figure 4.8 Illustration of accentuation on subject and object in broad focus structure, with more or less sustained high pitch. Speaker K4. .................................................. 99

Figure 4.9 Bar plot for the proportional distribution of accented (blue) versus stressed (grey) syllables for each group................................................................. 100

Figure 4.10 Illustration of accentuation on the narrow focused subject followed by post-focal deaccenting. Speaker B3. W indicates a weak post-tonic syllable.............................. 102

Figure 4.11 Illustration of accentuation on the narrow focused verb accompanied by pre- and post-focal deaccenting. Speaker K3. .................................................. 102

Figure 4.12 Illustration of accentuation on narrow focused object. The f0 peak on the nuclear focal mill is raised relative to the f0 associated with the prenuclear accent. Speaker B1. ........................................................................................................ 103
Figure 4.13 Illustration of accentuation on narrow focused subject followed by another accent realised in compressed pitch. Speaker K2. ................................................................. 103

Figure 4.14 Illustration of accentuation on narrow focused verb followed by another accent realised in compressed pitch. Speaker B3. ................................................................. 104

Figure 4.15 Illustration of re-phrasing strategy in narrow focus structure. The focused word *Nina* is produced as a separate IP. Speaker B2. ................................................................. 105

Figure 4.16 Illustration of re-phrasing strategy in narrow focus structure. The focused word *borrow* is produced as an IP, separated by pauses from the rest of the utterance. Speaker B4. ................................................................. 105

Figure 4.17 Box and whiskers plot for pause duration in milliseconds grouped by speaker. Blue boxes correspond to BE speakers. Green boxes correspond to KE speakers. The thick horizontal line shows the median, while the extent of the whiskers indicate the range of distribution. ........................................................................ 106

Figure 4.18 Bar plot for the proportional distribution of the main focus marking strategies presented for each speaker. Green corresponds to deaccenting, grey correspond to re-phrasing and blue corresponds to pitch range compression without deaccenting. .......... 108

Figure 4.19 Box and whiskers plot for the distribution of *f0* height in semitones by prosodic position (S - subject, V - verb, O - object) presented by group (KE - top panel, BE - bottom panel). ........................................................................ 110

Figure 4.20 Plots showing mean *f0* values (semitones) in three accent conditions (stressed, accented and focally accented), presented by speaker and averaged across the prosodic positions. Values for BE speakers are shown to the left in dark blue, values for KE speakers are shown to the right in dark green. ........................................................................ 111

Figure 4.21 Box and whiskers plot for the distribution of *f0* excursion in semitones by prosodic position (S - subject, V - verb, O - object) presented by group (KE - top panel, BE - bottom panel). ................................................................. 112

Figure 4.22 Box and whiskers plot for the distribution of dB-RMS values by prosodic position (S - subject, V - verb, O - object) presented by group (KE - top panel, BE - bottom panel). ................................................................. 113

Figure 4.23 Plots showing mean dB-RMS values in three accent conditions (stressed, accented and focally accented) averaged across the prosodic positions, presented for each group with BE speakers to the left and KE speakers to the right. ......................... 114

Figure 4.24 Box and whiskers plot for the distribution of syllable duration in ms by word structure (1S - monosyllabic, 2S - bisyllabic, 3S - trisyllabic) across three conditions (S - stressed, A - accented, F - focal), presented by group (KE - top panel, BE - bottom panel). ................................................................. 115

Figure 4.25 Box and whiskers plot for the distribution of *F1* values in Hz by vowel type presented by group (BE - top panel, KE - bottom panel) and condition (S - stressed, A - accented, F - focal). I - /ɻ/, E - /ɛ/, O - /ɔ/, u - /u(ː)/, ai - /au/, i - /iː/. .................................................. 118

Figure 4.26 Box and whiskers plot for the distribution of *F2* values in Hz by vowel type presented by group (BE - top panel, KE - bottom panel) condition (S - stressed, A - accented, F - focal). I - /ɻ/, E - /ɛ/, O - /ɔ/, u - /u(ː)/, ai - /au/, i - /iː/. .................................................. 120

Figure 4.27 Vowels /ɻ/, /ɛ/, /ɔ/ and /u(ː)/ for speaker B1 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz. .................................................. 122
Figure 4.28 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker B2 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz..............................................122

Figure 4.29 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker B3 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values. ...............................................122

Figure 4.30 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker B4 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values Hz..............................................123

Figure 4.31 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker K1 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values Hz.................................123

Figure 4.32 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker K2 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in two stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values Hz..............................................123

Figure 4.33 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker K3 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values Hz..............................................123

Figure 4.34 Vowels /ɪ/, /ɛ/, /ʊ/ and /ʊ(ɨ)/ for speaker K4 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values Hz..............................................123

Figure 5.1 Schematic representation of the phonetic realisation of L+H* and L*+H. The grey boxes represent the accented syllable, while the solid line represents the f0 movement corresponding to the rising gesture.................................................................138

Figure 5.2 Schematic representation of main measurements, shown relative to the f0 movement and the segmental landmarks.................................................................149

Figure 5.3 Box plot for the distribution of proportional ContoL distance as a function of L1 group (BE versus KE), presented by target word....................................................151

Figure 5.4 Mean duration of nasals /m/ and /n/ in milliseconds, presented by group..........151

Figure 5.5 Box plot for the distribution of proportional VontoL distance as a function of L1 group (BE versus KE) grouped by target word....................................................152

Figure 5.6 Box plot for the distribution of proportional peak delay (ContoH) as a function of four KE speakers, presented by word.................................................................153

Figure 5.7 Box plot for the distribution of proportional peak delay (ContoH) as a function of four BE speakers by word.................................................................153

Figure 5.8 Mean duration of the target vowels in milliseconds produced by BE (blue dotted line) and KE (green solid line) speakers. The labels on the x-axis follow standard phonetic-acoustic labelling criteria (Croot & Taylor, 1995) and correspond to the following vowels: A – /æ/ a: – /æ:/ E – /ɛ/ u(ı) – /u(ı)/..................................................155

Figure 5.9 Scatter plot for mean peak relative to word offset and mean accented syllable duration (HtoWoff), both in milliseconds..................................................156

Figure 5.10 Scatter plot for mean peak delay (ContoH) and mean accented syllable duration, both in milliseconds, presented by L1 group..................................................156
Figure 5.11 Box plot for the distribution of proportional rise duration as a function of L1 group (BE versus KE) by target word. .......................................................... 156
Figure 5.12 Scatter plot for mean peak delay (ContoH) and mean low delay (ContoL) in milliseconds, presented by L1 group. .......................................................... 156
Figure 5.13 Box plot for the distribution of mean L scaling in semitones, presented by word for each group. .......................................................... 158
Figure 5.14 Mean f0 for L scaling (blue dashed line) relative to the highest f0 (black dashed line) and the lowest f0 (black solid line) presented by speaker. .................................................. 159
Figure 5.15 Box plot for the distribution of mean H scaling in semitones, presented by word for each group. .................................................................................. 159
Figure 5.16 Box plot for the distribution of rise magnitude in semitones, presented by word for each group. .................................................................................. 160
Figure 5.17 Scatter plot for mean rise duration and mean rise magnitude by L1 group........ 161
Figure 5.18 Box plot for the distribution of proportional mean ContoL distance presented by speaker. .................................................................................. 164
Figure 5.19 Mean duration of /l/ and /m/ in milliseconds presented by speaker. .............. 164
Figure 5.20 Box plot for the distribution of proportional mean peak delay as a function of L1 group (BE versus KE). .......................................................... 165
Figure 5.21 Scatter plot of mean peak delay and accented syllable duration, both in milliseconds, presented by L1 group. .......................................................... 166
Figure 5.22 Box plot for the distribution of proportional mean rise duration as a function of L1 group (BE versus KE) grouped by target word. ...................... 167
Figure 5.23 Scatter plot of mean peak delay (ContoH) and mean low delay (ContoL) distance, both in milliseconds, presented by group. ...................... 168
Figure 5.24 Scatter plot of mean rise duration in milliseconds and mean rise magnitude in semitones, presented by group. ................................................ 168
Figure 5.25 Box plot for the distribution of proportional mean ContoL as a function of L1 group (BE versus KE) and prosodic position. .................................................. 170
Figure 5.26 Box plot for the distribution of proportional mean peak delay as a function of L1 group (BE versus KE) and prosodic position. .................................................. 170
Figure 5.27 Box plot for the distribution of mean accented syllable duration as a function of L1 group (BE versus KE) and prosodic position. .................................................. 170
Figure 5.28 Box plot for the distribution of proportional mean rise duration as a function of L1 group (BE versus KE) and prosodic position. .................................................. 171
Figure 5.29 Box plot for the distribution of mean L scaling in semitones, presented by group and prosodic position. Blue boxes correspond to nuclear focal, green boxes correspond to prenuclear accented words. .................................................. 172
Figure 5.30 Box plot for the distribution of mean H scaling in semitones, presented by group and prosodic position. Blue boxes correspond to nuclear focal, green boxes correspond to prenuclear accented words. .................................................. 172
Figure 5.31 Box plot for the distribution of mean rise magnitude as a function of L1 group (BE versus KE) and prosodic position. Blue boxes correspond to nuclear focal, green boxes correspond to prenuclear accented words. .................................................. 172
Figure 6.1 Box and whiskers plot for the distribution of final lengthening by boundary type (word - W, ip - intermediate phrase, IP- intonational phrase), presented by group. Data points shown in the y-axis are presented in milliseconds. 

Figure 7.1 Prosodic hierarchy with the associated intonational events, illustrated for the utterance Lara lives in Lilydale. The utterance is produced as one intermediate and one intonational phrase with pitch accents on Lara and Lilydale. The intonational structure of the utterance using ToBI annotation is L+H* L*+!H L-L%. 

Figure 7.2 f0 contour and annotations of the utterance Lara lives in Lilydale produced by speaker B1 with broad focus structure. 

Figure 7.3 Intonational contour of a question produced by speaker B2 showing the occurrence of an H* accent on the prenuclear accented words will and mellow. 

Figure 7.4 Two intonational contours illustrating an H* and an !H* pitch accents produced by speaker K3. 

Figure 7.5 Illustration of a delayed high pitch accent in nuclear position in a declarative produced by speaker K4. 

Figure 7.6 Illustration of an L+H* accent in a complex declarative produced as a series of IPs, each starting with a rising accent. Speaker B4. 

Figure 7.7 Illustration of an L+H* accent occurring on the nuclear focal word really, produced by speaker K2 in spontaneous speech. 

Figure 7.8 An illustration of an L*+H in nuclear focal position (borrow), produced by speaker B2. 

Figure 7.9 Illustration of an L* accent in a polar question produced by speaker K4. 

Figure 7.10 Proportional distribution of pitch accents for each speaker. The types of pitch accents are colour-coded: H* - light green, L* - dark green, L+H* - light blue, L*+H - dark blue. 

Figure 7.11 Proportional distribution of accented (blue) versus stressed (grey) syllables for BE to the left and KE to the right. 

Figure 7.12 Illustration of a wh-question produced by speaker B1, showing the occurrence of an L- phrase accent. 

Figure 7.13 Illustration of the occurrence of an L- phrase accent at the end of a coordination question produced by speaker K1. 

Figure 7.14 An illustration of the intonational contour of a polar question produced by speaker K4, showing two occurrences of an H- phrase accent at the end of each IP. 

Figure 7.15 An illustration of an H- before the clause, produced by speaker B1 in the complex declarative. 

Figure 7.16 Illustration of the occurrence of an !H- after a bitonal pitch accent in the first IP of a question produced by speaker B1. 

Figure 7.17 Illustration of the occurrence of an L-L% boundary configuration with different nuclear accents in each IP. Speaker B2. 

Figure 7.18 Illustration of the occurrence of an L-H% boundary configuration following a delayed rising accent. Speaker B2. The word Nina is in narrow focus. 

Figure 7.19 Illustration of the occurrence of an H-H% boundary configuration in combination with L* accent. Speaker B1. 

Figure 7.20 Illustration of the occurrence of an H-L% boundary configuration in the first IP. Speaker B4.
Figure 7.21 Proportional distribution of the tunes in declaratives for BE (left) and KE (right). Colour codes with corresponding tune types are listed below the bar plots.........................215

Figure 7.22 Illustration of the occurrence of a low rising tune (L* H-H%) in the first and the second IP. Speaker B3. .....................................................................................215

Figure 7.23 Illustration of the occurrence of a falling tune (H* L-L%) in the second IP. Speaker K1. ..................................................................................................................216

Figure 7.24 Illustration of a rising-falling tune (L+H* L-L%), with the word move in narrow focus. Speaker K4. ........................................................................................................217

Figure 7.25 Illustration of the occurrence of a delayed rising-falling tune (L*+H L-L%) with the word borrow in narrow focus. Speaker B2. .................................................................218

Figure 7.26 Proportional distribution of nuclear tunes in polar questions for BE (left) and KE (right). .........................................................................................................................219

Figure 7.27 Illustration of the occurrence of a low rising tune (L* H-H%). The f0 value of the H% tone is 320 Hz. Speaker B1. ....................................................................................219

Figure 7.28 Illustration of the occurrence of a high rising tune (H* H-H%). Speaker K2. ...220

Figure 7.29 Illustration of the occurrence of a falling tune (!H L-L%), with the nuclear !H* accent. Speaker K4. ............................................................220

Figure 7.30 Illustration of the occurrence of a rising-falling tune (L+H* L-L%) in a polar question. Speaker B4. ..........................................................221

Figure 7.31 Proportional distribution of nuclear tunes in wh-questions for BE (left) and KE (right). LH in the bar plot to the left (LH L-L%) corresponds to the use of both L*+H and L+H* accents in nuclear position. ......................................................222

Figure 7.32 Illustration of the occurrence of a delayed rising-falling tune with a downstepped nuclear accent (L+!H* L-L%) in a wh-question. Speaker B1. .................................222

Figure 7.33 Illustration of the occurrence of a falling tune (!H* L-L%) with a nuclear !H* in a wh-question. Speaker K3. .................................................................223

Figure 7.34 Proportional distribution of L+H* (rising - blue) and H* (high - green) on wh-words. ..............................................................223

Figure 7.35 Box plot showing the distribution, outliers and median of the f0 peak on the first accented word in declaratives (D), polar questions (PQ) and wh-questions (WhQ), presented by group. BE boxes are labelled in blue, KE boxes - in green. The f0 values on the y-axis are shown in semitones. ...............................................................224

Figure 7.36 Illustration of the occurrence of a low-to-mid rising tune (L* L-H%) in the first IP of the coordination question. Speaker K1. .................................................226

Figure 7.37 Illustration of the occurrence of a falling tune (H* L-L%) in the second IP of the coordination question. Speaker B2 ..................................................227

Figure 7.38 Illustration of the occurrence of a rising-falling tune (L+H* L-L%) in the second IP of the coordination question. Speaker B3. ..................................................227

Figure 7.39 Illustrations of the occurrence of a downstepped fall (!H* L-L%) in the second IP, corresponding to the embedded imperative. Speaker K3. ........................................228

Figure 7.40 Proportional distribution of nuclear tunes in BE imperatives........................................228

Figure 7.41 Illustrations of two tune types in imperatives. The sentence on the left shows the occurrence of !H* L-L% (speaker B2) and the second sentence on the right shows the occurrence of L+H* L-L% (speaker B3)........229
Figure 7.42 Box plot showing the distribution, outliers and median of the $f_0$ peak on the first accented word in declaratives (D) and imperatives (I), presented by group. BE results are shown in blue, while KE results are in green. The $f_0$ values on the y-axis are presented in semitones.
List of Tables

Table 2.1 Parameters within the AM framework with corresponding research questions........50
Table 3.1 Detailed information about the participants, showing speakers’ L1, other languages
spoken, city and state of residence in India, age of learning English, languages used
most of the time, and length of residency in Australia.............................................54
Table 3.2 Prompt questions and stimuli sentences with the target words underlined to elicit
narrow focus on subject, verb and object, and broad focus structures.........................57
Table 3.3 Target words with their corresponding phonemic transcription (Maxwell &
Fletcher, 2009, 2010) used in Experiment 1..........................................................59
Table 3.4 Summary of the phonetic correlates with the corresponding measurements and
description investigated in Experiment 1..............................................................64
Table 3.5 Phonemic transcription of the target words used in Experiment 2.......................67
Table 3.6 Number of stimuli recorded and analysed, presented by syntactic structure type..75
Table 5.1 Hypotheses with corresponding analyses or predictions.................................148
Table 5.2 Rising pitch accented categories presented for each group by prosodic position..174
Table 7.1 Tonal categories in ToBI, their associations and general pitch description proposed
for the intonational phonologies of BE and KE....................................................195
Table 7.2 Major nuclear tunes presented as tune type, ToBI annotational categories, schematic
contour and contexts occurring in BE and KE.......................................................212
Table B.1 LMM results for the absolute f0 height in stressed, accented and focal syllables.277
Table B.2 Mean and standard deviation of the absolute f0 values (Hz) in focal versus accented
syllables presented in three prosodic positions (subject, verb, object) for each speaker.
Shaded grey areas indicate that the differences were non-significant..........................277
Table B.3 LMM results for the comparison of pitch excursion in focal and accented syllables.
........................................................................................................................................278
Table B.4 LMM results for the RMS-amplitude measurements in stressed, accented and focal
syllables .......................................................................................................................278
Table B.5 LMM results for syllable duration in stressed, accented and focal syllables. ......279
Table B.6 LMM results for F1 measurements in stressed, accented and focal vowels............279
Table B.7 LMM results for F2 measurements in stressed, accented and focal vowels............280
Table C.1 LMM results for the temporal interval from syllable onset to L (ContoL) in
prenuclear rises........................................................................................................281
Table C.2 LMM results for the temporal interval from vowel onset to L (VontoL) in
prenuclear rises........................................................................................................281
Table C.3 LMM results for the temporal interval from syllable onset to H (ContoH ) in
prenuclear rises........................................................................................................282
Table C.4 LMM results for the temporal interval from H to word offset (WofftoH) in
prenuclear rises........................................................................................................282
Table C.5 LMM results for the distance between L and H (LtoH) in prenuclear rises........282
Table C.6 LMM results for L scaling in prenuclear rises..................................................283
Table C.7 LMM results for H scaling in prenuclear rises..................................................283
Table C.8 LMM results for the scaling difference between L and H in prenuclear rises. .....283
Table C.9 LMM results for the temporal interval from syllable onset to L (ContoL) in nuclear focal rises. ...................................................................284
Table C.10 LMM results for the temporal interval from syllable onset to H (ContoH) in nuclear focal rises. ...................................................................284
Table C.11 LMM results for the temporal interval from H to word offset (WofftoH) in nuclear focal rises........................................................................................................284
Table C.12 LMM results for the distance between L and H (LtoH) in nuclear focal rises...285
Table C.13 LMM results for the temporal interval between syllable onset and L (ContoL) in prenuclear versus nuclear focal rises. ...........................................................................285
Table C.14 LMM results for the temporal interval between syllable onset and H (ContoH) in prenuclear versus nuclear focal rises. ...........................................................................285
Table C.15 LMM results for accented syllable duration in prenuclear versus nuclear focal rises. ......................................................................................................................286
Table C.16 LMM results for the distance between L and H (LtoH) in prenuclear versus nuclear focal rises. ........................................................................................................286
Table C.17 LMM results for L scaling in prenuclear versus nuclear focal rises. ......................286
Table C.18 LMM results for H scaling in prenuclear versus nuclear focal rises. ......................287
Table C.19 LMM results for the scaling difference between L and H in prenuclear versus nuclear focal rises. ........................................................................................................287
Table D.1 LMM results for final syllable duration across three boundary types (W - word, ip - intermediate phrase and IP - intonation phrase). ..............................................288
Table E.1 BE and KE tonal categories and total number of observations for each category, based on ToBI annotations in Experiment 4. .................................................................289
Table E.2 LMM results for the f0 peak scaling in the first accented word across three sentence types (D - declaratives, WhQ - wh-questions and PQ - polar questions). .........................289
Table E.3 LMM results for the f0 peak scaling in the first accented word in two sentence types (D - declaratives, I - imperative). ..................................................................................289
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AmE</td>
<td>American English</td>
</tr>
<tr>
<td>AssamE</td>
<td>Assamese English</td>
</tr>
<tr>
<td>AusE</td>
<td>Australian English</td>
</tr>
<tr>
<td>BE</td>
<td>Bengali English</td>
</tr>
<tr>
<td>BISAfE</td>
<td>Black South African English</td>
</tr>
<tr>
<td>BrE</td>
<td>British English</td>
</tr>
<tr>
<td>GE</td>
<td>Gujarati English</td>
</tr>
<tr>
<td>HE</td>
<td>Hindi English</td>
</tr>
<tr>
<td>IndE</td>
<td>Indian English</td>
</tr>
<tr>
<td>KE</td>
<td>Kannada English</td>
</tr>
<tr>
<td>MAE</td>
<td>Mainstream American English</td>
</tr>
<tr>
<td>NE</td>
<td>new Englishes</td>
</tr>
<tr>
<td>NigE</td>
<td>Nigerian English</td>
</tr>
<tr>
<td>NZE</td>
<td>New Zealand English</td>
</tr>
<tr>
<td>RP</td>
<td>Received Pronunciation</td>
</tr>
<tr>
<td>SBE</td>
<td>Southern British English</td>
</tr>
<tr>
<td>SLA</td>
<td>Second Language Acquisition</td>
</tr>
<tr>
<td>SSE</td>
<td>Scottish Standard English</td>
</tr>
<tr>
<td>TE</td>
<td>Tamil English</td>
</tr>
<tr>
<td>TBE</td>
<td>Tripura Bangla English</td>
</tr>
<tr>
<td>TelE</td>
<td>Telugu English</td>
</tr>
<tr>
<td>ToBI</td>
<td>Tones and Break Indices</td>
</tr>
<tr>
<td>WE</td>
<td>world Englishes</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Context

Research suggests that despite some intonational universals, intonation can vary significantly across languages and language varieties (Ladd, 2008; Grice & Baumann, 2007). To date, the literature includes very detailed descriptions of the intonation systems of well-established English language varieties such as British English (BrE), American English (AmE) and Australian English (AusE). In recent years, with the rapidly increasing number of second language (L2) speakers of English around the world and, especially, the recognition of several varieties within the world Englishes framework, researchers have turned their interest to the prosodic and intonational features of relatively new varieties (Setter, 2006 on Hong Kong English; Detering, 2001 on Singaporean English; Simo Bobda, 2004 on Cameroon English; Gut, 2005 on Nigerian English; Zerbian, 2013 on South African English) and found strong interference from first languages (L1) in the prosodic structure of these varieties.

Indian English (IndE) takes a special place among the newer varieties of English around the world. The use of the English language in the subcontinent dates back to the 16th century and has developed its own distinct grammatical, lexical, stylistic and phonological features. In addition, this is one of the most widely spoken varieties of English, which has spread across several continents including large diasporas in Australia, Canada, Britain and other countries. Despite substantial interest in IndE phonological features, there has been very little systematic research into the intonation of this variety. Conclusions have often been made on the basis of earlier impressionistic studies, where the intonation patterns were described as the use of rise and fall in different types of utterances or ‘faulty’ chunking of an utterance into tone groups (Bansal, 1990; Latha, 1978 on Malayalam English).

Acoustic phonetic studies mostly focused on such aspects as lexical prominence or reported preliminary results on the use of phrasal prominence and pitch movements on accented words (Pickering & Wiltshire, 2000; Wiltshire & Moon, 2003; Wiltshire & Harnsberger, 2006). The Wiltshire and Harnsberger study (2006), investigating the differences and similarities in intonation patterns across IndE speakers of Gujarati and Tamil L1 backgrounds, confirmed a substantial degree of L1 influence but also showed that some of the features could be shared by speakers from
different L1 backgrounds. The source of the similar features could not be determined. They could have been the result of similarities in the intonational phonologies within one language family, the nature and use of English in the subcontinent, or even the effect of transfer from the substratum language, Hindi. These findings indicate the need for more detailed and systematic research in this field in order to describe IndE intonation and, most importantly, the degree of variation within this variety of English.

It is important to take into account the role of intonation in communication, speaker identity and discourse function. Speakers usually rely on certain norms acceptable in their variety and, despite the robustness of variation in intonation patterns (Grabe, Kochanski & Coleman, 2005), will not necessarily have an understanding or tolerance of the patterns more common in another variety, especially if these exhibit strong influence from a language they are not familiar with. The use of inappropriate or ‘unconventional’ intonation patterns may lead to a number of misunderstandings (Mennen, 2007). Gumperz (1982) showed how differences in the use of phrasal prominence and the mismatch in the use of final falls and rises contributed to misinterpretation of intended messages between speakers of IndE and speakers of AmE and BrE, resulting in negative attitudes between the groups. Wiltshire and Harnsberger (2006, p. 102) also raised an issue of “intelligibility problems in communication between IndE speakers and those of other English dialects” due to the differences in IndE.

There are large Indian communities across various geographical locations in Australia, and there are a number of domains where speakers of IndE interact with speakers of AusE. The Australian medical system, for example, relies heavily on the recruitment of doctors from overseas to fill the demand in the public health sector, and India is one of the leading countries for these doctors. For example, in 2003, the NorthWestern Mental Health network in Victoria had around 38% of its doctors recruited from India and the numbers were growing (Barton, Hawthorne, Singh & Little, 2003). In clinical settings, misunderstandings due to the use of intonation could be attributed to an inadequate level of linguistic or even professional skills, and may be seen to hinder the data gathering process or cause problems in establishing rapport with a patient. A more systematic investigation into the intonation patterns of IndE will help shed more light on a range of features responsible for miscommunication in cross-dialectal contexts and could, in the future, contribute to developing training
programs and materials in medical or other contexts.

1.2 Definition of key terms

‘Prosody’ or ‘prosodic features’ are quite often used synonymously with ‘suprasegmental features’; however, they describe different properties of a language. The term ‘suprasegmental features’ includes the phonetic features that span over more than one segment and include such aspects as stress, tone, melodic patterns, rhythm and tempo. These phonetic properties traditionally correspond to duration, fundamental frequency ($f_0$) and intensity (Ladd, 2008). Suprasegmental effects, however, cannot always be separated from the production of individual sounds, and the differences in these aspects can be due to the different aspects of the prosodic structure of a language. Therefore, prosody can be defined as “a set of higher-level organisation structures that account for variations in pitch, loudness, duration, spectral tilt, segment reduction and their associated articulatory parameters” (Clark, Yallop & Fletcher, 2007, p.327). Prosody refers to the phonological structure of a language and is concerned with aspects such as metrical structure, prominence relations and levels of phrasing.

As noted by Gut, Trouvain and Barry (2007), intonation can be used in theoretical research with a differing scope. In a narrow definition, ‘intonation’ refers to pitch modulation or “an ensemble of pitch variations in the course of an utterance” (‘t Hart, Collier & Cohen, 1990, p.10). This modulation of pitch is described as a sequence of rises and falls which correspond to tonal patterns on a phrasal level (Arvaniti, 2011). In other words, intonation functions beyond segments and words, and is concerned with sentence- or ‘post-lexical’-level tonal events. These tonal events are linguistically structured and are used to convey pragmatic meanings in a language (Ladd, 2008), depending on its inherent phonological structure. The linguistic form of intonation is created with the help of categorically distinct entities and relations between strong and weak elements. The assumption is that each language has a set of pitch patterns and tones which correspond to discrete meanings, with an underlying relationship between linguistic form and function (Gut, 2009).

In addition to its linguistic function, intonation can also be used to project such paralinguistic phenomena as a speaker’s emotional states and basic aspects of interpersonal interaction. In some of the literature, a broad definition of ‘intonation’
covers both linguistic and paralinguistic features which correspond to the physical parameters of tempo, voice quality and loudness (Fox, 2001). Intonation, together with syntax, phrasing, semantics and pragmatics, contributes to the interpretation of an utterance and helps in structuring interaction (Wichmann, 2000). It carries pragmatic and communicative functions, conveys attitudes and sentence modality (such as questions versus statements), helps project information structure and focus, indicates continuity or finality, and is used for phrasing and discourse segmentation (Himmelmann & Ladd, 2008; Cruttenden, 1997). The distinction between linguistic and paralinguistic, however, is not always straightforward, in that paralinguistic and intonational features often interact, and paralanguage is difficult to distinguish from the linguistic aspects of intonation (Ladd, 2008; House, 2006). Moreover, the same linguistic form may correspond to different paralinguistic meanings across languages (Ladd, 2008).

The definition adopted in this study focuses on the linguistic aspects of intonation and the post-lexical tonal events associated with phonologically specified landmarks of an utterance; that is, prominent syllables or prosodic boundaries. In addition, the term intonation will be used to refer primarily to $f_0$, the physical correlate of perceived pitch created by the rate of vibration of the vocal folds.

The term ‘intonational phonology’ was coined by Ladd (1996) and simply refers to a set of phenomena and not a single approach or school of thought. There are several approaches to analysis in intonational phonology. These depend on different traditions, the focus or the aspects under investigation, and the type of data obtained (auditory, perceptual or acoustic). Intonational phonology usually refers to post-lexical phonology and phonetic realisation of intonational events (Ladd 2008). In this dissertation, the term is applied to describe a number of intonational categories of a language or its varieties.

1.3 Indian English
This section defines the concepts ‘world Englishes’ and ‘Indian English’ and provides some background information on English in India, including its sociolinguistic variation, development and use. It also provides an explanation for the treatment of IndE as a complex multifaceted variety and the need for further research on the
phonological features within IndE in order to expand our knowledge about the ‘range and depth’ of variation (Kachru, 1986).

1.3.1 Indian English and world Englishes framework

Braj Kachru (1982, 1986) introduced the concept of ‘world Englishes’ (WE) in the late 1970s and early 1980s. The term was subsequently used to propose a framework that would include multiple different models of the English language that can be used across different cultures and countries (Kachru, 1992). The WE framework not only includes such well-known varieties as British or American English, spoken in countries where English originated and is spoken as an L1, but also accommodates newer L1 and L2 varieties that have emerged in various former British or American colonies (such as Singapore, the Philippines, India, Pakistan, Sri Lanka, Nigeria) and other countries around the world where English has been institutionalised and now plays an important part in education, administration, the media and the economy. The notion of WE implies inclusivity and does not privilege one variety over another. The main goal is to create a framework for better understanding of the use and functions of the English language reflective of its globalisation (Kachru & Nelson, 2006). Research into WE has grown since the introduction of the concept, generating many publications and leading to the formation of the International Association of World Englishes.

Indian English belongs to the group of WE varieties under the category of South Asian Englishes (Kachru & Nelson, 2006). It is estimated that there are over thirty million proficient speakers of English in India (Melchers & Shaw, 2003), and it can be used as first1 or second language. The number of L1 speakers of English in India, however, is estimated to be a relatively small proportion of English speakers overall. In addition, the majority of the population is either bilingual or multilingual. As a result, English may constitute one of two or more languages spoken.

After Independence in 1947 and a gradual decline in the prestige of Received Pronunciation (RP), there was an attempt to move away from the RP of BrE and replace the RP model with a native variety of English that would be more attainable and more practical for education and communication purposes (Pandey, 1981, 1994).

---

1 Wells (1982) and Coelho (1997) use the term Anglo-Indian English to refer to the English spoken as an L1 by the descendants of Anglo-Indian lineage.
The term General or Generalised Indian English was introduced to describe a variety of English that was taught in schools and universities (Bansal, 1969, 1976; CIEFL, 1972). As a result, ‘General’ (Wells, 1982) or ‘Standard’ (Kachru, 1983) Indian English is the term applied to the language spoken by educated Indians as well as in the Indian diaspora (Bansal, 1990). It is considered to be more ‘prestigious’ and different from a number of other varieties - for example, Babu English (Valentine, 2001; Kachru, 2005) or Butler English - which are traditionally associated with the spoken language used by low socioeconomic groups with limited or no education. Today, the word ‘General’ or ‘Standard’ is often omitted, resulting in a more common use of ‘Indian English’. In the last few decades, IndE has been used as a cover term for a number of varieties of English spoken in India and has become widely recognised both in the literature and more recently among the population in the subcontinent (Gupta, 2001).

1.3.2 Complexity and variation within IndE

Among the new varieties of world Englishes, IndE has had the longest history of use, with the numbers of its speakers constantly growing. English was first introduced into India by the British in the 16th century and it has maintained its power since the country’s independence. It has become one of India’s official languages. In the 1960s, the introduction of the Three Language Formula as part of the pre-university educational system (Gargesh, 2006; Melchers & Shaw, 2003) further integrated English into the educational system, subsequently contributing to its status and development. In the Three-Language-Formula approach, English is taught alongside the mother tongue of the state or the regional standard and Hindi or, in some Hindi-speaking states, another Indian language. At present, English is not only the language of education and administration but is also used in the media, business, religion, literary writing, informal communication, intimate contexts and more (D’souza, 2001).

Despite its widespread usage, Schneider (2007) defined IndE as a rather complex, elusive and problematic variety. This could be due to the interplay of several factors. Firstly, IndE is not always used for informal conversations with native speakers of well-established varieties of English (such as AmE or BrE). English is mostly spoken as a lingua franca, whereby the speakers of one language community
use English to interact with the speakers from another community whose mother
tongue is different from their own. In this communication, speakers from different
groups need to find common ground and understanding and are therefore not concerned with linguistic forms and pronunciation norms as long as the spoken
English brings successful interaction. Second despite the accessibility of American
and British media in India, English is taught to speakers of IndE by other speakers of
IndE. Therefore, it is most likely to be acquired as the ‘nativised’ or ‘indiginised’
(Sridhar & Sridhar, 1992) variety influenced by the indigenous languages spoken in
the subcontinent. Wiltshire and Harnsberger (2006) referred to IndE as a ‘transplanted
variety’, where the native-like system (for example, BrE) has not been fully acquired.
It is evident that this variety has developed distinct grammatical, phonological, lexical and discourse features which are mutually intelligible between the speakers of
IndE from various regions and L1 backgrounds (Bansal, 1990; Wells, 1982; Nihalani,
Tongue, Hosali & Crowther, 2004).

Third, the linguistic profile of India is characterised by a great number of
languages and dialects. Each state may have speakers of up to 12 indigenous
languages (Choudhry, 2003). The phonology of IndE could be closer to these native
languages than to any other longstanding varieties, such as RP, which could explain
the large degree of variation across its phonological features. This variation has been
reflected in the literature and has generated a number of studies investigating its
nature and range. A number of earlier studies defined regional varieties mostly based
on L1: for example, Malayalam English (Latha, 1978; Gopalakkrishnan, 1996),
Hindustani English (Pandey, 1980), Punjabi English (Sethi, 1980) or Telugu English
(Prabhakar Babu, 1971). However, the role of L1 influence and the definition of sub-
varieties cannot be applied in a straightforward manner. Being part of South Asia,
India belongs to a linguistic area (Emenau, 1956; Masica, 1976, 2005) where many
typologically different languages share a number of common features as a result of
language contact (such as retroflex consonants in Indo-Aryan and Dravidian
languages). According to Gargesh (2006), present day India has a high percentage of

---

2 Except for the Hindi belt area, where Hindi or Hindustani is often used as lingua franca among various communities.

3 The issue of whether IndE has developed its own phonological system as a variety remains ongoing in current experimental research (Wiltshire & Harnsberger, 2006; Maxwell & Fletcher, 2010; Sirsa & Redford, 2013).
multilingualism where all major languages exist beyond their home territory, and border areas show ‘a state of diffusion’, resulting in contact-induced convergence.

In spite of the indication in the literature that we are not dealing with a single variety that is uniform across all language groups, the task of mapping this variation remains a central issue in current research. In her study on the English language spoken by Tibeto-Burman speakers, Wiltshire (2005) found a number of similarities in the segmental phonology of the speakers from three L1s belonging to the same language family. She raised the question as to whether the variation should be defined on the basis of the speakers’ language family or whether it is more fine-grained and can be applied on the basis of the speakers’ L1, the model taught or other factors. One of the predisposing criteria could be geographical location (Bansal, 1970 on Uttar-Pradesh English; Dhamija, 1976 on Rajasthani English), whereby the speakers of IndE from the same region exhibit a number of shared features despite different L1s. Maxwell and Fletcher (2009, 2010), for example, found a number of similarities in the production of vowels in English spoken by L1 Hindi and Punjabi speakers who came from an area of close relative proximity to Delhi. Sirsa and Redford (2013) suggested that social and regional factors could be the driving force for variation within IndE, leading to the development of multiple varieties.

It is also important to take into account such factors as degree of exposure and use, proficiency, educational background, and belonging to a particular social, ethnic or religious group. This complexity poses a number of challenges for researchers, and the answers can be obtained though more experimental studies using speakers of IndE from various linguistic, regional, ethnic and social communities. In addition, there is a continuum of proficiency levels across IndE (Kachru, 2005). As noted by Kachru (1983), this is a ‘nativised’ hybrid with a mix of various developmental features as well as L1 influence. On the basis of these factors, the educational, ethnic, social and educational background of the participants of this research was kept relatively stable.

1.4 Research aims and design

This study addresses a gap in research on intonation in IndE and investigates several intonational features and prosodic structure of this variety. In order to present a detailed description of IndE intonation and examine variation, the data set is based on the speech of educated speakers from two typologically distinct language groups,
Bengali and Kannada. In this study, the terms Bengali English (BE) and Kannada English (KE) will be used as a point of reference. There is a possibility that some of the aspects of KE may be applicable to other L1 speakers from the state of Karnataka where Kannada is spoken and, similarly, some of the features described in BE may also be found for speakers of English with other L1s backgrounds living in West Bengal and Bangladesh.\(^4\) Furthermore, the study does not make the assumption that all speakers of IndE whose L1 is Kannada or Bengali will demonstrate those features described in the current investigation. Nonetheless, similar to previous research on segmental features, separation of IndE into sub-varieties (BE and KE) is a step towards a better understanding of variation in the IndE intonational system until further research is undertaken involving more L1 groups at different proficiency levels and more speakers from other regions and socioeconomic backgrounds.

The first aim of the study is to describe the features shared by both groups which could potentially be common for IndE speakers across a number of backgrounds (due to language contact, external factors, use of English in India, and so on), as well as some universal processes involved in the development of L2 prosody. The second aim is to present the features characteristic for each group. In this case, the differences could be attributed but not limited to the speakers’ L1s or regional affiliations. The third aim of this study is to outline similarities and differences with other well-described English varieties, such as Australian, British or American English and other languages, in order to add to the typology of English intonation and cross-linguistic or cross-dialectal use of intonation.

The study adopts an experimental design where the subjects are audio-recorded while performing a number of production tasks. The phenomena under investigation are prosodic structure, prominence and accentuation, phrasing, tonal composition, and the use of pitch range. These phenomena are analysed within the Autosegmental-Metrical (AM) framework (Pierrehumbert, 1980; Ladd, 2008), described in detail in §2.1 of the following chapter. This approach will allow for a separate investigation of phonological and phonetic features and will help determine the degree and the source of variation. The proposed descriptions of KE and BE can then be readily compared to other varieties of English as well as other languages.

\(^4\) A description of Bengali (Bangla) dialects will be presented in §2.3 of the following chapter.
1.5 Outline of the study

The structure of the study is as follows. The literature review in Chapter 2 provides a brief overview of the AM framework and the fundamentals of English intonation, previous research on L2 prosody, IndE prosodic features, and the intonational features of the L1s in this study and several other Indian languages to show similarities within and differences from the intonational phonologies of well-established Englishes.

Chapter 3 outlines the methodology used in this research. It describes the experimental design that is employed in Chapters 4, 5, 6 and 7. Chapter 4 is the first experimental chapter, and addresses the intonational realisation of information structure, including accentuation, prosodic marking of focus, and phonetic cues to accentual and focal prominence. Chapter 5 investigates the tonal alignment and scaling characteristics of rising accents. In this chapter, the phonetic realisation of the rising gesture on prenuclear and nuclear accented words is examined in order to posit pitch accent category or categories.

Chapter 6 examines the levels of prosodic constituency in IndE based on the amount of boundary related lengthening. Chapter 7 presents the description of Bengali English and Kannada English intonation. It draws on the findings of the previous experimental chapters and includes a set of additional analyses. The chapter outlines the prosodic structure, intonational contours and nuclear tunes in various types of sentences, as well as the tone inventory for BE and KE. Chapter 8 presents a summary of the findings addressing the research questions presented at the end of the literature review. It also highlights implications of the present study and aspects for future research. References and appendices follow the concluding chapter.
Chapter 2: Literature review

2.1 Intonational phonology within the Autosegmental-Metrical framework

Taking into account that the AM framework has had a number of detailed accounts in the literature (Ladd, 1996, 2008; Beckman & Venditti, 2011), the following section briefly introduces the framework, outlines its main components, and explains its processes or features in relation to the current study. It also presents the fundamentals of the intonational system of English, highlighting differences across competing analyses, and discusses differences and similarities across varieties of English around the world.

2.1.1 Overview of the AM framework and its main components

The AM approach to intonational phonology is one of the frameworks developed for the analysis of intonational systems and prosodic organisation across a variety of languages (Beckman & Venditti, 2011). The term was introduced and coined by Ladd (1996). The foundations of this approach were built on Bruce’s (1977) work on the tonal structure of Swedish, which investigated two lexical pitch accents and their phonetic realisation, and on Pierrehumbert’s (1980) analysis of AmE intonation. Pierrehumbert’s model of English intonation was further revised and modified (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman 1988). It is important to note that there are other approaches to the analysis of English intonation (Gussenhoven, 2004; Grabe, Post, Nolan & Farrar, 2000; Grabe et al., 2005) and these will be discussed in the context of the fundamentals of English intonational system (§2.1.2). In the 1990s, the AM framework was incorporated into a prosodic transcription system, the Tones and Break Indices (ToBI), developed for Mainstream American English (MAE) (Beckman & Ayers-Elam, 1994, 1997; Beckman, Hirschberg & Shattuck-Hufnagel, 2005). Since its development, ToBI has become the tool used for annotating intonation in other varieties of English, such as Glasgow English ToBI (Mayo, Aylett & Ladd, 1997) or AusE ToBI (Fletcher & Harrington, 2001), and for other languages and dialects. This has resulted in the introduction of multiple compatible ToBI models; for example, ToBI for Lebanese Arabic (Chahal,
In the AM framework, the intonation contour is analysed as sequences of underlying low (L) and high (H) turning points on an autosegmental tier, where L and H tones correspond to different tonal events and the pitch between them is generated by interpolation. In many languages, including English, these tonal events are divided into pitch accents, signalling accentual prominence, and edge tones which mark the edges of prosodic boundaries. The tones are distributed across utterances in ways allowed by the prosodic structure of that language, assuming a phonological organisation of intonation. The abstract phonological representation of the underlying tones is mapped onto their phonetic realisation. This theory recognises four major components for describing intonational and prosodic structure: accentual prominence, phrasing, tune-to-text alignment, and pitch range.

Of particular relevance for English varieties, the notion of ‘stress’ is a lexically specified distinction between strong and weak syllables and is a property of the word. Prominence is concerned with the ‘stress pattern’ within an utterance and usually corresponds to several degrees of perceived prominence (Ladd, 2008). Metrically strong (stressed) syllables are associated with pitch accents. This term first introduced by Bolinger (1958) is now commonly used after its re-introduction by Pierrehumbert (1980). Pitch accents are post-lexical pitch movements that mark accentual prominence and, together with edge tones, contribute to the intonational pattern of an utterance. Pitch accents have a starred tone to indicate their association with the stressed syllable, and can be represented by a single tone, high (H*) or low (L*), or a combination of two tones which are represented by bitonal accents, for example L+H* or L*+H. The last accented word in a phrase carries a nuclear accent, which represents a higher level of intonational prominence compared to a prenuclear pitch accent.

In the AM approach, there is a very close relationship between prominence and phrasing. While accentual prominence helps project the information, phrasing is responsible for structuring that information by ‘chunking’ it and dividing it into the relevant prosodic units. Both prominence relations and prosodic structure are realised with the help of such suprasegmental features as pitch, duration, amplitude and the realisation of segments (Jun, 2005b). The dynamic relationship between prominence, phrasing and acoustic parameters is demonstrated in the realisation of focus. This is
known to have a prosodic effect in many languages (Frota, 2002) and is expressed by different prosodic means that may vary in the way the focus domain is marked. Some of these prosodic means include pitch accent type (Beckman & Pierrehumbert, 1986 on Japanese; Frota, 2002 and Gussenhoven, 2004 on focus realisation and its typology; Khan, 2008, 2014 on Bengali), pitch excursion (Cooper, Eady & Mueller, 1985 on AmE; Gussenhoven, 2004; Féry & Kügler, 2008 on German), post-focal pitch compression (Xu & Xu, 2005 on English; Patil et al., 2008 on Hindi), change in phrasing (Jun & Lee, 1998; Jun, 2005a on Korean) and lengthening of the focus domain (Gussenhoven 2004; Cooper et al., 1985 on AmE; Baumann, Becker, Grice & Mücke, 2007 on German).

The prominence relations within a prosodic phrase are bound by the lexical and post-lexical prosodic structure of a language, and languages have different levels of prosodic structure despite the universal function of prosody to chunk information (Ladd, 2008). For example, Grice, Baumann and Benzmußler (2005) posited two levels of phrasing above the word in German – an intermediate and an intonational phrase. Khan (2008) in his analysis of Bangladeshi Bengali suggested an intermediate phrase in addition to the accentual or phonological and intonational phrases posited for other Bengali dialects (Hayes & Lahiri, 1991 for Kolkata Bengali; Selkirk, 2006).

Previous research has shown that prosodic boundaries corresponding to types of prosodic phrases have different acoustic properties reflecting the strength of those boundaries. For example, the amount of final lengthening is greater in intonational phrases compared to minor prosodic units (Beckman & Edwards, 1994; Beckman, Edward & Fletcher, 1992; Turk & Shattuck-Hufnagel, 2000; Dimitrova & Turk, 2012; Turk, 2012). Similarly, prosodic structure and the level of phrasing correlate with the degree of initial strengthening and are reflected in the realisation of acoustic and articulatory parameters (Keating, Cho, Fougeron & Chai-Shune, 2003; Keating, 2006; Byrd, Krivokapić & Sungbok, 2006). The position of the tonal target in the prosodic structure also affects the scaling of the tones (Liberman & Pierrehumbert, 1984; Truckenrodt, 2002; Ladd, 1986, 2008). A more detailed discussion on these topics will be included in the first section of Chapter 4, examining accentual and focal prominence and the phonetic cues to prominence.

The relationship between text and tune is another fundamental concept in the AM framework for a number of reasons. First, it is crucial in determining the
phonological organisation of the tune. Second, it reveals that the relationship between
the phonological association and phonetic alignment of the phonological categories is
more complex than was assumed in the earlier autosegmental literature (Arvaniti,
2012). In recent years, attention has been drawn to the tonal alignment of pitch
accents and their internal structure across languages. The studies have shown that a
tone or tones in pitch accents may align in different ways with the segmental material.
This can take place across and within languages. For example, both tones in L*+H
accents align outside the accented syllable in Greek (Arvaniti & Mennen, 1998;
Arvaniti & Baltazani, 2005), while in English, the L tone of L*+H aligns with the
accented syllable (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988; Arvaniti
& Garding, 2007). The alignment of the L and the H in rising accents differ
between Minnesotan English and Southern Californian English (Arvaniti & Garding, 2007)
and between northern and southern dialects of German (Atterer & Ladd, 2004).
Moreover, studies have shown that tonal alignment is affected by vowel duration
(Ladd, Mennen & Schepman, 2000 on Dutch; Ladd, Schepman, White, Quarmby &
Stackhouse, 2009 on BrE dialects), syllable structure (D’Imperio, Petrone & Nguyen,
2007 on Neapolitan Italian; Prieto & Torreira, 2007 on Spanish), speech rate (Xu,
1998 on Mandarin; Prieto & Torreira, 2007 on Spanish) and many other factors.
Chapter 5 of this dissertation focuses on tonal alignment and will include a full
discussion on the alignment and anchoring of tonal targets relative to the segments.

In addition to the timing of tonal events in relation to segmental landmarks,
pitch is analysed in terms of its ‘vertical’ dimension or pitch range. Pitch range and its
use within the AM framework involve abstract phonological tone levels and phonetic
rules for their realisation at the f0 level. The H or L tones are considered in reference
to the speaker’s use of ‘tonal space’, which are speaker-specific reference points for
high and low f0 values (Ladd, 2008). In any given language, speakers tend to lower
their pitch throughout an utterance (Pierrehumbert & Beckman, 1988). Scholars have
recognised different types of downward trend, including declination and downstep.
Despite their recognition, the treatment of downtrends together with their distinction
from one another and the underlying sources has been a source of ongoing debate
within AM models (Gussenhoven, 2004; Ladd, 2008; Connell, 2011).

Declination is one of the most recognised phenomena and is now considered
to be a universal feature in the phonetic realisation of pitch found in tonal and
intonational languages, at least for declarative intonation (Pike, 1945; Ladd, 1984).
Declination is used to refer to a gradual lowering of the phonologically specified f0 targets in a speaker’s tonal space through the course of an utterance (Connell & Ladd, 1990). It may be subject to suspension in non-statements in some languages (Connell, 2011). Downstep, on the other hand, is the lowering of H tones in a sequence of pitch accents which creates the effect of ‘terracing’, usually triggered by the influence of the L tone on the following H tone. The term was originally introduced in reference to tonal languages spoken in Africa (Winston, 1960), and was later adopted by Pierrehumbert (1980), Liberman and Pierrehumbert (1984) and Pierrehumbert and Beckman (1988) in the description of English intonation. This led to its implementation into the system of describing other languages. While discussing downstep, it is important to mention final lowering, distinguished by Liberman and Pierrehumbert (1984) from downstep. In their study of English intonation, they found that the last pitch accent in a phrase was subject to final lowering. This is a phonetic rule affecting the end of utterances, where the phrase-final pitch accent is scaled lower than the values predicted by the downtrend pattern. Subsequently, a study by Arvaniti (2007) on MAE and BrE confirmed that final lowering is independent of declination and targets the final accent of an utterance at the edge of the major phrase boundary.

The notion of downstep and the ways of characterising it remain an unresolved issue. First, in non-tonal languages it is often difficult to distinguish declination from downstep (Ladd, 2008). Second, there are different views on the nature of downstep, such as whether or not it can be explained in terms of its phonological organisation (Snider, 1999; Ladd, 1996, 2008) or its phonetic realisation (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988). Ladd (1996), for instance, argues that register differences in downstep can be phonological. Ladd’s proposal is based on the notion of relative strength in metrical phonology. Third, the latter approach assumes paralinguistic or gradient treatment of f0 modifications, raising an ongoing issue of distinguishing linguistic from paralinguistic features in intonation analysis.

Despite the controversies about the linguistic nature of tone scaling and the processes involved (see Ladd, 1996, 2008; Gussenhoven, 2004; Beckman & Venditti, 2011; Dilley, 2010), the AM approach to pitch range has made it possible to move beyond some of the previously irresolvable theoretical questions and to answer empirical questions about the scaling of tonal targets (Ladd, 2008). In the present study, pitch range and the use of tonal space will be examined in relation to two L1
groups in order to investigate the use of pitch in narrow focus marking, global $f_0$ trends in questions and statements, and tonal categories and their phonetic realisation.

2.1.2 English intonation within the AM framework

As indicated earlier, since Pierrehumbert’s (1980) analysis of AmE intonation, revisions to the model have been made (Beckman & Pierrehumbert, 1986) and the ToBI annotational system for MAE has been developed (Beckman & Ayers-Elam, 1994). For the purposes of the present study, this section presents the fundamentals of English intonation and describes some of the key elements in ToBI analysis, mostly focusing on tonal inventory and prosodic constituents (Beckman & Ayers-Elam, 1994, 1997; Beckman et al., 2005). There are other approaches to the analysis of intonation in English within the AM, based on Southern British English (SBE) (Gussenhoven, 1984, 2004; Grabe et al. 2000; Grabe et al., 2005), in addition to the widely used approaches to intonation analysis referred to as the British School (Halliday 1967, 1970; O’Connor & Arnold, 1961). A modified Pierrehumbert and Beckman (P&B) model was chosen due to its resemblance to the ToBI developed for AusE by Fletcher and Harrington (2001) and its relevance in the analysis of the corpus in the study.\(^5\) This section also addresses the main differences between competing analyses as well as a number of controversial issues in this field.

The P&B model includes the following pitch accents: $H^*$, $L^*$, $L+H^*$ and $L^*+H$, $!H^*$, $L+!H^*$, $L^*+!H$, and $H+!H^*$. The pitch accents where a high tone is marked with “!” correspond to the downstepped phonetic realisations of the three main accents, $H^*$, $L+H^*$ and $L^*+H$. Following the process of downstep (described in the previous section), downstepped accents are lowered relative to the preceding high tone. They never occur on the first accented word of a phrase in English, unless it is an $H+H!^*$ accent. Downstep, originally thought to have been triggered by a bitonal accent, was subsequently reviewed (Pierrehumbert, 2000; Ladd, 2008) and is now also recognised after a simple high accent, $H^*$. The falling accents in the original Pierrehumbert (1980) analysis were excluded after revision of the downstep process. That is, $H^*+L$ was subsumed by $H^*$, and an $H+L^*$ accent that lowers down from the preceding high pitch was subsequently substituted by $H+!H^*$.

\(^5\) The data was collected in Australia, and the speakers are all first generation migrants residing in the State of Victoria, Australia.
It is important to mention that English employs both prosodic and phonetic means of marking narrow focus structure, where greater prominence is placed on one word in a sentence as opposed to broad focus structure where several words can be prominent. The post-focal (and at times pre-focal) material undergoes the process of deaccenting, where only the focused word is assigned a pitch accent, and all the words following it do not bear pitch accents (Beckman & Pierrehumbert, 1988; Gussenhoven, 2004; Ladd, 1996, 2008). Focus in English leads to differences in accentuation patterns but does not introduce differences in phrasing, as is the case in such languages as Korean (Jun, 2005a) or Japanese (Pierrehumbert & Beckman, 1988). In addition, the focused syllables have higher pitch excursion and greater duration and intensity compared to accented syllables in non-focal contexts (Cooper et al., 1985; Eady & Cooper, 1986; Breen, Fedorenko, Wagner & Gibson, 2010).

The edges of prosodic boundaries are marked with the help of high (H) and low (L) tones. Two levels of prosodic constituents above the word have been posited in the P&B model. These are an intermediate phrase (ip) and an intonational phrase (IP). The right edge of an ip is marked with a phrase accent of H-, L- or a downstepped high !H-. The edges of an IP, a major prosodic constituent, are marked with a combination of an edge tone in configuration with a boundary tone, as either H% or L%. The use of tones corresponding to the relevant levels of prosodic constituency and pitch movement is shown below. In this example, a short sentence forms one ip within an IP. The pitch contour at the right edge is falling and is represented by a configuration of two low tones, L-L%.

\[
\text{[[Mary is studying ]} \text{ip}] \text{IP} \\
\text{H*} \quad \text{!H*} \quad \text{L-L%}
\]

It is important to note that a high phrase accent triggers the process of upstep, whereby the following boundary tone scales higher. In the configuration H-H%, which corresponds to a rising pitch at the edge of the phrase, the H% boundary tone is higher than the H- phrase accent. Similarly, a high phrase accent raises a low boundary tone in the H-L% configuration, thus bringing the low tone (L%) to the level of the phrase accent. This corresponds to a level or a plateau-like pitch movement and has been used in the analyses of AmE varieties and AusE (Fletcher & Harrington, 2001; Fletcher, Stirling, Mushin & Wales, 2002; Fletcher, Grabe &
Warren, 2005), but not in Glasgow English (Mayo, 1996; Mayo et al., 1997) or other languages, for example German (Grice et al., 2005) and Greek (Arvaniti & Baltazani, 2005).

The definition of a phrase accent and its association with the elements of the prosodic structure remains an unresolved issue in intonational research (Ladd, 1983, 2008; Gussenhoven, 2004; Grice, Arvaniti & Ladd, 2000). The P&B model posited that a phrase accent is a tonal event between a pitch accent and a boundary tone, and that the phrase accent is seen as the edge tone for a level of prosodic structure smaller than the intonation phrase. Based on their work on Japanese tonal structure, Pierrehumbert and Beckman also raised the possibility of a secondary association for phrase accents, and suggested that the phrase accent in English could be linked not only to the right edge of the intermediate phrase but also to the right edge of the nuclear word. This type of double association is a way of accounting for the fact that in English the phrase accent may correspond to a stretch of pitch rather than to a single turning point. Grice et al. (2000) considered a number of European languages to address this issue and investigated the possibility of secondary association. They showed that in the absence of a post-nuclear stressed syllable available as an anchor, a phrase accent occurred close to the phrase edge. The researchers left the question of the underlying source of the phrase accent and its relationship with the levels of prosodic structure for future research.

The difficulty in determining the nature of the phrase accent could in part be due to the competing analyses of the prosodic structure in English and a lack of consensus on the levels of prosodic constituents. The intermediate phrase, integral in the P&B model and supported by other analyses (Ladd, 1996, 2008; Mayo et al., 1997; Fletcher et al., 2002; Dainora, 2002; Arvaniti & Garding, 2007) is not considered in the work of Gussenhoven (1984, 2004), nor of Grabe and her colleagues (Grabe, 1998; Grabe et al., 2000; Grabe, Post & Nolan, 2001; Grabe et al., 2005). Gussenhoven’s (1984, 2004) analysis of Southern British English (SBE) proposes two levels of prosodic constituents above the word: a phonological phrase and an intonational phrase. The phonological phrase has no tones at its edges. Therefore, the nuclear pitch accent is followed by a single tone marking the edge of an IP. Grabe (1998) excludes an intermediate phrase level and applies a modified version of Gussenhoven’s analysis. An intonation phrase boundary can be associated with a high tone (H%), a low tone (L%) or no pitch movement at a boundary (0%).
Another debatable subject in the intonation of English is the phonological contrast between the two pitch accents H* and L+H*. These pitch accents are often difficult to differentiate in spontaneous speech, especially when a speaker uses narrow pitch or when the accents are located in utterance-initial positions. In Ladd’s (1983) view, L+H* is not a distinct accent but a version of an H* accent that adds more emphasis. Further experimental work on English by Ladd and his colleagues (Ladd & Schepman, 2003; Dilley, Ladd & Schepman, 2005) proposed that H* and L+H* are not two distinct phonological categories but a gradient variation. By contrast, Dainora (2002) found that in her corpus, the accents annotated as H* and L+H* were in different distributions. She therefore concluded that they were likely to be distinct phonological categories in the intonation of English. Arvaniti and Garding (2007), who investigated high and rising accents in AmE, found that a three-way distinction between H*, L*+H and L+H* included in the ToBI conventions for MAE is valid, but may not be applicable across all dialects. This may offer a partial explanation for the fact that H* and L+H* are often confused in ToBI annotations. The findings of Arvaniti and Garding (2007) emphasise the importance of research into cross-dialectal variation in English intonation and is the subject of the following section.

2.1.3 Variation across English language varieties

Studies looking at the differences and similarities across English language varieties constitute a growing body of research. There have been descriptions of English prosodic and intonational features in varieties of BrE (Grabe et al., 2000; Grabe, 2004; Grabe et al., 2005; Ladd & Schepman, 2003; Ladd et al., 2009), AmE (Diano, 2002; Hedberg & Sosa, 2002; Arvaniti & Garding, 2007), AusE (Fletcher & Harrington 2001, Fletcher et al., 2002), New Zealand English (NZE) (Warren & Britain, 2000), Scottish English (Mayo, 1996; Mayo et al., 1997), and regional varieties of Irish English (Kalaldeh, Dorn & Ní Chasaide, 2009; O’Reilly, Dorn & Ní Chasaide, 2010), as well comparisons across main varieties (Fletcher et al., 2005).

These studies examined several prosodic aspects, investigating tunes and possible semantic differences, phonological representation of tunes and tonal inventory, phonetics of tune realisation, and tonal alignment of pitch accents. The work of Grabe and her colleagues (Grabe, 2002; Grabe, 2004; Grabe et al., 2005) under the Intonational Variation in English (IViE) project includes an examination of
the phonological and phonetic features in the intonation of nine dialects spoken on the British Isles. Grabe (2002) found phonological differences in tune composition for several English varieties. Declaratives in Leeds, Bradford and Dublin English, for example, were produced with a fall (H* L%) but in Belfast English were mostly characterised by a rise (L* H%). Polar questions were produced with a rise in Belfast English (L* H% and L*H H%), compared with falling and rising nuclear contours observed in Leeds English (L*H % and H* L%). Grabe et al.’s (2005) study demonstrated broad similarities as well as some cross-dialectal differences. The data showed that in Cambridge or London dialects, the distinction between utterance types involves different accent patterns, while in other dialects (Leeds and Belfast) there were no localised differences in pitch contour.

Fletcher et al. (2005) compared the high rising tune in varieties of BrE, AusE and NZE and showed that in Glasgow English and Belfast English the rises form a ‘standard’ declarative tune. They also found that the phonetic realisation of the rising tunes and rises in general were not identical in AusE and NZE. Grabe et al. (2000) compared the phonetics and phonology of two pitch accents in phrase-final positions across Leeds English, Newcastle English and Belfast English. The results showed a potential difference in the realisation of a particular ‘intonological’ specification in different varieties of British English. For example, Cambridge English and Newcastle English compress (steep fall in the f0) rising and falling accents, but in Leeds English these accents are truncated (shallow fall in the f0 as a result of incomplete production of the contour). Moreover, the authors proposed that in some varieties, truncation or compression apply to all pitch accents available, but in others truncation can be tied to one pitch accent and compression to another.

Experimental research on the inventory and distribution of pitch accents in BrE has shown the preference for a rising pitch on accented words in Glasgow English and some varieties of Irish English as well as a smaller inventory of pitch accents. Mayo et al. (1997) proposed three main accents for Glasgow English: H*, L* and L*H, where the L*H accent indicates its association with a rising tune, thus eliminating the contrast between the bitonal accents (L+H* and L*+H). This echoes the findings of Arvaniti and Garding (2007) discussed in the previous section whereby not all varieties of English include three types of pitch accents with a high tone (H*, L+H* and L*+H). O’Reilly et al. (2010) reported that in Donegal English, spoken in the north of Ireland, the prevalent pitch accent type in both nuclear and
prenuclear positions was L*H, with only some occurrences of high (H*) or falling (H*L) accents. Belfast English is also characterised by a rise on accented words (Grabe et al., 2005).

Investigation of the alignment patterns of pitch accents across varieties has shown that the same phonological categories can have differences in their phonetic realisation. Ladd et al. (2009) found that both nuclear and prenuclear peaks were aligned later in Scottish Standard English (SSE) than in RP. Similarly, speakers of AmE from southern California produced high tonal targets with later alignment (Arvaniti & Garding, 2007). In Irish English (Kalaldeh et al., 2009), Dublin and Drogheda English appear to have fixed peak alignment in both nuclear and prenuclear conditions, while Donegal English shows a drift in peak alignment in both prosodic conditions.

Overall, previous research on well-described varieties of English demonstrates that similar cross-linguistic variation in intonation differences may be observed on several levels and may have different underlying sources. According to Ladd’s typology (2008), these differences can be: a) semantic, arising from differences in the meaning or use of the same tune (such as declarative rises in AusE and Glasgow English); b) systemic, related to the differences in phonological categories, as, for example, the pitch accent inventory in Glasgow English versus AmE varieties; c) realisational, corresponding to the differences in the phonetic realisation of the same tune (truncation versus compression in varieties of BrE); or d) phonotactic, comprising differences in tune-text association (alignment of pitch accents).

Unlike the longstanding varieties, an emerging group of new Englishes (NE) has only recently become the subject of experimental research. A great deal of attention has, to date, been given to the rhythmic properties of English and research has shown that such varieties as Singapore English (Low, Grabe & Nolan, 2000; Detering, 2001), East African English (Schmied, 2004 for English spoken in Kenya, Uganda and Tanzania), Nigerian English (NigE) (Udofot, 2003; Gut, 2005), Hong Kong English (Setter, 2006) or Black South African English (BlSAfE) (Coetzee & Wissing, 2007) do not necessarily employ, or employ to a lesser degree, the same canonical principles of English vowel reduction in unstressed syllables. They also exhibit differences in the durational properties of stressed versus unstressed syllables compared to BrE, possibly due to the interference from L1 prosodic phonology.

Limited accounts of the intonational systems and prosodic structures of new
varieties of English are available. Among the features found in a number of NE varieties are their accentuation patterns and use of accentual prominence, as well as their striking difference from well-established varieties. NigE, for instance, exhibits a lack of deaccenting of given information and accent placement on many words in an utterance, including function words (Udofot, 2003; Gut, 2004, 2005). There is a higher frequency of simple pitch movements on accented words and an interaction between accent and tone for speakers of English whose L1 is a tonal language. A preference for high pitch on stressed syllables has been reported for NigE (Jowitt, 2000; Gut, 2005), where function words are often produced with a low or mid pitch and content words are mostly produced with a high pitch. The results of a perception study by Gussenhoven and Udofot (2010) on NigE confirmed the discrimination of low and high pitch depending on the grammatical function of words, perhaps as result of L1 influence. The authors raised the issue of accenting and deaccenting in this variety.

Another perception study (Swertz & Zerbian, 2010) investigating focus marking and prosodic boundaries in BISAfE reported that the degree of L1 influence was related to the proficiency level of the participants whose L1 was Zulu. Moreover, the greatest difference was observed in the focus marking, reflecting systemic differences between English and Zulu in structuring information. Zulu does not use prosodic means for focus marking. When looking at the prosodic marking of focus in modified noun phrases across varieties of South African English, comparing white English speakers of General South African English with the speakers of the acrolect, postacrolect and upper mesolect of BISAfE, Zerbian (2013) found significant differences in the use of intensity and f0 across the four varieties. The mesolect and acrolect did not manipulate both parameters on the basis of focus, and showed the recognisable influence of Bantu languages. Vowel lengthening on focused nouns was observed only in the acrolectal variety of BISAfE. In addition, the difference in acoustic cues was found even in the newly emerged postacrolectal variety, where speakers used f0 but not intensity, unlike the speakers of GenSAfE who manipulated all three parameters under focus structure.

In summary, the results of the studies on several prosodic and intonational features have shown that NE often exhibit the features of L1s as a result of prosodic transfer. The degree of L1 influence, however, is not always easy to account for due to greater differences within varieties based upon language development and language
contact. Research on NE also echoes the findings of studies on L2 intonation, where proficiency plays an important role in the presence or absence of the linguistic features of L1. The lower the proficiency is the greater the amount of L1 influence. As a result, such features as the use of prominence, its acoustic cues, and accentuation patterns, taken for granted in longstanding varieties of English and other West-Germanic languages, are not necessarily present or may be only partially included in the intonation of other WE varieties. There is also a large degree of inter-speaker and inter-group variation, reflective of the typological differences in intonational systems across languages. More fine-grained acoustic analyses are needed to better understand the processes involved in the formation of the intonational systems of NE and to describe variation within these systems. Previous research has reported systemic and realisational differences. It is important to look at other categories proposed in Ladd’s typology, such as semantic or phonotactic differences, in order to extend the field of research and increase our understanding of typological differences across new and longstanding varieties.

This study is a step in this direction, and will draw on comparisons of the relevant categories and features with other varieties of English. With ongoing developments in ToBI models for languages around the world, the choice of using the AM theory for describing the sociophonetic variation in IndE intonation appears to be the most appropriate and justifiable.

2.2 L2 prosody and intonation

As argued in §1.3.2, L1 influence, together with a number of other factors, could account for linguistic variation within IndE. In linguistics, the role of L1 and its influence on L2 has been one of the core issues of experimental and theoretical research in Second Language Acquisition (SLA) (Eckman, 2004). The following section gives an outline of the SLA framework⁶ and highlights the differences in theory and application between SLA and WE paradigms. It also discusses the findings of previous studies on the prosodic and intonational features of L2, explains the main

---

⁶ A monolingual approach with an L1/L2 distinction has been criticised in the recent literature (see Pennycook, 2012; Canagarajah, 2013). The scholars argue that many speakers around the world experience language development and practice as a multilingual phenomenon. Canagarajah (2013) has highlighted that in the age of mobility and globalisation there is a need to consider language use from the perspective of ‘translingual’ practices, and has illustrated this approach in the context of linguistically diverse South Asia.
terms, and demonstrates how recent developments in the research on L2 intonation may be relevant when examining intonational features of IndE.

2.2.1 World Englishes and Second Language Acquisition

SLA was established as a field of research at the end of the 1960s and is built on a number of theories. This framework is concerned with the description of L2 language and its development, focusing on the acquisition of various aspects of language use and structure, and is therefore not limited to phonology. The term ‘second’ is applied to any language other than the first language, and can be a third or, at times, a fourth or further language. The aim of the research in this field is to describe L2 patterns with the help of empirical research in order to develop models or theories that explain these patterns in a systematic way. There are a number of main questions within SLA, such as what the speakers of an L2 acquire, how they acquire certain features of the L2, what kind of differences can be observed across individual speakers and the effect of instruction on the acquisition process (Ellis, 1994). Previous experimental research has a) investigated the characteristics of L2 speech at a certain point in time, looking at speakers with various proficiency levels, and b) examined changes in L2 phonology over a period of time. The factors considered in L2 acquisition have been categorised as internal (L1 transfer and linguistic universals) or external (input and interaction, social), as well as learner specific, which includes age, attitude and motivation, cognition, language aptitude, length of residency, and so on (Ellis, 1994; Flege, 1995).

The WE framework, similar to SLA, is concerned with the issues of language learning and acquisition but in completely different contexts (Kachru & Nelson, 2006). First, the concepts of ‘target’ and ‘native-speaker’-like proficiency are no longer appropriate. Speakers of NE varieties – often referred to as ‘indigenised’ varieties of English (Sridhar, 1986; Sridhar & Sridhar, 1992) as in Singaporean, Indian or Sri Lankan English – will use the same NE variety to communicate with other speakers of English as an L2. Second, English may function as a complementary language with the mother tongue used at home. Consequently, there is a high degree of code switching, code mixing and diglossia which are integral parts of WE in general (Kachru, 1983). Third, the SLA paradigm assumes extensive and sufficient input to learn a full range of features, while in WE this input is often more restricted and is linked to the way English is taught (including by L2 speakers) or to be used.
Motivation, which is described as integrative (Gardner, 1985) in SLA, is most likely instrumental in WE (Shaw, 1984), where speakers from one community need to know English in order to communicate with the speakers of another community.

There are a number of other differences that arise in the comparison between the SLA and the WE frameworks. According to Sridhar (1986, p. 11), “it is unrealistic to evaluate the success of second-language learning with reference to transfer-free norms in such contexts”. Similar to other indigenised varieties of English, IndE can be analysed neither as an L2 variety by simply applying SLA models nor under models of native variations (Sharma, 2005; Kachru & Nelson, 2006). A more detailed account on this issue is presented in Sridhar (1986), Sridhar and Sridhar (1992) and Kachru and Nelson (2006).

Such terms as ‘target’ or ‘native speaker’ could be inappropriate when describing IndE and other SAEs. Due to several assumptions, the SLA approach is difficult to apply to new and emerging varieties of WE. Nevertheless, taking into account the dearth of studies on IndE intonation, it is crucial to look at the possible influence of L1, examine its degree, and determine whether these differences can be explained in terms of phonological or phonetic features.

The role of L1 in terms of interfering with or facilitating L2 acquisition remains relevant and multifaceted in current linguistic research (Gut, 2009). Moreover, this is an ongoing field of enquiry with a number of questions that have not been fully explained. The latest developments in the research on L2 prosodic features have confirmed the effect of L1 on L2 intonation, found a number of variables contributing to its development, and shown the influence of L2 on L1 and vice versa.

2.2.2 Phonological and phonetic influences on L2 prosody

A large number of studies on L2 phonology have shown that “native speakers perceive and produce words and utterances of L2 through a phonetic or phonological ‘filter’ of their native language (L1)” (Nguyễn, Ingram & Pensalfini, 2008, p. 158). It is important to note that for decades, most attention has been drawn to the segmental features of L2 speech. Despite substantial interest in prosody in the last two decades, the acquisition of a number of intonational features, such as the use of pitch and temporal structure important for constructing meaning, have been neglected (Trouvain & Gut, 2007).
The term ‘non-native prosody’ or ‘L2 prosody’ has been introduced recently to refer to the empirical research that investigates and describes the prosodic features of L2 speakers in order to develop possible descriptions. This is a relatively new but growing field where, to date, a limited number of studies have closely examined the properties of L2 prosody production by speakers from different language backgrounds. The results are compelling and provide greater insight into the development of L2 prosodic features. For example, Ueyama (2000) studied the prosodic transfer of L2 English speakers of Japanese and L2 Japanese speakers of English, focusing on three prosodic phenomena at the level of word prosody. These phenomena were the contrast between lexically accented and unaccented vowels, the contrasts between English tense and lax vowels and between Japanese short and long vowels, and temporal organisation across syllables. The results showed that the transfer patterns of L1 prosodic features into L2 prosody exhibit a great deal of variation, and different transfer patterns in the learner’s production can be explained by a difference between L1 and L2 in terms of the phonological status of a relevant prosodic feature. In addition, Ueyama found that interaction between the prosodic and the segmental levels in L1 influence L2 speech development. Most importantly, the results of the research indicate that the prosodic system of an L2 speaker does not necessarily develop in parallel with the L1 for different dimensions of prosody.

Intonational patterns have received the least amount of interest in the recent research, with only a few production and perception studies investigating intonation and tonal events (Gut, 2009). Several earlier studies concerned with post-lexical level prosody were based on the acquisition of English as an L2. They reported a number of similarities7 across speakers from various L1 backgrounds, such as a narrower pitch range (Willems, 1982; Jenner, 1976) or differences in pitch shape on accented syllables (Willems, 1982; McGory, 1997). The development of the AM model and the description of a greater number of languages and language varieties using the same framework have provided a firmer ground for investigating cross-linguistic intonational differences and documenting these differences. What follows is a review of the relevant experimental studies conducted within the AM framework which highlight the complex role and interplay of L1 features in L2 intonation.

McGory (1997) explored the acquisition of intonational prominence in English

---

7 For a comprehensive summary of similarities in L2 English intonational features reported in earlier studies and/or based on other approaches to intonation analysis, see Mennen (2007).
by Seoul Korean and Mandarin Chinese speakers. She examined the phonetic correlates of prominence (f0, intensity and duration), global f0, and f0 movements on accented words. She found that both groups of L2 speakers appeared to have difficulties producing native English prominence relations. Overall, the patterns of the L1 intonational system were evident in the production of L2 English intonation, indicating the strong influence of L1. The degree of influence differed, however, depending on the proficiency level. In terms of the phonological aspects, unlike speakers of well-established varieties of English where pitch accents are placed on prominent target words only, the speakers in the study placed accents on more words. Significant differences were observed based on different L1 backgrounds. L1 Mandarin speakers associated high pitch not only with accented syllables but also with stressed syllables (not bearing pitch accents), possibly transferring Mandarin lexical tones into English intonation. L1 Korean speakers used the same acoustic cue to accentual prominence as L1 speakers, increasing duration in stressed syllables but failing to reduce the unstressed syllables. A possible explanation could be that Korean does not have the same prosodic characteristics as English, i.e. it is not a stress accent language.

Ueyama and Jun (1998) investigated the realisation of focus in Japanese English (Tokyo Japanese as L1) and Korean English (Seoul Korean as L1) intonation at different proficiency levels. They looked closely at phonological features, namely accentuation and the underlying tonal sequence in narrow focus, along with phonetic features in the actual realisation of tonal sequences. The expectations were that the speakers would use a strategy for marking focus similar to their L1. In Korean, focus is characterised by dephrasing, where a high phrase-initial tone is followed by a low tone, together with an absence of phrase tone after the word in focus until the right edge of the boundary. In Japanese, the process is quite similar but dephrasing is not always complete. Moreover the f0 on the accented word can be realised with lower f0 relative to the previous accent. Although in both languages focus is linked to a high tone while the post-focal material is associated with a low f0, the duration of the sustained high pitch differs. This suggests potential phonetic differences between the speakers of Japanese English and Korean English. Post-focal dephrasing was easier for Japanese than for Korean speakers at the same level of proficiency, and was attributed to the positive influence of L1. The results also indicated that the degree of dephrasing correlated with proficiency levels, wherein the number of post-focal pitch
accents decreased with the increase in proficiency level. The realisation of high tones and the duration of high plateaux also varied depending on L2 proficiency. The findings of the study demonstrate that the effect of L1 on L2 intonation is complex. Not all features directly contribute to the organisation of L2 intonation, and L1 features seem to interact with universal constraints on speech production, such as forming shorter phrases at the beginner level or a lack of fluctuation in the f0 height.

Nguyễn, Ingram and Pensalfini (2008) considered the nature of L1 prosodic transfer in Vietnamese acquisition of English contrastive stress patterns and compared L1 Vietnamese speakers with speakers of AusE. The experiment examined acoustic parameters (f0, intensity and duration) and perceptual strategies to differentiate the three stress patterns in compound, broad focus and narrow focus noun phrases. On the basis of the description of Vietnamese and English intonation, the authors hypothesised that Vietnamese speakers of English would be able to manipulate f0 and intensity but would have difficulty producing durational contrasts for marking prominence due to the limited use of this parameter in Vietnamese. Similar to Ueyama and Jun (1998), this study showed the importance of experience on the acquisition of prosodic features. Advanced speakers showed a greater ability to deaccent words in narrow focus and compound patterns and, to a certain degree, to compress compound words. The results also confirmed the effect of L1 on acoustic prosodic cues (Ueyama, 2000) and some phonological features. Only advanced Vietnamese L1 speakers of English were able to apply post-focal deaccenting and to discriminate compounds from noun phrases by means of duration. For all L2 speakers, however, there was a lack of syllable reduction and therefore temporal compression in compound words as well as the differences in the f0 shape in the accented words. Unlike AusE speakers but similar to Mandarin speakers of English in McGory (1997), Vietnamese L1 speakers used more rising pitch patterns on accented and stressed syllables, perhaps associating stressed syllables with a high rising tone in Vietnamese, thus demonstrating the influence of functional or phonological prosodic patterns from L1 in L2 intonation.

Mennen’s (2004) study, unlike the majority of previous research on L2 intonation, looked at the intonational features in Greek L1 speakers of Dutch and in Dutch L1 speakers of English. Her investigation concerned the phonetic realisation of the phonologically identical rise, common in non-final and prenuclear positions in both Dutch and Greek. The location of the peak in Dutch rises is earlier compared to
the same rising accent in Greek. As well, peak timing of the rise can be affected by vowel duration in the accented syllable in Dutch but not in Greek. The results showed that most L2 speakers in the study were unable to produce the $f_0$ peak alignment in L2 intonation, suggesting that it is difficult to realise the phonetic aspects of intonation even when phonological categories are relatively similar. Additionally, the findings showed that L2 intonational features had a strong effect on the timing difference between the alignment of the L and the H tones in the speakers’ L1, thus indicating that that L1 and L2 intonation may influence one another, similar to segmental features (Flege & Hillenbrand, 1984; MacKay, Flege, Piske, & Schirru, 2001).

Kim and Lu (2011) investigated tonal alignment of the rising L*+H pitch accent in English produced by L1 speakers of Mandarin. Similarly to Mennen (2004), the researchers found a strong influence of L1 on the phonetic realisation of the tonal category. Despite the production of a rising pitch accent by L1 Mandarin speakers, the alignment of L and H targets was associated with the edges of the target syllable, closely following the pattern found in the Mandarin LH tone sequence. Stella (2012) considered the alignment differences of German rising pitch accents produced by L1 speakers of Lecce Italian, spoken in the south of Italy. His acoustic and articulatory study investigated both acoustic and articulatory patterns of tonal alignment, and included two German L2 speakers, one with a low and one with a high proficiency level. The findings showed that the higher proficiency L2 speaker produced alignment patterns similar to the native German speaker. The alignment of the peak produced by the lower level German L2 speaker showed similarity to its alignment in German but the alignment of the low target was closer to the speaker’s L1, indicating partial L1 transfer on the phonetic alignment of tonal categories.

To summarise, these studies have shown how L1 intonation could place constraints on the production of L2 intonational patterns and that L2 speakers may have difficulties learning the full set of categories. However, the process of acquiring prosodic and intonational features is as complex as the acquisition of the segmental phonology. First, a speaker’s proficiency level plays an important part in the degree of L1 influence. This could potentially be applicable to IndE intonation patterns where variation is defined on the continuum of ‘cline proficiency’ (Kachru, 2005), which ranges from educated at one end to broken English at the other end of the continuum. Second, not all features present in L1 can necessarily contribute to and be manifested in L2 production. Moreover, the research suggests that L1 and L2 intonation can
influence one another. Third, universal constraints of speech production have to be taken into account and may interact with the L1-L2 interplay. Finally, the influence of L1 can be observed at the levels of both the phonological representation - as in the inventory of tones or tunes, and phonetic realisation - as in the temporal difference in tonal alignment of the phonologically identical tonal category.

The next section discusses previous research on IndE prosody and intonation, gives a brief description of the L1s in this study including prosodic and intonational features, and highlights several aspects found to be common across a number of indigenous languages spoken in the subcontinent.

2.3 Previous research on IndE prosodic features

Since the introduction of the term and the acceptance of IndE as a variety, its phonology has been the subject of ongoing research. This has generated a number of models and descriptions (Bansal, 1969; CIEFL, 1972; Wells, 1982; Bansal, 1990; Nihalani et al., 2004; Gargesh, 2004; Trudgill & Hannah, 2008), mostly focusing on segmental aspects. A small number of studies examined the prosodic features of this variety (Pandey, 1980; Vijayakrishnan, 1978; Das, 2001; Pickering & Wiltshire, 2000; Moon, 2002; Wiltshire & Moon, 2003; Wiltshire & Harnsberger, 2006; Fuchs, 2013).

Gargesh (2004), in his description of IndE, placed previous studies on IndE phonology into five main categories. These were: a) studies describing the phonetic aspects; b) studies attempting to describe a variety of IndE (mostly based on L1) comparing its features to the RP system and, in recent years, to AmE; c) studies that describe an Indian language, comparing it with the sounds of RP while attempting to outline the features of the particular IndE variety based on that Indian language; d) studies that describe IndE from a sociolinguistic perspective; and e) studies that investigate the issues of intelligibility and attitudes toward IndE pronunciation. For the purpose of this dissertation, I will consider the first three of these categories, beginning with word-level prosody and rhythm, and then proceeding to impressionistic and experimental studies focusing on several prosodic and intonational phenomena.
2.3.1 Word-level prominence and rhythm

Lexical prominence seems to have received the major attention of IndE researchers. Vijayakrishnan (1978), Nair (1996), Pandey (1980) and Das (2001), for example, investigated the placement of lexical stress in order to formulate the rules for stress in IndE and its varieties. Other studies have examined the location of stress and compared it with RP stress patterns (Chaudhary, 1993; Nair, 1996). In addition, several studies on segmental phonology have included information on stress assignment (Sethi, 1980; Bansal, 1969, 1970; CIEFL, 1972; Prabhakar Babu, 1971; Pandey, 1994; Mahanta, 2001; Das, 2001; Wiltshire, 2005).

Previous research has shown that speakers of IndE prefer to place prominence on the basis of segmental and syllabic structure. In other words, the rules of stress are based on quantity and position, where the number and size of the syllables in a word are more important than morphological category or structure. Das (2001), for example, reported that the words produced by Tripura Bangla English (TBE) speakers were produced with stress on heavy syllables; moreover, a final heavy syllable could initiate a foot. Mahanta (2001) suggested that, similar to Assamese, lexical stress in Assamese English (AssamE) is word initial in a sequence of light syllables but moves to heavy syllables in a sequence of light-heavy syllables. In addition, the lack of long vowels in the speakers’ L1 may lead to significant differences in the placement of stress in AssamE. For example, words like taboo or rupee have lexical stress on the first syllable as a result of the unexpected short vowels in the second syllables.

According to Gargesh (2004), word stress in IndE is influenced by filter language or languages and, similar to segmental features, could exhibit potential differences in its assignment and ‘rules’ across speakers from different speech communities. In studies examining Telugu English (Prabhakar Babu, 1971) and Rajasthani English (Dhamija, 1976), the results suggested strong evidence for placing stress on the first syllable, while Punjabi English speakers in Sethi’s study (1980) placed prominence on the second syllable, especially in words consisting of three syllables. An additional feature of lexical prominence in IndE is the unchanged stress position in words with the same spelling but different grammatical function and meaning such as record and re’cord, which are distinct in well-established English varieties. The latter feature could be applicable to IndE in general (CIEFL, 1972; 8

---

8 In this context, a filter language refers to an indigenous language.
Wiltshire, 2005) and is perhaps a reflection of the ‘self-replicating’ nature of English in India (Kachru, 1983).

Little research has been conducted on the acoustic cues to lexical prominence. An earlier investigation conducted by Vijayakrishnan (1978) showed that in Tamil English, prominence is cued by a distinct increase in duration. Two studies (Pickering & Wiltshire, 2000; Wilshire & Moon, 2003) report conflicting results. Pickering and Wiltshire (2000) compared the acoustic correlates of prominence between IndE and AmE. Speakers of English from India participating in the study were of several L1 backgrounds including Bengali, Tamil and Hindi-Urdu. The corpus consisted of video- and audio-recordings of spontaneous speech. In addition to this speech, the speakers of IndE produced a number of sentences in their L1s for further comparison of lexical cues in Indian languages. The analysis was based on accented and unaccented syllables, extracted from two- and three-syllable words, and examined $f_0$ measured at the middle of the accented syllable, and amplitude measured at its peak value in each syllable.

The researchers found that unlike in AmE where lexical prominence is cued by an increase in $f_0$, speakers of IndE had decreased values in stressed syllables. In addition, the speech of the AmE speakers showed a significantly greater, easily perceivable increase in amplitude, while the IndE speakers often did not increase amplitude on accented syllables or had a modest difference compared to unstressed syllables. Interestingly, comparison of amplitude measured in stressed syllables showed significant differences between IndE and AmE speakers and a lack of difference among the IndE speakers, suggesting that amplitude is not a reliable correlate in IndE. Moreover, the use of $f_0$ and amplitude in IndE could be the reflection of Indian languages’ word-prosodic features and could potentially be applicable to IndE in general. On the basis of this finding, the authors posited that similar to Japanese, which uses $f_0$ but not amplitude as a cue to lexical prominence, IndE is a therefore a pitch accent or a non-stress accent language (Beckman, 1986; Beckman & Venditti, 2011) where stressed syllables are characterised by a low pitch.

Wiltshire and Moon (2003) questioned these findings, suggesting that this could have been due to the misinterpretation of the position of lexical prominence in IndE words which was assumed to be the same as in AmE. As discussed earlier, the location of stress in IndE exhibits demonstrably different patterns from BrE or AmE. In their study, Wiltshire and Moon also examined the acoustic correlates of
prominence and compared the acoustic parameters of amplitude, duration and $f_0$ between the speakers of IndE and AmE. In addition, the researchers attempted to address the issue of location of phonological stress. Their recordings were based on a controlled read speech, where 60 target words were produced in a carrier sentence. The IndE speakers were of Telugu, Tamil, Gujarati and Hindi backgrounds, grouped according to their language family, Dravidian or Indo-Aryan.

Wiltshire and Moon (2003) also investigated any variation as a result of L1 differences. In line with previous research, the results showed differences in the location of stress between AmE and IndE. In terms of similarities in phonetic realisation, both groups used the same set of phonetic correlates to cue lexical prominence. The increase in amplitude, duration and $f_0$ on stressed syllables, however, was significantly greater for AmE speakers. It was noted that the degree of increase may have been due to the nature of the task and the recording conditions. There were no differences in any of the acoustic parameters between stressed and unstressed syllables across the IndE speakers.

The most remarkable finding was that the location of stress could have been influenced by the rules of the speakers’ L1s and were reflective of their language families. Telugu and Tamil speakers preferred to place stress on the first syllable, whereas a number of words were assigned stress on second syllable by the speakers of Gujarati and Hindi. The data, however, showed a large degree of inter-speaker variation, suggesting that lexical stress location could not be attributed to L1 influence alone and is therefore a more complex issue. The researchers also noted the difficulty in eliciting consistent judgment on phonological location of lexical stress from IndE speakers, confirming that morphological structure is less important in stress assignment.

Another related contentious issue in IndE is its rhythmic properties. The literature often states that unlike BrE or AmE, English spoken in India tends to be syllable-timed and not stress-timed (see Bansal, 1970; Chaudhary, 1989; Nihalani, Tongue, Hosali & Crowther, 1979; Wells, 1982; Gargesh, 2004; Trudgill & Hannah, 2008), where “syllables are uttered with almost equal prominence” (Gargesh, 2004, p. 1001). The schwa is not realised in weak positions and the vowels retain their quality and, often, duration, thus presenting different timing patterns from BrE or AmE. Kachru and Nelson (2006) noted that this feature could be shared across SAE varieties. The fact that the phonologies of a large number of languages spoken in the
subcontinent, including their rhythmic properties, are under-researched adds to the complexity of investigating rhythm in IndE. Moreover, with a renewed interest in developing metrics to measure linguistic rhythm in order to quantify differences across languages, the dichotomy between stress- and syllable-timed, which was taken for granted for a long time, has been questioned, resulting in various new approaches to investigating rhythm (see Roach 1982; Nespor, Shukla & Mehler, 2011; Fletcher, 2010 for a more detailed account; Fuchs, 2013).

Krivokapić (2013) compared the rhythmic properties of AmE and IndE, and examined rhythmic convergence between the speakers of these varieties. Despite the fact that both IndE and AmE showed patterns of stress- and syllable-timing, her findings indicated a more syllable-timed rhythm for IndE, where increase of foot duration with the number of syllables was larger compared to AmE. Krivokapić also found some evidence of convergence. One IndE speaker, who had lived in the US longer than the other subjects in the study, converged towards AmE in both syllable and foot duration measures. AmE speakers showed a tendency towards converging to IndE. The results did not reach significance but the duration of the foot showed an increase in synchronous speech (interaction with a speaker of the other variety). This suggests the foot squeezed the syllable less, which is a pattern indicative of more syllable-timing rhythm. Another possible interpretation given by the researcher was that the rhythmic properties of AmE may have been affected by the global speech rate in the synchronous condition.

A recent experimental study by Fuchs (2013) also addressed the claim that IndE is syllable-timed. Fuchs investigated rhythm in English spoken by twenty L1 speakers of Bengali, Hindi, Telugu and Malayalam, and compared it with the speakers of BrE from Southern England. In order to test potential differences in rhythm between BrE and educated IndE, the researcher followed a multi-dimensional model of rhythm and applied a range of metrics (e.g. Ramus, Nespor & Mehler, 1999; Low, et al., 2000; White & Mattys, 2007; and others) to investigate variability in loudness, intensity, f0, syllable durations, vocalic intervals and the variability of voiced and sonorant durations. The findings of the production study showed that educated IndE is more syllable-timed than BrE, based on a number of acoustic correlates. However, the rhythmic properties showed differences across metrics and between read and spontaneous speech styles. IndE exhibited less variability in syllable durations. The author suggested that duration may play a less important role in IndE than BrE in
marking prominence. Combined variability in intensity and duration was also found to be smaller in IndE. In IndE spontaneous speech, variability in average intensity increased with the rise in minimum number of intervals, showing that IndE speakers did not vary average intensity in longer utterances to the same extent BrE speakers did. Fuchs suggested that this could potentially be the feature that contributes to the perception of IndE being more monotonous; however, the conclusion was based on single utterances only, thus excluding the patterns of variability in intensity between the utterances.

Sonority measurement-based results (mean sonority and mean change in sonority) also indicated more syllable-timed rhythm for IndE. Variation in the size of $f0$ excursion and the mean of variation in duration and size of $f0$ excursion did not indicate significant differences between the speakers of IndE and BrE. The results on the variability of average and maximum intensity of vocalic intervals showed a lack of support for the idea that variability in intensity is greater for BrE. Despite the fact that unstressed vowels were shorter than stressed vowels in IndE speech, vowel elision in unstressed syllables was less frequent and unstressed vowels were also longer for IndE speakers, leading to the impression of syllable-timing. Vocalic intervals at the edges of intonational phrases were not lengthened to the same degree as those in BrE. In addition, a word-initial glottal stop insertion before vowels was more frequent in IndE.

Fuchs (2013) also reported that educated IndE was spoken more slowly than BrE, contrary to previous claims about IndE. The results of the perception study, which investigated the hierarchy of cues used to discriminate between IndE and BrE accents, showed that segments were the first most important cue, followed by intonation and rhythm.

To conclude this section, the research on lexical prominence has given support to the evidence that IndE speakers have different stress placement ‘rules’ compared to speakers of other varieties, reflecting L1 influence and the use of English in the subcontinent. Moreover, the phonetic parameters that cue lexical prominence in English, such as $f0$, duration and intensity, are manipulated differently in IndE, showing differences when compared with AmE and BrE. In addition, there is a possible variability in the use of these phonetic parameters depending on the speakers’ L1 background. IndE showed the pattern of a more syllable-timed rhythm compared to BrE or AmE. These findings on word-level prosody and rhythm could also be
reflective of potential differences in the use of such parameters as duration, f0 and intensity, on a post-lexical level, contributing to intonational features and prosodic structure.

2.3.2 Intonational phonology

Prior to the development and wide use of the AM framework, a very limited number of studies examined post-lexical prominence and intonation in IndE. The findings of these studies were mostly impressionistic and included somewhat descriptive accounts of the use of rise and fall in different types of utterances, and a different pattern of accentuation and division into intonational groups from well-established English varieties (mostly BrE). Several studies also examined the use of intonation patterns in regional varieties, once again, on a more or less descriptive basis (see Latha, 1978 on Malayalam English).

One of the most noticeable features of IndE is the use of accentual prominence. IndE speakers place an accent on a large number of words in an utterance, highlighting both content and function words including pronouns, prepositions, and so on (Latha, 1978; Bansal, 1969, 1990; Gumperz, 1982; Gargesh, 2004). Gumperz (1982) considered how the difference in phrasal prominence by IndE speakers compared to speakers of AmE or BrE could lead to miscommunication and even change in attitudes. Nihalani et al. (2004, p. 225) noted the placement of nuclear accent on the “wrong syllable”, compared to the phrase-final nuclear accent placement in BrE. Researchers examining varieties of IndE based on L1, however, found that accentuation patterns and nuclear accent placement may be reflective of the speakers’ L1s or regional affiliation. For example, while Rajasthani (Dhamija, 1976) and Punjabi speakers (Sethi, 1980) have a preference for placing greater prominence on the last word in a phrase, Telugu speakers place greater prominence on the first word in a phase (Prabhakar Babu, 1971).

Other noteworthy aspects in relation to prominence and phrasing in IndE are ‘incorrect’ grouping of words (Bansal, 1969; Dhamija, 1976) and inappropriate use of tune referring to pitch movement at the end of utterances (Bansal, 1969). Gargesh (2004) noted falling intonation in statements, and rising intonation in polar questions, some wh-questions and independent clauses. Also, compared to BrE, the narrower pitch range of IndE may contribute to a perception of speakers sounding aggressive
Two experimental studies conducted within the AM framework investigated several phonetic and phonological aspects of intonation in IndE, namely the use of accentual prominence, pitch movement on accented words, and phonetic cues to narrow focus (Wiltshire & Harnsberger, 2006; Moon, 2002). These studies included speakers from different L1s belonging to two major language families (Indo-Aryan and Dravidian) in order to consider L1 influence in an attempt to describe variation and present possible sub-varieties within this complex model.

Wiltshire and Harnsberger (2006) compared the intonational features of Gujarati and Tamil speakers of IndE with the features of AmE speakers. The main focus of the research was on the segmental differences, and it included only some preliminary results on the intonational differences between IndE and AmE and between Tamil English (TE) and Gujarati English (GE). The corpus was based on laboratory speech, and the materials included read sentences, a short read passage and short dialogues. In line with the claims made in previous instrumental or impressionistic studies, the utterances produced by IndE speakers showed many more pitch accents than one would expect in the same utterances produced by AmE speakers. Moreover, speakers of both L1s (Gujarati and Telugu) placed prominence on most or all content words before the right edge of the intonational phrase. This finding confirmed an earlier claim that IndE speakers exhibit multiple uses of pitch accents. However, the nature of this feature remains an issue for further research. Wiltshire and Harnsberger speculated that this important feature of IndE intonational phonology could be the result of influence from Tamil, Gujarati or even Hindi due to language contact. Taking into the account the linguistic complexity of IndE, it could also be linked to the similarity of some prosodic features between Dravidian and Indo-Aryan languages, or even the development of IndE as a variety.

The findings of Wiltshire and Harnsberger (2006), in addition, showed some systematic differences in pitch movements on accented words between the two L1 groups. Despite the use of four types of pitch shape associated with accented syllables (LH, HL, H or L) and very limited used of low pitch, Gujarati L1 speakers mostly relied on rising pitch, which would correspond to either L*+H or L+H* pitch accents in the AM framework, followed by high pitch which was possibly an H* pitch accent. L1 Tamil speakers of English used a falling pitch, corresponding to either an H+L* or
H*+L pitch accent, a high pitch (H*) or a rising pitch accent (L*+H or L+H*). The three types of pitch movement were used in relatively equal proportions.

Similarly, Moon (2002) found that the f0 contours on accented focused words varied between the speakers of English from two L1 backgrounds (Hindi and Telugu). In Hindi English (HE), there was a drop in the f0 at the beginning of the accented syllable followed by a sharp rise with the peak towards the end of that syllable. In order to capture this phonological category within the AM approach, the researcher suggested two possible interpretations of the rising f0, as either H* or L+H*. Moon leaned towards the latter of these because of the rise magnitude, the difference in scaling between the low and high turning points. For Telugu English (TelE) speakers, the f0 contour on the focused words showed no evident drop at the word onset, but had a gentle slope from the preceding word with the peak located in the middle or, on a few occasions, towards the end of the accented vowel, suggesting a H* accent. The author warns, however, that the interpretation of the results was preliminary and required further investigation, as the attempt was to present only a partial phonological analysis of the pitch accent inventory for IndE speakers. Further, the set of materials could have been rather restrictive for these purposes. Despite a range of vowels including monophthongs and diphthongs, all words were monosyllabic. Hence the alignment of the tones, especially the alignment patterns of the H, could not be fully examined. A rise in the f0 could also correspond to an L*+H accent. The location of the L at syllable onset is a reflection of a possible phonetic realisation of this accent. As noted by Ladd (1996), an L*+H accent requires at least two syllables for its realisation.

The main objective of Moon’s study (2002) was to examine the phonetic cues to focal prominence in English spoken by IndE speakers (Hindi and Telugu L1s) with the aims of describing the acoustic correlates of focus in this variety, comparing the two L1 groups, and examining potential differences in the use of phonetic correlates of focus expression between the speakers of IndE and AmE. Several standard phonetic parameters for prominence in English (Fry, 1958; Cooper et al., 1985; Chen et al., 2001) were investigated. These were f0, amplitude and duration. Applying the adjacency analysis, Moon compared focused target words and preceding unfocused words. The findings presented some similarities but also highlighted a number of compelling differences between IndE and AmE, and between HE and TelE (the two sub-varieties of IndE).
First, duration results suggested that unlike AmE speakers, IndE speakers of both L1 groups did not rely on duration as a cue to focus. Second, the $f_0$ results showed differences between AmE and the IndE groups as well as between the Hindi and Telugu L1 groups. While similar values were recorded for maximum $f_0$ across TelE, HE and AmE, there were significant differences in $f_0$ realisation at different temporal intervals of the accented vowel. This finding indicates differences in the location of max $f_0$ across the three groups and in pitch movement associated with the accented syllable. Greater lowering of $f_0$ at the beginning of the accented vowel was reported in HE as opposed to TelE and AmE. For TelE speakers, the $f_0$ at the onset showed a shallower rise compared to AmE and HE, with a peak in the middle of the vowel followed by a level $f_0$ or a continuation rise on the post-accentual material. Finally, the study reported that TelE speakers used much greater overall RMS-amplitude (measured over the entire accented vowel) compared to AmE and HE speakers.

Whilst these two recent studies included only a small number of aspects (distribution of pitch accents in a phrase, preliminary pitch accent inventory, acoustic cues to focal prominence), leaving other features unexplored, the findings reflect some interesting tendencies and help explain some of the features discussed in earlier descriptive work. Thus, differences in the accentuation patterns compared with other well-established Englishes seem to be a salient feature of IndE. Pitch shape differences associated with accented words observed by Dhamija (1976), Sethi (1980) and Prabhakar Babu (1971) in earlier studies have found further support in recent experimental research (Moon, 2002; Wiltshire & Harnsberger, 2006). The nature of this variation is not clear at this stage. It could result from the reflection of L1 varieties or even similarities between certain L1s. The pattern emerging from the studies of Moon (2002) and Wiltshire and Harnsberger (2006) shows the use of a rising pitch applied to accented words by IndE speakers whose L1s belong to the same language family.

A rise in pitch on accented words has been recorded in the intonational phonologies of a number of Indo-Aryan languages (Rajendran & Yegnanarayana, 1996; Patil et al., 2008 on Hindi; Hayes & Lahiri, 1991; Khan, 2008 on Bengali). However, it has also been observed in Tamil, which belongs to a Dravidian language family (Keane, 2007, 2014). Tamil L1 speakers of English in Wiltshire and Harnsberger’s (2006) study used a number of $f_0$ shapes on accented words and were
not limited to the use of a rising $f0$ on the accented syllables. This suggests that the use of rising pitch is more complex and cannot be explained by simply belonging to a language family. Furthermore, it is important to differentiate the pitch accents by their prosodic condition of nuclear or prenuclear. A number of languages employ different phonological categories for nuclear and prenuclear accents (see Arvaniti & Baltazani, 2005 on Greek).

In summary, IndE intonation has been a neglected field of study for a number of decades. The findings of recent experimental research have made significant contributions in addressing a number of features but were based on a small number of L1 speakers and mostly presented preliminary results. Such important features as phrasing, prosodic structure, tune and tone inventory, and pitch range have not been investigated. In addition, as shown by Moon (2002), there could be a number of phonetic differences between speakers of IndE and speakers of AmE or BrE. The acoustic correlates of focus marking used in well-described varieties of English (BrE, AmE, AusE), and long considered to be an integral part of the English intonational system, have been questioned in relation to IndE. The findings described above provide further support for differences across speakers of IndE. They also highlight the need for further systematic research, not only to consider the phonological categories but also to examine potential similarities and differences in the phonetic realisation of these categories in IndE from various L1 backgrounds in order to draw any empirically based conclusions.

2.4 Previous studies on the intonation of relevant Indian languages

The next few sections introduce the L1s of the speakers in this study and present a number of intonational features of Kannada and Bengali prosody, highlighting differences in the analyses of Bengali and its varieties. Section 2.4.3 also includes a summary of the patterns across a number of languages spoken in the subcontinent, in view of potential similarities in the intonational phonologies of the two L1 groups.

2.4.1 Profile of L1s in this study

The languages spoken in the subcontinent belong to four language families and could be divided into two groups: minor languages (Sino-Tibetan and Austro-Asiatic or...
Munda) and major languages (Dravidian and Indo-Aryan). The number of speakers of Indo-Aryan languages constitutes around 78.7% of the population, while the number of speakers of Dravidian languages is around 17.5% (Cardona & Dhanesh, 2003). The literature suggests that long coexistence of these four language groups has led to cross-linguistic influence and the development of features which are shared across South Asia, making it a Linguistic Area (Emenau, 1965; Masica, 1976, 2005). Moreover, there are speculations that Dravidian languages had influenced Indo-Aryan languages even during the development of proto-languages (Steever, 1998), and therefore the phonological system of Dravidian languages has affected the phonology of Indo-Aryan; hence, the significant degree of similarity between these two groups. Due to India being a linguistically defined area, some of the features of English spoken in India could be shared by speakers across the subcontinent, potentially leading to a number of generalisations about IndE phonology.

![A map of the states of India with the corresponding main official language.](http://www.corfizz.com/language.html)

Two languages have been chosen for this study. This choice was made first on the basis of accessibility to the speakers and the number of speakers belonging to one speech community, as Kannada and Bengali speakers constitute a large proportion of the Indian diaspora in Australia. Second, it was important to select L1s of English
speakers from different language groups spoken in areas remote from one another (See Figure 2.1). Kannada is a Dravidian language that, together with Tamil and Malayalam, belongs to the Southern branch of the Dravidian group (Krishnamurti, 2003). It is spoken in the state of Karnataka, in the south of India, and is one of the four main literary languages of the Dravidian family. Four major dialects have been recognised: southern, around the major cities of Bangalore and Mysore; northern, around the city of Dharwar; western, Mangalore district; and north-eastern, located in and around Bijapur. The northern and north-eastern dialects are known to be influenced by Marathi, an Indo-Aryan language. There is also sociolinguistic variation and diglossia (Sridhar, 1990; Steever, 1998). Standard colloquial Kannada as used in the grammars and descriptions of Kannada (Steever, 1998) reflects the speech of educated speakers in and around Bangalore and Mysore, and is the dialect of the speakers in this study.

Bengali or Bangla belongs to the Eastern Indo-Aryan languages, with around 58 million speakers in India and an estimated 107 million speakers in Bangladesh recorded by 1991 census (Dasgupta, 2003). Bengali is spoken in the north-east of India in the states of West Bengal and Assam, as well as the states Bihar, Jharkhand, Mizoram, and Tripura (Khan, 2008). There are a number of Bengali dialects, most commonly divided between eastern and western branches (Dasgupta, 2003; Grierson, 2005). The eastern dialects are spoken in Bangladesh, while western dialects are spoken between West Bengal and Bangladesh and include several sub-groups comprising the central, northern, western, and south-western dialects. Central Bengali is of particular interest to this study and contains the two most widely used forms of the language: the Nadia dialect, which is the basis of Standard Bengali, and the Kolkata dialect, which also has a great influence on the standard (Khan, 2008).

2.4.2 Intonational aspects of Kannada and Bengali

The intonational structure of the two languages relevant to this study has not received equal attention in previous research. Very few experimental or impressionistic accounts are available on Kannada prosody. Research has mostly focused on word-level prominence and speech rhythm. Steever (1998) suggested that lexical stress is non-contrastive and is associated with the first syllable of every word. In a perception investigation by Savithri (1995), word-initial syllables were identified as stressed
more often than other syllables. Savithri (1999) and Savithri, Rohini and Sairam (2005) found duration to be a strong cue for stress perception in Kannada. Sridhar (1990), on the other hand, suggested that stress does not play a role in the language except for emphasis on the phrasal level and that ‘emphatic intonation’ involves extra vowel lengthening. Intonation, focus prosody and the prosodic structure of Kannada are yet to be studied. A tentative account of tunes was given by Sridhar (1990). This found rising intonation is characteristic of polar questions, statements and imperatives are produced with a fall, and incredulity questions similarly have a final falling intonation. A preliminary production study by Mathew and Bhat (2010) generally confirmed these observations. Differences, however, were found in the production of interrogatives, where male speakers used a falling contour and female speakers produced rising contours.

Unlike for Kannada, a substantial amount of work has been carried out on Bengali intonation. This currently includes descriptions based on several dialects (Hayes & Lahiri, 1991; Lahiri & Fitzpatrick-Cole, 1999; Truckenbrodt, 2002; Michaels & Nelson, 2004; Selkirk, 2006; Khan, 2007, 2008, 2014). These analyses disagree on a number of fundamental features; for instance, focus marking, prosodic hierarchy, tonal inventory and the role of the obligatory contour principle (OCP) on tonal sequences. A detailed account of the differences across competing analyses is included in Khan (2008). This section outlines the features relevant to the present study.

Apart from Das (2001) and Shaw (1984), who suggested that lexical stress depends on the syllables’ weight, most of the studies agree that Bengali lexical stress is phonetically weak, completely predictable, and always falls on the first syllable (Kawasaki & Shattuck-Hufnagel, 1988; Hayes & Lahiri, 1991; Michaels & Nelson, 2004; Selkirk, 2006). Unlike in well-established Englishes, the phonetic cues to stress such as f0 and intensity seem to be less reliable in Bengali (Kawasaki & Shattuck-Hufnagel, 1988). Despite its phonetically weak realisation, lexical stress is phonologically significant (Khan, 2008, 2014). Initial stressed syllables are also associated with the post-lexical use of pitch and have pitch accents.

---

9 The authors did not specify what types of questions were selected (polar, wh-questions, and so on.)

10 OCP refers to the obligatory contour principle. In the intonational phonology of Bengali, this principle relates to the deletion of certain tones in order to avoid identical underlying tones in one sequence (Hayes & Lahiri, 1991).
In terms of phrasal prominence, there is a consensus that in non-focal positions, accented words are more often realised with a default low pitch accent (L*) followed by a high boundary tone at the right edge of the prosodic word.\textsuperscript{11} This tonal sequence corresponds to a prosodic constituent smaller than the intonation phrase, and has been given different definitions in the literature as the phonological phrase (Hayes & Lahiri, 1991), the major phrase (Selkirk, 2006) or the intermediate phrase (Michaels & Nelson, 2004). An utterance usually consists of a number of smaller phrases with the same rising pattern, often excluding the final phrase. In the most recent research, Khan (2007, 2008, 2014) proposed the use of the term ‘accentual phrase’ (AP), justifying it by the similarity of this unit in Bengali to the prosodic constituent defined as AP in languages such as Japanese (Beckman & Pierrehumbert, 1986), Korean (Jun, 2005a) or Farsi (Esposito & Barjam, 2007). Here, an AP contains a single pitch accent functioning as a head. In addition to introducing the term AP into the description of the prosodic structure, Khan (2007, 2008) also found evidence for three levels of tonally marked prosodic phrasing and posited an intermediate phrase (ip) to refer to the prosodic constituent between the AP and the intonational phrase (IP), not previously mentioned in the literature.

Accentuation in focal contexts is treated with greater controversy in current research on Bengali. Similar to accented non-focal words, focused words are realised with a rise in pitch; however, there are different approaches to the phonological treatment of this rise and the interpretation of post-focal tonal events or the lack of them. Hayes and Lahiri (1991) and Lahiri and Fitzpatrick-Cole (1999), who examined Kolkata Bengali, treated a low tone as a pitch accent (L*) followed by a high boundary tone (H\textsubscript{p}). The main difference with the non-focal contexts was the absence of additional tones between the phrase in focus and the boundary tone corresponding to IP. Similarly, Selkirk (2006) proposed a floating H tone associated with the right edge of focused elements.

Khan’s (2007, 2008, 2014) description, based on the eastern dialect of Bengali (Standard Bangladeshi Bengali), is in line with the findings of Michaels and Nelson (2004), who also investigated Eastern Bengali. Khan suggested a low rising pitch accent on a focused element (corrective focus and wh-answer focus), labelled as

\textsuperscript{11} Khan (2008) also observed a falling pattern (H* followed by a low phrase accent) in smaller phrases, but noted that it was more restricted in its distribution and could be more common for non-standard dialects.
L*+H. The researcher showed that a high tone associated with focus did not anchor to the right edge of the phrase. Rather, it was consistently realised in post-accented syllables and did not move to the right edge with the increase in the number of syllables post-focus. Moreover, apart from the instances of wh-words, the rise in L*+H was sharper and was produced with higher peak compared to L* followed by a high phrase accent. The author hypothesised that the tonal inventories of both non-standard and standard varieties of Bengali spoken in the east include an L*+H; however, there is a possibility that the tonal inventory of Kolkata Bengali does not. Kolkata Bengali may have a rising pitch accent but it was not found in the studies of Hayes and Lahiri (1991) and Selkirk (2006), which did not differentiate between different kinds of focus. Moreover, unlike earlier studies of focus expression (Hayes & Lahiri, 1991; Truckenbrodt, 2003; Selkirk, 2006), Khan’s analysis gave evidence of post-focal pitch compression and deaccenting. This feature is characteristic of focus expression in English where words following a focused item do not bear accents and show a stretch of smooth pitch interpolation.

In addition to a rise on the focal constituent, the literature includes other tonal events or sequences corresponding to different focus conditions in Bengali. For instance, Lahiri and Fitzpatrick-Cole (1999) also proposed a focus condition where the accented word is assigned L* and the emphatic clitic is realised with an H* tone. Khan’s (2007, 2008, 2014) description of focus prosody differentiates between a focused high pitch accent fH* used to convey surprising and unexpected information, and a focused rising AP where L* is followed by a high edge tone (fHa) and is used on constituents whose focused status is elicited by the use of focus clitics. These, however, were observed less frequently compared with L*+H.

Based on the literature review, the proposed descriptions of Bengali intonation show differences in tonal categories, in turn reflecting differences in the composition of tunes. Hayes and Lahiri (1991) included the downstepped nucleus, where the downstepped pitch (H*) in utterances with broad focus takes a form of a bitonal accent, L+H*. Furthermore, the same tonal categories are found in a different distribution. Khan (2008) and Michaels and Nelson (2004) suggested the use of L* in IP-final phrases in declarative intonation, while Hayes and Lahiri (1991), Lahiri and Fitzpatrick-Cole (1999) and Selkirk (2006) all proposed H* for a smaller phrase followed by a low boundary tone at the end of IPs. It is also important to note that Khan (2014) posited a non-focal L*+H accent in nuclear position.
In summary, there are a number of competing analyses of the intonational phonology and prosody of Bengali. Several aspects introduced in the current models have different phonological categories as well as differences in their interpretation and phonetic realisation. What is of importance to this study are the striking differences in the intonation and prosody between Bengali and well-described Englishes. First, lexical stress is cued by a number of phonetic parameters in English (see §4.2.2) but appears to be phonetically weak in Bengali. Second, accentuation shows a different pattern in Bengali and is more restricted in terms of the tonal realisation than in English.

Third, there are differences in the prosodic marking of focus between the two languages, whether it is L* followed by a high phrase accent or by a bitonal pitch accent. L* is not restricted to any prosodic conditions and the L*+H accent exists in the inventory of pitch accents in AusE, AmE and a number of BrE varieties, but is usually linked to a particular contextual meaning, such as uncertainty or incredulity (Ward & Hirschberg, 1985; Pierrehumbert & Hirschberg, 1990). In English, one would expect the focused word to be associated with L+H* or H* which is often also realised in a higher part of tonal space (Arvaniti & Garding, 2007). Finally, the prosodic hierarchy of English does not include a prosodic constituent (an accentual or a phonological phrase) the size of a prosodic word.

It will be important to investigate whether there will be any mapping of L1 phonological categories, any potential differences in the phonetic realisation of the same phonological categories, or even any modification of L1 features to ‘mimic’ the features of English intonation. In the case of the Kannada L1 group, it is rather challenging to make certain predictions in terms of the possible L1 influence due to the gap in research, and because some features could be characteristic for Kannada as a Dravidian language.

As discussed previously, a number of languages spoken in the subcontinent may show similarities across prosodic features, possibly as a result of language contact. A summary of the relevant features shared by well-documented Indian languages is presented in the next section.

2.4.3 Features shared across Indian languages

There has been considerable research on the languages spoken in India in general. Lexical stress seems to have drawn a great deal of attention, but it still remains a
controversial issue. Limited experimental research is available on intonation and prosody in Indian languages. Hindi (Indo-Aryan), perhaps, is one of the most described among them (Moore, 1965; Ohala, 1986; Harnsberger, 1994; Nair, 2001; Patil et al., 2008; Genzel & Kügler, 2010; Féry & Kentner, 2010) closely followed by Tamil (especially Ravisankar, 1994; Keane 2006a, 2007, 2014), which belongs to the Dravidian group. Current research also includes a few experimental studies on selected prosodic aspects of Assamese (Mahanta, 2001), Tulu (Dravidian) (Narra, Teja, Sneha & Dattatreya, 2012; Mathews & Bhat, 2010), Punjabi (Indo-Aryan) (Dawood, Shahid & Ahmed, 2004) and Konkani (Indo-Aryan) (Menezes, 2002). For the purpose of this study, it is necessary to mention the features that show a degree of similarity and which may be of relevance in the description of IndE intonational phonology and prosody.

Despite the fact that word-level prominence is beyond the scope of the present study, it is relevant because of the possible implications for the prosodic and phonetic realisations of post-lexical prominence, and for the association of pitch accents to stressed syllables in the AM framework. For Hindi, Ohala (1999) suggested that stress is not contrastive and that the concept of stress itself may be controversial and play a marginal role. By contrast, Nair (2001) and Dyrud (2001) reported that $f_0$ is a phonetic cue of stress in Hindi and that it correlates with relatively low pitch on the stressed syllable concomitant with an increase in syllable duration.

Mahanta (2001) carried out a preliminary investigation of lexical prominence patterns in Assamese and examined two acoustic parameters that usually cue stress in English, $f_0$ and duration. Assamese is an eastern Indo-Aryan language closely related to Bengali and does not have vowel length distinction. The researcher found that syllable weight plays a role in lexical prominence placement. In sequences of light syllables, stress falls on the first syllable, but in the instances of a heavy syllable following a light one, stress falls on the second syllable. The acoustic measurements of duration and $f_0$ yielded interesting results. Syllable duration proved to be a reliable measure of lexical prominence, while vowel duration showed no lengthening across syllables. The $f_0$ values showed a significant lowering on the stressed initial syllable, confirming the association of low pitch with lexical prominence in Assamese. In the cases of stress on heavy syllables following light ones, the $f_0$ had a plateau-like shape,
instead of a rise or a fall. Mahanta suggested a spreading of the low tone from the preceding light syllable as a possible explanation.

In a recent study on prominence in Tamil, Keane (2006a) found that duration and intensity did not prove to be reliable correlates of stress. Word-level prominence was marked mostly with the help of $f_0$, indicating either that Tamil does not distinguish prominence on a word level or that the initial syllable has an abstract lexical stress serving as an anchor for the pitch accent. Keane suggested the latter to be a more likely explanation. In his discussion of the typological distinction of word-level prosodic features, Ladd (2008) hypothesised that Bengali and possibly most of the languages in the subcontinent could belong to a non-stress accent group (after Beckman, 1986), where stress is not cued by duration and intensity but is only marked by pitch movement. Ladd further extended Beckman’s typology and proposed that these languages have no lexical specification of pitch, such as the case in Japanese, and pitch is only used on a post-lexical level.

Also of relevance to this study are the $f_0$ shapes associated with accented words and prosodic phrasing. These appear to be among the salient features for a number of languages which bear a resemblance to Bengali. In both Tamil (Keane, 2007, 2014) and Hindi (Harnsberger, 1994; Harnsberger & Judge, 1996; Patil et al., 2008), an utterance usually includes several smaller phrases the size of a prosodic word. Each phrase (AP for Tamil, Keane, 2014; phonological phrase for Hindi, Harnsberger & Judge, 1996; Patil et al., 2008) consists of a low pitch accent and a high phrase accent, thus creating a sequence of rises over the course of an utterance. Unlike in English, where the nuclear-accented word is always right-headed, the strongest word is always a left-most non-clitic word. Moreover, pragmatic focus has little reflection on the prosodic structure and does not necessarily lead to differences in accentuation or phrasing.

Keane (2014) found that the same utterances produced with narrow and broad focus in Tamil showed little difference in accent placement and overall pitch movement. The only distinguishing feature was a higher $f_0$ value of the H tone on narrow focused constituents. Harnsberger (1996) and Patil et al. (2008) suggest that in Hindi narrow focus marking, the prosodic phrases follow the same pattern with a gradual downstep in pitch after the focused constituent. This is a phonologically triggered register compression which lowers the $f_0$ targets for subsequent high tones, rather than raising the element in focus. Féry and Kentner (2010) found little
relationship between prosody and information structure in Hindi, confirming earlier findings that high tones are in a downstepped relationship and are marginally sensitive to information structure. It is possible that pitch compression plays a function here equivalent to deaccenting that is characteristic for other intonational languages such as English (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 2008) or German (Grice et al., 2005).

Genzel and Kügler (2010) investigated broad versus contrastive focus by Hindi speakers (where the speaker chooses an element from a set of alternatives) and reported different findings. Contrary to Patil et al. (2008) and Féry and Kentner (2010), focused words in the Genzel and Kügler study showed an increase in pitch span with both a lowering of the low tone (L) and a higher scaling of the high tone (H), together with significant lengthening of the focus domain. These results indicate that Hindi may potentially employ pitch expansion on the focused constituent similarly to English and a number of other languages. This expansion corresponds to a greater pitch excursion which includes both L and H tones and not just the high target found to be an integral part of focus marking in English (Cooper et al., 1985; Eady & Cooper, 1986; Breen et al., 2013). The researchers, however, compared the same non-phrase initial target words in broad and contrastive focus, and did not compare the f0 values of target words with phrase-initial accented words to investigate the global f0 contours in utterances with a broad versus narrow focus. Nor did they investigate the downstepped relationships of high tones reported in Patil et al. (2008).

Clearly, there are a number of unresolved issues regarding the prosodic features of Indian languages, making this a field of ongoing research. Nevertheless, the studies cited above have shown that the features and processes in the languages spoken in the subcontinent show distinct differences from the prosody and intonation of other longstanding varieties of English, such as the use of lexical prominence and its acoustic cues, tonal inventory, phrasal prominence, focus projection and prosodic structure. These could potentially be reflected in the prosodic phonology and intonation of IndE and are addressed in this study with reference to the four main components of the AM theory: accentual prominence; phrasing and prosodic constituency; tune-to-text alignment, especially the alignment of the rising gesture on accented words; and the use of pitch range in the form of pitch modulations depending on information structure, focus and modality.
2.5 Research questions

Based on the review of the literature and research aims, the present study raises several broad research questions:

1. What intonational features are shared by all speakers in this study and are possibly systematic for IndE intonation in general (due to language contact, developmental features, external factors, and so on)?

2. How do these features compare with other varieties of English?

3. What are the features specific for each group (BE and KE)?

4. What are the differences and similarities in intonation between the two L1 groups and compared with other Englishes? How can these be accounted for in terms of phonological representation, phonetic realisation or both?

5. How are these phonological categories and their phonetic realisation encoded in pragmatic meaning and information structure?

Focusing on the linguistic aspects of intonation and following the AM approach, the study addresses a number of specific questions. These are listed in Table 2.1 below, with hypotheses being presented at the beginning of the relevant experimental chapters. In the table, the main parameters of the AM theory and its details are listed in the left-hand column. Research questions corresponding to the relevant parameters are included in the right-hand column. On the basis of the speakers’ proficiency level and the literature on IndE (Moon, 2002), the research questions assume that the speakers in this study will have an English type system or focus to accent system.

Table 2.1 Parameters within the AM framework with corresponding research questions.

<table>
<thead>
<tr>
<th>AM parameter</th>
<th>Research questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prominence and information structure</td>
<td>1. What are the accentuation patterns in KE and BE?</td>
</tr>
<tr>
<td>• Accentual prominence and its phonetic correlates</td>
<td>2. What are the phonetic correlates of accentual prominence? What is the strength of each phonetic correlate that signals accentual prominence?</td>
</tr>
<tr>
<td>• Intonational realisation of focus</td>
<td>3. What is the effect of focus on accent placement and phrasing for each group?</td>
</tr>
<tr>
<td></td>
<td>4. What are the tonal and non-tonal characteristics of narrow focus?</td>
</tr>
</tbody>
</table>
| Phrasing and prosodic hierarchy | 5. What is the prosodic structure in KE and BE? How many levels of prosodic phrasing are relevant?  
| 6. What are the phonetic cues to prosodic boundaries? |
| Tune-to-text alignment | 7. What is the tone inventory and how do the tones align with the segmental material for KE and BE?  
| 8. What is the alignment of head and edge tones?  
| 9. What are the tune inventories and the use of tunes in different modality contexts? |
| Pitch range | 10. What is the role of pitch range modulation (for instance, pitch reset to mark prosodic constituents, $f0$ increase to mark prominence, and so on)? |
Chapter 3: Methodology

3.1 Introduction
The present chapter outlines methods used in the research project. It begins with a brief description of the laboratory phonology approach adopted in this study (§3.2). This is followed by an explanation of the experimental design which elaborates on the participants, use of materials and data collection process (§3.3). Sections 3.4-3.7 describe the sets of materials, annotation method, acoustic measurements and statistical analyses used in the four experiments. Section 3.8 concludes the chapter.

3.2 Laboratory phonology
Consistent with the majority of intonational studies, this project follows the laboratory phonology approach.12 Over the last two decades, laboratory phonology has expanded and now represents a community of researchers who investigate the human sound system and are interested in enhancing empirical research with the help of experimental methodologies and techniques. Laboratory phonology describes an approach or perspective to research that combines the fields of phonetics and phonology with a contribution from a number of disciplines investigating the use, structure and development of spoken language (Cohn, Fougeron & Huffman, 2012). It builds on the collective work of scholars and is represented by an integration of methodologies. The contributing fields include phonetics, sociolinguistics, speech sciences, second language acquisition, psycholinguistics, and computational linguistics.

The goal of laboratory phonology is to achieve a greater knowledge about phonological structure through experimental phonetic research. The researchers may adopt different phonological theories and disagree on some of the phonological concepts, but, most importantly, they all work towards a better understanding of phonological structure and processes with the help of experimental research that involves mixed methodologies. The advantage is that instead of a top-down explanation of phonological theories, laboratory phonology uses a bottom-up approach and builds on quantitative and experimental research.

12 This term was first introduced by Janet Pierrehumbert in 1987.
The growth of the laboratory phonology community and the expansion of knowledge about phonological theory have led to the development of several research tools for speech analysis together with documentation shared by the whole community (Cohn, 2010). One of these tools is ToBI, the prosodic annotation system described in §2.1.2.

3.3 Experimental design

3.3.1 Participants

Eight male speakers aged from 40 to 50 took part in the study. To ensure a balanced representation of L1 groups, four speakers spoke Bengali as their L1 and four speakers spoke Kannada. All participants were medical practitioners who migrated to Australia from India and were in Melbourne at the time of data collection. Seven speakers were practising psychiatry in Victoria and one was a medical practitioner undergoing training to become a surgeon, also in Victoria. Taking into account that the participants came to Australia on working visas and subsequently applied for residency status, their command of spoken English was very high. The researcher recruited all participants through personal and institutional contacts.

Table 3.1 below gives detailed information about each speaker. In order to maintain anonymity, each participant was assigned a number. The speakers were divided into two groups, where the letters “B” and “K” corresponded to their L1 background. Two L1 Bengali speakers (B1 and B4) came from Kolkata, West Bengal. The other two were born and raised in other cities also in the state of West Bengal. Speaker B3 was born and raised in Katak, located around 420 kilometres from Kolkata, and speaker B2 was from Durgapur, 160 kilometres from Kolkata. Two Kannada L1 speakers (K2 and K4) were from Bangalore, Karnataka. Speaker K1 grew up in Gulbarga, a city 575 kilometres north of Bangalore in Karnataka, which is not far from Hyderabad located in the state of Andhra Pradesh. Speaker K3 was the only participant who spent his early childhood years in Delhi, in the Hindi Belt\(^{13}\) region. His family moved to Puttur in Karnataka, around 314 kilometres from Bangalore, when he was 11.

\(^{13}\) Hindi Belt usually refers to the region of India where Hindi and Hindi dialects are spoken. It encompasses four states: Uttar Pradesh, Rajasthan, Madhya Pradesh and Bihar (Chand, 2009).
Table 3.1 Detailed information about the participants, showing speakers’ L1, other languages spoken, city and state of residence in India, age of learning English, languages used most of the time, and length of residency in Australia.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>L1</th>
<th>Other languages</th>
<th>City and state</th>
<th>Age of learning English</th>
<th>Most used languages</th>
<th>Length of residency</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>Bengali</td>
<td>Hindi, Punjabi, English</td>
<td>Kolkata, West Bengal</td>
<td>5</td>
<td>Bengali, English</td>
<td>7 years</td>
</tr>
<tr>
<td>B2</td>
<td>Bengali</td>
<td>Hindi, English</td>
<td>Durgapur, West Bengal</td>
<td>3</td>
<td>Bengali, English, Hindi</td>
<td>2 years</td>
</tr>
<tr>
<td>B3</td>
<td>Bengali</td>
<td>Hindi, Orya, Assamese, Punjabi, English</td>
<td>Katak, West Bengal</td>
<td>5</td>
<td>Bengali, English</td>
<td>7 years</td>
</tr>
<tr>
<td>B4</td>
<td>Bengali</td>
<td>Hindi, Orya, Assamese, English</td>
<td>Kolkata, West Bengal</td>
<td>5</td>
<td>Bengali, English</td>
<td>6 years</td>
</tr>
<tr>
<td>K1</td>
<td>Kannada</td>
<td>Hindi, Tamil, Marathi, English</td>
<td>Gulbarga, Karnataka</td>
<td>5</td>
<td>Kannada, English</td>
<td>8 years</td>
</tr>
<tr>
<td>K2</td>
<td>Kannada</td>
<td>Telugu, Hindi, English</td>
<td>Bangalore, Karnataka</td>
<td>13</td>
<td>Kannada, English</td>
<td>6 years</td>
</tr>
<tr>
<td>K3</td>
<td>Kannada</td>
<td>Hindi, Tamil, English</td>
<td>Delhi/Puttur, Karnataka</td>
<td>5</td>
<td>Kannada, English</td>
<td>10 years</td>
</tr>
<tr>
<td>K4</td>
<td>Kannada</td>
<td>Tamil, Malayalam, Telugu, English</td>
<td>Bangalore, Karnataka</td>
<td>13</td>
<td>Kannada, English</td>
<td>14 years</td>
</tr>
</tbody>
</table>

As can be seen from the third column of the table, all speakers were multilingual and spoke a number of languages. Most of the speakers were also fluent in Hindi. The most frequently used languages were the L1 and English. It was a preference during the subject selection process that the speakers spoke Bengali or Kannada at home and with their families. However, this was difficult to attain and small exceptions were allowed. Speaker B1 spoke Bengali with his wife and relatives but spoke English with his daughter. Speaker B2 used Bengali with most of his family but Hindi with his wife. In all cases, English was used at work and often to communicate with other emigrants from India whose L1 was different from the speaker’s L1. On average, the onset age for learning English was 5 years of age, except for speakers K2 and K4, who started much later at the age of 13, and speaker B2 who began at the age of 3. All subjects completed their medical degrees in India and had English as a medium of instruction at school and higher educational institutions.
The variable that was the most difficult to control was the length of speakers’ residency in Australia. This varied from 3 to 14 years, with a mean of 7.6 years. None of the speakers reported any hearing difficulties or speech problems. They were given a general overview of the research project and its aims, but did not receive the details about the intonational phenomena investigated in each of the experiments.

3.3.2 Corpus
The data analysed in the study consists of two types: read and spontaneous speech. Consistent with previous research on intonation (Xu & Xu, 2005; Grabe et al., 2005; Lickley, Schepman & Ladd, 2005; Ladd et al., 2009; Jun & Fletcher, 2014), read speech materials were designed to investigate a number of phenomena. Several sets of read materials were created, including short dialogues and read sentences. The samples which were aimed at eliciting narrow or broad focus structures were part of a short dialogue where the target response was preceded by a question (Ueyama & Jun 1998; Xu & Xu, 2005; Jun & Fletcher, 2014). Each data set, along with the annotation, analysis and acoustic measurements, will be described in relation to each experiment in §3.4-3.7.

In order to avoid gaps in the f0 contour, sonorant consonants /m, n, l, r/ were selected in the onset and coda positions of the target accented syllables, as well as at the onset of the following post-stress syllables. Research has shown that, unlike voiceless segments, sonorants show little interruption of the continuity of the f0 contour (Lehiste, 1970; van Santen & Hirschberg, 1994; Rose, 1998). The use of sonorants is a standard approach to designing materials for production experiments investigating intonational features (see Frota, 2002; Xu & Xu, 2005). Sonorants were therefore used as much as possible; however, in order to create cohesive sentences, it was impossible to avoid occasional non-sonorant consonants. Preference was given to voiced segments. Overall, the data sets included a small number of voiceless consonants, mainly /h/, /k/, and /s/, mostly in coda positions or outside the boundaries of target syllables.

The spontaneous speech consisted of a 10 minute narrative and a dialogue of around 8 minutes for each of the speakers. These were used as an additional resource for the investigation of overall intonational patterns, tones and tune inventories in BE and KE. Although not included in the quantitative analyses, they proved to be of great
value as the patterns observed in spontaneous speech matched and confirmed the results of the experiments. All materials are included in Appendix A.

3.3.3 Data collection and recording procedure

The data were collected at four locations around Victoria, in Melbourne, Shepparton, Traralgon and a small town on the Mornington Peninsula. In each case, the researcher travelled to the participants’ preferred destination. Each speaker was recorded individually. Recording sessions were carried out in a small, quiet room at either a participant’s house or workplace. Due to the diverse locations, the recordings were made in a number of one- to two-hour sessions in the period from February 2009 to March 2011. Each participant was recorded over two to four recording sessions. The data collected during the first two months came from two speakers (K2 and B4) and was mostly used as a pilot to test the materials for Experiment 1 and to evaluate the data collection process. Some of these samples were re-recorded in 2010.

The recordings were made with a digital Sony PCM-D50 linear PCM recorder with an external microphone. The microphone was placed on the table in front of each speaker at a distance of about 20 centimetres. The researcher used headphones connected to the recorder in order to listen to any obvious errors, misreading or mispronunciations, or a lack of post-sentence pause. If the researcher determined that a particular sentence was not produced properly, the recording was paused and the utterance was repeated.

For the read sentence data, two sets were recorded in a session, with a 5 to 7 minute interval in between each session. The speakers were given each set and had around 5 minutes to familiarise themselves with the content. The speakers were then asked to read the sentences at their normal speaking rate, as if they were talking to a friend. They were also instructed to leave a short pause after each sentence. The sentences were presented in a randomised order with a number of filler sentences. For the spontaneous speech corpus, the speakers were also encouraged to use their normal speaking rate and were given some time to prepare for the narrative and the dialogue.

The digital files were transferred to a MacBook Pro laptop. All sound files were converted to mono files and coded accordingly using Praat software version 5.1.35 (Boersma & Weenink, 1991-2010). In addition, the read speech files were divided into single-utterance files for each speaker, also with the use of Praat. Read
sentences containing any type of reading error were eliminated. The final set of digital files was imported into the EMU Speech Database System (Harrington, 2010) for further analyses.

3.4 Experiment 1: Intonational realisation of information structure

3.4.1 Materials

Experiment 1 is based on the set of materials closely modelled on the stimuli used by Xu and Xu (2005). The stimuli consisted of a number of simple declarative sentences divided into four main subsets. Each subset consisted of a number of target sentences with a corresponding question, asked by the researcher. The information structure was controlled by having the speakers say each target sentence as an answer to a prompt question, eliciting either narrow focus on a particular word or a broad focus reading. This method has been used successfully in previous studies on focus realisation in English (Cooper et al., 1985; Xu, 1999; Xu & Xu, 2005; Jun & Fletcher, 2014).

Two variables were kept constant: sentence length and sentence structure. Each sentence consisted of five words comprising three content words, one of them a proper name, and two pronouns. The decision to include the words ‘may’ and ‘my’, originally used in Xu and Xu (2005), was made on the basis of previous descriptive and recent experimental research on IndE which indicates accentuation of a great number of words in an utterance, including those that do not carry meaning (see §2.3.2). Traditionally, in AusE, BrE or AmE, these correspond to function words and remain unaccented unless in special contexts (produced with a narrow focus). Therefore, it was possible to investigate how these words were treated by the IndE speakers and whether they affected the accentuation patterns of the whole utterance. Table 3.2 below presents the stimuli in four subsets. It shows the prompt question and the answer read by the participants.

Table 3.2 Prompt questions and stimuli sentences with the target words underlined to elicit narrow focus on subject, verb and object, and broad focus structures.
In each subset, the underlined words correspond to target words that were used in the phonetic analyses and were rotated to form different sentences. The location of the focus was changed from sentence-initial, sentence-medial and sentence-final to neutral or broad focus. The sentences in subsets 1 to 3 were produced with a narrow focus, while the sentences in subset 4 were produced with a broad focus and had more word combinations. Four variables were manipulated for each utterance:

- focus placement
- target word length
- syllable structure
- and lexical stress location.

In addition to a prompt question, the list of printed sentences given to the participants had the target words underlined according to the anticipated focus structure. The speakers were also given an explanation of prominence on target words in narrow focus, simply linking the target word to the prompt questions. No details were given about intonation or accentuation patterns.

The target words included mono-, bi- and trisyllabic words. The number of syllables and the location of lexical stress in target words were varied to systematically change the syllabic position in the word and the distance to the next prominent syllable. In terms of the phonological length, stressed syllables in target words included simple onset with a long or a short vowel, with coda or no coda consonant. The exception was the word *umbrella*, which had a CCV structure. Table 3.3 below shows the segmental composition of each target word. Phonemic transcription of the vowel sounds followed Maxwell and Fletcher (2009, 2010). The target word *umbrella* includes two possible transcriptions to better reflect the phonemic inventory of IndE proposed by Maxwell and Fletcher (2009). The phonemic transcription of the vowel sounds will be revised in the results section (Chapter 4) to match the basis of their phonetic realisation of the vowels by speakers. The location of lexical prominence is indicated with the “‘" diacritic. Possible differences in stress patterns produced by the speakers will be discussed in Chapter 4.
Table 3.3 Target words with their corresponding phonemic transcription (Maxwell & Fletcher, 2009, 2010) used in Experiment 1.

<table>
<thead>
<tr>
<th>Target word</th>
<th>Phonemic transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>/ˈliː/</td>
</tr>
<tr>
<td>Nina</td>
<td>/ˈnɪnə/</td>
</tr>
<tr>
<td>Marina</td>
<td>/məˈrɪnə/</td>
</tr>
<tr>
<td>move</td>
<td>/ˈmuv/</td>
</tr>
<tr>
<td>borrow</td>
<td>/ˈbɔːrə/</td>
</tr>
<tr>
<td>minimise</td>
<td>/mɪnˈmaɪz/</td>
</tr>
<tr>
<td>mill</td>
<td>/ˈmɪl/</td>
</tr>
<tr>
<td>lily</td>
<td>/ˈlɪlɪ/</td>
</tr>
<tr>
<td>umbrella</td>
<td>/ˈʌmbrelə/ ~ /ˈʌmˈbrelə/</td>
</tr>
</tbody>
</table>

The sets were given to the speakers in random order. Only the response and the question corresponding to a particular focus structure were kept constant. In total, the stimuli in sets 1, 2 and 3 included three sentences, while set 4 comprised nine sentences. The target sentences were produced five times. A small number of utterances where the focus was placed on a word other than the target word were removed from subsequent analyses or reclassified to the corresponding focus-position category. The number of sentences recorded in each set is shown below.

- **Set 1** - 3 sentences x 5 repetitions x 8 speakers = 120
- **Set 2** - 3 sentences x 5 repetitions x 8 speakers = 120
- **Set 3** - 3 sentences x 5 repetitions x 8 speakers = 120
- **Set 4** - 9 sentences x 5 repetitions x 8 speakers = 360

**3.4.2 Annotation, acoustic measurements and analysis**

*Annotation*

The digitised files were annotated and analysed manually using the EMU Speech Database System (Harrington, 2010). The segmentation of the data was performed following a standard procedure of annotation criteria for acoustic-phonetic segmentation and word labelling (Peterson & Lehiste, 1960; Croot & Taylor, 1995 for ANDOSL). The segmental boundaries were determined on the basis of visual inspections of a waveform and a spectrogram as well as the auditory inspection of the corresponding .wav file. An illustration of a waveform and an f0 contour with
corresponding annotation tiers are presented in Figure 3.1. The EMU annotation included the four tiers of word, syllable, CV (segments) and tone.

Figure 3.1 Example of annotation in EMU with sound wave, labels and f0 track. The green panel in the top left corner shows the corresponding annotation tiers. Speaker K2.

The acoustic onset and offset of each word in a sentence file were identified and were marked on the word tier. The syllable tier was created to identify stressed (labelled as S), unstressed or weak (W) and accented syllables (A). The annotation of these levels of prominence was based on preliminary auditory analysis and was subject to investigation in this chapter. Taking into account that previous research shows that IndE speakers produce accents (Gumperz, 1982; Wiltshire & Harnsberger, 2006), words that were accentually prominent were identified first for each utterance. On auditory and visual examination, accented words were louder and showed greater f0 movement. In each target word with accentual prominence, stressed syllables were identified and were labelled as accented (A). Stressed syllables were also determined on the basis of auditory analysis: stressed syllables were louder and sounded somewhat more prominent. Whether the target vowel in accented and stressed syllables was produced as a fuller vowel compared to unstressed syllables was more difficult to determine on visual and auditory analyses. Vowel quality and possible reduction of vowels in weak positions will be discussed in the Chapter 4 (see §4.5.3.4). In the words that were not accented, stressed syllables were labelled as stressed (S). All syllables not bearing lexical prominence were labelled as weak (W).
The CV tier included segmentation of the syllables into vowels and consonants (C or V). Phonetic annotation of the segments was not necessary because the measurement could be extracted on the basis of target syllables and words. C1 was labelled at the start of the initial consonant in the accented syllable; V1 at the start of the vowel in the accented syllable, also the beginning of the rhyme; and C2 corresponded to the start of the initial consonant in the post-accented syllable or the coda consonant in closed syllables.

Finally, the tone tier included the annotation corresponding to the f0 contour of the sound file. The intonation contour was broken down into the sequences of low (L) and high (H) tones. At this stage, no assumptions were made in relation to the prosodic structure. The prosodic constituency and structure, based on the findings of Experiments 2 and 3, will be discussed in Chapter 7. The tone targets were placed in the location of the f0 turning points in the vicinity of the target syllables and were identified by applying the following criteria. To identify L targets, the f0 valley, defined as the local minimum (min f0), was taken in the vicinity of the stressed syllable before the rise in the f0 contour. For H tones, the f0 peak, defined as the local maximum (max f0), was taken at the highest point reached in the vicinity of the accented syllable immediately before the fall in the f0 contour. In the instances where no clear peak was visible, the H target was measured at the midpoint of the accented vowel. In addition, the L and H turning points were annotated at the right edges of prosodic boundaries, at the end of full intonational contours.

The digitised sound files were processed using the EMU Speech and Signal Processing tool to create the additional .sfb and .sf0 files required for building hierarchies. The hierarchies allowed the researcher to query the database to extract the necessary acoustic parameters for the relevant type of events, such as words, syllables and tones, and enabled further manipulations of the data within the R statistical package (Ihaka & Gentlemen, 1996) and the EMU-R interface (Harrington, 2010). Figure 3.2 is an illustration of a hierarchical tree of the same utterance presented in Figure 3.1. As shown in the figure, the labels are linked from top-to-bottom which is reflective of the hierarchical relationships between them. On the syllable tier, A stands for accented and S stands for stressed (unaccented) syllables. Each syllable is further linked to the consonant and vowel segments (C1, V1, C2), and each accented syllable is linked to the tonal target or targets found in the vicinity of that syllable. The relevant measurement - for example f0 events (L or H tones) associated with C and V
segments and also associated with accented (A) syllables, were extracted through the hierarchical relationships across the tiers.

![Figure 3.2 Example of the hierarchy in EMU, showing each label in a tier linked to the corresponding label in the tier below.](image)

**Qualitative analysis**

The first part of Experiment 1 investigated overall patterns of accentuation in broad and narrow focus, together with the strategies for focus marking in BE and KE. The following aspects were of interest:

- choice of words for accentual prominence placement
- number of accented words
- presence and degree of pre-focal and post-focal deaccenting
- phrasing
- presence of pauses in post-focal position and their duration
- and the f0 movement (L, H or LH) on the accented target words.

Relevant hypotheses will be presented in Chapter 4 (see §4.3). As discussed in §2.3.2, the literature on IndE indicates that speakers have a tendency to accent more words compared to the speakers of other well-established varieties of English. In order to determine whether there was such a phenomenon and whether there were any differences across the speakers, the number of words bearing accents in broad focus was calculated. Subsequently, a proportional number of accented versus stressed words was generated for each speaker and group in broad focus condition.

One of the aims of this experiment was to investigate and describe the strategies for focus marking used by BE and KE speakers. Taking into account that
deaccenting is one of the main strategies used for narrow focus structure in English (see §2.1.2), the absence of accent or accents before and after the narrow focused target word within one phrase was examined. Where a narrow focused word was accompanied by post-focal phrase break, the duration of the pause was measured.

**Acoustic analysis**

The second part of Experiment 1 aimed at investigating phonetic correlates of information structure, namely accentual and focal prominence. The measurement points used in the analyses were based on the four main correlates relevant for phonetic realisation of prominence in English (described in detail in §4.2.2). These were $f_0$, amplitude, duration, and vowel quality. Target tokens annotated as accented in three narrow focus structures (sentence-initial, sentence-medial and sentence-final) were compared to the same target tokens that were stressed unaccented. For example, the monosyllabic verb *move* in the sentence *Lee may move mill* in example 1 on the left is accented (A), while the same target word in broad focus reading as shown on the right is unaccented, labelled as stressed (S).

```
What may Lee do? - NF
What did you say? - BF

(1) Lee may move (A) my mill
(2) Lee may move (S) my mill.
```

For each target syllable, a number of acoustic measurements were extracted using the EMU-R interface (Harrington, 2010) and analysed with the help of the R statistical package (Ihaka & Gentlemen, 1996; Baayen, 2008). The measurements with their descriptions are presented in Table 3.4 below, and correspond to the four acoustic correlates of prominence in English. As shown in the table, the $f_0$ values were extracted in Hz. These values were then transformed to a semitones scale, which is a psychoacoustic logarithmic scale that captures a hearer’s perception of pitch differences more accurately and accommodates for differences in the use of pitch across speakers (Nolan, 2003). This scale has been used in cross-linguistic and cross-dialectal studies on intonation (e.g. Hellmuth, 2011; Karlsson, House, Svantensson & Tayanin, 2010; Keating & Kuo, 2012). The Hz values underwent conversion with the $100 \text{ Hz}$ value taken as a baseline (Karlsson et al., 2010): $St=12\log_2(\text{Hz}/100)$. All statistical analyses were performed on the $f_0$ values in semitones.
Table 3.4 Summary of the phonetic correlates with the corresponding measurements and description investigated in Experiment 1.

<table>
<thead>
<tr>
<th>Correlate</th>
<th>Measurement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_0$</td>
<td>Mean $f_0$ (Hz)</td>
<td>The mean $f_0$ in stressed, accented and focally accented syllables, extracted at the vowel midpoint.</td>
</tr>
<tr>
<td></td>
<td>$f_0$ trajectory (Hz)</td>
<td>The $f_0$ trajectory (pitch excursion) in the vicinity of accented and focally accented syllables, measured by subtracting the minimum $f_0$ (L tone target) from the maximum $f_0$ (H tone target). This acoustic measure was considered in the presence of a rising $f_0$, marked by the presence of L and H tones.</td>
</tr>
<tr>
<td>RMS-amplitude</td>
<td>dB-RMS values</td>
<td>The mean value measured in stressed, accented and focally accented syllables, extracted at the vowel midpoint.</td>
</tr>
<tr>
<td>Duration</td>
<td>Syllable duration (ms)</td>
<td>The mean durational value in syllables that were identified as stressed, accented and focally accented syllables.</td>
</tr>
<tr>
<td></td>
<td>Vowel duration (ms)</td>
<td></td>
</tr>
<tr>
<td>Vowel quality</td>
<td>F1 and F2 (Hz)</td>
<td>F1 and F2 values measured at the steady state of the vowels in stressed, accented and focally accented syllables. For the diphthong in the word <em>minimise</em>, the values were extracted in the first third part of the vowel, before the offglide.</td>
</tr>
</tbody>
</table>

In addition to the F1 and F2 analyses conducted separately, several target vowels were plotted in the F1xF2 plane for each speaker. Where the vowel showed overlap in its realisation across prominence levels (accented, focally accented or stressed), Euclidean distances measuring the acoustic straight line in the F1xF2 plane were examined following Harrington (2010). The hypotheses related to each of the measures will be included in Chapter 4 (§4.3) following a more detailed literature review on accentual prominence and focus in English (§4.2).
3.4.3 Statistical analyses

Taking into account the importance of experimentation, statistical methods are commonly employed in laboratory phonology research. With the increase in experimental production and perception studies and collaborative work with computational linguists, there has been a strong movement towards developing more sophisticated techniques for speech analyses and a wider use of the R statistical package (Ihaka & Gentlemen, 1996; Baayen, 2008; Johnson, 2008). According to Kingston (2012), R allows the user to get into ‘the nuts and bolts’ of the statistical analysis and provides a better understanding of the processes and outcomes of manipulations. One of the more recent explorations into statistical methods is the application of linear mixed-effects models (LMM) (see van de Ven & Gussenhoven, 2011 on final rises in Dutch intonation; Oh & Redford, 2012 on investigation of fake geminates in English).

van de Ven and Gussenhoven (2011) applied LMM procedures to investigate the alignment of the right-most L tone corresponding to the phrase-final rise in Dutch. Using two regression models and then comparing them in ANOVA for better ‘fit’ allowed the authors to assess the linear relations between the absolute timing of L and potential anchoring sites. As a result, they were able to identify the segmental points affecting the timing of the L. Unlike any another statistical calculations, LMM modelling allowed for the factoring of speaker differences and thus could account for intra-speaker variation.

In the present study, the main statistical manipulations were performed by a series of LMM analyses on a number of parameters with the help of the ‘lme4’ and ‘multcomp’ packages in R. In LMM, fixed factors are modelled by means of contrast while random effects are modelled as random variables with a mean of zero and unknown variance (Baayen, 2008). This approach reduces data variation as a result of speaker idiosyncrasies and examines the predicting factors by minimising the speaker effect, thus providing more robust results than traditional analyses (Bates, 2005). By applying this approach, it was possible to adjust for the differences between individual speakers (such as their speech rates). In a series of LMM analyses, SPEAKER was treated as a random factor, while L1 GROUP, prosodic POSITION (sentence-initial, -medial and -final), CONDITION (accented, focally accented and stressed) and VOWEL type (where relevant) were included as fixed factors. These
fixed factors were modelled for each of the measurements listed in Table 3.4 (RMS-amplitude, $f_0$, duration and F1/F2 values) separately. The results were considered significant when the $p$-value was smaller than 0.05. Post hoc Tukey tests were performed to investigate the effect of the relevant factors and possible interactions of factors. The full set of LMM analyses is included in Appendix B.

3.5 Experiment 2: Tonal alignment and scaling characteristics of rising accents

3.5.1 Materials

As discussed in §2.3.2 and §2.4.3, a rising pattern associated with accented words has been noted for IndE and is a feature shared by several Indian languages. Therefore, the intent of Experiment 2 was to investigate the alignment of the rising pitch movement on accented syllables in prenuclear and nuclear focal positions in order to determine the phonological category or categories. The speech materials used in this experiment consisted of two sets designed to elicit the production of target words in different prosodic conditions: broad and narrow focus.

The first set of read sentences investigated prenuclear accents and consisted of four sentences with a target word at the beginning. All sentences had the same modality and structure, in that they were simple declaratives consisting of four words. These sentences were also part of the materials used in Experiment 4. The list of sentences with the corresponding target word marked in bold is presented below.

**Set 1 - Prenuclear accents**

(W1) *Lara* lives in Lilydale.
(W2) *Lulu* lived in Melbourne.
(W3) *Nelly* likes yellow lemons.
(W4) *Maryellen* may feel mellow.

The second set consisted of question-answer pairs to elicit narrow focus. The target words were inserted in a carrier sentence with a simple declarative structure of various syntactic lengths. The same set of sentences was used to investigate the phonetic cues to prosodic boundaries and will be discussed in relation to Experiment 3 (§3.6 of this chapter). During the recording session, the researcher asked the
question and the participant had to respond to the question, placing focal prominence on the target word. The complete set with the prompt question is presented below. The target words (proper names) used for measurements and analysis are marked in bold.

Set 2 - Nuclear focal accents

Q. Who memorised Manuela’s manual?
(W1) Lulu memorised Manuela’s main/luminous/yellow-coloured/manual.
(W2) Maryellen memorised Manuela’s main/luminous/yellow-coloured/manual.

The target words from both sets contained open syllables in the accented position. Three target words had word-initial stress (Lara, Lulu and Nelly). The accented syllables contained short and long vowels. Phonemic transcription with the corresponding target word is included in Table 3.5 below. Similar to Experiment 1, the transcription of vowel sounds follows Maxwell and Fletcher (2009, 2010).

Table 3.5 Phonemic transcription of the target words used in Experiment 2.

<table>
<thead>
<tr>
<th>Target word</th>
<th>Phonemic transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lara</td>
<td>/læːra/</td>
</tr>
<tr>
<td>Lulu</td>
<td>/luːlu/</td>
</tr>
<tr>
<td>Nelly</td>
<td>/nɛlɪ/</td>
</tr>
<tr>
<td>Maryellen14</td>
<td>/ˌmærɪˈɛlɛn/</td>
</tr>
</tbody>
</table>

As already outlined in §3.3.2 of this chapter, all target words consisted of sonorants in order to avoid any visible perturbation or errors in the f0 tracking. It was also expected that using sonorants would minimise the differences in the alignment of the peak, reported for other consonants - for example, later alignment for fricatives in comparison to sonorants (Arvaniti et al., 1998). Previous research on AmE has shown

14 Possibly as a reflection of the differences in lexical stress placement in IndE (§2.3.1), all speakers produced the word Maryellen with primary stress placed on the first syllable.
that sonorant consonants behave similarly in terms of peak alignment properties (van Santen & Hirschberg, 1994).

Each speaker was asked to read the sentences five times in a natural manner at normal speed. The number of target words annotated is shown below.

\[
\begin{align*}
\text{Set 1} & - 4 \text{ tokens} \times 5 \text{ repetitions} \times 8 \text{ speakers} = 160 \\
\text{Set 2} & - 2 \text{ tokens} \times 5 \text{ subsets} \times 5 \text{ repetitions} \times 8 \text{ speakers} = 400
\end{align*}
\]

In prenuclear prosodic position, the speakers produced some of the repetitions with a sustained high pitch starting at the onset of the accented word. This was more characteristic of L1 Kannada speakers. These tokens were not included in the analyses of this chapter, and will be discussed in the pitch accents inventory in Chapter 7 (§7.5.1). A total of 146 target words were used in the analyses of Set 1 (prenuclear accents), and 329 target words were used in Set 2 (nuclear focal accents).

3.5.2 Annotation, acoustic measurements and analysis

Digitisation and labelling were performed following the method described in Experiment 1 (§3.4.2). The only difference in the present experiment was that the EMU annotation included the three tiers of word, phonetic, and target. Figure 3.3 shows an example of a waveform and an f0 contour with the corresponding annotation tiers. The phonetic tier included segmentation of the words into corresponding sounds. The segmentation of the sounds was performed following the procedure described in §3.4.2. The target tier was used for the measurement points corresponding to segmental boundaries (onset and offset) and tonal movements. For ease of further manipulation and statistical analyses, these measurement points were placed on the same tier.

The same acoustic landmarks were identified for each target word following the criteria described in §3.4.2; namely L and H targets, syllable onsets (C1, C2), vowel onset (V1). An additional label was added at the end of the accented target word (Woff). An example of the target word Lulu with the acoustic landmarks is shown in Figure 3.3 below. The f0 movement presents a clear rising pattern, where the low and the high turning points are labelled as L and H accordingly. The f0 landmarks were identified by eye. In general, the H target was relatively easy to
locate. It was usually identified as a peak located towards the end of the accented syllable or in the post-accented syllable. The L target sometimes showed perturbation or a break in the $f_0$ due to creaky modality, usually at the onset of the accented consonant. In these cases, the L measuring point was labelled at the beginning of visible $f_0$ contour. For each $f_0$ target (L or H), two values were extracted. These were the $f_0$ in Hz and the time point in milliseconds (ms). For the measuring points corresponding to segmental or word boundaries, only the duration in milliseconds was extracted.

![Figure 3.3 Waveform display, $f_0$ contour, and annotations in EMU corresponding to the word Lulu produced by speaker B1. The L target corresponds to the first elbow in the $f_0$ contour, while the H target corresponds to the $f_0$ peak.](image)

Following previous research on tonal alignment\textsuperscript{15} (Prieto & Torreira, 2005; Arvaniti, Ladd & Mennen, 2006; Ladd et al., 2009), in order to investigate the alignment of the two targets with the segmental material and the alignment of the rising gesture as a whole, as well as the scaling characteristics, the following temporal distances and tonal values were obtained:

- $C_{dur}$ - duration of the accented consonant, distance between C1 and V1
- $V_{dur}$ - duration of the accented vowel, distance between V1 and C2
- $S_{dur}$ - duration of the accented syllable, distance between C1 and C2

\textsuperscript{15} A more detailed review of the literature on tonal alignment will be presented in §5.2.
• ContoL - temporal interval between the accented syllable onset (C1) and the beginning of the rise (L)
• ContoH - temporal interval between the accented syllable onset (C1) and the f0 peak (H), also referred to as peak delay in Chapter 5
• VontoH - temporal interval between vowel onset or rhyme onset and the f0 peak (H)
• HtoVoff - peak location (H) relative to the offset of the accented vowel (C2) measured by subtracting C2 values from H values, with negative values recorded when the peak was realised before the vowel offset
• HtoWoff - peak (H) relative to the offset of the target word (Woff)
• LtoH - the temporal interval between the L and the H targets corresponding to the duration of the rising gesture
• and LHrise - f0 slope, the scaling difference between the L and the H target or rise magnitude, measured by subtracting the L value from the H value.

As shown in Figure 3.3, the peak was examined relative to four segmental landmarks (syllable onset, rhyme onset, accented vowel offset and word offset). The temporal interval between the accented vowel offset and H (HtoVoff) was chosen to investigate whether the peak aligned within the accented vowel or beyond its boundaries in order to posit a certain phonological category (Arvaniti & Garding, 2007; Ladd et al., 2009). The distance between the H target and the word offset was measured to test whether H aligned with the edge of the accented word. This helped determine the type of tonal event for the peak as either a pitch accent or a phrase tone delimiting the edge of a minor phrase.

Earlier research on tonal alignment (Silverman & Pierrehumbert, 1990) used rhyme onset as a reference point for the alignment of the f0 features, suggesting that the onset duration should not affect the location of the peak and the degree of peak delay. However, a number of studies have shown that the onset of the accented syllable also plays an important role in the alignment of putative targets. For example, longer onset together with a longer vowel contributes to a later peak location in English (van Santen & Hirschberg, 1994), in Mexican Spanish (Prieto et al., 1995) and in Greek (Arvaniti et al., 1998). In the present study, the decision was made to include both measures, VontoL and ContoL, and to investigate which of these would be more accurate in measuring the location of the peak.
Previous research on a number of languages indicates that there is a more complex relationship between the peak and the duration of the segments in accented syllables (Silverman & Pierrehumbert, 1990; Dilley, Ladd & Schepman, 2005; Ladd et al., 2009 for English). Moreover, a number of studies have shown that the alignment of the peak can be influenced by emphasis condition, leading to its delay (Ladd & Morton, 1997; Arvaniti & Garding, 2007). All temporal intervals (ContoL, ContoH, HtoVoff and LtoH) in the present study were converted into proportional measurements in order to measure the alignment of the targets relative to the duration of the accented syllable. Consistent with the methodology (see §3.4.2), all Hz values were converted to semitones. The list of hypotheses in relation to the measures discussed above will be presented in §5.3 after a comprehensive overview of the literature on alignment and scaling of tonal targets.

3.5.3 Statistical analyses

A series of LMM procedures with post hoc Tukey tests were used to investigate the alignment and scaling characteristics of prenuclear rises and nuclear focal rises, as well as to compare the two. For each of the measures (e.g. ContoH or LtoH interval), a model was fitted to assess the effect of the L1 group and the segmental duration or composition on the alignment of the targets. GROUP, WORD and VOWEL were included as fixed factors, while SPEAKER was treated as a random factor. For the comparative analysis of focal and prenuclear rises, both SPEAKER and WORD were included as random factors. This is because the main aim of this investigation was to look for potential differences between speakers of the L1 groups in two prosodic positions in order to justify the phonological category or categories for the rises proposed in separate analyses for prenuclear and nuclear focal rises. Less than 0.05 p-values were considered significant. The full set of LMM results is presented in Appendix C. In addition to LMM, correlation tests were performed for a number of temporal intervals and the f0 slope. These will be presented in the results section of Chapter 5.
3.6 Experiment 3: Durational cues to prosodic boundaries

3.6.1 Materials and measurements

Experiment 3 was designed to investigate boundary-related lengthening in BE and KE in order to examine the levels of prosodic constituency. The tonal alignment experiment (§3.5 above) enabled testing of whether the prosodic constituency of BE and KE includes a minor phrase, roughly the size of a prosodic word. It was thus important to determine whether the two sub-varieties had the intermediate phrase in addition to the intonational phrase level posited for AusE and AmE, or whether there was only one higher-level prosodic constituent in the form of the intonational phrase.

The materials used in this experiment were the same as those described in Set 2, Experiment 2 (§3.5.1). The target utterances were produced as an answer to a prompt question *Who memorised Manuela’s manual?* The words *main, luminous* and *yellow-coloured* were gradually inserted to vary the length of the noun phrase. In total, 15 combinations were created with 5 repetitions of each utterance. The words *Lulu, Emily, Marina and Maryellen* were rotated to form a number of sentences in the following template:

*Lulu/Emily/Marina/Maryellen memorised* **Manuela’s**
*main/luminous/yellow-coloured/manual.*

In order to examine potential differences in the amount of lengthening depending on the prosodic level, the duration of the final syllable in the target word *Manuela’s* was measured in milliseconds. By manipulating the duration of the segmental material after *Manuela’s*, it was possible to create different prosodic conditions. It was anticipated that the utterances with an extra short noun phrase (one word) would be produced with no boundary after *Manuela’s*, while the utterances with two words or more words included in the noun phrase (i.e. *main luminous manual*) would lead to changes in phrasing of the whole utterance and result in the word *Manuela’s* being produced in a separate phrase. Most likely, due to the target word being a proper name, all speakers accented the word *Manuela’s*. In addition, the word was produced with lexical prominence on the first syllable.

The digitised sound files with the corresponding annotation files from Experiment 3 were moved into a separate EMU database. The utterances were re-
examined and additional labels on the syllable tier were added to identify each prosodic level with which the final syllable was associated. These labels were:

- W - no indication of phrase boundary, corresponding to word level
- ip - minor phrase boundary usually with some degree of disjuncture, possible pitch reset after the target word
- and IP - major phrase boundary, often followed by a pause.

The three boundary types were identified following standard ToBI labelling procedures and determined on the basis of the degree of disjuncture (Beckman & Ayers-Elam, 1997; Cambier-Langeveld, 2000). Auditory and visual analyses were performed equivalent to a break index annotation reflective of perceived boundary strength: 1 was the lowest form of disjuncture corresponding to the prosodic word (W), 3 was the intermediate level of disjuncture (ip) and 4 was the major level of disjuncture (IP). A number of additional cues were used to determine the level of juncture. Among these were the presence or absence of a pause, strength of the edge tone after the pitch accented syllable on Manuela’s, and pitch range reset after the target word.

During the annotation process and based on impressionistic analysis, it was found that the utterances that did not have an expanded noun phrase (A) were produced with a level 1 disjuncture. The utterances with two or three words were produced with a minor phrase break, perceived as a level 3 juncture, corresponding to the ip in ToBI, shown in (B); and the utterances with three or more words included in the noun phrase were produced with a major phrase juncture corresponding to the intonational phrase or level 4 break index (C).

(A) [[Lulu memorised Manuela’s manual]\IP.
(B) [[Lulu memorised Manuela’s]IP [main luminous manual]\IP.
(C) [[Lulu memorised Manuela’s]IP [main luminous yellow-coloured manual]\IP.

3.6.2 Statistical analysis
Consistent with the methods used in this study, the durational values for the final syllable in Manuela’s across three prosodic levels were extracted using EMU/R and
subjected to LMM procedures with SPEAKER as a random factor and the prosodic POSITION for each final lengthening domain as a fixed factor. A total of 370 tokens were used in the analysis. Less than 0.05 p-values were considered significant. Any significant interactions were identified using the Tukey HSD method. Where necessary, raw values were submitted to a series of t-tests for individual speakers. The full set of statistics is included in Appendix D.

3.7 Experiment 4: Tone and tune inventory

3.7.1 Materials

Experiment 4 was designed for the analysis of intonational patterns in BE and KE in order to present a complete tone inventory and to describe the nuclear tunes applied in various modality contexts. Two types of data were included in the corpus: read speech and spontaneous speech.

Read speech

The first set of read speech materials consisted of a list of read sentences with different syntactic structures. These included simple and complex declaratives, polar questions, wh-questions, interrogative coordinations, and imperatives. Examples of a simple declarative and a polar question are provided below. The full list of sentences is attached in Appendix A.

Simple declarative

Nelly likes yellow lemons.

Polar question

Does Lara live in Lilydale?

Read sentences have been used successfully in laboratory speech research on intonation (Jun & Fletcher, 2014). Moreover, this design has been previously applied to investigate the tonal and nuclear tune inventory as well as other intonational phenomena in varieties of English. This set of materials was modelled on the IViE corpus investigating English intonation in the British Isles (Grabe, 2004; Grabe et al., 2005).

Consistent with the study design, speakers were asked to read each sentence five times. The sentences were presented in a randomised order, whereby questions were mixed with declaratives, coordinations were mixed with wh-questions, and so
on. Utterances that were produced with false starts or had mispronounced target words affecting meaning were removed from further annotation and analysis. Table 3.6 below shows the total number of stimuli recorded and analysed, presented by sentence type.

Table 3.6 Number of stimuli recorded and analysed, presented by syntactic structure type.

<table>
<thead>
<tr>
<th>Sentence type</th>
<th>Stimuli recorded</th>
<th>Stimuli analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declaratives</td>
<td>$6 \times 5 \times 8 = 240$</td>
<td>228</td>
</tr>
<tr>
<td>Polar Questions</td>
<td>$5 \times 5 \times 8 = 200$</td>
<td>181</td>
</tr>
<tr>
<td>Wh-Questions</td>
<td>$5 \times 5 \times 8 = 200$</td>
<td>187</td>
</tr>
<tr>
<td>Interrogative Coordinations</td>
<td>$5 \times 5 \times 8 = 200$</td>
<td>186</td>
</tr>
<tr>
<td>Imperatives</td>
<td>$1 \times 5 \times 8 = 40$</td>
<td>39</td>
</tr>
<tr>
<td><strong>Total: 821</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similar to the other three experiments, preference was given to sonorant consonants to avoid breaks in the $f0$ contour due to voiceless segment perturbation. However, it was not always possible to completely exclude voiceless consonants while ensuring the sentences were cohesive and meaningful. Set 1 included utterances of different syntactic length, with a small number of sentences having relative clauses. The target words also varied in the number of syllables and their syllabic composition. In terms of the information structure, all sentences assumed a broad focus production.

As can be seen in Table 3.6, this set of materials includes twice as many questions as statements. In order to balance the corpus and have a relatively even ratio of questions to statements, some of the recorded and annotated stimuli from Experiment 1 (§3.4.1, this chapter) were added to this analysis. It was also important to include more utterances with narrow focus to allow for the investigation of pitch accent types, boundary configurations and nuclear tunes typical of narrow focus structure, thus providing a more detailed overview of the intonational system for BE and KE. As a result, Set 2 of read speech encompassed simple declarative sentences in broad focus and narrow focus. Only a proportion of the large number of utterances analysed in Experiment 1 were selected for this experiment. These are listed below together with the elicitation questions.

- **BF** What did you say? – Lee/Nina may move my mill.
- **NF** Who may move my mill? – Lee/Nina may move my mill.
NF  What may Nina/Marina do? – Lee/Nina may move my mill.
NF  What may Lee/Nina/Marina move? – Lee/Nina may move my mill.

Five repetitions of each utterance were used in the subsequent analysis (9 sentences x 5 repetitions x 8 speakers = 360). Unlike Experiment 1, instances of ambiguity in the narrow focus placement or cases where an utterance was produced with an unexpected focus (such as when the verb was accented instead of the intended noun) were kept in the corpus for Experiment 4. Differences in narrow focus placement did not affect the analysis as the utterances were grouped by two criteria only. These criteria were broad or narrow, disregarding the prosodic position of the focused word. Only utterances with errors in the f0 tracking or false starts were removed from Set 2, leaving a total of 296 utterances for the analysis.

**Spontaneous speech**

Since the two controlled speech sets did not include the full range of syntactic structures possible in the English language and may not have shown the full extent of intonational variation, it was useful to draw on spontaneous speech, particularly when looking for further evidence of patterns emerging from controlled stimuli. The spontaneous speech sample consisted of dialogue data in a form of a role-play and a short narrative. For the dialogue data, participants were given a task similar to one they perform in their daily practice. This enabled recording of a more naturalistic speech, close to their working environment. Each participant was given the topic of depressive disorder and was asked to elicit current symptoms as well as explain the diagnosis of this disorder. The role of the patient was played either by the researcher or an educator on the recording site. Educators were given some coaching and necessary information by one of the participants at the beginning of the recording period. The dialogues recorded ranged from 8 to 10 minutes in duration, and included a wide range of questions and statements. The utterances with a relatively clear f0 shape were selected for annotation. Across all speakers, a total of 202 utterances were annotated, comprising 97 for the KE group and 105 for the BE group. It is important to note that these utterances were examined and checked for patterns but were not included in the quantitative analyses. They were used for illustrative purposes, as discussed in Chapter 7.
According to Arvaniti (2011), more research needs to be conducted to understand the relationship between melodic patterns and meaning. Spontaneous speech is a very complex area in the AM research with currently only a handful of studies addressing the pragmatic interpretation of tones and tunes in such data. A full annotation and investigation of spontaneous dialogues remains the subject of future research. In this study, the main aims are to describe the choice of nuclear contours in a range of statements and questions in order to present an inventory and to examine the frequency and proportional use of these tunes depending on the sentence type.

3.7.2 Annotation, acoustic measurements and analysis

The annotation and analyses of data in Experiment 4 were performed following the procedures described in detail in §3.4.2 and §3.5.2. The EMU annotation included four tiers: type, word, syllable and tone. An illustration of a waveform and an $f_0$ contour with corresponding annotation tiers are presented in Figure 3.4.

![Figure 3.4 Example of annotation in EMU with sound wave, labels and $f_0$ track. The green panel in the top left corner shows the corresponding annotation tiers. QC on the type tier stands for a polar question produced as one intonational phrase. The tonal categories are shown on the tone tier.](image)

In this experiment, the type tier included the segmentation of the utterances into intonational phrases, corresponding to full intonational contours (as shown in the figure below). Each utterance was coded for syntactic structure type - for example, simple declarative (SD), complex declarative (CD) or question (Q). In addition, the intonational phrases were coded for their position in an utterance as either medial (M)
or final (C). This enabled the investigation of nuclear tunes in utterance-medial and utterance-final position. The word tier was devised for word level segmentation, while the syllable tier included information on the types of syllables. While the segmentation of the word or syllable boundaries and identification of stressed (S), unstressed (W) and accented (A) syllables were the same as in §3.4.2 and §3.5.2, word-final syllables coinciding with a phrase edge were also annotated on the basis of prosodic boundary type (Fi and F). Here, Fi referred to the edge of a minor phrase boundary and F referred to the end of the final syllable coinciding with the edge of an intonational phrase.

On the tone tier, the \( f0 \) contours were first examined as sequences of low (L) and high (H) tonal events, consistent with the previous experiments. Subsequently, following ToBI annotation criteria for AmE and AusE (Beckman & Ayers-Elam, 1994; Fletcher & Harrington, 2001; Fletcher et al., 2002), three types of tonal events were identified. These were pitch accents associated with the metrically strong syllables of the accented words (such as \( H^* \)), phrase accents denoting a minor phrase boundary (H- or L-), and boundary tones denoting major intonation units or intonational phrases (L% or H%). The edge tones and the boundary tone at the end of intonational phrases were annotated as a configuration. This is shown in Figure 3.4 as H-H%.

In addition to the tonal events, the following \( f0 \) landmarks were annotated on the tone tier:

- **low \( f0 \) (Hz)** - the lowest point in the \( f0 \) contour measured at the right edge of an intonational phrase in utterance-final intonational phrases; those produced with a falling contour at the right edge
- **HiF0 (Hz)** - the highest point in the \( f0 \) associated with a pitch accent that had a high tonal component, either a monotonal \( H^* \) or a bitonal accent, identified for all declarative utterances.

For further analysis, all values in Hz were converted to semitones. The lowest \( f0 \) was included in the analysis of the speaker’s use of pitch and was used together with the HiF0 values to determine the pitch span window for each speaker.

In order to investigate possible pitch expansion on the first accented word observed in interrogative and imperative intonation, as noted in a number of languages including Bengali (Hayes & Lahiri, 1991) and Tamil (Keane, 2014), the
highest point in the $f_0$ associated with a pitch accent that had a high tonal component was also identified and labelled for other sentence types. These labels included:

- QHiF0 and WhHiF0 - measured at the H peak associated with the first high or rising pitch accent of the first intonational phrase in polar (Q) and wh-questions (Wh)
- DHiF0 and IHiF0 - measured at the H peak associated with the first high or rising pitch accent of the first intonational phrase in declaratives (D) and imperatives (I).

Using the EMU/R interface (Harrington, 2010), different types of tones on the tone tier were extracted for KE and BE separately. Similarly, the nuclear tunes consisting of a nuclear pitch accent with a following boundary tone configuration were extracted for each group. Analyses of pitch accents and nuclear tunes distribution were performed. The nuclear tune type and frequency data were extracted by sentence type and are presented in the relevant sections of Chapter 7. Illustrations provided were all made using Praat software. LMM analyses were performed on the HiF0 across different sentence types. The full set of LMM results is included in Appendix E.

3.8 Chapter summary

This chapter has outlined the design of the study, highlighting the materials and methods used for the investigation of several phenomena. It has described the laboratory phonology approach and justified the experimental methods and LMM analysis used for statistical manipulations. Detailed information on the speakers participating in the study was given, including their linguistic and sociogeographic profiles. The four experiments with their sets of materials, annotations, acoustic measurements and subsequent methods of analysis were presented. The experiments were designed to investigate several phenomena and were aimed to give the first full description of the intonational phonology of IndE, together with the variations and similarities between the two sub-varieties in this study, BE and KE. The experiments included:

1) information structure and acoustic cues to accentual and focal prominence, to be discussed in Chapter 4
2) tonal alignment of prenuclear and nuclear rises, discussed in Chapter 5
3) boundary-related lengthening to determine constituency level in Chapter 6
4) and tone and tune inventory in Chapter 7.

The hypotheses tested in each experiment will be included in the relevant chapters before the results sections. The results of the experiments are presented in the next four chapters and begin with the information structure and accentual prominence in BE and KE.
Chapter 4: Experiment 1 – Intonational realisation of information structure

4.1 Introduction

As discussed in §2.1.1, one of the essential components of intonational phonology within the AM framework is relative prominence, signalled by some syllables in an utterance being more prominent than others. Different degrees of prominence are recognised. These are lexical, in the form of stressed or unstressed syllables, and post-lexical. Stressed syllables bearing pitch accents correspond to a prominence level above the word, referred to as post-lexical or accentual prominence. The use of accentual prominence is linked to the meaning of an utterance and the interpretation within its context. This meaning, expressed with the help of prominence relations among its elements, corresponds to ‘information structure’. The present chapter is concerned with two main notions of information structure: ‘given’ and ‘new’ information, or givenness versus focus (Prince, 1981; Brown, 1983).

The literature accepts that focus is marked in a number of ways, such as syntactic, morphological, morphosyntactic or prosodic (Gussenhoven, 1999; Adamou & Arvaniti, 2010). Of particular interest to this chapter is the prosodic marking of focus, one of the most common strategies used across languages (Adamou & Arvaniti, 2010). Marking focus prosodically involves a number phonological as well as gradient phonetic ways (Frota, 2002). As discussed in detail in §2.1.1, among these are pitch accent type, change in phrasing, greater pitch excursion on accented words, post-focal pitch range compression or lengthening of the focal domain.

To date, a substantial amount of research has been conducted on the prosodic marking of information structure in English and its varieties such as AmE, BrE and AusE. In these varieties, the intonational realisation of focus structure is marked by the choice of relative prominence placement. While accented words are associated with new or important information, unaccented words usually carry given information. Thus, deaccenting is one of the main phonological strategies employed by the speakers of well-established varieties of English for marking givenness (Gussenhoven, 1983, 1999; Ladd, 1980, 1996, 2008; Selkirk, 1984, 1995; Xu & Xu, 2005). In addition, previous findings on the phonetics of focus realisation in English has shown higher $f_0$ peaks on the focused syllables as compared to other accented
syllables in an utterance (Liberman & Pierrehumbert, 1984; Cooper et al., 1985; O’Shaugnessy, 1979; O’Shaugnessy & Allan, 1983; Breen et al., 2010), as well as longer duration (Cooper et al., 1985; Wells, 1986; Chen et al., 2001; Breen et al., 2010), and a greater intensity (Sluiter and van Heuven, 1996b; Breen et al., 2010) or perceived loudness (Wells, 1986) of the segmental domain in focus.

While recent studies on the information structure, accentuation and focus marking in new varieties of English reported an influence of first languages on these varieties (Gusenhoven & Udofot, 2010 on Nigerian English; Zerbian, 2013 on varieties of BISAfE), little experimental research has been conducted on accentuation and marking of focus in IndE. Apart from Moon’s study (2002) on phonetic cues to focus in English spoken by Hindi and Telugu L1 speakers (see §2.3.2, Chapter 2), most of the work on accentual prominence and intonational realisation of focus in IndE has been descriptive or preliminary (Wiltshire & Harnsberger, 2006 on accentuation).

The present experiment addresses this gap. It is concerned with the surface structure realisation of focus and investigates the intonational and phonetic realisation of accentual and focal prominence, following an indirect-relationship approach (Pierrehumbert 1980; Ladd, 2008; Gusenhoven, 1983) where phonological categories facilitate the relationship between the acoustic features and meaning in discourse. This chapter aims to investigate the following:

a) accentuation patterns in focal and non-focal contexts
b) patterns of phrasing and how they relate to information structure
c) tonal and non-tonal cues to focus
d) and acoustic correlates of post-lexical prominence.

This chapter is outlined as follows. Section 4.2.1 introduces the phenomena of prominence and focus, and outlines some current theories on focus including ways of describing focus expression in English that were not covered in the literature review. Section 4.2.2 reviews a number of acoustic cues to prominence, with an emphasis on English. Section 4.2.3 gives an overview of focus marking cross-linguistically. Section 4.3 lists a number of hypotheses followed by a brief description of the experiment in §4.4. Section 4.5 presents the results of the experiment, beginning with a qualitative analysis of prominence and accentuation in non-focal contexts in §4.5.1. This is followed in §4.5.2 by findings on the prosodic realisation of focus structure.
that describe the strategies for focus marking in BE and KE. Subsequently, §4.5.3 presents the results of the acoustic correlates of prominence. Section 4.6 includes the discussion that addresses differences and similarities between the two L1 groups, as well as giving a comparison of the findings with previous research on IndE, other varieties of English and studies on L2 speakers of English. The chapter concludes in §4.7 with a summary of the findings and discussion points.

4.2 Accentual prominence and information structure

4.2.1 Prominence and focus

Due to the considerable body of research on information structure and focal prominence, there are various schools of thought about the treatment of focus in intonation and its functions. Such terms as ‘focus’, ‘narrow focus’, ‘contrastive focus’ and ‘broad focus’ may be interpreted differently depending on the theoretical approach taken. Moreover, as noted by Ladd (2008), the use of accentual prominence and focus expression cannot always be explained in a straightforward way, and the notion of ‘broad focus’ in particular has been a debatable issue in the literature for quite a long time.

Currently, the literature recognises two dimensions across which prosodic realisation of focus can vary. These are focus size and focus type. The distinction between narrow and broad focus (Ladd, 1980, 2008; Gussenhoven, 1983; Selkirk, 1995) corresponds to the first dimension and is defined by the size of the focused domain and the number of words involved. Narrow focus refers to the parts of an utterance produced with greater prominence on one word, while broad focus refers to the whole utterance and focuses on the ‘entire event’ (Breen et al., 2010). Examples of two focus structures are illustrated below.

(2) What did you say? – [Lee borrowed your umbrella]. Broad

As shown in example 1, the question Who borrowed my umbrella? is aimed at seeking new information about the subject, the person who may have borrowed the umbrella. Thus, in the answer, a special emphasis is expected to be placed on the

---

the help of a nuclear pitch accent. The segmental domain corresponding to narrow focus is limited to the word Lee. The rest of the post-focal words bear no pitch accents. The question in example 2, by contrast, is not aimed at eliciting any new information. It asks to restate the information given previously. In this case, the whole sentence forms one focal domain with two words bearing pitch accents, Lee and umbrella. Contrary to example 1, Lee is prenuclear accented, while umbrella is nuclear accented.

The second dimension mentioned in the literature is focus type, namely contrastive versus non-contrastive (Gussenhoven, 1983; Cooper et al., 1985; Ito, Speer & Beckman, 2004; Bishop, 2012). Contrastive or “corrective” focus refers to the instances where the focused element is one of the possible alternatives or a correction to the alternative presented in the question. Examples 3 and 4 show utterances with the word Lee in contrastive focus. Similar to the narrow focus utterance in (1), Lee forms the focal domain.


(4) Did Lulu or Lee borrow my umbrella? – [Lee] borrowed your umbrella. Contrastive

The distinction between contrastive and narrow focus remains a longstanding debatable issue based on the grammatical distinction between the two focus types. In addition, whether contrastive or non-contrastive narrow focus can be differentiated in production and perception has also been under debate, whereby some researchers argue that there is no difference, while others show that some of the acoustic features are different in contrastive versus narrow focus. For a more detailed account, see the discussion in Breen et al. (2010).

The notion of focus according to earlier work by Chomsky and Halle (1968) assumed one pattern of prominence, where prominence is placed on one word of a sentence and is subject to the phonological rule for that particular sentence. The researchers used the terms ‘normal’ and ‘contrastive stress’. ‘Normal stress’ does not carry any particular meaning or function and is a result of phonological rules applied to syntactic structures. In cases when a particular word in an utterance is given more emphasis, that utterance has ‘contrastive stress’ whereby the most prominent word is contrasted. Subsequently, Halliday (1970) described focus as means of expressing a pragmatic function in English, where the tone expresses a speech function and tonal
prominence expresses the structure of information. For Halliday, the placement of
tone and the division of an utterance into tone groups was related to how the
information is divided into units and in particular where the “new information” in an
utterance is located. Thus, tone and information grouping both contribute to the
structure of discourse.

By contrast, the highlighting view (Bolinger, 1958, 1972; Chafe, 1974, 1976)
assumes no rules that would determine the placement of pitch accent or accents in an
utterance. Thus, any word or words can bear a pitch accent and can be focused or
“highlighted” to reflect new, contrastive or any other context specific information.
The choice of highlighted word or words is determined by two factors - the speaker
and the context - and bears no relevance to grammatical structure. These factors
function on a paralinguistic level, presenting a continuum of accent pattern, from
normal to contrastive.

The introduction of Focus-to-Accent (FTA) (Ladd, 1980; Gussenhoven, 1983;
Selkirk, 1984) was an attempt to combine the approaches proposed by earlier schools
and presented a compromise between the normal stress view and the highlighting
view (Ladd, 2008). The projection of focus in this approach distinguishes between
two notions of focus. One is the semantic or pragmatic ‘focus’ and the other is the
phonological or phonetic ‘accent’. This approach is based on Bruce’s work (1977)
which re-introduced Bolinger’s (1958) term pitch accent. FTA was founded on the
notion of focus within generative theories of semantics and syntax, and subsequently
provided a foundation for the work within generative theories of semantic focus
(Williams, 1980; Krifka, 1991; Rooth, 1985, 1992). In FTA, similar to the
highlighting view, context is always taken into account in the use of broad or narrow
focus with the assumption that the location of prominence is always meaningful in
English. Focus manifests itself with the help of nuclear pitch accent, where a word
receives the nuclear accent. ‘Normal stress’ (Chomsky & Hale, 1968) or the
accentuation pattern that does not carry any particular meaning or function is
reinterpreted as broad focus, whereas contrastive focus is reinterpreted as narrow
focus. In addition, the focus domain can span constituents of any size, thus allowing
for neutral patterns that reflect the location of an accent on whole phrases in broad
focus structure.

The present experiment examines the size of focus domains and does not
investigate focus types. Following the FTA approach, the experiment is designed with
the assumption that focus is a constituent marked on the surface structure where the length of this constituent (focal domain) depends on the focus structure or scope. Accordingly, the experiment examines the phonological and phonetic realisation of two focus structures, broad and narrow. It investigates accentuation and phrasing as well as the phonetic means of marking focus. The following section discusses phonetic cues to accentual and focal prominence in English and reviews a number of relevant studies on the cues differentiating these prominence levels.

4.2.2 Phonetic cues to information structure

A number of studies have investigated the phonetic correlates of prominence in English at the lexical and post-lexical level. A number of acoustic parameters are traditionally accepted as the cues to prominence. These include:

1) duration (Fry, 1955, 1958; Lehiste, 1970; Cooper at al., 1985; Beckman, 1986; Crystal & House, 1988, 1990; Breen at al., 2010; Turk, 2012)

2) fundamental frequency (Fry 1955, 1958; Liberman, 1960; Copper at al., 1985; Breen et al., 2010)


4) and vowel formant pattern characteristics (Fry, 1965; Vanderslice & Ladefoged, 1972; Harrington et al., 1997; Harrington et al., 2000).

These acoustic dimensions correspond to the four perceptual dimensions of length, pitch, loudness and vowel quality.

Although the correlates listed above are widely accepted in current literature, it is essential to differentiate the phonological prominence levels under investigation. Beckman and Edwards (1994) defined four levels of prominence which correspond to nuclear accented syllables, (pre)nuclear accented syllables, stressed syllables and unstressed or weak syllables. Nuclear accented syllables can also have an additional emphasis when produced with narrow focus, henceforth referred to as nuclear focal syllables.
Cooper at al. (1985) investigated the effect of focus on $f_0$ height and duration. Their results showed that focus assignment had a more complex effect on $f_0$ height compared to the findings of earlier studies (O'Shaugnessy, 1979), whereby the major effect of focus is the $f_0$ lowering of post-focused syllables rather than an increase in $f_0$ height on the syllable in focus. The position of the syllable in a prosodic phrase also had an effect on the $f_0$ pattern. Sentence-initial and sentence-medial tokens under focus showed more prominent cues as a result of post-focal $f_0$ lowering, while sentence-final tokens where no post-focal drop was possible had less prominent cues to focus.

Earlier studies also did not take into account the structure of the tune containing target words. For example, when Eady and Cooper (1986) examined the $f_0$ cues to focal prominence, they found a difference in the results depending on the utterance types and respective tune structure - for example, declarative statements versus questions. Another issue arising from earlier research is the issue of vowel quality. Often these studies did not control for vowel quality, therefore differences in the duration, intensity and vowel formant characteristics found between the two vowels may have been the result of phonemic vowel differences. Thus, the $f_0$ of phonemically different vowels in stressed versus unstressed syllables may not be due to the different prosodic conditions, but due to the vowels not being phonemically identical.

Later studies (Beckman et al., 1992; Beckman & Edwards, 1994; Fletcher & McVeigh 1993; de Jong 1995, 2004; Sluijter & van Heuven, 1996a; Turk & White 1999; Harrington et al., 2000; Breen et al., 2010) have made an attempt to separate phonological prominence levels as well as control for tune structure and vowel identity. These studies have shown that while lexical stress is cued by non-tonal parameters such as duration, intensity and vowel formant characteristics (Beckman & Edwards, 1994), accented and nuclear accented syllables use $f_0$ as a distinctive cue in addition to the above mentioned parameters for lexical prominence level. In other words, $f_0$ functions on a post-lexical level and is simply phonotactically constrained to associate with the lexically strong syllable. Thus, $f_0$ movement is considered to be a more significant cue to accentual prominence than to lexical prominence in English (Beckman, 1996; Harrington at al., 2000). In an investigation of the acoustic correlates of stress and accent in AmE, Sluijter and van Heuven (1996a) found that duration and spectral balance correlated with lexical prominence, while overall
intensity correlated with accentual rather than lexical prominence. This, together with $f0$, was a reliable cue to focal prominence. Another study by Sluijter and van Heuven (1996b), investigating $f0$ contours and formant values in AmE, reported that $f0$ movements were more pronounced on focused words, whereas minor $f0$ movements accompanied non-focused words.

Duration appears to be one of the most salient prominence cues in English and has the strongest effect on the vowel portion of a syllable in most languages (Fletcher, 2010). This acoustic cue functions on a number of prominence levels. Crystal and House (1988), Beckman and Edwards (1994), and de Jong (2004) all showed that unstressed vowels are shorter than stressed vowels in unaccented syllables. Greenberg et al. (2003) reported that accented syllables were longer than unaccented syllables. Turk and Sawusch (1996) found that in AmE accented vowels were longer than unaccented stressed vowels. Cooper at al. (1985) investigated non-focused and focused words and found a difference in the duration of the accented words depending on the information structure, with significantly greater lengthening in focus. Similarly, Breen at al. (2010) reported a longer duration of words in focus structure. While the domains of lengthening measured across the above studies varied (word, syllable or vowel), the results for well-described Englishes ultimately show longer duration in stressed versus unstressed, accented versus stressed and focal versus accented conditions.

The effects of intrinsic vowel quality have to be taken into consideration when discussing cues to prominence. The target vowel may display differences in the F1 and F2 (vowel height and backness), $f0$ value, duration and intensity. Previous research shows that high vowels tend to have a shorter duration, higher $f0$ values and less intensity than low vowels (Peterson & Lehiste, 1960; Lehiste, 1970; Ladd & Silverman, 1984). Differences in the $f0$ can also be attributed to the intrinsic pitch differences across vowels (Lehiste, 1970; Ladd & Silverman, 1984) and these differences are more distinct for the vowels in nuclear accented and focal positions.

4.2.3 Cross-linguistic marking of focus

As discussed earlier, the prosodic means of signalling intended information structure differ across languages. This variation depends on the languages’ prosodic structures
and segmental phonologies (Koreman et al., 2008). Thus, focus may have a different effect on accent placement and phrasing across typologically diverse languages.

English and most West-Germanic languages follow head-marking patterns in signalling focus. In English, accents are placed on new or salient information in each phrase, while the rest of the phrase is deaccented and post-focal words bear no pitch accents (Ladd, 1980; Gussenhoven, 1983). Such languages as Korean (Jun & Lee, 1998; Jun 2005a; Jun & Kim, 2007), Japanese (Venditti, Jun & Beckman, 1996; Venditti, Maekawa & Beckman, 2008) and Hungarian (Vogel & Kenesel, 1987; Büring, 2012), on the other hand, rely on edge-marking strategies, where the focused element incurs changes in prosodic phrasing. Here, a narrow focus domain forms its own phonological phrase, thus resulting in dephrasing of post-focal material. Finally, some languages rely on a combination of focus marking strategies. For example, Komotoni Romani includes accentuation of a particular morpheme and deaccenting similar to English, as well as the stress-shift to an earlier syllable (Adamou & Arvaniti, 2010).

Among the languages spoken in India, Kolkata Bengali (Hayes & Lahiri, 1991), Hindi (Harnsberger, 1996; Patil et al., 2008) and Tamil (Keane, 2014) employ head and edge marking (Hayes & Lahiri, 1991) to signal focus (see §2.4.3 for more detail). Recent research on Bangladeshi Bengali shows the use of L*+H (Khan, 2008, 2014). This is different from many well-established varieties of English where a narrow focused word does not necessarily display a particular pitch accent; rather, it could be L+H* or H* (Watson, Tannenhaus & Gunlogson, 2008) or even L*+H for greater emphasis (Pierrehumbert & Hirschberg, 1990; Arvaniti & Garding, 2007). Moreover, a range of pitch accents can be used on the accented syllables in narrow and broad focus structures. As discussed in §2.3.4, another feature that sets well-established Englishes apart from a number of Indian languages is the absence of post-focal deaccenting in the latter (excluding Bangladeshi Bengali).

Languages may also exhibit differences in the phonetic parameters associated with focal and accentual prominence. For example, in Catalan (Astruc & Prieto, 2006), syllable duration, spectral balance and vowel quality are reliable acoustic correlates of lexical prominence. Accentual prominence is cued by pitch and overall intensity, which are the two parameters commonly used in English. Unlike English, however, vowel quality and vowel duration are only optional cues to accent. Barry et
al. (2007) found that both French and German use $f_0$ and duration as reliable cues to accentual prominence, but German speakers tend to use greater lengthening of the accented syllables compared to French speakers, most likely as a result of the absence of vowel length distinction in French and its presence in German.

4.3 Hypotheses

Experiment 1 sets out to answer the following research questions:

1. What are the accentuation and prosodic grouping patterns in narrow and broad focus structure in BE and KE?
2. What are the phonetic correlates of accentual and focal prominence levels in BE and KE?
3. What are potential phonological and phonetic differences between the two IndE varieties?

In view of the literature presented in the sections above as well as in §2.1, §2.3 and §2.4 of Chapter 2, a number of hypotheses have been formulated to address these research questions. The hypotheses are grouped into three sections in relation to each question.

**Accentuation and focus**

**Hypothesis 1**

According to previous research (Bansal, 1969, 1990; Latha, 1978; Gargesh, 2004; Wiltshire & Harnsberger, 2006), IndE speakers place accentual prominence on most content words in broad focus structure. Moreover, this is the feature that is applicable to IndE phonology in general. Hypothesis 1 suggests that speakers of BE and KE place accentual prominence on all or most content words and possibly also function words regardless of the speakers’ L1.

**Hypothesis 2**

This hypothesis proposes that, similarly to the speakers of well-described varieties of English who manipulate accentuation and phrasing patterns depending on focus structure (see Gussenhoven, 1983, 1999; Ladd, 1980, 1996; Xu & Xu, 2005), BE and KE speakers distinguish between broad and narrow focus. Hypothesis 2 suggests that the speakers produce the intonational realisation of information structure
similar to well-established Englishes and employ deaccenting of post-focal and possibly pre-focal words to signal narrow focus structure, without any changes in phrasing patterns.

**Hypothesis 3**

A number of languages spoken in the subcontinent do not rely on deaccenting as a means for the prosodic marking of focus. The patterns of accentuation are not necessarily changed under narrow focus. Instead, the content words are accented and the high tones in the vicinity of accented words are in a strict downstepped relationship (see Keane, 2007, 2014 on Tamil; Harnsberger, 1996 on Hindi; Patil et al., 2008 also on Hindi). It is only research on Bangladeshi Bengali that found evidence of post-focal deaccenting (Khan, 2008, 2014), whereas earlier research on Kolkata Bengali reported no effect of focus on accent placement (Hayes & Lahiri, 1991). Alternative to Hypothesis 2, Hypothesis 3 proposes that differences in information structure will not bring differences in accentuation and the speakers will rely on pitch range compression without deaccenting of post-focal material in narrow focus: in other words, the words following the focused one will be accented and produced with a downstepped $f0$ contour.

**Hypothesis 4**

It may be hypothesised that KE and BE speakers follow a “hybrid” system as a result of IndE development and the use of the prosodic marking of information structure being more systematic for IndE as a variety, influenced by the languages spoken in India. Hypothesis 4 suggests that the speakers have several ways of narrow focus marking and are not limited to post-focal deaccenting. Furthermore, the use of deaccenting proportional to other strategies may be reflective of the degree of L1 influence (Ueyama & Jun, 1998; Nguyễn et al., 2008; Swertz & Zerbian, 2010).

**Phonetic correlates to accentual and focal prominence**

**Hypothesis 5**

Four phonetic correlates to signal accentual and focal prominence have been found in well-described Englishes: duration, $f0$, intensity and vowel quality (Cooper et al., 1985; Beckman, 1986; Breen et al., 2010; Harrington et al., 2000; Harrington et al., 1997). It is hypothesised that the speakers of BE and KE will be able to
distinguish focally accented, accented and stressed unaccented prominence on the target syllables. Therefore, the four phonetic parameters will show an increase in values in focal versus accented syllables and accented versus stressed syllables.

Hypothesis 5A assumes that the duration of the segmental domain will increase with each level of prominence. Similarly, Hypothesis 5B assumes that the RMS amplitude will be higher in focal versus accented syllables and accented versus stressed unaccented syllables. According to Hypothesis 5C, the f0 height will increase in focal syllables compared to accented syllables and in accented syllables compared to stressed syllables.

As discussed in §2.4.3, Genzel and Kügler (2010) and Moon (2002) found that the increase in pitch was attributed not only to the higher peak but also to the lowering of the low turning point on the focal syllables in Hindi and Hindi English. Therefore, pitch excursion, measured from the low to the high point of the rise associated with the target syllable, could be a more reliable measure than an absolute f0 height for IndE speakers. Hypothesis 5D is applied to two levels of prominence. This hypothesis predicts greater pitch excursion in focal syllables compared to accented syllables.

Finally, Hypothesis 5E suggests that vowels in accented syllables are more peripheral than vowels in stressed unaccented syllables, and vowels in focal syllables are more peripheral compared to accented vowels (Sluijter & van Heuven, 1996b). In this case, the first formant (F1) will be realised lower for high vowels and higher for low vowels with a higher prominence level. The second formant (F2) will exhibit greater values for high vowels and lower values for low vowels in focal versus accented and accented versus stressed syllables.

Hypothesis 6

Given that IndE has developed into a distinct variety, Hypothesis 6 proposes that BE and KE speakers may exhibit a preference for some phonetic correlates over others. According to Hypothesis 6, the speakers in this study may not rely on all four phonetic parameters commonly found in well-described varieties of English. For example, the parameters of f0 or RMS amplitude could be more reliable cues in distinguishing focal and accentual prominence as opposed to duration (Moon, 2002 for IndE).
Potential differences between KE and BE

Hypothesis 7

Taking into account that prominence relations and focus prosody are not universal and that typologically distinct languages use different means for realising these (Frota, 2002; Koreman et al., 2008), there may be phonological and/or phonetic differences between KE and BE. Previous research on English as an L2 has shown evidence of L1 influence on the L2 intonation patterns (McGory, 1997; Ueyama & Jun, 1998; Nguyễn et al., 2008). It is hypothesised that speakers may employ different means to mark accentual and focal prominence, depending on their L1 background or the variety of IndE spoken in their region. BE speakers may show a greater amount of deaccenting than KE speakers, given that Bengali is the only language that was shown to have post-focal deaccenting (Khan, 2008, 2014). In addition, there may be differences in the manipulation of \( f_0 \), intensity, duration and F1/F2 vowel formants between the groups, similar to the findings of Moon (2002) on the acoustic parameters to mark focus in TelE and HE.

4.4 Overview of the method

Section 3.4 of Chapter 3 gives a detailed description of the method used in Experiment 1. It presents the set of materials with target words and syllables, and outlines the procedures for recording, annotation and analysis. It also includes the parameters under investigation and the acoustic measurements taken. As a reminder to the reader, the most important points are revised in this section with a focus on materials and analyses.

The set of materials is based on simple declaratives designed to elicit broad and narrow focus in three prosodic positions: subject (sentence-initial), verb (sentence-medial) and object (sentence-final). Each target sentence was produced in response to a question that would prompt broad or narrow focus answer. The sentence length and structure were kept constant (three content words), while the duration of the target words, together with their syllabic structure and lexical stress, were varied. The target sentence with all possible word combinations is illustrated below. The target words that were rotated are underlined.

Lee/Nina/Marina may move/borrow/minimise my mill/lily/umbrella.
A total of 720 utterances were examined for accent placement, prosodic grouping (based on the presence of pauses) and global $f_0$ movements in broad and narrow focus (see §5.1 and §5.2 for more detail). In addition to the accented words, the low and high turning points were annotated at the right edges of full intonational contours, marking prosodic boundaries. In order to determine the effect of prominence level on the parameters of $f_0$, duration, RMS amplitude and vowel quality, the following measurements were taken for each target syllable (see Table 3.3 in §3.4.1 for a full description):

- Absolute $f_0$ height (Hz) extracted at the highest peak in the vicinity of the target stressed syllable
- $f_0$ excursion (Hz) associated with the rise in the vicinity of the target stressed syllable
- Mean RMS amplitude (dB) extracted at the vowel midpoint of the stressed syllable
- Duration (ms) of the target stressed syllable and vowel
- F1 and F2 values (in Hz) measured at the steady state of the vowel in the target syllable, often corresponding to its mid-point.

The Hz values underwent conversion into the semitones scale with the 100 Hz value taken as a baseline (Karlsson et al., 2010). In the vowel analysis, Euclidean distances were calculated to investigate the level of vowel dispersion in accented, focal contexts versus unaccented for individual speakers where vowels showed overlap in their realisations (Harrington, 2010).

The values for each measure were combined in a single data frame using the R statistical package, with a total number of 1759 tokens. All statistical analyses presented in this chapter were performed using LMM procedures, with post hoc Tukey tests or paired t-tests where relevant. Main effects and interactions of LMM analysis are reported in the results section. The full set of statistics detailing the random and fixed effects for each phonetic parameter is presented in Appendix B.

Two types of box plots are included in the description of the results. These are: a) trellis plots, used to show different groups of data points as well as the effects and interactions of multiple factors, and b) regular box plots, used to show graphical summaries of distribution for one or more factors. In both types of plots, the
horizontal line or dot in the box presents the median, while whiskers indicate the end points of the distribution.

4.5 Results

4.5.1 Prominence and phrasing in broad focus

This section presents qualitative analyses of accentuation and prosodic grouping in broad focus declaratives sentences, testing Hypotheses 1 and 7. Prior to describing accentual prominence it is relevant to mention the patterns of lexical prominence across the speakers. Speaker B4 produced the word *Marina* with the stress on the first syllable, while the rest of the speakers placed prominence on the second syllable, which is a more typical pattern for AusE or AmE. The word *minimise* was produced with the stress on the first syllable by speakers B1 and B2. The other two BE speakers and all KE speakers placed lexical prominence on the final syllable.

4.5.1.1 Accentuation

The results showed that in broad focus structure, prominence was usually placed on the sentence-initial noun, a proper name. This pattern was relatively consistent across all speakers. The distribution of accents in the rest of the target sentences, however, varied between the two L1 groups.

In BE, the most typical pattern was to place accentual prominence on the first and the last content word of a declarative (subject and object). This pattern was observed in all of the utterances produced by speakers B1 and B3, and in around 80% of the utterances produced by speakers B2 and B4. The figures below illustrate this pattern and show the pitch tracks with annotations for speakers B1 (Figure 4.1), B2 (Figure 4.2) and B3 (Figure 4.3). The figures show that the \( f0 \) contour on accented words displays a number of pitch realisations, with a rising contour being the most common. Low pitch associated with the accented syllable was mostly used in nuclear positions, typically with a rising \( f0 \) contour at the boundary, as shown in Figure 4.2. Its use in sentence-initial positions is limited to a small number of repetitions.
Figure 4.1 Illustration of accentuation on subject and object in broad focus structure, with a rising pitch on the nuclear accented word. The $f_0$ track illustrates suspension of declination. Speaker B1.

Figure 4.2 Illustration of accentuation on subject and object in broad focus structure, with a rising pitch on the prenuclear accented word. Speaker B2.

Figure 4.3 Illustration of accentuation on subject and object in broad focus structure, with a rising pitch on prenuclear and nuclear accented words. The $f_0$ track illustrates suspension of declination on the nuclear accented word. Speaker B3.
The most striking feature of the declaratives produced by the BE group is that the declination pattern is not always present, and the final accented words are not necessarily produced with a downstep in pitch but display a suspension of declination (see Figures 4.1 and 4.3). Declination was more common for the utterances where all content words were accented and was more characteristic for speakers B2 and B4. Figure 4.4 shows an example of such an utterance with the accentuation on subject, verb and object.

Unlike the BE group, the utterances produced by KE speakers displayed differences in accent placement with more variation. The most typical pattern of accent distribution is shown in Figures 4.5 and 4.6, where the three target words are accented. Speakers K1 and K2 were the most consistent. It is worthwhile to note that speaker K2 always placed prominence on the pronoun my (Figure 4.5), including both broad and narrow focus contexts.

Another pattern found in a small proportion of this group was accentuation on the noun and the verb, deaccenting the object (as shown in Figure 4.7). Only speaker K4 deaccented the verb in a number of repetitions and placed accents on the subject and the object (Figure 4.8). This was the main pattern reported for the BE group. Overall, the proportion of the utterances produced by all KE speakers displaying less than three words accented was relatively small (under 20% across four speakers).
Figure 4.5 Illustration of accentuation on subject, verb and object in broad focus structure. Speaker K2.

Figure 4.6 Illustration of accentuation on subject, verb and object in broad focus structure. The f0 track illustrates declination throughout the utterance. Speaker K1.

Figure 4.7 Illustration of accentuation on subject and verb in broad focus structure, with a sustained high pitch or a ‘hat pattern’ on prenuclear accented words, followed by a sharp drop in f0 after the nuclear accent. Speaker K3.
Figure 4.8 Illustration of accentuation on subject and object in broad focus structure, with more or less sustained high pitch. Speaker K4.

In terms of the global $f0$ contours, the target sentences were often produced with a declination pattern in KE where the peak showed lowering throughout the utterance (as illustrated in Figure 4.6). A smaller number of repetitions showed $f0$ contours with sustained high pitch, corresponding to a ‘hat pattern’ or a high plateau in English. Stretches of sustained high pitch were more frequent for speakers K3 and K4 (Figures 4.7 and 4.8). Similar to the most frequent pitch shape associated with the accented words in BE, a rising $f0$ was also used by KE speakers on accented words, but the rises were often much shallower when compared to the rises produced in BE (see Figure 4.6). In the absence of a rising pitch movement on accented words, KE speakers used high pitch.

As a reflection of accent placement differences between the two groups, the number of words accented in declaratives with broad focus structure was greater for L1 Kannada speakers. Figure 4.9 below shows the total number of accented versus stressed syllables as a percentage of total utterances produced by each group. The blue colour in both bars corresponds to accented syllables and the grey colour corresponds to stressed syllables. As shown in the figure, the ratio of accented to stressed syllables corresponds to 47:53 for the BE group, while for the KE group, the ratio of accented to stressed is much higher at 64:36. This finding partially supports Hypothesis 1 that all IndE speakers place accents on a large number of words in an utterance. Furthermore, the prevalence of the pattern where three content words are accented among KE speakers, resulting in greater accentual density in KE compared to BE, supports Hypothesis 7 that predicted the differences between the two groups, possibly as a result of L1 influence.
Figure 4.9 Bar plot for the proportional distribution of accented (blue) versus stressed (grey) syllables for each group.

In terms of prosodic grouping, there were no differences between BE and KE. For both groups, most of the utterances were produced as one phrase. When an utterance was produced as two intonational phrases and an extra pause was inserted, the pause was more likely to occur after the noun, as illustrated in Figure 4.4. This was more common for speakers B1, B2 and K2.

4.5.1.2 Summary
To summarise, the accentuation patterns in broad focus simple declaratives partially support Hypothesis 1, in that placing accents on all content words may not be a feature shared by all IndE speakers. Instead, the results support Hypothesis 7 that suggests differences in accentuation patterns based on L1s. This is shown first through the fact that greater accentual density is a feature more applicable to KE but not necessarily to BE, thus indicating that Hypothesis 1 is true in relation to the KE variety. Second, it is shown that BE speakers have a strong preference for accenting the first and the last content word, which is a common pattern for broad focus in English (Ladd, 2008), with a relatively small number of repetitions where more words are accented. KE speakers, by contrast, are more likely to place accentual prominence on most content words or, as a second preferred mechanism, to place an accent on the first and second content word. There is also greater variability across the speakers for this L1 group.
More homogeneity in phrasing patterns was observed across the speakers. In most cases, simple declaratives were produced as one phrase, with a small number produced as two or more phrases, usually in the utterances with trisyllabic words. It is important to note that the present data set may have limitations, as all utterances were short and consisted of three content words. Therefore, accentuation and phrasing in broad focus will also be examined in a larger sample including complex declaratives and various types of questions (see Chapter 7) to provide further evidence for the findings of this experiment.

4.5.2 Prosodic marking of narrow focus

As discussed in §4.2.1 and §4.2.3 of this chapter, focus structure is usually reflected in the use of accentual prominence in most varieties of English. The next few sections investigate the prosodic realisation of narrow focus and examine the strategies for marking focus in BE and KE, testing Hypotheses 2, 3, 4 and 7. The qualitative analysis begins with deaccenting, the most salient prosodic strategy for focus marking in well-established Englishes.

4.5.2.1 Deaccenting

Figure 4.10 illustrates a narrow focus utterance produced by speaker B3. As shown in the figure, a single accent is placed on the word in focus (the subject) and the following words do not have accents. The focal word has a characteristic rising pitch followed by a gradual lowering throughout the utterance.

Despite the presence of deaccenting, this strategy was not used all the time across speakers and repetitions. K3 was the only speaker who used deaccenting in around 90% of the repetitions, thus producing the most English-like pattern. Speakers B1, B3, K3, and K4 used it more frequently compared to the rest. In sentence-medial and sentence-final contexts there were instances with pre-focal deaccenting; however, these were limited to utterances produced by speakers K3, K4, and a small proportion of the utterances produced by B3. An example of pre-focal deaccenting is shown in Figure 4.11 below. The verb *move* is in narrow focus and is the only word accented. The f0 shows a distinctive rise on the monosyllabic focal word.
Figure 4.10 Illustration of accentuation on the narrow focused subject followed by post-focal deaccenting. Speaker B3. W indicates a weak post-tonic syllable.

Figure 4.11 Illustration of accentuation on the narrow focused verb accompanied by pre- and post-focal deaccenting. Speaker K3.

Overall, however, deaccenting was used alongside a number of other strategies. Therefore, the results do not support Hypothesis 2 that predicted post-focal deaccenting as the main strategy to signal narrow focus structure, similar to well-established Englishes.

The results also showed that a more common accentuation pattern in narrow focus for the BE group was to accent both the subject and the object, similar to broad focus structure. In this instance, narrow focus was signalled by the increase in the height of the $f0$ peak on the focussed word. Figure 4.12 illustrates an utterance with a focus on mill. The $f0$ peak on the accented syllable in focus is higher than that of the non-focally accented peak in Lee.
Figure 4.12 Illustration of accentuation on narrow focused object. The $f_0$ peak on the nuclear focal *mill* is raised relative to the $f_0$ associated with the prenuclear accent. Speaker B1.

4.5.2.2 Pitch range compression without deaccenting

In the instances when post-focal material was not deaccented and the utterance was produced as one phrase, the words after the narrow focused syllable were realised in compressed pitch. No differences were observed based on L1. Speakers K1, K2, K4, B2 and B4 used this strategy in around 20-37% of utterances, while speakers B1, B3 and especially K3 used it less frequently, in under 20% of the total narrow focus utterances and mostly in sentence-medial focus structures.

Figure 4.13 illustrates this pattern in sentence-initial focus, where the verb *move* is also accented but the $f_0$ of the peak is lower relative to the preceding accent. In Figure 4.14, narrow focus is placed on the verb *borrow*. Similarly, there is no post-focal deaccenting and the pitch is compressed on the accented syllable of the object *lily*.

Figure 4.13 Illustration of accentuation on narrow focused subject followed by another accent realised in compressed pitch. Speaker K2.
This type of strategy is thought traditionally to be uncommon for narrow focus marking in English, where a narrow focused word is assumed to have a nuclear accent. This finding supports Hypothesis 3 which states that differences in information structure do not always bring differences in accentuation. Rather, these differences are more likely related to the pattern found in a number of languages spoken on the subcontinent, where focus has little effect on accent placement (Patil et al., 2008 on Hindi; Keane, 2014 on Tamil) and the accented words after the focused word are in a downstepped relationship produced in a compressed pitch range.

4.5.2.3 Re-phrasing

Another strategy for making the focal word more prominent is inserting a phrase break after the focal constituent, which is usually the focused word. This pattern was observed for two KE speakers (K1 and K2) and all BE speakers, with a small proportion of repetitions for B3. This pattern is illustrated in Figure 4.15 showing the pitch track of the utterance Nina may move my mill with a focus on Nina produced by speaker B2. The narrow focused noun Nina is produced as a separate intonational phrase (IP) and is realised as a full intonational contour (rise-fall-rise) followed by a pause of 130 ms.
In addition to a phrase break after the focused constituent, an optional major phrase break was at times inserted preceding it. In sentence-medial and -final contexts, a focused word (or focused word with an adjacent pronoun or modal verb) was also placed in a phrase with a full intonational contour. Figure 4.16 below is an example of prosodic phrasing with a sentence-medial focus structure produced by speaker B4. The pitch track of the utterance *Lee may borrow my lily* clearly shows the focused word *borrow* together with *my* forming one phrase with a full intonation contour, separated by pauses. The pre-focal pause measures 330 ms, while the post-focal pause duration is 160 ms.

In the instances of re-phrasing, the $f0$ height on the narrow focused word was often realised in a lower pitch range compared to the $f0$ on other accented words of that utterance (also shown in Figure 4.16). Thus, the speakers did not always produce higher peaks when the focused word was produced in a separate phrase.
The duration of the post-focal pauses after a focused noun or a focused verb varied across the speakers, ranging from an insignificant 80 ms to pauses with duration of around 350 ms. These durational values are illustrated in Figure 4.17 below. As shown, speakers B1, B4 and K1 inserted pauses with the longest duration. Moreover, their speech showed a greater amount of variation. The duration of the pauses was shortest for speakers B2 and B3. For speaker K3, there was only one account of pausing after a focused word, with a relative short duration of 57 ms. Despite this variability in pause duration, the $f0$ contours indicate that pauses are not related to disfluency but are an additional strategy.

Figure 4.17 Box and whiskers plot for pause duration in milliseconds grouped by speaker. Blue boxes correspond to BE speakers. Green boxes correspond to KE speakers. The thick horizontal line shows the median, while the extent of the whiskers indicates the range of distribution.

4.5.2.4 Pitch shape

The results showed that for all speakers the preferred pitch shape on the accented syllable in narrow focus was a rise (as shown in Figures 4.11-4.14). This is in contrast to the variability of pitch realisations on the accented syllables in broad focus or non-focally accented syllables in narrow focus contexts. The $f0$ shape on monosyllabic words, especially sentence-final position, did not show clear differences between the two groups. However, a closer examination of the $f0$ shape on focally accented bi- and trisyllabic words showed that BE speakers may have produced later peaks in focal rises compared to KE speakers. In BE, the peak of the rising gesture was often
realised in the post-tonic syllable on narrow focused words (borrow in Figure 4.14),
while in KE, the peak was somewhere in the accented syllable. This could potentially
indicate differences in the phonological categories or phonetic realisations of the rise
between BE and KE.

4.5.2.5 Summary
The results of the qualitative analysis of the prosodic marking of focus in IndE show
that speakers use a number of strategies and may have a hybrid system. The three
main strategies used by the speakers in this study include deaccenting, re-phrasing
and pitch range compression without deaccenting. This system bears some similarity
with other Englishes and other languages (such as with the use of deaccenting) but it
displays a number of features that may be characteristic to IndE intonational
phonology, thus supporting Hypothesis 4. The fact that speakers do not use
deaccenting as the only strategy disproves Hypothesis 2.

Unlike the patterns of broad focus, the differences in the use of strategies for
narrow focus marking between the groups are more discrete. This reflects individual
speakers’ differences rather than their belonging to a particular L1 group. Therefore,
Hypothesis 7 that suggested differences between the two L1 groups is not borne out in
the analysis of broad focus structure. The emerging picture indicates that regardless of
L1, speakers in this study prefer some focus strategies to others, with six speakers
relying on two strategies and eight speakers relying on three strategies.

Figure 4.18 below displays the proportional distribution of the three main
strategies used by the speakers as a percentage of the total utterances analysed. A
number of features are of note in the figure. First, deaccenting of post-focal material
(green) was used by all speakers, but was more characteristics for speakers K3, K4,
B1 and B3, with K3 being the most consistent. Second, six speakers employed re-phrasing (grey), in the form of the insertion of a major phrase boundary after and/or
before the focused word, at times in combination with the adjacent word. This type of
strategy seemed to be used less frequently by speakers B1 and B3, and was not
employed by K2 and K3. In addition, re-phrasing often compensated for the increase
in $f0$ height on the focused word. The pitch on the focal word may have been lowered
relative to other accented words in that utterance, but the focus was clearly marked by
separating the focused constituent into a phrase.
Figure 4.18 Bar plot for the proportional distribution of the main focus marking strategies presented for each speaker. Green corresponds to deaccenting, grey correspond to re-phrasing and blue corresponds to pitch range compression without deaccenting.

Third, compression of pitch range (blue) on post-focally accented words was another strategy used by all speakers. Its frequency was lower compared to deaccenting and was more common for speakers K1, K2 and B2. Nevertheless, the results support Hypothesis 3 and indicate that there is some influence here of the languages spoken in the subcontinent, ultimately indicating that IndE is a variety distinctly different from other Englishes. Moreover, pitch compression without deaccenting raises the question of defining the narrow focused accent as nuclear. This is an assumption made in most treatments of intonation in English.

There is also a possibility that both L1 groups use a rising f0 in narrow focus. The phonological nature of this rise – that is its precise alignment in both focal and non-focal contexts – will be investigated in the tonal alignment experiment (Chapter 5) to determine the phonological categories in BE and KE. Later alignment of the rising gesture associated with the focally accented syllable could indicate that BE speakers employ a delayed rising pitch accent.
4.5.3 Phonetic correlates of information structure

The following four sections present the findings of the phonetic correlates of prominence, namely $f0$, RMS amplitude, duration and vowel quality in the syllables across three conditions of stressed, accented and focally accented (focal). This will test Hypotheses 5 (A, B, C, D, E), 6 and 7. As noted in the overview of the materials (§4.4), the location of primary stress in the target words *minimise* and *Marina* varied across the speakers. In the figures presented below, the tokens of the words *Marina* include both first and second syllables, and the tokens for *minimise* include the first and the last syllable. In the section on vowel quality, the syllables are grouped according to the target vowels.

4.5.3.1 Fundamental frequency

Figure 4.19 shows the results of the $f0$ analysis for stressed (S), accented (A) and focal (F) syllables for the two groups. The plot presents the distribution of the $f0$ values (in semitones) measured at the peak for each accent condition by prosodic position (subject, verb and object). The dot inside a box corresponds to the median $f0$, and the outliers are shown as dots above and below the box whiskers. The top panel shows the values for KE speakers and the lower panel corresponds to the BE group.

As proposed in Hypothesis 5C, there was a significant effect of accent CONDITION (A, F or S) on the $f0$ values. Post hoc Tukey tests revealed differences across all three types of syllables, focal versus accented [$z=20.56$, $p<.001$], accented versus stressed [$z=23.66$, $p<.001$] and focal versus stressed [$z=41.52$, $p<.001$]. Greater increase in $f0$ height was observed between stressed and focal syllables, with a more modest increase in stressed versus accented syllables, especially in medial (V) and final focus (O). The figure also shows a large overlap in the $f0$ means between accented and focal syllables, suggesting that big differences exist between accented/focal and stressed syllables. For example, in sentence-medial tokens (V) produced by BE speakers, stressed syllables had a mean of 4 ST [SD=2.8 ST], accented syllables - a mean of 7.3 ST [SD=2.8 ST] and focal syllables - a mean of 12.6 ST [3.6 ST]. In the same sentence-medial tokens produced by the KE group the mean $f0$ height measured at 2.4 ST [SD=3.11 ST] in stressed syllables, 7.6 ST [SD=4.5 ST] in accented, and 8.8 ST [SD=2.8 ST] in focal syllables.

There was also an effect of sentence POSITION (S, V and O). Significant differences were reported for the $f0$ height in subject versus object [$z=6.08$, $p<.01$],
indicating higher scaling for the subject. In addition, there was an effect for VOWEL
TYPE, reflecting intrinsic vowel values and their interaction with \( f0 \). The effect was
significant for the back rounded vowel in *borrow* and three of the front vowels in the
words *umbrella, Lee,* and *Nina/mill/lily* \([/ɔ/-/ɛ]/: z=-2.88, p<.04; /ɔ/-/ɪ/: z=-3.76,
p<.003; /ɒ/-/u/: z=-2.868, p<.05\]. High vowels displayed higher \( f0 \) values.

Figure 4.19 Box and whiskers plot for the distribution of \( f0 \) height in semitones by
prosodic position (S - subject, V - verb, O - object) presented by group (KE - top
panel, BE - bottom panel).

Figure 4.19 also shows that despite a degree of inter-speaker variability and
overlap, GROUP had no effect on the \( f0 \) height \([t=-0.833, p>.05]\). There was,
however, an interaction between CONDITION and GROUP \([t=-7.067, p<.0001]\),
suggestive of a smaller \( f0 \) increase for the KE group. Further investigation of the \( f0 \)
height across accent conditions by speaker showed that two KE speakers (K2 and K3)
did not expand their pitch range in narrow focus (F). This is shown in Figure 4.20
below, which presents the mean \( f0 \) values (in semitones) across the three accent
conditions for each speaker and averaged across subject, verb and object.
Figure 4.20 Plots showing mean $f_0$ values (semitones) in three accent conditions (stressed, accented and focal), presented by speaker and averaged across the prosodic positions. Values for BE speakers are shown to the left in dark blue, values for KE speakers are shown to the right in dark green.

As shown, all BE speakers made distinctions across the three accent categories (stressed, accented and focal) by employing an increase in $f_0$, with somewhat greater differences between focal and stressed and between accented and stressed. In the plot to the right, however, the mean $f_0$ value of the target syllable is similar in accented (A) and focal (F) conditions for speaker K3, presented second from the top. Despite what appears to be a difference in mean $f_0$ of accented versus focal syllables for speaker K2 (top line of the plot to the right), a pairwise comparison of means in focal versus accented syllables across all three prosodic positions reported that the differences were not significant [p>.05]. The values for mean and standard deviation of $f_0$ height for each speaker presented by prosodic position (S, V and O) can be viewed in Appendix B.

To test whether the $f_0$ magnitude on focused words is associated with lowering of the low point as well as raising of the peak as posited in Hypothesis 5D, an additional $f_0$ analysis was performed. It investigated pitch excursion by measuring the difference in the $f_0$ between the low turning point and the peak in the vicinity of the accented (A) and focal (F) syllables. Stressed syllables were excluded from the analysis. Figure 4.21 shows the distribution for the $f_0$ excursion in focal versus accented syllables, presented by prosodic position for each group. As shown, pitch excursion was affected by accent CONDITION [$t=22.79$, $p<.0001$] with no effect of
GROUP [p>.05], suggesting that all speakers used a greater magnitude of excursion on narrow focused syllables.

In addition, an interaction between CONDITION and GROUP [t=-10.3, p<.0001] indicated that the rises produced by BE speakers were higher, as shown in Figure 4.21. The whiskers on the lower panel show considerable variation among BE speakers as well as a level of overlap between the boxplots in accented versus stressed conditions, reflecting individual use of tonal space and inter-speaker differences. Apart from a few outliers, KE speakers consistently produced shallower rises. The result also showed that, contrary to the analysis of the f0 height, the rises produced by the BE group were greater on both subjects and objects compared with verbs, which had significantly shallower rises [V-O: t=-5.503, p<.0001; V-S: t=-5.816, p<.0001].

**4.5.3.2 RMS Amplitude**

Figure 4.22 shows the results of the RMS amplitude analysis in decibels (dB) measured at the vowel midpoint of the target syllable for BE and KE across the three accent conditions. The plot shows the distribution of the data, where the dot indicates the median value for each accent condition, presented by prosodic position, and the
dots outside the whiskers correspond to outliers.

Figure 4.22 Box and whiskers plot for the distribution of dB-RMS values by prosodic position (S - subject, V - verb, O - object) presented by group (KE - top panel, BE - bottom panel).

Overall, the results showed a significant effect of CONDITION on amplitude [p<0.0001], thus indicating Hypothesis 5B is true. Tukey tests confirmed differences for each accent CONDITION and showed that these differences in RMS amplitude values were greater in focal versus stressed and accented versus stressed syllables compared to focal versus accented. [F-A: z=10.55, p<.0001; S-A: z=-20.08, p<.0001; S-F: z=-28.75 p<.0001]. No effect of GROUP was reported; however, an interaction between CONDITION and GROUP was significant, possibly as a result of a large degree of variation across the KE group [t=-4.05, p<.002] in all three conditions, especially in stressed syllables.

In terms of the prosodic POSITION, differences were significant between subject and object [z=6.649, p<.001], and between verb and object [z=3.664, p<.001], with no interaction between POSITION and GROUP. As a result of amplitude tracking the f0 contour, the results indicate that for all speakers the RMS amplitude values decreased towards the end of the utterance. There was also a significant effect
of vowel type, with the back vowel /u/ having lower intensity compared to a number of vowels [/u/-/e/: $z = -4.388$, $p < .001$; /u/-/i/: $z = -3.985$, $p < .001$; /u/-/o/: $z = -9.039$, $p < .001$], most likely as a result of the intrinsic RMS amplitude for vowels (usually lower for high back vowels).

Despite a significant effect of accent CONDITION, a clear difference in median amplitude between focal and accented syllables for the KE group is not obvious in Figure 4.22 above. As an additional analysis, RMS-amplitude was examined for each speaker across three conditions. These results are presented in Figure 4.23 which shows BE speakers in the plot on the left hand side and KE speakers to the right. Mean dB-RMS amplitude values were averaged across prosodic positions (subject, verb and object). As shown in the figure, speakers K2 and K3 produced focal syllables with the same RMS amplitude as accented syllables (for example, K2: focal syllables - mean=71 dB [SD=2.5 dB], accented syllables - mean=71 dB [SD=2.4 dB]). Welch t-tests confirmed that these KE speakers did not distinguish between focal and accented syllables on the basis of this parameter [$p > 0.05$]. The figure also shows similarity in intensity for BE speakers and greater interspeaker variation in the KE group, with overall lower intensity values for speaker K4, which must have also contributed to the degree of overlap in Figure 4.22 for this group. The results partially support Hypothesis 5B. Amplitude was used to mark accentual prominence level for all speakers, but was a reliable cue to focal prominence for six speakers.

Figure 4.23 Plots showing mean dB-RMS values in three accent conditions (stressed, accented and focal) averaged across the prosodic positions, presented for each group with BE speakers to the left and KE speakers to the right.
4.5.3.3 Duration

In order to investigate the role of duration as an acoustic cue to information structure and test Hypotheses 5A, the duration of the target syllable was examined with GROUP, accent CONDITION and POSITION as fixed factors, consistent with the analysis of f0 and RMS-amplitude. In addition, another fixed factor, WORD, was included to reflect monosyllabic (1S), bisyllabic (2S) and trisyllabic (3S) forms, and to investigate lengthening patterns in words of different length. Figure 4.24 shows the distribution, outliers and median duration value in milliseconds of focal, accented and stressed syllables. The results are presented by group (top and bottom panels) and by word (from left to right).

![Box and whiskers plot for the distribution of syllable duration in ms by word structure (1S - monosyllabic, 2S - bisyllabic, 3S - trisyllabic) across three conditions (S - stressed, A - accented, F - focal), presented by group (KE - top panel, BE - bottom panel).](image)

The presence of narrow focus as well as accent had an effect on the duration of the target syllables, although the durational difference between accented and stressed was smaller than between stressed versus focal and accented versus focal syllables [F-A: z=12.73, p<.0001; S-A: z=-5.04, p<.0001; S-F: z=-15.98 p<.0001]. As shown in Figure 4.24, there was no effect of GROUP [t=0.27, p>.05] and no interaction between GROUP and CONDITION [t=-0.243, p>.05]; however, the
results showed a large degree of variation, reflecting speaker-individual lengthening differences. There is also an overlap in the boxes for accented and focal syllables, which indicates that despite differences in means, the speakers showed much more variation in these two conditions (A and F). For example, monosyllabic words produced by speaker B1 in focal syllables showed a mean duration of 274 ms [SD=47 ms], while the same tokens produced by speaker B2 had a mean value of 330 ms [SD=58 ms]. Similarly, accented syllables had a mean value of 189 ms [SD=30] for B1 with a mean 290 ms [SD=56] for speaker B2. For KE speakers, the most amount of overlap is shown in the duration of syllables in trisyllabic words (3S) between focal and stressed conditions. This overlap corresponds to the repetitions produced by speakers K1 and K4 who showed variability in the duration of the stressed syllable in the word *umbrella*.

All syllables were significantly longer in monosyllabic words, as evident in the significant durational difference between monosyllabic and bisyllabic [1S-2S: \(z=54.82, p<.0001\)] as well as monosyllabic and trisyllabic words [1S-3S: \(z=36.58, p<.0001\)]. For example, focally accented syllables in monosyllabic words were 95 ms longer for the BE group and 93 ms longer for the KE group compared to the focal syllables in bisyllabic words. Syllabic composition may have contributed to the effect of prosodic position on durational increase. The differences were significant between verb and object [\(z=5.59, p<.001\)], and verb and subject [\(z=7.06, p<.001\)], showing greater lengthening in focal and accented syllables of the verbs, which included syllables with a coda consonant in *move* and *minimise*, in addition to the tense vowel or the diphthong.

An initial analysis on the effect of accent condition on duration included examination of the duration of the target vowels, and was performed in order to compare the present results with the findings on acoustic correlates in other IndE varieties (Moon, 2002). Similar to the syllable duration, vowel duration showed an increase in accented and focal conditions compared to stressed, but the differences reported were more modest, with a large degree of variation across the speakers in the realisation of focal and accented vowels, especially for the BE group. This variation could be attributed to variability in intrinsic vowel length and an absence of clear contrast between some of the tense and lax vowels across speakers and repetitions.
For example, the mean duration of the tense back vowel /uː/17 in move produced by BE speakers measured at 130 ms [SD=40] in accented syllables and 132 ms [SD=37] in focal syllables, and showed no significant durational differences with the vowel /ɔ/ in borrow [p<.05]. The same back vowel produced by KE speakers in focal syllables was produced as a relatively longer vowel [145 ms, SD=20 ms], and showed significant difference with the same vowel in accented condition [122 ms, SD=25 ms] as well as with the lax vowel /ɔ/ in focal [118 ms, SD=15] and accented [108 ms, SD=10] syllables. For the speakers in this study, syllable duration appeared to be a more reliable measure for testing lengthening effects than vowel duration.

4.5.3.4 Vowel quality
Figure 4.25 below shows the distribution, outliers and the medians of F1 values in Hz for target vowels, presented separately for BE (left hand side) and KE (right hand side) speakers. As noted in Table 3.4 in §3.4.2, both F1 and F2 values for the diphthong were extracted in the first third part of the vowel before the offglide. According to Hypothesis 5E, it was expected that high vowels would have lower F1 values and low vowels have higher F1 in accented versus stressed and accented versus focal syllables. The analysis showed that the F1 of focal and accented vowels had no significant differences [z=0.87, p>.05], thus indicating that this parameter is not used to cue accentual prominence. The differences were distinguishable only in stressed versus accented [z=6.227, p<.001] and stressed versus focal conditions [z=5.32, p<.001]. GROUP had no effect on F1 realisations [t=0.63, p>.05]. However, the overall picture is much more complex.

As can be seen from the figure, there was some variability between groups as well as inter-speaker variation for each group, reflected in wide whiskers on the plots for a number of vowels. Moreover, the observed differences between stressed and focal, and stressed and accented conditions did not necessarily correspond to the lowering of high vowels and raising of low vowels, and at times worked in the opposite direction.

17 A closer examination of this vowel in BE shows that the speakers produced a high back vowel with a lot of durational variation across repetitions and speakers. In the remainder of the chapter, the length diacritic will be placed in parenthesis for the phonemic representation of this vowel in application to BE or when referring to both groups together.
Figure 4.25 Box and whiskers plot for the distribution of F1 values in Hz by vowel type presented by group (BE - top panel, KE - bottom panel) and condition (S - stressed, A - accented, F - focal). I - /ɪ/, E - /ɛ/, O - /ɔ/, u - /u(ː)/, ai - /ɑɪ/, i - /i/.
F1 was consistently lower for the two high back vowels across the words produced by the BE group, despite variation, perhaps reflecting that fact that back vowels behave differently from front vowels and have lower F1 value when more prominent. For KE speakers, the F1 value for /u:/ showed an overlap between all three accent conditions, and the F1 of the vowel /e/ showed no difference between focal and stressed syllables [p>.05]. By contrast, F1 for the vowel /ɛ/ was raised in focal and accented positions by KE speakers. This vowel shows a large degree of overlap for BE speakers, with a lower F1 values in accented condition compared to stressed.

Similar to the F1 results, narrow focus had no effect on F2 [t=2.06, p<.05], with no significant difference between focal and accentual prominence levels, thus indicating that vowel quality is not potentially a reliable cue for narrow focus and may only be used to cue accentual prominence. Tukey tests reported significant differences between stressed and focal, and stressed and accented syllables [S-A: z=−8.207, p<.0001; S-F: z=−9.609, p<.0001], but not between focal and accented syllables [p>.05]. These results are presented in Figure 4.26 below. The figure shows the distribution, outliers and the medians of F2 in Hz for each target vowel for each L1 group. According to Hypothesis 5E, front vowels were expected to show higher F2 values and back vowels were expected to move further away from the centre of the vowel space, showing lower F2 values.

Despite statistical differences between stressed and accented, and between stressed and focal, the figure shows a large amount of overlap in vowel realisations. Unlike the F1 results, GROUP had a significant effect on F2 [t=−2.015, p<.05] with an interaction between GROUP and CONDITION [t=4.02, p<0.01]. As shown in the plot, the F2 results of the vowels produced by BE speakers show greater overlap in the distribution of stressed versus focal or accented vowels compared to the KE group. KE speakers were somewhat more consistent at manipulating the F2 parameter to cue accent, except for the high back vowel /u:/, where the differences between accented versus stressed syllables as well as focal versus stressed syllables were significant only for speakers K1 and K2 [p<.05].
Figure 4.26 Box and whiskers plot for the distribution of F2 values in Hz by vowel type presented by group (BE - top panel, KE - bottom panel) condition (S - stressed, A - accented, F - focal). I - /i/, E - /ɛ/, O - /ɔ/, u - /u(ː)/, ai - /ai/, i - /iː/.
Given such mixed results of the F1 and F2 analyses and the variation across speakers, four vowel targets were plotted in the F1 and F2 plane in Hz for each speaker individually in order to further investigate whether vowel quality is a cue to accentual prominence. These vowel plots are illustrated in Figures 4.27 through to 4.34 below. They show the values for the four vowel targets /i/, /e/, /ɔ/, /u(ː)/ in stressed condition (in black) and the combined tokens for focal and accented conditions (in red). Each point represents one repetition, while the ellipses show the distribution for each target vowel. Due to the accentuation of the pronoun instead of the object, some of the tokens for speaker K2 are missing (Figure 4.32).

As shown in the vowel plots, speakers B3, B4, K1, K3 and K4 centralised the stressed unaccented vowel /e/ and produced the focal or accented counterpart as a more peripheral vowel, with variation in its realisation across the speakers. The differences between the two conditions were not significant for speaker B1 [p>0.05] (Figure 4.27). The tokens of the /e/ vowel produced by speaker B2 in focal/accented condition were dispersed and variable in their realisation (Figure 4.28). As for the front lax vowel, inter-Euclidean distance measurements reported that four speakers (B1, B3, K1 and K3) centralised the high lax vowel /i/ in the stressed condition [p<0.05], while there was a lack of distinction between stressed and focal/accented vowels in the productions of speakers B2, B4 and K4.

The patterns for the back vowels are more complex and mirror the results reported for the F1 and F2 analyses, which showed inter-speaker variation and a degree of overlap for both groups. Two BE speakers made better distinction between stressed and focal/accented conditions for the vowel /ɔ/, as shown in Figures 4.29 and 4.30, while only one KE speaker (K4) made the vowel contrast in two accent conditions (Figure 4.34). The formants for the high back vowel displayed differences in the vowel quality depending on accent condition for two KE (Figures 4.31 and 4.32) and two BE speakers (Figures 4.29 and 4.30). For speaker K2, the difference was significant based on inter-Euclidean distances between the centroids of the vowel in two accent conditions (p<0.02), but this difference worked in the opposite direction, where the vowel in accented syllables was more raised and somewhat centralised. For speakers B1, B4, K3 and K4, the ellipse plots and the distribution of the tokens for the vowel /u(ː)/ showed a large degree of overlap, suggestive of no
effect of accent condition on the realisation of this vowel. Overall, the results indicate that vowel quality is not a cue to focal and accentual prominence, thus disproving Hypothesis 5E.

Figure 4.27 Vowels /ɪ/, /ɛ/, /ɔ/ and /u(ː)/ for speaker B1 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz.

Figure 4.28 Vowels /ɪ/, /ɛ/, /ɔ/ and /u(ː)/ for speaker B2 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz.

Figure 4.29 Vowels /ɪ/, /ɛ/, /ɔ/ and /u(ː)/ for speaker B3 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values.

Figure 4.30 Vowels /ɪ/, /ɛ/, /ɔ/ and /u(ː)/ for speaker B4 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values.
Figure 4.31 Vowels /ɪ/, /ɛ/, /ɔ/ and /uː/ for speaker K1 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values Hz.

Figure 4.32 Vowels /ɪ/, /ɔ/ and /uː/ for speaker K2 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in two stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz.

Figure 4.33 Vowels /ɪ/, /ɛ/, /ɔ/ and /uː/ for speaker K3 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz.

Figure 4.34 Vowels /ɪ/, /ɛ/, /ɔ/ and /uː/ for speaker K4 plotted in the F2 x F1 plane for the data extracted at the vowel midpoint in stressed and accented or focal conditions. The x-axis shows F2 values and the y-axis shows F1 values in Hz.
4.5.3.5 Summary

The results of the acoustic cues to prominence levels showed that absolute $f_0$ height was a reliable cue to accentual prominence for all speakers in this study. Marking of narrow focus, on the other hand, was cued by the $f_0$ height only for BE speakers and two KE speakers (excluding K2 and K3), showing higher peaks in focal versus accented conditions. This finding partially supports Hypothesis 5C that the absolute $f_0$ height scales higher with each level of prominence. Examination of pitch excursion in focal versus accented syllables, however, showed more consistent results with a significant increase in the magnitude of the rise in narrow focus for all speakers. The finding supports Hypothesis 5D that pitch movement on focally accented words involves lowering of the low target thus contributing to a greater rise in magnitude. Pitch excursion may be a more reliable measure to distinguish focal and accentual prominence levels for the speakers of IndE in this study, but this issue needs further investigation.

In addition, the $f_0$ results revealed differences in the use of pitch range and the degree of $f_0$ manipulation between the two groups, supporting Hypothesis 7 that predicted differences between the groups. The rises produced by the BE group were significantly higher. In addition, the utterances produced by KE speakers often followed declination patterns, which may have further contributed to shallower rises on the verbs and objects compared to BE speakers.

Similar to the absolute $f_0$ height, RMS amplitude was used to cue accentual prominence by all speakers but was only used to cue focal prominence by all BE and two KE speakers. This finding, in part, supports Hypothesis 5B that amplitude is used to cue accentual and focal prominence, as only six of the eight speakers employed it for focal prominence. The same two speakers (K2 and K3) who did not manipulate $f_0$ height made no distinction between focus and accent conditions based on RMS amplitude. The results also revealed that the differences in the RMS amplitude values depending on accent condition showed a large degree of overlap and variation across the speakers, especially for the KE group. The position of the word in the utterance had an effect on amplitude with the lowest dB-RMS values reported for the target syllables in the object. Additionally, the vowel /u(ː)/ in the target word move showed lower RMS amplitude compared to rest of the vowels. Both of these findings are
consistent with previous research (Peterson & Lehiste, 1960; Ladd & Silverman, 1996).

Despite a large degree of inter-speaker variation and overlap between the two accent conditions of accent and narrow focus, duration was a reliable cue to accentual and focal prominence for all speakers, showing the effect of accent condition on syllable length (Hypothesis 5A). The differences between stressed and accented and between stressed and focal were greater than between accented and focal syllables. Monosyllabic words showed a greater amount lengthening compared to bi- and trisyllabic words. Syllable duration also proved to be a more effective measure to accent and focus than vowel duration, most likely due to variation in vowel realisations and durational vowel contrasts for the BE group.

The F1 and F2 results presented a mixed results account. First, the vowel quality was non-distinguishable between accented and focal conditions. There were no significant differences in the F1 and F2 values between accented and focal conditions. Second, despite some distinction between stressed and accented or stressed and focal, there was no consistency across the speakers in the production of vowels in accented/focal versus stressed syllables. The speakers made a clear distinction for some vowels but not others, with no distinct pattern for each L1 group. For most speakers, the back vowels often showed no effect of accent on their realisation. The findings do not support Hypothesis 5E.

4.6 Discussion

Experiment 1 examined the intonational realisation of information structure in BE and KE, investigating the following features:

a) accentuation and phrasing in broad versus narrow focus
b) prosodic strategies for marking focus
c) and phonetic correlates to cue accentual and focal prominence levels.

The findings have shown that some of the features found in BE and KE are similar to well-established Englishes, and some could be applied to IndE as a variety. In addition, differences between BE and KE have presented a variable picture and do not cross over across all features investigated. A number of hypotheses have been tested and will be discussed below in relation to each feature or acoustic parameter.
IndE has always been known for accenting most words in an utterance, including content and often function words (Bansal, 1969; Gumperz, 1982; Wells, 1982; Gargesh, 2004; Wiltshire & Harnsberger, 2006). This pattern is also consistent with the use of accentual prominence in other new varieties of English (Udofot, 2003 on NigE; Gut, 2004, 2005). The accentuation patterns produced in simple declaratives by the speakers in this study do not fully support previous findings. Unlike in Gujarati and Tamil English (Wiltshire & Hansberger, 2006), where prominence is placed on most content words regardless of the speakers’ L1, the number of accented versus stressed syllables in broad focus was higher for KE speakers (64% to 36%) than for BE speakers (47% to 53%), with greater accentual density for KE speakers who were more likely to accent most content words in non-focal contexts, fitting with the generally observed accent placement in IndE. Other patterns in KE broad focus structure were the presence of pitch accents on the subject and the verb, and the subject and the object, but the proportion of utterances with these patterns were relatively small.

In contrast, BE speakers produced two patterns. The first one was the placement of accent on the first and the last content word, similar to a ‘canonical’ interpretation of broad focus in English (Ladd, 1996, 2008). The nuclear accent placement by BE speakers was also similar to Rajasthani (Dhamija, 1976) and Punjabi (Sethi, 1980) speakers of English, who place nuclear prominence on the last content word in a phrase. The second pattern, used to a lesser degree, was accentuation of all content words. The findings do not fully support Hypothesis 1 which predicted that all speakers would accent most or all content words, although they give evidence to Hypothesis 7 about the differences between BE and KE which could most likely be attributed to L1 influence. It is possible that accenting a large number of words is a feature present in IndE as a result of IndE development, as suggested by Wiltshire and Harnsberger (2006). However, this pattern of accenting is present on a continuum, as shown by the fact that BE speakers in this study produced some of the utterances with the accents on all content words. Baker and Bradlow (2007) found that IndE speakers deaccented the second mention of a word relative to its first mention in a discourse, thus indicating that not all words were accented, although second mention words were deaccented significantly less than the same words in the utterances produced by AmE speakers. Further experimental investigation with a wider range of L1s and educational backgrounds is needed in
order to make generalisations about this feature in IndE phonology. Given that the materials used in this experiment included short declaratives, accentuation will be re-examined in Chapter 7.

In narrow focus, the accentuation and prosodic grouping patterns did not show clear differences based on the speakers’ L1s, suggesting that Hypothesis 7 is not true for narrow focal contexts. Moreover, supporting Hypothesis 4, the speakers had a ‘hybrid’ system for marking focus with multiple strategies, including deaccenting, post-focal pitch range compression without deaccenting and re-phrasing. Hypothesis 2, stating that deaccenting is the main focus marking strategy similar to well-established varieties of English (Gussenhoven, 1983; Ladd, 1996, 2008; Xu & Xu, 2005), was not true. The speakers applied deaccenting on words following a narrow focused constituent as one of the ways to mark focus prosodically, but there was a large degree of inter-speaker variation, and differences in the proportional use of post-focal deaccenting compared to the other strategies. Speakers K3, K4, B1 and B3 used deaccenting in 70-90% of their samples, while for speakers K1, K2, B2 and B4 its use was under 50% in utterances where initial and medial words were in narrow focus. Pre-focal deaccenting was observed across the repetitions for some speakers, which is not surprising as the status of pre-focal deaccenting in English is optional (Xu & Xu, 2005).

The second strategy used for narrow focus marking which is distinctly different from AmE, BrE or AusE was post-focal pitch compression with the presence of accents after the narrow focused word. Its use varied from 7% to 35% of utterances across the speakers. The drop in $f0$ after the accented syllable was quite significant, with the $f0$ of the following accent 50-60 Hz lower when compared to the peak associated with the focally accented syllable. As discussed in §2.4.3 and §4.3, pitch compression without any changes in accentuation is characteristic for a number of languages spoken in India, where focus structure is distinguished on the basis of gradient phonetic parameters (Patil et al., 2008 on Hindi; Keane, 2014 on Tamil). This is also similar to Italian, where a nuclear focal accent is followed by another accent (Swertz et al., 1999; Grice et al., 2005). As suggested by Féry and Kentner (2010), pitch compression without changes in accentuation in Hindi plays a similar role as deaccenting in well-established Englishes. Pitch compression could be a way of
accommodation towards focus marking in English, more frequently used with deaccenting and less frequently without it.

In a number of NE varieties (Gut, 2005 on NigE; Grabe & Low, 1999 on SingE; Zerbian, 2013 on BlSAfE), the lack or degree of post-focal deaccenting use has been linked to language proficiency level, showing a correlation with the proportion of accented words. Similar findings have been reported in L2 intonation studies. For example, only advanced Vietnamese L1 speakers of English but not the speakers of lower levels used post-focal deaccenting (Nguyễn et al., 2008). In relation to IndE, there is also a question of variation within, based on educational level and socioeconomic background (Fuchs, 2013), which could correlate with the amount of L1 influence. Pitch compression without deaccenting could be used more frequently by less educated speakers of IndE or those who have less exposure to English.

In addition to pitch range compression, six speakers in this study (K1, K2, B1, B2, B3 and B4) relied on re-phrasing, isolating a focused word and at times an adjacent word into a full intonational phrase to signal narrow focus. This pattern has not been reported for the languages spoken in the subcontinent. Ueyama and Jun (1998) suggested that forming shorter sentences is more characteristic for lower levels of L2 speakers. The speakers in this study, however, were fluent in English. Previous research on IndE has pointed to the pattern of ‘incorrect’ grouping of words (Bansal, 1969; Dhamija, 1976) as one of its distinctive features but did not specify whether this was the case for narrow or broad focus structure. Re-phrasing could be a feature of IndE phonology that emerged in the process of its development, with some variability across IndE speakers.

The choice of pitch shape on a focused word was another prominent feature recorded for the IndE speakers in this study. The $\mu\theta$ contour on narrow focused words was a distinctive rise (LH), which is a pattern previously recorded for IndE speakers of L1 Tamil, L1 Gujarati and L1 Hindi, with greater use of rises for Gujarati and Hindi English speakers (Wiltshire & Harnsberger, 2006; Moon, 2002). As discussed in §2.4.3, this is a characteristic feature of the intonational phonologies of Bengali (Hayes & Lahiri, 1991; Khan, 2009, 2014), Hindi (Patil et al., 2008; Genzel & Kügler, 2010) and Tamil (Kean, 2007, 2014). For all BE and KE speakers, rising pitch was used across the three narrow focus structures of sentence-initial, sentence-medial and sentence-final. This is in contrast to broad focus sentences, which showed
variation in pitch shapes used on accented words, including a sustained high pitch and less frequently a low pitch, the latter more common in sentence-final contexts.

Impressionistic observations also revealed noticeable differences between the groups in the alignment of focal peaks. The nature of the high tone target (H) may indicate potential phonetic or phonological differences in the pitch shape associated with narrow focus. The groups could be using different types of pitch accents in narrow focus, similar to Moon’s (2002) preliminary analysis of pitch accent categories on focally accented words. In Hindi English (HE) the focused item was associated with an L+H*, with a greater magnitude of rise. By contrast, the association in Telugu English (TelE) is with an H*, with the f0 showing no evident drop at the onset of the accented syllable.

The results of the phonetic cues to post-lexical prominence show that BE and KE speakers in this study use f0 to cue accentual prominence, similar to the speakers of other well-researched varieties of English (Cooper et al., 1985; Beckman, 1996; Gussenhoven, 2004; Xu & Xu, 2005). As for the focal prominence level, focal syllables showed greater absolute f0 than accented ones in the productions of BE and two KE speakers (K1 and K4). However, all speakers produced focal syllables with a higher pitch excursion compared to that for accented syllables. Absolute f0 height may be a less reliable cue to focal prominence as opposed to f0 excursion, at least for the speakers in this study. Cooper at al. (1985) and Xu and Xu (2005) suggested that one of the main effects of focus on f0 in AmE is a lowering of f0 on post-focal words. Despite the fact that the f0 values after the post-focal words were not measured across the whole corpus, the utterances in narrow focus showed pitch range compression with or without deaccenting, as discussed previously.

While both L1 groups showed consistent differences in the f0 rises in focal versus accented syllables, the narrow focused accents produced by the KE speakers showed a smaller increase in the f0 trajectory of the rise, suggesting a narrower pitch range use compared to the BE group. For example, the mean rise on a focused subject measured at around 9 semitones for the BE group and 5.5 semitones for the KE group. This feature could be related to the speakers’ L1 influence. It is also noteworthy to mention that the sentences produced by the KE group in narrow focus showed a strong declination pattern, and the f0 on sentence-final words was lower compared to the f0 on sentence-medial words even in narrow focus contexts. For BE speakers, there was often a suspension of declination, and the rises on focused syllables in
sentence-final position were significantly higher than the rises on sentence-medial accented syllables.

The degree of prominence (stressed versus accented and accented versus focal) was cued by the durational increase of the target syllable. The highest increase occurred in stressed versus accented and in stressed versus focal syllables. There was no difference between the two L1 groups, but a large degree of variation across the speakers, especially in KE. This finding shows that the use of this parameter may be similar to that of well-established Englishes. Several studies on English reported an increase in duration for focused versus accented conditions (Cooper et al., 1985; Greenberg et al., 2003; Xu & Xu, 2005; Breen et al., 2010) and for accented versus stressed conditions (Turk & Sawusch, 1996). However, duration has more correlation with lexical prominence in English as opposed to f0 and intensity which are more reliable correlates of accentual prominence (Beckman & Edwards, 1994; Sluijter & van Heuven, 1996b). In addition, the results showed greater lengthening when the target word in narrow focus was shorter, in line with previous research on AmE (Cooper et al., 1985) which found syllables of monosyllabic focused words show significantly greater lengthening compared to the words with more than one syllable.

The durational results differ from the findings reported on other IndE varieties that suggested duration to be a marginal cue to prominence. Moon (2002) reported a lack of durational cue on the level of focal prominence in TelE and HE. These discrepancies in the results could be indicative of the speakers’ L1 influence. L2 English intonation studies have illustrated that duration is more likely to be used to mark accentual prominence if this parameter is present in the speakers’ L1 (see §2.2.2). For example, Korean L1 speakers in McGory’s (1997) study used duration as a cue to accentual prominence, but Vietnamese L1 speakers did not as a result of the limited use of this acoustic parameter in Vietnamese. For L1 speakers of Kannada and Bengali, durational contrasts distinguishing stressed, accented and focal levels may be present in their L1, and therefore used English. Alternatively, the choice of segmental material when measuring focal prominence may have led to the differences between the present study on IndE and previous research. Moon (2002) investigated the duration of vowels in accented and focal syllables. A preliminary study by Maxwell and Fletcher (2011) also examined the vowel duration in a smaller set of data and found that BE speakers did not lengthen vowels to mark focal prominence. In the present experiment, the duration of the vowel was initially investigated, with BE
speakers showing a large degree of inter- and intra-speaker variation in the intrinsic vowel duration and an inconsistency in tense-lax vowel contrasts. Syllables durations showed more consistent patterns.

RMS amplitude was higher in syllables that were accented and focal for all BE speakers and two KE speakers, thus supporting Hypothesis 5B for six speakers. This finding is consistent with previous research on English (Beckman, 1986; Harrington et al., 2000; Kochanski et al., 2005b; Breen et al., 2010). The results also support the findings of Moon (2002) who reported higher amplitude on narrow focused syllables in HE and TelE. Two speakers (K2 and K3) in this study did not make a distinction between focus and accent conditions based on RMS amplitude. Overall, the tokens produced by the KE group showed more variability across repetitions than the tokens produced by the BE group, in addition to a greater overlap in the distribution of RMS amplitude values between focal and accented syllables. This may have been a reflection of a less robust use of this parameter in Kannada. In the investigation of Kannada speech rhythm, Savithri (1995) found that only half of the target syllables that were perceived to be louder showed an increase in intensity compared to the preceding syllables. Keane (2006a) found that loudness was not a phonetic cue to lexical prominence in Tamil which, like Kannada, is a Dravidian language. Given the lack of research on the acoustics of prominence realisation in Kannada, especially on a post-lexical level, this is a tentative suggestion at this stage.

While RMS amplitude differences based on prominence levels were statistically significant for six speakers, the differences between accented and stressed syllables were greater than those between accented and focal, similar to the results on f0 height and syllable duration. Focally accented syllables showed a mean increase of 3-5 dB compared to accented syllables, which is comparable to AmE. Sluijter and van Heuven (1996b) reported a mean of 5 dB on narrow focused syllables for AmE and accented versus stressed syllables showed a mean increase of 8 dB. Harrington et al. (2000) reported an increase of 5-10 dB in the same two accent conditions for AusE. In addition, the position of the word had an effect on the RMS amplitude with the lowest values on the accented and focal tokens that were sentence-final. This could be explained by the fact that RMS amplitude values also track f0 values and the latter undergo the process of declination. In addition to the sentence position, intrinsic
vowel intensity had an effect on RMS amplitude with the lowest values found for the back high vowel /u(ː)/ (Peterson & Lehiste, 1960; Ladd & Silverman, 1984).

Vowel quality was not sharpened by the presence of narrow focus, indicating that this acoustic parameter is not a correlate of focal prominence for the IndE speakers in this study, unlike in AmE (Sluijter & van Heuven, 1996b) and AusE (Harrington & Fletcher, 1997). The main differences were observed in stressed vowels compared to either focal or accented vowels. These differences, however, did not always correspond to accented or focal vowels having a sharper acoustic or fuller quality and stressed vowels being more centralised. For some of the speakers, the accent condition had the opposite effect. Moreover the speakers made a clear distinction for some vowels but not others, regardless of the L1 group. The vowel /e/ was somewhat more peripheral in accented or focal contexts (with six speakers making the distinction), while the vowels /u(ː)/ and /a/ showed the greatest overlap when plotted in the F1 and F1 frame. Only speakers B1, B2 and K4 produced a fuller vowel /a/, while speakers K1, K2, B2, B3 produced a fuller vowel /u(ː)/ in accented or focal syllables. Unlike f0, RMS amplitude and duration, vowel quality had a marginal and variable affect on accent, and was not a reliable cue to either accentual or focal prominence for both L1 groups, thus disproving Hypothesis 5E.

4.7 Chapter summary

To conclude, the results of Experiment 1 give a detailed account of the intonational realisation of accentual prominence and focus in two varieties of IndE. The findings contribute to the intonational typology of Englishes around the world and show that some of the features reported for BE and KE are similar to other well-established varieties of English. The speakers place accents on certain words, and accent location is the key prosodic strategy. BE and KE intonation includes post-focal deaccenting and shows three acoustic correlates to cue accentual prominence level (duration, f0 and amplitude), similar to BrE, AmE or AusE.

The findings also support the notion that IndE has developed its own phonology (as argued in §1.3.2), with a more variable system of narrow focus marking compared to well-established Englishes that includes pitch range compression without deaccenting, re-phrasing and the absence of vowel quality to cue
both accentual and focal prominence. Similar to other IndE varieties (e.g. Gujarati or Tamil English), accentual density is high in KE, with most of the content words assigned an accent. In line with Moon’s (2002) findings on L1 Telugu and L1 Hindi speakers, $f_0$ height and amplitude were used to cue narrow focus for six speakers in this study.

In addition, not all of the results of Experiment 1 support previous research on IndE. A number of the features reported for the speakers in this study differ from generalisations about IndE as a variety. Unlike in HE or TelE (Moon, 2002), durational increase together with the accentuation is a cue to focal prominence in this study despite a large degree of variation. BE intonation shows neither accentuation of most content words nor the use of a narrow pitch range, contrary to the claims about IndE (see Nihalani et al., 2004; Wiltshire & Harnsberger, 2006; Gumperz, 1982). Greater rise magnitude was found on focused words produced by BE speakers, similar to L1 Hindi speakers but unlike L1 speakers of Telugu (Moon, 2002). This emphasises the need for fine-grained acoustic analyses in order to distinguish the extent of similarities and differences across IndE varieties.

Despite ample evidence of certain features being specific for each L1 background, other features show an overlap without being confined to one group and reflect inter-speaker variation. These include proportional use of focus marking strategies, manipulation of RMS amplitude and $f_0$ height. The differences between BE and KE appear to be more discrete and could be the result of intrinsic (L1) and extrinsic (language use, education, socioeconomic background) influence, more evident in the investigation of phonetic parameters, or even the result of universal constraints on speech production and the development of intonation (Ueyama & Jun, 1998; Mennen, 2007). Despite the fact that all speakers differentiated stressed from accented syllables, and accented from focal syllables, not all correlates were manipulated in the same way. The main cues to accentual and focal prominence were $f_0$, amplitude and duration, however, the findings showed a large degree of overlap across the speakers and repetitions. Pitch excursion was a more reliable cue to narrow focus than the absolute $f_0$ height, indicating that BE and KE speakers produced rises

---

18 The role of universal factors in the development of L2 intonation remains a less researched topic in the literature, with the majority of the studies based on L2 speakers of English. As argued by Mennen (2007), the similarities reported for L2 speakers of English from various L1 backgrounds could be due to the idiosyncrasies of English intonation. See Mennen (2007) and Mennen, Chen and Karlsson (2010) for a more detailed discussion.
on focally accented words. Moreover, BE rises showed greater magnitude, also found on focused words in HE (Moon, 2002). The rising gesture on focal and accented words is precisely what the next chapter will investigate. It will examine the phonetics of the tonal alignment and $f0$ scaling to determine the phonological pitch accent categories for BE and KE.
Chapter 5: Experiment 2 – Tonal alignment and scaling characteristics of rises

5.1 Introduction

The AM framework assumes that high (H) and low (L) tones associated with prominent syllables correspond to tonal targets and can be mapped onto the segments of an utterance. These tonal targets are defined along two dimensions or ‘coordinates’ (D’Imperio, 2012). These are: alignment, which is their temporal relationship with the segmental string or prosodic units; and scaling, corresponding to the f0 value of the target in the speaker’s pitch range.

The findings reported in §4.5.2.4 showed that the speakers in this study produced a rising pitch movement on accented words, involving the rise from a low to a high tone (LH). This pitch shape was observed on narrow focused words for both groups and a number of accented words in broad focus structure. As discussed in §2.4, rises on accented words were reported for a number of IndE varieties (Moon, 2002; Wiltshire & Harnsberger, 2006). These rises were interpreted as pitch accents (L*+H, L+H* or even H*) according to the AM approach. The analyses, however, were preliminary and required further investigation to posit the pitch accent inventory for IndE.

Experiment 2 was designed to examine the tonal alignment and scaling characteristics of the rise in three contexts: prenuclear, nuclear non-focal and nuclear focal. Nuclear non-focal tokens were subsequently excluded from the analyses due to variability in pitch movement realisations across speakers and repetitions. Nuclear accented phrase-final words were mostly produced with a high plateau or as a low tone preceding a rise at the boundary. Only two speakers (B1 and K2) consistently used a rising pitch movement on nuclear non-focal words under broad focus. In addition, there was variation in accent placement across the speakers. As a result, the current chapter examines the rises in two prosodic positions, prenuclear and nuclear focal only. Nuclear accents in broad focus structure will be discussed in the inventory of pitch accents in Chapter 7.

The aim of this experiment was to determine what type of pitch accent the speakers produced in two prosodic contexts and to investigate any differences or similarities across the speakers. The investigation includes the phonetic realisation of
the rising gesture in order to examine possible phonetic differences in the alignment and scaling of putative targets of the rise. It will also examine the effect of the segmental environment on peak alignment. This chapter addresses two main questions:

1. What are the alignment and scaling characteristics of the tonal targets and the rising gesture associated with the prominent syllable?
2. Are there any differences in tonal alignment in prenuclear versus nuclear focal rises between BE and KE?

Section 5.2 begins the chapter with a brief overview of the literature relevant to tonal alignment and scaling of tonal targets. Section 5.3 presents the research hypotheses. Next, §5.4 includes an overview of the materials and analyses, mainly to remind the reader of the measurements and temporal intervals examined in this experiment. The results sections are presented by prosodic position, prenuclear (§5.5.1) and nuclear focal (§5.5.2), followed by a comparison between the two positions in §5.5.3. Section 5.6 includes the discussion of the results, while §5.7 concludes the chapter with a summary of the findings and discussion points.

5.2 Previous research on tonal alignment and scaling

Bruce’s (1977) pioneering work on Swedish word accents showed that differences in the $f_0$ alignment of the falling pitch on the accented words corresponded to two types of lexical accents, where Accent I had an earlier alignment of the $f_0$ movement in relation to the accented syllable as compared to Accent II. Since his work, there have been a large number of studies investigating tonal alignment across different languages. These studies demonstrate that tonal alignment plays an important part in current theories of intonational phonology due to its relationship with intonational meaning distinctions and the typology of intonational categories (Prieto, 2011). Tonal alignment gives a fine-grained phonetic interpretation and helps determine whether there are any categorical intonation contrasts corresponding to differences in meaning, pragmatic in the case of English, or whether the existing contrasts are gradient and non-categorical. Despite a wide body of research, tonal alignment remains a complex notion in the AM framework with a number of competing unresolved issues (Arvaniti, 2012; Ladd, 2008). In order to lay the groundwork for the present experiment, the
following sections review the fundamental concepts relative to the dimensions of tonal targets: tonal association and alignment, stability of tonal alignment, ‘segmental anchoring hypothesis’, and tonal scaling.

5.2.1 Tonal association and tonal alignment

In the first AM account of English intonation, Pierrehumbert (1980) proposed two types of tones: pitch accents, marking metrically strong syllables; and boundary tones, marking utterance edges. At present, one of the fundamental concepts within the theory is that both pitch accents and edge tones have a primary and/or secondary association to the tone-bearing unit (TBU), the syllable in English.¹⁹ The notion of secondary association is not the focus of the present investigation and remains a debatable issue in the literature (Ladd 1983, 2008; Gussenhoven 2004; Grice et al., 2000; Prieto et al., 2006). For a brief discussion on this topic, see §2.1.2, Chapter 2. The present investigation is concerned with the primary phonological association, whereby pitch accents (for example H* or L*) are assumed to have a primary association with prominent (accented and stressed) syllables, and phrase accents (such as H-, H%) are primarily associated with the edges of prosodic constituents. Pitch accents and edge tones associated with certain TBUs are also aligned within those TBUs. As a result, intonational categories are created on the basis of tonal association as well as the temporal alignment of tones.

Experimental studies on the perception and production of bitonal pitch accents in AmE (Pierrehumbert, 1980; Pierrehumbert & Steele, 1989; Arvaniti & Garding, 2007) have shown that the alignment of L and H targets in relation to the accented syllable correspond to a categorical distinction between early aligned and late aligned pitch accents. As a result, these pitch accents (L+H* and L*+H) are treated as two separate phonological categories, where a star notation in the ToBI annotational system represents the primary association of the tone as well as its alignment with the prominent syllable. Thus, in L*+H, the L tone is aligned with the stressed syllable and is associated with that syllable, while the H is realised in the post-tonic syllable representing a trailing tone (Grice, 1995). In L+H*, the H tone is timed within the stressed syllable, and the L aligns somewhere before the stressed

¹⁹ The definition of the TBU varies across languages and can be identified as the mora or the segment.
syllable. A schematic representation is illustrated below, with the stressed syllable marked by the grey box (Figure 5.1).

![Schematic representation of L+H* and L*+H.](image)

Figure 5.1 Schematic representation of the phonetic realisation of L+H* and L*+H. The grey boxes represent the accented syllable, while the solid line represents the $f0$ movement corresponding to the rising gesture.

The relationship between the phonological association and phonetic alignment of L and H tones was taken for granted in the early autosegmental literature (Arvaniti, 2012). Studies by Silverman and Pierrehumbert (1990), van Santen and Hirschberg (1994) and Prieto et al. (1995) were the first to challenge this assumption and found that the temporal alignment of $f0$ events with the segmental material did not necessarily follow the association principle. These researchers found that sometimes the $f0$ peak corresponding to the H tone was delayed and aligned after the end of the stressed syllable it was phonologically associated with. Peak delay was found to be influenced by various prosodic and phonological effects, showing that the relationship between association and alignment is not as straightforward as was originally accepted. These effects are discussed in the next section.

### 5.2.2 Tonal stability and the ‘segmental anchoring hypothesis’

Following Steele’s (1986) study that reported variation in the realisation of the peak in nuclear versus prenuclear H* accents, Silverman and Pierrehumbert (1990) investigated the alignment of peaks in AmE prenuclear accents. The researchers addressed the question of the similarity or difference in prenuclear versus nuclear accents, and the stability of tonal alignment under different conditions. Firstly, they found that peak proportions showed more rule-governed regularities than absolute peak delays. Prenuclear peaks were aligned in proportion to the duration of the stressed syllable resulting in later peaks for longer syllable rhymes, thus confirming the effect of syllable duration on tonal alignment as opposed to the alignment of the H tone at a fixed distance from the rhyme onset. Secondly, speech rate and right hand
prosodic context had a strong effect on the alignment of the H tone. The peaks were aligned later when there were more unstressed syllables following the target syllable.

Silverman and Pierrehumbert concluded that the amount of prosodic lengthening determined the location of the peak, providing evidence for the differences in peak alignment between prenuclear and nuclear accents. With the assumption that prenuclear syllables are shorter than nuclear syllables, earlier peaks in nuclear accents were explained in terms of greater lengthening before the upcoming phrase boundary. Another phonetic explanation for earlier peaks was given in terms of gestural overlap and tonal repulsion. In nuclear accents, gestural overlap is a mechanism that describes an interruption of the articulatory gesture for the accent by the following tone marking the edge of the phrase. Tonal repulsion corresponds to the mechanism when the entire gesture for the first accent is shifted earlier when the next accent is too close in time. Subsequently, several studies have shown that peaks align earlier in nuclear contexts for a number of languages (Arvaniti & Baltazani, 2005 on Greek; Schepman, Lickley & Ladd, 2006 on Dutch; Ladd et al., 2009 on BrE; Dehé, 2010 on Icelandic).

van Santen and Hirschberg (1994) investigated prenuclear peaks of H* accents in AmE. They showed that the alignment of peaks could be modelled in terms of the segmental composition of the accented syllable and that the alignment of the turning points was attached to certain anchoring points located proportionally to the entire f0 movement of the pitch accent gesture. Prieto et al. (1995) examined H* accents in Mexican Spanish and manipulated the prosodic condition, the accented syllable position and the distance in syllables to the next stressed syllable. Unlike Pierrehumbert and Silverman (1990), they examined the alignment of both the start of the rise (f0 valley) and the f0 peak. They found that L aligned at the onset of the stressed syllable with a slight move to the right when the stressed syllable immediately preceded the accented syllable. They attributed this to tonal repulsion from a left hand prosodic context. Overall, however, valley location was found to be more constant and less affected by the prosodic position and the distance to the next accented syllable, contrary to their results on peak alignment.

With this in mind, the authors suggested the onset of the syllable to be a more reliable measure of peak delay as opposed to the onset of the rhyme. Similar to the findings of Pierrehumbert and Silverman (1990), the alignment of the peak showed greater variation. The temporal interval of peak delay measured from the syllable
onset increased with an increase in vowel and onset durations. Peak alignment was affected by a number of prosodic factors, such as adjacency to word or phrase boundaries and the subsequent stress clash, giving strong evidence that both syllable duration and right prosodic context have an effect on H alignment. In addition, the alignment and scaling of the pitch gesture as a whole were variable, which was contradictory to the fixed rise-time hypothesis (Fujisaki, 1983; Collier & Cohen, 1990). The study showed that syllables with earlier peaks had shorter rise durations and a steeper rise in English.

Unlike earlier studies mentioned above which examined H* accents, Arvaniti, Ladd and Menn (1998) investigated rising L*+H accents in Greek. These authors found that neither of the tones aligned within the accented syllable, thus challenging the assumption about the phonological association of tones to prominent syllables. Their findings showed that neither of the tones aligned with respect to the other, contradicting the original analysis of bitonal accents proposed by Pierrehumbert (1980), and Pierrehumbert and Beckman (1988). Instead, the temporal interval between the two tones correlated with the duration of the accented syllable. For some of the speakers, the alignment of the peak was affected by ‘tonal crowding’, corresponding to f0 movements being too close to each other, such as a pitch accent closely followed by another accented syllable or an accent immediately following or preceding a phrase boundary. Building on the findings of van Santen and Hirschberg (1994) and Prieto et al. (1995), Arvaniti and her colleagues concluded that the beginning and end points of the f0 movement were anchored to specific locations and that they independently aligned with specifiable points in the segmental string. This phenomenon was introduced as the ‘segmental anchoring hypothesis’. In this view, the slope and duration of the f0 movement are not constant but vary depending on the segmental material they are aligned with.

Since the introduction of the segmental anchoring hypothesis, a number of studies have replicated the results and examined the segmental anchoring of tonal targets. Some of the studies even questioned the validity of the hypothesis in its strict view (see Prieto & Torreira, 2007). Phonological vowel length and syllable structure have been shown to have a strong effect on peak alignment (D’Imperio, 2000; Ladd et. al, 2000; Schepman et al., 2006; Prieto & Torreira, 2007; Prieto, 2009; Ladd et al., 2009). D’Imperio (2000), for example, showed that in Neapolitan Italian the peak aligned differently depending on the syllabic structure. In open syllables, the peaks
were located at the end of the accented vowel, while in closed syllables the peak aligned with the coda consonant. Similarly, the effect of the coda consonant was reported for falling nuclear accents in Catalan (Prieto, 2009) and prenuclear rising accents in Icelandic (Dehé, 2010). In the investigation of prenuclear rising pitch accents in Dutch, Ladd et al. (2000) found that the alignment of H differed depending on whether the vowel was phonologically short or long. The peak was aligned at the end of the following consonant when the accented vowel was short. On the other hand, when the vowel was long, the peak was anchored at the end of this vowel. They suggested that the difference could be due to the syllabic structure, such as the syllabification of single intervocalic consonants with a following onset after a long vowel. A subsequent study investigating nuclear accents in Dutch (Schepman et al., 2006) failed to find supporting evidence that syllable structure had an effect on peak alignment. Instead, vowel length had a greater effect on the alignment. In line with Silverman and Pierrehumbert (1990), the researchers also suggested that, at least in the nuclear position, peak alignment could be predicted as a proportional measure.

Ladd et al. (2009) examined the alignment of both prenuclear and nuclear peaks in BrE varieties and found different patterns in the alignment of nuclear versus prenuclear accents. They reported that prenuclear peaks aligned later with phonologically short vowels, and similar to Ladd et al. (2000) but contrary to Ladd et al. (2006), this effect was attributed to the syllabic structure and not the vowel length alone. The timing of nuclear peaks, on the other hand, was not affected by syllable structure. Rather it was influenced by the proximity of the accented syllable to the right edges of a prosodic boundary. Prieto and Torreira (2007) examined the effect of syllable structure and speech rate on peak alignment of LH* prenuclear accents in Peninsular Spanish. These authors also found that the tonal targets did not necessarily align with segmental anchors but were affected by syllabic structure as well speech rate. Prieto and Torreira suggested that segmental anchoring could be interpreted in a less constrained way. There was clearly a coordination of the f0 movement with the accompanying segmental material, but the phonetic alignment of this movement showed variability.

Contrary to the original interpretation of the segmental anchoring hypothesis that assumed that the two targets aligned independently, the findings of Atterer and Ladd (2004) suggested that the L and the H targets of a rising accent in two varieties of German did not align independently, but moved in parallel as part of the same
phonological gesture, with the later alignment of L leading to greater peak delay. Similarly, Frota (2002) also found a relationship between the two targets when examining European Portuguese nuclear falling accents where the L location was dependent on peak location.

In summary, the research to date has shown that tonal alignment and especially the timing of f0 peaks can be affected by a number of factors including tonal crowding, speech rate, syllable structure and prosodic boundaries. Thus, in view of more recent research, the segmental anchoring hypothesis in its original sense has been reconsidered and is now accepted in a less strict notion (D’Imperio, 2012). An alternative approach to f0 movement is from the point of view of articulatory phonology (Browman & Goldstein, 1986), whereby pitch accents are treated as speech gestures (Ladd, 2006; Prieto & Torreira, 2007). According to Ladd (2006), segmental anchoring is an empirical finding and not a phonological hypothesis. Studies on segmental anchoring have clearly shown that the f0 gesture is related to the duration of the unit it is associated with, and that the low target has a more stable alignment than the peak (excluding Sullivan, 2007, on Belfast Irish).

More stable alignment could also apply to the first tonal target, not limited to the valley. Prieto (2006, 2009) demonstrated that in Catalan nuclear falling accents, the beginning of the fall was consistently timed with the onset of the accented syllable while the L location was more variable. A more stable timing of the first tonal target (L in rises or H in falls) could be explained in terms of its more synchronous supralaryngeal movement starting at the syllable onset, while the timing of the second target occurs towards the end of the syllable and is less tightly coordinated and more variable (Prieto & Torreira, 2007). Despite a vast body of research, however, “the details of how segmental anchoring is affected by phonological structure are [still] poorly understood” (Ladd et al., 2009, p. 2). To add to the complexity of segmental anchoring and alignment patterns, the f0 gesture and the timing of tonal targets in relation to segments appear to be language specific and dialect specific. The following section reviews some of the literature on tonal alignment of peaks, with a focus on rising accents.
5.2.3 Cross-linguistic and cross-varietal differences

A number of studies have demonstrated that the same phonological categories of rising accents may show differences in their phonetic alignment (Arvaniti et al., 1998 on Greek; Arvaniti & Garding, 2007 on AmE; Ladd et al., 2009 on BrE; Face, 2006 on Castilian Spanish; Prieto, D’Imperio & Gili-Fivela, 2006 on Spanish, Catalan and Neapolitan Italian; Atterer & Ladd, 2004 on German; Kalaldeh et al., 2009 on Hiberno-English; Nieman, Mücke, Nam, Goldstein & Grice, 2011 on German).

Arvaniti and colleagues (1998) found that, unlike in English (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988), the L and the H targets align beyond the stressed syllable in L*+H accents in Greek, with the L precedes the accented syllable and the H aligns with the post-accented syllable. In Spanish and Neapolitan Italian rising accents (Prieto et al., 2006), both tonal targets are located in the stressed syllable, where the L tone aligns with the onset of the stressed consonant. In Catalan (Prieto et al., 2006) and Castilian Spanish (Face, 2006), the alignment of L and H tones in rising accents is much more complex than originally proposed by Pierrehumbert (1980) for AmE, suggesting the possibility of a three-way contrast. For example, Catalan rises can be produced with a delayed peak, a non-delayed peak and a post-tonic rise. Atterer and Ladd (2004) investigated L+H* prenuclear accents in German and found that the German rising accents aligned later compared to the same category in BrE. The same German speakers recorded in English carried this phenomenon over into their pronunciation of English, showing later alignment patterns possibly as a result of L1 transfer on the precise phonetic alignment. In addition, there were differences in the phonetic timing of peaks between southern and northern varieties of German. The H tone showed more variability in general, but the alignment of the L was consistently later for the southern variety.

Arvaniti and Garding’s (2007) study on varieties of AmE showed differences in the phonetic realisation of rising accents (L+H* and L*+H) in comparison to the same accent categories in Pierrehumbert’s (1980) and Pierrehumbert and Steele’s (1986) data on AmE. Arvaniti and Garding found that in L+H* accents, the L aligned in the accented syllable, contrary to earlier observations of AmE where the L aligned

---

20 See section in §2.2.2 (pp. 28-29) for a more detailed discussion on L1 influence on tonal alignment in L2 speech.
before the stressed syllable. This study also reported that the duration of the rise was shorter than the duration reported in earlier studies on AmE (Pierrehumbert, 1980; Pierrehumbert & Steele, 1986) and were attributed to dialectal differences in AmE. Southern Californian speakers showed a much later H timing of L*+H compared to speakers of English from Minnesota. Ladd and colleagues (2009) also found differences in the alignment patterns depending on BrE variety. The H tone of nuclear and prenuclear accents in Standard Scottish English (SSE) aligned later than in Southern British English or RP. Similarly, in varieties of Hiberno-English, the peaks of L*H accent were aligned earlier in Drogheda English than in Dublin English, but both varieties had a relatively stable peak alignment in nuclear and prenuclear positions. On the other hand, Donegal English showed greater variability with a drift in peak alignment in both prenuclear and nuclear-accented words.

The studies discussed above point out that the alignment of tonal targets is not only affected by various phonological and prosodic contexts, but also shows variation based on a particular language or variety. The source of this variation is a debatable issue in the literature and could be explained in terms of a phonetic continuum in alignment across languages and language varieties (Ladd, 2006) or in terms of innate language- or dialect-specific phonological structures affecting the precise alignment of the putative targets (Nieman et al., 2011).

5.2.4 Pitch range and scaling of tonal targets

As mentioned previously, the use of pitch range and its variation contributes to the information structure and meaning across a number of languages. It is well accepted in the AM framework that edge tones have greater low or high f0 values compared to pitch accents, and phrase accents scale relatively higher or lower compared to boundary tones, reflecting the prosodic hierarchy (Pierrehumbert & Beckman, 1988; Beckman & Ayers-Elam, 1997). However, unlike the research on tonal alignment which shows the importance of the timing of tonal targets in categorical distinction, studies on the use of pitch range and f0 peak scaling have reported conflicting results (see Gussenhoven & Rietveld, 1988; Ladd & Morton, 1997; Dilley & Brown, 2007; Dilley, 2010; Borrás-Comes, Vanrell & Prieto, 2010). For example, Ladd and Morton (1997) in their perception study on the peaks of normal and emphatic accents found evidence of abrupt shifts in identification with an increase in pitch from normal to
emphatic, but there was a lack of evidence in the discrimination of the peak in stimulus pairs. The authors suggested that the distinction between normal and emphatic accents may only be interpreted categorically, but are not categorically perceived. Borràs-Comes, Vanrell and Prieto’s (2010) perception study on Catalan investigated the L+H* accent in three different sentence types and found a categorical difference on the basis of f0 shape associated with L+H* between statement interpretation and echo question interpretation, but the difference between a statement interpretation and a focus interpretation was gradient. They concluded that the scaling of L+H* in Catalan has both a categorical and a gradient effect on interpretation.

Some experimental work has been conducted on the scaling of high and rising pitch accents in English (all on AmE). Arvaniti and Garding (2007) found that the L scaled lower and the H scaled higher in L*+H compared to the low and high targets in L+H*. As a result, the scaling difference between L and H was greater in L*+H. The lower scaling of the L is in line with Pierrehumbert’s (1980) predictions, which suggested that the lower scaling of the L in L*+H could be due to the starred L tone not being affected by the following H tone, as in an L+H* accent where the starred H tones raise the L. Dilley (2010) specifically examined whether pitch range variation plays a role in determining a three-way phonological contrast in English (H*, L+H* and L*+H) and reported continuous variation in pitch range with no evidence of categorical effects for any of the stimulus sets. Moreover, following Ladd and Schepman (2003), Dilley questioned the distinction between the H* and L+H* accents proposed for English on the basis of the lower scaling in L+H* versus mid-pitch scaling of the low target in H*.

Clearly, this is the area of research that needs more attention, especially in relation to NE varieties and IndE, with the latter having more or less preliminary interpretations of pitch accent categories within the AM framework. In the present experiment, the scaling of L and H are part of the analyses. Taking a more traditional approach (Pierrehumbert, 1980; Beckman & Ayers-Elam, 1997), the L scaling is considered to provide additional support for a bitonal versus monotononal pitch accent analysis. As for the peak scaling, there is a possibility that the scaling of the H will not contribute to the distinction of pitch accent categories in prenuclear or nuclear focal syllables, but will simply reflect the focus condition.
5.3 Hypotheses

On the basis of the literature review, a number of hypotheses have been formulated to address the two research questions listed in the introduction to this chapter (§5.1).

Hypothesis 1: Segmental anchoring hypothesis

Based on the tonal alignment research presented in §5.5.2, the present study adopts a weaker view of the segmental anchoring hypothesis (Ladd, 2006; Ladd et al., 2009; D’Imperio, 2012). Hypothesis 1 suggests that the tonal targets of the rising gesture (or gestures) are anchored to certain landmarks in the segmental string but are not of constant slope and duration. The L alignment will show more stability. Segmental duration and composition will have an effect on the location of the peak. Therefore, the peak delay measured from onset to accented syllable or the rhyme and the duration of the rise will correlate with the duration of the accented syllable. The right-hand prosodic context, such as tonal repulsion or tonal crowding, may influence the alignment of the targets, leading to earlier alignment in nuclear accents (Silverman & Pierrehumbert, 1990; Prieto et al., 1995; Ladd et al., 2009). In addition, shorter rises will exhibit higher f0 slopes.

Hypothesis 2: Bitonal pitch accent analysis (L*+H or L+H*)

In the case of a bitonal pitch accent gesture, the distance between the L and H targets is hypothesised to be a relatively stable pitch excursion. The duration of the distance between syllable onset and the L target will correlate with the duration of the distance between syllable onset and the H target, indicating that the two targets move together in synchronicity as one tonal event (Atterer & Ladd, 2004; Ladd, 2006; Frota, 2002). In addition, the L target will scale in the low region of the speakers’ pitch range. According to the AM framework, two analyses will be considered for this hypothesis.

Hypothesis 2A supports an L*+H analysis where the low tone is aligned within the accented syllable and H is a trailing tone aligned with the following syllable. The L will align after the onset of the accented syllable, most likely in the accented vowel (Pierrehumbert, 1980; Pierrehumbert & Steele, 1989; Arvaniti & Garding, 2007 all on AmE).
Hypothesis 2B supports an L+H* analysis. The L will align at the onset of the target syllable, while the H tone will align earlier in relation to the accented syllable offset. It is hypothesised that the peak will align in the accented syllable, with possible differences in the alignment due to vowel length or syllable structure (Prieto et al., 1995; Ladd et al., 2009; Prieto & Torreira, 2007; Hellmuth, 2006), for example, peaks moving into a post-stress consonant when the syllables do not contain a coda consonant.

Hypothesis 3: Low pitch accent and an edge tone analysis (L* (Hp))

Taking into account previous research on Bengali (Hayes & Lahiri, 1991 on Kolkata Bengali; see §2.4.2 for details) and Tamil (Keane, 2014), a language closely related to Kannada, it is possible that the rising gesture is not a bitonal accent but a low pitch accent (L*) followed by a high tone that demarcates the right edge of a prosodic phrase (Hp). In this analysis, it is hypothesised that the initial low target will be anchored to the accented syllable, but the alignment of the H will be a fixed distance from the end of the word. The peak will be realised further away from the accented syllable and will align closer to the edge of the prosodic word with an increase in the number of post-accented syllables.

Hypothesis 4: Simple high accent analysis (H*)

Excluding a clearly delayed rise on nuclear focal words produced by BE speakers, it is feasible that the pitch accent gesture or gestures produced in any of the other prosodic positions is a monotonal accent corresponding to an H* pitch accent. Hypothesis 4 proposes that the speakers produce an H* accent. In this analysis, the L and the H targets will align with certain segmental landmarks, and similar to the rising accent, the alignment of the peak will correlate strongly with the accented syllable duration. However, there will be no relationship between the targets. The peak will move independently of L, as a reflection of the L tone corresponding to a phrase-initial boundary tone in phrase-initial target words. There is also a possibly that H will be realised in the post-tonic consonant, based on previous research showing a common pattern of peak delay in simple high accents in varieties of English (Silverman & Pierrehumbert, 1990; Ladd et al., 2009). In addition to a less synchronised movement between L and H, the scaling of the L target will be
somewhere in a mid level region of the speakers’ pitch range resulting in a shallow rise.

**Hypothesis 5: Potential differences between BE and KE**

Hypothesis 5 suggests that each L1 group has its own inventory of pitch accents and show different categories altogether or those specific to prenuclear and nuclear focal positions. Alternatively, the speakers could produce similar pitch accent phonological inventories but the targets may show differences in the phonetic alignment as a result of L1 influence, given the findings on the phonetics of tonal alignment across languages and dialects.

**Summary table**

Hypotheses 1 to 5 with the corresponding analyses and predictions are summarised in Table 5.1 below.

<table>
<thead>
<tr>
<th>Hypothesis 1</th>
<th>Segmental anchoring hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 2A</td>
<td>L*+H analysis</td>
</tr>
<tr>
<td>Hypothesis 2B</td>
<td>L+H* analysis</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>L* (Hp) analysis</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>H* analysis</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>Different tonal categories for BE and KE</td>
</tr>
</tbody>
</table>

**5.4 Overview of the method**

A full description of the method used in the present experiment is outlined in §3.5 (Chapter 3). It presents the materials, recording procedure, annotations and acoustic analysis. The present section includes a summary of the materials with a focus on target words, measurements and variables for the statistical analyses.

The speech materials used in this experiment consisted of two sets designed to elicit the production of the target words in different prosodic conditions, namely prenuclear and nuclear focal. In order to maintain consistency with the prenuclear accents subset and control for the prosodic environment in narrow focus, it was decided to look at phrase-initial contexts. The first set of read sentences investigated
prenuclear accents and consisted of four sentences with a target word in the initial position. All sentences had the same modality and structure, being simple declaratives consisting of four words. The four words following the target one varied in terms of their phonological composition and structure. The target words included Lara, Lulu, Nelly and Maryellen, all with an open syllable structure. The second set of materials consisted of question-answer pairs to elicit narrow focus on the target tokens. The target words Lulu and Maryellen were inserted in a carrier sentence with a simple declarative structure. Details about segmental composition of the target words can be found in Table 3.5 (§3.5.1). The segmentation of the data was performed following a standard procedure for acoustic-phonetic segmentation and word labelling (Peterson & Lehiste, 1960; Croot & Taylor, 1995).

The EMU annotation for this experiment included the three tiers of word, phonetic, and target. Segments were annotations on the phonetic tier, while the location of the tonal targets and the segmental landmarks were annotated on the target tier. The measurements taken for further analyses are described in detail in §3.5.2. As a reminder to the reader, a schematic representation of these measurements is presented below in Figure 5.2.

![Figure 5.2 Schematic representation of main measurements, shown relative to the f0 movement and the segmental landmarks.](image)

(a) ContoL  
(b) ContoH  
(c) LtoH  
(d) f0 slope  
(e) VontoH  
(f) VofftoH

---

21 As mentioned in §3.5, the target word Maryellen was consistently produced with lexical stress placed on the first syllable.
Timing measures included the temporal interval from syllable onset to L (a) and H (b), duration of the distance between the two tonal targets (c), the distance between the vowel (or rhyme) onset and H (e), the distance between the vowel offset and H (e), and the distance between H and word offset (not shown in the figure). In addition, L and H were examined relative to each other (ContoL and ContoH). As discussed in §3.5.2, all timing measures involving tonal targets relative to various segmental landmarks were analysed as a proportional duration of the target accented syllable (Silverman & Pierrehumbert, 1990; Schepman et al., 2006).

The $f0$ measures included the absolute $f0$ of the L and the H targets as well as the magnitude of the rise corresponding to the $f0$ slope shown in Figure 5.2 (d). Consistent with the analyses presented in the previous chapter, all $f0$ values were converted into semitones relative to the 100 Hz baseline.

A total of 475 target words were included in the statistical analyses. For each of the temporal and scaling measures, LMM procedures were performed to assess the effect of L1 group and segmental duration or composition on the alignment of the targets. In the analysis for each prosodic position, prenuclear or nuclear focal, GROUP and WORD were included as fixed factors and SPEAKER was treated as a random factor. In some of the analyses, additional fixed factors of VOWEL and SYLLABLE duration were added to further investigate segmental effects. In the comparative analysis of nuclear focal versus prenuclear rises, both SPEAKER and WORD were treated as random effect factors. The main aim was to look for potential differences between the groups in two prosodic positions in order to justify any phonological category or categories for the rises proposed in separate analyses.

The main effects and interactions of the LMM procedures are reported in the results section. The full set of statistics on random and fixed effects on each timing or $f0$ measure can be viewed in Appendix C. Two types of box plots are included in the description of the results. These are: trellis plots showing different groups of data points, as well as the effects and interaction of multiple factors; and regular box plots showing graphic summaries of distribution for one or more factors. In both types of plots, the horizontal line or dot in the box represents the median and whiskers indicate the end points of the distribution.
5.5 Results

5.5.1 Prenuclear rises

5.5.1.1 Alignment of L

Figure 5.3 presents the temporal interval from syllable onset to L target (ContoL) as a proportion of the accented syllable duration and is presented by group and by target word. GROUP showed no statistically significant effect on L alignment [t=-0.679, p>0.5]. Mean values of 51 ms [SD=26] were recorded for the BE group, corresponding to a 0.27 ratio mean [SD=0.11]. For the KE group, mean values of 52 ms [SD=19] were reported corresponding to the ratio duration mean of 0.29 [SD=.09]. Variation across BE speakers was mainly due to the productions by B1 and B2 who showed earlier alignment of L in some of the words (Lulu for B1, and Nelly and Maryellen for both B1 and B2).

![Figure 5.3 Box plot for the distribution of proportional ContoL distance as a function of L1 group (BE versus KE), presented by target word.](image)

![Figure 5.4 Mean duration of nasals /m/ and /n/ in milliseconds, presented by group.](image)

Figure 5.3 also shows that there was an interaction between WORD and GROUP. The Tukey test revealed that the difference in the alignment for BE and KE holds in the stressed syllables of the target words Nelly [t=3.232, p<.001] and Maryellen [t=1.969, p<.05], with the latter difference being more modest. This could be a reflection of consonant duration in the target syllable. Despite some variation across BE speakers, the mean duration of the lateral (Lulu and Lara) showed no significant differences, with the mean measuring at around 90 ms [BE: SD=30 ms; KE: SD=11 ms]. Figure 5.4 shows that the duration of both nasals was significantly
shorter for Kannada L1 speakers. Comparison of means revealed significant differences. For /n/, \( t=2.3354 \) with \( p<.05 \), while for /m/, \( t=2.6516, p<.05 \).

Taking into account that the L target location is the basis of the contrast between L+H* or H* and L*+H in English, L alignment was also examined relative to the onset of the rhyme (VontoL distance) to investigate whether the low target was anchored to the accented consonant or accented vowel. The durational values of the VontoL distance as a proportion of the accented syllable duration are presented in Figure 5.5 below.

As shown in Figure 5.5, the L appeared to be anchored to the consonant of the accented syllable for all speakers, regardless of L1 group. The L target was usually located before the onset of the accented vowel, represented by the 0.0 mark on the y-axis. The number of outliers realised in the accented vowel (above the 0.0 mark in the box plot) was very small. BE speakers showed a greater degree of variation for most target words. The LMM analyses mirrored the results reported for the temporal interval measured from syllable onset to L (ContoL). There was no significant difference between the L1 groups [\( t=-0.7, p>.05 \)], but there was a strong interaction between GROUP and WORD, reflective of the later L location in the words *Maryellen* [\( t=1.969, p<.005 \)] and *Nelly* [\( t=3.232, p<.001 \)] for KE speakers. This is consistent with the ContoL alignment results.
5.5.1.2 Alignment of H

Initially, three types of the measurements were considered when investigating the peak delay of the rising gesture. The H was measured:

a) relative to the accented vowel or rhyme onset (following Silverman & Pierrehumbert, 1990)

b) relative to the accented consonant onset (Caspers & van Heuven, 1993; Prieto et al., 1995; Arvaniti et al., 1998)

c) relative to the accented vowel offset (Arvaniti & Garding, 2007).

The results of peak delay in relation to the syllable onset (b) and the rhyme onset (a) yielded similar results. The measurement of the peak delay relative to the vowel offset (c) was also not relevant and somewhat redundant. Taking into account that the L was consistently realised in the accented consonant by speakers from both groups and that Prieto et al. (1995) reported a correlation between onset duration and peak delay in their examination of tonal alignment in Spanish, suggestive of a more reliable measure, this section includes the results of peak delay calculated relative to the syllable onset (b), thus excluding the other two measures.

Unlike the L alignment, the alignment of the peak (ContoH) showed considerable variation in the productions across speakers and is therefore presented by group and by speaker. Figures 5.6 and 5.7 illustrate the ContoH distance as a proportion of the accented syllable duration for KE and BE speakers respectively, where 1.0 on the y-axis indicates the end of the accented syllable. The values below

![Figure 5.6 Box plot for the distribution of proportional peak delay (ContoH) as a function of four KE speakers, presented by word.](image1)

![Figure 5.7 Box plot for the distribution of proportional peak delay (ContoH) as a function of four BE speakers, presented by word.](image2)
1.0 show the peaks within the accented vowel and the values higher than 1.0 indicate that the H peak is located in the post-accented syllable.

As shown, the H was mostly located after the accented vowel offset. KE speakers were more consistent across different target words, especially in Lara. In most of the repetitions of Lulu produced by speaker K2 and some of the repetitions by K3, the peak was realised towards the end of the accented vowel and is shown close to 1.0 in Figure 5.6. The BE group showed much greater variability across the speakers and repetitions (Figure 5.7). For example, for speaker B1, the peak delay in Maryellen was more delayed compared to the peaks in the rest of the target words. Speaker B2 produced earlier peaks in the words Nelly and Maryellen, and B4 showed variability across most of the tokens.

Taking into account the inter-speaker variation shown in the figures above, analysis confirmed that GROUP had no effect on the location of the peak \([t=0.933, p>0.5]\). On average, peaks were realised at the offset of the post-accented consonant. In a small number of repetitions, the H target moved into the post-tonic vowel. There was an effect of WORD, with further investigation revealing a significant effect for the word Lulu \([t=-2.815, p<0.005]\) showing earlier peaks, especially in the KE group. The H timing in the target word Lara was the most consistent, most likely as a result of the consistently long vowel in this target word. Further investigation found a strong effect of VOWEL duration \([t=-3.439, p<0.0006]\) on H alignment, showing later peaks in Maryellen and Nelly, with the exclusion of speaker B2. Examination of the duration of the target vowels (Figure 5.8) reported that only KE speakers maintained a clear contrast between long and short vowels.

As shown in Figure 5.8 below, Kannada L1 speakers produced a much longer vowel in Lulu, with some variation across speakers \([\text{mean}=138, \text{SD}=33]\). This may have contributed to the location of earlier peaks when measured as a proportion of the accented syllable, as well as a more consistent overall peak delay for this group. The Bengali L1 group produced a short vowel \([\text{u}]^{22}\) in Lulu \([\text{mean}=81 \text{ ms}, \text{SD}=12]\) and made no durational distinction between \([\text{u}]\) and /æ/ as well as \([\text{u}]\) and /ɛ/ (where pairwise comparisons did not reach statistical significance with \(p>.05\)). Only /ɑː/ was

---

22 The assumption that BE speakers produced a short variant of the [u] vowel and not [ʊ] is based on the F1 and F2 analysis in §4.5.3.4.
produced consistently as a long vowel. This finding mirrors the result of Experiment 1 presented in §4.5.3, emphasising the lack of durational contrast for some vowels in BE.

![Figure 5.8 Mean duration of the target vowels in milliseconds produced by BE (blue dotted line) and KE (green solid line) speakers. The labels on the x-axis follow standard phonetic-acoustic labelling criteria (Croot & Taylor, 1995) and correspond to the following vowels: A - /æ/, a: - /ɑː/, E - /ɛ/, u(:) - /u(:)/.

The examination of the interval between the peak and the end of the word did not support the possibility of H being a phrase tone as proposed in Hypothesis 3. The peak location was relatively stable. An increase in the number of post-accented syllables did not lead to the peak moving toward the word offset, thus confirming that the peak was part of a pitch accent gesture. This held for all speakers. There was no effect of GROUP [t=0.348, p>0.5], nor any interaction between WORD and GROUP [p>0.5]. The WORD effect of Maryellen was significant [t=25.128, p<.0001]. Tukey tests for multiple comparisons of means revealed differences between Maryellen and the other three words [Maryellen-Lara: z=25.13, p<.001; Maryellen-Lulu: z=22.07, p<.001; Maryellen-Nelly: z=23, p<.001]. Figure 5.9 below shows the interval between H and the end of the word (HtoWoff) as a function of word duration.

The interval HtoWoff was longer with the increase in the number of post-accented syllables [BE: R²=0.92, p<.0001; KE: R²=0.89, p<.0001], thus indicating that H is not a phrase accent but is part of the pitch accent gesture. In addition, peak delay correlated with the duration of the accented syllable for both groups (Figure 5.10 below), [BE: R²=0.71, p<.0001; KE: R²=0.87, p<.0001], supporting Hypothesis...
1 on segmental anchoring. Higher R-squared values for KE speakers could be explained by greater consistency in peak timing for this group.

Figure 5.9 Scatter plot for mean peak relative to word offset and mean accented syllable duration (HtoWoff), both in milliseconds.

Figure 5.10 Scatter plot for mean peak delay (ContoH) and mean accented syllable duration, both in milliseconds, presented by L1 group.

5.5.1.3 Alignment of the rising gesture

Figure 5.11 below illustrates the duration of the rising gesture (LtoH) as a proportion of the accented syllable duration. The temporal interval between the tonal targets was not affected by GROUP \[ t=0.766, p>.05 \], but was affected by WORD \[ Maryellen: t=2.13, p<.04; Lulu: t=-1.89, p<.05 \], with no interaction between GROUP and WORD.

Figure 5.11 Box plot for the distribution of proportional rise duration as a function of L1 group (BE versus KE) by target word.

Figure 5.12 Scatter plot for mean peak delay (ContoH) and mean low delay (ContoL) in milliseconds, presented by L1 group.
The main differences were reported between the pairs Maryellen-Lulu \([z=3.893, p<.001]\) and Nelly-Lulu \([z=-3.056, p<.02]\), indicating shorter rises for the target word Lulu, but with a greater amount of variation across KE speakers. In BE, the greater variation in rise duration found in the words Maryellen and Nelly was due to the production of these tokens by speaker B2, who consistently showed earlier H alignment leading to a shorter duration of the rise.

For both groups, the duration of the rise correlated highly with the duration of the accented syllable \([\text{BE}: R^2=0.75, p<.0001; \text{KE}: R^2=0.89, p<.0001]\). These correlations can be seen in the scatter plot in Figure 5.12 above. This finding is in line with the segmental anchoring hypothesis, which accepts the notion that the tonal gesture is not of constant duration but is affected by phonological structure, namely the duration of the segmental material (Silverman & Pierrehumbert, 1990; Prieto et al., 1995; Arvaniti et al., 1998; Ladd et al., 2000; Prieto & Torreira 2007; Ladd et al., 2009). A stronger correlation between peak delay and syllable duration for the KE group simply shows that the alignment of the rising gesture is timed more precisely for KE speakers, as found for peak timing in the previous section.

The interval from syllable onset to L target (ContoL) and syllable onset to H target (ContoH) investigated a possible relationship between the two targets. The correlation was significant, suggesting that Hypothesis 2 is true and the speakers produced a bitonal pitch accent. Despite greater variability in H alignment, a higher R-squared value was reported for the BE group \([R^2=0.66: p<.0001; \text{KE}: R^2=0.46, p<.001]\).

5.5.1.4 Scaling of tonal targets

The scaling of the L tone targets that were part of the LH movements were investigated in order to test Hypotheses 2 and 4. Figure 5.13 below shows the results of L scaling in prenuclear position. The y-axis may show negative semitones values due to 100 Hz being a baseline for Hz-semitone conversion.

The analysis showed no effect of GROUP on the scaling of the L target \([t=-0.279, p>.05]\). There was a strong effect of WORD \([t=-5.456, p<.001]\), revealing the differences between Lulu and the words Nelly and Maryellen. Tukey tests on pairwise comparisons of means confirmed significant differences for Nelly-Lulu \([z=3.056, p<.01]\) and Maryellen-Lulu \([z=3.893, p<.001]\), indicating lower f0 values for the word Lulu.
There was no interaction between GROUP and WORD. Instead, a large degree of inter-speaker variation was observed in the results of L scaling, reflecting intrinsic properties of pitch range use. For five speakers (B1, B2, B3, K3, K4), the L target mean values were somewhere between 2 and 5 semitones (120-135 Hz), with speaker B1 showing the most variation. Two speakers produced the highest L values with the mean of 8.5 semitones for K2 and 7.8 semitones for B4.

Taking into account such a degree of variation, an examination of L scaling in relation to the speakers’ individual pitch ranges was performed. Pitch range was measured as the highest and the lowest f0 points in the f0 across all experiments of this study. This enabled the establishment of pitch range differences more accurately, as it was based on a larger sample of data. The mean low f0, mean L scaling in prenuclear accents and mean high f0 values are presented for each speaker in Figure 5.14 below. The bottom (dashed) and top (dashed) black lines correspond to the high and low ends of the pitch range. The blue solid line in the middle represents L scaling.

As shown in the figure, the L tones produced by speakers B1, B2, B3, K1, K3 and K4 scale below 5 semitones. For speakers B4 and K2, L scaling in the target words is higher but these speakers have an overall higher baseline (lowest f0). On the whole, L targets are located low relative to the speakers’ baseline of low f0, suggesting that L scales in the lower region of the speakers’ pitch range. This is indicative of Hypothesis 2, a bitonal pitch accent analysis.
Figure 5.14 Mean $f0$ for L scaling (blue solid line) relative to the highest $f0$ (top black dashed line) and the lowest $f0$ (bottom black dashed line) presented by speaker.

Similar to L scaling, GROUP had no effect on H scaling [$t=-1.009$, $p>.05$] and showed a large degree of variation across the speakers. Speakers B1, B4 and K2 displayed the highest $f0$ values for the peak, ranging from 8 to 15 semitones (180-270 Hz). The lowest values were reported for speakers K1 and B3.

Figure 5.15 Box plot for the distribution of mean H scaling in semitones, presented by word for each group.

As shown in Figure 5.15, WORD had a significant effect on the scaling of high targets [$t=6.3$, $p<.0001$], most likely due to intrinsic vowel differences and their effect on $f0$. The main differences were observed in the pairs Lara-Lulu [$z=3.783$, $p<.001$], Lara-Nelly [$z=6.195$, $p<.001$], Lara-Maryellen [$z=7.030$, $p<.001$], and Lulu-Maryellen [$z=3.437$, $p<.004$].
The scaling difference between L and H (i.e. rise magnitude) revealed a better picture and showed a more consistent pattern than L and H scaling alone. There was no effect of GROUP or WORD, but the difference in the rise magnitude depending on L1 group approached significance [t=−1.911, p<.056]. These results are illustrated in Figure 5.16. The lack of significance in LMM results may be due to the relatively shallow rises (around 6 semitones) produced by speaker B3 in a proportion of repetitions. Nevertheless, the f0 results show a clear pattern for KE speakers to produce shallower rises compared to the BE group and is consistent with the findings reported in the previous chapter (§4.5.3.1).

Figure 5.16 Box plot for the distribution of rise magnitude in semitones, presented by word for each group.

WORD showed a strong effect on rise magnitude [t=−3.766, p<.001]. The Tukey contrasts showed that scaling of the rise was most different in the pairs Lara-Maryellen [z=3.766, p<.001] and Lulu-Maryellen [z=3.671, p<.002], with the steepest rise in Lara, echoing the results of H scaling. Apart from B4, most speakers produced a more consistent, and shallower, rise on the target word Maryellen, confirming the effect of word length on the magnitude of the rise reported in the literature (see §5.2). The longer the word, the smaller the f0 slope and vice versa. As mentioned previously, all speakers placed primary stress on the first syllable in Maryellen.

The relationship between the duration of the rise and the scaling of the rising gesture for each group was also examined and is shown in Figure 5.17 below. A correlation between rise duration (L to H interval) and rise magnitude (f0 slope) was
found for BE speakers [\(R^2=0.36, p<.01\)], as well as for the KE group [\(R^2=0.47, p<.001\)]. KE speakers produced more consistent rises. For both groups, the rising gesture was not of constant slope and duration, supporting Hypothesis 1 on segmental anchoring and indicating that shorter rises are most likely to be produced with greater rise magnitude.

Figure 5.17 Scatter plot for mean rise duration and mean rise magnitude by L1 group.

5.5.1.5 Summary

Reminiscent of the studies on tonal alignment and proving Hypothesis 1 on segmental anchoring, the alignment of L was stable and was timed with the onset of the accented syllable. In spite of no significant differences in L alignment between the two L1 groups, the interval between syllable onset and L target was shorter in syllables containing nasals in KE. The speakers of this group produced /m/ and /n/ consistently shorter compared to the same target consonants produced by L1 speakers of Bengali.

Several H tone alignment measures were examined. It was discovered that first, the findings based on the results disprove Hypothesis 3, excluding the possibility of the L* (Hp) analysis. If the hypothesis were true, the location of the H peak would have moved closer to the word offset with the increased number of post-accented syllables. Second, H alignment relative to the accented syllable onset, as well the temporal interval between L and H, showed that the duration of peak delay and the rising gesture correlated with the accented syllable duration for all speakers, where the location of the peak was affected by the durational properties of the segmental material. In addition, the pitch excursion of the accent gesture was not of fixed slope or duration, consistent with segmental anchoring hypothesis (Hypothesis 1).
Third, H alignment did not show a statistically significant difference between the two L1 groups and, similar to the L alignment, disproved Hypothesis 5 which suggested potential differences in the alignment of the targets. The findings, however, showed that KE speakers were more consistent in peak alignment. The H aligned right in the post-accented consonant for all KE speakers, with a small proportion of the tokens showing H alignment at the right edge of the accented syllable. The H was anchored to the post-accented consonant with short vowels, and was timed at the end of the accented vowel when the vowels were long. This pattern is consistent with the alignment of H* and L+H* accents in a number of English varieties, as well as in other languages. BE speakers, to the contrary, showed a large degree of variation across repetitions and speakers. The location of H in target words included earlier peaks timed within the accented vowel, as well as later peaks timed in the post-accented vowel. This may have contributed to a weaker correlation between peak delay and syllable duration compared to KE. There was also a lot of variability in the realisation of vowel durations by BE speakers.

The examination of the relationship between L and H gave support to a bitonal pitch accent analysis (Hypothesis 2). The correlation between the temporal interval from syllable onset to L and from syllable onset to H was significant for both groups, indicating that the targets are part of the same rising gesture. L scaling results further supported the analysis proposed in Hypothesis 2. On average, the L target scaled in the lower part of the speakers’ pitch range. These findings still leave two possible analyses for BE and KE, as either L*+H or L+H*.

Based on more consistent H alignment patterns and the effect of vowel length in open syllables, an L+H* interpretation is considered for KE. In the target words with short vowels (such as /e/ in Nelly), the H was timed outside the accented syllable. However, in the words with a significantly longer vowel (Lara and Lulu), the H was timed just before the accented syllable offset. A similar alignment pattern has been observed for the rising accent in AusE (Fletcher, 2004), Egyptian Arabic (Hellmuth, 2006), Dutch (Ladd et al., 2000), Peninsular Spanish (Prieto & Torreira, 2007) and Neapolitan Italian (D’Imperio, 2000). There is also a possibility that a left-

---

23 As mentioned previously (§2.1.2), not all of the literature agrees on an L+H* and H* distinction (see Ladd, 1996, 2008; Dilley, 2009).
hand phrase boundary had an effect on H alignment, leading to somewhat later realisations of peak.

For BE, a rather unstable anchoring of the peak makes it more difficult to posit a pitch accent category. On the one hand, the rising gesture could be interpreted as an L*+H pitch accent analysis, as proposed in Hypothesis 2A and based on the late alignment of H. On the other hand, earlier valleys and a proportion of earlier peaks would be better fitted to the L+H* accent analysis proposed in Hypothesis 2B. A possible interpretation is that BE speakers do not necessarily produce one single category in the prenuclear position, but use two types of pitch accents, L+H* and L*+H. This interpretation of the results would also align with previous work on English (Pierrehumbert & Beckman, 1988; Arvaniti & Garding, 2007) and ToBI annotations for English intonation (Beckman & Ayers-Elam, 1997).

The scaling results showed no significant differences between the groups. There was some variation in H scaling across the words, perhaps as a reflection of the intrinsic qualities of vowels, as well as across the speakers which reflected individual differences in the use of pitch range. Similar to the findings of the previous chapter, KE speakers produced shallower rises compared to BE speakers, which could be specific to KE and possibly linked based on the degree of pitch modulation in their L1.

Further investigation of nuclear focal rises (§5.5.3) and the comparison of the rises in two prosodic positions (§5.2.3) will expand on the findings presented in this section and will re-evaluate the proposed pitch accent categories.

5.5.2 Nuclear focal rises results

5.5.2.1 Alignment of L

The examination of the L target relative to the rhyme onset (VontoL) confirmed that, similar to prenuclear accents, L aligned within the stressed consonant and not within the vowel. Only speaker B3 was an exception and produced several repetitions where L aligned at the beginning of the stressed vowel. Therefore, the results presented in this section include L alignment patterns measured from the onset of the accented syllable (ContoL).

WORD had no effect on L alignment \[t=0.419, p>.05\]. Similarly, no differences in L alignment were reported based on L1 GROUP \[t=-0.536, p>.05\], with
some variation across the speakers. As a result, the values for L alignment are presented by speaker (x-axis) and not by group. Figure 5.18 illustrates the distance from syllable onset to L as a proportion of the accented syllable for each speaker.

![Figure 5.18 Box plot for the distribution of proportional mean ContoL distance presented by speaker.](image)

Depending on the speaker, L aligned at 0.18 to 0.38 ratio duration of the stressed syllable, or at 20 to 60 ms interval from syllable onset. The words produced by speaker B4 displayed the latest mean alignment of L, and speaker K1 produced a later peak compared to K2, K3 and K4 which could have been due to the durational differences in the production of the sonorants across speakers. The mean values in milliseconds for /l/ and /m/ for each speaker are shown in Figure 5.19 below.

![Figure 5.19 Mean duration of /l/ and /m/ in milliseconds presented by speaker.](image)
These two consonants had the longest duration for speakers K1 and B4 who also produced later L alignment. A correlation test between ContoL distance and the duration of the stressed consonant confirmed a strong relationship for all speakers \( R^2 = 0.73; p < .0001 \). This finding indicates that L is anchored to the consonant for both groups and explains the later L alignment for some of the speakers.

5.5.2.2 Alignment of H

Figure 5.20 displays peak delay relative to syllable onset (ContoH) in focally accented syllables of two target words as a proportion of the stressed syllable. There was no effect of WORD and a lack of interaction between GROUP and WORD. Although, unlike the alignment patterns reported for prenuclear rises (§5.5.1), GROUP had a strong effect on the alignment of the peak \( t = -3.115, p < .002 \), showing significantly earlier alignment for the KE group. On average, the interval ContoH measured at 1.38 ratio of the stressed syllable [238 ms, SD=46] for the BE group, indicating that the H target was anchored well into the post-tonic rhyme. Mean peak delay for the KE group measured at 1.2 ratio of the stressed syllable [175 ms, SD=34 ms], and was realised in the post-tonic consonant, similar to peak delay in the prenuclear tonal gesture. Speaker K2 consistently produced the earliest peaks. A number of outliers for the KE group corresponds to some of the repetitions produced by speaker K1, who produced a small number of the repetitions with a later peak.

![Box plot for the distribution of proportional mean peak delay as a function of L1 group (BE versus KE).](image)

Figure 5.20 Box plot for the distribution of proportional mean peak delay as a function of L1 group (BE versus KE).
The LMM analyses also showed that H alignment was affected by the duration of the stressed VOWEL \([t=-5.890, p<.0001]\), indicative of the segmental effects on peak delay. As shown in Figure 5.21 below, H alignment strongly correlated with the length of the stressed syllable for both language groups [BE: \(R^2=0.87, p<.0001\); KE: \(R^2=0.87, p<.0001\)], supporting the segmental anchoring hypothesis.

![Figure 5.21 Scatter plot of mean peak delay and accented syllable duration, both in milliseconds, presented by L1 group.](image)

In addition, there is ample evidence that H is part of a pitch accent gesture and does not demarcate a minor phrase. The distance from the peak to word offset was affected by WORD \([t=53.42, p<.000]\), indicating a greater distance of H to word offset for Maryellen, with no interaction between GROUP and WORD. These results do not support Hypothesis 3 that states that H is a phrase tone. The number of syllables in Maryellen did not lead to H moving toward the end of the word edge. As anticipated, GROUP showed a significant effect \([t=3.057, p<.002]\) on H alignment relative to word offset, mirroring the result of H alignment relative to syllable onset reported above due to KE speakers producing significantly earlier peaks.

### 5.5.2.3 Alignment of the rising gesture

Analyses of the temporal interval between L and H (LtoH distance) are similar to the analyses presented in the preceding section on peak delay. Rise duration was affected by the GROUP factor \([t=-5.243, p<.001]\) and unaffected by WORD. There was also a strong effect of accented syllable duration \([t=-9.742, p<.000]\). The mean proportional durations of the rise are presented in Figure 5.22. On average, BE speakers produced
a rise of 1.1 ratio of the stressed syllable duration or a mean of 191 ms [SD=33],
while the rise produced by KE speakers was shorter than the duration of the stressed
syllable with a ratio of 0.97 [155 ms, SD=32].

Figure 5.22 Box plot for the distribution of proportional mean rise duration as a
function of L1 group (BE versus KE) grouped by target word.

Consistent with the rest of the results and in line with the segmental anchoring
hypothesis, rise duration correlated with the duration of the accented syllable [BE:
$R^2=0.62$, $p<0.0001$; KE: $R^2=0.71$, $p<0.0001$]. The correlations were somewhat weaker
compared to the correlation between accented syllable duration and peak delay
(shown in Figure 5.21) but were significant.

In order to further examine the rising gesture as a whole, a correlation was run
between the L target relative to the syllable onset interval (ContoL) and peak delay
(ContoH) (Figure 5.23 below). The correlation was significant for both groups, giving
further evidence that Hypothesis 2 is true and the focal rise is a bitonal pitch accent.
For the BE group, there was a much stronger correlation between the two intervals
[BE $R^2=0.82$, $p<0.0001$; KE $R^2=0.51$, $p<0.0001$], suggesting a more synchronised
movement of the tonal targets for this group, also in comparison to the rise in
prenuclear position.

Similar to the results reported in the section on prenuclear rises (§5.5.1), there
was a correlation between rise magnitude and rise duration, in support of Hypothesis
1 that the pitch accent gesture is not of constant slope and duration. Figure 5.24 below
illustrates this correlation for each group [BE: $R^2=0.34$, $p<0.001$; KE: $R^2=0.40$,
p<.001]. In addition, the figure shows that BE speakers produced higher rises on focal words, consistent with the results on f0 excursion presented in Chapter 4.

Figure 5.23 Scatter plot of mean peak delay (ContoH) and mean low delay (ContoL) distance, both in milliseconds, presented by group.

Figure 5.24 Scatter plot of mean rise duration in milliseconds and mean rise magnitude in semitones, presented by group.

In order to avoid repetition in the results presentation, L and H scaling as well as the scaling of the rising gesture in nuclear focal rises will be included in the comparison of prenuclear and nuclear focal accents (§5.5.3).

5.5.2.4 Summary
The results of the alignment of the rising gesture presented above mirror some of the results reported for prenuclear accents in §5.5.1. For both groups:

a) the tonal alignment was affected by the duration of segmental material (Hypothesis 1)

b) the L target was anchored to the accented consonant

c) the H did not align with the edge of the prosodic word (null Hypothesis 3)

d) the two targets displayed a coupling indicating a relatively synchronised movement (Hypothesis 2; null Hypothesis 4).

Differences between the L1 groups were found in the alignment of the peak and the duration of the rise. In BE, the H consistently aligned within the post-tonic vowel, usually at its midpoint, with little variation across the speakers. The most appropriate pitch accent category according to the AM framework is L*+H, in line with the analysis proposed in Hypothesis 2A. A stronger correlation between the
ContoL and ContoH interval for this group could be additional evidence towards the L*+H accent, where H is a trailing tone (Pierrehumbert, 1980; Grice et al., 1995).

For KE speakers, the results of H alignment showed that the peak was timed within the post-tonic consonant or just before the end of the accented vowel. This pattern is rather similar to that produced in prenuclear contexts. Moreover, vowel length had a strong effect on H alignment. In CV structure (Maryellen), it was timed significantly earlier compared to the CVV structure (Lulu). Therefore, two hypotheses are considered for KE speakers: an L+H* pitch accent (Hypothesis 2B) and an H* accent (Hypothesis 3). In order to support one of these hypotheses, the scaling of the L target will be examined in the next section. In addition, the rising gesture will be examined against the rising gesture in prenuclear position.

Finally, the differences in H alignment are indicative of phonological and not gradient differences between the speakers, giving proof to Hypothesis 5 that predicted the speakers to exhibit phonetic and/or phonological differences based on their L1.

5.5.3 Comparison of nuclear focal and prenuclear rises

This section presents a comparison of the accent gesture in prenuclear and nuclear focal prosodic positions. For this set of results, the effects of SPEAKER and WORD were treated as random factors.

5.5.3.1 Alignment of tonal targets

Figure 5.25 shows a comparison of ContoL as a proportion of the accented syllable duration presented for each group by the prosodic position: prenuclear (Pre) and nuclear focal (NF). As can be seen, the alignment of L relative to syllable onset showed no differences between the groups [t=-0.753, p>.05]. Despite no overall effect of CONDITION on L alignment [t=-0.259, p>.05], there was an interaction between GROUP and CONDITION [t=3.193, p<.002], where earlier alignment was found in prenuclear rises produced by the KE speakers. The valleys in nuclear focal accents aligned earlier for this group (17 ms), most likely as a result of segmental duration differences. Despite a modest difference, the L target was still timed within the stressed consonant for all speakers.

As illustrated in Figure 5.26 below, both GROUP [t=-2.601, p<.01] and CONDITION affected the H peak alignment [t=-6.347, p<.001]. Moreover, the
alignment of H showed a strong interaction between GROUP and CONDITION \([t=8.138, p<.001]\). Peaks in nuclear focal words were realised later by BE speakers, on average 30 ms longer. Peak delay in BE prenuclear accents, on the other hand, was similar to the peak delay produced by KE speakers in both nuclear focal and prenuclear accents. A follow up Tukey test confirmed the lack of significant difference in peak alignment between nuclear and prenuclear accents for the KE group \([p>.05]\).

As can be seen from Figure 5.27 below, nuclear focal syllables were lengthened, regardless of the group. The analyses confirmed the effect of CONDITION \([t=2.752, p<.03]\), indicating longer syllables in nuclear focal rises which is consistent with the lengthening of segmental domain in narrow focus reported in Chapter 4. No effect of GROUP \([t=-0.934, p>.05]\) nor any interaction between GROUP and CONDITION \([t=1.219, p>.05]\) were found.

Figure 5.28 illustrates the duration of L to H temporal interval. The LMM analyses confirmed that similar to peak delay results (§5.5.2.2), the duration of the rise was affected by GROUP \([t=-2.114, p<.04]\) and CONDITION \([t=-5.103, p<.001]\), with no interaction between GROUP and CONDITION \([t=5.469, p<.001]\). The rises in narrow focused words produced by BE speakers were significantly longer than the rises in prenuclear position produced by this group, as well as being longer than prenuclear and nuclear rises produced by KE speakers. Later peak alignment in focal
positions supports Hypothesis 2A for the BE group (L*+H). The lack of difference for KE speakers suggests that they may be producing the same phonological category in prenuclear and nuclear rises.

5.5.3.2 Scaling of tonal targets

Figure 5.29 below includes the scaling of L presented by group and prosodic position. Blue boxes correspond to the nuclear focal rises and green boxes correspond to the prenuclear rises. The LMM analyses showed that GROUP had no effect on the scaling of L [p>.05]. Moreover, there was a large degree of variation across the speakers, echoing the results on prenuclear accents (§5.5.1.4). Interaction between GROUP and CONDITION was observed for KE speakers [t=4.291, p<.001], however, it simply reflected L scaling differences in two prosodic positions without showing any clear patterns. Most importantly, the lack of any differences in L scaling between two prosodic positions showed that the L target scaled in the lower part of the speakers’ pitch range, consistent with the bitonal pitch analysis proposed in Hypothesis 2.

Figure 5.30 illustrates the scaling of the H target in two prosodic positions. The results reported no effect of GROUP on H scaling [t=−1.454, p>0.14], more likely due to the differences in pitch range across speakers. H scaling was affected by CONDITION [t=−5.511, p<.001], but this was only the case for BE speakers. An interaction between GROUP and CONDITION confirmed higher H scaling in nuclear focal rises [t=4.645, p<.001] for the BE group.
The scaling difference between L and H, or the rise magnitude, showed a strong effect of GROUP \([t=-2.99, p>.003]\) and CONDITION \([t=-9.701, <.0001]\). The interaction between GROUP and CONDITION confirmed that BE speakers produced higher rises in nuclear focal accents \([t=7.725, p<.001]\). These effects and the interaction of the effects are illustrated in Figure 5.31 below.
5.5.3.3 Summary

The results presented the comparison of the rises in nuclear focal versus prenuclear positions and showed differences between the two groups in the realisation of nuclear focal rises. While L alignment was anchored to the accented consonant for all speakers regardless of the prosodic position, the alignment of H showed significant delay in BE nuclear focal rises and was timed within the post-accented vowel. Consequently, the duration of the rise was significantly longer in narrow focus for this group. This finding supports Hypothesis 2A: BE speakers produce an L*+H pitch accent on focally accented words. Similar to the findings on f0 manipulation in narrow focus structure (§4.5.3.1), the H tone on focally accented syllables scaled higher for BE speakers, reflecting the use of pitch range as a strategy to mark narrow focus as well as a greater magnitude of the rises for this group.

As for the prenuclear rises, which showed a great deal of variation in peak alignment, §5.5.1.5 proposed L+H* and L*+H (Hypotheses 2A and 2B) as two types of accents in BE that are used interchangeably. This seems to be the most feasible explanation at this stage. The comparison of H alignment and the proportional duration of the rise in nuclear focal and prenuclear accents further support this analysis. Both types of rising pitch accents will be closely looked at in the description of the tone inventory (§7.5.2 of Chapter 7).

In KE, the alignment of H, along with the alignment of the rising gesture as a whole and the f0 scaling results showed little difference between the two prosodic conditions. This suggests the same phonological category for these speakers in prenuclear and nuclear focal contexts. Based on previous research on English L+H* accents, the analysis posited a rising L+H* accent in KE on prenuclear and nuclear words (Hypothesis 2B). This finding points out that despite some variability in the phonetic realisation of rises, English spoken by Kannada L1 speakers does not include a categorical distinction between L+H* and L*+H accents.

The results of L and H scaling displayed somewhat puzzling results for the KE group. Focus showed no effect on f0. This is in contrast to the findings of Experiment 1 (§4.5.3.1), which showed that two speakers produced higher peaks and all speakers produced higher rises on the focally accented syllables. One of the possible explanations for the lack of f0 differences in prenuclear and nuclear focal contexts could be the sentence-initial position of the target words. Prenuclear accented tokens in Experiment 1 included all pitch shapes, while prenuclear accents in this experiment...
were limited to rising \( f_0 \) gestures, excluding pitch movements realised as a sustained high pitch.

A summary table of the proposed pitch accent categories is included below (Table 5.2). Prenuclear pitch accents may not be limited to \( L+H^* \) in KE or \( L^*+H \) and \( L+H^* \) in BE, and will be further discussed in the tone inventory (§7.5.2) together with the possible pitch accents in nuclear non-focal contexts.

Table 5.2 Rising pitch accented categories presented for each group by prosodic position.

<table>
<thead>
<tr>
<th>Group</th>
<th>Prosodic Position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prenuclear</td>
</tr>
<tr>
<td>BE</td>
<td>( L+H^* ) or ( L^*+H )</td>
</tr>
<tr>
<td>KE</td>
<td>( L+H^* )</td>
</tr>
</tbody>
</table>

5.6 Discussion

The experiment described in this chapter has presented the alignment and scaling patterns of the rising gesture produced by IndE speakers in prenuclear and nuclear focal positions. In line with the literature on tonal alignment and segmental anchoring (Arvaniti et al., 1998; Ladd et al., 2000; Ladd, 2006; D’Imperio, 2012), the findings provided evidence that the rising gesture aligned with the segmental landmarks and was affected by the segmental composition and duration of the accented syllable. Consistent with the segmental anchoring hypothesis, the accent gesture was not of fixed slope or duration, thus supporting Hypothesis 1. In the investigated rises, correlations between the duration of the rise and the magnitude of the rise were significant for prenuclear and nuclear focal accents produced by both groups. A much smaller correlation between \( L \) to \( H \) distance and \( f_0 \) slope (\( R^2=0.17, p<.03 \)) was reported for nuclear focal accents in BE speakers, possibly reflecting a distinctive phonetic realisation of \( L^*+H \) in this variety.

The valleys were sensitive to the length of the accentuated consonant but nonetheless showed consistent anchoring to that consonant. A stable alignment of the first tonal target in bitonal accents has been reported in cross-linguistic research on tonal alignment (such as an \( L \) tone in rising accents in Prieto et al., 1995 on Mexican Spanish; Arvaniti et al., 1998 on Greek; Atterer & Ladd, 2004 on German; Prieto & Torreira, 2007 on Spanish; or an \( H \) tone in falling accents in Frota, 2002 on European
Portuguese; Prieto, 2009 on Catalan). There were a few limitations in Experiment 2 that were justified in the methodology chapter (§3.5). These were first that the target words were all phrase-initial with stress on the first syllable, and second that the makeup of the stressed syllables was limited to sonorants. It is possible that the L target may show differences in alignment in other prosodic (sentence-medially or following another accent) and phonological (fricatives or stops) contexts. For example, Sullivan (2007) found that in rising accents of Belfast English, the valleys were affected by anacrusis, sentence type, segmental structure and speech style. This is similar to the findings generally reported for the behaviour of the H target across a number of languages (as argued in §5.2.2).

The evidence presented in Experiment 2 also showed that consistent with previous research, the H target had greater variability in alignment affected by segmental duration. In addition, accented syllable onset was shown to be a reliable measure for peak delay, giving proof that onset together with rhyme contribute to peak timing (Prieto et al., 1995; Arvaniti et al., 1998, Prieto & Torreira, 2007). The timing of the prenuclear and nuclear peaks produced by KE speakers was consistent, but subject to vowel length in syllables with a long vowel showing earlier peaks. Ladd and colleagues (2009) reported a similar phenomenon for prenuclear peaks in BrE. A number of studies point out that syllable structure rather than vowel duration (tense versus lax) alone contribute to peak alignment (Hellmuth, 2006 on Egyptian Arabic; Ladd at al., 2000 on Dutch; Prieto & Torreira, 2007 on Spanish; Ladd et al., 2009 on BrE dialects). Ladd and colleagues (2009), for example, suggest that earlier alignment could be the result of syllabification of a single intervocalic consonant with a following onset after long vowels, but could be ambisyllabic after short vowels. Prieto and Torreira (2007) suggest that the differences in peak location between CV and CVC syllables could be better explained in terms of the gestural coordination and articulatory phonology model (Browman & Goldstein, 1986, 2003; as discussed in §5.2.2).

Hypothesis 3 predicted that the speakers would produce a low pitch accent with a high right edge tone in narrow focus was not confirmed. Supporting Hypothesis 2 that posited a bitonal pitch accent analysis, there was a degree of

24 Anacrusis is a sequence of unstressed syllables preceding the accented syllable (Hirst & De Cristo, 1998). Sullivan followed Ladd (1996) and re-interpreted anacrusis as a sequence of unaccented syllables before the accented syllable.
synchronisation between the tonal targets, replicating the results of other studies (Atterer & Ladd, 2004; Frota, 2002; Prieto, 2006, 2009). The two tonal targets consistently aligned with certain points in the segmental string with differences in rise duration, but any slight movement of the L led to the movement of the H tone. This pattern of coupling confirmed that the two targets were not independently aligned but were part of the same phonological event. The correlation between the low and the high targets was much stronger for the nuclear focal accents produced by the BE group.

Based on the phonetics of L and H alignment and scaling, the following pitch accents categories were posited for each group. The KE tone inventory includes an L+H* accent for both prenuclear and nuclear focal accents (Hypothesis 2B). The BE inventory includes two types of rising accents, corresponding to the English L*+H and L+H* (Hypothesis 2A and 2B). Narrow focus structure was signalled by L*+H on the accented word in BE (Hypothesis 2A).

On the one hand, the results show that the speakers in this study used rising pitch accents included in the inventory of a number well-established Englishes (Pierrehumbert & Beckman, 1988; Fletcher & Harrington; 2001; Arvaniti & Garding, 2007). As mentioned previously, rising pitch accents have also been found in several IndE varieties, such as Gujarati, Tamil (Wiltshire & Harnsberger, 2006), Hindi and Telugu English (Moon, 2002). The phonological pitch accent categories (L*+H versus L+H*) proposed for these varieties, however, remain preliminary and were not based on experimental investigations of tonal alignment. Nevertheless, it appears that rising pitch accent or accents are part of the intonational phonology of IndE. The phonetic realisation of the rise does not reflect the pattern that would be indicative of an L* accent followed by a phrase tone, characteristic for a number of Indian languages (Patil et al, 2008; Harnsberger, 1999; Hayes & Lahiri, 1991; Keane, 2014).

The findings also supported Hypothesis 5 that proposed differences in the inventory of prenuclear and nuclear focal accents as a result of L1 influence. A delayed peak accent in narrow focus produced by the BE speakers is similar to the rising L*+H found in Bangladeshi Bengali (Khan, 2008, 2014). According to Khan (2014), “speakers of Bangladeshi Standard Bengali, especially those in Dhaka, are strongly influenced by Kolkata Standard and other prestigious varieties through the media…” (p. 83). The use of an L*+H accent in BE is also similar to the L*H accent reported for Donegal English and Glasgow English (Mayo et al., 1997) more likely
due to the influence of L1s on these varieties (Kalaldeh et al., 2009).

On the other hand, in well-established varieties of English, the L*+H accent has a marked status (Arvaniti & Garding, 2007) and is used less frequently compared to L+H* and H* accents (Pierrehumbert & Steele, 1989; Dainora, 2002; Fletcher & Harrington, 2001; Fletcher, 2004) and is not restricted to nuclear positions. Moreover, together with the falling boundary tone, this pitch shape corresponds to a particular pragmatic meaning and is often used to express greater incredulity or uncertainty (Pierrehumbert & Hirschberg, 1990; Watson et al., 2008). Therefore, BE shows significant difference with such varieties as AusE AmE and several BrE dialects based on the function of L*+H. The use of a delayed rising accent in BE could potentially have an impact on the communication process and the interpretation of a given utterance.

The fact that KE speakers lack the L*+H accent and have a single rising accent category may be related to Kannada intonational phonology. Despite the lack of studies on Kannada intonation, research on Tamil, a Dravidian language related to Kannada, shows that Tamil has a smaller tonal inventory (Keane, 2014) compared with Bengali (Khan 2008, 2014; Hayes & Lahiri, 1991) or Hindi (Harnsberger, 1996, 1999). A smaller tonal inventory in the speakers’ L1 could potentially account for a lack of contrast between L*+H and L+H* in this variety.

The findings of this experiment also reveal that the same phonological categories of rising accents in KE and BE differ in their realisation of L*+H and L+H* from varieties of AmE and BrE. This difference in the phonetic detail could also be a reflection of speakers’ L1s intonational phonologies and the patterns of phonetic alignment in their L1 (Willems, 1982; McGory, 1997; Mennen, 2004; Mennen, 2007; Kim & Lu, 2011; Stella, 2012). The alignment characteristic of the accents produced by BE and KE speakers in this study exhibited a number of differences in the phonetic realisation of the accents compared with other well-established Englishes. First, the results show that the L target in both L*+H and L+H* is anchored to the accented nasal consonant, with no significant differences in L timing between the early and late peak accents as usually occurs in AmE (Pierrehumbert, 1980; Arvaniti & Garding, 2007; Beckman & Ayers-Elam, 1997) or AusE (Fletcher, 2004). This pattern could be a limitation of the data set, which included only sentence-initial tokens. However, a more likely explanation would be the presence of a phonetic difference in the realisation of pitch accents by the IndE
speakers, reflecting cross-varietal differences in tonal categories. Similar to the findings of this study, Arvaniti and Garding (2007) found that in Minnesotan and Southern Californian English, the L target in L+H* accents aligned in the accented syllable onset, and not before the accented syllable (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988; Beckman & Ayers-Elam 1997).

Second, in the canonical interpretation of an L*+H pitch accent in English, the L aligns with the accented vowel (Pierrehumbert, 1980; Beckman & Ayers-Elam, 1997). BE speakers in this study timed the L target with the onset of the accented syllable, unlike the speakers of AmE in studies by Pierrehumbert (1980), Pierrehumbert and Steele (1989), and Arvaniti and Garding (2007). Pierrehumbert and Steele (1989) found a number of repetitions where the L target of the L*+H category was located in the accented consonant and not a vowel, and simply attributed this to greater variability compared to Pierrehumbert’s (1980) earlier data. In addition to alignment, the L tone of L*+H is usually characterised by a low stretch. In the present study, the L of the L*+H accent produced by the BE group at times showed an optional low region before the rise, but more frequently was produced with a sharp rise at the L target, similar to the one in the L+H* accent. This pattern resembles the phonetic realisation of rising accents in Spanish and Neapolitan Italian (Prieto et al., 2006), where the L aligns with the syllable onset in both L*+H and L+H*. It is the H alignment that defines these two distinct categories, similar to the alignment patterns reported for BE rises in this experiment. As mentioned in §5.5.3.3, L alignment of the L*+H accent will be further examined in non-sentence-initial contexts in Chapter 7.

It has been shown that right-hand prosodic context affects the H tone, wherein the alignment of the peak is timed earlier in nuclear accents as compared to prenuclear ones for a number of languages (Schepman et al., 2006; Ladd et al., 2009; Arvaniti & Baltazani, 2005). The results presented in this chapter show that peaks produced by KE speakers exhibited similar patterns, in that the ones in prenuclear accents were located later than those in nuclear focal words. However, this H timing difference between nuclear focal and prenuclear was very modest. On the other hand, the peak alignment pattern in nuclear focal L*+H accents produced by BE speakers showed a substantial peak delay. Here, the H aligned in the post-accented vowel, which is similar to an L*+H accent category in Greek (Arvaniti et al., 1998). It remains to be investigated whether BE includes this accent category in nuclear non-focal position.
In addition to the H alignment differences, the duration of the rising gesture in IndE varied from other varieties of English. In this experiment, BE speakers produced L*+H accents that had an average duration of 182 ms, with a number of the tokens showing a rise duration of 260-270 ms. In AmE, Pierrehumbert (1980) reported an average rise distance interval of 200 ms for L*+H, while Arvaniti and Garding (2007) reported a shorter rise distance of 172 ms for Minnesotan and Southern Californian English speakers in addition to a later alignment of the targets for Southern Californian speakers.

In terms of f0 manipulation, the scaling of the targets and the use of pitch range were a reflection of the speakers’ individual use of tonal space. The results did not display significant differences in the scaling of L between two prosodic contexts, but the L target measured relative to the speakers’ lower region of pitch range supported the bitonal pitch accent analysis (Hypothesis 2). Despite speaker-specific differences in the use of pitch range, the magnitude of the rise was greater in BE nuclear and prenuclear accents compared to the L+H* accents in KE. This finding is consistent with the f0 results reported in Experiment 1 (Chapter 4).

5.7 Chapter summary

Experiment 2 examined the tonal and scaling characteristics of prenuclear and nuclear focal rises and determined the intonational categories of rising accents in BE and KE. In summary, the findings have shown that speakers in this study produce the tonal categories found in a number of English varieties as well as in IndE (Moon, 2002; Wiltshire & Harnsberger, 2006). All speakers produced a rising pitch accent in nuclear focal and prenuclear accented words, thus supporting Hypothesis 2 and disproving Hypothesis 4 (H* analysis). Hypothesis 3 predicting a low pitch accent followed by a phrase tone (L* (Hp)) on accented words was also disproved. This finding indicates that IndE intonational phonology does not bear a close resemblance to the languages spoken in the subcontinent. In addition, the results show a number of characteristics previously described in the literature on tonal alignment for a number of languages. These include: the strong effect of syllable composition on the alignment of the peak and the accent gesture, supporting the segmental anchoring hypothesis (Hypothesis 1); a more stable L target alignment compared to the peak;
and earlier peaks on nuclear focal L+H* accents compared to prenuclear accents in KE.

As proposed in Hypothesis 5, differences in tonal categories were found between the two L1 groups. Earlier analyses of the data (§4.5.2.4) showed that BE speakers produced later peaks in narrow focus. The results of this chapter have confirmed this observation, in that BE speakers apply an L*+H accent in nuclear focal positions, and use both L*+H and L+H* in prenuclear positions. However, the use of the focal L*+H in BE differs from the use and distribution of this accent in Aus or AmE. The KE tonal inventory on the other hand includes only one rising pitch accent, L+H*.

The results presented in this chapter also show variation in the phonetic realisation of accent categories and add to the ongoing research on English language varieties. First, the L target aligns in the syllable onset for both L*+H and L+H* accents, which is similar to AmE speakers from Southern California and Minnesota (Arvaniti & Garding, 2007) but different to several other English varieties. In the BE L*+H accent, the L target does not necessarily display a characteristic low f0 stretch or the alignment in the accented vowel reported for L*+H accents in well-established varieties of English, showing cross-varietal differences in tune-text alignment. In addition, the duration of the rising gesture differs from the rise of the same pitch accent category in AmE varieties. The following chapter builds on the present results and examines boundary-related lengthening to further determine the levels of prosodic constituency for BE and KE.
Chapter 6: Experiment 3 – Boundary-related lengthening

6.1 Introduction

One of the research questions set out at the beginning of this study was how many levels of prosodic phrasing could be relevant for the two varieties of IndE in this study and possibly applicable to educated varieties of IndE. There has been little previous experimental research into this area. The present chapter is concerned with the amount of final lengthening at boundary edges to determine the number of prosodic constituents above the word.

As discussed in the literature review (§2.1.2), this study follows Pierrehumbert and Beckman’s (1988) framework for intonational analysis. This approach recognises that the prosodic hierarchy of English includes the intermediate phrase (ip), a minor prosodic category above the word, and the intonational phrase (IP), a major prosodic category (Pierrehumbert & Beckman, 1988; Ladd, 1996, 2008; Fletcher et al., 2002; Dainora, 2002). In English, the ip refers to the unit that groups words together with at least one pitch accent. The prosodic hierarchy of several described Indian languages also includes two levels above the word: the phonological, often referred to as the accentual phrase, and the IP (§2.4.4). This is with the exception of Bangladeshi Bengali that includes three levels: the accentual, the intermediate and the intonational phrase (Khan, 2008, 2014). The main difference with English is that the minor prosodic unit above the word (the accentual or phonological phrase) is much smaller in size than the intermediate phrase (§2.4.2). As discussed in §2.4.4, a small size prosodic constituent in languages such as Tamil, Hindi and Bengali usually consists of a single lexical word with a possible addition of a clitic or non-content word, similar to Japanese (2005) and Korean (Jun & Fougeron, 2002). This basic prosodic unit is characterised by a pitch accent followed by a phrase tone, more often corresponding to a rise in pitch (Khan, 2014; Keane, 2014, Patil et al., 2008).

The results of the tonal alignment experiment (§5.5) determined that the rising gestures on accented words were rising pitch accents and not low pitch accents followed by a high word-edge boundary tone. The speakers tended to group words into somewhat larger phrases. What remains to be investigated is whether these

25 The intermediate phrase in the Pierrehumbert and Beckman model is similar to the phonological phrase proposed by Nespor and Vogel (1983).
phrases correspond to one post-lexical prosodic constituent above the word in the form of the IP (i.e. major intonational phrase), or whether the varieties of IndE in this study have two levels of post-lexical prosodic constituency that correspond to the ip and the IP levels similar to several well-established Englishes.

Prosodic phrase boundaries in a language are usually marked by a number of acoustic cues, including strong pitch movements at the right edges of the phrases, pitch declination, the presence of pauses, and pre-boundary lengthening. Duration is considered to be the most reliable cue to the demarcation of prosodic structure (Turk, 2012). The amount of lengthening at the edges of prosodic constituents, or boundary-related lengthening, cues the levels of prosodic structure in a number of languages, including English (Klatt, 1975; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992; Byrd, 2006; Turk & Shattuck-Hufnagel, 2007; Krivokapić & Byrd, 2012). Thus, higher level or major prosodic constituents exhibit greater amount of final lengthening.

Previous research has shown that a phrase-final word is the segmental domain that undergoes lengthening. A number of studies have looked into the amount and distribution of final lengthening within the final word. Wightman et al. (1992) found the rhyme of the syllable preceding the boundary in AmE to be the domain of boundary-related lengthening. Cambier-Langeveld (2000 on Dutch) found the effect of lengthening at the beginning of the main stress syllable, while White (2002 on BrE) reported final lengthening in a penultimate syllable and even in antepenultimate main-stressed syllable coda. Turk and Shattuck-Hufnagel (2007 on AmE) reported that the distribution of lengthening within a final word could be more complex than considered in earlier literature. While most of the duration increase was found in the final syllable, lengthening effects were observed in the main-stressed syllable but only when this syllable was not final. Lengthening also occurred in several other regions but with a more sporadic pattern.

In the present study, an experiment was designed to investigate boundary strength on the basis of final lengthening. Word final syllables at the right edge of different boundary types were investigated. Despite several approaches and views on the segmental domains, the final syllable usually undergoes the greatest amount of lengthening and is a more reliable cue to the prosodic hierarchy levels (Turk &

---

26 See Turk and Shattuck-Hufnagel (2007) for a detailed account of three major approaches to determining the domain of final lengthening.
The chapter outline is as follows. Section 6.2 presents the research hypotheses tested in the experiment. Section 6.3 gives a brief overview of the method, including materials and acoustic analysis. The analyses are presented in §6.4. These are followed by the discussion of the results in §6.5, while §6.6 includes the summary of the chapter.

6.2 Hypotheses

Hypothesis 1

Hypothesis 1 suggests that the speakers in this study use final lengthening as a cue to prosodic boundary strength. The durational increase at the right-hand boundary will be associated with the two levels of prosodic constituency above the word, the ip and the IP, posited for English (Pierrehumbert & Beckman, 1988 for AmE; Fletcher & Harrington, 2001 for AusE). The duration of the final syllable will be longer at the right edge of the ip than at the right edge of the prosodic word. Similarly, the durational increase will be greater in final syllables at the right edge of IPs compared to the final syllables at the right edge of ips.

Hypothesis 2

Hypothesis 2 also assumes that final lengthening is a cue to prosodic boundary strength in BE and KE but is alternative to Hypothesis 1. The hypothesis tests a possibility that speakers in this study do not differentiate between the ip and IP levels, and the IP is the only prosodic constituent above the word. In this case, there will be no differences between the duration of the final syllables at phrase edges. An increase in duration is expected at the phrase edge when compared to the words phrase-medially, thus distinguishing the word from the IP boundary.

Hypothesis 3

Similar to the previous two experiments, it is anticipated that speakers may exhibit variation based on their L1 background. Hypothesis 3A suggests that BE or KE may have differences in the number of prosodic units above the word (phonological variation). In line with more recent research (Khan, 2014), BE may
include both the ip and the IP prosodic constituents, while this may not be the case for KE. There is also a possibility of phonetic variation. Hypothesis 3B predicts that BE and KE may not show the same extent of final lengthening across boundary types.

6.3 Overview of the method

The full description of Experiment 3 was presented in §3.6. The following section briefly outlines the method with a focus on materials, annotation and analysis. The materials used in this experiment consisted of the sentences where the duration of the noun phrase *Manuela’s manual* varied from two to five words. The full template is shown below. These materials were also used in Experiment 2 and were produced with narrow focus. For the purpose of this chapter, only the target word *Manuela’s* (in bold) was annotated and used in the subsequent statistical analysis.

*Lulu/Emily/Marina/Maryellen memorised Manuela’s (main/luminous/yellow-coloured) manual.*

The syllable under investigation was the last syllable not bearing lexical stress in the word *Manuela’s*. Similar to the word Maryellen, it was found that all speakers placed lexical stress on the first syllable of the word. In addition, the target word was always accented, creating comparable conditions across the speakers and repetitions. Below is the phonemic transcription of the word following Maxwell and Fletcher (2009, 2010), reflecting the pronunciation of the target word by the speakers in the study.

/ˈmɛnjuəlæz/

The duration of the final syllable was measured in three conditions, corresponding to different boundary types. It was anticipated that the sentence with a short noun phrase would be produced with no phrase boundary after *Manuela’s* while the sentence with two and more words in the noun phrase would be produced with a phrase break after the target word. Annotations were performed using EMU, consistent with the methodology used in the study. The final syllable was labelled according to prosodic level TYPE:

- W - word-final and phrase-medial
- ip - indicating the edge of an ip
• IP - corresponding to the location of the syllables at the IP edge.

The type of boundaries (W, ip or IP) were identified following standard ToBI labelling procedures and determined on the basis of the degree of disjuncture (Beckman & Ayers-Elam, 1997; Cambier-Langeveld, 2000). As described in §3.6.1, the analysis equivalent of a break index annotation was performed, where 1 was the lowest form of disjuncture corresponding to the prosodic word (W), 3 was the intermediate level of disjuncture (ip) and 4 was the major level of disjuncture (IP). A total of 370 tokens were included in statistical analysis. LMM procedures were performed on the duration of the final syllable to determine the extent of pre-boundary lengthening across the three boundary types (W, ip and IP). Boundary TYPE and GROUP (BE and KE) were included in the model as fixed factors and SPEAKER as a random factor. The interpretation of LMM results and post hoc Tukey tests were carried out following the method described in §3.6. The full results of LMM analysis can be found in Appendix D.

6.4 Results

The analysis showed consistent final lengthening at the edges of phrases. The duration of final syllables phrase-medially was significantly smaller compared to the duration of phrase-final syllables \( t=27.15, p<.00001 \), confirming that speakers in this study relied on pre-boundary-related lengthening. The results also showed a strong effect of boundary TYPE on the amount of final lengthening. Post hoc Tukey tests revealed significant differences across the three types of prosodic boundaries: prosodic word and ip \( z=20.17, p<.0001 \); ip and IP \( z=16.96, p<.0001 \); and prosodic word and IP \( z=-34.02, p<.0001 \). These findings disprove Hypotheses 2 and 3A, and support Hypothesis 1 which proposed two prosodic constituents above the word. The results of final lengthening by boundary TYPE and by GROUP are illustrated in Figure 6.1 below. The y-axis shows the duration of the final syllable in milliseconds. The three boundary types are shown in the x-axis with the corresponding labels for each group. Results for BE are shown in blue, and for KE in green.

As can be seen from the distribution of mean syllable duration in the figure, ip- and IP-final words show progressive lengthening, with the largest increase at the IP boundary signalling greater depth for the higher prosodic constituent. The figure
also shows the effect of GROUP on the extent of final lengthening \( t=2.86, p<.004 \) with an interaction between GROUP and TYPE \( t=-7.90, p<.001 \). This finding supports Hypothesis 3B which proposed phonetic differences between the two L1 groups.

These differences are shown, first, in that the target final syllable phrase-medially produced by KE speakers exhibited significantly longer duration \( t=10.12, p<0.0001 \) than for BE speakers. Second, a smaller durational increase was found for KE speakers in the amount of final lengthening at the edge of IPs and a greater increase was found for the ip-final syllables, leading to smaller differences in the durational increase between the two boundary types. The mean duration of the final syllable at the right edge of the ip in KE measured at 296 ms [SD=29 ms], while the mean duration of the syllable at the IP edge was 327 ms [SD=28 ms], showing a mean increase of 31 ms for this group. There was also a degree of variation in the length of the final syllable at the boundaries among the four KE speakers. In BE, the mean duration of IP-final syllables measured at 337 ms [SD=25 ms], while a mean duration of 269 ms [SD=15 ms] was reported for the ip-final syllables. Thus, the mean increase

---

Figure 6.1 Box and whiskers plot for the distribution of final lengthening by boundary type (word - W, ip - intermediate phrase, IP- intonational phrase), presented by group. Data points shown in the y-axis are presented in milliseconds.
in final lengthening was around of 68 ms, significantly greater than in KE [p<0.001]. Similarly, a more modest durational increase was found between the prosodic word and the IP in KE [mean=124 ms] compared with BE [mean=134 ms].

6.5 Discussion

The results of Experiment 3 have shown that similar to English (Wightman et al., 1991; Fletcher & MacVeigh, 1993; Turk & Shattuck-Hufnagel, 2007) and a number of other languages (Turk, 2012; Cambier-Langeveld, 2000; Berkovits, 1994), final syllables at the right edges of intonational phrases undergo boundary-related lengthening. The increase between phrase-final and phrase-medial final syllables is significant for BE and KE. In addition, final lengthening is progressive in relation to the prosodic boundary depth, with the largest amount of lengthening occurring in the IP-final syllables. This is consistent with previous research which shows boundary-relating lengthening functions as a cue to the levels of prosodic boundaries (Turk & Shattuck-Hufnagel, 2007 for AmE; Wightman et al., 1992 for AmE; Cambier-Langeveld, 1997, 2000 for Dutch; Ladd, 1988; Krivokapić & Byrd, 2012 for AmE).

KE speakers produced the final syllable of Manuela’s phrase-medially with greater duration, indicating possible differences in segmental phonology between the two varieties of IndE previously as reported in the literature (Wiltshire & Harnsberger, 2006; Wiltshire, 2005; Sirsa & Redford, 2013). Consequently, despite the similarity in the number of boundary types, differences between the two L1 groups were observed in the extent of the durational increase in ip versus IP, and W versus IP, with significantly greater lengthening for BE speakers across these boundary types. The present investigation did not make a direct comparison with any of the well-established English varieties. In his study on IndE rhythm, Fuchs (2013) found that vocal intervals at the end of intonational phrases did not show the same extent of lengthening in IndE as in BrE. The durational increase in BE and KE seems to be comparable with that found in AmE (Turk & Shattuck-Hufnagel, 2007). The speakers in this study produced phrase-final syllables with a 124 ms (KE) and 134 ms (BE) increase compared to the syllables phrase-medially.

Turk and Shattuck-Hufnagel reported that final syllables in three-syllable words with antepenultimate stress (Madison, Mendelson) were 118 ms longer phrase-finally than phrase-medially. Most of the increase was associated with the syllable
rhyme (111 ms). The reduced final syllable of the target words included a coda nasal, while in the present experiment the coda consonant was a fricative. This may have lead to a greater durational increase for IndE speakers in the study. Berkovits (1993) found that final fricatives showed almost four times as much utterance-final lengthening as the preceding stressed vowel. Moreover final fricatives showed greater lengthening than stops. More research is needed to examine boundary-related lengthening across different syllable types. The main aim of the present experiment was to look at boundary-related lengthening in order to determine the levels of prosodic constituency and not to examine the domain of final lengthening itself.

On the basis of the analysis, two levels above the word have been posited for the IndE speakers in this study, the intermediate phrase and the intonational phrase, similar to AusE (Fletcher & Harrington, 2001; Fletcher et al., 2002) and AmE (Pierrehumbert & Beckman, 1988; Beckman & Ayers-Elam, 1997). The lack of research on boundary types and the prosodic constituency levels in IndE makes it impossible to compare the results with other varieties of IndE. The current interpretation of the results proposes that educated speakers of IndE rely on final lengthening and a more tentative conclusion can be made that IndE prosodic structure is similar to other well-established varieties or English. Whether this is applicable to other IndE varieties remains to be the issue of future research.

6.6 Chapter summary

In summary, an examination the findings have shown the existence of boundary-related lengthening in KE and BE. Differences between the two L1 groups were observed in the degree of manipulation of the acoustic correlate of duration. While phrase-medial syllables were longer in KE, the extent of final lengthening differentiating word-IP and word-ip levels were significantly greater for the BE group. On the basis of the findings, two levels of prosodic constituency above the word have been found for BE and KE as ip and IP, similar to several well-established Englishes. In the remainder of the thesis, conventional ToBI labels will be used to mark the right edges of ip (“-“) and IP (“%”) boundaries when annotating pitch tracks.
Chapter 7: Description of Bengali English and Kannada English intonation

7.1 Introduction

The present chapter gives a detailed description of the intonational phonology of IndE, focusing on the varieties analysed in this study. In terms of the AM parameters and corresponding research questions, this chapter is concerned with tonal events, tune-to-text alignment and prosodic structure. The description of the IndE intonational phonology includes overall intonational patterns, tone inventory, and nuclear tunes in various modality contexts, namely declaratives, imperatives, polar questions, wh-questions, and coordinations. In addition, it outlines the prosodic constituency levels for the two varieties.

While a number of phonological categories proposed herein are based on the analyses of Experiment 4 (§3.7), this is a summative chapter that builds on the results of Experiments 1, 2, and 3 presented in previous chapters. Taking into account the limited literature and mostly preliminary descriptions of the tonal categories and prosodic structure in IndE, there will be no hypotheses tested in this chapter. The main aims are to give the first comprehensive account of the IndE tone/tune inventory and prosodic structure, outline features that may be characteristic for IndE intonational phonology, and explore the degree of variation between the varieties in this study.

The chapter is organised as follows. Section 7.2 presents a brief overview of materials and summarises the annotation and analysis. Section 7.3 describes the prosodic structure and tune-text association. Sections 7.4 and 7.5 provide the intonational systems for BE and KE, drawing on similarities and differences between the two. Section 7.4 focuses on the tone inventory, divided into pitch accents and boundary tones. Section 7.5 includes the intonational contours and major nuclear tunes for declarative and interrogative intonation based on different sentence types. After these detailed descriptions of BE and KE intonation, §7.6 discusses the phonological inventories proposed and gives a cross-comparative overview of the two varieties, drawing on previous research on English and IndE as well as a number of other languages. Section 7.7 concludes the chapter by presenting a summary of the findings and discussion points, raising issues for further research.
7.2 Overview of the method

Experiment 4, designed to investigate the inventory of tones and nuclear tunes, is described in full in §3.7 of Chapter 3. As a reminder to the reader, the following section briefly outlines the materials, annotation and analysis. The materials used in the experiment included two types of data: read and spontaneous speech. Spontaneous speech consisted of a short narrative and a dialogue in the form of a mock doctor-patient interview (see §3.7.1 for full account). A total of 202 utterances were examined and annotated. These were used to support the patterns observed in the read stimuli. A full investigation of the spontaneous speech data is the subject of future research. The proportional distribution of nuclear tunes is based on the read speech stimuli outlined below.

Read speech is commonly used in intonational research (Grabe et al., 2005; Jun & Fletcher, 2014) and formed the basis of the previous experiments. An additional corpus of read sentences with different modalities was analysed in this experiment. The sentences included both simple and complex declaratives, polar questions, wh-questions, interrogative coordinations and imperatives. The full list of sentences can be viewed in Appendix A under Experiment 4. These sentences were given to the speakers in randomised order with the instruction to read each sentence aloud as naturally as possible. Consistent with the methodology, the words in target sentences included as many sonorants as possible, to avoid breaks in the f0 contour. A total of 821 sentences were used in the analysis.

As mentioned in §3.7, the proportional distribution of questions was greater than that of declaratives. Moreover all sentences assumed a broad focus reading. Therefore, in order to balance the read speech materials, 296 declarative sentences were added from the materials used for Experiment 1 (described in §3.4), including those with both broad and narrow focus. These included read simple declaratives elicited with the help of a question. The sentences that formed part of the analysis in this chapter can be viewed in §3.7.1.

The annotation procedure was consistent with that for Experiments 1, 2, and 3 (§3.4.2, §3.5.2 and §3.6.2) and was performed using the EMU speech database. A small proportion of utterances was cross-labelled with another phonetician familiar with ToBI annotation. An illustration of the EMU annotation with the corresponding labels can be found in §3.7.2. The main difference with the preceding experiments is
the annotation on the tone tier. Three types of tonal events were identified: pitch accents, phrase accents and boundary tones. In addition, the highest point in the $f0$ (HiF0) associated with a pitch accent that had a high tonal component was taken for each declarative sentence. In imperatives, wh- and polar questions, the highest $f0$ point was also measured at the peak of the first high or rising pitch accent to compare pitch expansion on the initial peaks across several modality contexts.

Pitch accents and nuclear tones (a nuclear pitch accent in combination with boundary tones) were extracted using the EMU/R interface. Illustrations of pitch tracks were generated using Praat, showing $f0$ contour on the Hz scale. Similar to the previous experiments, statistical analyses were performed with the help of LMM procedures. The full set of LMM analyses is shown in Appendix E. Any statistical manipulations of the $f0$ values involved the values being converted to the semitones scale, with the 100 Hz point taken as a baseline.

7.3 Prosodic structure and tune-text association

Based on the results of boundary-related lengthening (Chapter 6) and the tonal alignment of rising pitch accents (Chapter 5), there are two main post-lexical constituents in BE and KE: the intonational phrase (IP) and the intermediate phrase (ip). Figure 7.1 is an illustration of the prosodic constituency hierarchy based on the utterance Lara lives in Lilydale.

The figure shows the prosodic constituency levels starting with the highest level at the top (IP) and the lowest at the bottom (syllable, either strong ($\sigma_s$) or weak ($\sigma_w$)). The metrical tree shows the associations between the stressed syllable ($\sigma_s$) and the pitch accents, as well as the association of the phrase accent and boundary tone to the boundaries of the intermediate and the intonational phrases. The tonal events corresponding to the intonational contour of the utterance and their tune-to-text alignment are shown below the syllable on the tone level. Figure 7.2 is an illustration of the corresponding pitch track with word, syllable and tone tier annotations for speaker B1. The utterance was produced with broad focus structure.
Figure 7.1 Prosodic hierarchy with the associated intonational events, illustrated for the utterance *Lara lives in Lilydale*. The utterance is produced as one intermediate and one intonational phrase with pitch accents on *Lara* and *Lilydale*. The intonational structure of the utterance using ToBI annotation is L+H* L*+!H L-L%.

Figure 7.2 $f0$ contour and annotations of the utterance *Lara lives in Lilydale* produced by speaker B1 with broad focus structure.

The IP is the highest level of the constituency hierarchy and is marked by a boundary tone at the right edge (L% or H%). It is also the domain of declination, the downward drift towards the end of the intonational phrase where the high peaks of simple high or bitonal accents show a lowering pattern compared to the preceding ones until the end of the phrase. Consistent with well-established Englishes, utterance-final intonational phrases are affected by final lowering towards the end (Liberman & Pierrehumbert, 1984). The intonational phrase contains at least one intermediate phrase. In the example given above, the right boundary of the intonational phrase coincides with the right edge of the intermediate phrase.
The ip is a prosodic constituent below the intonational phrase in the prosodic constituency hierarchy. The edges of intermediate phrases are marked by phrase accents (L-, H-, !H-), and are located between the pitch accent and the boundary tone. Prominence relations define the intermediate phrase, wherein the last accented syllable in the phrase is the most prominent corresponding to nuclear accent prominence. The number of prenuclear accents is optional and reflects the intonational and segmental structure of the phrase. Similar to well-established Englishes, the intermediate phrase is the domain of the downstep.

Following the prosodic approach to constituency levels (Pierrehumbert & Beckman, 1988; Beckman & Ayers-Elam, 1997; Arvaniti et al., 2005), the identification and differentiation between the intonational phrase and the intermediate phrase levels are described in phonetic terms and are based on purely phonetic cues. The intonational phrase is defined by a number of characteristics. In line with previous research and based on the results of Experiment 3 (§6.4), one of the most reliable cues is a greater degree of lengthening of the IP-final syllable (Wightman et al., 1992; Edwards, et al., 1991; Turk & Shattuck-Hufnagel, 2007). In addition, intonational phrase boundaries are often accompanied by a pause, even in utterance non-final position. The tonal events associated with the intonational phrase often have more complex contours and show greater scaling in the f0 compared to the tones associated with the intermediate phrase. These differences will be illustrated in the section on tone inventory in this chapter (§7.5.2).

The tonal alignment experiment (§5.5-§5.6) showed no evidence of a prosodic constituent of the size of a prosodic word for either of the L1 groups in this study. The analyses clearly showed that the H peak was not used to demarcate the right edge of the word but was part of a pitch accent. As illustrated in Figure 7.1 above, the pitch accents are assigned to prosodic words, have no tonal marking and are automatically associated with the stressed syllable of the prosodic word. In line with canonical English intonation, the occurrence of multiple pitch accents depends on the metrical structure of the segmental material to which the tune is associated (Ladd, 2008).

Despite such close similarity in the prosodic constituency structure with AusE and AmE, there is a striking difference in IndE in the pattern of word grouping into intermediate and intonational phrases. Speakers in this study tended to separate noun phrases with phrase breaks in places not typical for speakers of English. There was a
large degree of variation, with speakers B2, B4, K2 displaying greater amount of ips and IPs within an utterance.

Another difference from well-established English varieties is the relationship between prosodic constituents and prominence levels. According to the prominence hierarchy proposed for English (Pierrehumbert & Beckman, 1988; Beckman & Ayers-Elam, 1997; Ladd, 2008), an intermediate phrase needs at least one accented word to allow for the nuclear accent prominence level and, most importantly, nuclear accents are associated with the rightmost accented stressed syllable of the intermediate phrase. For the most part, the analysis of narrow focus data (see §4.5.2) showed a right-headed assignment of the nuclear accent, where the focal word bearing a pitch accent was the last accented word in the phrase regardless of its position in an utterance. This lack of a post-nuclear pitch accent is an indication that BE and KE nuclear prominence assignment is similar to the one traditionally posited for well-established varieties of English. However, a proportion of the data with sentence-initial or sentence-medial narrow focus showed post-focal placement of a pitch accent with a concomitant pitch compression on post-focal items within the same IP. As argued in §4.6, this feature could be applicable to IndE in general due to the effect of interaction with the speakers’ L1 systems or even the influence of the substratum language Hindi (Rajendran & Yegnanarayana, 1996). In future research, it would be advisable to investigate nuclear accent placement as well as prosodic grouping patterns in English spoken by speakers of IndE from other educational backgrounds and L1s.

7.4 Tone inventory

7.4.1 Tonal categories

Table 7.2 presents the tonal categories with pitch descriptions and their association to intonational events proposed for BE and KE. The inventory of tones is based on the results of Experiments 2 and 3 as well as Experiment 4 of this chapter. The total number of observations for each tonal category based on the ToBI annotations in Experiment 4 can be found in Table E.1 (Appendix E). High tones occurred when the $f0$ turning points were in the top three-quarters of the speaker’s range, while the low tones were approximately in the bottom quarter of the range and perceived as low.
As shown in the table, both varieties exhibit the same inventory of tonal categories associated with ip and IP boundaries, consisting of three phrase accents and two boundary tones. It is the pitch accent inventory that sets the two varieties apart. English spoken by the Bengali L1 group includes four basic pitch accents (H*, L*, L+H*, L*+H) with three downstepped realisations (!H*, L+!H*, L*+!H), while Kannada English has three basic accents (H*, L*, L+H*) and only one downstepped accent (!H*).

Table 7.1 Tonal categories in ToBI, their associations and general pitch description proposed for the intonational phonologies of BE and KE.

<table>
<thead>
<tr>
<th>Association</th>
<th>Pitch description</th>
<th>Tone - BE</th>
<th>Tone - KE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch accents</td>
<td>Simple high</td>
<td>H*</td>
<td>H*</td>
</tr>
<tr>
<td>Simple low</td>
<td>L*</td>
<td>L*</td>
<td></td>
</tr>
<tr>
<td>Rising</td>
<td>L+H*</td>
<td>L+H*</td>
<td></td>
</tr>
<tr>
<td>Delayed rising</td>
<td>L*+H</td>
<td>L*+H</td>
<td></td>
</tr>
<tr>
<td>Downstepped high</td>
<td>!H*</td>
<td>!H*</td>
<td></td>
</tr>
<tr>
<td>Downstepped rising</td>
<td>L+!H*</td>
<td>L+!H*</td>
<td></td>
</tr>
<tr>
<td>Downstepped delayed rising</td>
<td>L*+!H</td>
<td>L*+!H</td>
<td></td>
</tr>
<tr>
<td>ip boundary</td>
<td>High</td>
<td>H-</td>
<td>H-</td>
</tr>
<tr>
<td>Low</td>
<td>L-</td>
<td>L-</td>
<td></td>
</tr>
<tr>
<td>Downstepped high-mid</td>
<td>!H-</td>
<td>!H-</td>
<td></td>
</tr>
<tr>
<td>IP boundary</td>
<td>High</td>
<td>H%</td>
<td>H%</td>
</tr>
<tr>
<td>Low</td>
<td>L%</td>
<td>L%</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the differences in the pitch accents inventory, the results also indicate differences in the use of pitch accent types in prenuclear versus nuclear positions. These will be discussed later in the next few sections as well as in the description of intonational contours and nuclear tunes (§7.5).

The following sections describe the tonal categories grouped by their tune-to-text associations, namely pitch accents, phrase accents (ip boundary) and boundary tones (IP boundary). The phonetic characteristics of each tonal event, patterns of distribution and prosodic positions (prenuclear and nuclear) are considered.
7.4.2 Pitch accents

As shown above, BE and KE have three basic pitch accents: high (H*), rising (L+H*) and low (L*). BE also includes a delayed rising accent, L*+H.

H*

A simple high accent (H*) was usually perceived as a high pitch, and was realised in the middle or upper part of the speaker’s f0 range. It was used in both nuclear and prenuclear positions. Similar to the simple high in AmE and AusE (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988; Arvaniti & Garding, 2007; Fletcher et al., 2005), a high tone was at times preceded by a shallow rise. In BE, this was more common phrase-initially, coinciding with the pitch reset at the beginning of an intermediate phrase, while in KE shallow rises were also observed phrase-medially. The starting point of the rise scaled in the middle of the speaker’s f0 range, while the peak scaled from the mid to upper part of the range depending on the position of the accented syllable within a phrase. A shallow rise characteristic of the H* accent can be observed in Figure 7.3, showing the pitch track of a polar question produced as two ips within one IP.

![Figure 7.3 Intonational contour of a question produced by speaker B2 showing the occurrence of an H* accent on the prenuclear accented words will and mellow.](image)

The H* accents shown in this figure are associated with the prenuclear accented words will and mellow, and show a rise in the f0 measured at around 25 Hz. In both cases the rises were shallower on will and mellow compared to the rising bitonal accents on the words Maryellen and May. In addition, a sequence of H* accents often exhibited no dip in the f0 between the high tonal targets, resembling a ‘flat hat’ pattern reported for a number of English varieties including AusE (Fletcher
For BE speakers, the use of H* and sequences of H* accents were more frequent in questions, while KE speakers tended to produce this pattern in both questions and declaratives. Figure 7.4 illustrates the pitch track of an utterance produced by a KE speaker. All accented words bear a high pitch accent. The f0 on the accented syllables of the words *Lulu* and *lived* show a very shallow rise at word onset but are produced as a stretch of high pitch.

![Figure 7.4](image.png)

Figure 7.4 Two intonational contours illustrating an H* and an !H* pitch accents produced by speaker K3.

Figure 7.4 also illustrates different phonetic realisations of the simple high accent. It shows a downstepped high accent (!H*) which was more common for nuclear words in KE (on *Melbourne*) and constituted the most common tune. Consistent with the AM framework, the tonal target has a lower f0 relative to the preceding high tone and can only occur within the same intermediate phrase. The preceding accent can be either high or rising (see §2.1.1). In phrase-final prosodic positions, H* was typically realised as a plateau regardless of the L1 group. Such a plateau is shown on the nuclear-accented word *Melbourne*. Most frequently, the duration of the plateau coincided with the duration of the accented syllable but at times it spanned over the post-accented syllable.

In the instances where the H* accent showed a clear peak, the location of the peak varied and was usually influenced by the location of a pitch accent within an IP. When the nuclear H* accent was phrase-final, the location of the peak coincided with the accented syllable. In the instances when prenuclear H* was utterance-initial, the peak was often delayed into the post-accented syllable, which is common cross-linguistically (Nolan & Farrar, 1999 on BrE; Silverman & Pierrehumbert, 1990 on AmE; Frota, 2002 on European Portuguese). The delayed high accent was labelled...
following the ToBI conventions for English with the help of a ‘<’ diacritic. This phonetic realisation of a simple high accent is shown Figure 7.5, where the H peak on Nelly is within the post-accented syllable of the word. Delayed peaks never expanded beyond the boundaries of the post-accented syllable. Because KE speakers used H* more frequently, high accents with delayed peaks were more common in the KE data.

![Figure 7.5 Illustration of a delayed high pitch accent in nuclear position in a declarative produced by speaker K4.](image)

**L+H***

Based on the tonal alignment experiment results (§5.5), both KE and BE tonal inventories include an L+H* accent. Unlike the shallow dip that may precede a high tone in H*, the rise in L+H* was usually steep and sharp with a distinctive fall after the peak where the L and H tonal targets exhibited a tight coupling. In most cases, the tones were anchored to the accented syllable. Similar to the findings of Arvaniti and Garding (2007), the L tone was found at the beginning of the accented syllable. The rise in the f0 usually spaned the accented syllable and the H aligned in the middle or second half of the accented vowel, depending on the prosodic condition. Later peaks were observed in IP-initial syllables. Despite similarity in its phonetic realisation, the L+H* tone differed in its context and prosodic position between the L1 groups.

In BE, L+H* was widely used on non-focal prenuclear and nuclear words. For speakers B2 and B4, it occurred more often on nuclear accented words. L+H* was also the most frequent pitch accent type on wh-words and command verbs in imperatives for this group. Figure 7.6 illustrates the occurrence of an L+H* accent on phrase-initial prenuclear words *my, know* and *lived* in a complex declarative. The f0 contour on the accented syllables in each of these words begins with a low target located at the beginning of that syllable. The scaling difference between the L and the
H, together with the $f_0$ of the L target below the middle of the speakers’ pitch range were additional cues to a bitonal pitch accent. The magnitude of the rise on the word my, for example, measured at 109 Hz. The $f_0$ rise on the word know measured at 86 Hz, while the differences between the L and H values on the word lived reached 106 Hz.

![Figure 7.6 Illustration of an L+H* accent in a complex declarative produced as a series of IPs, each starting with a rising accent. Speaker B4.](image)

In KE, the L+H* accent was mostly used in nuclear focal positions. Figure 7.7 illustrates an example of a narrow focus structure with an intonation contour of an utterance produced by speaker K2 in spontaneous speech. The word really is placed in narrow focus. The rise begins at the onset of the accented syllable, with a slight perturbation in the rhotic, and peaks within the vowel. The peak reaches around 200 Hz and falls throughout the post-accented syllable. Relative to the speaker’s pitch range, the L tone is located in the lower part of the tonal space. This speaker has a higher low region (§5.5.1.4).

![Figure 7.7 Illustration of an L+H* accent occurring on the nuclear focal word really, produced by speaker K2 in spontaneous speech.](image)
In addition to narrow focus marking, a prenuclear L+H* was used in several syntactic structures in KE. These included some of declaratives, wh-words in wh-questions, and command verbs in imperatives, all phrase-initial. It is possible that due the limited use of L+H* by KE speakers (phrase- initial prenuclear or nuclear focal), this type of accent does not have a downstepped phonetic realisation, unlike the high accent (!H*). There is also a possibility that this type of accent is used to indicate more emphasis.

$L^*+H$

A delayed rising accent L*+H in BE intonation was characterised by a low pitch movement (L tone) followed by a rise to a peak (H tone). The results of the tonal alignment (§5.5.2) showed that the L tone was anchored to the onset of the accented syllable, most likely a consonant, while the H tone was realised in the post-accented syllable. The delayed H tone in L*+H peaked beyond the boundaries of the post-accented syllable when the post-accented syllable was open and contained a short vowel. As determined in the tonal alignment experiment (§5.5), this type of accent was one of the prosodic means to mark narrow focus structure. An example of the L*+H pitch accent on the word *borrow*, produced in narrow focus, is shown in Figure 7.8.

![Figure 7.8 An illustration of an L*+H in nuclear focal position (borrow), produced by speaker B2.](image)

The pitch track shows that the low $f_0$ starts to rise at the beginning of the accented vowel in *borrow* and reaches its peak in the post-accented syllable. The magnitude of the rise is around 150 Hz, and the H tone is realised in the uppermost part of the pitch range for this speaker, similar to the other bitonal accent (L+H*)
described in the previous section. The L*+H shows a later rise starting point and a clear valley for the L tone. This could have been the effect of the bilabial stop. Most of the instances of L*+H were produced without a valley.

As argued in §5.5.1.5, the L*+H tone was used in nuclear focal and prenuclear contexts by BE speakers. The analysis of the corpus in the present chapter shows that this type of accent also occurred in nuclear non-focal contexts. The non-focal L*+H was often realised in compressed pitch, annotated as L+!H* (See Figure 7.2). Another feature found in BE intonation, possibly linked to phrasing, was the absence of a sequence of bitonal downstepped accents. The BE rises, L*+H and L+H*, do not extend to more than two downstepped accents in a single phrase.

$L^*$

The L* accent was produced in the lower part of the pitch range and was perceived as low. The low f0 span associated with L* often depended on the number of syllables in the accented word and the prosodic position. In nuclear accented words with more than one post-accented syllable, the f0 spanned over the whole stressed syllable and into the post-stress syllable. In monosyllabic words or words with stress on the final syllable, the f0 valley was usually quite short and showed a sharp rise corresponding to the boundary tone configuration. Distinguishing L* from !H* in IP-final position was at times challenging as a result of the lowering in the f0 towards the end of an utterance. This was more the case with KE speech due to a narrower pitch range use. Figure 7.9 illustrates an example of the L* accent in the L* H-H% tune used in a polar question.

Figure 7.9 Illustration of an L* accent in a polar question produced by speaker K4.
In BE, the L* occurred in prenuclear and nuclear contexts. A more common use of the L* accent, however, was observed in nuclear low-to-mid rises (L* L-H%), as well as low rises (L* H-H%). In KE, this accent type occurred less frequently. Moreover, L* was used by KE speakers solely on nuclear accented words in polar questions and coordination questions, and occasionally in wh-questions.

**Pitch accent distribution and accentuation**

Figure 7.10 presents the distribution of pitch accent types by speaker. Each pitch accent category includes possible phonetic realisations of the accent (such as a downstep or a peak delay), where applicable.

![Figure 7.10](image)

Figure 7.10 Proportional distribution of pitch accents for each speaker. The types of pitch accents are colour-coded: H* - light green, L* - dark green, L+H* - light blue, L*+H - dark blue.

As shown in the figure, the most frequently used pitch accents in BE were a simple high H*, a rising L+H* and a delayed rising L*+H accents. The proportional use of L* was similar in this group, apart from speaker B1 whose productions showed slightly higher usage of this accent. The figure also shows that for BE speakers the combined use L*+H and L+H* is greater than the use of H*.

The predominant pitch accent for the KE group was a high accent (H*). It was widely used for both nuclear and prenuclear accented words. A rising L+H* is also included in the inventory but was usually a) associated with greater emphasis and
narrow focus, and b) used in prenuclear positions sentence-initially in some declaratives and in questions. KE speakers displayed little use of the low accent, especially speakers K1 and K2.

Experiment 1 showed that KE intonation was characterised by greater accentual density than BE (§4.5.1.1). The results were based on sentences consisting of three content words. The patterns were to be reviewed by looking at a larger sample of data with more complex syntactic structures. Similar to Experiment 1, the present analysis of pitch accents involved calculating the total number of accented (A) and stressed (S) syllables for each group. This time, the analysis included annotation of 1117 sentences in the form of declaratives and interrogates. The proportional distribution of accented versus stressed syllables is shown in Figure 7.11 below.

![Figure 7.11 Proportional distribution of accented (blue) versus stressed (grey) syllables for BE to the left and KE to the right.](image)

Figure 7.11 Proportional distribution of accented (blue) versus stressed (grey) syllables for BE to the left and KE to the right.

Mirroring the results of Experiment 1 (See Figure 4.9 in §4.5.1.1), KE speakers placed accents on a larger number of words, thus confirming somewhat greater accentual density for KE intonation. As shown in the figure, the ratio of accented versus stressed syllables corresponded to 52:48 for BE and 63:37 for KE.

### 7.4.3 Boundary tones

**Phrase accents**

The main phrase accents in BE and KE are L-, H- and !H-. These three accents are contrastive in terms of the $f0$ level in a speaker’s pitch range associated with the
accent. The analyses presented in this chapter have followed an approach adopted by Pierrehumbert and Beckman (1988), and Beckman and Ayers-Elam (1997) in the treatment of phrase accents in English. That is, a phrase accent does not necessarily align at the edge of the phrase but is a tonal event occurring between the pitch accent and the edge of the full intonational phrase boundary. Pierrehumbert and Beckman (1988) have distinguished the peripheral association of the phrase accent with the prosodic phrase edge, but have also suggested the possibility of a secondary association with the specific prosodic domain (the edge of the nuclear accented word). As mentioned in §2.2.1, investigation of this possibility for the two varieties of IndE is beyond the scope of the current study.

The L- was a phrase accent realised as a low pitch target relative to the pitch range of the phrase. Its phonetic realisation varied from a sharp low turning point to a distinct low \( f_0 \) stretch. This variability was linked to the location of the nuclear pitch accent in a phrase, as well as to the choice of the following boundary tone. In the instances of a nuclear accented syllable located in non-phrase-final position (when the accented syllable was followed by a number of stressed syllables) in L-L% boundary configurations, the L- accent was characterised by a stretch of low \( f_0 \) to a falling or drooping \( f_0 \) until the end of the phrase, as shown in Figure 7.12 below.

![Figure 7.12 Illustration of a wh-question produced by speaker B1, showing the occurrence of an L- phrase accent.](image)

When a nuclear accented syllable was close to the right edge of the phrase, there was a sharp fall in the \( f_0 \) from the high target of the pitch accent, as illustrated in the second IP in Figure 7.13. When a low phrase accent was followed by a high boundary tone, the \( f_0 \) showed a sharper transition, often with a clear turning point regardless of the length of post-nuclear segmental material.
As shown in the figure below, the turning point at the end of the first IP is located at the onset of the nuclear-accented syllable of the word *lime* and this is where the $f0$ starts to rise. Across the whole corpus, instances of L- in the middle of IPs were, however, very rare. There were only a few found in BE intonation, suggesting that this type of phrase accent is more likely to be used in combination with an IP boundary tone. The speakers in this study relied on high or downstepped high phrase accents to signal an ip mid-utterance.

![Figure 7.13 Illustration of the occurrence of an L- phrase accent at the end of a coordination question produced by speaker K1.](image)

The H- phrase accent corresponded to the tonal target realised in the upper part of the speaker’s range, ranging from mid-level to high pitch. Its phonetic realisation also depended on the amount of segmental material after the nuclear accented syllable and the type of boundary tone to follow. There was no difference between KE and BE. The $f0$ shape associated with the H- was either a plateau pattern (as shown in Figure 7.14) or a distinct rise.

![Figure 7.14 An illustration of the intonational contour of a polar question produced by speaker K4, showing two occurrences of an H- phrase accent at the end of each IP.](image)
In Figure 7.14, the high pitch after a shallow rise at the onset of the accented syllable in *Mary* is sustained and the $f0$ shape resembles a plateau until the end of the prosodic word. In the second intonational phrase *live in Melbourne*, the nuclear syllable is assigned a low pitch accent and the $f0$ rises sharply through the nuclear accented syllable to plateau through the rest of the phrase. Another example of a sharp transition from a low pitch accent to a high phrase accent is shown in Figure 7.15. The nuclear accented syllable in *Mary* has a low pitch accent, manifested by a low valley throughout the stressed syllable. The $f0$ then has a sharp rise at the end of the nuclear-accented word. Subsequently, at the beginning of the next ip, the $f0$ falls for the upcoming L+H* accent.

![Figure 7.15 An illustration of an H- before the clause, produced by speaker B1 in the complex declarative.](image)

Similar to the downstepped pitch accents, !H- is used to describe a phrase accent realised somewhere in the middle of a speaker’s pitch range. It occurs after the pitch undergoes the process of downstep. In the data, this was mostly triggered by a bitonal accent, either L*+H or L+H*. Sequences of H* !H- occurred but in a smaller proportion. The downstepped phrase accent exhibited phonetic variation, depending on the length of the segmental material after the nuclear accent and the type of boundary tone, either high or low. A longer stretch in $f0$ after the !H- was typical when there was more segmental material between the nuclear accented syllable and the right edge of the phrase in combination with a low boundary tone.

As shown in Figure 7.16, the first IP has a rising nuclear accent followed by a sharp fall through the post-stress syllable *my*. The $f0$ then gradually lowers and has a slightly falling plateau through the rest of the phrase. On auditory analysis, this was perceived as being level pitch. This pattern was more common for BE speakers.
Boundary configurations

Two boundary tones denoting the right edge of the IP were posited for KE and BE as L% and H%. In combination with phrase accents, these boundary tones formed a number of configurations (L-L%, L-H%, H-H%, H-L%, and !H-L%). Depending on the L1 group, these configurations exhibited a number of differences in their realisation and most importantly in their distribution.

One of the boundary configurations was a sequence of L- and L% tones. This configuration was more common in declaratives, followed by a proportion across imperatives and wh-questions. In KE, it was used in over 60% of the corpus, and often followed !H*. In BE, the L-L% configuration followed a pitch accent with a high target (H*, L+H*, L*+H plus their downstepped realisations). The L-L% configuration was never preceded by a low accent, which is not surprising. The L* L-L% tune is also rare in other well-described varieties of English (e.g., Dainora, 2002 on AmE).

Consistent with the ToBI guidelines, the L% was associated with the lowest part of the speaker’s pitch range, and was located lower in the tonal space than the L-tone as a result of final lowering (Liberman & Pierrehumbert, 1984). Figure 7.17 illustrates two examples of L-L%. As can be seen from the pitch track and annotations, the imperative utterance was divided into two intonational phrases, with a shorter tail in the first IP.
On both words bearing a nuclear pitch accent (\textit{said} and \textit{clean}), the $f_0$ drops sharply through the post-accented syllable. In the second intonational phrase, however, the low $f_0$ resembles a low plateau until the right edge of the phrase. This pattern of low, slightly falling $f_0$ was rather common in utterances with non-phrase final nuclear accents. In BE, changes in the $f_0$ and the amount of final lowering at the right edge of declaratives were usually greater than in KE. Moreover, they were accompanied by an audible difference. This difference in the phonetic realisation of L-L\% could be based on the speakers’ L1 and is perhaps linked to the use of a narrower pitch range in KE.

The L-H\% tonal configuration was usually represented by a low turning point in the $f_0$ followed by a sharp rise somewhere to the middle of the speaker’s pitch range. The $f_0$ associated with the edge of the phrase was not typically higher than the highest tonal target associated with the pitch accent of that phrase. In KE, the use of L-H\% was limited to L* L-H\% and H* L-H\% tunes. The latter were observed in the productions of speakers K1, K3 and K4, but not K2.

The L-H\% configuration was mostly associated with ‘continuation’ rises in BE intonation, including declaratives, as well as polar, wh- and coordination questions. The spontaneous speech data included a large number of declaratives and questions with this type of tonal sequence, giving additional evidence for its wide usage in BE. The L-H\% sequences were preceded by any type of pitch accent, with L* and L+H* being the most common. Figure 7.18 illustrates a rising pitch accent followed by an L-H\% configuration. The word \textit{Nina} forms a single IP corresponding to a full intonational contour of L*+H L-H\%. 

![Figure 7.17 Illustration of the occurrence of an L-L\% boundary configuration with different nuclear accents in each IP. Speaker B2.](image)

![Pitch (Hz)](image)
Figure 7.18 Illustration of the occurrence of an L-H% boundary configuration following a delayed rising accent. Speaker B2. The word Nina is in narrow focus.

The main difference between the H-H% and the L-H% configurations was the magnitude of the rise. The former was phonetically realised in the upper level of the speaker’s pitch range, while the latter only reached mid-level. The H% tone of the H-H% tonal sequence was affected by upstep, regardless of the nuclear accent preceding it (Pierrehumbert, 1980). In other words, the H% after the high phrase accent scaled higher than the preceding high phrase accent. In BE, the sequence of H- and H% usually occurred in polar questions as well as a number of wh- and coordination questions, thus indicating a preference for a rising pattern in questions overall. It was more commonly preceded by a low pitch accent, but also occurred after the H* and L+H* accents. Figure 7.19 below is an example of an H-H% configuration used in a rising contour in the question Does Nelly like yellow lemons?

Figure 7.19 Illustration of the occurrence of an H-H% boundary configuration in combination with L* accent. Speaker B1.

The nuclear accent was placed on the first syllable in the word yellow, and was produced with an L* accent. At the offset of the post-nuclear syllable, the f0 started to
rise smoothly and peaked in the nasal of the final syllable in *lemons*. The H target associated with the boundary tone measured at 237 Hz, extending higher in the tonal space than the f0 recorded for the H target of the phrase-initial pitch accent on the word *does* (182 Hz).

In BE, H-H% was also observed at the end of declarative utterances, both simple and complex. The rise tended to be preceded by the L* accent rather than the H*. This pattern resembled the use of high rising terminals (HRT) in AusE (Fletcher & Harrington, 2001; Fletcher et al., 2005). An example of a rise with a nuclear L* used in declarative intonation is illustrated in Figure 7.15 earlier. The f0 in the second intonational phrase (*who lived in Melbourne*) had a sharp rise to the upper level of the speaker’s pitch range, with the final f0 value of 260 Hz. There was a degree of variation across BE speakers, with rises in statements more typical for speakers B1 and B4. To the contrary, KE speakers produced a limited number of L* H-H% tunes. The use of H-H% for this group was mostly restricted to polar questions and interrogative coordinations, often in combination with H*.

Similar to AusE (Fletcher & Harrington, 2001; Fletcher et al., 2005) and AmE (Pierrehumbert, 1980; Beckman & Ayers-Elam, 1997) and unlike German (Grice et al., 2006), Greek (Arvaniti & Baltazani, 2005) or Glasgow English (Mayo, 1996; Mayo et al., 1997), the low boundary tone in the H-L% configuration underwent an upstep after the H-, thus bringing the L% to the level of the phrase accent. The f0 stretch between the two tonal events resembled a plateau or level contour. The use of H-L% was similar in BE and KE, but with higher frequency for BE speakers. An example is shown in Figure 7.20. The words *Does Nelly like* form one IP with a high plateau tune.

![Figure 7.20 Illustration of the occurrence of an H-L% boundary configuration in the first IP. Speaker B4.](image-url)
The H-L% configuration also occurred in compressed pitch and was a variation of the H-L% boundary configuration, affected by the process of downstep on the phrase accent. The !H-L% was more common for BE intonation after a bitonal pitch accent. The f0 representation of this contour was usually manifested by a drop in pitch followed by a plateau stretch in the middle of the speaker’s pitch range (see Figure 7.16 earlier in the section).

7.5 Intonational contours and nuclear tunes

7.5.1 Major nuclear tunes in BE and KE

Table 7.2 below presents the major nuclear tunes found for the speakers in this study. The nuclear tunes are described in terms of their AM phonological composition as well as more general pitch configurations (such as a fall, a rise or fall-rise), equivalent to the British style of nuclear tone categories (Halliday 1967; Cruttenden, 1986). The table includes contour types, their ToBI tonal categories and a schematic representation of each tune. In addition, the last two columns outline the contexts in which the tunes can be used in BE and KE. In the schematic contour of the tunes, grey-coloured boxes correspond to accented syllables and light-coloured boxes correspond to post-accented material. The upper edge of the box also functions as the baseline of the speakers’ pitch range. Here, the closer the lines are to the elongated box, the lower the pitch.

On the pitch contour, the heavy lines show pitch movements associated with pitch accents and are aligned in relation to the accented syllables. Where applicable, the dotted lines prior to the heavy line indicate the pitch shape just before the accented syllable. The remainder of the line shows pitch movement associated with the boundary configuration. All drawings of nuclear tunes assume that the nuclear-accented syllable is followed by at least one post-stress syllable.

It is assumed that a number of these tunes may have a downstepped realisation of the pitch accent or phrase accent with a high target, such as !H* L-L%, in addition to H* L-L%. The analysis of the current corpus shows no differences in meaning between major tunes and their possible phonetic realisations. However, this does not exclude the possibility of a more restricted use for some of the tunes composed with a downstepped tone.
Table 7.2 Major nuclear tunes presented as tune type, ToBI annotational categories, schematic contour and contexts occurring in BE and KE.

<table>
<thead>
<tr>
<th>Tune type</th>
<th>ToBI</th>
<th>Schematic contour</th>
<th>Context - BE</th>
<th>Context - KE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fall</td>
<td>H* L-L%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Declaratives, wh-questions, polar questions</td>
<td>Declaratives, polar, wh-questions, coordinations, imperatives</td>
</tr>
<tr>
<td>2. Fall-rise</td>
<td>H* L-H%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Declaratives, wh-questions, imperatives, continuation</td>
<td>Mostly continuation</td>
</tr>
<tr>
<td>3. Rise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a. High rise</td>
<td>H* H-H%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Polar and coordination questions</td>
<td>Polar questions</td>
</tr>
<tr>
<td>3b. Low rise</td>
<td>L* H-H%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Polar and coordination questions, declaratives</td>
<td>Polar questions and coordinations</td>
</tr>
<tr>
<td>3c. Low-to-mid rise</td>
<td>L* L-H%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Declaratives, wh-questions, continuation</td>
<td>Declaratives, wh-questions and continuation</td>
</tr>
<tr>
<td>4. Rise-fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a. Rise-fall</td>
<td>L+H* L-L%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Declaratives, wh- and polar questions</td>
<td>Narrow focus declaratives</td>
</tr>
<tr>
<td>4b. Scooped rise-fall</td>
<td>L*+H L-L%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Narrow focus, declaratives, wh-, polar and coordination questions</td>
<td>NA</td>
</tr>
<tr>
<td>5. Rise-fall-rise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a. Rise-fall-rise</td>
<td>L+H* L-H%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Some wh-questions and declaratives</td>
<td>NA</td>
</tr>
<tr>
<td>5b. Scooped rise-fall-rise</td>
<td>L*+H L-H%</td>
<td><img src="image" alt="Schematic" /></td>
<td>Declaratives and some wh-questions</td>
<td>NA</td>
</tr>
</tbody>
</table>
Similar to a number of English language varieties (e.g., AmE or AusE), the tune inventory for both L1 groups includes a range of nuclear contours: a simple fall (category 1, H* L-L%), a rise (category 3, L* L-H% or L* H-H%), and a rise-fall (category 4, L+H* L-L%). For both groups, the falls encountered in the corpus are those with a high tone on the nuclear accented syllable, corresponding to a simple high or the high tone of a bitonal pitch accent. It seems that low pitch accents are always followed by a rise, either to the upper (category 3b, L* H-H%) or to the mid level (category 3c, L* L-H%) in the speaker’s pitch range.

Table 7.2 also shows several differences in the tune inventory between the two groups. First, BE exhibits a larger inventory of nuclear tunes. Second, there is a strong reliance on rising, rising-falling and rising-falling-rising intonation (categories 3, 4 and 5) across several modality contexts in BE. Third, the corpus shows that BE intonation includes a large proportion of nuclear tunes with a level contour or a plateau (category 6). These can have a simple high, a downstepped high or a rising pitch accent. Plateau tunes are found in the utterances in which the speaker inserts an additional phrase break followed by an audible pause. This is the case with both statements and questions in BE, where sentences are frequently broken down into two or more IPs. It appears that nuclear level contours are often used to signal
continuation of the information, showing there is more to come together with the more canonical continuation tunes that include a final rise to mid level: a fall-rise (category 2, H* L-H%), a rise-fall-rise (category 5, L*+H L-H% or L+H* L-H%) or a low-to-mid rise (category 3c, L* L-H%).

KE intonation includes 5 major contour types: a fall, a fall-rise, a rise, a rise-fall, and a plateau contour. This group therefore has fewer nuclear contours in their inventory. Similar to BE, KE speakers use three types of nuclear rises, H* H-H%, L* H-H%, and L* L-H%. This is consistent with the literature on well-researched varieties of English (Ladd, 2008). The number of fall-rises (H* L-H%) for the KE group, however, was limited and was mostly used for medial tunes reserved for instances of continuation mid-utterance, consistent with English (Fletcher & Harrington, 2001; Ladd, 2008).

The data analysis shows that there was no evidence of a rise-fall-rise contour (category 5) in KE. The most widely used nuclear contour was H* L-L%. KE speakers tended to apply this contour across various sentence modalities including statements (broad focus), questions, coordinations and imperatives. The KE inventory also includes a rise-fall (category 4, L+H* L-L%) but it is limited to certain contexts, observed less frequently in the data. The sections to follow present the analysis of intonational contours and the distribution of nuclear tunes in different modality types starting with declaratives.

7.5.2 Declaratives

Figure 7.21 below shows the types of nuclear tunes and their proportion across the corpus in simple and complex declaratives for BE (left) and KE (right). The analysis of tunes is based on sentence-medial and sentence-final intonational phrases. As mentioned above, level-like or plateau tunes were mostly associated with sentence-medial contexts. These were more common for complex declaratives, longer sentences and sentences with a number of multisyllabic words (leading to greater amount of phrasing). Each tune is colour-coded, and the tune types with corresponding colours are shown in a bar plot for each L1 group.
As can be seen from the figure, all speakers produced a rise (L* L-H% in KE and L* L-H% as well as L* H-H% in BE, shown in yellow), a fall (H* L-L%, dark blue), a rise-fall (L+H* L-L%, purple) and a high plateau (H* H-L%, bright green). BE, however, shows a wider range of contours with differences in the distribution of nuclear tunes. Four types of tunes account for 77% of repetitions. These include two rise-falls, L*+H L-L % (shown in grey) and L+H* L-L% (purple), and two rises, L* H-H% and L* L-H% (combined in the yellow bar). BE speakers used two types of rising tune more frequently than KE speakers used L* L-H%, at a rate of 23% of utterances for BE versus 6% for KE. The realisation of a low rising tune with the high boundary tone realised in the upper level of the speaker’s pitch range is shown in Figure 7.22 below.
One of the possible explanations of \( L* \ L-H\% \) usage could be the effect of laboratory speech and the choice of read sentences as stimuli, which may lead to a ‘listing’ style of intonation. However, the frequency of its use was relatively high (especially in BE) and this pattern emerged in the spontaneous speech data. Low range rises are common in AusE declarative intonation and often indicate a forward-looking function (Fletcher et al., 2002).

The other tunes were used to a lesser degree in BE. The nuclear contours \( L+H* \ L-H\% \) and \( H* \ L-H\% \) were combined due to a relatively small proportion of repetitions of the \( H* \ L-H\% \) and are shown as the \( L(H*) \) \( L-H\% \) category (in dark green). The analyses also indicate a degree of variation across the speakers. Speaker B4, for example, used more rises at the right boundary, involving the \( L+H* \ L-H\% \) (dark green) and \( L* \ L-H\% \) contours (yellow). Speaker B3 tended to use more falling tunes (\( H* \ L-L\% \), dark blue) compared to the rest of the speakers. Speakers B1 and B2 had greater reliance on rising-falling (\( L+H* \ L-L\% \), purple) and delayed rising-falling (\( L*+H \ L-L\% \), grey) tunes.

By contrast, KE speakers showed less variability and mostly relied on two main tunes in declarative intonation, \( L+H* \ L-L\% \) (purple) and \( H* \ L-L\% \) (dark blue), corresponding to 44% and 45% respectively of total use. The results report a very small percentage of \( L* \ L-H\% \) (yellow), \( H* \ H-L\% \) (bright green) and \( L+H* \ H-L\% \) (not shown in the figure), but the majority of tunes are falling. The most widely used tonal configuration for KE was the \( H* \ L-L\% \) sequence, corresponding to the falling nuclear contour. This pattern is presented in Figure 7.23, showing a pitch track of the utterance *Nelly likes yellow lemons*.

![Figure 7.23 Illustration of the occurrence of a falling tune (H* L-L%) in the second IP. Speaker K1.](image)
The whole utterance is produced with a series of high accents. This was the most common pattern in KE declarative intonation, and consisted of a ‘pointed’ or a ‘flat’ hat pattern. The utterance in the figure above is realised as two intonational phrases, where the second phrase shows a final fall. High $f0$ is associated with the nuclear accented syllable, and the $f0$ starts to fall at the onset of the post-stress syllable. The $f0$ associated with the L% tone measures at 80Hz. This tune was frequently realised as !H* L-L% in the corpus. Figure 7.23 also illustrates a level tune with a downstepped accent, !H* H-L% (bright green). This type of nuclear contour was more characteristic for speakers K1 and K2. In spontaneous speech, high plateau tunes often corresponded to unfinished utterances.

All speakers used a rising-falling tune (shown in purple) which consisted of a rising accent L+H* followed by a falling boundary configuration (L-L%). In BE, this type of tune contributed to 19% of usage in the read data and was used in broad focus. In KE, the rise-fall was used in narrow focus contexts. Figure 7.24 below is an illustration of this contour and shows a pitch track of the sentence Nina may move my mill, produced with narrow focus on move.

The pitch track shows that the $f0$ starts to rise sharply at the onset of the monosyllabic word move and peaks towards the end of the rhyme. The $f0$ then sharply falls through the post-stress syllable and continues falling gradually until the right edge of the phrase. This type of tune allows for a long, short or no tail depending on the position of the nuclear focal syllable in the phrase. In the instances of a long tail, the $f0$ exhibits a sharp rise in the nuclear accented syllable followed by a sharp fall through the post-accented syllable.
In terms of BE, the tonal alignment experiment determined that the speakers use a delayed or ‘scooped’ pitch accent on a focused word. As a result, one of the main declarative tunes in narrow focus was a delayed rising-falling tune (L*+H L-L%, grey). An example of sentence-medial narrow focus is presented in Figure 7.25.

This type of tune allows for both a long and a short tail, depending on the length of the post-nuclear material. This tune constituted around 35% of total use for BE speakers. It was used in utterances with narrow focus structure, and was also observed in non-focal contexts, especially in the production of speakers B1 and B3. As shown in Figure 7.21, it was the most frequently used tune in BE declaratives. An example of L*+H L-L% in broad focus is illustrated in Figure 7.2 earlier in the chapter, showing a downstepped realisation of the nuclear L*+H on the word Lilydale.

In summary, the analysis showed a strong preference for rising-falling and rising nuclear tunes in BE, but mostly falling tunes in KE (except for narrow focus). A falling nuclear contour is very common in a number of English varieties (Pierrehumbert & Beckman, 1988; Grabe et al., 2005), but in BE it constitutes only 8.5% of total utterances. Moreover, it has been observed in the speech of speakers B2, B3 and B4, but not in speaker B1 who consistently produced rise-falls.

7.5.3 Polar and wh-questions questions

Polar Questions

The types of tunes and their distribution in polar questions are illustrated in Figure 7.26. Unlike the intonation patterns in declaratives, both groups showed a smaller
tune inventory for polar questions consisting of three main tunes: a high rise (H* H-H%, grey), a low rise (L* H-H%, red) and a fall (H* L-L%, dark blue). BE speakers (on the left) also used an additional rise-fall L+H* L-L%, shown in yellow.

Figure 7.26 Proportional distribution of nuclear tunes in polar questions for BE (left) and KE (right).

As shown in the figure, differences between BE and KE are reflected in the proportional distribution of the tunes. In BE, the most common final nuclear contour was a low rise in expanded pitch range (L* H-H%, red). It accounted for 57% of the total use and was produced by all speakers. An example of this contour is illustrated in Figure 7.27. The pitch movement starts in the lower part of the speaker’s pitch range with the nuclear accented syllable in *Melbourne* bearing a simple low pitch accent. This is followed by a sharp rise at the right edge of the phrase. The high boundary tone reaches the upper part of the speaker’s pitch range.

Figure 7.27 Illustration of the occurrence of a low rising tune (L* H-H%). The f0 value of the H% tone is 320 Hz. Speaker B1.
This type of tune was used to a much lesser degree in KE (14%). The preferred choice for KE polar questions was a rise with a high onset (H* H-H%, grey), as shown in Figure 7.28. The overall f0 shape in the IP in this figure shows a pattern typical for KE intonation in polar questions. Apart from a shallow phrase-initial dip in the f0, the utterance is produced with a series of simple high accents. Most of the questions that were produced with a high rising tune (grey) showed a sustained high pitch throughout the IP with a sharp rise starting around the nuclear or post-nuclear syllable.

Figure 7.28 Illustration of the occurrence of a high rising tune (H* H-H%). Speaker K2.

The falling nuclear contour (dark blue) was the second preferred tune type for KE. Moreover, it was used more frequently in KE polar questions than in BE, at a rate of 40% compared with 10%. As a result of the downdrift throughout the IP, the falls were often produced with an !H*, corresponding to the downstepped fall (!H* L-L% sequence). An example of this pattern is shown at the end of the IP in Figure 7.29.

Figure 7.29 Illustration of the occurrence of a falling tune (!H L-L%), with the nuclear !H* accent. Speaker K4.
In addition to the difference in tune distribution between the groups, there was also a large degree of variation across all speakers, regardless of their L1. For instance, not all KE speakers produced the three types of tunes – high rises (H* H-H%, grey), low rises (L* H-H%, red) and falls (H* L-L%, dark blue) – shown in Figure 7.26. Speakers K1 and K4 predominantly produced falls in polar questions, with a very small number of rises with low or high nuclear accents. Speakers K2 and K3 only relied on the rises, with a high rising tune being the preferred choice for speaker K2.

All BE speakers produced two types of rises (H* H-H% and L* H-H%), while falling (H* L-L%, dark blue) and rising-falling (L+H* L-L%, yellow) tunes were observed in the productions of speakers B1 and B2, but not speakers B3 and B4. Similar to declaratives, a nuclear pitch accent in questions was often placed on the last content word, leading to a short tailed realisation of a rising-falling tune (L+H* L-L%, yellow) in BE intonation, echoing the results of the accentuation pattern in broad focus (§4.5.1.1). This pattern is illustrated in Figure 7.30. The last content word May has a nuclear accent. The pitch begins to rise at the onset of May, then peaks and falls sharply at the right edge.

![Figure 7.30 Illustration of the occurrence of a rising-falling tune (L+H* L-L%) in a polar question. Speaker B4.](image)

**Wh-questions**

Figure 7.31 shows the types of tunes in wh-questions and their distribution in BE (left) and KE (right). Overall, the most common f0 movement at the right edge of the intonational phrase was a falling pitch. BE speakers used a fall (32%, dark blue) and a rise-fall (46%, purple), with the latter including both L*+H and L+H* nuclear accents. A relatively small proportion of wh-questions was produced with a low-to-
mid rising contour (L* L-H% yellow) by BE (14%) and KE speakers (6%). KE speakers mostly relied on the falling nuclear contour (93%).

![Proportional distribution of nuclear tunes in wh-questions for BE (left) and KE (right).](image)

Figure 7.31 Proportional distribution of nuclear tunes in wh-questions for BE (left) and KE (right). LH in the bar plot to the left (LH L-L%) corresponds to the use of both L*+H and L+H* accents in nuclear position.

An example of a delayed rise-fall (L*+H L-L%) with the nuclear accent placed on the last content word is shown in Figure 7.32. This type of tune was mostly produced with a short tail due to accent placement on the last content word.

![Illustration of the occurrence of a delayed rising-falling tune with a downstepped nuclear accent (L+!H* L-L%) in a wh-question. Speaker B1.](image)

Figure 7.32 Illustration of the occurrence of a delayed rising-falling tune with a downstepped nuclear accent (L+!H* L-L%) in a wh-question. Speaker B1.

Figure 7.33 is an illustration of a falling tune, preferred by KE speakers. The figure shows the pitch track of a wh-question produced with a sequence of high accents. The nuclear accent on Maya is realised in compressed pitch.
Another point of interest in wh-questions was the use of the prenuclear accent type on wh-words (Figures 7.32 and 7.33), which was not restrictive of a particular L1 group. There were a handful of repetitions across the corpus where the speakers did not place an accent on wh-words. An additional pitch accent was at times placed on the auxiliary following a wh-word (as shown in Figure 7.33). The difference between KE and BE appears to be based on the proportional use of rising versus high accents on wh-words. Figure 7.34 presents the proportional distribution of these two pitch accents for each group. BE speakers used L+H* on wh-words in around 85% of productions, while KE speakers used it in around 36% of their repetitions. The preferred pitch accent in KE was the H* (64%).

Figure 7.33 Illustration of the occurrence of a falling tune (!H* L-L%) with a nuclear !H* in a wh-question. Speaker K3.

Figure 7.34 Proportional distribution of L+H* (rising - blue) and H* (high - green) on wh-words.
**Pitch range expansion**

Taking into account that in English and a number of other languages, wh-questions show an expansion of pitch range on the first accented word, leading to higher $f0$ peaks on the first pitch accent in wh-questions compared to declaratives, this phenomenon was investigated for the speakers in this study. The highest point of the $f0$ in the vicinity of the first accented word was taken across three sentence types: declaratives, polar questions and wh-questions. LMM procedures were performed with GROUP and sentence TYPE as fixed factors, and SPEAKER as a random factor. These results are presented in Figure 7.35. The box plot shows the distribution, outliers and the medians of the $f0$ scaling (in semitones) for each sentence type, with D standing for declaratives, PQ for polar questions and WhQ for wh-questions. Blue boxes correspond to BE, while green boxed show the results for the KE group.

![Box plot showing the distribution, outliers and median of the $f0$ peak on the first accented word in declaratives (D), polar questions (PQ) and wh-questions (WhQ), presented by group. BE boxes are labelled in blue, KE boxes - in green. The $f0$ values on the y-axis are shown in semitones.](image)

Despite no effect of GROUP [p>.05], there was significant effect of sentence TYPE [p<.001]. Tukey post hoc test revealed differences across the three modality types [PQ-D: $z=7.6$, p<.001; WhQ-D: $z=12.8$, p<.0001; WhQ-PQ: $z=5$, p<.00]. As shown above, the increase in the $f0$ height did not work in the same direction for KE and BE. The results confirmed a strong interaction between GROUP and TYPE [p<.001], and showed lower scaling of the $f0$ peak in wh-questions [t=-5.04, p<.001]
and higher scaling in polar questions \([t=8.8, p.<.0001]\) for KE speakers. This is in contrast to a gradual increase in \(f0\) from declarative to wh-questions in BE.

In addition, there was a large degree of overlap in the distribution of the \(f0\) values between wh-questions and declaratives for KE speakers. Further t-test analyses performed for each speaker individually showed higher \(f0\) scaling in wh-questions only for speakers K3 \([t = 4.84, p< .001]\) and K4 \([t = 10.76, p< .00001]\), and not for K1 and K2 \([p>.05]\).

### 7.5.4 Interrogative coordinations

Coordinations refer to the syntactic constructions with two or more units or conjuncts combined into a larger unit. Conjunction constructions occur in most languages, but their formation shows cross-linguistic variation (Haspelmath, 2000). English has several types of conjunctions. In order to compare the intonation contours of coordinations with other English language varieties, preference was given to interrogative coordination in this study, similar to the IViE corpus (Grabe, 2004; Grabe et al., 2005). Below is an example of a coordination structure in a question. The two conjuncts, *lime* and *lemons*, form a larger syntactic unit. Both words have the same syntactic relation with the other elements of the question.

*Does Nelly like lime or lemons?*

The intonation patterns of coordinations in BE and KE showed similarity in sentence-medial tune choice. Both conjuncts were usually accented. The first word always carried a pitch accent, and the second conjunct was accented in around 95% of productions. The speakers often placed the two conjunct words into separate IPs, at times separated by an extra short pause (around 60 ms). This pattern is illustrated in Figure 7.36 below. The sentence-medial nuclear tune association with the first conjunct word (*lime*) is produced with a low-to-mid rise (L* L-H%). The word *lime* bears a nuclear accent, wherein a low \(f0\) at the onset of the syllable is followed by a gradual rise (of 35 Hz) until the phrase boundary.
In the sentence-final contexts of interrogative coordinations, there were differences between the groups with KE speakers relying on two utterance-final nuclear contours of H* L-L% (as shown in the second IP in Figure 7.36) and L* H-H%. The fall contributed to 93% of the tunes distribution, and the low rise contributed to 7%. Moreover the low rise was only observed in the productions of two KE speakers (K2 and K3).

On the other hand, BE speakers exhibited a total of four nuclear tunes in coordinations. These were a low rise (L* H-H%), a high rise (H* H-H%), a fall (H* L-L%), and a rise-fall (L+H* L-L%). There was a large degree of inter-speaker variation in the BE group, reflecting a preference for different tune type or types. Speakers B1 and B4, for example, used a low onset rise (L* H-H%) which is also found in polar questions. Speaker B1 produced a number of repetitions with a high rising contour (H* H-H%). The nuclear tunes produced by the other two speakers (B2 and B3) included a fall in the f0 at the right edge, but varied in their pitch accent choice associated with the nuclear syllable. Figures below show the same utterance produced by speaker B2 (Figure 7.37) with a falling tune (H* L-L%) and speaker B3 (Figure 7.38) with a rising-falling tune (L+H* L-L%) in the second IP.
Figure 7.37 Illustration of the occurrence of a falling tune (H* L-L%) in the second IP of the coordination question. Speaker B2.

Figure 7.38 Illustration of the occurrence of a rising-falling tune (L+H* L-L%) in the second IP of the coordination question. Speaker B3.

The figures above also illustrate the presence of a prenuclear accent on the auxiliary verb does. In both instances the word was produced with a rising pitch accent by the BE speakers. Similarly, the pitch track of the coordination question produced by speaker K1 in Figure 7.36 earlier shows accentuation of the auxiliary word. This feature of accenting auxiliaries was common for both groups, resembling prenuclear accent placement in polar questions (see Figures 7.27-7.30) and frequent accentuation of auxiliary verbs in wh-questions (see Figures 7.32 & 7.33).

7.5.5 Imperatives

Imperatives refer to the verbal mood category in which the speaker wants the hearer to perform certain action. The syntactic structure of imperative utterances used in the stimuli consisted of an indirect speech sample, where the embedded expression was an imperative, as in *He said to Mary, “Clean some yellow lemons!”*. The intonational patterns of the embedded speech have been included in the analysis of imperatives.
In KE, the single tonal configuration used in imperatives was a fall (H* L-L%). The most common realisation of the nuclear high tone was a downstep, shown in the second IP in Figure 7.39. The IP is produced as a sequence of two high pitch accents, with the second one realised lower to the preceding one. The f0 falls sharply after the peak on the nuclear accented syllable in *yellow*.

![Figure 7.39 Illustrations of the occurrence of a downstepped fall (H* L-L%) in the second IP, corresponding to the embedded imperative. Speaker K3.](image)

BE speakers showed more variation in the choices of nuclear tunes (Figure 7.40). The most common tune in BE was a rise-fall (L+H* L-L%, purple) followed by a fall (H* L-L%, dark blue) and a fall-rise (H* L-H%, dark green).

![Figure 7.40 Proportional distribution of nuclear tunes in BE imperatives.](image)

In BE, a rise-fall (L+H* L-L%, purple) contributed to around 61% of the repetitions. For speaker B3, this contour was used in all of the repetitions. The pitch track of the imperative *Clean some yellow lemons* in the second IP in Figure 7.41 illustrates a rise-fall (L+H* L-L%) in an imperative. Compared to the prenuclear high...
accent on *Clean*, the nuclear accent is produced with a higher $f0$ and shows a distinct rise of 95 Hz. This reverse of declination with L+H* pitch accent in nuclear position occurred in around 20% of BE imperatives.

Figure 7.41 Illustrations of two tune types in imperatives. The sentence on the left shows the occurrence of !H* L-L% (speaker B2) and the second sentence on the right shows the occurrence of L+H* L-L% (speaker B3).

The realisation of the falling tune in BE, used in around 22% of productions, was very similar to the falls produced by the KE speakers. This pattern is illustrated in the first IP corresponding to the imperative produced by speaker B2 in Figure 7.41 above. The nuclear accent on *yellow* is realised in compressed pitch (!H*).

**Pitch range expansion**

On visual examination, imperatives produced by both groups exhibited expanded pitch range at the left edge of the intonational phrase compared to declaratives, which is common for English (Pierrehumbert, 1980) and a number of other languages (e.g., Estebas-Vilaplana & Prieto 2009 for Castillian Spanish). To investigate this issue further, the $f0$ of the peak associated with the first pitch accent in declaratives was compared with the $f0$ of the peak on the command verb (also the first accented word) in imperatives. The box plot in Figure 7.42 shows the distribution, outliers and medians of the $f0$ scaling (in semitones) for declaratives (D) and imperatives (I). BE results are shown in blue to the left, while KE results are in green to the right.
Figure 7.42 Box plot showing the distribution, outliers and median of the $f0$ peak on the first accented word in declaratives (D) and imperatives (I), presented by group. BE results are shown in blue, while KE results are in green. The $f0$ values on the y-axis are presented in semitones.

The LMM analysis confirmed a significant raising of the $f0$ in imperatives compared to declaratives [I-D: $t=13.4$, $p<.0001$]. There was no effect of GROUP, but a significant interaction between GROUP and sentence TYPE [$t=-3.688$, $p<.02$], reflective of the lower pitch top line values in KE. As can be seen from the figure, the difference in the $f0$ scaling on the first accented word between the two sentence types was more modest among KE speakers. Moreover, the box size and the whiskers indicate a large degree of variation for this group. The mean increase in the $f0$ ranged from 2.8 to 4.3 semitones, with speaker K3 producing the highest peaks. In BE, the mean increase was around 4.9 semitones for speaker B1 and around 7.5 semitones for speakers B2, B3 and B4. This increase was significant, even taking into account raising of the peak on the nuclear accent by BE speakers in some of the repetitions.

7.6 Discussion

One of the findings of this study is that both varieties of IndE have the same prosodic structure. As determined in Experiments 2 and 3, and presented in §7.3 of this chapter, there are two tonally marked prosodic constituency levels above the word: the intermediate phrase and the intonational phrase. In addition, BE and KE show the same tonal events associated with the intermediate and intonational phrase boundaries
in the form of three phrase accents (H-, L- and !H-) and two boundary tones (H% and L%). These results are consistent with a number of well-established Englishes (described in §2.1.2). The differences between KE and BE have been found in the inventory of tones and tunes, the distribution of tune types depending on modality, and the distribution of pitch accent types.

In narrow focus statements, both varieties rely on a rise-fall, with some rise-fall-rises among BE speakers. In BE, the narrow focused tune consists of a delayed rise (L*+H) in nuclear position, as opposed to the L+H* in KE, showing systemic differences in the use of rising-falling tunes. A limitation of the present study is that the stimuli did not include interrogatives with narrow focus structure. This issue is worth investigating in the future.

The intonational patterns in broad focus present a more complex picture. In declaratives, KE speakers use three types of tunes, with a preference for a falling tune (H* L-L%), similar to bilingual Punjabi-English speakers from Bradford, UK (Grabe, 2002; Grabe et al., 2005). This is also the most common nuclear tune in declaratives in some varieties of American (Pierrehumbert & Hirschberg, 1990) and British English, such as Cambridge, Newcastle, Leeds, London and Dublin English (Grabe, et al., 2005; Grabe, 2004). In the case of KE, this could potentially be linked to the speakers’ L1. Future research is needed that would directly compare KE and Kannada intonation.

BE declaratives show the use of seven tunes, with rising-falling (L+H* L-L%, L*+H L-L%) and rising (L* L-H% and L* H-H%) being the most frequent. The distinct use of rising-falling contours for this group could be linked to the rise-falls in the declarative intonation of Kolkata Bengali (Hayes & Lahiri, 1991). In addition, the speakers of BE produced more complex pitch movements in nuclear positions, such as rise-fall-rises. The use of rise-falls and rise-fall-rises has also been found in Kolkata (Hayes & Lahiri, 1991) and Bangladeshi Bengali (Khan, 2014; represented as L* followed by HL% and HLH% boundary configuration). BE declarative intonation also includes a proportion of low rises (L* H-H%). This tune type is used in statements in AusE (Fletcher & Harrington, 2002) and is similar to the ‘standard’ declarative tune in Glasgow and Belfast English (Fletcher et al., 2005; Grabe et al., 2005).

Mid-utterance KE speakers used a more canonical English ‘continuation contour’, the H* L-H% tune (Ladd, 2008; Fletcher & Harrington, 2002; Fletcher et al.,
However, the H% tone was realised at mid-pitch level, unlike in AusE where the boundary tone can be realised at both high and mid level pitch (Fletcher & Harrington, 2001). In BE, the fall-rise was used less frequently to indicate continuation, and was often replaced by L* L-H% or two types of rise-fall-rises, showing differences in tune choice between BE and KE.

All speakers used a number of tune types in polar questions, with a preference for a rising tune, which is consistent with English (e.g., Pierrehumbert & Hirschberg, 1990). In BE, the L* H-H% is a more likely choice, showing the highest use (57%). This tune type is similar to the L* H% in English spoken by Punjabi-English bilingual speakers (Grabe, 2004), and unlike a rise-fall with the L* nuclear accent in the speakers’ L1, Bengali (Hayes & Lahiri, 1991). In KE, polar questions are often produced with a high rising tune (H* H-H%), while in BE this type of tune is used as a second choice. Furthermore, the tune H* L-L% and its realisation with a downstepped nucleus (!H* L-L%), common for KE declaratives, may also be used for polar questions in KE and is the first choice in wh-questions (93%), again similar to English used by Punjabi-English bilinguals (Grabe, 2004; Grabe et al., 2005). This provides an additional indication of semantic differences between the varieties investigated in the study.

BE wh-question intonation shows the presence of several tunes, similar to BrE (Grabe et al., 2005) and AmE (Pierrehumbert & Hirschberg, 1990; Hedberg et al., 2010) and Bengali, which include falls, rise-falls and rises in wh-questions (Khan, 2008, 2014; Michaels & Nelson, 2004; Hayes & Lahiri, 1991). The preferred tunes for this type of question, however, are rising-falling and delayed rising-falling (L+H* L-L% and L*+H L-L%).

It appears that the most common boundary tone configuration in BE and KE is falling (L-L%), which is consistent with other English varieties (Hedberg et al., 2010; Grabe et al., 2005) and IndE (Gargesh, 2004). However, the choice of rising versus high accents in nuclear position contributes to the difference between the two varieties. The preference for a rising-falling tune in wh-questions found in BE is also common in Hindi (Harnsberger, 1994). As mentioned previously (§2.4.2), there is insufficient research on Kannada to date. Preliminary findings showed that Kannada interrogatives produced by male speakers have a falling f0, while female speakers used rising intonation, but it was not clear what the phonological categories of pitch accents and boundary tones were (Mathew & Bhat, 2010). This pattern, however, is
similar to KE declaratives and similar to Tamil, a language related to Kannada. According to Keane (2006b, 2014) wh-questions in Tamil are not distinguished from declaratives by nuclear pitch accent and tune type, but show phonetic differences based on the f0 height at the beginning of the wh-questions and the f0 scaling at the end of wh-questions. Similarly, six speakers (excluding two KE speakers) in this study relied on the f0 expansion on wh-words, showing higher peaks compared to the IP-initial accents in declaratives. The increase in the f0 on the first accented word in wh-questions (and imperatives) compared to declaratives is significantly smaller for the KE group, reflecting differences in the use of f0 modulation between KE and BE reported in earlier experiments (shallower rises in KE).

The intonation of wh-questions also shows differences in the use of prenuclear pitch accents in BE and KE. Contrary to AmE and BrE patterns, but similar to English spoken by Punjabi-English bilinguals (Grabe, 2004; Grabe et al., 2005), the wh-word is accented most of the time, regardless of the speakers’ L1. Similar to nuclear accents, KE speakers used H* more frequently in this context, while BE speakers showed a preference for a rising L+H* accent. The latter could be linked to the speakers’ L1, as Bengali wh-questions are usually focused and the wh-word receives the nuclear accent with a rising contour (Hayes & Lahiri, 1991). Hedberg and Sosa (2010), studying the spontaneous speech of AmE speakers, found that a simple high (H*) or delayed rising accent (L*+H) were more common on wh-words, with only 6.5% of accented wh-words showing sharp rises associated with L+H*.

Overall, the emerging picture is that the two varieties of IndE in this study use a number of tunes and show variations that fall into systemic, semantic and realisational categories according to Ladd’s (2008) typology (§2.1.3). One of the most important findings of this chapter is the distribution of rising-falling versus falling tunes. In this, KE speakers mostly relied on H* L-L% or !H* L-L% across several modalities, limiting the use of L+H* L-L% to narrow focus. In contrast, the use of H* L-L% was less frequent in BE and the speakers often preferred L+H* L-L%, L*+H L-L% or L* H-H%. Taking into account that the boundary tone inventory is the same for the two varieties and is similar to the one posited for AusE (Fletcher & Harrington, 2001; Fletcher et al., 2002) and AmE (Pierrehumbert, 1980; Beckman & Ayers-Elam, 1997), this tune distribution difference is linked to pitch accent inventory and the distribution of pitch accent types.
As determined in the tonal alignment experiment (§5.5.3) and the present results, BE has a three-way contrast across pitch accents with a high tonal target. This use of H*, L+H* and L*+H is likewise found in some varieties of BrE and AmE as well as in AusE. Contrastive L+H* and L*+H also exist in a number of other languages, such as German (Grice et al., 2005), Neapolitan Italian (Prieto et al., 2006), and Greek (Arvaniti & Baltazani, 2005). Similar to the lack of contrast between rising accents in Glasgow (Mayo et al., 1997), Donegal (O’Reily et al., 2010) and Belfast English (Grabe et al, 2005), KE includes only one rising accent, L+H*, contrasting with H* on the basis of pragmatic use and the extent of the f0 rise.

This chapter also highlights a number of new findings. First, both varieties of IndE have a low accent but it is used infrequently in BE compared to other accents and has limited use in KE. L* often occurs in L* L-H% tunes, reflecting the higher probability of L-H% after a low accent as in English. Dainora’s (2002) study for AmE showed that an L* accent has a greater likelihood of occurring preceding L- phrase accent in combination with the H% boundary tone. In addition, less frequent use of L*, reserved for nuclear positions, is similar to well-established Englishes, which are referred to as an H* language (Sosa, 1999; Hedberg & Sosa, 2007). Unlike Bengali (Hayes & Lahiri, 1991; Michaels & Nelson, 2004; Khan, 2008, 2014) and other Indian languages (Patil et al., 2008 on Hindi; Keane, 2007, 2014 on Tamil) that are known for a high use of the L* accent and the rise that stretches over the entire lower level prosodic constituent, BE and KE rises are realised as rising pitch accents, further confirming previous research that IndE has developed its own phonology as a variety among WE varieties.

Second, L*+H in BE is not restricted to narrow focal and prenuclear accents, but also occurs in nuclear non-focal contexts, declaratives and wh-questions. Speakers B1 and B3 showed the most use of this accent compared to the other two BE speakers. Nuclear non-focal L*+H has also been found in Bangladeshi Bengali (Khan, 2014). The present study did not investigate the alignment of rising accents in nuclear non-focal position, therefore the conclusion that the speakers use two distinctive types of rises (L+H* and L*+H) on nuclear non-focal words is preliminary at this stage and needs further analysis by looking at the tonal alignment. Regardless of the phonological category of nuclear non-focal rises, BE intonation is still characteristic for its rising accents, indicating that this is a somewhat ‘rising’ variety. As discussed in §2.3.2 and §5.6, the preference for rising accents could be a feature of Indo-Aryan
languages. Wiltshire and Harnsberger (2006) also found that Gujarati English showed a greater number of rises (either L*+H or L+H*) and Moon (2002) has suggested that Hindi English speakers use a rising pitch accent in narrow focus.

Third, this data shows a high frequency of the H* accent or its downstepped realisation in KE. This use of H* is in fact similar to a number of well-established varieties of English, however, it could also be due to the tonal categories in the speakers’ L1. Further investigation into Kannada intonation is needed. It is interesting that the intonational phonology of Tamil (Keane, 2014), a language related to Kannada, includes sequences of rises (L* followed by high phrase accent), similar to Hindi and Bengali (§2.4.2, §2.4.3), but unlike Bengali, does not include rising pitch accent or accents.

The results of Experiment 1 (§4.5.1.1) showed that in simple declaratives KE speakers placed accents on a greater number of words than BE speakers. This finding has been confirmed in the additional analysis presented in this chapter. The evidence also points to a similarity among the speakers that is noted for IndE generally (Latha, 1978; Bansal, 1969, 1990; Wiltshire & Harnsberger, 2006; Gumperz, 1982), wherein pitch accents can be placed on function words, including auxiliaries and prepositions, in both BE and KE. Because the present study did not compare prominence placement directly with other English varieties, it is impossible to establish the degree of this difference, but clearly KE intonation has higher accentual density than BE. This suggests that KE differs from well-established Englishes and also differs from BE on the basis of accent distribution, which is another parameter of cross-linguistic prosodic variation (Hellmuth, 2007; 2010). An unusual aspect is that all speakers in this study were fluent in English. Consistent with previous research on L2 intonation and the link between L2 proficiency and L1 influence (§2.2.2), Hellmuth (2010) found that advanced or more fluent speakers of L1 Egyptian Arabic, a language with accents on every word, did not transfer this feature into English. Greater accentual density was more typical for lower level speakers. This is may not be the case for IndE, and accentual density may be more characteristic for some L1 backgrounds. This feature needs to be explored in the future across different proficiency levels and educational backgrounds.

Pitch accent inventory and the distribution of accent types may provide an explanation of the differences in pitch range and pitch span, the amount of f0 modulation, between BE and KE. The results of Experiments 1, 2 and 4 showed that
KE speakers use a narrower pitch span. Rise magnitude on accented words is overall more modest for KE speakers. By contrast, BE shows greater pitch modulation with significantly higher rises on accented words indicating a wider use of pitch span and overall pitch range. Mennen, Schaeffler & Docherty (2012), comparing English and German intonation, showed that different languages or dialects may use characteristically different f0 ranges, and that the f0 range can influence the realisation of tones at different points in the intonation contour as well as contribute to differences in the incidence of distribution of tones. In other words, the use of f0 may be linked to the tonal categories of BE and KE with a high density of H* pitch accents in KE and a large number of rising accents in both prenuclear and nuclear positions in BE.

Another feature related to the use of f0, common across well-established varieties of English within the AM theory, is the process of downstepping within an intermediate phrase in non-focal contexts. Downstep was produced by the speakers in this study but with differences between the two L1 groups. The intonational patterns of KE show that only H* undergoes downstep while L+H* does not, perhaps due the function of the latter (focus and greater emphasis). This is different from BE intonation where all accents with high targets can be realised as a downstep. The most common pattern in BE is one or two downstepped accents in a phrase. This is in contrast to AusE or AmE, where a phrase can be produced as a sequence of downstepped accents (either L+H* or H*). It also differs from Bengali dialects, where each subsequent smaller phrase in one IP is produced with a following lowered H tone (Hayes & Lahiri, 1991, Michaels & Nelson, 2004; Khan, 2008, 2014) in a downtrend equivalent to the process of downstep in Germanic languages (Khan, 2014). The most likely explanation for the lack of such sequences in BE is the pattern of prosodic grouping and formation of several phrases in a single utterance, not typical for well-established Englishes but noted previously for IndE (Bansal, 1969; Dhamija, 1976). In addition, some declarative and wh-questions were produced with a reverse or suspension of declination in BE, perhaps similar to the L+H* accent in non-focal contexts instead of a downstepped nuclear accent in Bengali (Hayes and Lahiri, 1991). In English and in a number of other languages, this feature is more common for polar questions. As for KE, less ‘chunking’ of information and the greater amount of accented words may explain the presence of downstepped sequences.
7.7 Chapter summary

Based on the analysis presented in this chapter, KE and BE show some similarities with other Englishes as well as similarities with each other, namely prosodic structure and constituency levels, inventory of boundary tones, modelling of the boundary tones H% and L% as an upstep in configurations with a high phrase accent, a range of nuclear tunes to indicate different modality contexts, a frequent use of L-L% in wh-questions, and a more restricted use of the L* accent (mostly in nuclear position). In addition, most speakers manipulated the f0 in different types of sentences, showing differences in the phonetic realisation of intonational contours as opposed to the phonological distinctions between declaratives and questions, and declaratives and imperatives.

The findings have also shown semantic, systemic and realisational differences in the intonational patterns between KE and BE, possibly as a result of L1 influence. These differences are in the choices of nuclear tunes and the distribution of tune types in interrogative and declarative intonation. First, BE includes a wider range of nuclear tunes with several (rising)-falling-rising nuclear contours. Second, KE has a strong preference for falling nuclear contours, while BE has a large number of rising-falling tunes, as a reflection of rising accents in nuclear position. Third, despite the presence of low (L* H-H%) and high (H* H-H%) rises, BE intonation is characterised by a greater use of low onset rises, including declaratives, and KE shows greater use of high onset rises. In addition, both groups use rising-falling tunes in narrow focus structure, but the realisation of this tune differs (BE L*+H L-L% versus KE L+H* L-L%). Fourth, these differences in tune choices between BE and KE are linked to the preference for rising versus high pitch accents in prenuclear and nuclear positions, which together with accent distribution could be a factor in pitch range differences.

To conclude, the comparison of these two varieties suggests that IndE has its own intonational phonology comparable with other well-established varieties. However, similar to dialectal variation in English, BE and KE have a number of distinctive features further supporting variation within IndE.
Chapter 8: Conclusion

The main aims of this experimental study were: a) to present the first full description of the intonational phonology of IndE using the Autosegmental-Metrical framework of intonation analysis; b) to outline similarities across the speakers of BE and KE to determine the features possibly applicable to IndE in general; and c) to compare the findings with previous research on IndE and a number of other languages, including English varieties, in order to explore cross-dialectical and cross-language variation.

8.1 Summary of findings

In the introductory chapter, it was argued that with more speakers of English around the world and the growth of the world Englishes framework, further research is needed into the intonation and prosody of these varieties. IndE is one of the more longstanding varieties and is characterised by having its own grammatical, lexical, stylistic and phonological features which are mutually intelligible between IndE speakers (Bansal, 1990; Wells, 1982; Nihalani et al., 2004). English is one of the official languages in India and is used in several domains including business, media, education, literary writing, and so on. In the literature, it is referred to as a complex (Schneider, 2007), indigenised (Sridhar & Sridhar, 1992) and transplanted (Wiltshire & Harnsberger, 2006) variety due to its use as lingua franca by various linguistic communities, the influence of indigenous languages spoken in the subcontinent upon it, and the development of its own phonological system distinctly different from AmE or BrE.

Recent experimental research on prosodic (Fuchs, 2013; Krivokapić, 2013) and intonational (Moon, 2002; Wiltshire & Harnsberger, 2006) features of IndE has shown a number of differences between IndE and such well-established varieties as BrE and AmE. While some of the findings disproved earlier claims about IndE phonology, for example, the notion that IndE is spoken faster than BrE (Fuchs, 2013), others confirmed earlier impressionistic investigations and showed large variation within IndE as a variety, consistent with earlier research on both segmental and intonational phonology. One of the open questions in relation to IndE remains the degree of variation and the sources of this variation, which could be based on the speakers’ L1s, their belonging to a certain language family (for example Dravidian,
Indo-Aryan or Tibeto-Burman), their regional affiliation, or their social, economic and educational backgrounds. Contact-induced convergence across the languages spoken in the sub-continent (Gargesh, 2006) may also play a part. The present study considered educated IndE speakers of L1 Kannada (Dravidian, spoken in the south of the country) and L1 Bengali (Indo-Aryan, spoken in the north-east). The terms Bengali English and Kannada English were used loosely, with an awareness that other L1 speakers from the two regions may exhibit the features reported for each variety.

Chapter 2 outlined the fundamentals of the AM framework and English intonation, highlighting variation in intonation and prosody among well-established (such as AmE or BrE) and new (e.g. NigE or BISAfE) English varieties. Previous research has shown that the AM framework, with its approach to intonational analysis (Pierrehumbert, 1980; Ladd, 1996, 2008; Beckman & Venditti, 2011), is the most justifiable for describing variation across dialects and languages. The present study follows Pierrehumbert and Beckman’s (1988) model. Chapter 2 also discussed previous research on L2 intonation and presented a comprehensive overview of the studies on IndE prosodic features, including lexical prominence, rhythm and intonation. Further, it gave a detailed account of Bengali and Kannada linguistic profiles, and discussed the intonational features shared across the Indian languages which could potentially have contributed to the development of IndE intonational phonology.

Chapter 3 provided details on the methodology used in the study. Consistent with the majority of intonational research, the project followed a laboratory phonology approach. Four experiments were designed to investigate the following AM parameters: prominence and accentuation, phrasing, tune-to-text alignment and the use of pitch range. Four L1 speakers of Bengali and four L1 speakers of Kannada, aged 40 to 50, were audio-recorded performing a number of tasks. The speakers were well-educated medical practitioners who had lived in Australia for an average of 7.6 years at the time of data collection.

In Chapter 4, accentual and focal prominence, prosodic strategies for narrow focus, and phonetic cues to post-lexical prominence were investigated. The findings on accentuation patterns partially support previous claims about IndE. Contrary to what has been found in the literature (Bansal, 1969; Well, 1982; Gargesh, 2004; Wiltshire & Harnsberger, 2006), BE and KE speakers do not necessarily accent almost all content words and show differences in accent distribution between the two
L1 groups, with KE having a greater density of pitch accents compared to BE in broad focus structure. Accentuation was re-examined in Chapter 6 based on the analysis of complex syntactic structures, various sentence modalities and focus structures. The findings showed that in line with previous research on IndE, the speakers placed accents on some function words, such as wh-words, auxiliaries and prepositions. The patterns, however, were not consistent. Moreover, the results further confirmed greater accentual density in KE, indicating that accentual prominence placement may not be uniform across IndE speakers and may need an examination across different L1, social and educational backgrounds.

Similar to well-established Englishes, the IndE speakers in this study applied deaccenting to non-focal material in narrow focus. This strategy, however, was used alongside pitch compression or the lowering of pitch on post-focally accented words, and re-phrasing, which involves placing a focal word (at times with an adjacent word) into a separate phrase. Pitch compression is a pattern similar to that found in a number of languages spoken in the subcontinent (including Tamil, Hindi and Kolkata Bengali). This feature could be an influence of Indian languages and may be shared by IndE speakers from various L1 backgrounds. Re-phrasing has not been previously reported for IndE but has been found among L2 speakers of English from several L1 backgrounds, where it is often associated with a lower proficiency level (see Nguyên et al., 2008; Zerbian, 2013; Gut, 2005). A lack of proficiency is not the case for the speakers in this study who were fluent in English, thus suggesting that re-phrasing could also be part of IndE intonational phonology more broadly. Overall, the presence of several prosodic strategies for narrow focus marking indicates a more diverse system for the prosodic marking of focus in BE and KE compared to AmE, AusE or BrE.

The findings on the acoustic cues to prominence showed that BE and KE distinguish accentual and nuclear focal prominence. This is consistent with the canonical English intonation which uses duration, intensity, $f0$ and vowel quality to cue prominence. Despite accent and focus distinction, differences between BE and KE as well as with other IndE varieties were found both across the phonetic parameters and in terms of the extent of manipulation for each parameter. For all speakers, syllable duration was a reliable cue to accentual and focal prominence, unlike for the Telugu English (TelE) and Hindi English (HE) speakers in Moon’s (2002) study. Duration was closely followed by $f0$ height as an indicative prominence
marker, with $f_0$ excursion being a more accurate measure than the absolute $f_0$ height, especially for KE speakers. Similar to TelE and HE (Moon, 2002), RMS amplitude was used to signal accentual and focal prominence, with the exception of two KE speakers who did not distinguish between accent and focus based on this parameter. This could be linked to its more subtle use of in Kannada. In her study of Kannada speech rhythm, Savithri (1995) reported that only half of the syllables perceived as being louder showed an actual increase in intensity. Contrary to well-described Englishes, vowel quality was not found to be a reliable cue to focal prominence in BE and KE. Moreover, even the distinction between stressed and accented syllables appeared to be rather marginal, being maintained for some vowels but not others, with a large degree of inter-speaker variation regardless of L1. Back vowels showed the most overlap in the F1 and F2 space between stressed and accented syllables.

Chapter 5 examined the rising gesture on prenuclear and nuclear accented words, and showed that BE and KE speakers use pitch accent categories found in several varieties of English, but with differences in distribution and function. In BE, the L*+H accent was used on nuclear focal as well as prenuclear words, and the L+H* was used in prenuclear position. Consistent use of L*+H in narrow focus by BE speakers may be the result of L1 influence or the mapping of the closest phonological category that would show a delayed rising gesture (Khan, 2008, 2014 on L*+H in Bangladeshi Bengali; Hayes & Lahiri, 1991 on L* Hp in Kolkata Bengali). KE did not maintain the L*+H and L+H* distinction, showing similarity to the single rising accent found in Donegal English (Kalaldeh et al., 2009) and Glasgow English (Mayo et al., 1997). Based on the alignment patterns, L+H* was posited as the preferred rising accent for KE. There is a possibility that this use of a single rising accent may be due to a smaller pitch accent inventory in the speakers’ L1.

The findings reported in Chapter 5 were also consistent with previous research on tonal alignment of pitch accents (Silverman & Pierrehumbert, 1990; Prieto et al., 1995; Prieto & Torreira, 2007; Ladd et al., 2009; D’Imperio, 2012). It was shown that the rising gesture aligned with segmental landmarks, was not of constant slope and duration, and showed strong correlation with the accented syllable duration. Differences were found in the phonetic realisation of the rising accents in comparison to the same categories in varieties of AmE and BrE (see Pierrehumbert, 1980; Arvaniti & Garding, 2007; Ladd et al., 2009) and other languages (see Arvaniti et al., 1998; Atterer & Ladd, 2004; Face, 2002), such as the anchoring of the L tone to
the accented syllable onset in both L+H* and L*+H, and the lack of a clear valley in L*+H. The phonetic realisation of the same phonological categories could have been the result of L1 influence: phonetic aspects are more likely to show variability in L2 intonation even when the same categories are present in L1 and L2 (Mennen, 2004).

The tonal alignment experiment in Chapter 5 also showed that the rise on accented words was not across the prosodic constituent but localised to a stressed syllable, consistent with many other English varieties. This finding disproved the presence of a minor prosodic constituent roughly the size of a word in BE and KE. Chapter 6 examined boundary-related lengthening to determine the number of prosodic constituents above the word. First, the findings confirmed that similar to English (Edwards et al., 1991; Wightman et al., 1992; Turk & Shattuck-Hufnagel, 2007), IndE speakers in this study relied on final lengthening at phrase edges. Second, the speakers differentiated between minor and major prosodic levels, in the form of the intermediate phrase and the intonational phrase, by manipulating the degree of pre-boundary lengthening as well as pitch range reset, suggesting that IndE (or at least the varieties analysed in this study) may have a prosodic structure similar to other well-established varieties (such as AmE and AusE). Variation across the two L1 groups was found in the amount of boundary-related lengthening, with BE showing greater durational increase on phrase-final syllables.

The tone inventory, intonational contours and nuclear tunes across different modalities for KE and BE were described in Chapter 7. Both varieties showed similarity in phonological categories associated with phrase boundaries, with three phrase accents (H-, !H- and L-) and two boundary tones (H% and L%). Further research is needed to look into the phonetic realisation of these tones in spontaneous speech. The pitch accent category varied between KE and BE. While KE included H*, L*, L+H* and !H*, the BE pitch accent inventory featured these accents in addition to L*+H and downstepped counterparts for rising pitch accents (L+!H*, L*+!H). The rising accents were used in nuclear and prenuclear positions for all speakers but constituted the highest rate of distribution in BE. In KE, L+H* was used on nuclear focal words and prenuclear phrase-initial words, while H* was by far the most common pitch accent in nuclear non-focal position. The lack of research on Kannada intonation does not allow for a comparison with the speakers’ L1. However, taking into account a smaller tonal inventory in Tamil (Keane, 2014) compared to Bengali
(Hayes & Lahiri, 1991; Michaels & Nelson, 2004; Khan, 2008, 2014), the differences between KE and BE may be related to the speakers’ L1.

Fewer pitch accents and a frequent use of nuclear rising accents in BE compared to the nuclear simple high accents of KE in non-focal contexts explain the differences in major nuclear tunes between the two varieties and with a number of well-established Englishes. In statements, BE showed frequent use of rise-falls (L*+H L-L% or L+H* L-L%), rises (L* L-H% or L* H-H*) and rise-fall-rises (L+H* L-H% or L*+H L-H%) across several modalities including declaratives, wh- and polar questions, the contours found in Bengali dialects (Hayes & Lahiri, 1991; Khan, 2014). Falling nuclear tunes (H* L-L%) were more characteristic for KE declaratives, similar to English spoken by bilingual Punjabi-English speakers from Bradford (Grabe, 2004; Grabe et al., 2005). Interrogative intonation included several tunes for both varieties but also showed frequent use of falls in KE and rise-falls in BE. Polar question rises with the H* target (H* H-H%) were more typical for KE, while BE polar questions were realised with a nuclear L* (L* H-H%), the latter also common for polar questions in English spoken by Punjabi-English bilinguals (Grabe, 2004). Moreover, low onset rises (L* H-H%) may be part of BE declarative intonation, similar to AusE, Glasgow English and Belfast English (Fletcher & Harrington, 2001; Fletcher et al., 2005; Grabe et al., 2005).

Although these results were not directly compared to well-established Englishes, previous claims that all IndE speakers use a narrower pitch range compared to BrE or AmE have found partial support in this study. The investigation of $f0$ ranges for the speakers in this study revealed that despite relatively high pitch level (overall $f0$ height) for most speakers, pitch span (the amount of $f0$ modulation within a phrase) varied significantly based on the L1 group. KE speakers produced shallower rises throughout the corpus compared to the high rises produced by the BE group. As suggested by Mennen et al. (2012), different $f0$ ranges could be the combined result of accentuation and pitch accent type. KE mostly relied on H* and had higher accentual density, while BE intonation displayed a great number of rising pitch accents (L+H* or L*+H) in an intonational phrase with a smaller amount of accented words.

8.2 Implications
The present study has contributed to the research on IndE intonational phonology and presented the first detailed account of the intonational and prosodic features of educated IndE speakers, based on Bengali and Kannada L1 backgrounds. Through a series of experiments, several phonetic phenomena within the AM framework were investigated to provide and expand the understanding of IndE intonation neglected in previous experimental research (Wells, 1982; Gargesh, 2004). Tonal categories, nuclear tunes, phrasing and prosodic structure have been among the least researched aspects. The laboratory phonology approach and the use of the AM framework have enabled a determination of IndE phonological categories on the basis of phonetic criteria and have for the first time added a detailed description of IndE intonation to the body of work on intonation across WE.

Several similarities found between BE and KE confirm that IndE has developed its own intonational phonology and suggest that English spoken in India has some resemblance to well-established Englishes. For example, the results showed that the prosodic constituency of both BE and KE includes the intermediate phrase and the intonational phrase, and that both varieties have an inventory of phrase accents and boundary tones found in AmE and AusE. However, accentuation, prosodic marking of focus, the distribution of pitch accents and nuclear tunes across various modalities suggests that IndE intonational phonology presents a somewhat ‘hybrid’ system, possibly as the result of the influence of indigenous languages as well as its development throughout history.

According to Jun’s (2005b) prosodic typology, languages can be grouped into head-prominence, edge-prominence and head/edge-prominence marking. The examination of prominence and phrasing patterns in the current research showed that for the most part KE and BE speakers used deaccenting in narrow focus, found in canonical English intonation (BrE, AmE and AusE). However, the use of pitch compression without deaccenting, common for several Indian languages, together with the use of re-phrasing suggest that IndE may not be fully classified as a head-prominence language. Reliance on phrasing in both BE and KE could be similar to Kolkata Bengali, Hindi and Tamil, all of which fall into the category of head/edge-prominence marking.

IndE speakers in this study distinguish accented syllables from nuclear focal, but the reliable phonetic cues to prominence only include three parameters (duration, RMS-amplitude and f0) and exclude vowel quality sharpening. This raises the issue of
whether IndE speakers produce reduced vowels in weak syllables, and is related to the phonetic realisation of lexical prominence which has been found to be different from AmE (Pickering & Wiltshire, 2000; Wiltshire & Moon, 2003). Similarly, the high and rising pitch accents posited for IndE speakers in this study are commonly found in well-established Englishes. However, the use of pitch accents in well-established varieties of English is more flexible and not restricted to a particular pitch accent type for focus marking, contrary to the L+H* in KE and L*+H in BE used for narrow focus.

The present study has demonstrated that despite L1 influence, IndE features may not always be explained in terms of the processes applicable to L2 intonation, and therefore IndE should not be treated as an L2 variety (Sridhar & Sridhar, 1986; Sharma, 2005; Kachru & Nelson, 2006). L1 is just one of many factors contributing to IndE phonology. The fact that the highly educated and fluent speakers in this study still placed accents on more words and at times grouped utterances into more phrases in ways often linked to a lower level of proficiency in L2 studies (Hellmuth, 2010; Ueyama & Jun, 1998) indicate that these features could be part of IndE intonational phonology.

The present study has also extended the knowledge of intonational variation within IndE. The findings support the more recent experimental research suggesting that English spoken in India is not uniform (Wiltshire, 2005; Wiltshire & Harnsberger, 2006; Maxwell & Fletcher, 2009, 2010; Sirsa & Redford, 2013). Differences have been found between the two varieties described in this study as well as with other IndE varieties (i.e., TE, HE or TelE). Such variation is common for the dialects of AmE and BrE and have been extensively studied in the literature.

An examination of the intonational features of BE and KE in relation to Ladd’s (2008) typology, which was developed to describe intonational differences among languages and dialects, shows that differences between the two varieties can be found in three categories. These include:

a) semantic, such as L*+H (BE) and L+H* (KE) used in narrow focus, or the use of the same tune in different contexts, for example L* H-H% to signal both statements and questions in BE and restricted to questions in KE;

b) systemic, such as the lack of an L+H* or L*+H contrast and a smaller inventory of tones and tunes in KE, together with the use of rise-fall-rises (L*+H L-H% and L+H* L-H%) in BE; and
c) realisational, for example, the extent of boundary-related lengthening, the degree of $f0$ modulation on accented words and the use of pitch range in BE and KE, and also evidence of a reversal of declination in wh-questions and declaratives in BE.

The findings indicate that there is a possibility of discerning typological distinctions within intonational phonology of IndE, not just phonetic variation.

Differences were also found according to Jun’s (2012) extended prosodic typology which includes the criteria of macro-rhythm, a tonal rhythm perceived by the changes in $f0$ within an intonational phrase, defined on the basis of a) the inventory of possible phrase-medial pitch accents; b) the type of pitch accent or boundary tone; and c) the number of pitch accents or boundary tones. Based on these three parameters, KE would be categorised as having a stronger macro-rhythm, while BE would belong to the group with a medium micro-rhythm, as a result of having few accented words and a larger tonal inventory. In this way, it is more consistent with AusE, AmE and several dialects of BrE.

More generally, this study contributes to the research on intonation and provides evidence that certain processes in intonation and prosody may be applicable to a wide range of languages and varieties. IndE exhibits a number of features that have been found across typologically distinct languages. These include boundary-related lengthening, pitch expansion on initial pitch accents in polar questions and imperatives compared to declaratives, a wide range of tunes used in declarative and interrogative intonation, a frequent use of a falling boundary configuration in declaratives with patterns of downdrift and final lowering, prosodic marking of focus, and the alignment of tonal targets in relation to the segmental landmarks.

8.3 Future directions

Despite the obvious contribution this study makes to the field of intonational phonology, the results indicate a number of possibilities for future research. One of the ongoing debates on IndE phonology concerns the sort of features that could be applicable to the variety in general, together with the degree and source of variation within it. Given that this first description of IndE intonational phonology is based on speakers from two L1s with similar social and educational backgrounds, more experimental research is needed involving speakers with various L1s and different
geographical affiliations from a range of socio-economic and educational backgrounds. Building on the present findings and experimental research by Moon (2002) and Wiltshire and Harnsberger (2006), the prosodic structure, prominence, tone and tune inventory need further investigation to determine the typology of IndE within WE.

Unlike Fuchs’ (2013) investigation of IndE rhythm, the production experiments did not include a direct comparison of speech data with any of the well-established Englishes. Moreover, conclusions about L1 influence in KE were made on the basis of the assumption that a more likely source of variation between BE and KE would be the speakers’ L1s. Future comparison with the intonation of Kannada will shed more light on the findings reported in the study. Additionally, lexical prominence was not part of this research and needs to be examined in view of the present findings, with a particular focus on vowel quality. The precise phonetic alignment of nuclear non-focal rises (L*+H) in BE remains to be investigated. A semantic analysis of L+H* and L*+H for BE will be useful to determine whether there are differences in meaning between the two pitch accents in non-focal contexts, as is the case in other well-established Englishes (Pierrehumbert & Hirschberg, 1990). In addition, perception experiments may provide additional support for the interpretation of the results and the phonological categories posited in the study.

On a final note, another area worth considering and not addressed in the present study is the use of intonation across dialects and varieties. As argued by Gumperz (1982) and Wiltshire and Harnsberger (2006), differences between IndE and well-established varieties of English can be linked to the issues of intelligibility and overall success in communication. Falling intonation and narrow pitch range use in KE, and frequent use of L*+H and delayed rise-falls in BE could have an effect on communication. Spontaneous speech data from this study can be used to conduct a more detailed investigation of intonation and its meaning in different contexts. This will allow for a more comprehensive comparison of IndE with other English varieties and will help further determine systemic, semantic or realisational differences, and capture which features and to what degree may potentially contribute to miscommunication between the speakers of IndE and other Englishes.
References


Language and education in India. In J. Bourne & E. Reid, (Eds.), Language
education: World yearbook of education 2003 (pp. 227-238). London, UK:
Kogan Page.
CIEFL.
Clark, J., Yallop, C., & Fletcher, J. (2007). An introduction to phonetics and
Cohn, A. (2010). Laboratory Phonology: Past successes and current questions,
challenges, and goals. In C. Fougeron, B. Kühnert, M. D’Imperio, & N. Vallée
(Eds.), Laboratory Phonology 10 (pp. 3-30). Berlin: De Gruyter.
Cohn, C., Fougeron, C., & Huffman, M. (Eds.). (2012). The Oxford Handbook of
(Eds.), The Blackwell Companion to Phonology (Vol 2) (pp. 824-847). Oxford,
UK: Wiley-Blackwell.
7, 1-29.
question answer contexts. Journal of the Acoustical Society of America, 77(6),
2142-2156.
labelling in the Australian National Database of Spoken Language. Retrieved
Crystal, T., & House, A. (1988). Segmental durations in connected-speech signals:
Crystal, T., & House, A. (1990). Articulation rate and the duration of syllables and
stress groups in connected speech. Journal of the Acoustical Society of America,
88, 101-112.


D’Imperio, M. (2002). Language-specific and universal constraints on tonal alignment: The nature of targets and ‘anchors’. In B. Bel & I. Marlien (Eds.),


Esposito, C., & Barjam, P. (2007). The intonation of question in Farsi: Wh-questions, yes/no-questions, and echo questions, unpublished manuscript, Department of Linguistics, UCLA.


Fletcher, J. (2004, December). E-ToBI Intonational Annotation. Workshop presented at the Australian Language Technology Association Summer School, Macquarie University, North Ryde, NSW.


Appendices

Appendix A

The appendix includes read speech materials relating to four experiments presented in chapters 4, 5, 6 and 7.

Experiment 1

Sentence template

Lee/Nina/Marina may move/borrow/minimise my mill/lily/umbrella.
Lee/Nina/Marina may move/borrow/minimise my mill/lily/umbrella.
Lee/Nina/Marina may move/borrow/minimise my mill/lily/umbrella.

Example of a sentence with a prompt question and the corresponding focus structure

Who may move my mill?  
Lee may move my mill.  
F – subject
What may Lee do?  
Lee may move my mill.  
F – verb
What may Lee move?  
Lee may move my mill.  
F – object
What did you say?  
Lee may move my mill.  
F – broad

Experiment 2

Set 1 - Prenuclear accents

Lara lives in Lilydale.
Lulu lived in Melbourne.
Nelly likes yellow lemons.
Maryellen may feel mellow.

Set 2 – Nuclear focal accents

Lulu/Maryellen memorised Manuela’s main/luminous/yellow-coloured/manual.

Example of a sentence with a prompt question


Experiment 3

Sentence template

Lulu/Emily/Marina/Maryellen memorised Manuela’s main/luminous/yellow-coloured/manual.

Example of a sentence with a prompt question

Experiment 4

Simple declaratives
- Lara lives in Lilydale.
- Lulu lived in Melbourne.
- My niece Maya knows Mary.
- Maryellen may feel mellow.
- Nelly likes yellow lemons.

Complex declaratives
- My niece Maya knows Mary who lived in Melbourne.
- Lara lives in Lilydale, which is close to Melbourne and far from Narooma.
- Mary knows that Nelly likes yellow lemons with his meal.
- Maryellen may feel mellow even when in Melbourne.
- My niece Maya remembered the lemon tree my mum had planted in early May.

Polar questions
- Does Lara live in Lilydale?
- Did Mary live in Melbourne?
- Does Nelly like yellow lemons?
- Will Maryellen feel mellow in May?
- Will my niece know where to find Mary?

Wh-questions
- Why will Mary feel mellow?
- What does Mary know about Nelly?
- Who lives in Lilydale now?
- What does my niece Maya know?
- When did Mary use to live in Melbourne?

Coordination questions
- Does Nelly like lime or lemons?
- Did Mary live in Melbourne or Narooma?
- Does my niece know Mary or Nelly?
- Will Maryellen feel mellow or hollow?

Imperatives
- He said to Mary, “Clean some yellow lemons”.
Appendix B: LMM analyses for the results in Chapter 4

Table B.1 LMM results for the absolute $f_0$ height in stressed, accented and focal syllables.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>15.1472</td>
<td>3.8919</td>
</tr>
<tr>
<td>Residual</td>
<td>5.6023</td>
<td>2.3669</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>8.5280</td>
<td>1.9732</td>
<td>4.322</td>
<td>0.001</td>
</tr>
<tr>
<td>Condition - Focal</td>
<td>3.9939</td>
<td>0.1943</td>
<td>20.557</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Stressed</td>
<td>-4.5829</td>
<td>0.1937</td>
<td>-23.664</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-2.2965</td>
<td>2.7571</td>
<td>-0.833</td>
<td>0.404</td>
</tr>
<tr>
<td>Position - Subject</td>
<td>1.1949</td>
<td>0.1965</td>
<td>6.082</td>
<td>0.000</td>
</tr>
<tr>
<td>Position - Verb</td>
<td>0.4376</td>
<td>0.2910</td>
<td>1.504</td>
<td>0.132</td>
</tr>
<tr>
<td>Vowel /e/</td>
<td>0.5071</td>
<td>0.3693</td>
<td>1.373</td>
<td>0.169</td>
</tr>
<tr>
<td>Vowel /iː/</td>
<td>0.8640</td>
<td>0.3764</td>
<td>2.296</td>
<td>0.02</td>
</tr>
<tr>
<td>Vowel /sit/</td>
<td>0.3376</td>
<td>0.3122</td>
<td>1.081</td>
<td>0.279</td>
</tr>
<tr>
<td>Vowel /uː/</td>
<td>-0.6211</td>
<td>0.2517</td>
<td>-2.467</td>
<td>0.01</td>
</tr>
<tr>
<td>Vowel /u/</td>
<td>-0.1006</td>
<td>0.2447</td>
<td>-0.411</td>
<td>0.6</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>-2.0186</td>
<td>0.2856</td>
<td>-7.067</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition Stressed x Group</td>
<td>0.1319</td>
<td>0.2809</td>
<td>0.470</td>
<td>0.6</td>
</tr>
</tbody>
</table>

AIC 8100  BIC 8182  logLik -4035

Table B.2 Mean and standard deviation of the absolute $f_0$ values (Hz) in focal versus accented syllables presented in three prosodic positions (subject, verb, object) for each speaker. Shaded grey areas indicate that the differences were non-significant.

<table>
<thead>
<tr>
<th></th>
<th>Subject</th>
<th></th>
<th>Verb</th>
<th></th>
<th>Object</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Focal</td>
<td>Accented</td>
<td>Focal</td>
<td>Accented</td>
<td>Focal</td>
<td>Accented</td>
</tr>
<tr>
<td>B1</td>
<td>188</td>
<td>143</td>
<td>171</td>
<td>125</td>
<td>176</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>SD=20</td>
<td>SD=10</td>
<td>SD=24</td>
<td>SD=8</td>
<td>SD=20</td>
<td>SD=10</td>
</tr>
<tr>
<td>B2</td>
<td>220</td>
<td>159</td>
<td>211</td>
<td>145</td>
<td>215</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>SD=13</td>
<td>SD=29</td>
<td>SD=33</td>
<td>SD=18</td>
<td>SD=27</td>
<td>SD=32</td>
</tr>
<tr>
<td>B3</td>
<td>193</td>
<td>157</td>
<td>195</td>
<td>153</td>
<td>239</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>SD=17</td>
<td>SD=11</td>
<td>SD=17</td>
<td>SD=11</td>
<td>SD=47</td>
<td>SD=23</td>
</tr>
<tr>
<td>B4</td>
<td>327</td>
<td>279</td>
<td>269</td>
<td>182</td>
<td>257</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>SD=15</td>
<td>SD=23</td>
<td>SD=41</td>
<td>SD=20</td>
<td>SD=49</td>
<td>SD=45</td>
</tr>
<tr>
<td>K1</td>
<td>186</td>
<td>120</td>
<td>176</td>
<td>149</td>
<td>118</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>SD=28</td>
<td>SD=9</td>
<td>SD=17</td>
<td>SD=12</td>
<td>SD=10</td>
<td>SD=9</td>
</tr>
<tr>
<td>K2</td>
<td>133</td>
<td>141</td>
<td>132</td>
<td>137</td>
<td>149</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>SD=33</td>
<td>SD=14</td>
<td>SD=14</td>
<td>SD=14</td>
<td>SD=18</td>
<td>SD=26</td>
</tr>
<tr>
<td>K3</td>
<td>191</td>
<td>176</td>
<td>172</td>
<td>182</td>
<td>144</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>SD=18</td>
<td>SD=41</td>
<td>SD=22</td>
<td>SD=26</td>
<td>SD=5</td>
<td>SD=12</td>
</tr>
<tr>
<td>K4</td>
<td>173</td>
<td>166</td>
<td>132</td>
<td>144</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD=12</td>
<td>SD=9</td>
<td>SD=14</td>
<td>SD=14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.3 LMM results for the comparison of pitch excursion in focal and accented syllables.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>1.3125</td>
<td>1.1456</td>
</tr>
<tr>
<td>Residual</td>
<td>5.0471</td>
<td>2.2466</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.01541</td>
<td>0.59205</td>
<td>8.471</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Focal</td>
<td>4.23771</td>
<td>0.18596</td>
<td>22.788</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-1.34537</td>
<td>0.82790</td>
<td>-1.625</td>
<td>0.104</td>
</tr>
<tr>
<td>Position - Subject</td>
<td>-0.05112</td>
<td>0.15895</td>
<td>-0.322</td>
<td>0.747</td>
</tr>
<tr>
<td>Position - Verb</td>
<td>-0.97133</td>
<td>0.17651</td>
<td>-5.503</td>
<td>0.001</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>-2.80765</td>
<td>0.27268</td>
<td>-10.297</td>
<td>0.000</td>
</tr>
</tbody>
</table>

AIC | 5362 |
BIC | 5402 |
logLik | -2673 |

Table B.4 LMM results for the RMS-amplitude measurements in stressed, accented and focal syllables.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>48.203</td>
<td>6.9428</td>
</tr>
<tr>
<td>Residual</td>
<td>9.703</td>
<td>3.1150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>67.1128</td>
<td>3.4972</td>
<td>19.190</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Focal</td>
<td>2.6979</td>
<td>0.2557</td>
<td>10.552</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Stressed</td>
<td>-5.1170</td>
<td>0.2549</td>
<td>-20.077</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>0.3462</td>
<td>4.9132</td>
<td>0.070</td>
<td>0.9</td>
</tr>
<tr>
<td>Position - Subject</td>
<td>1.719</td>
<td>0.2586</td>
<td>6.649</td>
<td>0.000</td>
</tr>
<tr>
<td>Position - Verb</td>
<td>1.4031</td>
<td>0.3829</td>
<td>3.664</td>
<td>0.001</td>
</tr>
<tr>
<td>Vowel /ɛ/</td>
<td>0.2477</td>
<td>0.4861</td>
<td>0.510</td>
<td>0.6</td>
</tr>
<tr>
<td>Vowel /i/</td>
<td>-0.6539</td>
<td>0.4954</td>
<td>-1.320</td>
<td>0.1</td>
</tr>
<tr>
<td>Vowel /u/</td>
<td>-0.2660</td>
<td>0.4109</td>
<td>-0.647</td>
<td>0.6</td>
</tr>
<tr>
<td>Vowel /ɔ/</td>
<td>0.6490</td>
<td>0.3313</td>
<td>1.959</td>
<td>0.05</td>
</tr>
<tr>
<td>Vowel /u(ː)/</td>
<td>-1.9678</td>
<td>0.3220</td>
<td>-6.111</td>
<td>0.001</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>0.1565</td>
<td>0.3759</td>
<td>0.416</td>
<td>0.677</td>
</tr>
<tr>
<td>Condition Stressed x Group</td>
<td>-1.4985</td>
<td>0.3697</td>
<td>-4.054</td>
<td>0.002</td>
</tr>
</tbody>
</table>

AIC | 9063 |
BIC | 9145 |
logLik | -4516 |
Table B.5 LMM results for syllable duration in stressed, accented and focal syllables.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>1264.8</td>
<td>35.564</td>
</tr>
<tr>
<td>Residual</td>
<td>3148.7</td>
<td>56.113</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>247.212</td>
<td>18.121</td>
<td>13.64</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Focal</td>
<td>42.754</td>
<td>3.358</td>
<td>12.73</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Stressed</td>
<td>-18.035</td>
<td>3.576</td>
<td>-5.04</td>
<td>0.001</td>
</tr>
<tr>
<td>Target word – one syllable</td>
<td>186.405</td>
<td>3.400</td>
<td>54.82</td>
<td>0.000</td>
</tr>
<tr>
<td>Target word – three syllables</td>
<td>51.667</td>
<td>3.163</td>
<td>16.34</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>6.856</td>
<td>25.279</td>
<td>0.27</td>
<td>0.8</td>
</tr>
<tr>
<td>Position - Subject</td>
<td>-7.872</td>
<td>3.659</td>
<td>-2.15</td>
<td>0.01</td>
</tr>
<tr>
<td>Position - Verb</td>
<td>18.078</td>
<td>3.230</td>
<td>5.60</td>
<td>0.001</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>5.990</td>
<td>6.981</td>
<td>-0.243</td>
<td>0.8</td>
</tr>
</tbody>
</table>

AIC 19168  BIC 19223  logLik -9574

Table B.6 LMM results for F1 measurements in stressed, accented and focal vowels.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>801.41</td>
<td>28.309</td>
</tr>
<tr>
<td>Residual</td>
<td>3089.67</td>
<td>55.585</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>563.315</td>
<td>15.157</td>
<td>37.17</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition – Focal</td>
<td>2.213</td>
<td>4.553</td>
<td>0.49</td>
<td>0.6</td>
</tr>
<tr>
<td>Condition - Stressed</td>
<td>27.473</td>
<td>4.412</td>
<td>6.23</td>
<td>0.0001</td>
</tr>
<tr>
<td>Group</td>
<td>9.769</td>
<td>20.454</td>
<td>0.48</td>
<td>0.6</td>
</tr>
<tr>
<td>Vowel /ɛ/</td>
<td>-54.761</td>
<td>6.544</td>
<td>-8.37</td>
<td>0.000</td>
</tr>
<tr>
<td>Vowel /i/</td>
<td>-270.784</td>
<td>5.705</td>
<td>-47.46</td>
<td>0.000</td>
</tr>
<tr>
<td>Vowel /u/</td>
<td>-228.516</td>
<td>4.998</td>
<td>-45.72</td>
<td>0.000</td>
</tr>
<tr>
<td>Vowel /uː/</td>
<td>-44.803</td>
<td>5.780</td>
<td>-7.75</td>
<td>0.0001</td>
</tr>
<tr>
<td>Vowel /u拗/</td>
<td>-284.582</td>
<td>5.632</td>
<td>-50.53</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>-6.989</td>
<td>6.698</td>
<td>-1.04</td>
<td>0.3</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>-29.290</td>
<td>6.571</td>
<td>-4.46</td>
<td>0.001</td>
</tr>
</tbody>
</table>

AIC 19117  BIC 19188  logLik -9545
Table B.7 LMM results for F2 measurements in stressed, accented and focal vowels.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Groups</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speaker (Intercept)</td>
<td>3017.2</td>
<td>54.929</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>23574.2</td>
<td>153.539</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1769.29</td>
<td>31.28</td>
<td>56.57</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition - Focal</td>
<td>25.96</td>
<td>12.58</td>
<td>2.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Condition Stressed</td>
<td>-99.98</td>
<td>12.18</td>
<td>-8.21</td>
<td>0.001</td>
</tr>
<tr>
<td>Group</td>
<td>-81.69</td>
<td>40.54</td>
<td>-2.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Vowel /ɛ/</td>
<td>-26.02</td>
<td>18.07</td>
<td>-1.44</td>
<td>0.2</td>
</tr>
<tr>
<td>Vowel /i/</td>
<td>546.93</td>
<td>15.75</td>
<td>34.73</td>
<td>0.000</td>
</tr>
<tr>
<td>Vowel /ɪ/</td>
<td>359.13</td>
<td>13.79</td>
<td>26.04</td>
<td>0.000</td>
</tr>
<tr>
<td>Vowel /ɔ/</td>
<td>-734.86</td>
<td>15.96</td>
<td>-46.04</td>
<td>0.000</td>
</tr>
<tr>
<td>Vowel /u(ː)/</td>
<td>-976.38</td>
<td>15.55</td>
<td>-62.79</td>
<td>0.000</td>
</tr>
<tr>
<td>Condition Focal x Group</td>
<td>-11.42</td>
<td>18.50</td>
<td>-0.62</td>
<td>0.6</td>
</tr>
<tr>
<td>Condition stressed x Group</td>
<td>72.84</td>
<td>18.14</td>
<td>4.02</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>22665</td>
<td>22736</td>
<td>-11319</td>
</tr>
</tbody>
</table>
Appendix C: LMM analyses for the results in Chapter 5

Table C.1 LMM results for the temporal interval from syllable onset to L (ContoL) in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0038343</td>
<td>0.061922</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0082042</td>
<td>0.090577</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.296796</td>
<td>0.036995</td>
<td>8.023</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.03552</td>
<td>0.052319</td>
<td>-0.679</td>
<td>0.5</td>
</tr>
<tr>
<td>Word - Lulu</td>
<td>0.006879</td>
<td>0.028643</td>
<td>0.240</td>
<td>0.8</td>
</tr>
<tr>
<td>Word - Maryellen</td>
<td>-0.04511</td>
<td>0.030563</td>
<td>-1.476</td>
<td>0.1</td>
</tr>
<tr>
<td>Word - Nelly</td>
<td>-0.03309</td>
<td>0.030070</td>
<td>-1.000</td>
<td>0.3</td>
</tr>
<tr>
<td>Group x Word Lulu</td>
<td>0.024525</td>
<td>0.040784</td>
<td>0.601</td>
<td>0.6</td>
</tr>
<tr>
<td>Group x Word Maryellen</td>
<td>0.077295</td>
<td>0.043132</td>
<td>1.972</td>
<td>0.05</td>
</tr>
<tr>
<td>Group x Word Nelly</td>
<td>0.147992</td>
<td>0.042088</td>
<td>3.516</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>-214.6</td>
<td>-184.7</td>
<td>117.3</td>
</tr>
</tbody>
</table>

Table C.2 LMM results for the temporal interval from vowel onset to L (VontoL) in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0055724</td>
<td>0.074648</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0092650</td>
<td>0.096255</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.16957</td>
<td>0.04308</td>
<td>-3.936</td>
<td>0.001</td>
</tr>
<tr>
<td>Group</td>
<td>-0.04264</td>
<td>0.06093</td>
<td>-0.700</td>
<td>0.5</td>
</tr>
<tr>
<td>Word - Lulu</td>
<td>-0.08986</td>
<td>0.03044</td>
<td>-2.952</td>
<td>0.03</td>
</tr>
<tr>
<td>Word - Maryellen</td>
<td>-0.06185</td>
<td>0.03248</td>
<td>-1.904</td>
<td>0.057</td>
</tr>
<tr>
<td>Word - Nelly</td>
<td>-0.08190</td>
<td>0.03196</td>
<td>-2.563</td>
<td>0.02</td>
</tr>
<tr>
<td>Group x Word Lulu</td>
<td>0.07243</td>
<td>0.04334</td>
<td>1.671</td>
<td>0.09</td>
</tr>
<tr>
<td>Group x Word Maryellen</td>
<td>0.09026</td>
<td>0.04584</td>
<td>1.969</td>
<td>0.05</td>
</tr>
<tr>
<td>Group x Word Nelly</td>
<td>0.14458</td>
<td>0.04473</td>
<td>3.232</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>-196.4</td>
<td>-166.6</td>
<td>108.2</td>
</tr>
</tbody>
</table>
Table C.3 LMM results for the temporal interval from syllable onset to H (ContoH) in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Groups</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0085219</td>
<td>0.092314</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.0254886</td>
<td>0.159651</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Vowel duration</td>
</tr>
<tr>
<td>Word - Lulu</td>
</tr>
<tr>
<td>Word - Maryellen</td>
</tr>
<tr>
<td>Word - Nelly</td>
</tr>
</tbody>
</table>

AIC: -60.32  BIC: -36.45  logLik: 38.16

Table C.4 LMM results for the temporal interval from H to word offset (WofftoH) in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Groups</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>2671.8</td>
<td>51.690</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>2225.1</td>
<td>47.171</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Word - Lulu</td>
</tr>
<tr>
<td>Word - Maryellen</td>
</tr>
<tr>
<td>Word - Nelly</td>
</tr>
</tbody>
</table>

AIC: 1538  BIC: 1558  logLik: -761.8

Table C.5 LMM results for the distance between L and H (LtH) in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Groups</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0058328</td>
<td>0.076373</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.0276919</td>
<td>0.166409</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Word - Lulu</td>
</tr>
<tr>
<td>Word - Maryellen</td>
</tr>
<tr>
<td>Word - Nelly</td>
</tr>
</tbody>
</table>

AIC: -64.19  BIC: -43.3  logLik: 39.09
### Table C.6 LMM results for L scaling in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>9.15841</td>
<td>3.02629</td>
</tr>
<tr>
<td>Residual</td>
<td>0.75932</td>
<td>0.87139</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.3628</td>
<td>1.5210</td>
<td>3.526</td>
<td>0.004</td>
</tr>
<tr>
<td>Group</td>
<td>-0.5989</td>
<td>2.1446</td>
<td>-0.279</td>
<td>0.8</td>
</tr>
<tr>
<td>Word - Lulu</td>
<td>-1.0703</td>
<td>0.1962</td>
<td>-5.456</td>
<td>0.0001</td>
</tr>
<tr>
<td>Word - Maryellen</td>
<td>-1.0487</td>
<td>0.2076</td>
<td>-5.052</td>
<td>0.0001</td>
</tr>
<tr>
<td>Word - Nelly</td>
<td>-1.1407</td>
<td>0.2025</td>
<td>-5.633</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>425.6</td>
<td>446.5</td>
<td>-205.8</td>
</tr>
</tbody>
</table>

### Table C.7 LMM results for H scaling in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>12.0441</td>
<td>3.4705</td>
</tr>
<tr>
<td>Residual</td>
<td>1.6451</td>
<td>1.2826</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>11.1481</td>
<td>1.7502</td>
<td>6.370</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-2.4854</td>
<td>2.4631</td>
<td>-1.009</td>
<td>0.4</td>
</tr>
<tr>
<td>Word - Lulu</td>
<td>-1.0924</td>
<td>0.2887</td>
<td>-3.783</td>
<td>0.001</td>
</tr>
<tr>
<td>Word - Maryellen</td>
<td>-2.1478</td>
<td>0.3055</td>
<td>-7.030</td>
<td>0.0001</td>
</tr>
<tr>
<td>Word - Nelly</td>
<td>-1.8462</td>
<td>0.2980</td>
<td>-6.195</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>531.6</td>
<td>552.5</td>
<td>-258.8</td>
</tr>
</tbody>
</table>

### Table C.8 LMM results for the scaling difference between L and H in prenuclear rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>1.8827</td>
<td>1.3721</td>
</tr>
<tr>
<td>Residual</td>
<td>1.4765</td>
<td>1.2151</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>5.78918</td>
<td>0.71957</td>
<td>8.045</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-1.89417</td>
<td>0.99109</td>
<td>-1.911</td>
<td>0.056</td>
</tr>
<tr>
<td>Word - Lulu</td>
<td>-0.02191</td>
<td>0.27354</td>
<td>-0.080</td>
<td>1</td>
</tr>
<tr>
<td>Word - Maryellen</td>
<td>-1.08991</td>
<td>0.28940</td>
<td>-3.766</td>
<td>0.004</td>
</tr>
<tr>
<td>Word - Nelly</td>
<td>-0.69618</td>
<td>0.28232</td>
<td>-2.466</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>506.1</td>
<td>527</td>
<td>-246.1</td>
</tr>
</tbody>
</table>
Table C.9 LMM results for the temporal interval from syllable onset to L (ContoL) in nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0049616</td>
<td>0.070439</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0079672</td>
<td>0.089259</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.276640</td>
<td>0.036129</td>
<td>7.657</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.027225</td>
<td>0.050825</td>
<td>-0.536</td>
<td>0.6</td>
</tr>
<tr>
<td>Word</td>
<td>0.004158</td>
<td>0.009923</td>
<td>0.419</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>-606.1</td>
<td>-587.1</td>
<td>308.1</td>
</tr>
</tbody>
</table>

Table C.10 LMM results for the temporal interval from syllable onset to H (ContoH) in nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0062212</td>
<td>0.078874</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0175590</td>
<td>0.132511</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.6211066</td>
<td>0.0569643</td>
<td>28.458</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.179905</td>
<td>0.0577624</td>
<td>-3.115</td>
<td>0.002</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>-0.0002883</td>
<td>0.0004895</td>
<td>-5.890</td>
<td>0.0001</td>
</tr>
<tr>
<td>Word</td>
<td>0.0141100</td>
<td>0.0147754</td>
<td>0.955</td>
<td>0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>-337.3</td>
<td>-314.5</td>
<td>174.7</td>
</tr>
</tbody>
</table>

Table C.11 LMM results for the temporal interval from H to word offset (WofftoH) in nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker (Intercept)</td>
<td>881.97</td>
<td>29.698</td>
</tr>
<tr>
<td>Residual</td>
<td>2054.30</td>
<td>45.324</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>59.998</td>
<td>15.400</td>
<td>3.90</td>
<td>0.001</td>
</tr>
<tr>
<td>Group</td>
<td>66.087</td>
<td>21.615</td>
<td>3.06</td>
<td>0.002</td>
</tr>
<tr>
<td>Word - Maryellen</td>
<td>269.190</td>
<td>53.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>3454</td>
<td>3473</td>
<td>-1722</td>
</tr>
</tbody>
</table>
Table C.12 LMM results for the distance between L and H (LtoH) in nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0070409</td>
<td>0.08391</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0187502</td>
<td>0.13693</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.7223223</td>
<td>0.0757539</td>
<td>22.736</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.534700</td>
<td>0.1019901</td>
<td>-5.243</td>
<td>0.001</td>
</tr>
<tr>
<td>Syllable duration</td>
<td>0.0034732</td>
<td>0.0003565</td>
<td>9.742</td>
<td>0.000</td>
</tr>
<tr>
<td>Word</td>
<td>-0.0188379</td>
<td>0.0158428</td>
<td>-1.189</td>
<td>0.23</td>
</tr>
<tr>
<td>Group x Syllable duration</td>
<td>0.0021223</td>
<td>0.0004911</td>
<td>4.321</td>
<td>0.001</td>
</tr>
</tbody>
</table>

AIC | -299.8 |
BIC | -273.2 |
logLik | 156.9 |

Table C.13 LMM results for the temporal interval between syllable onset and L (ContoL) in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0033364</td>
<td>0.057761</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>0.0000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0092449</td>
<td>0.096150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.280022</td>
<td>0.029807</td>
<td>9.394</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.031809</td>
<td>0.042244</td>
<td>-0.753</td>
<td>0.5</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.003485</td>
<td>0.013458</td>
<td>-0.259</td>
<td>0.8</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>0.061254</td>
<td>0.019186</td>
<td>3.193</td>
<td>0.002</td>
</tr>
</tbody>
</table>

AIC | -818 |
BIC | -788.9 |
logLik | 416 |

Table C.14 LMM results for the temporal interval between syllable onset and H (ContoH) in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0087190</td>
<td>0.093376</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>0.0028061</td>
<td>0.052972</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0223965</td>
<td>0.149655</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.38563</td>
<td>0.05603</td>
<td>24.731</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.17724</td>
<td>0.06813</td>
<td>-2.601</td>
<td>0.1</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.15130</td>
<td>0.02384</td>
<td>-6.347</td>
<td>0.001</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>0.24316</td>
<td>0.02988</td>
<td>8.138</td>
<td>0.001</td>
</tr>
</tbody>
</table>

AIC | -394.2 |
BIC | -365 |
logLik | 204.1 |
Table C.15 LMM results for accented syllable duration in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>469.61</td>
<td>21.671</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>157.04</td>
<td>12.532</td>
</tr>
<tr>
<td>Residual</td>
<td>1145.81</td>
<td>33.850</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>171.153</td>
<td>13.036</td>
<td>13.129</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-14.746</td>
<td>15.789</td>
<td>-0.934</td>
<td>0.3</td>
</tr>
<tr>
<td>Condition</td>
<td>11.468</td>
<td>5.401</td>
<td>2.123</td>
<td>0.04</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>8.239</td>
<td>6.758</td>
<td>1.219</td>
<td>0.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>4713</td>
<td>4742</td>
<td>-2350</td>
</tr>
</tbody>
</table>

Table C.16 LMM results for the distance between L and H (LtoH) in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>0.0086675</td>
<td>0.0025641</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>0.0025641</td>
<td>0.0050637</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0283058</td>
<td>0.0168243</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.09239</td>
<td>0.05589</td>
<td>19.545</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-0.14478</td>
<td>0.06849</td>
<td>-2.114</td>
<td>0.04</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.13573</td>
<td>0.02660</td>
<td>-5.103</td>
<td>0.0001</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>0.18369</td>
<td>0.03359</td>
<td>5.469</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>-286.1</td>
<td>-256.9</td>
<td>150.0</td>
</tr>
</tbody>
</table>

Table C.17 LMM results for L scaling in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>446.8598</td>
<td>21.1391</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>8.1801</td>
<td>2.8601</td>
</tr>
<tr>
<td>Residual</td>
<td>42.8021</td>
<td>6.5423</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>125.025</td>
<td>10.690</td>
<td>11.696</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>2.147</td>
<td>14.966</td>
<td>0.143</td>
<td>0.9</td>
</tr>
<tr>
<td>Condition</td>
<td>1.591</td>
<td>1.050</td>
<td>1.279</td>
<td>0.1</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>-5.772</td>
<td>1.306</td>
<td>-4.418</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>3185</td>
<td>3214</td>
<td>-1586</td>
</tr>
</tbody>
</table>
Table C.18 LMM results for H scaling in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>12.98934</td>
<td>3.60407</td>
</tr>
<tr>
<td>Word (intercept)</td>
<td>0.42897</td>
<td>0.65496</td>
</tr>
<tr>
<td>Residual</td>
<td>1.72712</td>
<td>1.31420</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>11.0663</td>
<td>1.8372</td>
<td>6.023</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-3.7124</td>
<td>2.5527</td>
<td>-1.454</td>
<td>0.14</td>
</tr>
<tr>
<td>Condition</td>
<td>-1.1664</td>
<td>0.2116</td>
<td>-5.511</td>
<td>0.0001</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>1.2190</td>
<td>0.2624</td>
<td>4.645</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

AIC: 1672  BIC: 1701  logLik: -829

Table C.19 LMM results for the scaling difference between L and H in prenuclear versus nuclear focal rises.

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>3.1907</td>
<td>1.7862</td>
</tr>
<tr>
<td>Word (Intercept)</td>
<td>0.2329</td>
<td>0.4826</td>
</tr>
<tr>
<td>Residual</td>
<td>1.6217</td>
<td>1.2734</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.2617</td>
<td>0.9353</td>
<td>7.764</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-3.8013</td>
<td>1.2712</td>
<td>-2.990</td>
<td>0.003</td>
</tr>
<tr>
<td>Condition</td>
<td>-1.9728</td>
<td>0.2034</td>
<td>-9.701</td>
<td>0.0001</td>
</tr>
<tr>
<td>Group x Condition</td>
<td>1.9643</td>
<td>0.2543</td>
<td>7.725</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

AIC: 1633  BIC: 1662  logLik: -809.3
Appendix D: LMM analyses for the results in Chapter 6

Table D.1 LMM results for final syllable duration across three boundary types (W - word, ip - intermediate phrase and IP - intonational phrase).

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (Intercept)</td>
<td>159.67</td>
<td>12.636</td>
</tr>
<tr>
<td>Residual</td>
<td>379.09</td>
<td>19.470</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>269.351</td>
<td>6.744</td>
<td>39.94</td>
<td>0.000</td>
</tr>
<tr>
<td>Boundary Type - IP</td>
<td>67.598</td>
<td>3.985</td>
<td>16.96</td>
<td>0.000</td>
</tr>
<tr>
<td>Boundary Type - W</td>
<td>-66.587</td>
<td>3.302</td>
<td>-27.15</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>27.297</td>
<td>9.540</td>
<td>2.86</td>
<td>0.04</td>
</tr>
<tr>
<td>Boundary Type x Group</td>
<td>-41.656</td>
<td>5.270</td>
<td>-7.90</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>3253</td>
<td>3284</td>
<td>-1618</td>
</tr>
</tbody>
</table>
Appendix E: Analyses for the results in Chapter 7

Table E.1 BE and KE tonal categories and total number of observations for each category, based on ToBI annotations in Experiment 4.

<table>
<thead>
<tr>
<th>Association</th>
<th>Tone - BE</th>
<th>Total number</th>
<th>Tone - KE</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch accents</td>
<td>H*</td>
<td>274</td>
<td>H*</td>
<td>495</td>
</tr>
<tr>
<td></td>
<td>L*</td>
<td>290</td>
<td>L*</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>L+H*</td>
<td>314</td>
<td>L+H*</td>
<td>275</td>
</tr>
<tr>
<td></td>
<td>L*+H</td>
<td>209</td>
<td>!H*</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>L*+!H</td>
<td>31</td>
<td>!H*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>!H*</td>
<td></td>
</tr>
<tr>
<td>ip boundary</td>
<td>H-</td>
<td>128</td>
<td>H-</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>L-</td>
<td>532</td>
<td>L-</td>
<td>553</td>
</tr>
<tr>
<td></td>
<td>!H-</td>
<td>43</td>
<td>!H-</td>
<td>50</td>
</tr>
<tr>
<td>IP boundary</td>
<td>H%</td>
<td>208</td>
<td>H%</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>L%</td>
<td>499</td>
<td>L%</td>
<td>569</td>
</tr>
</tbody>
</table>

Table E.2 LMM results for the \( f0 \) peak scaling in the first accented word across three sentence types (D - declaratives, WhQ - wh-questions and PQ - polar questions).

| Random effects | | | |
|----------------|-----------------|-----------------|
| Groups         | Variance        | Std. dev.       |
| Speaker (Intercept) | 6.4923 | 2.5480 |
| Residual       | 4.8013          | 2.1912          |

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.9899</td>
<td>1.3035</td>
<td>6.130</td>
<td>0.001</td>
</tr>
<tr>
<td>Type - PQ</td>
<td>2.9361</td>
<td>0.3857</td>
<td>7.613</td>
<td>0.001</td>
</tr>
<tr>
<td>Type - WhQ</td>
<td>4.7248</td>
<td>0.3691</td>
<td>12.801</td>
<td>0.0001</td>
</tr>
<tr>
<td>Group</td>
<td>-1.7691</td>
<td>1.8416</td>
<td>-0.961</td>
<td>0.7</td>
</tr>
<tr>
<td>Type PQ x Group</td>
<td>4.8234</td>
<td>0.5476</td>
<td>8.808</td>
<td>0.001</td>
</tr>
<tr>
<td>Type WhQ x Group</td>
<td>-2.5905</td>
<td>0.5140</td>
<td>-5.040</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>1973</td>
<td>-962</td>
</tr>
</tbody>
</table>

Table E.3 LMM results for the \( f0 \) peak scaling in the first accented word in two sentence types (D - declaratives, I - imperative).

| Random effects | | | |
|----------------|-----------------|-----------------|
| Groups         | Variance        | Std. dev.       |
| Speaker (Intercept) | 10.5875 | 3.2538 |
| Residual       | 3.2687          | 1.8080          |

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7.8760</td>
<td>1.6435</td>
<td>4.792</td>
<td>0.001</td>
</tr>
<tr>
<td>Type</td>
<td>6.2575</td>
<td>0.4669</td>
<td>13.403</td>
<td>0.000</td>
</tr>
<tr>
<td>Group</td>
<td>-1.7030</td>
<td>2.3228</td>
<td>-0.733</td>
<td>0.46</td>
</tr>
<tr>
<td>Type x Group</td>
<td>-2.4413</td>
<td>0.6620</td>
<td>-3.688</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
</tr>
</thead>
<tbody>
<tr>
<td>727.5</td>
<td>746.4</td>
<td>-357.7</td>
</tr>
</tbody>
</table>
Author/s: MAXWELL, OLGA

Title: The intonational phonology of Indian English: an autosegmental-metrical analysis based on Bengali and Kannada English

Date: 2014

Citation: Maxwell, O. (2014). The intonational phonology of Indian English: an autosegmental-metrical analysis based on Bengali and Kannada English. PhD thesis, School of Languages and Linguistics, Faculty of Arts, The University of Melbourne.

Persistent Link: http://hdl.handle.net/11343/39964

File Description: The intonational phonology of Indian English: an autosegmental-metrical analysis based on Bengali and Kannada English