“Men call him Procrustes, or the Stretcher,” said the girl–and she talked low and fast. “He is a robber. He brings hither all the strangers that he finds traveling through the mountains. He puts them on his iron bed. He robs them of all they have. No one who comes into his house ever goes out again.”

“Why do they call him the Stretcher? And what is that iron bed of his?" asked Theseus, in no wise alarmed.

“Did he not tell you that it fits all guests?” said the girl; “and most truly it does fit them. For if a traveler is too long, Procrustes hews off his legs until he is of the right length; but if he is too short, as is the case with most guests, then he stretches his limbs and body with ropes until he is long enough. It is for this reason that men call him the Stretcher.”

James Baldwin  
Old Greek Stories

Introduction

This dissertation provides an account of the phonology and phonetics of intonation in Derry City within the Autosegmental Metrical (AM) framework. Derry City is located in the northwest of Northern Ireland near the border with Donegal in the Republic of Ireland and is the second largest urban area in Northern Ireland. Derry City English (DCE) belongs to the northern Irish English variety (nIE), which is spoken across most of Northern Ireland, in Donegal, and in northern parts of counties Monaghan and Cavan. The AM framework views the pitch contour as the systematic implementation of a sequence of underlying phonological primitive tones (L for low and H for high) inside an intonational phrase (IP). The majority of AM research on northern Irish English has focused on Belfast English, with a much smaller body of research on Donegal English. Only one study of intonation in Derry City has been carried out in the last forty years, and that presented an analysis of two speakers within the framework of the British Tradition of intonational analysis. There has until now, been no larger scale research on Derry City intonation, nor any research conducted within the AM framework, even though this approach has been the dominant mode of intonational analysis for the last thirty years.

The first broad aim of this work, therefore, is to offer a description of Derry City English within the AM framework so that it is amenable to comparison with other studies of nIE intonation, with the intonation of other varieties of English, and with intonation in other languages. It catalogues the phonological inventory and phonetic features of pitch events in relation to a variety of formal and functional conditions. On the formal side, it analyses intonation in unmarked declaratives under a range of metrical conditions, namely variation in foot size, anacrusis (unstressed content before the first stressed syllable), and lexical boundary conditions. On the functional side, it analyses intonation across sentence modes (declaratives, binary questions, wh-questions, and declarative questions) and under different of focus conditions. In this context, focus refers any semantic element which the speaker brings to prominence in an utterance, through linguistic and implementational strategies.

The second key aim of the research is theoretical in nature. It originates in questions raised by the pervasive use of rising intonation in unmarked contexts in northern Irish English, while most other varieties of English (and other languages) tend to have a fall in pitch. Firstly, this raises questions about the phonology and the phonetics of intonation, in terms both of description and function. For example, within AM, it is common to divide the pitch contour into a linguistic component which can be described in terms of intonational phonology, and a paralinguistic component which exists (quasi-)independently of the linguistic element. The linguistic/paralinguistic distinction may be easy to maintain when unmarked declaratives and binary question use pitch accents involving different pitch trajectories, i.e., a falling pitch in the nuclear contour (H\*L %) and binary questions use a rising pitch (L\*H % or L\*H H%). However, if the same rising contour dominates both in declaratives (L\*H %) and binary questions (L\*H %), one might ask if intonation has any linguistic function at all in nIE. Alternatively, one might wonder if the AM description of the phonology fails to provide a sufficient account of the linguistic component.

A standard AM explanation would be that phonetic traces in the pitch contour reflect attitudinal shifts across sentence modes, with question forms exploiting progressively higher pitch registers as the amount of grammatical or semantic question marking decreases. Such changes would be identified as paralinguistic. However, this would imply that nIE speakers use pitch paralinguistically for questions forms while speakers of other varieties use pitch linguistically. This would in turn imply a major typographic distinction, which feels intuitively implausible. A more plausible explanation may be that there exists a separate (linguistic) register tier which the speaker employs to shift overall pitch range upward under certain grammatical conditions. This could apply both to question forms and the narrow focus. Do note that evidence for a register tier would not rule out the presence of paralinguistic pitch effects nor would it suggest that the register tier is unique to nIE.

The second theoretical concern relates to the primacy of the H tone in the analysis of intonation in English (and other languages). Issues such as peak alignment, peak drift, and peak identification have surfaced frequently in AM research. For English, this is due largely to the fact that pitch accents are most typically associated with peak prominence (H\*) in the two most widely studied varieties of English, Southern British English and General American English. This tonal element of the pitch accent is commonly called the starred tone. In nIE, however, L\*H dominates in the nuclear contour of unmarked declaratives, with L being the starred tone. The formal analysis of intonation in DCE will allow us to see if there are any systematic changes to the pitch accent which provide evidence for the special nature of the H tone, or if the attention accorded to the H tone may simply be due to its dominance in other varieties of English.

In order to deal with the two key aims, two different approaches to the analysis have been adopted. The second approach is motivated in large by a reflection on the insights gained from the first approach.

The first approach is, in essence a phonology-first approach, and might also be viewed as a top-down approach. It assumes that there is previously established set of phonological pitch events and pitch accents which can be labelled using the established system. As such, the first pass of this analysis involves evaluating the pitch accents and boundary tones in each utterance, while the second pass involves identifying tonal targets associated with those pitch accents for the phonetic analysis. The phonetic analysis of the phonology-first approach adopts an understanding of tonal targets in terms of the local *f*0 minima and maxima. The phonology-first approach is more orthodox and establishes baseline phonological and phonetic features of DCE, thus facilitating cross-variety comparisons with other AM analyses of intonation.

The second approach is a phonetic-first approach, which might also be described as a bottom-up approach. The first pass here involves identifying the minimal number of H and L turning points required to resynthesize the contour so that it is perceptually identical to the original. Having identified the turning points, they are then associated with intonational events, such as pitch accents and boundary tones. This analysis proposes an extra pitch event, the secondary tone. A secondary tone is an optional tonal event associated with another tone, typically indicating the onset or extent of the primary tone (Secondary Tone Hypothesis or STH). This analysis revisits and reanalyses some of the data from the formal analysis and the analysis of sentence modes to test the extent to which this alternative form of analysis can account more fully for DCE within the AM approach. The second approach also adopts a more rigorous definition of the tonal target. Here, it is viewed as the intended target of a pitch trajectory and is associated with an elbow (or turning point) in the contour rather than simply *f*0 minimum or maximum. Of course, minima and maxima can themselves also be elbows. Given time constraints, the analysis of focus is conducted using the second approach alone.

The issues raised above have been distilled into the research questions (RQs) outlined in Table 2.1. Note that the first three RQs focus largely on descriptive concerns, while RQs four to six relate to theoretical concerns.

Table .Research Questions

|  |
| --- |
| **Descriptive Concerns**   1. What are the phonological and phonetic characteristics of pitch accents in DCE in unmarked speech under variations in metrical (anacrusis and foot size) and lexical structure? 2. What are the phonological and phonetic characteristics of nuclear pitch contours in DCE across sentence modes?   **Theoretical concerns**   1. Is there evidence in the realisation of PN pitch accents in DCE for the special status of H tones? 2. Does a register tier provide a plausible phonological explanation for variation across sentence modes in DCE? 3. Does the Secondary Tone Hypothesis provide a more stable analysis of the phonology and phonetics of Intonation in DCE? |

Overall, it is hoped the phonology-first and phonetics-first approaches to the analysis of DCE will offer further insight into the nature of nIE rises while at the same time providing a critique of and advancing our understanding of both the phonology and phonetics of intonation within an AM framework. Moreover, if the Secondary Tone Hypothesis provides meaningful insight into intonation structure within AM, it may aid to the effective use of AM analysis in the field of speech synthesis, helping generate natural-sounding pitch movements generated from a parsimonious set of pitch parameters. Finally, it will be argued that the STH approach can help resolve—or at least provide a bridge between—some disagreements within AM

In order to answer these questions, the dissertation has been organised as follows. Chapters 2 and 3 provide an overview of the literature while Chapter 4 introduces the research questions and modes of thought stemming from the literature review. Chapter 5 focuses on methodologies, specifically in relation to corpus development and statistical methods. Chapter 6 provides an analysis of Metrical and Lexical effects on intonation form (RQ1) and also provides insights relating to the status of H targets (RQ4). Chapter 7 analyses the function of intonation in sentence modes (RQ2) and the issue of a register tier (RQ5). Chapter 8 represents a turning point from the phonology-first to the phonetics-first approach. It offers critical analysis the phonology-first approach from the previous chapters and introduces the Secondary Tone Hypothesis (STH) along with the tools employed for the analysis. Chapter 9 then uses the STH to re-evaluate some of the problematic data from Chapters 6 and 7 along with focus data and in the process trials the effectiveness of the STH. Finally, Chapter 10 provide a short critical summary of the research, considers its significance with the AM approach and nIE studies, and suggests directions for future research.

# Theoretical Context: Intonation

This chapter focuses on the theoretical framework governing the analyses described in this dissertation. It provides a summary of the two main approaches to Intonation studies in English, before delving into more detail on the Autosegmental Metrical (AM) approach, the framework which is adopted for this study. The discussion of AM focuses on its core principles and accounts for the development of divergent views within AM itself. These different views affect core issues, including the structure of the Intonational Phrase, the underlying phonology, and its phonetic implementation.

## What is intonation?

Following Ladd (2008, p. 4), intonation is understood here as the post-lexical linguistically structured use of pitch to convey meaning, and is expanded upon in the following paragraphs. It is post-lexical in that it may be implemented across a domain larger than the individual word. This distinguishes it from other linguistic uses of pitch, such as lexical pitch (e.g. in Japanese or Swedish) and lexical tones (e.g. in Chinese or Thai), where pitch is a property of the word and so helps distinguish between lexical items. It is linguistically structured in that its components can be categorised into discrete linguistic entities (high or low), with meaningful intonational events referred to as *pitch accents* (Cruttenden, 1997; Ladd, 2008). This distinguishes it from gradient, indexical, and paralinguistic uses of pitch. However, as paralinguistic uses of pitch and linguistic uses of pitch operate largely along a single phonetic dimension (*f0*), it can be difficult to distinguish between them (see 2.3.5 below). For the purposes of the research outlined here, I have attempted to maintain the linguistic/paralinguistic distinction, both in terms of the theoretical framework and data collection. Finally, intonation is meaningful in the sense that it forms part of a language’s grammar, helps signal affect, and contributes towards information structure in discourse (Tench, 1996; Cruttenden, 1997).

Intonation needs to be considered independently of other prosodic features with which it is sometimes confused, namely prominence and stress. Throughout this dissertation, prominence refers to any phonetic effect which causes a syllable to stand out in an utterance. This could be a function of pitch, loudness, length, or voice quality. Stress refers exclusively to lexical stress, which is viewed as a component of the lexical word encoded in the lexicon along with its phonemic structure. Lexical stress may be associated with a prominence, but this is not necessarily the case, especially in continuous speech. [example?]

Like stress, pitch accents are associated with prominence, but they are not simply manifestations of prominence. The view taken here is that, by and large, pitch accents occur in association with prominent syllables and thus may operate as *cues to prominence*. This view originates from Ladd (2008, pp. 50–54), who argues that pitch does not itself lend prominence to a syllable—as per ’t Hart *et al* (1990, p. 96)—but that, because pitch accents are associated with prominent stressed syllables, they help cue the percept of prominence. Effectively, this allows the linguistic function of the pitch accent to be viewed independently of prominence signalling. This distinction may, however, breakdown when considering the topic of narrow focus, in which a specific semantic element within the IP is given greater prominence than other elements. In narrow focus, accentuation and de-accentuation are understood to assist in the manifestation of narrow-focus process [REFs]. (See also section 2.2.2.)

### Functions of Intonation

We can identify two key linguistic functions of intonation in speech: grammatical and discoursal.

In English, the grammatical function is typically reflected in the tendency of intonation contours to contrast binary questions, typically with rising intonation, and statements, typically with falling intonation. However, as we shall see in Chapter 3, this distinction does not always hold. The discourse functions can be understood in terms of information structure and speaker interaction.

In terms of information structure, intonation can be used to signal completeness or incompleteness. Typically, completeness is indicated with a low boundary (or fall) while incompleteness is indicated using a high (or rising) boundary [REF?]. The pitch movement associated with incompleteness is typically referred to as the ‘continuation rise’ [REF?].

Intonation can also be used to select specific semantic content for focus. This can be understood in several ways. The semantic content under focus may be understood as new information for the listener—and may be accented—while the other content may be viewed as old, shared, or given information, and thus may be deaccented. This can be described as narrow focus. Alternatively, the focus may be corrective, in that the speaker wants to correct or alter the listener’s knowledge about a shared topic. Focus can also be achieved through semantic and syntactic means, e.g., with implicitly corrective adverbials such as “actually” or via cleft structures such as “It was X who did Y”, as in the phrase, “It was actually John who rescued the children.” However, even when semantic and syntactic features are employed, so too is pitch [REF].

Given the inherently communicative function of language, both of these discourse functions—completeness and focus—can be related to interaction. Signalling incompleteness and completeness helps signal to the listener when it is appropriate to initiate a turn, while the selection of semantic components for focus depends on the speaker inferences about listener knowledge and the extent to which they share the same information. As such, functions of intonation have also been generalised in the form of abstractions, so that distinctions between grammatical and discourse functions may be collapsed into more abstract categories.

Brazil, Gussenhoven, and Cruttenden all offer more abstract views of intonation function. Brazil (1995), focusing on functional uses of intonation in social interactions, defines two broad types of meaning: proclaiming and referring, where referring intonation references given information (rise, or fall-rise) and proclaiming intonation references new information (fall, rise-fall). Brazil also includes a second dimension, that of dominance. He argues that dominance is a function of asymmetry in spoken interactions, and that a speaker can assert current control over the discourse using a dominant tonal pattern (erceived he sense of finality is ded) which a final fallnuation rises ase include question forms (ceding dominance) and statemreferring rise and the proclaiming rise-fall) or cede control using a non-dominant pattern (referring fall-rise and the proclaiming fall). This discourse approach can be generalised to show how so-called continuation rises can assert dominance and present information as given (note: presented **as** given) while a final fall can cede control (of which the sense of finality is a by-product). Cruttenden adopts an even more abstract distinction between open rises and closed falls. This allows for a generalisation in which continuation rises and YNQ rises are collapsed into the open category, while proclaiming falls and completion falls are collapsed into the closed one ( Cruttenden, 1997, p. 119). Gussenhoven’s abstractions relate to the origins of intonation functions, specifically the argument that universal tendencies in intonation arise from the phonologization of biological codes. These are considered in more detail in section 2.3.5.

It should be noted that these abstractions invoke the sense that there is (quasi-)universal link between the pitch movement and meaning. Gussenhoven, however, does observe that biological codes subjected to phonologization become part of an arbitrary linguistic system, and so the meaning of the linguistic form is no longer constrained by its biological progenitor.

### Acoustic Measurement of Intonational events

Pitch is very strongly correlated with the fundamental frequency (*f*0) of the speech waveform, which reflects the rate of vibration of the vocal folds, the source of voicing, and thus *f*0 is the main parameter used in the acoustic analysis of pitch. At the same time, there is also a psychoacoustic component to pitch. For example, *f*0 is subject to segmental effects from consonants and vowels (Ohala, 1976) [REF] but listeners tend to filter out these effects as they interpret the pitch contour. This reinforces that fact that pitch and *f*0 do not correlate completely.

It is also important to remember that even though *f*0 is very strongly correlated with pitch, it is only a measurement of first harmonic in the rich spectrum of voiced speech, and that other harmonics repeat throughout the spectrum multiples of *f*0. Thus, the acoustic information which is interpreted as pitch is not simply encoded in the spectrum at the fundamental frequency, rather it is present throughout the spectrum [REF].

The SI unit for frequency is Hertz, which measures the number of oscillations per second of a periodic waveform. However, our perception of pitch is largely logarithmic, meaning that we hear each doubling of the frequency (in Hertz) as representing an equal ‘distance’ in terms of pitch. For example, in music, each doubling of *f*0 represents an increase of one octave. Even though our perception of individual frequencies up to about 500 Hz is roughly linear rather than logarithmic [REF], we need to remember that *f*0 is only one—albeit a very important one—of a rich combination of harmonic frequencies contributing to our perception of pitch. Therefore, it is not surprising that Nolan (2003) found that logarithmic or quasi-logarithmic frequency scale better reflected intuitions about pitch equivalence in an imitation experiment. As a result, *f*0 will be analysed in terms of semitones (ST), a log 2 measurement of frequency reflecting the twelve-tone equal temperament found in most Western music styles, wherein each octave comprises twelve perceptually and logarithmically equidistant tones. Given the general familiarity with this system, the description of *f*0 in ST units is also more likely to be easily interpreted.

## Theoretical Frameworks for Intonation Analysis

There are a large number of different models of intonation, but this section summarises the two approaches which have most influenced the study of intonation in English, namely the British tradition (Halliday, 1967; Arnold and O’Connor, 1973; Tench, 1996; Cruttenden, 1997) and the Autosegmental Metrical (AM) approach (J. B. Pierrehumbert, 1980; Gussenhoven, 2004; Ladd, 2008). After a brief overview of both, focus will shift back to the AM approach, since it is adopted in this study.

### The British Tradition

The British tradition posits the existence of an intonation group (or tone group) (Cruttenden, 1997). The intonation group must have at least a nucleus—or tonic syllable—which contains the most prominent pitch movement, also known as a pitch accent, of the phrase. A nucleus may have a tail, which is a sequence of syllables following the nucleus. In addition, there may also be a pre-tonic segment with a pre-head and/or a head. The pre-head is the stretch of unstressed syllables before the first stressed syllable, i.e. anacrusis, while the head is the stretch from the first stressed syllable up to (but not including) the nucleus. The intonation group structure is summarised in Table 2.1 (after Tench, 1996, p. 14). Within this framework, the nucleus is described in terms of tone height (high, low) and pitch glide (rising, falling, and sometimes level). Thus, one might describe the nucleus as a high-fall or a rise-fall.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table . Formal structure of the intonation group   |  |  |  |  | | --- | --- | --- | --- | | pre-tonic segment | | tonic/nuclear segment | | | pre-head | head | tonic/nucleus | tail | | (P) | (H) | N | (T) | |  |  |  |  | |

It should be noted that the intonation group fuses prominence, pitch, and lexical stress features and analyses them on a syllable-by-syllable bases. For example, Cruttenden (1997) identifies four degrees of accent/stress. Only the first two of these contain pitch movements, and so only they can be described as (pitch) accents. These are the primary stress/accent, which contains main prominence of the intonation group, i.e., the nucleus, and the secondary stress/accent, which contains a non-nuclear pitch movement. The other two degrees of accent/stress are tertiary stress—indicating a loudness or length related prominence—and the unstressed syllable. The visual implementation of this approach can be seen in the use of interlinear tonetic transcription, also known as tadpole diagrams, as shown in Figure 2.1 below. This represents a fall-rise nuclear contour, where the high fall nucleus occurs in *thought* and identified by the large black dot and its falling tail, with the rise represented by the sequence of rising dots for each syllable through the rest of the phrase. The lexically stressed syllable in *married* (also marked by a large black dot)only contains a tertiary stress, and so does not carry a pitch accent. All the other syllables are identified as unstressed using the small dots.

|  |
| --- |
| Figure . Example of interlinear tonetic transcription of intonation. (Cruttenden, 1997) |

One criticism of the British tradition is that by dividing the intonation unit into a distinct hierarchical structure, it creates a level complexity which requires a separate description of each intonational phrase, and that the structurally large intonation group suggests a degree of planning which may not be psychologically plausible (Taylor, 1992). Another criticism of the British tradition is that it makes assumptions about form-function relationships, thus excluding patterns *a-priori* which do not fit the assumptions (Ladd, 1983). This second criticism, however, as we shall see later, cannot be aimed exclusively at the British tradition [SECTION].

A criticism perhaps even more pertinent to the current study is that the British tradition bundles prominence, stress, and pitch accent into a single structural unit. This may obscure the way in which functionally equivalent pitch accents can be implemented across a wide range of utterances containing very different stress patterns, or alternatively it may obscure ways in which speakers can vary the location of prominence in phrases which are otherwise structurally equivalent. Thus, the following section introduces the AM approach, which—among other things—separates the strands of (lexical) stress, pitch movement, and prominence.

### The Autosegmental Metrical Approach to Intonation

The Autosegmental Metrical (AM) approach to intonation has its origins in work by Liberman (1975) and Pierrehumbert (1980), and has become the most widely used approach in the description and analysis of intonation in English varieties both within and beyond Europe. We will briefly outline the four core tenets of the AM approach, as summarised by Ladd (2008, pp. 44–45), while also indicating how it differs from the British tradition.

1. **Sequential tone structure**. Intonation is viewed as a string of tonal targets rather than a series of dynamic pitch movements as in the British tradition. Thus, pitch glides are considered epiphenomenal transitions occurring in between target tones, so pitch accents are important *tonal* events[[1]](#footnote-2). The string of tones occurs within an intonational phrase (IP) and may be marked by initial or final edge tones at prosodic boundaries. Pitch accents are ‘associated with prominent syllables in the segmental string’ (Ladd, 2008, p. 44). This means that there is a tone tier independent of but still linked to the segmental string. Moreover, there can be a one-to-many or a many-to-one association of tone to units on the segmental string. The autonomy of pitch and the segmental string distinguishes AM from the British tradition, which—as previously noted—fuses degrees of stress, prominence, and pitch movements into a single structure.
2. **Distinction between pitch accent and stress***.* Pitch accents are separate phenomena from stress and prominence (underscoring the distinction made in section 2.1); however, as their location is associated with stressed syllables, they may serve as cues to prominence. Again, this facilitates an analysis of the intonational structure of a phrase independently of the segmental string and stress tier.
3. **Analysis of pitch accents in terms of level tones**. Tones are analysed as intonational primitives, i.e., either high (H) or low (L) tonal targets. This reinforces the point, that the intonational contour is a manifestation of a string of tonal primitives rather than a sequence of both tones and glides.
4. **Local source for global trends**. The phonetic realisation of H and L targets across an utterance is accounted for by a variety of local factors, such as downstep. This contrasts with the idea, for example, that individual pitch events are superimposed over a global downward trend, as in the Fujisaki model (Fujisaki, 2004)[[2]](#footnote-3). The crucial point here, however, is that there are rules governing the phonetic realisation of the underlying phonological sequence of tones, and these account for differences in the scaling and timing of H and L tones.

The AM approach has been adopted for this research project for several reasons. Firstly, as the dominant approach to the study of intonation in English, it facilitates comparison with recent and contemporary studies, including those on nIE, such as Lowry (Lowry, 2001, 2002, 2011), Dorn (2006) Sullivan (Sullivan, 2007, 2010, 2012), or Jespersen (2018). Secondly, the view that intonational structure can be analysed independently of—or at least discretely from—segmental and metrical structure is appealing in that it lends itself to a scientific approach which isolates and manipulates target variables[[3]](#footnote-4). Finally, there is much to be said for viewing contours as epiphenomena associated with a string of phonological tones. This is not least because it helps reduce much of the noise in the signal—of which there is always a considerable amount in the *f0*—and offers a more parsimonious account of the intonational events. It also, in principle at least, provides insight into how an underlying phonological structure may be accounted for via a set of implementational rules. Such rules—among other things—help explain how the string of tones is linked to the segmental string at different locations without compromising the underlying intonational structure and, as a corollary, why otherwise functionally identical pitch events may seem superficially different.

## AM studies of Intonation

While the AM approach is very appealing, it is not without its own internal divisions and issues. Therefore, this section outlines the development the AM approach, highlights disagreements within AM, offers a critique, and explains how these issues have contoured the research presented in the subsequent chapters. Much of the description and critique which follows is indebted to the monographs by Ladd (2008), Cruttenden (1997), and Gussenhoven (2004).

### Pierrehumbert (1980)

Pierrehumbert’s (1980) doctoral thesis is a seminal work in the AM tradition of intonation analysis, and the labelling system it used has provided the basis for AM labelling systems since. Within Pierrehumbert’s original framework, an intonational phrase (IP) comprises a sequence of high and low tones (H and L) which are associated with three different structural units. These are the pitch accent, the boundary tone, and the phrase accent. A pitch accent is an intonational pitch event within the IP and is associated with the metrically strongest syllable in the foot in which it occurs. (Note that Pierrehumbert points out the there is disagreement about how pragmatic and syntactic considerations may affect the strength of syllables in the foot.) Both the phrase accent and boundary tone are edge tones. Boundary tones can occur at the beginning or end of the IP, while the phrase accent occurs before the final boundary tone of the IP.

Pierrehumbert (1980) uses a range of symbols after the tone to show how the tones are associated both with each other and with the text. Boundary tones are indicated by a percentage sign (L%, H%), while a macron indicates a phrase accent (H¯, L¯). An asterisk indicates which tone in a pitch accent which is associated with the metrically strongest syllable in the foot (H\*, L\*). In Pierrehumbert’s analysis, pitch accents can be monotonal or bitonal, and bitonal pitch accents can be either left- or right-headed, and a plus sign links leading and trailing tones to the pitch accent with which they are associated. The macron is also used for these tones, as in L¯+H\* or L\*+H¯. The conventional use of the asterisk and the percentage sign continues today; however, the macron is no longer used, and—where they are used [SEE XX]—phrase accents are indicated with a hyphen instead. The plus sign is still in use, but not in all labelling systems or AM analyses [SEE XXX].

Pierrehumbert views the intonational grammar as a finite state, shown in the schematic in Figure 2.2. Note that the empty line on the left boundary tone indicates that an initial boundary tone it is optional. The leftward-pointing arrow at the top of the Pitch Accents section indicates that PAs are iterative.

|  |
| --- |
| Figure . Pierrehumbert’s (1980) finite state grammar of the intonational phrase. |

In addition to the structural component of the PA and its tune-to-text associations, Pierrehumbert (1980) also posits rules that governed the phonetic implementation of the tones within the IP, three of which will be covered here. Firstly, the gradual reduction in pitch scaling across an utterance is motivated by downstep, a lowering of the phonetic realisation of the H tone as an automatic consequence of an HL or LH sequence of pitch accents, examples of which are shown in Figure 2.3, panels a. and b. Secondly, the height of the final boundary is determined by the preceding phrase tone. That is, after a H¯ phrase accent, the H% boundary tone is up-stepped; however the L% essentially becomes null in such cases, representing no to little lowering of *f*0 in the contour. The contrast between H¯L% and L¯H% can be seen clearly in panels b. and c. of Figure 2.3.

|  |  |  |  |
| --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | a. | b. | c. |   Figure . Examples of pitch contours and tonal targets from Pierrehumbert (1980) |

It should also be remembered that Pierrehumbert’s system only allows for two tone states, H or L. However, it needs to account for at least three different boundary phenomena: the final fall of the unmarked declarative (panel a), the down-stepped plateau of the calling contour (panel b.), and the high boundary of the binary question (panel c.). As such, the phrase accent and boundary tone rules are required to permit the two-tone system to account for such phenomena.

Finally, Pierrehumbert accounts for stretches of contour between boundary tones and pitch accents by viewing them as linear interpolations between pitch events.

### Critique of Pierrehumbert’s (1980) Upstep and Downstep

Several problems with this analysis have been identified in Pierrehumbert’s analysis both of downstep and upstep.

Firstly, it has been noted that Pierrehumbert’s account of downstep means that wherever downstep is observed [REF], an intervening L target must be posited to motivate—or justify—it, even if there is little evidence of one in the surface realisation. This can be seen in Figure 2.3, panel a., where the second PA must be described as H\*+L¯ in order to explain why the final H\* is downstepped. Ladd (1983) comments on this issue and also notes that sometimes the same phonological structure appears to give rise to very different surface contours. He suggests that the surface phenomena can be represented more sensibly if downstep is viewed as a speaker-motivated phenomenon (labelled as !H) rather than an obligatory implementational feature.

A similar problem may also be seen in relation to the phrase accent in panels a. and c. In panel a., the final pitch accent is H\* followed by the edge-tone sequence L¯L%. The phrase accent L¯ is required in order to explain the fall from the H\* to the boundary, since, without the intervention of the phrase accent, the boundary L% would not trigger a fall, but would reflect be a continuation of the H\*, since boundary L% essentially represents a null state. Thus in order to show the phrase final fall, L% must—by virtue of the argument Pierrehumbert presents—be accompanied by a phrase accent, leading to L¯L%. Unfortunately, there seems to be little phonetic evidence to indicate that there are indeed two tones here. Rather, it is the theory which requires the presence of the L¯ rather than the data which suggests it. Grabe (1998a) presents an alternative solution to this, which is to suggest that the final boundary tone is fully relational and does not need to be specified at all. In this way, a final H% always triggers up-step, L% always indicates final lowering, while the unspecified final boundary (or 0%) indicates a continuation of the final tone in the pitch accent. In effect, this obviates the need for the phrase boundary at all.

Together, Ladd’s proposal for a motivated rather than obligatory downstep (!H) and Grabe’s proposals for boundary tones make it is possible to describe the sequence of tones in an intonational contour more parsimoniously, and without the need for injecting extra tones into the tonal string for which there may be insufficient empirical evidence.

### Modifications to PA structure: a more hierarchical approach

Two modifications to PA structure as originally proposed by Pierrehumbert have also been suggested. Firstly, Pierrehumbert and Beckman (1986) propose that an intermediate phrase (small *ip*) can be embedded within the IP, wherein each *ip* ends with a phrase accent, but only the IP has a boundary tone. In this way, a hierarchical component is added to the IP, wherein the final accent of the IP becomes more akin to the nuclear tone of the British tradition, while the intermediate phrase is similar to the head of the tone group. Within AM approaches, it is now standard to view the final PA as the nuclear pitch accent, and the preceding pitch accents as pre-nuclear [REF]. This too reflects a more hierarchical structure. Moreover, the final pitch accent plus the final edge tones may be described as the nuclear contour (c.f., Gussenhoven, 2004, chapter 11). Such evolutions represent a partial synthesis of the original AM approach with the British tradition of a hierarchically structured Intonation Group.

Despite these modifications, it should still be borne in mind that the tonal string still belongs fundamentally to a separate tier (autosegment), the tones of which are associated with events in the segmental string, which itself is organised within a metrical hierarchically. After all, this is still the underlying view that led to the Autosegmental Metrical approach.

### Gussenhoven

A different approach to AM analysis is presented by Gussenhoven (1983, 2004, 2016). Reflecting this difference, Gussenhoven’s analysis of Standard Southern British English diverges from the Pierrehumbert school of thought in several key areas. Aside from a few notational differences[[4]](#footnote-5), Gussenhoven’s analysis differs in many fundamentals, summarised as follows:

1. **Off-ramp analysis**. All pitch accents are left-headed, following the British Tradition (c.f. Gussenhoven, 1983), unlike Pierrehumbert’s approach which permits both left- and right-headed PAs.
2. **IP boundary constituents**. The initial IP boundary (Lι, Hι) is obligatory but—similar to the IViE approach—the final IP boundary is optional (Lι, Hι, Øι) and the height of Hι is relative to the preceding tone. Although obligatory, initial L*ι* refers to a mid to low tone.
3. **Absence of phrase accents**. There is no phrase accent tone since the off-ramp analysis accounts for pitch movements which are otherwise required by the phrase-accent analysis [SEE XXX?].
4. **Double alignment**. Initial boundary and monotonal pitch accents extend rightwards towards the onset of the next pitch event, so that each Lι, Hι, L\*, and H\* tone has both a left-hand tonal target and a right-hand one (Gussenhoven, 2016). This means that, in effect, this tone continues rightward until the next tonal event. Trailing tones also display double alignment, but this is only typically seen before a final Tι. Final Tι is not double aligned, and only represents a single event.
5. **Intra-accentual interpolation.** Interpolation only occurs within the pitch accent, not across pitch accents. This in stark contrast to Pierrehumbert’s view of interpolation, which occurs between PAs.
6. **Rightward displacement**. Within a pitch accent, the trailing tone can drift rightwards towards the next tonal event, although the timing and scaling of the drift may be variable. This implementational feature is not a feature of Pierrehumbert’s approach.
7. **Morphological and PA-internal downstep**. Downstep applies only to H\* and can be motivated by a [downstep] morpheme applying to every H in the IP, or it can be triggered PA-internally when the H\* is prefixed by an H tone. (See 9 below.) Note that this essentially accommodated both Ladd’s speaker-motivated and Pierrehumbert’s obligatory downstep.
8. **Tri-tonal pre-nuclear pitch accents**. Pre-nuclear pitch accents can take the form H\*LH, wherein the L will be implemented as a fall after the H\* but final trailing H tone can drift rightwards towards the next pitch event. While the use of tritonal pitch accents is not in itself of great note, what is more note-worthy is that pre-nuclear pitch accents and nuclear pitch accents do not have access to the same inventory. Again, this reinforces a hierarchy in which the final (nuclear) pitch accent has special status.
9. **Tone prefixation**. All pitch accents can be prefixed with an L tone, while nuclear pitch accents can also be prefixed by an H tone or a HL sequence. Gussenhoven argues that this can explain the contrast between scooped rise, sometimes labelled L+H\*, and the unscooped H\* [REF?]. For example, if an H\*L tone is prefixed with an L\*, the H\* is displaced with L\* takings its place as the starred tone, thus becoming L\*HL. Again, the final L in the PA is free to drift rightwards. Similarly, an H prefixed to a nuclear pitch accent leads to obligatory downstep of the H\* tone. This equates to the H+!H\* of the ToBI-style analysis (see section 2.3.6 below).
10. **noSlump**. This is an obligatory feature of SSBE, which prevents the final Lι from exhibiting the kind of slump found in other varieties of English, including nIE (Lowry, 2001, 2002)

Gussenhoven’s intonational grammar of Standard British English is summarised in (1) below. Note that, there are only four pitch fundamental accents: L\*, H\*, L\*H, and H\*L. All other features are functions of implementation and the effect of prefixal tones. Even the “interloper” pre-nuclear H\*LH, Gussenhoven argues, is likely an artefact of an earlier IP form H\*L Hι (2004, pp. 302,  305–306).

While the formalization of the Grammar is relatively succinct, Gussenhoven’s analysis of English intonation appears considerably more complex overall, requiring the inclusion of the double alignment and rightward displacement rules as well as the noSlump rule, delay triggered by prefixal L\*, prefixal tones in general, and both morphological and PA-internal downstep. At the same time, it should be noted that Gussenhoven’s approach aims to integrate a limited PA inventory into a more complex set of implementations reflecting the complexity of pitch contours as they are realized.

### Linguistic and paralinguistic uses of pitch

The matter of linguistic versus paralinguistic use of pitch is important in the study of intonation. The most common distinction between paralinguistic and linguistic use of pitch is, as noted in section 2.1, between the gradient and categorical. For example, the paralinguistic use may be observed in the indexical relationship between the affect of surprise and the height of the pitch excursion, with higher excursions indicating greater surprise. In contrast, a linguistic use may be found in the categorical distinction between the falling intonation of declarative statements and the rising intonation of binary questions (H\*L % and L\*H H% or H\* H% respectively). Unfortunately, this distinction can be hard to maintain, since the paralinguistic use of pitch height appears to bleed into to the linguistic, especially in cases where the size of the excursion or overall height of *f*0 may vary depending on the grammatical, semantic, or pragmatic context as well as on affect and attitude.

Building on Ohala (1983), Gussenhoven (1999, 2004) proposes a framework aimed at accommodating both the phonological elements of pitch and the paralinguistic component. Ohala proposed that the near universal use of high or rising pitch has evolutionary biological origins, in which low pitch is associated with dominance and larger size while high pitch indicates smallness. He described this as a Frequency Code, which this leads to the tendency for low or falling pitch to be associated with statements (indicting certainty) and high or rising pitch with questions (indicating uncertainty and deferring to the listener). Expanding on this, Gussenhoven argues that there are three biological codes which do indeed motivate the apparent aforementioned universality of pitch movements. These are the Effort Code and the Production Code. The Effort Code associates greater overall effort in speech with an increase in *f*0 while the Production Code reflects the gradual decrease in *f*0 overtime during sustained voicing.

Gussenhoven argues that such codes may be phonologized in an otherwise arbitrary system of linguistic symbols. Thus, the frequency code, associated with size, may have a universal informational interpretation contrasting certainty with uncertainty, but lead to a phonologization which associates H% boundaries with questions and L% boundaries with statements. The effort code, where higher pitch may be universally associated with a greater sense of urgency, may lead to phonologization wherein H\* is associated with focus and L tones with backgrounding or prominence loss. The production code, where pitch gradually decreases, may be associated universally with a movement from newness or towards completion, may be phonologized so that a high final boundary H% indicates continuation or incompleteness while a low L% boundary is associated with completion. Because these universal tendencies have been phonologized, there is also the potential that the phonological units may change their associations and thus subvert the putative universality of function of pitch in terms of biological codes. In addition to phonologization, however, Gussenhoven also argues that a parasitic phonetic trace often remains, meaning that an utterance may still contain an indexical paralinguistic element in the pitch contour.

An example of this interpretation might be found in Haan’s (2002) PhD dissertation on the intonation of question forms in Dutch, in which she provides evidence of linguistic and paralinguistic components in question intonation. She finds that H% boundaries are associated systematically with question forms; however, she also finds paralinguistic effects in their implementation. She notes that declarative questions have the highest average *f*0, polar questions lower average *f*0, and wh-questions the lowest. This, Haan notes, confirms her hypothesis that there is an inverse correlation between pitch height and the amount of non-acoustic—i.e., lexical, semantic, or morphological—marking in the utterance.

In addition to the issue of pitch height is the alignment of tonal targets. For example, Gussenhoven refers to an observed distinction between the later less precise alignment of pitch peaks and an earlier more precise alignment inside stressed syllables in the Zagreb variety of Serbo-Croat, where the more precise alignment occurs when the lexical word is in focus. He notes that this may reflect a difference in the selection of pitch accent (L+H\* v H\*) to reflect different kind of focus, or it may simply be the implementational effect of a tendency to align the peak more precisely when speaking more carefully, such as might happen when a word is in focus. In other words, he points out that the distinction may reflect a phonological choice or be an effect of implementation, and that it may be difficult to decide between the phonological and the implementational interpretations. Whether this is viewed as phonological or paralinguistic, the more precise alignment of the peak with the stressed syllable might be understood as originating from the effort code[[5]](#footnote-6). (Gussenhoven, 2004, pp. 60–60, in reference to Smiljaníc and Hualde, 2000)

### Labelling in Contemporary AM analyses of English Intonation: ToBI[[6]](#footnote-7) and IViE

It is clear that the labelling of PAs and boundary tones within an IP ought not be viewed merely as a matter of convention or an arbitrary selection of pre-determined pitch accent types. Rather, it is driven by underlying principles, beginning with the understanding that tones are sequences of Hs and Ls in an autonomous tonal tier.

Currently, within the AM research on English, there are two main different approaches to intonational labelling, and they depend largely on whether or not one accepts the existence of and need for phrase accents. In a closely related manner, they also depend on whether the pitch accent is analysed as being right-headed or as left-headed, or—as Gussenhoven (2004, p. 128) puts it—whether the analysis of the pitch accent follows and on-ramp or an off-ramp approach. Broadly, the phrase accent on-ramp approach is mostly followed in ToBI (Tone and Breaks Index) labelling and labelling systems derived from it (Beckman and Gayle, 1997; Beckman, Hirschberg and Shattuck-Hufnagel, 2005). On the other hand, the phrase-accent-free off-ramp approach is adopted in studies associated with the Intonational Variation in English (IViE) project. The IViE project was developed to collect corpora from several urban areas in Britain and Ireland, including Belfast (Grabe, Nolan and Farrar, 1998; Gussenhoven, 2004), and the labelling system bears some principles in common with Gussenhoven’s analysis.

ToBI analyses pitch accents as monotonal or bitonal, with bitonal pitch accents being either left- or right-headed, e.g., L+H\* or L\*+H. It takes a more theory-neutral approach to down step triggers, and uses downstepped H tones (!H\* and H+!H\*) in pitch accents (Beckman, Hirschberg and Shattuck-Hufnagel, 2005). However, downstep is still, as elsewhere, understood to trigger an overall reduction in scaling of H in the IP. Because downstepped H is available in ToBI, there is no need for the H\*+L PA used in Pierrehumbert (1980) since the trailing +L was essentially viewed as a downstep trigger. Further, H+!H\* replaces the H+L\* originally proposed by Pierrehumbert. In line with Pierrehumbert and Beckman (1986), ToBI also includes an obligatory phrase accent, occurring at the right edge of each *ip* and before the final boundary tone. In this way, the edge tones combined resemble the tail of the British Tradition. The final boundary tone retains the same features as the original Pierrehumbert analysis, with H% reflecting an upstep after H- and a high after L-, while L% represents a null state, with little to no downward drift in *f*0.

IViE labelling, like ToBI, has both monotonal and bitonal pitch accents and permits the use of downstepped !H. However, IViE also permits tritonal pitch accents, L\*HL and H\*LH, where the first two targets occur on the stressed and following syllable, with the third target following. This is similar to Gussenhoven, although Gussenhoven only permits tritonal pitch accents in pre-nuclear positions. Further, like Gussenhoven, IViE rejects the phrase accent hypothesis, and instead adopts the fully relational approach to boundary tones suggested by Grabe (1998a). This is in part motivated by the observation that ToBI cannot effectively represent some contours commonly found in nIE, most noticeably the nuclear rise-plateau-slump [see XXX later] (Grabe, Nolan and Farrar, 1998). Moreover, it allows for a more parsimonious labelling system which can facilitate the identification and comparison of parallel pitch accents and boundary structures across varieties of English.

### Phonetic analysis of Intonation in AM: tones and tonal targets

In addition to the phonological analysis of intonation, AM analyses also generally evaluate the phonetic implementation of tones, or tonal targets. In its simplest formulation, a tonal target is a surface realisation of an underlying phonological tone, described in terms of both its *f0* and its temporal alignment in relation to a segmental landmark, frequently the onset of the vowel in the lexically stressed syllable [REFS – explanation needed?].

As discussed in section 2.3.5, the alignment and scaling of tonal targets may be influenced by the pragmatic or linguistic function of an utterance. However, formal structure features are also understood affect the *f*0 scaling and temporal alignment of tonal targets. These include compression and truncation effects, tone drift, and tonal crowding.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| a. L\*H realised fully. | b. L\*H realised with compression effects. | c. L\*H realised with truncation effects. |

Figure . Three stylizations of L\*H pitch accents reflecting truncation and compression effects. The grey area on the right of each box can represent unvoiced segmental material or simply the end of speech.

It has been argued that a pitch contour may be truncated or compressed as a function of available voiced material (Grabe, 1998a, 1998b). In cases of compression, the tonal target may be fully realised but within a shorter time window to compensate for the reduced amount voiced segmental material, as shown in Figure 2.4, panel b. With truncation, however, the pitch movement may be foreshortened, as represented in panel c. [REFS?], so the tonal target is not realised fully. It also possible that pitch contours can be subject to truncation and compression effects simultaneously, such as those found for Belfast English in Sullivan (2012). Truncation/compression trends vary across English dialects, so one cannot assume that effects which in one variety will hold in another (Grabe, et al., 2000; Ladd, 2008).

It also appears that tonal targets may also be fixed or subject to drift. When a target is fixed, it is typically anchored to a specific landmark in the phrase, typically associated with a stressed syllable or the foot [REFS], or sometimes with lexical boundaries [REF]. However, it has been found that, given enough segmental material, and conditions by foot size and anacrusis, a tonal target may drift rightwards or leftwards [REFS]. Analysis of variation in peak alignment strategies has shown that they can be a marker of dialect differentiation, both in English (Arvaniti and Garding, 2007; Kalaldeh *et al.*, 2009) and other in languages (Bruce and Thelander, 2001; Atterer and Ladd, 2004; Dalton and Ní Chasaide, 2007).

The third formal effect is tonal crowding (Silverman and Pierrehumbert, 1990; Arvaniti, Ladd and Mennen, 1998; Ladd *et al.*, 2009). This occurs when tonal targets must be realised in close proximity to each other, such as when two starred tones occur in adjacent syllables. It has been found that in such cases, there can be interactions affecting the alignment or the alignment and scaling of targets, most likely as adjustments must be made to accommodation both tones within a small window of time.

Without consideration for Truncation/compression effects, the effects of tonal drift, and tonal crowding, it may be possible to misidentify fundamentally identical pitch accents as different. Through systematic analysis of pitch accent timing and scaling under different segmental and/or metrical conditions—longer feet meaning more segmental material—one can assess whether truncation or compression effects are in operation, if targets are subject to drift under different formal conditions, or if (and how) tonal crowd affects the realisation of tonal targets.

## Issues for AM analysis of the phonology and phonetics of intonation

So far, the main difficulties discussed regarding AM analysis have focused on the theoretical divisions motivating ToBI-style and IViE-style analysis. However, there are a few other fundamental issues which need to be considered in relation to AM analysis. The following section describes three of them, indicating why they matter, and explaining how they have influenced the research project.

### Tonal targets and implementational domains

Within the AM approach, the domain of the starred tone is, by definition, the metrically stressed syllable since the star simply represents the stressed syllable which links the segmental string to the tonal sequence. There are however, several problems with this assumption, one of the most significant being that the starred tone does not in fact always align with the stressed syllable (e.g. Nolan and Farrar, 1999; Arvaniti, Ladd and Mennen, 2000). For example, Arvaniti, Ladd, and Mennen analysed pre-nuclear tones in Greek, which appear to have a low tonal target before the stressed syllable and a high tonal target after it, making it difficult to identify a single tone associated with the stressed syllable, and this make it difficult to land on an appropriate label (L\*H, L+H\*, [LH]\*, H\*, etc.).

The failure of starred tones to align consistently and neatly with the metrically stressed syllable could stem from several sources. It could be an effect of segmental pressure, i.e., it might sometimes be difficult to coordinate pitch targets with metrical targets when there is less segmental material available in which to implement the pitch movement [REFS]. Alternatively, it could be that tonal targets are subject to leftward (or rightward) drift when there is an excess of segmental material before the metrically stressed syllable, as in the case of anacrusis [REFS]. Or it could be pragmatic, where the speaker aligns the target only as precisely as they believe necessary for the speaker to interpret the intended pitch accent correctly. Of course, it could well be a combination of all of these, wherein the speaker may be more or less likely to cede to segmental pressure or to let the target drift based on the perceived need to realise the pitch accent.

A closely related issue is the implementational domain of the pitch accent. This is generally taken to be the foot. However, as Arvaniti, Ladd, and Mennen (2000) point out, it is not always clear how the trailing tone is supposed to be aligned. There is evidence that starred tones and trailing tones in English may be aligned partially according to the left and right boundaries of lexical word in which foot stress occurs rather than the foot alone (Silverman and Pierrehumbert, 1990). There is also evidence that pitch alignment may be partially lexically motivated in that it plays a role in the segmentation of words (Ladd and Schepman, 2003).

There is also debate as to how to identify tonal targets in the first place. The simple view takes the local *f0* maxima and minima as the tonal targets. However, while this seems sensible at first, there can be complicating factors. For example, a study of British English by Knight and Nolan (2006) found that the most stable landmark of the H tone was not the peak but the end of the H plateau 4% below the *f0* maximum (i.e. the end of the effective duration of the H tone). Greater consistency in the alignment of the end of the plateau rather than that of peak *f*0 may suggest that the speakers too may intuitively aim at a target near the end of the plateau rather than simply the peak itself. In other cases, the listener may be able to identify the percept of an L target but there may not be an easily identifiable local *f0* minimum in the contour. A stylized example of this problem is shown in Figure 2.5. In such cases, the point in the contour where a distinct change in *f0* trajectory may be selected, as shown in the first panel of Figure 2.5. Such points are described as *elbows* or *turning points*, and have been used to identify L targets in several studies (D’Imperio, 2000; Welby, 2003; Shosted, Giudice and Arvaniti, 2006). In fact, the use of turning points can be generalised to interpret all tonal targets in this manner. Frequently, as in the second panel of Figure 2.5, *f*0 maxima and minima do align with turning points but this is not always the case. The topic of elbows is taken up further in Chapter 10, which outlines a technique for identifying turning points.

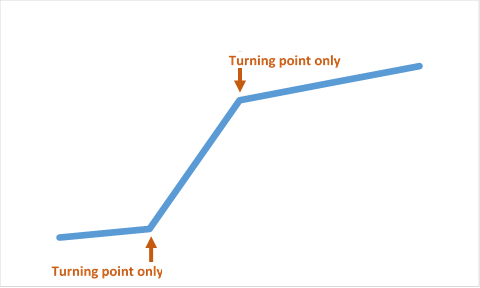
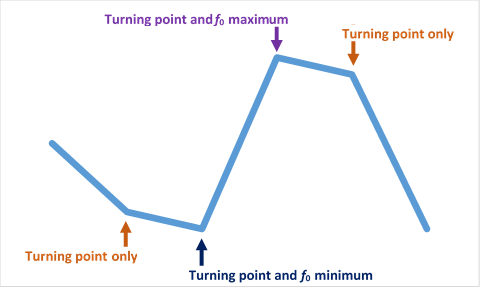
a.  b. 

Figure . Stylized representations of an f0 contour, indicating turning points (elbows) and f0 minima and maxima. Note that an f0 minimum or maximum may also be a turning point.

In summary, we can see that neither the identity implementational domain of the turning point nor its most salient identifier be taken for granted.

### Form-function mismatches and failures in phonological and phonetic analysis

AM analysis of intonation typically involves identifying functions associated with intonation and then describing them in terms of intonation structure. Examples of this can be found in the analysis of Eastern European Question Intonation [REF], question types in Dutch [REF], Greek [REF], German [REF], and Italo-Romance varieties [REF]. In English, this can be seen around the discussion of question modes, calling contours, and focus [REFS]. However, as noted above in section 2.3.5, pitch contours sometimes vary according to function in a manner which cannot easily attributed to different combinations of pitch accents and edge tones. Such variation is often viewed as gradient and is typically attributed to a paralinguistic effect. However, there is a very clear methodological problem with this approach. In essence, it means that pitch features associated with functions are identified as linguistic only in so much as they fit the theory. When the pitch contours do not fit the theory, they are in effect cordoned off as paralinguistic. This occurs even in cases where there is a clear and consistent correlation between grammatical function and the pitch contour, which suggests that the different contours do in fact serve a linguistic function too. While there is clearly a paralinguistic component to pitch, the problem here lies in the fact that the theory is essentially prioritized over the data, when in fact it may be that the theory needs to be adapted or reimagined so that it can explain the data. In short, the availability of the paralinguistic ‘out’ may hinder progress in the AM analysis of intonational function.

Grice *et al.* (2017) illustrates and deals with the problems of discreteness, gradience, and function in a particularly insightful manner. In an analysis of narrow and corrective focus in Standard German, the researchers found—as they expected—a relatively low correlation between intended focus type and the pitch accent employed, but they found that the focus type was still more likely to be interpreted as intended. More importantly, however, they found that, despite the lack of consistency in pitch accent choice, speakers did still employ similar implementational strategies. Specifically, peaks were aligned later and with higher *f*0 in contrastive focus when compared to narrow focus, and in narrow focus when compared to broad focus. The authors note the difficulties their results pose for linguistic traditions based on discrete categories of analysis, since their research identifies a mismatch between the phonological categories and their function while at the same time identifying clear patterns and gradient shifts associated with changes in function. They argue that there is a need to integrate discrete and continuous features in an intonational grammar, since both belong to one single system.

As the authors suggest, it may well be necessary to carefully consider both discrete and gradient features in a systematic analysis of function in intonation. However, it is also necessary to consider possible limitations to the phonological framework within which the analysis is conducted and to consider possible adjustments to it. For example, in the study described above, discrete categories were generally poor predictors of function, but an overall systematic trend was found in the alignment and scaling of *f*0 targets from broad to narrow and from narrow to contrastive focus. This should, perhaps, not only signal the importance of assessing both categorical and gradient patterns in the analysis of intonation, but it should also signal that the current phonological description may not adequately identify and reflect phonological features which are easily interpreted by listeners during conversion.

In short, it is important to avoid the temptation to simpy perform a phonological analysis which dismisses some phenomena as paralinguistic [i.e. scaling in declaratives and questions] rather than linguistic. Therefore, if one identifies a systematic relationship between function and an ostensibly gradient feature, one must consider if such a feature can in fact be accommodated by adjusting the phonological theory instead excluding it from the phonological.

### Contours versus Targets

At its core, the AM framework relies on the assumption that the pitch contour is the physical manifestation of an underlying sequence of tonal primitives. As such, glides and curves should be understood as side effects of this implementation. However, this does not always appear to be the case. For example, D’Imperio (2000) found that the perception of peak alignment—and thus of pitch accents—can be influenced by the shape of the contour rather than tonal target alignment alone, while Knight (2008) found that the perception of pitch height and pitch prominence is influenced by the shape of the pitch accent, specifically the duration of the plateau associated with the target pitch accent.

Building on such insights, Barnes *et al* (2012, 2021) argue that the perception of tonal targets may not be associated so much with turning points but with a global measurement, Tonal Centre of Gravity (TCoG). TCoG abstracts away from the alignment and scaling of *f*0 by integrating them into a single function which calculates the time point which the authors describe as the “centre of gravity” of the contour. It is in effect an integral function which calculates the time at which the area under the curve is balanced on either side of the curve, described more precisely as TCoG-t in Barnes *et al* (2021). Figure 2.6 helps illustrate this point. Each shaded area represents half the area under the curve between the onset and offset of the pitch movement. The frequency domain equivalent of TCoG-t is TCoG-*f*0, which measures the mean *f*0 across the contour. Whenever the glide to the peak is more domed, TCoG-*f0* is higher and TCoG-t is earlier. Conversely, whenever the glide to the peak is more scooped, TCoG-*f0* is lower and TCoG-t is earlier. These effects are reflected by the different positions of the dots in Figure 2.6.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| a. | b. | c. |

Figure . Schematic representations of different pitch curves with different TCoG. Adapted from Barnes et al (2021) to illustrate how TCoG-t reflects timing of the balance of the area under the curve.

In their 2012 paper, the authors were able to demonstrate that the TCoG-t was a better predictor of annotators’ PA categorization choices than turning point timing and scaling parameters. In the 2021 paper, the authors used forced choice match-to-sample tasks with utterances containing resynthesized pitch contours. They demonstrated that the shape of both the rise and fall influenced listeners’ categorical judgments.

The alignment effects predicted by TCoG appear to reflect the findings of D’Imperio (2000), in that a domed rise leads to the perception of an earlier peak. Similarly, *f*0 effects of TCoG appear to reinforce the findings of Knight (2008), where the more plateau-like domed contours are associated with a higher TCoG-*f*0.

Both papers present a serious challenge to the AM approach. They use AM conventions to describe the PAs, which appears to validate the fundamental principle of the AM framework and the importance of underlying L and H primitives along with their associated tonal targets. However, the analyses suggest that listeners do not categorise pitch accents based on the implicit identification of tonal targets, but rather on the interpretation of the tonal centre of gravity. If this is indeed the case, i.e., that listeners use TCoG rather than tonal targets to identify and categorise pitch accents, then it stands to reason that the interpretation of intonation events in terms of L and H primitives is wrong-headed. As such, it brings into question the whole AM project and might suggest that the use of terms such as L+H\* or L\*+H are useful only in so far as they are convenient placeholders for PA categories, while the L and H terms themselves hold little theoretical merit. [Could there be an underlying LH phonology with a realisation in contour shape]

This train of thought is not merely a panicked catastrophizing, since the TCoG analysis does indeed appear to reflect listener’s ability to categorize pitch contours without recourse to tonal targets or the underlying primitive with which they are associated. However, there are two major considerations which may help ease the distress of such catastrophic thinking.

The first consideration relates to the psychological plausibility of TCoG. It would require a set of complex cognitive processes from the listener for it to work:

1. Continuous summation of *f*0 over time (a cognitive integral function);
2. Retrospective assessment of the timing of the onset and offset of the complete contour;
3. Retrospectivecalculation of the time at which the area under the curve can be divided equally in half.

In addition, the speaker would need a complementary set of production strategies, including the planning of the contour so as to calculate the area under the *f*0 curve and an implementation strategy to generate the target TCoG so that it can be interpreted correctly by the listener. While, such cognitive processes *may* occurin the production and perception of PAs, it feels more practical to assume a simpler process in which the pitch contour is produced through the implementation of a linear sequence of L and H primitives. In short, it is unlikely that the predictive value of TCoG is due to the fact that it directly represents the underlying phonology. Rather, it is more likely that it is an epiphenomenon which works as an effective heuristic in the categorisation of pitch contours.

The second consideration is the way in which tonal targets are identified. In Barnes *et al* (2012), tonal targets are associated with *f*0 minima and maxima. The authors explicitly reject the use of elbows as tonal targets. They argue that, since the most common algorithm for estimating elbows is sensitive to the contour shape much in the same was as TCoG, the use of elbows “in essence amounts to smuggling global contour shape in through the model’s back door.” (Barnes *et al.*, 2012, p. 379). However, it is very possible that this observation, relegated to a footnote, explains the apparent effectiveness of TCoG. That is, TCoG may appears so effective is that it reflects the way in which tonal targets are realised: as elbows rather than as *f*0 minima and maxima. In this case, the apparent challenge to the AM framework might be an artefact of the methodology. In other words, the problem might lie in the technique used to identify tonal targets.

The same concern arises in Barnes *et al* (2021). Because the study described in this paper used synthesized *f*0 contours, low *f*0 targets could easily be measured at the point at which the synthesized contour began to rise and at the point at which it was complete, while high *f*0targets were measured at the *f*0 peak. Considering that the scooped rise begins gradually, it is possible that the *f*0 minima is not as immediately salient as in the domed rise, where *f*0 rises suddenly and dramatically. Again, the experimental method for identifying the tonal target may not accurately reflect the listener’s percept.

The conclusion one might draw from this discussion of TCoG is that tonal targets are best measured in terms of elbows rather than *f*0 minima and maxima. However, it was not until later in the project—once I had designed the methodology, conducted the analyses, and was reflecting on the results—that I began to consider more fully the implications of TCoG and the limitations of using *f*0 minima and maxima as proxies for tonal targets. As such, the issue of turning points and TCoG is given fuller consideration in Chapters 8 and 9.

## Summary and conclusion

The dominant mode of intonation analysis in English is the Autosegmental Metrical (AM) approach, and it has been adopted for the current project. In essence, there are two main schools of thought in AM research. This first is a direct descendent of Pierrehumbert’s original (1980) approach, and has been propagated though ToBI and its offshoots (Beckman, Hirschberg and Shattuck-Hufnagel, 2005). The second is more closely aligned with the work of Gussenhoven (1983, 2004), Grabe (1998a), and the IViE project (Grabe, Nolan and Farrar, 1998; Grabe and Post, 2002). The features for each approach are summarised in Table 2.2. The terms ToBI-like and IViE-like are used for convenience, and it should also be noted that while both ToBI and IViE are intended as largely theory-neutral approach, both are still influenced by competing theoretical perspectives.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table . Comparison of ToBI-like and IViE-like analysis   |  |  |  | | --- | --- | --- | | Features | ToBI-like | IViE-like | | Onramp PA |  |  | | Offramp PA |  |  | | Phrase accent |  |  | | Intermediate phrase |  |  | | Obligatory final boundary |  |  | |  |  |  | |

Fundamentally, both approaches analyse the same set of phenomena, follow the same overall theoretical framework, and share the same general understanding of the relationship between the underlying phonology and the surface form. Both approaches agree that there is a string of underlying tones which are linked to the segmental string through metrical events. The key difference lies in how the events in the tonal string are associated with events within the IP. As can been seen from Table 2.2, this comes down to whether or not the approach uses an exclusively offramp approach, whether it accepts the existence of the phrase accent and the intermediate phrase, and whether or not the final boundary tone is obligatory.

### Labelling choices for research

ToBI-style analysis requires the labelling of the PA final edges in terms of phrase-accents and boundary tones, even in cases where there is little evidence for one whereas the solution adopted by Grabe, Gussenhoven, and the developers of the IViE project feels intuitively more appealing. It does away with phrase accent, allows for an optional the final boundary (L, H, Ø), and includes an off-ramp only approach. This creates an overall more economic labelling system and appears more suited to capturing parallelisms across pitch accent and boundary conditions than the obligatory phrase-accent and boundary tone approach, which can lead to fundamentally similar tunes being analysed as if they are very different.

Gussenhoven (2016) argues that an off-ramp analysis which does away with the phrase accent and permits optional phrase-final boundary tones is able to capture a wider range of attested nuclear contours than ToBi-like annotation. However, if one is to adopt Gussenhoven’s approach, it is also necessary to accept a long list of additional claims about the nature of the implementation of the phonology. These may be absolutely correct; however, in terms of analysis, it feels unwise to adopt, wholesale, Gussenhoven’s approach, as this could, again, lead to a prioritisation of the theory over the data.

For the purposes of labelling, the IViE system seems preferable overall compared to ToBI. After all, IViE was designed for labelling varieties of English, including nIE, which is the focus of this dissertation, whereas ToBI (specifically MAE\_ToBI) was designed with North American English in mind. The fact that IViE has a more flexible approach to final boundaries also makes it more amenable to the analysis of northern Irish English, as shall be seen in Chapter 3. Finally, like Gussenhoven, there is no phrase accent in IViE, which means that there is no need to include labels for tones for which there may not be empirical evidence.

For the reasons outlined above, IViE labelling is used for the analyses reported in Chapters 6 and 7, described in the introduction as the phonology-first approach.

### Strategy for identifying tonal targets

For the purposes of identifying tonal targets, the analyses in Chapters 6 and 7 use *f*0 maxima and minima rather than *f*0 turning points. This decision was based on three things. Firstly, minima and maxima are intuitively easier to reconcile with concepts of High and Low than turning points. Secondly, they are commonly used in the AM literature [REFS]. Lastly, they are much easier to identify than turning points. As mentioned in 2.4.3 above, however, it was only on reflection and with the benefit of experience that the potential advantage of using turning points over *f*0 maxima and minima became apparent.

### Analysis of form and function

As noted in 2.4.1, the alignment of tonal targets may not always reflect the theoretical ideal, i.e., starred tones may not be realised in the stressed syllable and trailing tones may sometimes be associated with lexical boundaries rather than metrical ones. This demonstrates the need for an analysis of formal (metrical and lexical) effects on the implementation of pitch accents. Chapter 3 will outline some of the formal effects found on the realisation of pitch accents in nIE, while Chapter 6 offers an analysis of the metrical and lexical boundaries effects on pitch accent inventories and the alignment and scaling of tonal targets in DCE.

Section 2.4.2 points out that the partitioning of pitch events into phonological/linguistic and paralinguistic categories may mean that some intonational phenomena are inadvertently dismissed as paralinguistic because they do not ‘fit’ the theory, even though they still systematically reflect a communicative function. To avoid such a procrustean trap, it is suggested that one must consider the limitations of the theory and see if reasonable accommodations can be made to the theory so that it can explain systematic patterns in the data more completely. [REGISTER!] Chapter 3 will indicate how a strict AM analysis of sentence modes in nIE might lead to phonological differences being ascribed to paralanguage. In such cases, adjustments to the theoretical approach may demonstrate how they too are, in fact, phonological. Potential adjustments to the phonology are considered in the analysis of sentence modes in chapter 7.

# Local Context: northern Irish English and Derry City

This chapter focuses on Derry City and northern Irish English (2.4), outlining the geopolitical and linguistic contexts, and summarizing previous research on Intonation in northern Irish English and Derry City. It begins by presenting some general background information about Derry City, and northern Irish English, before moving on to discuss a British Tradition analysis of DCE intonation and AM studies of nIE intonation. It ends with a consideration of how the issues within AM discussed in the previous chapter (section 2.4) related to the current study of DCE intonation.

## Derry City and northern Irish English

Derry City is located in the Northwest of Northern Ireland, close to the border with Donegal in the Republic of Ireland (see Figure 3.1). It has a population of 83,125 (NISRA, 2015). The oldest settlements of the city were along the west bank of the river Foyle, although the city now straddles both the eastern and western banks of the river. The area on the western side is referred to locally as the City Side, and the area along the eastern bank is called the Waterside. Before the 1612 Charter, the site along the West Bank of the river Foyle was part of Donegal (Lacy *et al.*, 1983), and historically there has been a close relationship between Derry City and county Donegal.

As in Northern Ireland in general, Derry City is in many ways demographically homogenous. In the 2011 census, 98% identified as white, 97% as Christian or as having been brought up as Christian. 89% were born in Northern Ireland, with less than 3% of the population born outside of either the UK or the Republic of Ireland. Two very strong markers of identity in Northern Ireland, however, are nationality and religion, and there is a general tendency for Roman Catholics to identify as Irish and for Protestants as British (Zwickl, 2002, pp. 72–101). The city has a large Roman Catholic majority (78% in 2011 census), with most of the Protestant population (19% in total) living on the Water Side. 59% of the population identify as Irish, 34% as Northern Irish, and 21% as British. (The total is over 100% as the census allows people to identify with several nationalities.)

Varieties of English spoken in the northern part of the island are quite distinct from southern varieties; unfortunately, the term Northern Irish English—often found in the literature—is intrinsically ambiguous. That is, it may refer the Irish English spoken in the geographical north or to English spoken in Northern Ireland, the political jurisdiction, the border of which is shown by the thick black line in Figure 3.1. In some cases, the distinction between Irish English of the geographical north and the English of (the political entity) Northern Ireland is blurred, so a northern variety of Irish English—such as Belfast English in Grabe, Kochanski and Coleman (2005)—might be contrasted with English in the Republic of Ireland, in such a way as to imply that the political border and the linguistic borders coincide. The confusion of conflating the political boundary with the isogloss can also been seen in Folley, Gibbon, and Peppé, who observe that “[a]lthough statements have a high terminal in Northern Ireland […], this is not the case in the Irish Republic” (2010, p. 23). Conversely, Moritz (2016) includes northern varieties spoken in the Republic in her discussion of English in Northern Ireland. The two-book survey of Irish English by Corrigan (2010) and Kallen (2013) explicitly splits the burden of description into a volume on Northern Ireland and a volume on the Republic of Ireland, but both books exclude the English of Donegal, which is the geographically the northernmost county on the island but also in the Republic of Ireland.

|  |
| --- |
| Derry city  Ulster border  North-south isogloss  NI-ROI border  Figure .. Approximate boundaries of northern Hiberno-English dialects (Hickey, 2007, p. 442 after Harris, 1985, p. 16). Derry City is highlighted in red. |

Harris (1985) avoids such geopolitical confusion, conflation, and exclusion when he identifies an isogloss running roughly from Donegal Bay in the west to Carlingford Lough in the East, which separates northern varieties from southern varieties of Irish English, This is shown in Figure 3.1 by the dashed purple line. Harris describes the northern variety as Ulster English and identifies three main language groups: Southern Ulster English, Ulster Scots, and Mid-Ulster English (MUE), each influenced to varying degrees by the influx of Scots speakers and English speakers from the English Midlands during the seventeenth century plantations. It should be noted that the isogloss is not coterminous with the Ulster boundary (shown by the pink dotted line), especially in south Ulster. Therefore, following McCafferty (2001), lowercase northern Irish English (nIE) will be used to describe the varieties of Irish English spoken in the northern part of the island.

As can be seen from Figure 3.1, Derry city is in the MUE region. Thus, one expects broad similarities between DCE phonology and the phonology of this region in general. In fact, one well recognised similarity in across all nIE varieties is the use of rising nuclear pitch accents in neutral declarative sentences [SEE CHAPTER XX.X.X]. It is this specific feature of nIE which motivated this research, since, as shall be seen in Section 3.4, the prevalence of rising intonation patterns raises questions about form and function, and about linguistic/paralinguistic boundaries of pitch.

## British Tradition analyses of northern Irish English Intonation

In the 1970s and early eighties, two studies of northern Irish intonation were conducted within the British Tradition, the first focusing on Belfast English and the second on Derry English.

In their study of Belfast English, Jarman and Cruttenden (1976) (J&C) found that 70% of all nuclei contained by rising tones, including semantically unmarked (i.e. simple declarative) phrases. Working within the British tradition, they observed that low rising tones (tone 1) were common in declaratives while high rising tones (tone 2) were common in questions. In each case, they noted the post-tonic stretch formed a plateau. Although the rise of semantically unmarked nuclei was less dramatic than that of questions, the importance of their findings at the time was that they drew attention to the weakness of the assumption that unmarked intonation is ‘universally’ indicated by a falling tone.

McElholm’s (1986) study of intonation in Derry English found that it was similar Belfast intonation as described in J&C, especially in relation to the use of rising tones in unmarked forms. However, unlike J&C, McElholm found no examples of falls in nuclear syllables, and he also found that the low-rising tone was used for wh-questions. Where McElholm’s analysis differs from the Belfast study, McElholm suggests it may be due to a sparsity of data or a possible discrepancy in social class between the Derry informants and the Belfast informant. Therefore, McElholm’s overall inventory of tones, summarised in Table 3.1, is slightly different to that of J&C.

Table . Summary of McElholm's inventory of nuclei for Derry English (adapted from McElholm, 1986, p. 56). Parentheses indicate corresponding tonal forms in Jarman and Cruttenden (Jarman and Cruttenden, 1976).

|  |  |  |  |
| --- | --- | --- | --- |
| Tone | stylization | tonic movement | general meaning |
| A (1) |  | low rising | natural for all major speech functions except YNQs; also used for secondary information |
| B (2) |  | high rising | neutral for YNQs |
| C (3) |  | rising-falling | contrastive |
| D |  | extra-high rising-falling | assertive or surprised [NB very rarely attested] |
| E (5) |  | rising plus rising | conveying new plus secondary information |
| F |  | extra-rising rising-falling plus rising | as tone D plus secondary information |
|  |  |  |  |

Like J&C, McElholm carried out his analysis in the British Tradition, and was largely an impressionistic analysis of two speakers. The differences in theoretical approach and the lack of quantitative data make it is difficult to compare it with AM studies beyond a few broad phonological comparisons.

## AM Studies of northern Irish English Intonation

Within the AM approach, the majority studies of nIE are based on IViE data, and thus focus on Belfast English. A few studies of Donegal English have been conducted based on work by O’Reilly, Dorn and Ní Chasaide (Dorn, 2006; Kalaldeh, Dorn and Ní Chasaide, 2009; O’Reilly, Dorn and Ní Chasaide, 2010). The only other AM study of nIE of which I am aware is Moritz (2016), which compares Southern Ulster English, Ulster Scots, and Mid-Ulster English, with Belfast English representing Mid-Ulster English.

The IViE project collected speech corpora from secondary school pupils in different urban areas across Britain and Ireland (Grabe and Post, 2002), including Belfast. The Belfast corpus includes speech from 12 speakers (6 female, 6 male, aged 17) attending one of two schools near the city centre, all of whom had been born and grown up in Northern Ireland. Speech was elicited using five tasks to elicit a range of speaking styles, speech functions, and interactions.

### Phonological analysis of Belfast English

Lowry’s initial (2001) analysis of nIE followed the ToBI labelling approach and found that adjustments needed to be made to it in order to accommodate features of nIE which are not found in the General American English for which ToBI was designed, most notably in ToBI’s inability to adequately label the rise-plateau-slump of Belfast English. Later, using the IViE system, she identified four different nuclear accents in Belfast English, labelled L\*H %, L\*H L%, L\*H H%, and H\*L % respectively, as shown in Table 3.2 (Lowry, 2002). These patterns have been attested in a number of analyses of the IViE corpora (Lowry, 2002; Grabe, 2004; Grabe, Kochanski and Coleman, 2005) as well as in more recent corpora (Sullivan, 2010, 2012; Jespersen, 2018). L\*H % is by far the most common tone across sentence modes. However, differences in the distribution of nuclear tones across speech style, gender, and sentence mode have also been found.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table . Nuclear patterns of Belfast English, adapted from Lowry (Lowry, 2002).   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Schematic representation |  |  |  |  | | Impressionistic description | rise-plateau | rise-plateau-slump | high rise | fall | | IViE labelling | L\*H % | L\*H L% | L\*H H% | L\*H % | |

Lowry’s (2002) study of style shift found that rising nuclei dominate all styles of speech, and their use increases dramatically in less careful speaking styles, especially among the female speakers. Grabe’s (2004) analysis of the IViE read-sentence corpus found that L\*H % accounted for a least 83% of nuclear tones across all sentence types. H\*L % was found in only 4.2% of declaratives (DECs) and 5.6% of wh-questions (WHQs) while low boundary tones, L\*H L%, occurred only in 12.5% of DECs. The simple rise, L\*H H%, was found in yes-no questions (YNQs) but was not the dominant form, accounting for 5.6% of YNQs and 16.9% of declarative questions (DCQs). Sullivan (Sullivan, 2010) found L\*H H% in DECs as well as YNQs. She also found L\*H L% was more likely to occur in question forms than DECs. In a more recent study, Jespersen (2018) has also found occurrence of simple rises.

Another study of the IViE data (Grabe, Kochanski and Coleman, 2005) analysed PAs across utterances in both pre-nuclear and nuclear position. They found pre-nuclear accent types (H\*, L\*H, H\*L, and L\*), with H\* being by far the most common, occurring in 78% of declaratives all utterances. In sharp contrast to this, L\* occurred in only 2.2% of PN pitch accents.

Table 3.3 summaries the inventory of nuclear accents attested in Belfast English across sentence modes. If the same similarity exists today between Belfast and Derry City English as described in J&C and McElholm, one might expect a similar distribution of pitch accents, though with fewer—if any—instances of H\*L Derry City English.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table . Summary of Pitch accents attested in Belfast English across sentence modes.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Nuclear | DEC | WHQ | YNQ | DCQ | | L\*H % |  |  |  |  | | L\*H L% |  |  |  |  | | L\* H% |  |  |  |  | | L\*H H% |  |  |  |  | | H\*L % |  |  |  |  | |

Unmarked rising nuclei are also attested other varieties of English, such as Liverpool, Manchester, Tyneside, and Glasgow, and the phenomenon has been described as Urban Northern British intonation (UNBI) (Cruttenden, 1995, 2001, 2007). In the IViE corpus, there are four sets for data from Britain in the UNB category: Liverpool, Leeds, Bradford, and Newcastle upon Tyne. However, nuclear rises in declarative statements were attested only in the Newcastle corpus. Even still, Newcastle had 17% L\*H in declaratives compared with 96% in Belfast (Grabe, 2004). Cruttenden also (2007) notes the evidence for the greater dominance of unmarked rises in Belfast English when compared to other varieties.

### Alignment, compression and Truncation in nIE

Nolan and Farrar (1999) used the IViE corpus to study peak lag in pre-nuclear accents, which was defined as the occurrence of the tonal target after the stressed syllable. They found that it was very common in Belfast but was less so in the presence of anacrusis. Sullivan (2007) focused on the alignment of nuclear valleys, also using the IViE corpus. She found a significant effect of anacrusis on the alignment of L, causing earlier alignment. Gender was also a significant factor but sentence mode had only a limited effect.

A comparative study of nuclear and pre-nuclear alignment in varieties of Irish English (Kalaldeh, Dorn and Ní Chasaide, 2009) examined L\*H PNs in Donegal English rather than H\*. The study found that anacrusis caused rightward drift of the pre-nuclear peak, which appears to be the opposite of Nolan and Farrar (1999). The Donegal data also indicate rightward drift of peaks in nuclear L\*H pitch accents as the syllable count in the foot increases. For both nuclear and pre-nuclear accents, L targets were stable.

### Proposed source of declarative L\*H dominance in nIE and elsewhere

Cruttenden suggests that the UNBI phenomenon may have originated in migration from Belfast and other regions of northern Ireland. Offering an alternative hypothesis to account for the existence of rises in Tyneside—which was not subject to the same degree of migration from Ireland—Hirst (1998, 2013) has suggested that the declarative rise may have originated in migration and settlement from Scandinavia during the Viking raids starting in the eighth century. However, the distribution of Scandinavian settlements do not reflect the distribution of the rises in Ireland or in other parts of England either. So, while Tyneside English may have retained prosodic features adopted during Norse settlement, it seems unlikely that it can account for the occurrence of nuclear rises elsewhere. Therefore, it is seems more reasonable to suggest that the Scandinavian hypothesis is plausible for the declarative rise in Tyneside English while the northern Irish migration hypothesis is more reasonable for areas such as Liverpool and Glasgow.

This still does not explain the origin of the rise in nIE. In fact, rising nuclear tones dominate not just nIE but also northern varieties of Irish Gaelic (Dalton and Ní Chasaide, 2007; O’Reilly, Dorn and Ní Chasaide, 2010). The realignment hypothesis proposes a possible source for this, namely that diachronic rightward drift of the H\*(L) peak has led to a phonological realignment of H, and thus the rise—so to speak—of L\*H accents. Dalton and Ní Chasaide (2005) found the realignment hypothesis unlikely for Donegal Irish, and Sullivan (2010) found little evidence for it in Belfast English. Sullivan, however, proposes a transfer hypothesis, which states that the unmarked use of L\*H is due to transfer from other functional domains, such as continuation forms or question rises. Sullivan found more phonetic similarity between statements and continuation forms than between statements and questions, suggesting transfer from the continuation function to declarative forms.

### Quantitative analysis of intonation in Belfast

Grabe, Kochanski, and Coleman (2003) used the IViE read speech corpus to conduct a quantitative analysis of pitch trajectories across sentence modes and dialects, including Belfast English. They used time- and *f0*-normalised data to facilitate comparison of contours. The study found that, as with many previous studies (including Sullivan, 2010, 2012), mean *f0* increased in question forms, with DCQs having the highest average *f0* in all dialects (see Figure 3.2). The authors were able to show that the *f0* average and *f0* slope contribute to the distinction between sentence modes across dialects. This could be interpreted as a phonetic pitch raising effect similar to that found in Haan’s (2002) study of Dutch question forms. Such findings also reinforce the argument from Grice *et al* (2017) that the analysis of phonetic/gradient data must be integrated with analysis of phonological/categorical data to fully capture similarities and difference in intonation across different functions.

|  |
| --- |
| -0.15  -0.10  -0.05  0.00  0.05  -0.15  -0.10  0.00  0.05  -0.05  C0: average *f*0  C1: *f*0 slope  DECQs  DECs  WHQs  and YNQs  Belfast  Bradford  Cambridge  Dublin  Leeds  London  Newcastle  Figure . Average f0 (x-axis) plotted against the global slope of f0 (y-axis) for four utterance types and seven dialects. Adapted from Grabe, Kochanski and Coleman (2003) |

## Implications for an AM analysis of intonation in DCE

Chapter two focussed exclusively on intonation theory, with a particular focus on the AM approach. This included a discussion of some problem areas within AM and ended it an argument for the importance of both a formal and functional analysis of intonation in a target language or language variety. This chapter has dealt with intonation in relation to nIE and Derry City English. It is now time to consider the issues raised in chapter two might inform the research aims of a phonetic and phonological analysis of DCE.

### Does nIE provide evidence for the special status of H targets [AND NUCS?]

Aside from the study by Sullivan (2007), the vast majority of studies on tonal alignment focus on H targets, looking at issues such as peak alignment, peak lag, and even tonal centre of gravity, which is essentially a means of identifying (perceptual) peaks. This is primarily because H targets dominate unmarked starred tones in a quasi-universal manner, i.e., H\* and H\*L are much more common as unmarked PAs than L\*H and L\*. However, all the evidence suggests that L\*H is the dominant nuclear PA in nIE. At the same time, as noted in section 3.3.1, an analysis of the Belfast IViE data indicated greater variation pre-nuclear pitch accents.

Increased variation in PN pitch accents could indicate that pre-nuclear PAs also signal variation in meaning; however, given that nuclear pitch accents are more closely associated with communicative function than pre-nuclear accents, this seems unlikely that they would contain more variation in order to signal a greater variety of meanings. It is more likely, therefore, that variation across PN pitch accents stems from the fact that they are actually communicatively less important, and therefore, speakers are less apt to realise them with the same care and consistency as the nuclear pitch accent. If it is the case that PN pitch accents vary in form because they have a much lower functional value, it is very interesting to note that, in the IViE Belfast data, where they are attested, they are very likely to be H\* and very unlikely to be L\*. This is interesting because it might speak to the importance of H tones over L tones, in that, when given the choice of either deleting a tone in L\*H or employing L\* or H\*, speakers prefer H\*. Such a finding would imply that PN H\* pitch accents are, at least some of the time, essentially a reduced for of L\*H.

In the Donegal English data discussed in section 3.3.2 [REF], PNs were identified exclusively as L\*H, and the temporal alignment of the L target was remarkable stable, even under varying foot-size and anacrusis conditions. Therefore, it cannot be argued that de-prioritization of L targets is a general feature across nIE varieties, let alone that it is universal.

Given the general importance of H tones—and especially H\* PAs—across languages and language varieties, it is definitely worth examining trends in the realisation of PN pitch accents in DCE, both in terms of phonological inventory and phonetic implementation. It may also be valuable as tool for evaluating evidence for the special status of H tones, even in DCE, where L\*H is generally expected to dominate. This provides an additional motivation for the analysis of formal effects on pitch accent realisation previously discussed in Chapter 2, (2.4.1 and 2.5.3).

### Is there evidence for a register tier in nIE

McElholm’s (1986) study of DCE identified a functional contrast between low and high rising pitch accents reflecting a contrast between unmarked declaratives and unmarked YNQs. AM analyses of nIE varieties, however, has found that one nuclear contour dominates across sentence mode functions, namely L\*H %. L\*H H%—which can be interpreted as analogous to McElholm’s high rise—has also been found in DCQs and YNQs, but L\*H % still dominates. In terms of the prosodic signalling of function, this suggests that speakers generally expect the listener to rely on inference to interpret the illocutionary force of DCQs correctly, at least if we look at PA contrasts alone. When we also take into consideration continuous parameters, such as the slope and scaling of *f*0 across different functions, it is clear that they help signal the difference between DECs and DCQs not only in nIE but in other varieties of English in Britain and Ireland (section 3.3.4). In fact, the analysis by Grabe, Kochanski and Coleman (2003) almost implies a cross-dialect categorical contrast between DECs and DCQs in the scaling and slope of *f*0. Of course, in most varieties in this analysis, this is generally accompanied by a phonological contrast between H\*L % in DCQs and L\*H H% or L\*H % in DCQs. In nIE, however, there no such parity between the continuous and the discrete data has been identified.

The AM argument, championed by Gussenhoven (see sections 2.3.4 and 2.3.5), maintains that parasitic traces of pre- or paralinguistic uses of pitch remain in the pitch trace alongside the linguistic component. These are distinct from the categorical grammatical uses of pitch accents and boundary tones. When this argument is considered in relation to nIE, it means that nIE speakers (and listeners) use pitch paralinguistically to distinguish between sentence modes (DECs and DCQs), whereas speakers of othefr English varieties use linguistic structures. In other words, it strongly implies that, functionally, nIE varieties lack a chunk of the intonational phonology available to other varieties of English. Such an apparent void in nIE intonational grammar has the appearance of a major typological distinction. While this could of course be the case, seems more prudent to assume that nIE varieties employ intonation for categorical contrasts much in the same way as other varieties do. If we begin from the premise that nIE varieties of English employ intonational phonological contrasts much in the same way as other varieties, then we might wonder if the theory underpinning the available inventory of labels has not successfully managed to capture all the linguistic / categorical contrasts available.

AM approaches to intonation originating largely from Pierrehumbert (1980) employ a single tonal tier to explain the phonological use of pitch; however, other AM descriptions of phonological pitch have argued for the use of a register tier. For example, it has been argued that features of Hausa intonation cannot simply be accounted for with reference to a single tonal tier alone, but rather that there appears also to be a register tier which controls the raising (or non-raising) of the pitch contour (Leben, Inkelas and Cobler, 1989; Inkelas and Leben, 1990). Haan (2002) notes the possibility of a register tier explanation in her discussion of pitch register shifts in Dutch question intonation, but she ultimately rejects the register tier hypothesis in favour of the paralinguistic trace view.

Gussenhoven explicitly argues against the existence of a phonological register tier. He claims that changes in pitch register are a matter of phonetic implementation “subject to purposeful speaker control” and that therefore “variation in pitch range and register are not represented in the phonology” (Gussenhoven, 2004, p. 116). This view appears to work from an assumption that the phonology is *not* subject to “purposeful speaker control”. However, if we consider any aspect of a language’s grammar, the speaker may have little control over the grammaticality of a particular string of morphemes yet does have control over the selection of grammatical structures, much in the way a diner at an à la carte restaurant is largely free to choose any item from the menu even if they cannot choose which items appear on the menu in the first place. Furthermore, ‘purposeful’ includes not only a sense of intent but also of function; that is, all speech serves a communicative function, and that function is largely determined by speaker intent. If a set of features occur consistently and repeatedly in association with a specific linguistic function, it therefore seems reasonable to assume that those features form part of the grammar, i.e., they are linguistic rather than paralinguistic.

If we allow for the possible existence of a register tier in the phonology, this might well explain how nIE maintains a phonological distinction between DECs and DCQs. This is in contrast to the view that nIE must employ gradient features and context alone to aid discrimination between the two sentence modes. Therefore, one key object of this research is to assess the validity of a register tier hypothesis.

## Conclusions

This chapter has provided information on Derry City, nIE, and previous research into intonation in nIE varieties. It then considered how theoretical concerns regarding AM raised in chapter two might influence the current study. One concern is formal and asks if there might be evidence for the special status of H tones in the realisation of PN tones in DCE. This second is of a functional nature and asks if an analysis of sentence modes in DCE might provide evidence for a phonological register tier, which could explain changes in pitch scaling across sentence modes which might otherwise be treated as purely paralinguistic. Such a consideration is important, since—following on from arguments made in chapter 2—one must be prepared to make changes to the model in order to best explain the data rather than cordoning off certain sets of data which don’t neatly fit the phonological model.

The following short chapter will outline the research questions generates by the issues raised both in this chapter and the previous chapter.

# Prospectus

The primary purpose of this short chapter is to outline the research questions (RQs) for this project. It will identify how, in answering them, the project aims to contribute to our body of knowledge of northern Irish English, how it hopes to provide insights which may help reconcile some areas of disagreement in AM theory, and how it will provide a framework for future AM intonation research. However, a research project of this size is a long-term effort, and over time, one’s understanding of the topic develops, especially after having lived with the data for a long time. Therefore, the secondary aim of this chapter is to outline how my thinking developed, and how this led to a change in approach to the analysis. The initial approach is described as the phonology-first approach, while the second is the phonetics-first approach, and each will be discussed briefly at the end of the chapter. However, a broader discussion of each will take place in Chapter 9 and Chapter 10.

## Research Questions

The research questions are divided into two categories: descriptive and theoretical. The research was originally motivated by the descriptive aims, while the first two theoretical concerns developed out of a consideration of the questions raised about the AM approach in the light of the most striking feature of nIE in general, the prevalence of L\*H pitch accents, as outlined in Chapter 3, Section 3.4. A third theoretical concern, however, arose out of the original analysis of the data, called a Phonology-First approach. This is the approach adopted in Chapters 6 and 7. Reflection and consideration of the assumptions and strategies employed in this approach led to the very different AM analytical approach. This approach is described as a Phonetics-first approach, and it is presented in Chapters 8 and 9.

### Descriptive Concerns: form and function

In the description of DCE for this project, the key formal concern is the effect of metrical structure and lexical boundaries in the choice of and implementation of pitch accents. It is a formal rather than a functional concern in that it assesses changes to intonation patterns in the absence of any changes in communicative function. Metrical structure here refers specifically to the number of syllables in the foot and the number of syllables in anacrusis. (For the sake of convenience, anacrusis is taken to include unstressed syllables before the nuclear pitch accent as well as those before the first stressed syllable in the foot.) The analysis of lexical boundaries is limited to PN pitch accents, in part because effects on PN tonal targets have been observed previously in the literature (e.g., Silverman and Pierrehumbert, 1990), and in part because potential lexical boundary effects on PN accents were observed during the analysis of pilot data during the corpus development phase (see Chapter 5).

The functional role of intonation is assessed in relation to question modes and focus. The description of the phonology and phonetics of intonation across sentence modes aims to establish a foundation from which delve into the theoretical issues outlined in Chapter 3 (section 3.4.2) and below in XXX. The analysis of focus lends itself to the same theoretical concerns as the analysis of question modes, and as with sentence modes, a baseline descriptive analysis is necessary before dealing with the theoretical concerns.

Following standard AM procedures, all analyses include both a description of phonological inventories and their phonetic implementation, essentially continuous data. The descriptive concerns of this project are outlines in the three research questions (RQS) below:

1. What are the phonological and phonetic characteristics of pitch accents in DCE in unmarked speech under variation in metrical (anacrusis and foot size) and lexical structure?
2. What are the phonological and phonetic characteristics of nuclear pitch contours in DCE across sentence modes?

As outlined in Chapter 3 (section 3.3), some of these descriptive concerns have been addressed for other varieties of nIE, most notably in Belfast English and Donegal English, but until now there has been no AM intonation analysis of DCE. As noted in 3.1, Derry City is physically and socio-culturally close to Donegal, but is an urban area in Northern Ireland, and so has many features in common with Belfast (political and social infrastructure, education system, and so on). In answering RQs 1-2, this research adds to the body of knowledge of nIE but will also indicate the extent to which DCE seems to pattern with Belfast English and the extent to which appears to pattern with Donegal English.

### Theoretical concerns: H tones and register tiers

Two theoretical concerns were outlined at the end of Chapter 3 (section 3.4). The first relates to the status of the H tone in intonation. L\*H dominates as the unmarked tone in nIE, but this is relatively unusual in English and in most other languages. It was also noted that H\* has been shown to dominate in PN positions, at least in Belfast English. It was postulated that the reason that H\* dominates in PN position is that there is less communicative pressure to realise PN pitch accents with the same degree of consistency as nuclear pitch accents, since it is the nuclear pitch accent and not the PN which carries most of the communicative burden. Therefore, it is possible that H\* in PN position is in fact a reduced form of L\*H. The analysis of metrical and lexical effects on the intonation of PN pitch accents, therefore, may shed light on this, because one would expect occurrences of L\*H to increase as most segmental material becomes available, and occurrences of H\* to increase when there is less segmental material available. If there is a tendency for H to be retains and L to be deleted, this would demonstrate that, in DCE at least, the maintenance of the H tone is more important that of the L tone, and thus, even though L\*H dominates in unmarked nuclear pitch accents, H tones still retain a privileged status. Thus, the first research question with a more theoretical bent is as follows:

1. Is there evidence in the realisation of PN pitch accents in DCE for the special status of H tones?

The second theoretical concern also derives from the dominance of L\*H in nIE. Given that L\*H has been found to dominate across all sentence mode functions, including declarative statements (DECs) and declarative questions (DCQs), one might—following the approach championed by Gussenhoven (see Chapter 2, sections 2.3.4 and 2.3.5)—assume that the intonational difference between DECs and DCQs is purely paralinguistic. However, other varieties of English exhibit a phonological contrast between the two, since DECs are likely to be realised with an H\*L % but DCQs with an L\*H H% nuclear pattern. In Chapter 3 (section 3.4.2), it was suggested that the difference in nIE may too be phonological but that it is controlled by changes triggered in a register tier. However, the argument for a register tier is not commonly invoked; however, Gussenhoven claims that implementational rules and the tonal tier are sufficient to explain register shifts in the pitch contour. It is hoped, therefore, that the descriptive analysis of the phonology and phonetics of intonation in sentence modes present evidence for (or against) the presence of a register tier as the best explanation for functional changes in the phonology of DCE. Thus, the second theoretically oriented RQ is as follows:

1. Does a register tier provide a plausible phonological explanation for variation across sentence modes in DCE?

### Phonology-first, Phonetics-first, and the final research question

RQs 1, 2, 3, and 4 are answered to a great extent by the Phonology-First approach. This portion of the research is viewed, retrospectively, as Phonology-First since it takes the PA inventories of nIE as described in the IViE project and other AM analyses of nIE as a starting point (see Chapter 3, section 3.3). It also takes the view that an *f*0 peak is the most likely realisation of an underlying H tone and an *f*0 minimum as the mostly likely realisation of an L tone. Finally, it assumes that phonological tones are realised as tonal targets with a single important landmark (they are after all, *targets*). While I believe that the analytical approach and the findings presented in Part II are valid, on reflection, some of these views seemed to reflect a naïve view of the relationship between underlying tones and their implementation in the pitch contour. A more detailed critique of the research presented Chapter 8.

Chapters 8 and 9 offer an alternative approach to the analysis of the data, the Phonetics Approach. Chapter 8 provides the rationale for this approach but begins with two fundamental changes to the analysis. Firstly, it takes a view of tonal targets which is more akin to Gussenhoven’s view that tonal targets can be left and right aligned; that is, it assumes that two tonal landmarks can be associated a single phonological tone (see Chapter 2, section 2.3.4). It does not, however, completely follow Gussenhoven’s view, and to some extent aims to reconcile differences between the ToBI-like and IViE-like approaches to intonational analysis, specifically on the issue of phrase accents and onramp/offramp approaches (see Chapter 2, section 2.5). Secondly, it replaces the *f*0 minima and maxima approach to tonal targets with a turning points approach (see Chapter 2, section 2.3.7). In fact, this essentially follows from this first change since tonal landmarks are more generalizable as *f*0 turning points rather than maxima or minima. Together, these two changes to the analytical approach are described as the Secondary Tone Hypothesis (STH). In this formulation, the secondary tone represents an option tonal landmark associated with another tone in the underlying phonology. It is a hypothesis, since it is a proposed new approach which is tested through the analyses described in Chapter 9.

The STH analysis is fundamentally phonetic-first since it begins by identifying the timing and *f*0 scaling of a minimal number of turning points required to adequately reproduce the pitch contour. This bottom-up approach then works to identify the most likely underlying phonology, inferring the most likely identity (if any) of the turning points as tonal targets in terms of boundary tones, starred and trailing tones, and secondary tones. Thus, the final theoretically motivated RQ is:

1. Does the Secondary Tone Hypothesis provide a more stable analysis of the phonology and phonetics of Intonation in DCE?

In order to answer this question, some of the data used in chapters 6 and 7 is revisited in Chapter 9, and as such, offers insights for RQs 1, 2, 3, and 4. The thinking which gave rise to the STH developed out the analysis of metrical and lexical effect and the analysis of sentence modes. However, some focus data is also analysed. While an analysis of focus is not one of the main objects of this dissertation, the analysis of focus from an STH perspective will help identify some of the benefits of this phonetics-first approach.

The tools developed for the STH analysis are based on clear phonetic and phonological principles (see Chapter 8), can facilitate the analysis of several intonational phenomena of other corpora, and are publicly available for download and use (Rodgers, 2020). Given that these tools are publicly available as are principles behind them, it is hoped that the STH approach can facilitate the analysis of intonation in future research projects.

# Methodologies

This chapter summaries the methodologies adopted for corpus development, testing, and recordings, as well as methodologies for statistical analysis which are used in every subsequent chapter.

## Corpus Development

This study is largely concerned with differences in the phonetic implementation of pitch accents under different metrical conditions and for different communicative functions. Of great interest is the fact that previous studies of nIE have shown that L\*H dominates in casual or colloquial speech regardless of communicative function. Therefore, it was decided to elicit colloquial speech patterns as much as possible and to attempt to minimize the style-shifts to careful or formal speech patterns. At the same time, it was also important to ensure that a representative set of tokens for each variable could be collected without placing too large a burden on volunteers. Furthermore, it was necessary to ensure that the data collected would be amenable to phonetic analysis. This chapter outlines how the development of the corpus, describing the construction and development of tasks, the piloting and iteration process, the cohort of volunteers, the recording process, and finally the annotation and data processing procedures used to generate an analysable database.

### Task design

Four tasks were chosen to facilitate the collection of analysable data: a read-speech task, an interactive goal-oriented task, a story telling task, and a contour imitation task.

Read speech allows key variables could be controlled systematically, which ensures that there is sufficient coverage of each variable and that each token offers maximally informative. However, read speech is also more likely to lead to style-shifting to a more formal style. To mitigate against this, several strategies were employed. Firstly, the target phrases were embedded in short dialogues, which in turn were placed in a plausible everyday context, such as talking about a holiday or talking about family members. In this way, the presentation of the stimuli, the semantic content, the pragmatic context, and the subject domain of all target phrases were controlled so as to maximise the chance of approximating casual speech styles. Secondly, it was decided to record volunteers in pairs and to ensure that they already knew each other. Each volunteer was asked to comment whenever they noticed their partner switching to a more *telephone* style of speech and to encourage them to speak in their *everyday voice*. This setup was maintained throughout the recording process.

For the *spot-the-differences* activity, each participant was presented with a picture, with each picture differing in a number of details. The partners were asked to identify as many differences as possible within a time limit without looking at the other’s picture. This kind of activity encourages different kinds of speech acts, including description, clarifying statements, checking questions, binary questions, and wh-questions. The pictures and instructions can be found in Appendix X.

The story-telling task required volunteers to tell a story about a local character or childhood memory from growing up in the city. The instructions for this task were presented before the recording session began but it was the last activity recorded. This gave speakers time to think about their story. Local themes were selected to encourage the activation of a “Derry City” schema in the hope that speakers would more likely maintain a DCE speech pattern.

The final task, contour imitation, involved the production of semantically empty pitch contours (fall, rise-fall, and so on), with an aim to establishing baseline *f*0 ~ voice quality interactions. Speakers were presented with a visual representation of a pitch contour and a verbal description. They were then asked to produce the contour described visually and verbally using the nonsense utterance *DAdada DAdada*. This activity is reproduced in Appendix X.

### Piloting and Refinement

The tasks were trialled in two phases. The first phase used colleagues in the Phonetics and Speech Laboratory and on family members as guinea pigs. The main aim of this was to check the clarity of the instructions and task presentation, the simplicity of the tasks, and the burden each task placed on the participants. The second phase involved trialling the activities on DCE speakers. This phase had two main aims. Firstly, it checked that the stimuli elicited the target language effectively. Secondly, it permitted provisional analysis of the data to identify any weaknesses or oversights, specifically regarding the DCE speaking cohort. During each phase, stimuli were modified based on participants’ ability to complete the tasks effectively. Modifications included changes to the manner of instruction, presentation of the text, and slight alterations to the read-speech and spot-the-differences stimuli.

The trial process led to several modifications to the stimuli and their intended function. The *spot-the-differences* and the story-telling tasks only produced a small amount of analysable data; however, they were retained for comparative purposes. That is, they could be compared with the read speech data to confirm that speakers maintained a casual speech style during the read speech task and that—impressionistically at least—the intonation patterns elicited during the read-speech activity reflected the patterns produced during the less controlled activities. The contour-imitation task proved too challenging for all the volunteers, so was not used for analysis. However, the volunteers found the task enjoyable, so it was retained to punctuate the otherwise potentially monotonous read speech task with something more entertaining. During the provisional analysis of the data, it was noted that some DCE speakers appeared to align PN peaks with word boundaries while others aligned them with the right edge of the foot. This led to the addition of a few more stimuli to analyse potential lexical boundary effects tonal alignment.

### Read speech stimuli

In the end, only the read-speech stimuli were chosen for detailed analysis. There are five sets of stimuli, as shown in Table 5.1, which outlines the purpose of each. A detailed description of each set is presented in the materials and methods section of the chapters where the specific sub-corpus is analysed. The stimuli themselves can be found in appendix XXX.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table .. Stimulus sets for the read-sentence corpus   |  |  |  | | --- | --- | --- | | Set | Purpose | Associated Chapter | | A | Analysis of alignment in pre-nuclear (PN) and nuclear pitch (NUC) accents. | Chapter 6, Chapter 11 | | H | Subset of A to test alignment effects of word boundaries in PNs | Chapter 6, Chapter 11 | | M | Analysis of pitch accents in sentence modes. | Chapter 7, Chapter 11 | | F | Analysis of broad focus and narrow focus | Chapter 12 | |  |  |  | |

### Participants

Speakers were recruited initially by talking to staff and volunteers at the Verbal Arts Centre on Bishop Street Within in the city centre, and then through word of mouth. In total, there were 24 participants. However, only eleven participants could be used in the analysis. There were several reasons for this. Some chose to take on the interlocutor role only and did not produce the target phrases, while a two did not complete the recording session. Two participants turned out not to be from Derry City, two had speech impairments which made completion of the task difficult, three persistently style-shifted during the recording, and one was not used due to persistent nasality in her speech.

Of the eleven remaining participants, there were 6 females and 5 males, with a mean age of 40 (standard deviation 9.9). All had at least one parent from the city, and—except for M10, who had lived in London until age seven—all speakers had been also born in the city. All participants except M04 and F17 had spent three to five years living outside Derry City as adults, either for work or education; however, all participants (including M10) had spent at least 85% of their lives in the city and had been living and working there continuously for at least the last 12 years. All self-identified as having a distinctive Derry City accent, and their speaking partners agreed. All participants described themselves as middle class or as middle class from a working-class background. All grew up in Roman Catholic areas and communities, although F17 came from a mixed Roman Catholic and Protestant family and felt she had grown up with “the best of both worlds.” Given the difficulty in recruiting volunteers and acquiring suitable data for analysis, the final cohort of speakers used in the corpus is not as homogenous as one would like. However, based on their own judgment, that of their speaking partners, and my own, all had distinctive Derry City accents. A summary of the details for each participant can be found in Table 5.2, while the map in Figure 5.1 shows the rough location of each local area along with the location of the Verbal Arts Centre.

Table .. Biodata for participants used in analyses. (See Figure 5.1 for approximate location for each area.)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Code | Age | Sex | Highest level of education | local area | Recording date | Pairing | Relationship to partner |
| F05 | 37 | F | 3rd level | Waterside | 08 Nov 2017 | F04-F05 | colleague |
| F06 | 35 | F | post-grad | Rosemount | 08 Nov 2017 | F06-F07 | ex-colleague |
| M04 | 60 | M | 2nd level | Bogside | 19 Dec 2017 | M04-M05 | colleague |
| M05 | 45 | M | 2nd level | Bogside | 19 Dec 2017 | M04-M05 | colleague |
| F12 | 57 | F | 3rd level | Creggan | 24 Jan 2018 | F12-F13 | colleague |
| M08 | 54 | M | 3rd level | Creggan | 02 Feb 2018 | M08-M08i | friend |
| M09 | 45 | M | 2nd level | Brandywell | 08 Feb 2018 | M09-F14 | colleague |
| F15 | 44 | F | 2nd level | Waterside | 16 Feb 2018 | F15-F16 | sister |
| F16 | 35 | F | 2nd level | Waterside | 16 Feb 2018 | F15-F16 | sister |
| M10 | 54 | M | 2nd level | Strand Road | 23 Feb 2018 | M10-F17 | friend |
| F17 | 62 | F | 3rd level | Strand Road | 23 Feb 2018 | M10-F17 | friend |
|  |  |  |  |  |  |  |  |

|  |
| --- |
| Bogside (2 M)  Brandywell (1 M)  Creggan (2 F, 1 M)  Rosemount (1 F)  Strand Road (1 M, 1 F)  Waterside (3 F)  Verbal Arts Centre  Bogside  Brandywell  Creggan  Rosemount  Strand Rd  Waterside |
| Figure .. Rough demarcation of local area for speakers in the final corpus. Yellow star marks the location of the Verbal Arts Centre. Derry City map from Northern Ireland Statistics and Research Agency (2016). |

### Recording

The staff at the Verbal Arts Centre generously offered free use of their recording studio. All recordings were carried out there. Participants were recorded with a Röde NT1000 microphone using the Cubase software. Recordings were downsampled and rendered as *.wav* files with a 44.1 kHz sample rate.

Before the recordings, roughly 10 minutes was spent chatting with participants, explaining the recording procedure, and collecting biodata. They were encouraged to think about the story they might tell at the end of the session. Each participant also read and signed the consent form and was given a copy for their own reference (see Appendix XXXX). Participants were provided with water or tea during the recording and were encouraged to take breaks whenever they wanted. Three breaks were also incorporated into the reading tasks, as this was the most tiring portion of the recording session. Sound levels were checked before recording, and speakers were encouraged to sit ‘an arm’s length’ away from the microphone. For each task, participants were asked to avoid talking over each other, and to avoid moving around while speaking. The researcher gave instructions in the live room but sat in the control room during the recording and was able to communicate with the participants via headphones. Participants were reminded that they could end the recording session whenever they wished.

The tasks were always presented in the following order:

1. Read speech and pitch contour tasks
2. Interactive task
3. Story telling task

For the read speech task, dialogue prompts were randomised in advance. The prompts were presented to the participants in a PDF document on an iPad, which they operated themselves. Each prompt appeared on a separate page which showed the general context of the dialogue, a question prompt, followed by the target phrase, as shown in Figure 5.2 below. The task was divided into four sections, with a written prompt encouraging participants to take a break at the end of each section. The pitch movement task was placed at the start of each section. Ten repetitions were recorded for each prompt, and participants swapping roles after the first 5 repetitions. To avoid the risk of one of the speakers always setting a baseline for pronunciation, participants took turns going first for each new prompt. Participants were asked to count to five in between repetitions. They were also encouraged to comment if they felt the other was speaking in his or her ‘telephone voice’. Whenever there were any noticeable problems with a repetition—including misreading, speaker overlap, and excess background noise—one or two extra repetitions were recorded. Problematic utterances were discounted; however, very occasionally, this meant that there were six reasonable repetitions for one target phrase.

|  |
| --- |
| Graphical user interface, text, application  Description automatically generated |
| Figure .. Example of read-speech corpus prompt for target phrase A3422. |

## Annotation and Data Processing

All data was annotated in Praat (Boersma and Weenink, 2022) using several scripts to help speed up the process (see github.com/AERodgers/PhD-Scripts and Appendix XXX). The annotation tiers used for analysis are shown in Table 5.3 and an example of the annotation window in Figure 5.3.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table .. Annotation tiers used to facilitate phonological and phonetic analysis of IPs.   |  |  |  |  | | --- | --- | --- | --- | |  | Tier | Type | Function | | 1 | Orthographic | interval | Mark orthography and word boundaries | | 2 | Syllable | Interval | Mark syllables and syllable boundaries. | | 3 | Rhythmic | Point | Mark IP boundaries, metrical stress, and boundary tones | | 4 | Phonological | Interval | Annotate pitch accents - absense of expected PA marked as (\*) | | 5 | Vowel | Interval | Segment stressed vowels and vowels in syllables with tonal targets | | 6 | Tone | Point | H and L tonal target, onset (S) and offset (E) of voicing in phrase. | | 7 | Comments | Interval | Additional comments and observations | |

Annotation was done in two stages: a segmentation stage and a pitch annotation stage. In the segmentation stage, the **syllable** tier was annotated manually based on a visual inspection of the spectrogram. In cases of ambisyllabicity (Harris, 1994; Hayes, 2009), ambisyllabic segments were annotated as part of the rhythmically stressed syllable. **Orthographic** and **rhythmic** tiers were generated automatically from the syllable tier using the script create\_more\_tiers, which also generated blank **phonological** and **vowel** tiers.

|  |
| --- |
| Graphical user interface, timeline  Description automatically generated with medium confidence  Figure .. Example of annotation window in Praat with annotation tiers. |

In the pitch annotation stage, the phonological tier was annotated manually using the IViE labelling system as a foundation, with a focus on auditory analysis aided by visual analysis of the spectrogram and pitch contour. However, during the annotation of each sub-corpus, adjustments were made to the IViE labelling. Typically, the need for such adjustments only became apparent during the labelling process itself, so there is a degree of circularity to the methodology. That is, IViE was adopted as a labelling methodology, but during annotation, it was felt that alterations to the labelling system were needed to capture the phonology of the utterances more appropriately. For this reason, adjustments to the phonological analysis belong partially to methodology and partially to results and analysis. They are dealt with more fully in each chapter, in a section named Phonological Labelling (c.f. Sections 6.4 and 7.4).

A second trained phonetician (MOR) who specialises in intonation also judged the pitch accents, and, wherever there was disagreement, consensus was reached through discussion and repeated listening.

After the phonological analysis stage, the vowel and tone tiers were annotated. The fix\_pitch script was used, to manually correct *f*0 estimation errors such as pitch halving or pitch doubling. This script also allowed the used to remove gross perturbations in the contour caused by segmental effects such as fricatives. This approach was preferred over automated strategies, such as Xu’s *f*0 trimming algorithm (Xu, 1999), which sometimes appeared to overcorrect the contour and remove pitch points which appeared to be part of the intended intonation contour, especially when there were sharp rises or falls in *f*0. To facilitate analysis, the curve was then interpolated to replace missing *f*0 points. It was then smoothed using Praat’s Smooth function, with the bandwidth parameters set at 19 Hz. This is a minimal amount of smoothing, and essentially removed any remaining micro perturbations from the contour. Finally, the corrected pitch contour was saved for analysis. An example of a corrected contour is shown in Figure 5.4.

|  |
| --- |
| Chart, histogram  Description automatically generated |
| Figure . Example of corrected f0 contour after running correction script. The corrected contour is in blue and the original in grey. |

After pitch correction, the onset (S) and offset (E) of voicing were annotated in the **tone** tier. L and H tonal targets were also marked in the tonetier according to the local minimum and maximum *f*0 in the appropriate portion of contour. In cases where there were several potential maxima or minima, a candidate in the most vowel-like portion of the syllable was chosen. For example, in Figure 5.3 above, there are two potential L targets in the RIV- syllable, one in the approximant /ɹ/ and on in the vowel /ɪ/. In this case, /ɪ/ was preferred, since the low in /ɹ/ may have been the result of a minor segmental effect.

A Praat script called process\_texgrids [REF] was used to tabulate the data across all the text grids for each Corpus. This data is used for the analytical experiments detailed in Parts II and III.

## Visual analysis of count data

All the sub-corpora contain missing data points, so there is a risk that tabulated raw phonological data will be misleading because it will over- or under-represent specific speakers and target utterances. To avoid misrepresentation of the data, especially for visual analysis of trends, raw values have been adjusted to project a more balanced predicted distribution of counts per speaker and per stimulus. This takes into consideration variation in the number of repetitions per speaker, number of speakers per target for cases where a speaker produced no analysable utterances, and in some cases the number of tokens per gender.

The process for this comprises three stages:

1. Counting the number of phonological tokens per stimulus per speaker.
2. Converting counts to a proportion per stimulus per speaker.
3. Convert the proportion per stimulus per speaker to a proportion per stimulus per target feature (e.g., mode or foot size).

The proportions in this table are converted to counts based on a projected ideal count of five utterances per speaker per target (and an equal number of male and female speakers). From this, three tables of adjusted data counts are generated:

1. Phonological tokens by speaker (adjusted)
2. Phonological tokens by target feature (adjusted)
3. Phonological tokens by target feature and gender (adjusted)

Nearly all visual representations of the count data use these adjusted counts, although it is always specified if the count is based on raw or adjusted values.

The code for these procedures embedded in the R code for each chapter (found in REFS). Do note, however, that this process is only used to facilitate data visualisation and is not used for any of the inferential statistical analyses.

## Inferential statistical methods: Linear Mix Effects models

All inferential statistical analysis was conducted in R (R Core Team, 2022), using packages outlined below alongside some purpose written functions. Microsoft Office Excel (REF) was also used for some data visualisation and the calculation of means and standard deviations in some cases. All data, R Code, and Excel workbooks can be found in APPENDIX REF and GITHUB REF.

Mixed Effects models have been used throughout the dissertation for inferential statistical analysis. These have several advantages over non-mixed effects models, but there are two which are especially important here. First, they are good at coping with multiple repetitions of a target phrase, including cases where there are missing data points, as is the case here. Secondly—and closely connected to the first—they can, unlike ANOVAs for example, cope with multiple random factors. That is, mixed effects models can compensate for variance caused by multiple random effects, and thus compensate for the amount of error or noise they add to the estimates.

A random factor represents a factor which is known to influence the result but is not of interest in the analysis. Most precisely, it represents a factor within the data set, the levels of which are taken sample of the whole population. For example, Chapter Six analyses—among other things—the effects of anacrusis and foot size on the temporal alignment of tonal targets. Foot size and anacrusis are fixed effects with a limited number of levels, which are controlled experimentally. However, the data are taken from eleven speakers, representing a random sampling of the population, and each has their own speech idiosyncrasies. Thus, speaker-specific variation adds noise to the model. By treating speaker as a random effect, we can reduce the effect of this noise. Sometimes, however, features which are typically random effects can be of interest and are thus included as fixed factors. Conversely, factors which are—by strict definition—fixed effects may be of little interest and can be included as random factors instead.

Mixed Effects models are not without their flaws, however. For the purposes here, two issues must be highlighted. The first is that they are prone to convergence errors, especially when the model is complex. A convergence error means that the algorithm for resolving the model has not achieved a stable solution for the model and that the estimates may not be reliable. The second problem with missed effects models is singularity. This refers to cases where the variance-covariance matrix is equal to zero or one (i.e., perfect correlation), suggesting that the model has been over-fitted.

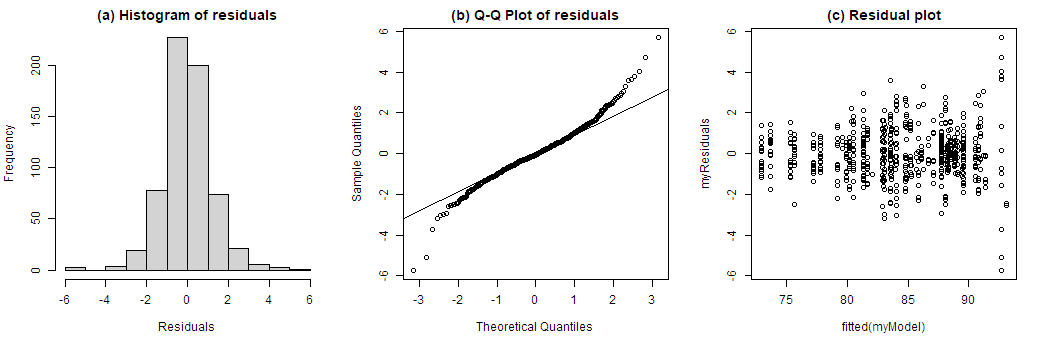
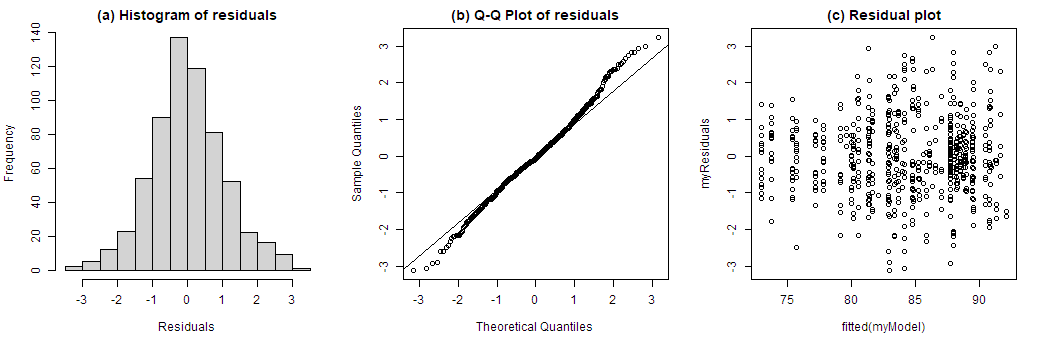
Two types of Mixed Effects model are used here, Linear Mixed Effects (LMEs) modelling for continuous phonetic variables and Bayesian Generalised Linear Mixed Effects models (BGLMs) for categorical phonological data. Originally, Bayesian models were not to be used at all; however, they provided the best solution to another problem which came up during the analysis of categorical data (see 5.4.2 below) that does not occur in the continuous data.

### LMEs and the analysis of continuous parameters

Linear mixed effects (LME) models were chosen to evaluate the continuous phonetic parameters and used the lme4 package (Bates *et al.*, 2015). For each analysis, an ideal maximal model was identified which was believed to best describe the relationship between the independent variable(s), random factors, and the dependent (outcome) variable. The independent factors under analysis were treated as fixed factors, while other factors were treated as random factors. However, these models were sometimes prone to convergence and singularity problems.

To mitigate against convergence—and following advice on convergence issues in the lme4 package—the package’s allFit() function was employed, which runs the model through all available optimizers. Optimizers are algorithms which try to optimize the solution to the mixed model formula, and allFit() outputs the negative log-likelihood of the differences between all optimizers. If the model outputs are all very similar (i.e., very low log likelihood), one can assume that the convergence errors were spurious and that the model is acceptable. Unfortunately, such felicitous outputs never occurred, and other approaches were needed. As a rule, one of two optimisers was employed, either optimx (Nash and Varadhan, 2011; Nash, 2014) or bobyqa (Powell, 2009). optimx was the default choice, with bobyqa being tested when optimx failed. In cases where singularity issues arose, the isSingular() function from lme4 was used, with a tolerance set at 1e-5, i.e., with a threshold for singularity tolerance closer to zero than the default. If the model was not identified as singular at this lower threshold, the model was accepted.

If there were still convergence errors or singularity issues, the model was simplified, and this was generally necessary. Each model was simplified in two stated. First, the step() function from the lmerTest package (Kuznetsova, 2017) was employed to help reduce model complexity. This works by automatically performing a series of tests using stepwise backward reduction of fixed and random effects from the original model. This helps identify simplify the model by removing non-significant effects from the original formula. However, sometimes, non-significant results are of interest (i.e., sometimes we want to check that a factor has very little effect on the outcome), so they were not always removed from the model. In cases where the backward reduction still led to convergence errors, models which removed random slopes one-by-one were tested manually to find most complete model which did not generate convergence or singularity errors. Typically, this ended up being to a random-intercepts-only optimal working model.



b.

a.

Figure . Residuals of the model l\_f0 ~ mode + fin\_phon + gender + (1 + mode | speaker). Panel a. shows the models without trimming, where outliers can clearly be seen to create a very spikey histogram, and the Q-Q plot shows large deviation from the expected correlation between sample and theoretical quartiles (i.e., the individual dots stray from the diagonal line at the edges). Panel b. shows the model residual plots after removing observations with residuals greater than 3 standard deviations in the original model, with a loss of 9 out of 632 total observations in the m-corpus.

Once the final model was established, the residuals of the model were examined visually using a histogram of residuals, Q-Q plots of residuals, and a residual plot (see Figure 7.1 above). In cases where the histogram was skewed or leptokurtic (i.e., spikey), the Q-Q plot indicated extreme outliers, or the residual plot indicated heteroskedasticity (i.e., the distribution changes along the plot), the dataset was trimmed of outliers. This was done following Baayen (2008, pp. 279–282), by removing observations with residuals beyond a standard deviation threshold. The standard deviation threshold was set for each individual model to minimise the number of observations trimmed while producing a more reasonable distribution of residuals. An example of visual residual analysis is shown in Figure 7.1, with panel a. indicating the model residuals before trimming and the panel b. showing the residuals after trimming observations using a standard deviation threshold of 2.5 (though do note, in the second plot, heteroskedasticity has not been completely removed). The process of determining the most informative model for each parameter is documented in [REF TO MarkDowns], while the models themselves are discussed in the results sections of each relevant chapter.

### BGLMs and the analysis of discrete categories

As with all generalized linear models, Generalized linear mixed-effects models (GLMs) use logistic regression analysis. That is, they estimate the odds ratio of a binary outcome as a function of the predictor variable or variables. As mixed effects models, however, they incorporate both fixed effects (predictor variables) and random effects (known variables which contribute to the error in the model).

For simplicity can consistency, the originally plan was to use generalised linear mixed effects (GLMM) models from the lme4 package to evaluate categorical phonological data; however, this was not possible because of a problem called complete separation. This occurs when one or more levels of a predictor variable perfectly predicts the outcome. For example, in one analysis of mode (see 7.5.1 below), MDC in is never associated with an H% boundary, so in a binomial analysis, the odds of H% are 0:1 against, while the odds for the alternative are infinity:1, generating an infinitely small non-zero odds ratio of 0/∞. While the probability of H% is in fact zero for MDC, the model generates log odds Confidence Intervals (CIs, see 5.4.3 below) tending towards plus and minus infinity, giving the erroneous impression that the likelihood of H% cannot be predicted.

This issue was resolved by adopting a Bayesian generalised mixed effects (BGLM) model, using the package bglme (Chung and Rabe-Hesketh, 2013). This package is based on lme4, which makes the switch from one to the other ease, as both use the same syntax for the modelling. The BGLM allows the user to impose priors on the model. These are in essence expectations based on prior experience and knowledge. As we know that the confidence intervals of the models is not 0 or infinity, so we can impose zero-mean normal priors on the model (Bolker, 2018), i.e., inform the model the priors expect a model with a normal distribution around the mean. This prevents extreme odds or CI estimation, and thus mitigate against the problem of infinitely wide CIs which occurs in GLMMs in cases of complete separation.

As with the LME models, a maximal model with random slopes and intercepts was tested first, but this was reduced to a random-intercepts-only model if it generated convergence or singularity issues. The final model was always the most informative model which did not generate errors or warnings.

### p values and confidence intervals

A GLMM (or any kind of logistic regression) estimates the likelihood of one outcome over another. If log odds are zero, each outcome is equally likely, and if the 95% confidence intervals (CIs) for the estimated mean log odds cross the zero—i.e., from positive to negative or from negative to positive—the model fails the test of significance at a level of p < 0.05. However, this does not mean that the log odds are unreliable. It means rather, that one cannot claim that one outcome is more likely than the other. In an analysis of a categorical independent variable (such as syllable count or sentence mode) estimated log odds of 0 in the intercept (β0) indicates a 50% probability of either target outcome for the intercept category (e.g., a one-syllable foot, or a yes-no question). Such a result can be very informative, say for example if the predicted probability of an outcome is 50% under one condition but much lower under others. 95% CIs, on the other hand, indicate the values between which we are 95% confidence that true log odds ratio lies. Therefore, we should be more concerned about large CIs for intercepts than simply their p values.

This also holds for the estimated means of the intercept of a dependent variable in LME analysis. The p value here indicates if the estimated mean of the intercept is significantly different from zero. However, an intercept of zero may be of interest, such as when the estimated mean slope of the linear regression of an *f*0 contour is 0 ST/sec. Again, the size of the confidence intervals is more important than the p value, since very large confidence intervals indicate that the true mean lies across a large range of values.

The p values of the slopes (β1) of the model, however, are more important, since they indicate if we can confidently reject the hypothesis that there is no difference—for the analyses conducted here—between two levels of an independent variable. For example, we might want to decide if there is a statistically significant difference between the mean *f*0 of a declarative statement (MDC) and the mean *f*0 of a declarative question (MDQ). If MDC is the intercept, the estimated slope (β1) of MDQ is the estimated difference between the two. The closer to zero, the less difference between the two. Similar to the log odds ratios, if the 95% CIs cross zero, we cannot confidently reject the possibility that there no significant difference between the two at a significance level of p < 0.05.

This goes for log odds ratios of slopes (β1) in GLMMs as well. If we want to consider the likelihood that a speaker uses and L% based on their gender, we look at the slope, typically with female as the intercept and male as the slope. If there is no gender difference, the estimated slope should be close, with CIs crossing zero. Alternatively, the estimate may be far from zero, but the CIs may be very large. Both these outcomes will be non-significant at a level of p < 0.05. If the CIs are small and the estimate is close to zero, however, we be more confidence in accepting the null hypothesis that there is no gender difference. (Despite superficial appearances, this is quite different from cases where non-significant p value indicates that we cannot *reject* the null hypothesis.)

Finally, there can be some confusion in the use of the terms *slope* and *mean*, which needs to be resolved. In cases in the results section where the targets parameter is the slope of a linear regression of an *f*0 contour. As the statistical methods used here all also involve different kinds of regression (linear or logistic), the results of the statistical methods also have slopes. Therefore, in cases where this might be confusion, the statistical formalism β1 is added to indicate a reference the slope of the regression model, as opposed to the slope of the *f*0 contour. Similarly, in the regression analysis conducted here, the intercept (β0) always represents the estimated mean of a specific level of an independent factor. Therefore, when it might be unclear whether the term *mean* refers to a dependent parameter (*f*0 mean) or the intercept of the regression model, the intercept of the regression model is indicated by the addition of the β0 formalism.

### Shared procedures

Once the most informative working model was established for either the LME or BGLMM analysis and outliers has been removed, the purpose-written functions analyseModel() and getModelFixedFX() [REF] were used to perform a set of shared processes.

analyseModel() produces tidy summaries of the models and a series of visuals to help analyse the results. In the case of BGLMs, this includes predicted probabilities of for each level of the independent factors. It also calculates marginal and conditional r-squared values for each model using the r2\_nakagawa() function from the performance package (Nakagawa and Schielzeth, 2013; Nakagawa, Johnson and Schielzeth, 2017; Lüdecke *et al.*, 2021). Marginal r-squared (r2m) indicates the amount of variance explained by the fixed effects only, while conditional r-squared (r2c) reflects the amount of variance explained by the whole model, i.e., it includes both random and fixed effects. Thus, for example, a marginal r2 of 0.05 indicates that 5% of the variance in the target variable can be explained by fixed effects, while a conditional r2 of 0.95 indicates that 95% of the variance is explained by the whole model.

getModelFixedFX()calculates the intercepts (β0) for each level of the independent factor(s) along with slopes (β1) for pairwise comparisons between each level of the predictor variables. In this way, for example, it is possible to generate the estimated mean values (or odds ratios) of a specific target variablefor each level of a predictor variable as well as the estimated slopes between each level of the target variable. For factors with only two levels, e.g., gender, only slopes are calculated.

Finally, after all analyses are complete for a specific analysis, post-hoc adjustment of p values for each analysis set is conducted to account for multiple testing. This is because, as the more and more statistical tests are conducted, the likely of getting a spurious false positive result—or Type I error—increases. Therefore, a mechanism to compensate for this possibility is required. The Benjamini and Hochberg (BH) method was used for this, since it minimize the number of Type I errors while also mitigating against Type II errors (false negatives) by controlling the false discovery rate (FDR) (Benjamini and Hochberg, 1995). This was implemented via a purpose-written function adjustP\_posthoc() which pools together results from connected analyses and adjusts the p values using the base R function p.adjust(). For example, it generates a list of every p value generated in the pairwise analysis of categorical phonological features of mode and adjusts them accordingly. Both the original and adjusted p value are included in the full results of the statistical tests (see appendices and online). However, only the adjusted values are reported in the body of the dissertation.

# Analysis of Form: Metrical and Lexical Effects

This chapter focuses on the effects of metrical and lexical structure on the phonology and phonetic implementation of the tonal tier in declarative statements in DCE. In other words, it aims to answer the question of form raised in RQ1 [REF]. It is important to answer the question of form first because it will establish a baseline before moving on to questions of function. That is, it will help identify the extent to which metrical structure and word boundaries influence the pitch accent inventory and how each affects the phonetic parameters associated with the realisation of pitch accents. Armed with this knowledge, it will be easier to isolate phonological and phonetic components associated with function and to avoid misinterpreting formal effects as functional ones. As with all chapters in parts II and III, this chapter contributes to answering RQ4, “How well does the phonological description of intonation in AM capture the salient difference in linguistic function?”

The chapter is organised as follows. Section 6.1 focuses on the expected outcomes of the analysis in this chapter. Section 6.2 describes the methods and materials, outlining the sub-corpora used, how they were validated, and how they were analysed. presents the results of the studies, focusing first on phonology and then on phonetic implementation. It includes both a general overview and descriptions of results for individual speakers. It incorporates both descriptive and inferential statistical analyses aimed at answering RQ1. Section 6.7 considers the results within the broader regional and theoretical contexts and assesses the extent to which the results can be explained within current AM theories, thus contributing to answering RQ4.

## Hypotheses

When we consider the intonational phonology, it is clear from previous research [REF] that nuclear pitch accent inventories tend to be much more limited when compared to PN inventories. It is plausible that the larger PN inventory and narrower nuclear pitch accent inventory may be explained by differences in communicative pressure. The thinking goes as follows. The nuclear pitch accent tends to carry most of the communicative function, so the speaker has a greater need to realise it more precisely for effective communication. This pressure, however, is absent from the pre-nuclear pitch accent, so it is potentially more prone to non-linguistically motivated variation. Variation in pitch accents thus may have two competing sources:

1. Communicative intent
2. Metrical and lexical structure

If variation is motivated by communicative intent, we should expect to see little correlation between the intonational phonology, anacrusis[[7]](#footnote-8), and foot size; however, if variation is conditioned by the metrical or lexical structure, we expect to see pitch accents vary as a function of anacrusis and foot size. It is hypothesized that communicative intent will determine the inventory of nuclear pitch accents, which will be almost exclusively L\*H, but that this will not be the case for pre-nuclear pitch accents, where there is less communicative pressure. However, it is expected that as anacrusis and foot size increase, L\*H will be seen to dominate in pre-nuclear position as well. Thus, in answering **the phonological component of RQ1**, we have three hypotheses to consider:

1. L\*H is the dominant pitch accent in nuclear and pre-nuclear position
2. Variation in metrical context has no effect on the inventory of nuclear pitch accents
3. Variation in metrical context has a strong effect on the (surface) phonology of pre-nuclear pitch accents. Specifically, given hypothesis (1), increases in foot size and anacrusis will be associated with an increase in instances of L\*H.

Note in (3) the reference to *(surface) phonology*. This is because the expectations re the dominance of L\*H are expected to hold true both in nuclear and PN position. However, if variation in pre-nuclear pitch accents is found to vary as a function of metrical and lexical conditions, it follows that the recorded inventory of pitch accent labels does indeed reflect variation in a *surface* realisation of an underlying L\*H pitch accent. This would be analogous to the way that segmental allophones are variant realisations of underlying phonemes. i.e., just as /t/ is realised as [tʰ] in <time> [tʰaɪm] but as [ɾ] in <butter> [bʌɾɚ] and [t=] in <stay> [st=eɪ], we might—for example—see that /L\*H/ is realised as [H\*] when there is little metrical content, but as [L\*H] when there is sufficient material to permit its realization.

Looking at previous research on tonal alignment in nIE, as discussed in chapter XXX, one can expect metrical and lexical context to condition the timing of PN tonal targets. This assumption was reinforced by the preliminary analysis of the pilot data, where it appeared that speakers sometimes use word boundaries as an anchor point for PN H tones. Further, an early analysis of temporal alignment of PN accents from six speakers in the corpus—presented at BAAP 2019 (Rodgers, 2018, Appendix XX)—also indicated that this was the case. In fact, this partial study suggested two temporal alignment strategies for H targets. In one, the right foot boundary acts as an anchor point, while in the other the lexical boundary does. Thus, in the analysis of the full dataset, we expect H targets to be aligned later when the foot is longer and when the right word boundary occurs in a later syllable. When it comes to anacrusis effects, it is much less clear what the effects will be. For example, while a study by Nolan and Farrar (1999) found that the addition of anacrusis was associated with earlier peaks in Belfast English, a study by Kalaldeh, Dorn, and Ní Chasaide (2009) found the opposite in Donegal English, i.e., that anacrusis effects correlate positively with peak alignment in PN pitch accents.

In nuclear pitch accents, it is expected that H targets will align later as foot size increases, but it is presumed that an increase in syllables preceding the stressed syllable will have little effect on H target timing. It is expected that L targets will be relatively stable across conditions, but there is a possibility that tonal crowding may influence alignment, either from the preceding target or—assuming L\*H dominated—from the trailing tone. That is, since there is increased communicative weight associated with nuclear pitch accents, it is in the speaker’s interest to realise all tonal targets associated with the nuclear pitch accent, and—as such—the speaker may crowd the tones into a shorter time period if necessary. Furthermore, if there is a specified L% boundary, crowding effect should also be observed. (Remember that, taken together, the nuclear pitch accent and boundary tone are viewed as the nuclear contour.) This is in essence an assumption that tonal crowding will be reflected in compression effects. However, there may also be truncation effects, or a combination of the two, similar to those which Sullivan found in her analysis of Belfast English (2010).

In answering the **phonetic component of RQ1**, therefore, we have three strong hypotheses:

1. The tonal alignment in **PN accents** will be more vulnerable to **metrical and lexical** **effects** compared to tonal alignment of nuclear pitch accents,
2. There are competing strategies for H target anchoring in **PN pitch accents**, one using the right lexical boundary and the other using the right **foot** boundary as an anchor point.
3. Compression effects in the guise of tonal crowding will be observed in **nuclear pitch accents** and **nuclear pitch contours**. This will be an effect reduced **foot size** and fewer unstressed syllables in the **preceding foot**.

In addition, there are two weaker hypotheses:

1. **Anacrusis** will affect the alignment of **PN H targets**; however, the direction of this effect is not predictable in advance.
2. Truncation effects will be observed in nuclear **pitch accents**.

## Materials

The A and H sub-corpora are used for analysis in this chapter. The A-Corpus was designed primarily to investigate metrical effects on the inventory of nuclear and pre-nuclear pitch accents and on alignment of tonal targets, while the H-Corpus was designed to compare the alignment of H targets in PN pitch accents under variation in lexical boundaries.

Each target phrase in the A-Corpus contains two lexically stressed syllables, the first of which may be associated with a pre-nuclear pitch accent, and the second with the nuclear pitch accent. The A-Corpus contains 11 stimuli, and they are used to assess metrical effects across four variables:

1. Anacrusis, from zero to three syllables.
2. Size of the foot associated with the pre-nuclear pitch accent (PN foot size), from one to four syllables.
3. Unstressed syllables in the foot preceding the stressed syllable associated with the nuclear pitch accent (labelled ‘preceding’).
4. Size of the foot associated with the nuclear pitch accent (NUC foot size), from one to four syllables.

The target utterances for the A-Corpus are listed in Table 6.1. The underlined section of the target utterance indicates the portion of the utterance under analysis. Note that some utterances, such as A0221, are used to analyse several variables.

Table . A-corpus stimuli and parameter conditions. Note that “preceding” refers to the number of unstressed syllables preceding the stressed syllable in the second foot.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Code | Target utterance | Pitch Accent | Variable | Syll. count | Meter |
| A0423 | **Val**erie's is valid. | PN | Anacrusis | 0 | \*... \*. |
| A1422 | The **vall**ey's by the river. | PN | Anacrusis | 1 | .\*... \*. |
| A2422 | There's a **vall**ey with a river. | PN | Anacrusis | 2 | ..\*... \*. |
| A3422 | There was a **vall**ey with a river. | PN | Anacrusis | 3 | ...\*... \*. |
| A0131 | **Val's** valuables. | PN | Foot size | 1 | \* \*.. |
| A0221 | **Val's** is valid. | PN | Foot size | 2 | \*. \*. |
| A0321 | **Val's** is invalid. | PN | Foot size | 3 | \*.. \*. |
| A0423 | **Val**erie's is valid. | PN | Foot size | 4 | \*... \*. |
| A1111 | They know **Val**. | NUC | Preceding | 0 | .\* \* |
| A0221 | Val's is **val**id. | NUC | Preceding | 1 | \*. \*. |
| A0321 | Val's is in**val**id. | NUC | Preceding | 2 | \*.. \*. |
| A0423 | Valerie's is **val**id. | NUC | Preceding | 3 | \*... \*. |
| A1211 | He lives with **Val**. | NUC | Foot size | 1 | .\*. \* |
| A0221 | Val's is **val**id. | NUC | Foot size | 2 | \*. \*. |
| A1231 | I live with **Val**erie. | NUC | Foot size | 3 | .\*. \*.. |
| A1241 | They need e**val**uating. | NUC | Foot size | 4 | .\*. \*... |
|  |  |  |  |  |  |

The H-Corpus was designed to test if changes in the location of the word boundary within the foot affected the temporal alignment of the H target in PN pitch accents. As can be seen in Table 6.2, there are three pairs of phrases in the H corpus, with a different configuration of anacrusis and foot size parameters for each pair. Within each pair, however, the variable of interest—i.e., the lexically stressed PN word boundary—changes. The variation within and across pairs will help assess if variation in H alignment is influenced by word boundary effects more than anacrusis of foot-size effects.

Table . H-corpus stimuli and parameter conditions. The vertical bar in ‘meter and lexical boundary’ indicates the final boundary of the stressed word in the first foot.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Pairing | Code | Target Utterance | Anacrusis (syllables) | PN Foot Size | PN word-end syllable | | Meter and lexical boundary |
| 1 | A0321 | Val's is invalid. | 0 | 3 | 1 | \*|. . \* . | |
| H0322 | Lally's is valid. | 0 | 3 | 2 | \* .|. \* . | |
| 2 | H0422 | Lally's is invalid. | 0 | 4 | 2 | \* .|. . \* . | |
| A0423 | Valerie's is valid. | 0 | 4 | 3 | \* . .|. \* . | |
| 3 | H1321 | Elaine was a nanny. | 1 | 3 | 1 | . \*|. . \* . | |
| H1322 | Elaina's a nanny. | 1 | 3 | 2 | . \* .|. \* . | |

### Annotation and data extraction

The utterances were annotated as outlined in 5.6 with IViE labelling conventions used for the phonological labelling (Grabe, 2001). Most of the annotation was routine, but the phonological labelling, particularly of prenuclear pitch accents proved more difficult, and adjustments to the IViE labelling were to reflect this. These adjustments are discussed below in 6.4.

### Data extraction, pruning, and preparation

Once the annotation stage was complete and the phonology had been agreed upon, the data from the A and H corpora were tabulated using the process\_textgrids Praat script [APPENDIX XX]. However, not all annotated data were amenable to analysis. That is, each stimulus was designed to elicit two-foot phrases with a specific syllable count, but some utterances were not realised as such, typically due to elision or the addition of special stress to subject pronouns. A Praat script, corpus\_audit, was written to prune ineligible utterances automatically [APPENDIX XX]. This removed 45 utterances from the data set, so the total number of valid utterances fell from 833 to 788.

Table 6.3 summarizes the final distribution of valid utterances by stimulus and speaker and also identifies the data subsets to which each target belongs. As can be seen from the cells highlighted in pink and red, A0423 (*Valerie’s is valid)* and A1231 (*I live with Valerie*) have the greatest data loss, with 38 and 35 total utterances respectively as opposed to the target 55. With the exception of speakers F16 and F6, there was also some data loss for each speaker.

Table .. Summary of valid A and H corpus utterances by target utterance, dataset and speaker. Pink and red indicate utterances with less than five tokens each. Green indicates targets with six valid tokens.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker / dataset | A0131 | A0221 | A0321 | A0423 | A1111 | A1211 | A1231 | A1241 | A1422 | A2422 | A3422 | H0322 | H0422 | H1321 | H1322 | TOTAL |
| **pn\_ana** |  |  |  | ü |  |  |  |  | ü | ü | ü |  |  |  |  | 209 |
| **pn\_foot** | ü | ü | ü | ü |  |  |  |  |  |  |  |  |  |  |  | 203 |
| **nuc\_pre** |  | ü | ü | ü | ü |  |  |  |  |  |  |  |  |  |  | 202 |
| **nuc\_foot** |  | ü |  |  |  | ü | ü | ü |  |  |  |  |  |  |  | 196 |
| **pn\_lex** |  |  | ü | ü |  |  |  |  |  |  |  | ü | ü | ü | ü | 316 |
| F5 | 5 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 5 | 5 | **75** |
| F6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **75** |
| F12 | 5 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **73** |
| F15 | 5 | 5 | 5 | 5 | 5 | 5 | 0 | 5 | 6 | 5 | 5 | 5 | 5 | 6 | 6 | **73** |
| F16 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **75** |
| F17 | 5 | 5 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **68** |
| M4 | 5 | 5 | 5 | 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **71** |
| M5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **73** |
| M8 | 5 | 5 | 5 | 3 | 5 | 1 | 0 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **64** |
| M9 | 5 | 5 | 5 | 0 | 5 | 5 | 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | **66** |
| M10 | 5 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 6 | 5 | 5 | 5 | 5 | 5 | **75** |
| **tot.** | **55** | **55** | **55** | **38** | **54** | **51** | **35** | **55** | **56** | **56** | **55** | **55** | **56** | **56** | **56** | **788** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

The A and H corpora were split into five datasets to facilitate the analysis of metrical and lexical effects on the phonology and phonetic implementation of tonal targets. These were labelled as shown in Table 6.4. (The abbreviation on the left of the underscore referring to the target pitch accent and the abbreviation the right referring to the treatment variable.) For each dataset, the phonological data was tabulated to provide an overview of the distribution of pitch accents and contours as a function of the treatment variable (foot size or anacrusis). The overall per-speaker distributions of pitch accents and contours were also tabulated. To remove excessive noise in the PN data, downstep was ignored, since it can be view as an effect of an initial high boundary tone rather than a reflection of a unique PN pitch accent.

Table . List of datasets used for the analysis of lexical and metrical effects on pitch accent phonology and phonetic implementation of tonal targets.

|  |  |  |  |
| --- | --- | --- | --- |
| Data | Target pitch event | Treatment variable(s) | Function |
| nuc\_foot | Nuclear pitch contour | Foot size (syllable count) | Analysis foot size effects on nuclear pitch accents and pitch contours. |
| nuc\_pre | Nuclear pitch contour | unstressed syllables preceding the stressed syllable. | Analysis of effects syllable count in syllables preceding on nuclear pitch accents and pitch contours. |
| pn\_foot | Pre-nuclear pitch accent | Foot size (syllable count) | Analysis of foot size effects on pre-nuclear nuclear pitch accents. |
| pn\_ana | Pre-nuclear pitch accent | Anacrusis (syllable count) | Analysis of anacrusis effects on pre-nuclear nuclear pitch accents. |
| pn\_lex | Pre-nuclear pitch accent | Anacrusis, foot size, location of word-end syllable | Analysis of word boundaries effects on pre-nuclear nuclear pitch accents. |
|  |  |  |  |

## Methods

The data set contains missing data points, so there is a risk that averaged or tabulated raw phonological data will be misleading as it will over- or under-represent specific speakers and target utterances. To avoid misleading representations, raw values have been adjusted to project a more balanced predicted distribution of pitch accents per condition and per speaker. This takes into consideration variation in both the number of repetitions per speaker and the number of speakers per target (in cases where a speaker produced no analysable utterances). The outline for this procedure is described in Appendix XXX and can be accessed at githubref.

For the inferential statistical analysis of the phonological data, generalized linear mixed effects models (GLMMs) were employed. As with all generalized linear models, they use logistic regression analysis. That is, they estimate the odds of a discrete outcome—in this case a particular pitch accent—occurring as a function of the treatment variable (or variables). Thus, a value of 1 (1:1) means that the model’s ability to predict the outcome variable is equal to chance. As the odds decrease below one, it means that the model’s ability to predict the outcome variable(s) becomes increasingly worse than chance, while values greater than one indicate odds increasingly greater than chance. As they are mixed effects models, GLMMS incorporate both fixed effects (the predictor variables) and random effects (known variables which contribute to the error in the model). This allows the model to compensate for variance caused by random effects, thus reducing the amount of error, or noise, they generate. For example, in the data analysed here, there are fixed effects such as foot size and anacrusis. This are effects with a limited number of levels, which can be controlled experimentally. However, the data are taken from eleven speakers, representing a random sampling of the population, and each has their own speech idiosyncrasies. Thus, each speaker adds noise to the model. By treating speaker as a random effect, we can reduce the effect of this noise. Moreover, we also reduce the distortion generated by the imbalance in the number of repetitions per speaker.

For the phonetic analysis, *f*0 mean, standard deviation, and range for each speaker was calculated using the A and H corpora in both Hertz and semitones re 1 Hz. The semitone scale was used for all *f*0 analyses. (see Chapter X, section Y).

## Phonological labelling

For the nuclear contours, labelling was very straightforward; however, pre-nuclear pitch accents presented more of a challenge. To help illustrate this, Table 6.5 presents stylizations of PN contours. In many cases, H\* was visually and auditorily salient, as shown in panels one to three. If there was a noticeable rise from a low boundary to the peak (Table 6.5, Panel one), the boundary of these PN H\*s was labelled as %L. These were labelled as such because, during the annotation, it seemed that in some cases the L of an L\*H may have become re-associated with the boundary rather than the stressed syllable. However, they did not appear to give rise to the percept of an L\*H pitch accent, nor did they seem to indicate a change in intonation function. In other cases of H\*, there was a plateau-like *f*0 stretch from the boundary towards a peak prominence in the stressed syllable (panel three). These were labelled as %H H\* to reflect that the plateau structure. Note that this use of %H differs ToBI labelling, in which an initial %H followed by another H target indicates a sharp drop from a very high boundary tone.

L\*H PN pitch accents were often auditorily salient (panels four to six), despite some variation in excursion size and the height of the L target. For example, sometimes the L target was visually less prominent (panel six) but was nonetheless perceived as an L\*H, both by the author and by MOR, the intonation specialist who checked the phonological analyses. The initial boundaries for L\*H tended to have a phonetically low or mid-range *f*0 and added no auditorily salient variation to the pitch contour, so these were labelled as unspecified (%).

In a few marginal cases, there was a rise from an initial mid-range *f*0 to a peak, which was audibly and visually later than the peak of the typical H\* pitch accent but not necessarily as late as the peaks in the L\*H pitch accents. This is illustrated in panel seven. In other cases, as stylized in panel eight, there was a phonetically high initial *f*0 followed by a plateau which ended after the stressed syllable. In both cases, there was no auditory percept of a low target, and, in fact, there was typically no visual clue to the presence of an L target either, unlike in L\*H PAs. Overall, these contours sounded neither quite like L\*H nor the typical H\*. They were labelled as >H\* to reflect the salient high quality, their later peak alignment, and the lack of any auditory or visual cue to an accompanying L target.

Table .. Stylized representation of contours typically found in PN position.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Stylization | Comment | Label |
| 1. |  | Phonetically low intial boundary rising to a peak in or near the end of the lexially stressed syllable. Auditorily salient peak prominence in the stressed syllable. | %L H\* |
| 2. |  | Phonetically mid-range intial boundary rising to a peak in or near the end of the lexially stressed syllable. Auditorily salient peak prominence in the stressed syllable. | % H\* |
| 3. |  | Phonetically high boundary, leading to a small peak or plateau-like prominence in the stressed syllable before falling slightly | %H H\* |
| 4. |  | Phonetically low or mid-range initial boundary with a clearly identifiable L\*H. | % L\*H |
| 5. |  |
| 6. |  | Low or mid-range initial boundary. L target is not as clearly defined as in X and Y; however, it is easily perceived as an L\*H |
| 7. |  | Mid-range initial boundary rising to a peak somewhere after the stressed syllable. Sounds somewhat intermediate between H\* and L\*H. However, there is no visually or auditorily salient L target. | %(H) >H\* |
| 8. |  | High initial boundary rising to a peak somewhere after the stressed syllable. Sounds somewhat intermediate in between H\* and L\*H. However, there is no visually or auditorily salient L target. |
|  |  |  |  |

## Results and analysis: phonology

### Phonology of nuclear pitch contours

Summaries of the distribution of nuclear pitch contours as a function both of foot size (foot\_syls) and of syllables preceding the stressed syllable (pre\_syls) are shown in Table 6.6 and

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| foot\_syls | prompt code | Raw counts (n=202) | | | Adjusted counts (n=220) | |
| L\*H % | L\*H L% | total | L\*H % | L\*H L% |
| 1 | A1211 | 50 | 1 | 51 | 50 | 5 |
| 2 | A0221\* | 55 | 0 | 55 | 55 | 0 |
| 3 | A1231 | 30 | 5 | 35 | 49 | 6 |
| 4 | A1241 | 54 | 1 | 55 | 54 | 1 |
|  | **total** | **189** | **7** | **196** | **208** | **12** |
|  |  |  |  |  |  |  |

Table 6.7 respectively, while the summary of contours per speaker is presented in Table 6.8. Each table shows the raw counts from the unbalanced data, with the adjusted projected counts in parentheses.

Table . Summary of nuclear contours across foot size conditions (number of syllables in the foot). Asterisk indicates the same prompt used in both nuc\_foot and nuc\_pre subsets.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| foot\_syls | prompt code | Raw counts (n=202) | | | Adjusted counts (n=220) | |
| L\*H % | L\*H L% | total | L\*H % | L\*H L% |
| 1 | A1211 | 50 | 1 | 51 | 50 | 5 |
| 2 | A0221\* | 55 | 0 | 55 | 55 | 0 |
| 3 | A1231 | 30 | 5 | 35 | 49 | 6 |
| 4 | A1241 | 54 | 1 | 55 | 54 | 1 |
|  | **total** | **189** | **7** | **196** | **208** | **12** |
|  |  |  |  |  |  |  |

Table . Summary of nuclear contours across conditions re syllables preceding the stressed syllable. Asterisk indicates the same prompt used in both nuc\_foot and nuc\_pre subsets.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| pre\_syls | prompt code | Raw counts (n=202) | | | Adjusted counts (n=220) | |
| L\*H % | L\*H L% | total | L\*H % | L\*H L% |
| 1 | A1111 | 50 | 4 | 54 | 51 | 4 |
| 2 | A0221\* | 55 | 0 | 55 | 55 | 0 |
| 3 | A0321 | 55 | 0 | 55 | 55 | 0 |
| 4 | A0423 | 37 | 1 | 38 | 53 | 2 |
|  | **total** | **197** | **5** | **202** | **214** | **6** |

We see that L\*H was used exclusively for nuclear pitch accents for all speakers across all conditions. This confirms the first half of hypothesis (2)**Error! Reference source not found.**, that L\*H is the dominant pitch accent in nuclear position. It also confirms hypothesis (3), that variation in metrical context has no effect on the inventory of nuclear pitch accents. However, two different boundary tones were produced, L% and 0%. L% accounts for only 12 out of 343 utterances, i.e., 3.5% of the raw utterance count (3.4% adjusted).

Table . Summary of nuclear contours by speaker across the nuc\_foot and nuc\_pre subsets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| speaker | Raw counts (n=343) | | | PAs per speaker (%) | |
| L\*H % | L\*H L% | total | L\*H % | L\*H L% |
| F5 | 34 | 0 | 34 | 100% | 0% |
| F6 | 35 | 0 | 35 | 100% | 0% |
| F12 | 23 | 10 | 33 | 69% | 31% |
| F15 | 30 | 0 | 30 | 100% | 0% |
| F16 | 35 | 0 | 35 | 100% | 0% |
| F17 | 28 | 0 | 28 | 100% | 0% |
| M4 | 31 | 0 | 31 | 100% | 0% |
| M5 | 33 | 0 | 33 | 100% | 0% |
| M8 | 23 | 1 | 24 | 97% | 3% |
| M9 | 25 | 1 | 26 | 97% | 3% |
| M10 | 34 | 0 | 34 | 100% | 0% |
| **total** | **331** | **12** | **343** | **97%** | **3%** |
|  |  |  |  |  |  |

As outlined in Chapter 3, section 3.3.1, low-frequency occurrences of L\*H L% were expected in declaratives, based previous AM studies of nIE. Furthermore, McElholm’s (1986) two-speaker study of DCE found instances of a rising-falling nuclear contour (tone C). The distribution of boundary tones does not appear to be an effect of the treatment variable, as we see that the counts rise and fall across conditions with no discernible trend, as shown in both Table 6.6 and

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| foot\_syls | prompt code | Raw counts (n=202) | | | Adjusted counts (n=220) | |
| L\*H % | L\*H L% | total | L\*H % | L\*H L% |
| 1 | A1211 | 50 | 1 | 51 | 50 | 5 |
| 2 | A0221\* | 55 | 0 | 55 | 55 | 0 |
| 3 | A1231 | 30 | 5 | 35 | 49 | 6 |
| 4 | A1241 | 54 | 1 | 55 | 54 | 1 |
|  | **total** | **189** | **7** | **196** | **208** | **12** |
|  |  |  |  |  |  |  |

Table 6.7, as well as in figure XXX. In fact, when we look at the summary of nuclear contours by speaker (Table 6.8), we see that most L% boundaries (10/12) are produced by a single speaker (F12), with the remaining two coming from one speaker each (M8, M9).

It is possible that occurrences of L\*H L% is largely an effect of F12’s idiolect but based on my own impressionistic interpretation of the nuclear contour—as well as with intuitions elicited from DCE speakers and others familiar with DCE—L\*H L% seems to serve a different communicative function from L\*H %. Namely, it gives the impression that the speaker is clarifying something they believe the listener should already know, as if they were implying the idea, “…and I thought you already knew that.” More technically, the speaker appears to be signalling to the listener that the propositional content is already given rather than new. In fact, when we consider the conversational contexts in which F12 uses the L\*H L% contour (shown in Table 6.9 and Table 6.10), it is reasonable to think that F12 intended the listener to interpret the utterances as, “I live with Valerie [and I thought you knew that]” and “They know Val [and I thought you knew that].” However, as the main aim of this analysis is to establish the effects of metrical and lexical effects rather than establish the function of different nuclear contours, it is sufficient to note that the occurrence of L\*H L% appears to be a matter of function rather than form.

Table 6.9 Foot size and nuclear pitch contours for F12

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| foot\_syls | code | verbal prompt and target utterance | L\*H % | L\*H L% |
| 1 | A1211 | Q: Who does he live with?  A: **He lives with Val.** | 5 | 0 |
| 2 | A0221 | Q: What did you say about Val's travel card?  A: **Val's is valid.** | 5 | 0 |
| 3 | A1231 | Q: Who does you live with?  A: **I live with Valerie**. | 0 | 5 |
| 4 | A1241 | Q: What's happening with those job applications?  A: **They need evaluating**. | 5 | 0 |
|  |  |  |  |  |

Table 6.10 Preceding syllables and nuclear pitch contours for F12

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| pre\_syls | code | verbal prompt and target utterance | L\*H % | L\*H L% |
| 0 | A1111 | Q: Have they met any of the others yet?  A: **They know Val.** | 1 | 4 |
| 1 | A0221 | Q: What did you say about Val's travel card?  A: **Val's is valid.** | 5 | 0 |
| 2 | A0321 | Q: What did you say about Val's travel card?  A: **Val's is invalid.** | 5 | 0 |
| 3 | A0423 | Q: What did you say about Valerie’s travel card?  A: **Valerie's is valid.** | 2 | 1 |
|  |  |  |  |  |

### Phonology of pre-nuclear pitch accents: foot-size and anacrusis effects

Summaries of the distribution of pre-nuclear pitch accents as a function of foot size (foot\_syls) and anacrusis (ana\_syls) are shown in Table 6.11 and Table 6.12 respectively. Each table shows the raw counts from the unbalanced data, with the adjusted projected counts in parentheses.

Table . Summary of pre-nuclear pitch accents across foot size conditions. Asterisk indicates the same prompt used in both pn\_foot and pn\_ana subsets

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| foot\_syls | code | Raw counts (n=203) | | | | | | Adjusted counts (n=220) | | | | |
| (\*) | L\* | H\* | >H\* | L\*H | tot. | (\*) | L\* | H\* | >H\* | L\*H |
| 1 | A0131 | 9 | 5 | 19 | 2 | 20 | 55 | 9 | 5 | 19 | 2 | 20 |
| 2 | A0221 | 2 | 2 | 21 | 3 | 27 | 55 | 2 | 2 | 21 | 3 | 27 |
| 3 | A0321 | 1 | 0 | 13 | 6 | 35 | 55 | 1 | 0 | 13 | 6 | 35 |
| 4 | A0423\* | 0 | 0 | 0 | 3 | 35 | 38 | 0 | 0 | 0 | 5 | 50 |
|  | **total** | **12** | **7** | **53** | **14** | **117** | **203** | **12** | **7** | **53** | **16** | **132** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

Table . Summary of nuclear contours across anacrusis conditions (number of syllables). Asterisk indicates the same prompt used in both pn\_foot and pn\_ana subsets.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ana\_syls | code | Actual counts (n=105) | | | | Adjusted counts (n=220) | | |
| H\* | >H\* | L\*H | tot. | H\* | >H\* | L\*H |
| 0 | A0423\* | 0 | 3 | 35 | 38 | 0 | 5 | 50 |
| 1 | A1422 | 9 | 5 | 42 | 56 | 9 | 5 | 41 |
| 2 | A2422 | 0 | 4 | 52 | 56 | 0 | 4 | 51 |
| 3 | A3422 | 0 | 4 | 41 | 45 | 0 | 4 | 51 |
|  | total | 9 | 16 | 170 | 195 | 9 | 18 | 193 |

The most noticeable feature of the distribution of PN pitch accents is that there are simply more of them, namely L\*, H\*, >H\*, and L\*H, along with several cases without any pitch accent, represented by (\*). This is in sharp contrast to the nuclear pitch accents, which were exclusively L\*H. While all five token types—including unaccented cases—are found in the pn\_foot data, only three occur in the pn\_ana data, namely H\*, >H\*, and L\*H. This increases in the inventory of pitch accents lends credence to hypothesis (4), that variation in metrical context has a strong effect on the (surface) phonology of pre-nuclear pitch accents.

The second striking feature is observed in the pn\_foot data, where we can see that as foot size increases there is an overall proportional increase in the occurrence of L\*H while instances of non-accentuation and L\* decrease. This pattern is particularly salient in Figure 6.1, which shows the adjusted PA distributions across PNs across foot-size conditions, and clearly indicate that non-accentuation and L\* occurrences drop off sharply while L\*H becomes more dominant as foot-size increases. This lend more support to hypothesis (4), as it does appear that increased foot size is associated with increased occurrence of L\*H.

Chart, bar chart, histogram

Description automatically generated

Figure . PNs across foot-size conditions (adjusted)

This trend is not mirrored in the pn\_ana data, where L\*H dominates throughout, as reflected in Figure 6.2, which shows the adjusted distributions of PA tokens across anacrusis conditions. However, all target phrases in the pn\_ana data already contain a four-syllable initial foot, so seems that foot size conditions are sufficient, anacrusis has little effect on the phonology. Unfortunately, the extent of the foot size effect had not been anticipated, so was not build into the corpus design.

Chart, histogram

Description automatically generated

Figure . PNs across anacrusis conditions (adjusted)

When we consider the overall distribution of pre-nuclear pitch accents by speaker—as presented in Table 6.13—we can see that L\*H dominates overall (70% of the average per-speaker proportion). This further confirms hypothesis (1), that L\*H is the dominant pitch accent in both nuclear and pre-nuclear position. We also see that as foot size increases, the likelihood of L\*H occurrences also increases. This trend holds true among all the female participants, but it is a mixed picture for the male speakers. M8 and M10 both show a clear preference for L\*H; however, M5 has an almost even split between L\*H and H\* (49% and 51% respectively), while M4 shows a preference for L\*H (42%), followed closely by H\* (32%). Unlike all the other speakers, M9 has a clear preference for H\* (67%), with only marginal use of L\*H (7%). In fact, after H\*, M9 uses the ambiguous >H\* most frequently (27%**)**.

Overall, this give the impression that there is an effect of gender, in that th

Table . Summary of pre-nuclear pitch accents by speaker across the pn\_ana and pn\_foot subsets.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| speaker | Raw values (n) | | | | | | PAs per speaker (%) | | | | |
| (\*) | L\* | H\* | >H\* | L\*H | **tot.** | (\*) | L\* | H\* | >H\* | L\*H |
| F5 | 4 | 0 | 5 | 2 | 24 | **35** | 11% | 0% | 14% | 6% | 69% |
| F6 | 3 | 0 | 0 | 1 | 31 | **35** | 9% | 0% | 0% | 3% | 89% |
| F12 | 0 | 1 | 0 | 0 | 32 | **33** | 0% | 3% | 0% | 0% | 97% |
| F15 | 0 | 0 | 2 | 2 | 32 | **36** | 0% | 0% | 6% | 6% | 89% |
| F16 | 0 | 0 | 0 | 0 | 35 | **35** | 0% | 0% | 0% | 0% | 100% |
| F17 | 1 | 6 | 0 | 0 | 25 | **32** | 3% | 19% | 0% | 0% | 78% |
| M4 | 4 | 0 | 10 | 4 | 13 | **31** | 13% | 0% | 32% | 13% | 42% |
| M5 | 0 | 0 | 17 | 0 | 18 | **35** | 0% | 0% | 49% | 0% | 51% |
| M8 | 0 | 0 | 2 | 9 | 22 | **33** | 0% | 0% | 6% | 27% | 67% |
| M9 | 0 | 0 | 20 | 8 | 2 | **30** | 0% | 0% | 67% | 27% | 7% |
| M10 | 0 | 0 | 6 | 1 | 28 | **35** | 0% | 0% | 17% | 3% | 80% |
| Tot. raw / ave. % | 12 | 7 | 62 | 27 | 262 | 370 | 3% | 2% | 17% | 8% | 70% |
|  |  |  |  |  |  |  |  |  |  |  |  |

Inter-speaker variation in the realisation of pitch accents—especially considering the outlier M9—demonstrates that the correlation between foot size and L\*H realisation suggested by the summary data (Figure 6.1) is not representative of all individual trends. Thus, the overall trend could simply be an over-simplification produced by the aggregation of speaker data, which in effect treats all speakers as one single unit. The more likely explanation, however, is that the aggregated data do indeed reflect the overall trend, but speakers such as M9 (and to a lesser extend M5 and M4) simply do not follow it. It is possible that these speakers continue to produce H\* because, for them, it is functionally contrastive with L\*H, or there is no intended functional contrast, but underlying /L\*H/ in PN position is much continues to be affected by metrical conditions. Without access to the black box of speaker intent, however, it is difficult to assert that one or the other is correct.

… further investigations of adjusted data:

1. Mean speech rate appears to correlate with proportion of occurrences of L\*H and H\*
2. Gender and acc\_phon appear to be correlated
3. Gender and speech rate cannot be correlated within an exceptable p value (p.=0.8926)
4. Since gender and SR do not appear to be correlated, a GLMM using gender\*SR doesn’tmake sense to me … most likely highly explanatory model is:

isLH ~ foot\_syls + speech\_rate + gender + (1 | speaker)

…GLMM analysis:

1. GLMM models should assess effect of gender, speech rate, and foot\_syls individually [bf adj = 3]
2. If we consider that foot\_syls and speech rate may be conflated into duration, we could also consider duration as a factor [bf adj = 1]
3. Gender + foot\_syls + speech\_Rate should also be considered (even though gender does not vary within speaker… here at least) [bf adj = 1]
4. Leads to Gender + goot\_duration as another test [bf adj = 1]
5. Foot\_syls alone should be considered for pairwise comparison to test original hypothesis, even if effects of SR and gender also turn out to be important. [bf adj = 3]
6. Total Bonferroni adjustment = 9 🡪 even with this very high adjustment, the results of the less detailed analysis still hold!

Conclusion:

Foot\_syls, gender, and SR matter. So too does foot duration, but it has a cumulative effect (R2m) less than foot\_syls + SR (0.58 vs 0.47, which suggests that foot\_syls effects may be not simply be a proxy for the duration of the foot, but that foot\_syls and SR may effect PA choice at the level of planning(????)

Marginal R2 of [foot syls] + [gender] + [speech\_rate] ~= Marginal R2 of [foot syls + gender + speech\_rate], which may be a nice way of showing how each of these independently influences the Pitch accents.

Chart, bar chart

Description automatically generated

Chart

Description automatically generatedA picture containing chart

Description automatically generated

### Statistical modelling of foot size effects on PN pitch accents

To assess the strength of the apparent foot size effects on PN pitch accent distribution, a generalized linear mixed effects model (GLMM) analysis was conducted. As with all generalized linear models, this uses logistic regression analysis. That is, it estimates the odds of a discrete outcome—in this case a particular pitch accent—occurring as a function of the treatment variable (or variables). Thus, a value of 1 (1:1) means that the model’s ability to predict the outcome variable is equal to chance. As the odds decrease below one, it means that the model’s ability to predict the outcome variable(s) becomes increasingly worse than chance, while values greater than one indicate odds increasingly greater than chance. As a mixed effects model, a GLMM incorporates both fixed effects (the predictor variables) and random effects (known variables which contribute to the error in the model). This allows the model to compensate for variance caused by random effects, thus reducing the amount of error, or noise, they generate. In the data analysed here, there is fixed effect of foot size. This has a limited number of experimentally controlled levels. The speakers, however, represent a random sampling of the population, and each has their own speech idiosyncrasies which influence the results. Thus, each speaker adds noise to the model. By treating speaker as a random effect, we can reduce the effect of this noise. Moreover, we also reduce the distortion generated by the imbalance in the number of repetitions per speaker.

The statistical analysis was conducted in R (R Core Team, 2022). The lme4 package (Bates *et al.*, 2015) was employed for the GLMM modelling along with optimx (Nash and Varadhan, 2011; Nash, 2014), to provide improved optimization to avoid convergence problems with the modelling. The MuMIn (Bartoń, 2022) package was used to provide a pseudo-r-Squared statistic for the model. The tidyverse, broomExtra, sjPlot, formattable, knitr, and mefa4 packages were employed to facilitate tidy coding and presentation of the data (Sólymos, 2009; Xie, 2014; Wickham *et al.*, 2019; Lüdecke, 2021; Ren and Russell, 2021; Patil, 2022). To permit pairwise comparison, the model was tested four times using each treatment condition in turn as the intercept, and p values were Bonferroni adjusted accordingly (p × n = 4). The code for this analysis can be found at GitHubREF and is also reproduced in Appendix XX.

To test the likelihood of L\*H occurrences increasing with foot size in PNs following GLM model was tested, with speaker as a random intercept:

1. isLH ~ foot\_syls + (1 | speaker)

where isLH true if the pitch accent is L\*H. A likelihood ratio test of the model against a model without foot\_syls as a fixed effect indicated a significant difference between the two models (χ2(3) = 39.348, p < 0.0001).

Marginal and conditional pseudo-r2 estimates (r2m=0.178, r2c=0.658) indicate that the fixed effect (foot size) accounts for 18% of the variance while the whole model (fixed and random effects included) account for 66% of the variance in the data. The difference in the marginal and conditional r2 suggests that individual inter-speaker differences (random effects) have a noticeable influence on the likelihood the realisation of L\*H, and that, while not as great, foot size does make a distinct contribution.

Summary statistics of the model are presented in Table 6.14 and a plot of the model in Figure 6.3, with foot\_syls1 as the intercept. The plot shows the log odds for each condition, with the red line at zero indicating chance. We see that the odds of the one-syllable condition (the intercept) predicting L\*H are below chance (estimate = -0.97). The likelihood of each subsequent condition to predict L\*H when compared with the foot\_syls1 rises above chance and increases with foot size. However, the effect of the intercept (p.=0.2139) is not statistically significant, nor is there a statistically significant difference between the intercept and foot\_syl2 (9=0.0727). Only for foot\_syl3 and foot\_syl4 are the slopes significantly significant. In the figure, the overlapping standard errors reflect the lack of statistical significance in the first two terms.

Diagram

Description automatically generated

Figure . plot of the model isLH ~ foot\_syls + (1 | speaker). Values indicate log odds for each condition, while the black line indicates standard error. The red line at zero represents chance.

Table .. Summary statistics for model isLH ~ foot\_syls + (1 | speaker). Values indicate log odds for each condition, while the black line indicates standard error. p.adjusted is the p. value after Bonferroni adjustment (n=4).

| term | estimate | std.error | z.value | p.adjusted | signif. (adj.) |
| --- | --- | --- | --- | --- | --- |
| (Intercept) | -0.966 | 0.777 | -1.243 | 0.8554 |  |
| foot\_syls2 | 0.925 | 0.516 | 1.795 | 0.2909 |  |
| foot\_syls3 | 1.968 | 0.546 | 3.608 | 0.0012 | p < 0.01 |
| foot\_syls4 | 3.955 | 0.821 | 4.815 | 5.89e-06 | p < 0.001 |
|  |  |  |  |  |  |

For the pairwise comparisons (Table 6.15 and Table 6.16), the intercept was significant only for foot\_syls4 (p < 0.01). The slope for foot\_syls4 was always significant, as was the slope for foot\_syls3 when foot\_syls1 was the intercept (p < 0.01). This suggests that only foot\_syls4 can be relied upon as a good predictor of L\*H, and that when there is a different of at least two syllables between conditions, the foot with the higher syllable count is comparatively a significantly better predictor than the one with the lower count. i.e., foot\_syls2 is probably not a better predictor of L\*H than foot\_syls1, but foot\_syls3 is.

Table . Intercept (β0) estimates from pairwise comparison tests for model isLH ~ foot\_syls + (1 | speaker). p.adjusted is the p. value after Bonferroni adjustment (n=4)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **intercept** | **estimate** | **std.error** | **z.value** | **p.adjusted** | **signif. (adj.)** |
| foot\_syls1 | -0.966 | 0.777 | -1.243 | 0.8554 |  |
| foot\_syls2 | -0.041 | 0.770 | -0.053 | 1 |  |
| foot\_syls3 | 1.002 | 0.776 | 1.291 | 0.7874 |  |
| foot\_syls4 | 2.989 | 0.986 | 3.031 | 0.0098 | p < 0.01 |
|  |  |  |  |  |  |

Table . Slope extimates (β1) re each Intercept (β0) from pairwise comparison tests for model isLH ~ foot\_syls + (1 | speaker). p.adjusted is the p. value after Bonferroni adjustment (n=4)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **intercept** | **term** | **estimate** | **std.error** | **z.value** | **p.adjusted** | **signif. (adj.)** |
| foot\_syls1 | foot\_syls2 | 0.925 | 0.516 | 1.795 | 0.2909 |  |
| foot\_syls1 | foot\_syls3 | 1.968 | 0.546 | 3.608 | 0.0012 | p < 0.05 |
| foot\_syls1 | foot\_syls4 | 3.955 | 0.821 | 4.815 | 5.89e-06 | p < 0.001 |
| foot\_syls2 | foot\_syls3 | 1.043 | 0.515 | 2.025 | 0.1715 |  |
| foot\_syls2 | foot\_syls4 | 3.030 | 0.795 | 3.810 | 5.56e-04 | p < 0.001 |
| foot\_syls3 | foot\_syls4 | 1.987 | 0.783 | 2.538 | 0.0446 | p < 0.05 |
|  |  |  |  |  |  |  |

### Summary of Phonological analysis

Taken together, the results confirm the three phonological hypotheses. L\*H is the only nuclear pitch accent, while L\*H dominates in prenuclear position even though other pitch accents occur. The aggregation of prenuclear pitch accent data across foot-size and anacrusis conditions indicates that with increased foot size, occurrences of de-accentuation and L\* decreases, while instances of L\*H steadily increase. However, there is no evidence for anacrusis effects. This is mainly due to the fact that each target utterance in the pn\_ana dataset also had a four-syllable foot, the effect of which was much more dramatic than anticipated. (i.e., four-syllable feet are already saturated with L\*H regardless of anacrusis size). To test the prediction that foot-size has a significant effect on the realisation of L\*H, a GLMM analysis of foot size as a predictor of PN L\*H with speaker as a random effect was conducted. The results indicate that as foot size increases so too does the likelihood L\*H occurrence. However, foot\_syls4 alone is a statistically significant predictor of the L\*H, a fact already suggested by the saturation of foot\_syls4 with L\*H. Furthermore, the predictive strength of any condition in comparison with another condition appears to be significant only when the foot is at least two syllables longer than the alternative. That is, the likelihood of a three-syllable foot having an L\*H PN accent in contrast to a one-syllable foot having one is statistically significant (p.=0.012, p < 0.05), but this is not the case when a three-syllable foot is contrasted with a two-syllable foot (p.=1.175, p>0.05).

\*\*\* Something about nucs and pns being different; is PN different or is it just a variant? Does the speaker change PA or just shift alignment?\*\*\*\*

## Results and analysis: phonetic parameters

### Alignment and Scaling of nuclear pitch accents

#### Foot size

Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated

The L target of the L\*H % appears to be barely affected by changes on foot size, while the H target appears to increase with increased foot size.

[CHECK TO SEE THE TIMING RE SYLLABLE BOUNDARIES]

In L\*H L%, There only appears to be backward pressure on the alignment of the L\* in the single syllable foot, which is understandable as it has the least amount of segmental space. In all cases, the H target appears to be pushed backwards, by 58 ms in the 1-syl conditions, and steadily decreasing to 41 ms by the 4-syl condition. This decreasing differential may be accounted for by the fact that there is more segmental material with increased footsize and so slightly less backwards pressure on the H target in order to accommodate the L target, which can be realised later. Coincidence of L% and 0% target alignment is a measurement effect, as it was taken at the end of voicing.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, bar chart

Description automatically generated

We see that the L target in L\*H % tends to slightly higher as foot size increases overall, but that it is slightly lower in the 3-syl condition. However, these changes are unlikely to be significant [SEE??]. The H target increases steadily from the one-syllable to the three syllable condition (0.81 to 1.76 standard deviations), but the seems to fall slightly in syl=4. This suggests that after three syllables, *f*0 is unlikely to keep rising. It also suggests that if we see truncation effects, they are unlikely to occur in the syl-3 and syl-4 conditions. The L\*H % boundary *f*0 is slightly lower in the 1-syl condition when compared to other conditions, where it does not change much.

It is more difficult to interpret the effects of foot size in the L\*H L% PAs. The L target from the 1-syl slightly between the 1-syl and 3-syl condition, but it increases noticeably in the 4-syl condition. With increased foot size, the H target appears to be become lower, while the *f*0 of the boundary tones drops between syl-1 and syl-3 (from -2 to -3 SDs) then increases sharply for syl-4 (-1 SD). However, there were no L\*H L% tokens in the 3-syl condition, few in total, and the vast majority of them came from a single speaker. That is, 7 out of 196 (3.6%) of all utterances in nuc\_foot but 5 out of 20 (25%) of F12’s. Therefore, it is unwise to generalise from such limited data.

Chart, line chart

Description automatically generatedChart, bar chart

Description automatically generated

When we consider the (stylised) shape of the contours, we note that L\*H % becomes slightly compressed between the 4-syl and the 3-syl conditions; with little effect on *f*0. However, as the foot size further decreases, we can see that both the timing scaling of the rise decrease while the slope remains steady. This suggests that in the 1-syl and 2-syl contexts truncation effects occur, while in the 3-syl condition, there is a compression effect. (Of course, this assumes that 4-syl is a realisation of the PA with the least amount of compression or truncation pressure)

Looking at L\*H L%, we see that there appears to be an overall compression effect, in which the H target is aligned earlier, and actually increases slightly. The L% target the 4-syl condition is much higher than that of the 1-syl and 3-syl, but the mean timing between the peak target and the boundary remains very similar (108, 121, and 114 ms for syl-1, syl-3, and syl-4 respectively). This could imply that, when the boundary is specified, the co-ordination of timing between the high target and the low boundary is more important than it is when the boundary is unspecified. However, again, there is little data for L\*H L% nuclear contours.

#### Preceding syllables

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

It appears that the addition of unstressed syllables before the stressed nuclear syllable has almost no effect on the timing of tonal targets in L\*H %, although there may be a very slight downward trend in their realisation. All targets in L\*H L% are slightly earlier than in L\*H %, with the L\* and the H targets in pre-0 being somewhat earlier than those in pre-3 (58 and 54 ms compared to 25 and 30 ms respectively). Again, giving the paucity of L\*H L%, these are interesting and expected observations, but it is unwise to draw conclusions from them.

Chart, line chart

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, bar chart

Description automatically generated

The mean *f*0 data suggests little effect of the pre-syls conditions on the L target *f*0. In L\*H %, average *f*0 is lowest in the pre-0 conditions (1.64 SDs), and is slightly higher in the other conditions (-1.24, -1.45, and -1.35 respectively). [IS THE stress clash (“They KNOW VAL.”) IN ANA-0 IMPORTANT?] Average *f*0 of H targets increase with the addition of unstressed syllables before the nucleus, with the most noticeable jump being from pre-0 to pre-1 (0.22 to 1.09 SDs). There appears to be little change between pre-1 and pre-2, however, with an increase of 0.58 SDs increase ana-3 to ana-4. Mean *f*0 targets L\*H L% follow roughly the same pattern, although the L and H targets are lower than those of the L\*H % contours.

The overall increase in *f*0 across conditions is particularly interesting because there is almost no change in the timing of the targets. Moreover, due to an oversight in the selection of targets for the pre-syls conditions, pre-0 ends in a one-syllable foot (“VAL”), while the others end in a two-syllable foot (“VALid”). Despite the difference in foot size, the timing of targets does not change much. This would suggest that the speakers extend the duration of the one syllable foot, in effect matching text to tune [REF]. It also suggests that the different in *f*0 between pre-0 and pre-1 is not best explained by foot duration (in ms) but rather by foot size (in syllables).

Chart

Description automatically generatedChart, bar chart

Description automatically generated

When we consider the shape of the contours, we see that mean timing of all targets in L\*H % changes little, but that there is an overall increase in scaling from ana-0 to ana-3. This suggests that the addition of more unstressed material before the nuclear stress might permit a planning window in which the speaker is more apt to adjust the H target toward a higher, more salient *f*0. This is reflected in the slopes of L\*H % and L\*H L%, which rise steadily with the addition of preceding syllables.

### timing and Scaling of prenuclear pitch accents

#### Foot size

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, bar chart

Description automatically generated

#### Anacrusis

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart

Description automatically generated

#### Word boundaries

## Discussion

### Phonology

##### General comments and observations

##### Issues with phonological inventory used

##### Proposal for phonological reanalysis re peak delay

### Phonetics and phonology

##### General observations

##### Issues regarding tonal targets

## Conclusion

### Summary

### Proposal: possibility of separate category of >H\*

##### L\* deletion?

##### Delayed H\* Target due to segmental or syllable pressure?

## Conclusions from Analysis of Form in Unmarked Sentences

### Summary of findings (RQ1, RQ3)

### Limitations of underspecified inventories and tonal targets (RQ4)

# Analysis of Function: Sentence modes

Chapter 3 observed that AM analyses of nIE attests to the dominance of L\*H pitch accents in IP nuclear position across declarative and question forms, including declarative questions, which are lexically and grammatically identical to declarative statements. This contrasts with other (standard) varieties of English, where speakers tend to employ a falling nuclear contour for statements (H\* L% or H\*L %) and a rising contour for questions (L\*H % or L\*H H%). Even in varieties of English which contain declarative rises (e.g., varieties in the North of England), the declarative rise does not dominate to the extent that has been found in nIE.

It was also observed that, if one follows the AM approach championed by Gussenhoven [REF] and illustrated in Haan’s study of Dutch question intonation [REF], the distinction between declarative intonation and question intonation can be ascribed to paralinguistic effects. That is, as the number of lexical and grammatical cues to interrogativity decrease, the overall scaling of *f*0 is likely to increase in a gradient fashion. Thus, in nIE, while L\*H may dominate across all sentence modes, interrogativity will be signalled paralinguistically.

Two problems were noted with the paralinguistic/linguistic split in the description of intonation. Firstly, it was argued that if there is a consistent correlation between *f*0 contours and grammatical function, it is important to investigate seriously the possibility that the correlation reflects a phonological event and is not simply a paralinguistic effect. Not to do this may lead to a corollary danger. Namely, by partitioning off uncooperative data as paralinguistic, we are excluding data which challenge the theory’s ability to adequately describe our observations. This helps preserve the phonological theory as is rather than allowing it to evolve to provide a fuller description of the data.

The second problem relates to the way in which the paralinguistic/linguistic split suggests an unlikely typological division between nIE and other varieties of English. That is, if question forms are not typically expressed using intonational phonology but rather must be signalled by paralinguistic means in nIE, then this would imply that nIE does not have recourse to a chunk of the English Grammar system that other varieties of English do. Of course, we do expect that different varieties of any language will vary in the form, application, and distribution of structural components; however, it is a different proposition altogether to imply that a feature is altogether missing. Thus, with nIE, it seems unlikely that—even as L\*H dominates—it offers no consistent recourse to a phonological intonational form to signal the difference between declaratives and interrogatives.

The proposed solution is that nIE speakers (speakers from Derry City in this case) exploit a phonological register tier to distinguish between interrogative and declarative forms, called here the register tier hypothesis. Note, it is possible that speakers of other varieties also exploit a register tier, but—as there is also recourse to a distinction in the distribution of pitch accents—this may largely go unnoticed or may appear redundant.

Thus, this chapter serves to assess the viability of the register tier hypothesis in the description of the phonology and phonetics of sentence mode in Derry City English. That is, it aims to answer the following research questions:

1. **Descriptive:** What are the phonological and phonetic characteristics of nuclear pitch contours in DCE across sentence modes?
2. **Theoretical:** Does a **register tier** provide a plausible phonological explanation for variation across sentence modes in DCE?

Although utterance-wide intonational features are not central to the core questions, an examination of them should be useful, if for nothing else than for the sake of completeness. Therefore, the chapter will also consider global (or utterance-wide) phonological and phonetic features of intonation and mode[[8]](#footnote-9).

## Expectations

There will be two sets of analyses of the phonetics and phonology of intonation in relation to mode in this chapter. One will assume a null hypothesis where there is no register tier, while the other will assume that the register tier exists. The first will be called the non-register-tier analysis, and the second the register-tier analysis. In employng the two approaches, the plausibility (or lack thereof) of the register tier hypothesis should become clearer. That is, for the register tier hypothesis to hold water, we need to establish if the register-tier analysis can provide a more coherent, efficient, and transparent explanation of the intonational phonology and its phonetic implementation than the non-register-tier analysis can.

There will be a degree of overlap in the results of the two sets of analyses. Therefore, to avoid redundancy and repetition, wherever the same expectation applies to both sets, it will not be discussed in full in both.

### Phonological analysis

Setting aside the resister-tier component, L\*H % is expected to dominate as the nuclear contour across modes. However, both L\*H % and L\*H L% were observed in the analysis of formal effects on declaratives, so L\*H L% is also expected. Given that there were no instances of H\* % or H\*L% in the H and A corpora, they are unlikely to occur in statements and exceedingly unlikely in yes-no or declarative questions, with which these forms are not typically associated in other English varieties.

Broadly, most final boundaries in the M-corpus should be unmarked; however, we should expect some differences in the distribution of boundary tones compared to the A- and H-corpora of the previous chapter. Based on studies of Belfast English, we can expect L\*H H% in the nuclear contour—even if relatively rare—increasing in frequency from MYN to MDQ. That is, we should expect H% to be used to reinforce interrogativity. We should also expect to see L% in boundary tones, as previously; however, we should also expect a different distribution across modes.

In chapter 6 (6.5.1), it was suggested that L% is used for discourse functional purposes, namely that the speaker appears use L% to signal that a previous expectation of givenness conflicts with a newer understanding of the shared knowledge in the discourse space. For example, the statement “I live with Valerie”, if it ends with an L%, seems to imply the additional meaning of “I live with Valerie [L% = and I thought you already knew that].” We can also expect L% to occur in MDQs, where the speaker is questioning the propositional content of the whole utterance. For example, in the question, “You live in the valley?”—which can be interpreted as a checking question—the speaker might want to indicate that the new information (embedded in the propositional content of the sentence) conflicts with what the expected to be true, and they might use L% to indicate this, i.e., “You live in the valley? [L% = I’m surprised. I’d never have guessed.]” However, this does not mean that L% signals interrogativity, rather that the discourse function it represents is more compatible with the interrogativity of declarative questions. Finally, the use of a low boundary will be the same regardless of whether it is in service to a register-tier or non-register-tier analysis.

This intuition regarding L% is in essence a re-articulation of the surprise and redundancy contour described by Sag and Liberman in GenAm, which also involves a final fall (1975). However, it is not that surprise and redundancy share the same intonational contour or, as Sag an Liberman suggest it might (p. 497), that redundancy is a secondary effect. Here, the use of L% instantiates the same discourse mechanise. The L% is interpretable as surprise in the echo question—which, as an echo, is inherently also redundant—since the speaker uses it show surprise that the propositional content in the sentence conflicts with their understanding of already establish shared knowledge in the discourse. In the declarative statement responding to the question, “Where do you live?” it is also surprise since the speaker is indicating surprise that the information was not already established fact, and also therefore—from the speaker’s point of view—the response is (or should have been) redundant.

In associating L% and H% with different functions, this reflects and implicit view that boundary tones are compositional in relation to meaning. H% reinforces the question status of the utterance, while L% indicates a conflict between the speaker’s understanding of givenness and the current state of the discourse. How these two conflicting boundary-tone functions might manifest in a single question utterance, however, is unclear. One possibility is that a speaker may use a compound HL% boundary to signal interrogativity and indicate the conflict in their understanding of shared knowledge.

Expectations specific to the non-register tier analysis of intonational phonology are as follows:

1. L\*H % will dominate across sentence modes in the non-register tier analysis.
2. H% can re-enforce interrogativity in the non-register tier analysis, leading to increasing frequency of H% in MYNs and MDQs.
3. Compound HL% boundaries *may* also occur in the register tier analysis.

When considering the register-tier analysis, we expect that the H% will generally not be required since the register tier will already account for the higher scaling in the nucleus. We also expect that the register-tier analysis will provide a better account of phonological change across sentence modes than the non-register tier account. The expectations from the register-tier phonological analysis are as follows:

1. Patterns will occur which are adequately explained only with reference to both a register tier and the tonal tier.
2. The register-tier analysis will indicate phonological changes across sentence mode more effectively than the non-register-tier analysis.

For both H% and high register, the expectation is that neither will occur in the nuclear pitch accent of MDC or MWH but will occur with increasing frequency in MYN and MDQ.

The low boundary is hypothesised NOT to be primarily a function of mode, unlike either H% or high register. Therefore—if we include compound boundaries HL% from the non-register tier analysis as a variant of L%—we expect no difference in the use and distribution L% in either approach. The expectation from L% is:

1. L% will occur in all modes but more frequently in MDQs.

### Phonetic analysis of tonal targets

If we reject the register tier hypothesis, we should expect to observe gradient *f*0 scaling of tonal targets in pitch L\*H pitch accents as a function of sentence mode. That is:

1. Non-register tier analysis: *f0(*MDC) <= *f0(*MWH) < *f0(*MYN) < *f0(*MDQ)

If, on the other hand, the register tier hypothesis is valid, we should expect to see those differences scaling effects disappear once the register tier has been incorporated into the model. i.e., in a register tier analysis, *f*0 scaling will more appropriately be associated changes in pitch accent and register tier[[9]](#footnote-10) rather than mode.

Further, we should expect to find significant differences in the scaling of pitch accents themselves when they are subject to register shift. Thus, the final two expectations are as follows:

1. Apparent paralinguistic differences in scaling of tonal targets across modes will disappear in a model incorporating the effects of the register tier.
2. There will be significant differences in the scaling of tonal targets across pitch accents due to register tier effects.

### Phonetics and phonology of Utterances

In both the phonological and phonetic analysis of the utterance, we should expect MDC and MWH forms to be similar. This is because wh-questions are most highly saturated with lexical and morphosyntactic cues to interrogativity and thus are least likely to require intonational support to distinguish them from statements. However, question words are typically prominent, so they are more likely to be associated with pre-nuclear H\* PAs. H\* is less likely to occur in declaratives, where PN accentuation may be less common. If we follow Haan’s hierarchy [REF], we should expect to see IP mean *f*0 increase from wh-questions to yes-no questions to declarative questions as other cues to interrogativity disappear. Given the greater likelihood of wh-questions beginning with an H\*, we can also expect the slope of the IP in wh-questions to be lower than both declaratives and other question forms. We should also expect the slope to increase from statements to yes-no questions to declarative questions. Typically, this is attributed to gradient effects of paralinguistic biological codes. However, it could equally be an effect of the deployment of the proposed register tier in the nuclear contour or across the utterance.

If we accept the register tier hypothesis, we should expect to see increased use of high register in YNQ and DCQs, which will parallel changes in the scaling of *f*0 in tonal targets. That is, an increase in the use of high register in question forms should increase in inverse proportion to the amount of syntactic and lexical markers of interrogativity, as described in expectation 7) above.

## Materials

The M-Corpus is used for the analyses in this chapter. The stimuli were designed to assess phonological and phonetic variation across sentence modes, namely, declarative statements (MDC), wh-questions, (MWH) yes-no questions (MYN), and declarative questions (MDQ). There are three phrases per sentence mode, ending with the word *valley*, *vases*, and *valuables* in turn. These were chosen to have both two- and three-syllable final feet. *Valley* was chosen as it ends in a fully voiced syllable, while *vases* has a much higher chance of being phonetically devoiced. By systematically varying syllable count and the potential amount of voiced material in the tail of the nucleus, it is possible to identify if these effect the timing or height of trailing tones, i.e., they can be used in future work to assess truncation and compression effects. As with all the read speech stimuli, each target utterance is a response to a stimulus read by the participant’s speaking partner and embedded within a short dialogue (see 5.1 and 5.3).

|  |
| --- |
| Table . Target phrases for sentence mode analysis. |
| |  |  |  | | --- | --- | --- | | **Code** | **Mode** | **target response (B)** | | MDC\_1 | declarative | I valued the **vases**. | | MDC\_2 | declarative | I live in the **valley**. | | MDC\_3 | declarative | I've hidden the **valuables** | | MWH\_1 | wh-question | Who valued the **vases**? | | MWH\_2 | wh-question | Why do you live in the **Valley**? | | MWH\_3 | wh-question | Where have you hidden the **valuables**? | | MYN\_1 | yes-no question | Have you valued the **vases**? | | MYN\_2 | yes-no question | Do you live in the **Valley**? | | MYN\_3 | yes-no question | Have you hidden the **valuables**? | | MDQ\_1 | declarative question | You valued the **vases**? | | MDQ\_2 | declarative question | You live in the **valley**? | | MDQ\_3 | declarative question | You've hidden the **valuables**? | |  |  |  | |

Ideally, there would be 660 utterances in the M-corpus. As previously, if there speaker or recording errors, speakers were asked to record one or two extra repetitions, and only the good repetitions were retained. Unfortunately, some errors weren’t noticed until later, so there is still some data loss. After repetitions with disfluencies or speaker and recording errors were excluded, there were a total of 639 utterances, as shown in Table 7.2. (Note that there are no repetitions for M9\_MDQ1, as the interlocuters accidentally skipped this prompt, an error which was missed at the time of recording.)

Table .. Summary of valid utterances in M-Corpus. Red indicates missing utterances, green superfluous ones.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | MDC1 | MDC2 | MDC3 | MDQ1 | MDQ2 | MDQ3 | MWH1 | MWH2 | MWH3 | MYN1 | MYN2 | MYN3 | TOTAL |
| F5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 59 |
| F6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 60 |
| F12 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 60 |
| F15 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 60 |
| F16 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 60 |
| F17 | 5 | 4 | 5 | 5 | 5 | 5 | 5 | 1 | 5 | 5 | 5 | 5 | 55 |
| M4 | 5 | 5 | 5 | 5 | 2 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 57 |
| M5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 60 |
| M8 | 5 | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 61 |
| M9 | 5 | 3 | 5 | 0 | 3 | 3 | 5 | 5 | 5 | 5 | 3 | 5 | 47 |
| M10 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 60 |
| Total | 55 | 53 | 55 | 50 | 50 | 53 | 55 | 51 | 55 | 55 | 53 | 54 | 639 |

## Methods

As with each corpus in Part II, utterances were annotated in Praat and a data table was extracted using the process\_texgrid script, as described in Chapter 5. The data table for M-Corpus can be found at [REF]. IViE labelling conventions were used as the basis for the phonological labelling (Grabe, 2001). However, during the process of annotation, IViE labelling proved inadequate for labelling distinctive pitch patterns—which is be expected given the register tier hypothesis—so modifications were made which both accommodated the register tier hypothesis while preserving the underlying IViE labelling system. These adjustments and the rationale behind them are outlined in the following section.

For representative visualisation of count data, raw counts were adjusted to be proportionally representative, using the process outlined in Chapter Five (5.3). For inferential statistical analyses, phonetic parameters are analysed using Linear Mixed Effects models (LMEs) while phonological categories are analysed using Bayesian Generalised Linear Mixed Effects models (BGLMs), as outlined in Chapter Five (5.4). As with all chapters, all inferential statistical analyses were conduct in R (R Core Team, 2022). All code and markdown for this chapter can be found in [GitHub refs] and [appendix refs].

## Phonological labelling

As with the A and H sub-corpora, L\*H dominated the nuclear position. However, there were many cases where simply labelling them as L\*H would have been misleading, and an alternative which incorporated the register tier into the labelling was developed. Firstly, there were often cases the L\*H simply occur at a distinctly high register. In fact, the raised register sometimes covered the whole IP in such a way that the whole contour of MDQ was essentially a copy of the MDC raised a several semitones. In such cases, the nuclear contour L\*H % was highly salient in each case but so too was the IP-wide raised register. In other cases, the raised register was limited to the nuclear pitch accent, or possibly even just a single tone. In fact, while the L\*H quality of the nuclear pitch accent was very salient, it was difficult to label the data without also accounting for local and global changes in register. Several examples are provided below to further demonstrate the issue, including the problems raised by non-register tier alternatives. After this, the approach to labelling adopted hereon in is outlined.

Figure 7.2, which presents all the pitch contours for F5 for MDC2 and MDQ2, in which we can see that there is a noticeable shift to high register in MDQs and that they are realised at a categorically different register in the speaker’s range.

|  |
| --- |
|  |
| Figure . An illustration of IP-wide register shift in MDQ across all repetitions. |

Figure 7.3 illustrate cases where the raised register occurs on the PA but not on the pre-nuclear stretch or on the boundary tone. Each phrase has a distinct L\*H nuclear pitch accent, which we see in the rise out of the stressed syllable in the final foot at roughly 65-75% within the utterance. One MDC has an unspecified boundary while the other has a clear L%, and the same is true of the two MDQ contours. We can also see that during the prenuclear stretch (up to roughly 55%), each contour realised an H\* pitch accent with a similar pitch range. However, in the nuclear pitch accent of the MDQs, there is only a very slight fall in *f*0 in the stressed syllable before it begins to rise. This is quite distinct form the MDCs, where the fall is much greater. Despite the different extents of the *f*0 fall before the nuclear pitch accent, we see that the rise in each MDQ is essentially a copy of the MDC pitch accent but just in at a higher register. For the MDQ and the MDC with the unspecified boundaries, there is just a slight drop in *f*0 just before the offset of voicing. However, in each utterance with an L% boundary, *f*0 falls to the speaker’s baseline (roughly 2.11 log10 *f*0, or 130 Hz). This suggests that the process which has affected the shift in pitch in the nuclear pitch accent is absent at the boundary. The general impression from these example contours is that there is a motivated upshift in register affecting only the scaling of the tonal targets of the nuclear pitch accent.

|  |
| --- |
|  |
| Figure . An illustration of register shift in nuclear PA which does not affect the PN stretch or the final boundary. |

One might be tempted simply to ignore the changes in register and transcribe all PAs as L\*H with only variation in the boundary, but this would ignore the salient scaling difference the changes in *f*0 to a register shift. Alternatively, one might choose to label them differently. As such, once might decide to interpret the MDQ H\* H%, while the MDQ with the boundary fall could be viewed as H\* HL%. However, there are several arguments against this.

Firstly, we can see from the illustration of the contours in Figure 7.3 that the nuclear pitch accent of each MDQ is essentially a raised version of the corresponding MDC, even though the beginning of the rise is slightly. It seems odd, therefore, that where the MDC would be L\*H %, the MDQ with a similar contour shape would be H\* H%. In fact, it requires that the H target at the end of each rise is ascribed to a different structural unit of the IP, the pitch accent in the case of MDC and the boundary in the case of MDQ, but there is no evidence for such an analysis. Contrary to the high boundary analysis, in these particular cases, the peaks of the MDQs are aligned slightly earlier their corresponding MDCs, which actually goes against expectations of a peak associated with the boundary rather than the PA.

Similarly, when the MDC is L\*H L%, its non-register-tier-analysis counterpart of MDQ H\* HL% would become, with the H peak again reassigned to the boundary. This again would also suggest a different underlying structure. It would also require a compound boundary, but this is, as proposed in section 7.1.1, not an unreasonable possibility. However, this does not account for the distinctiveness of the boundary fall with its large negative excursion, which suggests that there is more happening in at the boundary than the regular relative low of an L tone alone. Of course, to represent the size of the fall, one could suggest that the boundary is in fact a complex HLL%, but this looks like an overly complicated procrustean interpretation of the contour serving only to a desire to reject the possibility of a register tier while simultaneously using the boundary sequence to provide a more *phonetic* representation of the large fall.

Finally, the shapes of the *f*0 contours in the MDQ pitch accents do not look anything like prototypical low targets. Each contour is concave and has the appearance of rising out of a low target, features which are typical in the realisation of an L tone. In contrast, the H tone is prototypically associated with a convex shape (as in a peak) in the *f*0 contour, but there is no such evidence for this around the lexically stressed syllable. In fact, if there were a H\* tones in nuclear pitch accent, we might expect to see a sagging pitch contour between the pre-nuclear H\*s and the nuclear pitch accent. Rather, what we see quite different. After the PN H\*s, the *f*0 contour falls until the nucleus of following lexically stressed syllable before beginning to rise. All in all, it hard to interpret either MDQ as H\* H% or H\* HL%.

The most sensible reading of the example MDQ contours presented above is, I believe, to interpret each as L\*H with a raised register. That is, each should be understood as an instantiation of a register tier shift from an (unmarked) L register to the H. The IViE labelling conventions have been adapted to incorporate interpretation, as follows:

1. The caret symbol ^ is used indicate register shift
2. square brackets [] to indicate the scope of the register shift.

In other words, an upshift in the register tier is indicated by ^[...] in the labelling, while the normal (low) register is unlabelled. In this way, the four example contours in Figure 7.3 are represented symbolically as follows:

1. MDC 1: %0 H\* ^[L\*H] 0%
2. MDC 2: %0 H\* ^[L\*H] L%
3. MDQ 1: %0 H\* ^[L\*H] 0%
4. MDQ 2: %0 H\* ^[L\*H] L%.

This approach preserves the clear similarity in the contour shape across the four examples while reflecting the distinct rise in the nuclear PAs of the MDQs. Moreover, since it limits the scope of register shift to the PAs and not the boundary tones, it neatly captures the dramatic fall at the boundary of one of the MDQs. In cases where the whole IP is affected by high register, the scope of the effect can be indicated using square brackets. This is exemplified in the following labelling for the two IPs represented in Figure 7.2:

1. MDC: ^[% H\* L\*H %]
2. MDQ: ^[% H\* L\*H %]

One reasonable criticism of this approach is that it implicitly rejects the null hypothesis, yet, as outlined above, it was impossible to ignore the effects of register shift during labelling. Basically, the difficulties encounter even during the labelling process already began to provide evidence for expectation 4, that “patterns will occur which are adequately explained only with reference to both a register tier and the tonal tier” (section 7.1.1). That said, two strategies were adopted to avoid falling into the self-fulfilling hypothesis trap.

The first strategy was to generate a set of labels which excludes the register tier. This was done by replacing register-tier/PA combinations with alternatives which can be expressed adequately in terms of PA+boundary and then removing all register tier labelling from the data. In only two contexts, however, did relabelling seem reasonable, namely L\*^[H] % and L\*^[H] L%. These are reinterpreted as L\*H H% and L\*H HL% respectively. The other potential scenario for relabelling was in cases where an apparent upward register shift in the nuclear PA (as illustrated in the MDQs of Figure 7.3) might possibly be interpreted as H\* H% and H\* HL%. However, as noted above, this would represent a procrustean stretching of data which phonetically and phonologically still retain an L\*H (L)% quality. Admittedly, this will weaken the case for the non-register tier analysis, but it would have been dishonest to use the H\* H(L)% labels. Table 7.3 summarises register-tier labelling, the non-register-tier alternative, and the rejected alternatives.

The second strategy was the decision to analyse the phonetic parameters of each PA/register tier combination. This will help us assess the reliability and validity of labelling choices such as L\*H, ^[L\*H], allowing us to see if there does appear to be a consistent distinction between them.

|  |
| --- |
| Table . Differences between register-tier and non-register tier labelling used in phonological analysis. Rejected alternative non-register tier labels are also included. |
| |  |  |  | | --- | --- | --- | | Register-tier labelling | Non-register-tier alternative | Rejected alternative | | L\*^[H] 0% | L\*H H% |  | | L\*^[H] L% | L\*H HL% |  | | ^[L]\* H L% | L\*H L% |  | | ^[L \* H] 0% | L\*H 0% | H\* H% | | [L\* H] L% | L\*H L% | H\* HL% | | ^[L\* H L%] | L\*H L% | H\* HL% H\* HLL% | |

### Interpreting initial boundaries and prenuclear pitch accents

While this chapter focuses primarily on the nuclear pitch accent of the IP, it also considers utterance-wide phonology. Again, as with the previous chapter, prenuclear pitch accents which gave rise to the most difficulty, specifically the second pre-nuclear pitch accent of wh-questions. These difficulties are outlined below, primarily for the sake of transparency, but also because the issues raised by these more challenging contours will be considered again later in XXXXX.

In most cases, the second prenuclear pitch accent was unproblematic, especially in cases where it was clearly an L\*H or a L\*!H (Figure 7.4, panels a. and b. respectively). In other cases, however, there is a salient L on the stressed word in the second foot, yet, even though a slight rise may be visible in the pitch contour, it is so dampened that there is no H percept at all, either auditorily or visually (Figure 7.4, panel c.), so it feels more appropriate to label it L\* rather than (\*), the label used for a stressed syllable not associated with pitch accent (see 5.2) Figure 7.5 superimposes each contour on top of the other, illustrating the distinctive differences between each PN accent.

In terms of function, there is little apparent difference between the L\*H and the L\*!H, and the speaker appears to be packaging the semantic content of the utterance in to three units: who, valued, and vases. However, in the utterance with L\*, it gives the impression that the speaker is dividing the information differently, essentially into the question who, and its complement, valued the vases. This issue of will be taken up again in XXX-YYY.

|  |
| --- |
| a. % H\* L\*H L\*H % |
| b. % H\* L\*!H L\*H % |
| c. % H\* L\* L\*H % |
| Figure . Three repetitions of MWH1, “Who valued the vases?” with three different prenuclear pitch accents in the second foot. Panels a-c show each contour with pitch and CPP. | | |

|  |
| --- |
| Figure . The contours in Figure 7.4 superimposed on each other (time normalised). |

Another issue with the prenuclear PA revolves around the difficulty distinguishing between accentuation and lack thereof, especially in the second foot. In some cases, there may be a slight perturbation in *f*0, but it is not strong enough to trigger the percept of a phonological pitch event. This issue is illustrated in Figure 7.6, where each line represents an *f*0 contour from F15’s repetitions of MYN2, “Do you live in the valley?”

In Figure 7.6a., *f*0 begins relatively low in the speaker’s range and there is no pitch event until the nuclear pitch accent, despite a slight rise in *f*0 around the word “live”. Utterances b. and c. are very similar, each with an initial high boundary but no salient pitch accent until the nuclear pitch accent. Even though there are some bumps in *f*0 around “live”, these do not trigger a percept of prominence. Only in d. is there a salient pitch event, on “you”, which has been transcribed as H\*. While c. too appears to have a bump in *f*0 around “you”, it seemed auditorily less salient and so was not labelled with a pitch accent. One should also note the high stretch from the boundary does not extend uniformly across before beginning to fall towards the L of the nuclear PA. It extends further in c. than in b. While this is not a concern for the analysis in this chapter, it is discussed further in [STH REFs].

|  |
| --- |
|  |
| Figure . Contours for M-corpus repetitions of MYN2 by F15, “Do you live in the valley?” |

The final issue for annotation relates to whether a pitch accent in the second foot should be interpreted as L\* or !H\*. While these two pitch accents share several similarities, L\* is perceptually slightly different from !H\*; moreover, each is typically accompanied by a different contour shape. When the L\* occurs after a previous H, the L\* looks has a concave elbow (or scooped fall [REF]), with the most salient element of the elbow occurring in the stressed syllable. This can be seen in Figure 7.7a., where the concave elbow of the *f*0 contour occurs in the /a/ vowel of *val-*. In the case of !H\*, a convex elbow occurs in the stressed syllable, as illustrated in Figure 7.7b. Figure 7.8 compares both stylized contours and highlights the difference between the concavity of the L\* and the convexity of the !H\*.

|  |  |
| --- | --- |
| a. M8\_WH1\_1 f0 contour, labelled %H L\* L\*H % | b. M8\_MWH1\_5, labelled %H !H\* L\*H % |
| Figure .. Two repetitions of MWH1 (“Who valued the vases?”) from M5. Dark red lines indicate a slightly smooth contour with high voicing and amplitude settings. | |

|  |
| --- |
| /a/  Figure . f0 stylized contours from Figure 7.7 compared, with time normalised to utterance duration. The grey bar indicates the vowel in “val-”. |

As with all labelling issues, things are sometimes even less clear cut. Figure 7.9, for example, shows two contours where there is a distinct PA in the second foot. However, in each case, the concave elbow occurs before the stressed word and the convex elbow after the stressed syllable, making it more difficult to interpret. In each case, there is obviously some kind of downstep, but the question is: is this downstep from H\* to !H\* or is it an L\* after the previous H\*. In each case, one could even opt for !H\* as a compromise, but the L\* interpretation has been preferred.

The reasoning for the L\* choice goes as follows. Overall, the contour appears to curve concavely toward the stressed syllable, and the low continues throughout the lexical word (“live” and “valued” respectively). *f*0 then drops slightly lower again at the beginning of the verbal complement (“in the valley” and “the valuables” respectively), as if in anticipation of the L in the L\*H of the nuclear pitch accent rather than because of a delayed peak. This anticipatory lowering is in fact quite common and will be revisited in REFS STH.

|  |  |
| --- | --- |
| a. | b. |

Figure . Two pitch contours labelled as % H\* L\* L\*H % which might arguably be labelled with !H\* or >!H\* instead of L\* as the second pitch accent.

Of course, all the examples discussed above reflect edge cases in categorical judgments, and many other utterances do not pose such problems. It is possible that different labellers will make slightly different judgments.

## Phonological results and analysis

Statistical analysis of the relationship between mode and intonational phonology was carried out using both the non-register-tier analysis labelling and the register-tier analysis labelling. The results of the non-register tier analysis are presented first, followed by the register tier analysis. These are then compared, before moving on to looking at IP-wide phonology. Table 7.3 summarises the differences between the non-register tier and the register tier phonology.

### Non-register-tier analysis

Without the register tier, nearly all nuclear PAs are L\*H, accounting for 98.9% (n=632/639) of the raw data. When adjusted, it accounts for the same proportion (see Appendix XX for full details.) The only cases without L\*H as the nuclear pitch accent are H\* (n.=2) and >H\*(n.=5) are both produced by the same speaker, M8. In short, in the non-register-tier analysis, the nuclear pitch accent alone does not contribute to sentence mode, as expected.

When we look at the distribution of nuclear contours (pitch accent plus boundary), we see a more interesting distribution, as illustrated in Figure 7.11, which shows the adjusted distribution of nuclear contours by mode. We can see that occurrences of L\*H L% begin to increase in MYN and are most common in DCQ. As expected, L\*H H% is not found at all in MDC or MWH, and it is more common in MDQ than MYN (2.6% as opposed to 0.8%). There are a few rare occurrences of L\*H HL% in MDQ. We can therefore assume that it is really only the boundary condition that is associated with mode, again, much as expected from a non-register-tier analysis of intonational phonology.

|  |
| --- |
|  |
| Figure .. Proportional distribution of pitch contours by mode, non- register tier analysis. |

Figure 7.12 shows the boundary tone distribution by mode alone. It is almost identical to Figure 7.11, but the added *noise* from the >H\* and H\* contours is lost. When we further break this down to show the distribution of boundary conditions by mode and gender, it appears that there is also an effect of gender (Figure 7.13) That is, the male speakers tend to use L% more frequently in MDQ than female speakers (5.8% of the adjusted count versus 1.5%), and conversely, they are less likely than female speakers to use the unspecified % boundary (6.0% versus 8.1%). Female speakers, on the other hand, appear more likely to use the high boundary for YNQ and DCQ than men, accounting for 3.5% of the adjusted count for the female speakers as opposed to 0.8% for the male speakers. Therefore, we can speculate that gender has an effect on the choice of boundary tone.

|  |  |
| --- | --- |
|  | |
| Figure .. Proportional distribution of boundary tones by mode, non- tier analysis. |

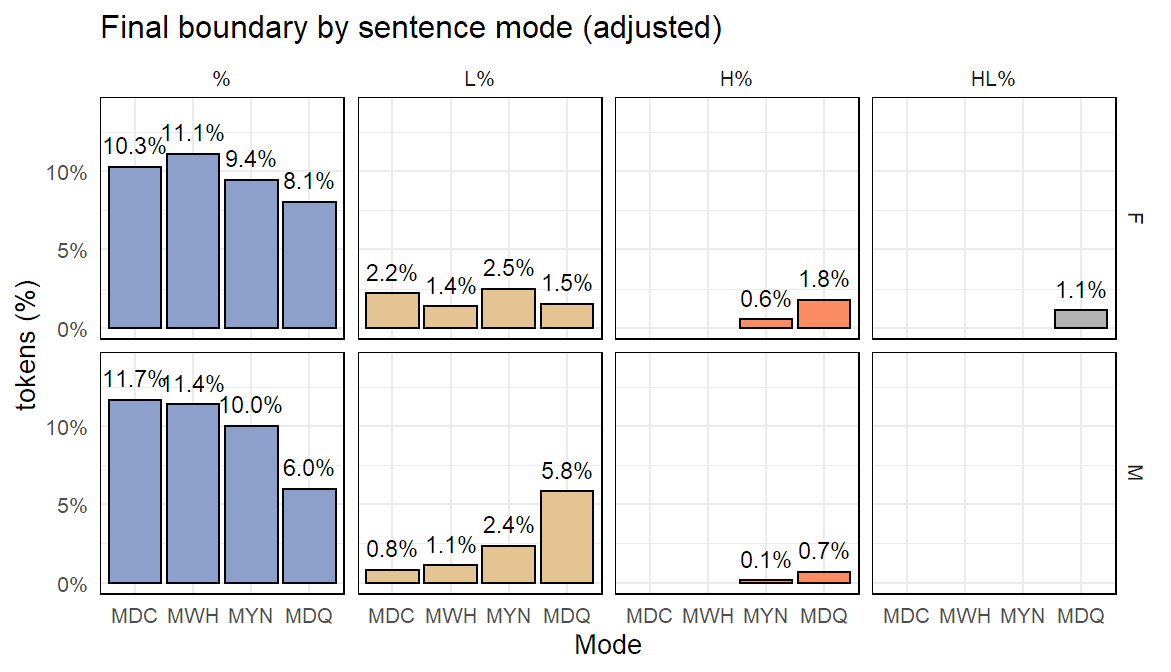


Figure .. Proportional distribution of boundary tones by mode and gender, non-register tier analysis.

A model was constructed to test the likelihood of an H%, with mode and gender as fixed factors and speaker and prompt as random intercepts. These models can be summarised as:

1. `is\_H\_boundary` ~ mode + gender + (1|speaker) + (1|prompt).

An ANOVA comparing this BGLM model a null model indicates that it is significant (χ2 (4) = 68.75, p < 0.0001). The model has a marginal r2 of 0.41 and condition r2 of 0.75. The high marginal r2 suggests that mode and gender account for 41% of the variance between % and H%, while the addition of the random speaker and prompt effects explains a further 34% of the variance.

Looking at the estimated intercepts of the log odds ratios for each level of mode in Figure 7.14a, it is clear that both MDC and MWH are exceedingly unlikely to be associated with H% (est. = ‑6.09 with upper CIs of ‑9.08 and ‑3.1, p.adj. < 0.001 for both). MYN is also unlikely to be H%, with an estimate of ‑4.23 and an upper CI of ‑2.06 (p.adj. < 0.001). While the log odds ratio for MDQ is lower than zero (‑2.027) the upper CI almost reaches zero (-0.12). (MDQ as intercept fails the test of significance once the value is adjusted (p.adj. = 0.1). As noted in 5.4.3, however, this simply means that we cannot claim that one outcome is more or less likely than the other, i.e., the true probability may lie around 50%).

When looking at the pairwise comparison of slopes (Figure 7.14b), we see that only MDQ is significantly more likely to have a H% when compared to the other levels of mode, having an estimated slope of 4.14 against MDC as intercept, 4.15 against with MWH, and 2.28 against MYN. The slope of GenderM has an estimated log odds ratio of -2.04 but an upper CI which extends above zero (CIs = -4.97—0.90), and so gender fails the test of significance (p.adj. = 0.21). That is, we cannot reject the possibility that gender has no effect on the choice H%.

|  |  |
| --- | --- |
| a. Intercepts for each level of mode. | b. pairwise comparison across slopes. |

Figure .. Graphical summary of intercepts and slopes of pairwise comparisons across levels of mode and gender for likelihood H% in the non-register tier analysis.

Considered in terms of probability (Figure 7.13), we see that the probability of H% in either MDC or MWH is almost zero (CIs = 0%—3% and 0%—2.6% respectively). The estimated probability of H% is 2%, (CIs = 0%—11.8%, while the estimate for MDQ is 13% (CIs = 2%—49.8%). All of this indicates that, while the probability of H% does increase in MYN and MDQ, it is still very low.

A picture containing graphical user interface

Description automatically generated

Figure .. Predicted probability of H% as an effect of mode in the non-register-tier analysis.

### Register-tier-analysis

Looking at the nuclear contour in the register-tier analysis labelling—shown in Table 7.4—we see a large range of possibilities. L\*H % still dominates, even after the inclusion of the register tier, with 419/639 tokens. This is followed its raised register counterpart, ^[L\*H] % with 81 tokens, and then by L\*H L% (59 tokens).

|  |
| --- |
| Table .. Nuclear contour / Register Tier tokens in M-Corpus |
| |  |  | | --- | --- | | nuclear contour | count | | L\*H % | 419 | | ^[L\*H] % | 81 | | L\*H L% | 59 | | L\*^[H] % | 22 | | ^[L\*H] L% | 19 | | ^[L\*H L%] | 18 | | ^[L\*]H L% | 6 | | L\*^[H L%] | 6 | | >H\* L% | 5 | | H\* L% | 2 | | L\*^[H] L% | 2 | | **Total** | **639** | |  |  | |

Seven tokens which not have an L\*H-like nuclear pitch accent. As we know from the non-register-tier analysis, these are >H\* L% (5 tokens), and H\* L% (2 tokens). These are the only tokens which resemble the falling nuclear contour of standard Southern British English and General American English. However, of the >H\* L% tokens, only two of them occur in declaratives, while the other three occur in either MYN (n.=2) or MDQ (n.=1). This is quite surprising and reinforces the idea that fall to the low boundary does not serve the same function in DCE as in standard varieties, or at least doesn’t serve to suggest finality in the same way. It does reinforce the view that L% serves a discourse function which is compatible with question forms but does not in itself signal a question. For this reason, the rest of this section will treat accent phonology separately from the boundary tone.

The M-corpus contains six different nuclear pitch accent/register tier combinations: L\*H, L\*^[H], ^[L\*]H, ^[L\*H], H\*, and >H\*. Table 7.5 presents the distribution of these pitch accents by mode (adjusted data). The vast majority of nuclear pitch accents are L\*H (n.adj.=506), with the next most common, ^[L\*H] having considerably fewer tokens (n.adj.=112). This is followed by L\*^[H] (raised H target only) (n.adj.=30). Three tokens are rarely attested, namely H\* (n=2), >H\* (n=5), and ^[L\*]H (n.=6). Aside the very rare tokens, MDC and MWH are otherwise exclusively L\*H (n.adj.=162 and 165 respectively). This falls to 117 for MYN and 62 for MDQ. Conversely, instances of L\*^[H] and ^[L\*H] appear in MYN (n.adj.=5 and 38 respectively), and then rise again for MDQ, with 25 tokens (adj.) for L\*^[H] and 74 for MDQ.

|  |  |
| --- | --- |
| Table .. Distribution of pitch accent by mode (adjusted). |  |
| Graphical user interface, table  Description automatically generated |  |

There is also considerable interspeaker variation in the use of pitch accents, which Figure 7.15 demonstrates. The three rare pitch accents, H\*, >H\*, and ^[L\*]H, it can be seen, were all produced by a single speaker, M8. Some speakers used only two different pitch accents. That is, M5, M9, and M10 used only L\*H and ^[L\*H]. More extremely, F16 used L\*H almost exclusively, with only one instance of L\*^[H]. The remaining speakers employed the three more common pitch accents to varying degrees. Superficially, there appears to be a gendered difference in PA production, as M4 is also the only male speaker who used L\*^[H]. However, given that three of the six female speakers (F12, F15, and F16) also used this token, it is difficult to interpret its use as a female trend.

Three speakers used raised register exclusively for MDQ. F12 used L\*^[H] once and ^[L\*H] the remaining 14 times, while M4 used L\*^[H] four times and ^[L\*H] the remaining eight times. M5 produced ^[L\*H] exclusively. In contrast to this, F16 only used raised register once across all IPs. The general impression, however, is that high register is associated with YNQs and DCQs, but that it is optional rather than obligatory.

|  |
| --- |
| Chart, box and whisker chart  Description automatically generated |
| Figure .. Proportional distribution of pitch accents by speaker and mode (adjusted data). Percentages refer to the individual speakers (listed on the right y-axis). |

The relationship between mode and high register was tested using the model in (eq. 3), with mode and gender fixed factors, and speaker and prompt intercepts as a random factor. The model has a marginal r2 of 0.69 and condition r2 of 0.79, and in an ANOVA comparing it against the null model, the results were significant (χ2 (5) = 321.66, p < 0.0001).

1. `high register in nucleus` ~ mode + gender + (1 | speaker) + (1 | prompt)

|  |  |
| --- | --- |
| A picture containing table  Description automatically generated  a. Intercepts for each level of mode. | Diagram  Description automatically generated with medium confidence  b. pairwise comparison across slopes. |

Figure .. Graphical summary of intercepts and slopes of pairwise comparisons across levels of mode and gender for likelihood of high register in register tier analysis.

As is very clear from Figure 7.16a, the estimated log odds ratios for high register were exceedingly low for both MDC at ‑6.39 and MWH at ‑6.38 (CIs = ‑8.51—‑4.27 and ‑8.50—-4.26 respectively). The estimate was higher for MYN (‑2.20) but with CIs still below zero (‑3.64—‑0.95). The log odds ratio for MDQ is almost zero (‑0.07, CIs = -1.22—1.08), indicating that the probability of high register in MDQ is roughly 50%.

When taking the slopes between each level of mode into account (Figure 7.16a), we see that MYN and MDQ are both more likely be associated with high register in the nuclear pitch accent when compared with either MDC or MWH (the CIs also do not fall below zero, and they are all statistically significant). The likelihood of H% in MWH when compared to MDC seems much less likely (est.=-1.71), but it has quite large CIs (-6.037—2.627), with the upper CI well above zero, rendering it statistically non-significant (p = 0.5).

The slope of gender-male against the intercept is 1.71 with a low CI just above zero (CIs = 0.12—3.28), but the adjusted p value fails to reach significance (p.adj. = 0.051). However, it is important to remember that there are no instances of high register in either MDC or MWH, which means that the effect of gender on high register incorporates two levels of mode where we already know that high register does not occur, which might have skewed the results. Therefore, it is worth considering the predicted probabilities of high register as a function both of gender and mode. This is shown in Figure 7.17 and Table 7.7.



Figure .. Predicted probability of high register in the nuclear pitch accent by gender across sentence modes.

As expected, we see that the probability of high register in both MDC and MWH is around zero both for male and female speakers but increases first for MYN and the again for MDQ. In fact, the probability of high register is noticeably higher for male speakers in both MYN and MDQ. The predicted probability of high register for female speakers is 11% for MYN (CIs = 3%—28%) and 50% for MDQ (24%—76%). For male speakers, however, there is a predicted probability of 39% for MYN (CIs = 15%—69%), and 85% for MDQ (60%—95%). Again, MYNs are less likely to be associated with high register. For the MDQ, both the lower and upper CIs are above 50%, indicating that high register in fact much more likely in MDQ than not for male speakers.

Table .. Predicted probability of high register in the nuclear pitch accent by gender for MYN and MDQ.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **group** | **mode** | **predicted** | **conf. low** | **conf. high** | **std. error** |
| F | MYN | 11% | 3% | 28% | 0.6217 |
| F | MDQ | 50% | 24% | 76% | 0.5944 |
| M | MYN | 39% | 15% | 69% | 0.6409 |
| M | MDQ | 85% | 60% | 95% | 0.6591 |
|  |  |  |  |  |  |

#### Boundary Tones (register tier analysis)

There are only two boundary conditions in the register tier analysis, L% and the unspecified boundary (%). The low boundary tone may be affected by high register, e.g., ^[L\*H L%] as opposed to ^[L\*H] L%. There are 24 tokens where L% is affected by high register and 21 where it is not. However, raised register does not (and cannot) occur at the boundary alone.[[10]](#footnote-11) Since raised register in the nuclear contour is also already exclusively associated with YNQ and DCQ, and since raised register in the boundary is parasitic on raised register preceding the boundary, the inclusion of ^[L%] in the analysis will be uninformative. Therefore, the high register L% and the low register L% have been collapsed into a single category for statistical analysis.

Figure 7.18 shows the (adjusted) distribution of boundaries by gender and mode for the register-tier analysis. The distribution here is almost identical to the non-register tier analysis, given that the only difference is the absence H%. Among female speakers, there is a very even distribution of L% and % boundaries across modes, with L% accounting for 17.4% of the female utteracnes and % for 82.6. L% is, however,slightly less common in MWH. Among the male speakers, there is a gradual increase in the use of L%, from 1.6% in MDC and 2.2% in MWH to 4.6% in MYN and 11.6% in MDQ, with L% accounting for 17.4% of the boundary tones in the male data. Despite the steady increase in use, L% still only accounts for 23.2% of boundaries in MDQ.

|  |
| --- |
| Chart  Description automatically generated |
| Figure .. Final boundary by mode and gender (adjusted data, register tier analysis). High and low register have been collapsed into a single category. |

To test the effects of mode and gender on the boundary in the register-tier analysis, a BGLMM analysis of the model in (eq. 4) was carried out.

1. L% boundary ~ mode + gender + (1 | speaker) + (1 | prompt)

An ANOVA comparing this model against the null model was significant (χ2 (4)=40.03, p.= 4.3×10-8). The marginal r2 was 0.05 and the conditional r2 0.77.

The results of the statistical analysis reflect the visual representation of the adjusted counts. Predicted estimates indicate that L% is very unlikely in either MDC and MWH, with a log odds ratio of ‑3.53 for MDC (CIs = ‑6.46—‑0.60) and ‑3.94 for MWH (CIs = ‑6.88—‑1.00). MYN is more likely to have L%, with a log odds ratio of ‑2.923 (CIs = ‑5.84—‑0.01, p.adj. = 0.08). MDQ is most likely to have an L%, with a log odds ratio estimate of -1.728 (CIs = ‑4.62—1.17), with an upper CI reaching into positive a log odds ratio.

Looking at the estimated slopes, there is little difference between MDC and MWH, with a log odds ratio just below zero (‑9.42, CIs = ‑1.21—0.38), suggesting L% is slightly less likely in MWH than MDC but with CIs indicating that this is not statistically significant. Conversely MYN is slightly more likely to have L% than MDC (est.=0.60, CIs = ‑0.13—1.33), but again the CIs indicate that this is not statistically significant. MDQ is always more likely to have L% than the other levels of mode, with a lower CIs always above zero, indicating that this is a statistically significant trend. MYN is also statistically more likely to have an L% than MWH (est.=1.02, CIs = (0.2—1.79). Finally, the effect of gender (slope of Male re Female as intercept) is almost zero, and the exceedingly wide CIs suggest that there is in fact no meaningful effect of gender at all and that differences which appeared to be gender-specific in the visual analysis are really speaker-specific (est.=-0.19, CIs = ‑3.83—3.45).

|  |  |
| --- | --- |
| a. Intercepts for each level of mode. | b. pairwise comparison across slopes. |

Figure .. Graphical summary of intercepts and slopes of pairwise comparisons across levels of mode and gender for likelihood of L% in register tier analysis.

Even though MDQ is more likely to be associated with L% than the other levels of mode, the probability of L% is still very low across all modes, and the fixed effects of the model (mode and gender) only account for 5% (r2m = ‑0.05) of the variance in L%. This means that while mode does influence the choice of the L%, its contribution is rather small (and we can dismiss gender given the proximity of the estimate to zero and the exceedingly large CIs).

To reinforce the limited effect of the model’s fixed factors, it is worth reinterpreting the log odds ratios of the intercepts in terms of probabilities. Figure 7.20 shows the predicted probabilities of L% by mode. Here, the highest predicted probability of L% (MDQ) still only reaches 15%. The lower CIs are all below 1%, and the higher CIs are quite large, even for MDC (36%) and MWH (27%). This contrasts quite dramatically with the predicted probabilities of high register (Figure 7.17), in which MYN and MDQ were clearly associated with increased probabilities of high register.

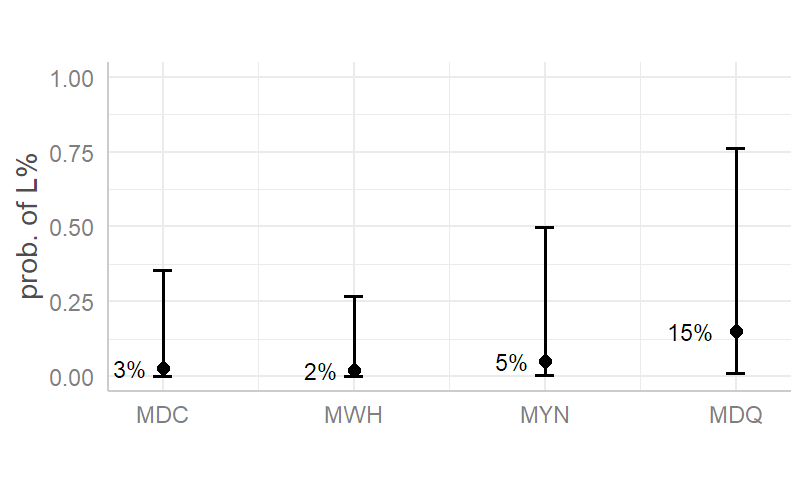


Figure .. Predicted probabilities of L% in a register-tier analysis of the M-corpus.

#### Summary

High register is extremely unlikely to occur in the nuclear pitch accent of either MDC or MWH. However, it is somewhat more likely to occur in MYN, and even more so in MDQ. There is a strong effect of gender on the use of high register in MYN and MDQ, with male speakers much more likely to use high register, especially in MDQ. There is a small effect of mode on the likelihood of L%, but it does not appear to be the main contributing factor.

### Comparing the non-register tier analysis and the register tier analysis

The fixed factors—speaker and gender—explain 42% of the variance in the non-register tier model analysis of H% as a marker of interrogativity, while they explain 69% of the variance in the register tier analysis. To assess the effect of speaker alone, each model was retested using gender as a random effect, but this created convergence issues, so gender was simply removed as a factor. In the retested models (see appendix XXX), the r2 values were as follows:

Table .. Marginal and conditional r-squared of models for H% in non-register tier analysis and high register in register tier analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **analysis** | **model** | **r2m** | **r2c** |
| Non-register tier | is\_H\_boundary ~ mode + (1 | speaker) + (1 | prompt) | 0.33 | 0.74 |
| Register tier | high\_reg\_in\_nuc ~ mode + (1 | speaker) + (1 | prompt) | 0.63 | 0.8 |
|  |  |  |  |

These results indicate that mode accounts for 33% of the variance in the non-register-tier H% model but for 63% of the variance in the register-tier analysis model.

Of course, the non-register this analysis is bound to be less informative since the only L\*^[H] (L)% was relabelled with a H boundary, i.e., it was interpreted as L\*H H(L)%. Other instances of L\*H with high register were excluded from relabelling with H(L)% before the fact because—in terms of contour shape and percept—they were still clearly L\*H. Had these ^[L\*H] (L)% contours been relabelled as H\* H(L)%, the results for the non-register tier analysis would have been identical. However, as already previously observed (6.4), this would have required an improbable re-association of tones with difference structural elements of the IP and represented a description of the data in service to the theory more than a true representation of the data.

The comparison of the non-register with the register tier analysis, therefore, does not prove the register tier hypothesis. Rather, it demonstrates that a register-tier analysis provides a greater degree of explanation of the data (63% as opposed to 33%) without compromising the descriptive integrity of the data. For the subsequent analysis of IP-wide phonology, therefore, the register tier analysis is maintained.

### Utterance-wide Phonology and Mode

Nearly all IPs in the M-corpus end with a nuclear L\*H pitch accent (with or without register shifts). However, there is a wide variety of IP-wide patterns. Table 7.9 shows the total number of tokens for IP-wide phonology accounting for at least 1.25% of the corpus (see REF for the full list). Overall, the most common IP-level intonation pattern is % L\*H % (n=78), which indicates no pre-nuclear accentuation. In fact, 23.3% (n=147) of all IPs contain no pre-nuclear accents.

Table . IP-level intonation by mode for tokens accounting for at least 1.25% of the M-corpus (raw counts).

Table

Description automatically generated

Table 7.10 summarises the raw counts for nuclear-pitch-accent-only utterances. Both MYNs and MDQs have the highest rate of nuclear-pitch-accent only utterances, at 34.6% and 35.3% respectively, while MDCs are around the average, at 22.7%. There is only one uttera where MWH lacks pre-nuclear accentuation. It is unsurprising that all but one MWH utterances have pre-nuclear accentuation since the wh-word at the front of the sentence represents the focus of the question, and thus one expects it to have a pitch accent. The tendency towards a greater lack of accentuation in MYN and MDQ tokens most likely reflects the fact that by de-accentuating the pre-nuclear content of the IP, greater salience is lent to the rise in the nucleus.

Table . IP level intonational phonology by mode for tokens with nuclear pitch accent only

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **MDC** | **MWH** | **MYN** | **MDQ** | **Total** |
| Count | 36 / 163 | 1 / 161 | 56 / 162 | 54 / 153 | 147 / 639 |
| percentage | 22.1% | 0.6% | 34.6% | 35.3% | 23.0% |
|  |  |  |  |  |  |

It appears, therefore, that speakers may sometimes use a strategy whereby they lend more salience to the nuclear rise in YNQ and DCQ forms by avoiding pre-nuclear accentuation. To test this, a model was constructed to assess the likelihood of nuclear-PA-only IPs as a function of mode and gender, given that gender was sometimes a meaningful factor in previous analyses:

1. nuc\_PA\_only ~ mode + gender + (1 | speaker) + (1 | prompt)

In an ANOVA, the model is significant when compared with the null model (χ2 (4)=120.3, p < 0.0001). It also has a marginal r2 of 0.44 and condition r2 of 0.77 (see app. Ref.)

There is a significant effect of gender in this model. The log odds ratio of the gender-Male slope is -2.53 (CIs = ‑4.642—‑0.41, p.adj. = 0.028), indicating that male speakers are significantly less likely than female speakers to use a nuclear-PA-only IP. The results of the pairwise comparison of slopes indicate that there is a significant difference between all levels of mode except for MYN and MDQ, between which there is almost difference (see Figure 7.21). This makes sense since there is little communicative pressure to use different intonational strategies to distinguish between them. However, when we consider the predicted probabilities of a nuclear-PA-only IP in terms of both mode and gender, we see a different picture.

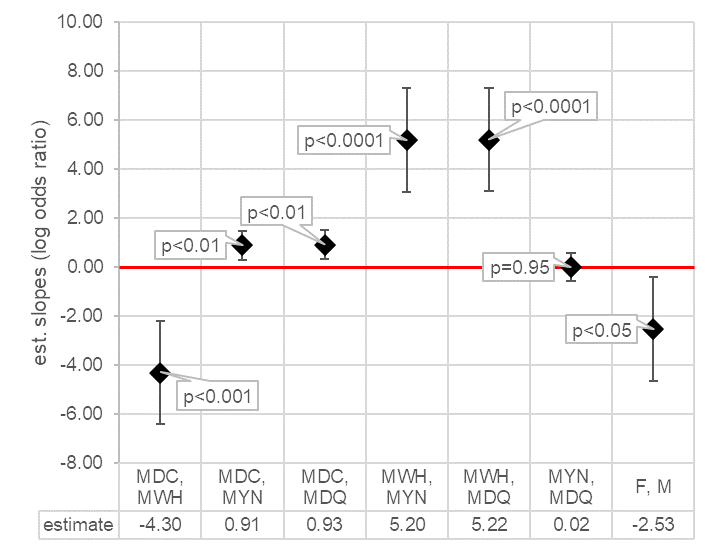


Figure .. Graphical summary of slopes of pairwise comparisons across levels of mode and gender for likelihood of nuclear-PA-only in IP in the M-corpus.

Figure 7.22 shows the predicted probabilities of a nuclear-PA-only IP as a function both of mode and gender. Here, we see the effect of gender more precisely. Both male and female speakers are very unlikely to have only one PA in MWH (less than 1%). Each is more likely to have a single PA in MDC and most likely to have only a single PA in MYN and MDQ. However, the increased probability of a nuclear-PA-only IP among females is much more dramatic, especially in MYN and MDQ (roughly 48.5% in each case, CIs = 10%—90%). At the same time, however, the CIs are also very large, reflecting a wide degree of variability.

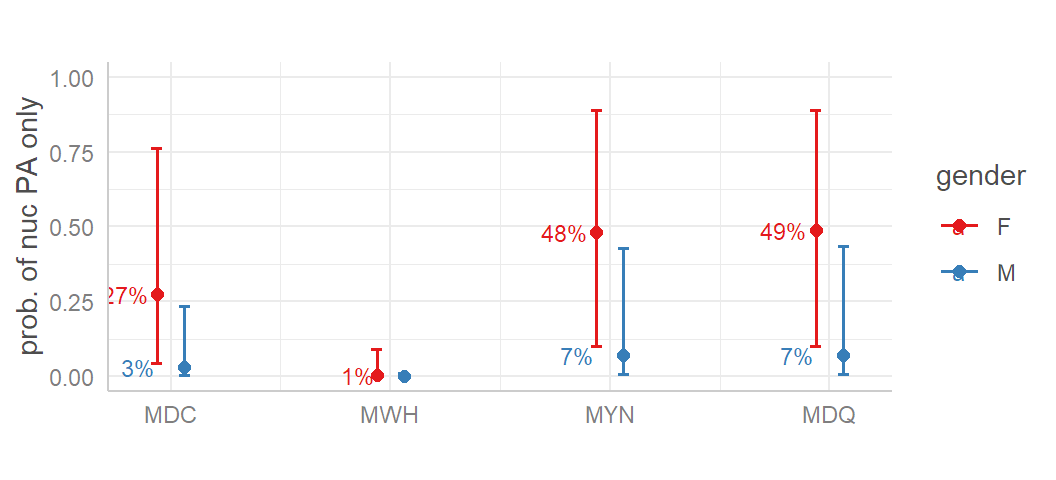


Figure .. Predicted probabilities of nuclear-PA-only IP by mode and gender in the M-corpus.

### Combining IP-wide and nuclear PA strategies

Females appear more likely to use the nuclear-PA-only strategy while males are more likely to use the high-register strategy to distinguish MYN and MDQ from other modes (see 7.5.2). While the choice of strategy is most likely gendered, it still appears that two phonological strategies are available to help identify question types and to distinguish MDC from MDQ. A raw count of instances of either strategy per mode (Table 7.10) indicates more than 50% use of at least one of these strategies in MYN (56.2%) and MDQ (79.7%), but both are practically absent in MWH (0.6%) and not even very common in MDC (22.1%)

|  |
| --- |
| Table . IP level intonational phonology by mode for tokens with either nuclear PA only or high register in the nuclear PA. |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | **MDC** | **MWH** | **MYN** | **MDQ** | **Total** | | Count | 36 / 163 | 1 / 161 | 91 / 162 | 122 / 153 | 250 / 639 | | percentage | 22.1% | 0.6% | 56.2% | 79.7% | 39.10% | |

The effect of gender and mode on the likelihood of the using of one of these strategies was tested using a BGLMM analysis of the following model:

1. nuc\_PA\_only\_OR\_H\_reg ~ mode + gender + (1 | speaker) + (1 | prompt)

An ANOVA of the model comparing it with the null model indicates that it is significant (χ2 (4)=339.67, p < 0.0001). It has a marginal r2 of 0.57 and condition r2 of 0.78.

In this model, gender now basically has no effect, as can be seen in the Gender-Male log odds ratio of almost zero (‑0.06) and narrow CIs (‑1.66—1.15), shown clear in Figure 7.23. This strongly suggests that the gendered effects of the two different strategies are neutralised when considered together. Moreover, in this mode, there is a significant difference in slope between every level of mode (p < 0.001).

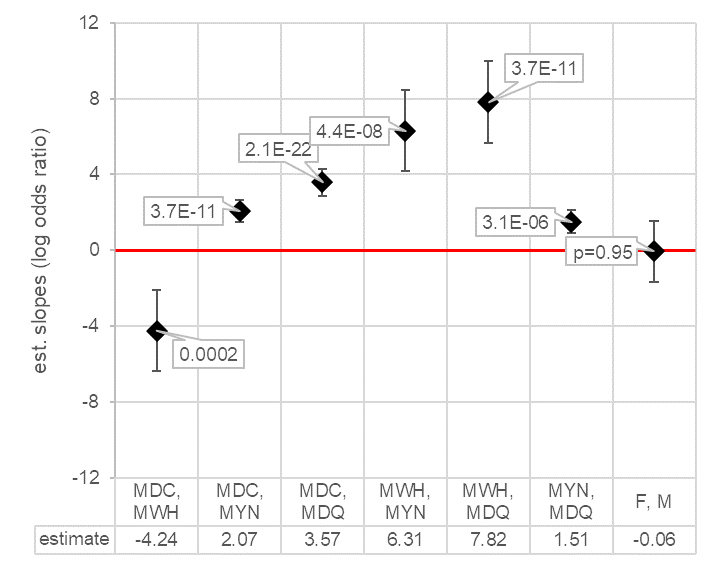


Figure .. Graphical summary of slopes of pairwise comparisons across levels of mode and gender for likelihood of nuclear-PA-only in the IP or raised register in the nuclear PA in the M-corpus.

Looking at the predicted probabilities of either strategy across modes (Figure 7.24), we see that there is only 0.3% probability of at least one of the two strategies occurring in MWH (CIs = 0.02%—4.1%). This increases somewhat for MDC to 15.6% (CIs = 2.8%—54.4%), and then against for MYN, at 59.4% (CIs = 19%—90.1%). The predicted probability for MDQ is highest at 86.8% (CIs = 50.7%—97.7%). While the CIs are still quite large (after all, we have collapsed two categories into one), we can see that there is a large different between MDC and MDQ especially. As these are the two categories for which there is the most communicative pressure to distinguish from each other, it is quite gratifying to see that speakers are likely to use phonological strategies to distinguish between the two.

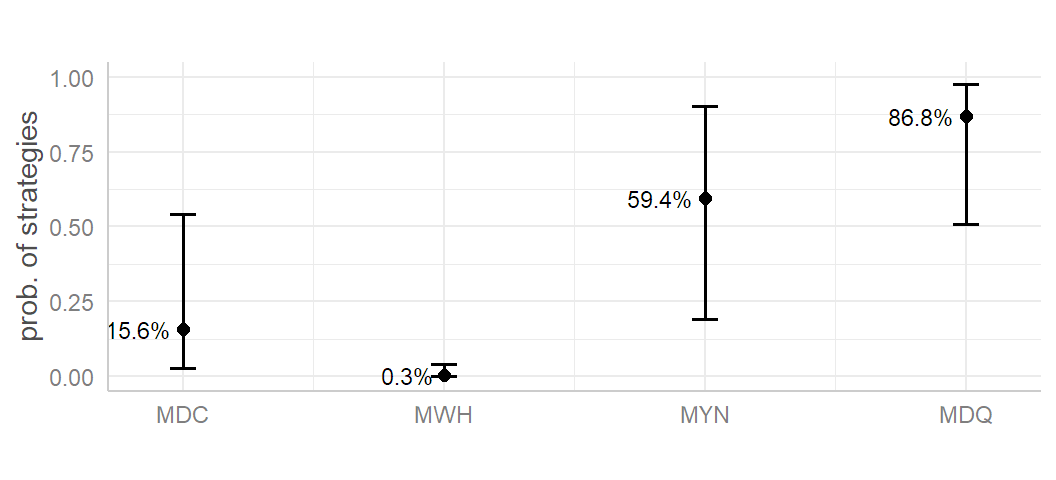


Figure .. Predicted probabilities of nuclear-PA-only IP or raised register in the IP in the M-corpus

### Summary

Nuclear contours and pitch accents were evaluated using both a non-register tier and a register-tier analysis. In the non-register tier analysis, H% was highly unlikely in MDC and MWH, while increasingly more likely in MYN and MDQ. In the register-tier analysis, high register was more likely to indicate a distinction between MDC and MDQ, with a very high likelihood of high register in MDQ for male speakers (85%). The register-tier analysis also demonstrated better explanatory value for phonological variation across modes than the non-register-tier analysis (r2m of 0.63 versus 0.33).

In the utterance-wide analysis, there were a large number of nuclear-PA-only IPs, which was unexpected. This may be a strategy which helps make the rise of the nuclear pitch accent more salient, thus reinforcing its function as a question.

It appears, therefore, that speakers have two phonological strategies available to help reinforce interrogativity in YNQs and DCQs, namely, register raising in the pitch accent or non-accentuation before the pitch accent. In assessing the likelihood that speakers use at least one of the of these two strategies to distinguish between modes, the predicted possibility for MDCs was 15.6%, while for MDQs it was 86.8%. Thus, in a register-tier analysis of the intonational phonology, we can see that speakers are likely to employ one of phonological strategies to distinguish between the two modes which are otherwise syntactically and semantically identical.

## Phonetic results and analysis

Four phonetic parameters were analysed. These were the *f*0 and temporal alignment at the *f*0 minimum in the nuclear PA and the *f*0 and temporal alignment of the *f*0 maximum, as summarised in Table 7.3. The combined alignment and *f*0 parameters were treated as the phonetic implementation of tonal targets.

Table . Phonetic parameters used in analysis of nuclear accent phonology.

|  |  |  |
| --- | --- | --- |
| **Parmeter type** | **Parameter code** | **Parameter description** |
| *f*0 parameters (ST re 1 Hz) | l\_f0 | *f*0 minimum (at L target) |
| h\_f0 | *f*0 maximum (at H target) |
| Time parameters (ms) | l\_t | Temporal alignment of L target re onset of vowel in stressed syllable |
| h\_t | Temporal alignment of H target re onset of vowel in stressed syllable |
|  |  |  |

For the analysis of utterance-wide effects, two global parameters were considered. The first is the utterance-mean *f*0 (utt\_mean\_f0) measured in ST re 1 Hz[[11]](#footnote-12). The second is the slope of the *f*0 contour(utt\_slope), which represents the slope of linear regression of *f*0(t) from the onset to the offset of voicing in the IP. Slope is measured in ST/sec.

Because the focus here is on paralinguistic and phonological changes to the rise, only utterances with L\*H-like PAs are. That is, the rare instances of H\* (n = 2) and >H\* (n = 5) are excluded, leaving L\*H with four different register tier patterns:

1. L\*H a nuclear rise with unmarked low register in the pitch accent.
2. ^[L\*]H a nuclear rise, where high register appears only to affect the low tone.
3. L\*^[H] a nuclear rise where high register only affects the H target, creating a large rise from a low *f*0.
4. ^[L\*H] a nuclear rise where the whole pitch accent is affected by high register.

Of these four, ^[L\*]H is the most contentious. It is only attested six times, all from the same speaker, and it is unclear might be there might be any meaningful—as opposed to purely formal—phonological distinction between the unmarked L\*H and ^[L]\*H, where high register tier does not affect the H target. However, as it was analysed as an L\*H-like pitch accent, it was retained.

### Mode and phonetic parameters of nuclear pitch accents

Two kinds of model were generated to assess the effects of mode on the contour in the nuclear pitch accent. The first type is a mode-only model, which treats mode (mode) as the sole fixed factor affecting the scaling and timing of tonal targets. This represents the effect of mode as an independent factor unmediated by phonological processes. In other words, it assumes that differences in the timing and scaling of tonal targets must be purely paralinguistic. The send type is a mode plus phonology model, which treats both accent phonology (acc\_phon) and mode as fixed factors affecting, i.e., it assumes that phonological processes also affect the alignment and scaling of *f*0. Since the choice of accent phonology may be associated with sentence mode, a mode\*acc\_phon interaction is expected in this type of model. If the register tier hypothesis is true, we should expect to see the apparent paralinguistic effects of mode in the mode-only models which largely disappear in the mode-plus-phonology models.

For each type of models, per-speaker random slopes and intercepts of mode (and acc\_phon) were included in the ideal maximal model. Even though fin\_phon has a finite set of levels, it was included as a random intercept since it is not a factor of interest but may still affect the target parameter. The target phrase (prompt) was also treated as a random intercept, given that each is essentially one of an infinite range of potential phrases. Gender was not of interest here and was excluded as a random factor since gender is already nested in speaker and will thus already be included in random speaker effects. The ideal maximal models are shown in Table 7.11.

Table .. Maximal models for LME analysis of mode of nuclear PAs, where x is the target parameter.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **factor(s) of  interest** | **random by-speaker factors** | **common random factors** |
| x ~ | mode + | (1 + mode | speaker) + | (1 | prompt) + (1 | fin\_phon) |
| mode\*acc\_phon + | (1 + mode + acc\_phon | speaker) + |

It was not possible to get these models to work on all four tonal target parameters, and, in the end, random intercepts-intercepts only models were the only ones which did not cause convergence or singularity issues. Prompt also caused singularity issues for some *f*0 parameters, so it was removed from all *f*0 models. The final working models are those shown in Table 7.12:

Table .. optimal working models for LME analysis of nuclear PA tonal targets.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **factor(s) of interest** | **random factors** | **shared random factor** |
| l\_f0 ~ | mode | + (1 | speaker) | + (1 |fin\_phon) |
| mode\*acc\_phon |
| h\_f0 ~ | mode |
| mode\*acc\_phon |
| l\_t ~ | mode | + (1 | speaker) + (1 | prompt) |
| mode\*acc\_phon |
| h\_t ~ | mode |
| mode\*acc\_phon |
|  |  |  |  |

Note that for each target parameter, the only difference between each type of parameter in the factor(s) of interest. This ensures that the results of each model can be fairly compared across similar parameters in the two different types of models. Since it caused singularity issues in some—but not all—models, prompt is excluded from the *f*0 models. Moreover, in the models where prompt did not cause a singularity issue (e.g., in the mode-plus-phonology model for h\_f0), step()estimated had it had a non-significant effect on the model when compared with a simpler model which excluded it, (LRT(1) = 0.47, p = 0.49). The negligible effect of prompt as a random intercept is illustrated in Figure 7.23 below, which shows the estimate means and CIs of each level of prompt (identified by the final word in the utterance).

Chart, scatter chart

Description automatically generated

Figure .. Random effect of prompt in the model h\_f0 ~ mode \* acc\_phon + (1 | speaker) + (1 | fin\_phon) + (1 | prompt). The estimated effects almost zero.

#### Mode-only models

ANOVAs for each model indicate that mode is a significant factor at a level of p < 0.0001 (see appendix XXX). The marginal r2 of the *f*0 mode-only models were 0.04 and 0.10 for l\_f0 and h\_f0 respectively, indicating that mode accounts for 4% of the variation in l\_f0 and 10% in h\_f0. However, we expect much of variance in *f*0 is an effect of by inter-speaker differences, within which gender differences are nested. Therefore, it is unsurprising that the conditional r-squared both for l\_f0 and h\_f0 is much higher than the marginal r-squared, at 0.93 and 0.88 respectively. The higher r2m for h\_f0 than l\_f0 may reflect the fact that the height of the H target is more important in the expression of mode than the height of the L targets. Conversely, r2m of l\_t is higher than that of h\_t, (0.05 vs 0.01) although each of these is very low, especially given that we do not expect inter-speaker variation to be as large for alignment as for *f0* height. The r2c for the time parameters is lower than for the *f*0 parameters, with l\_t at 0.77 and h\_t at 0.75. This suggests that the variance in h\_t is better explained by the whole model than variance in l\_t.

|  |
| --- |
| Table .. Marginal and Conditional r-squared of mode-only models for each target parameter. |
| |  |  |  | | --- | --- | --- | | parameter | R2m | R2c | | l\_f0 | 0.04 | 0.93 | | h\_f0 | 0.10 | 0.88 | | l\_t | 0.05 | 0.77 | | h\_t | 0.01 | 0.85 | |  |  |  | |

Mean estimates were calculated from the models using each level of mode as intercept. These are summarised in Figure 7.25. All estimated means were significant at a level of p < 0.01. In terms of timing (panel a.), there is very little difference between MDC, MWH, and MYN, with estimated low mean timing ranging between 64 and 66 ms and estimated high timing between 265 and 267 ms. MDQ is noticeably earlier, with the low mean aligned at 45 ms and the high at 250. This can be seen clearly in Figure 7.26, panels a. and b., which shows the pairwise comparisons for each level of mode. MDQ slopes and CIs for both l\_t and h\_t are all noticeably lower than zero, as indicated by the third, fifth and sixth markers. The other slopes are at or close to zero, with CIs crossing the zero boundary. Thus, only MDQ slopes l\_t and h\_t achieve significance (p < 0.0001, see appendix XXX).

|  |  |
| --- | --- |
| a. Estimated means for temporal alignment | b.Estimated means for f0 targets |

Figure .. Estimated mean tonal target parameters and CIs for mode-only models.

The results for the *f*0 parameters are slightly different (Figure 7.25b.). There is little difference between estimated means of the low and high *f*0 targets for MDC and MWH, with an estimated low for MWH of 84.4 ST, only 0.1 STs above the MDC estimate. The estimated mean for the high target *f*0 of MWH is 90.5, only 0.4 STs above the MDC estimate. Both the low and high *f*0 estimates for MYN are noticeably higher, at 85.9 and 91.9 ST respectively, ranging from 1.3 to 1.7 ST higher than MWH and MDC. MDQ is even higher again, with a low *f*0 estimate of 86.7 and a high of 94.6 ST. The high estimate is more than 2.5 ST above even the MYN high estimate. These results are reinforced by the pairwise comparison levels of mode, shown in Figure 7.26, panels c. and d. Here, we see that the slope of MWH against MDC (first marker) is almost zero, with the lower CI crossing the zero threshold, and neither parameter achieving significance at p < 0.05. (The proximity of the mean estimated slope to zero and the short CIs indicate that might accent the null hypothesis rather than simply fail to reject it.) The slopes of MYN against MDC and MWH are noticeably higher (second and fourth markers) again, with confidence intervals well above zero. The slopes of MDQ re MDC and MWH are even higher again, again with lower CIs well above the zero line. In fact, although the difference is slighter, MDQ is even significantly higher than MYN for both *f*0 parameters, at 0.9 for l\_f0 and 2.8 ST h\_f0. Unsurprisingly, all MYN and MDQ slopes are significance at a level of p < 0.0001.

|  |  |
| --- | --- |
| Chart  Description automatically generated  a. Estimated slopes for l\_t. | Chart  Description automatically generated  b. Estimated slopes for h\_t. |
| Chart, scatter chart  Description automatically generated  c. Estimated slopes for l\_f0. | Chart, scatter chart  Description automatically generated  d. Estimated slopes for h\_f0. |

Figure .. Summary of mode-only model output phonetic parameters for pairwise comparison of each level of mode. First term in each case is the intercept, and error bars indicate 95% CIs.

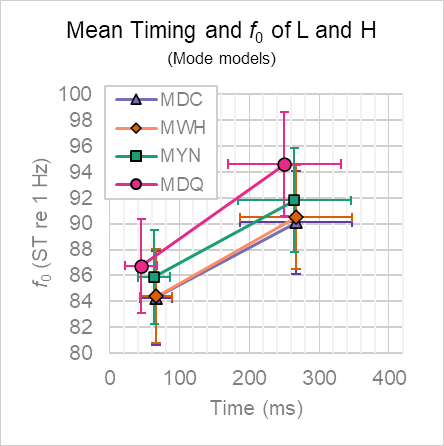


Figure .. Estimated means of phonetic parameters for each level of mode as intercept in mode-only analysis. Error bars indicate 95% CIs.

All told, the analysis of mode-only models gives the impression that there is indeed a difference in scaling from MWH/MDC to MYN to MDQ, much in the vein of Haan’s work on Dutch (2002). As such, these models support the view that there is a gradient increase in *f*0 across sentence modes as the amount of syntactic and lexical signals of interrogativity decrease. This is displayed very clearly in the visualisation of the tonal targets in Figure 7.27, where we can see that there is little difference between MDC and MWH, but that MYN and MDQ have increasingly higher *f*0 scaling.

#### Mode-and-phonology models

ANOVAs for each model indicate that acc\_phon is a significant factor at a level of p < 0. 01 for each parameter. mode is significant for the l\_f0, l\_t, and h\_f0 models at a level of p < 0.0001, and at a level of p < 0.05 for h\_t (p.adj. = 0.049). The acc\_phon:mode interaction is significant for all models (p < 0.05) except the h\_t model (p.adj. = 0.212). (See appendix for full results.).

Table 7.14 shows the marginal and conditional r-squared values for the mode-and-phonology models and, in parentheses, the change from the mode-only models. For each parameter, we can see that mode\*acc\_phon explains only a little more of the variance than the mode-only models. (Treated proportionally, these changes could be construed as large, since and increase from 0.04 to 0.06 in the marginal r2 of l\_f0 is a 50% improvement on the mode-only model; however, such interpretations are quite misleading as the actual change is from 4% of the variance to 6% of the variance.)

|  |
| --- |
| Table .. Marginal and conditional r-squared of mode-and-phonology models for each target parameter. Values in parentheses indicate change from mode-only models |
| |  |  |  | | --- | --- | --- | | parameter | R2m | R2c | | l\_f0 | 0.06 (+0.02) | 0.95 (+0.02) | | h\_f0 | 0.12 (+0.02) | 0.91 (+0.03) | | l\_t | 0.07 (+0.02) | 0.77 (+0.00) | | h\_t | 0.02 (+0.01) | 0.85 (+0.00) | |  |  |  | |

Figure 7.27 summarises the estimated means of the time and *f*0 parameters (β0) for the mode-and-phonology models, while Figure 7.28 shows the slopes (β1) between each level of mode. To facilitate comparison between the two models, estimated means from the mode-only models are shown in grey.

Looking at the time parameters (Figure 7.28a), we see no meaningful difference in the results when compared to the mode-only model. The estimated mean alignment of the low targets in the mode-and-phonology model for MDC, MWH, and MYN are almost identical, at 67, 68, and 69 ms respectively, with very similar CIs (from 46—48 to 89—91 ms). The estimated mean for MDQ is 20—22 ms earlier than these, at 47 ms (CIs = 25—69 ms). The estimated mean alignment of the high target is 168 ms for MDC, MWH and MYN (CIs ≈ 187.5—348.5 ms in each case) and is 16 ms earlier in MDQ at 252 ms (CIs = 192—333). The results of the pairwise comparison of slopes are unsurprising (Figure 7.28a. and b.), indicating no significant difference between in target alignment in MDC, MWH, and MYN (p.adj = 0.931—0.984 across interactions). The earlier alignment of MDQ re other levels of mode is distinct (third, fifth, and sixth markers) and statistically significant (p < 0.01 in each case).

|  |  |
| --- | --- |
| a. Estimated means for temporal alignment | b.Estimated means for f0 targets |

Figure .. Estimated mean tonal target parameters and 95% CIs for each level of mode mode-and-phonology models. Estimated means from the mode-only models are shown in grey.

|  |  |
| --- | --- |
| a. Estimated slopes for l\_t. | b. Estimated slopes for h\_t. |
| c. Estimated slopes for l\_f0. | d. Estimated slopes for h\_f0. |

Figure .. Summary of estimated mean difference between each level of mode (β1) for mode-and-phonology model and 95% CIs. Estimates for the mode-only models are shown in grey First term in each case is the intercept.

Changes in estimated mean *f*0 targets across levels of mode in the mode-and-phonology models are noticeably less pronounced than in mode-only model. This is clear from Figure 7.27b, where we see that the rise in estimated means of *f*0 targets do not rise as sharply from MWH to MYN to MDQ as they do in the mode-only models.

The estimated mean low targets for MDC (84.0 ST, CIs = 80.5—87.5 ST) and MWH (84.3 ST, CIs = 80.5—87.6 ST) are almost identical, with an estimated mean difference of 0.1 ST (Figure 7.28c), which is statistically non-significant at a level of p < 0.05 (CIs = ‑0.2—0.3, p.adj. = 0.617). There is still an *f*0 does still increase from MWH to MYN (85.2 ST, CIs = 81.6—88.7 ST), but the change is less pronounced than the same change in the mode-only model. That is, it is an increase of 1.1 ST rather than 1.7 ST, as shown in the MWH, MYN slope in panel c. (fourth column). Finally, the mean estimated low target *f*0 falls very slightly in MDQ to 85.5 ST (CIs = 81.5—88.6 ST) so that the mean estimated difference between MYN and MDQ is almost zero with short confidence intervals crossing the zero boundary, indicating that the difference between them is almost also non-significant (p.adj. = 0.589).

As can is clear from Figure 7.28c, there is still a significant difference between all levels of mode (p < 0.0001) except for MWH re MDC and MDQ re MYN. However, each slope is noticeably less steep in the mode-and-phonology models when compared with the mode-only models, with each of the significant slopes hovering around one semitone. The change between the modes-only and mode-and-phonology models is particularly distinctive for the slope of MDQ re both MWH and MDC. Whereas the slopes in the mode-only model were 2.5 and 2.6 ST respectively, they both fall to 1 ST in the mode-and-phonology model.

For the high *f*0 targets, the damping of the changes across mode is also very noticeable in the mode-and-phonology model. In Figure 7.27b, we see that the mean estimated *f*0 for high targets (orange squares) slowly increases from 80.2 in MDC to 92.5 in MDQ. However, this contrasts with the more sharply rising curve across levels of mode in the mode-only analysis (grey diamonds), ranging from 90.1 ST in MDC to 94.6 ST in MDQ. The mean estimated *f*0 for MDC and MWH are again very similar in the mode-and-phonology model, at 90.2 (CIs = 86.1—94.2 ST) and 90.6 ST (CIs = 86.5—94.6 ST) respectively. These results are also almost the identical to those in the mode-only analysis (0.1 STs lower in each case). The from MWH to MYN is only 0.7 ST to 91.3 ST (CIs = 87.3—95.3). This contrasts with a rise of 1.2 ST in the mode-only model. The rise from MYN to MDQ is 1.2 STs in the mode-and-phonology mode, with an estimated mean high target for MDQ of 92.5 ST (CIs = 88.4—96.5 ST). This contrasts with a rise of 2.8 ST in the mode-only model.

Looking at the pairwise comparison of levels of mode for high *f*0 targets (Figure 7.28d, orange squared), we see that they all—except for MWH re MDC, which still hovers around zero and is non-significant (p.adj. = 0.102)—much less pronounced in the mode-and-phonology model than in the mode-only model. Again, the most dramatic change is in the models in the slopes of MDQ re both MDC and MWH. Whereas in the mode-only models, these slopes were 4.5 and 4.1 ST respectively, in the mode-and-phonology model, they each fall dramatically to 2.3 and 1.9 ST.

#### Summary

The overall effect of modelling the low and high targets using a mode-and-phonology model is that we see the estimates for mode coming considerably closer together compared to the mode-only model. This The overall lessening of the contrasts is indicated very clearly in Figure 7.29. Panel a shows the *f*0—time plot of targets by mode in the mode-only models, while panel b shows the plots for the mode-and-phonology models. We can see in panel b how the low and high targets of MYN and MDQ have dropped considerably and are now much closer to the targets of MDC and MWH. While MDQ is still also timed a little earlier, the main concern is that the *f*0 values are closer together. This indicates that the apparent paralinguistic effects of mode on *f*0 scaling are indeed noticeably diminished once accent phonology—i.e., pitch accent and register tier effects—is incorporated into the model analysis.

|  |  |
| --- | --- |
| a. Mean tonal target estimates for the mode-only model. | b.Mean tonal targets for the mode+acc\_phon model. |

Figure .. Comparing the estimated means of phonetic parameters for each level of mode as intercept in the mode-only and mode+acc\_phon models Error bars indicate 95% CIs.

A simple but effective way of quantifying the extent to which mode-and-phonology models reduce the apparent effects of mode on *f*0 scaling is to compare the mean and standard deviation across the estimated slopes (β1) between each level of mode across each model. The closer each value is tozero, the less the difference in *f*0 across levels of mode. Table 7.13 shows the results of this analysis. We see the mean β1 of l\_f0 fall from 1.49 ST in mode-only model to 0.7 ST in the mode-and-phonology model, a drop of 0.79 ST. There is a slight decrease in standard deviation of 0.3 ST, from 0.9 ST to 0.6 ST. The mean β1 of h\_f0 falls from 2.48 ST in the mode-only model to 1.27 ST in the mode-and-phonology mode, indicating a mean decrease of 1.21 ST. There is a substantial drop in standard deviation 0.9 ST, from 1.6 ST in the mode-only model to 0.7 ST in the mode-and-phonology model.

Table .. β1 mean and standard deviation of all pairwise comparisons of mode for f0 parameters by model.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| β1 | mode-only | |  | mode-and-phonology | |  | change | |
|  | mean | std dev. |  | mean | std dev. |  | mean | std dev. |
| l\_f0 | 1.5 | 0.9 |  | 0.7 | 0.6 |  | -0.8 | -0.3 |
| h\_f0 | 2.5 | 1.6 |  | 1.3 | 0.7 |  | -1.2 | -0.9 |

These results suggest that the parasitic paralinguistic effect on *f*0 scaling existed in both the mode-only model and in the model which incorporates the phonological effects of register tier as well. However, the apparent effect is quantifiably weaker once the pitch accents and register tier have been incorporated into the model. Most striking is the overall decrease in variance in β0 estimates for h\_f0 from 2.62 to 0.5 ST.

Again, as with the phonological analysis, this comparison does not on its own demonstrate the existence of the register tier. However, it does demonstrate how a register tier analysis helps separate apparent paralinguistic effects from phonological effects, and that apparently distinct paralinguistic effects are greatly dampened when accent phonology is introduced to the model. However, paralinguistic effects do not disappear completely, as suggested in section XX.XX. They are still in operation but are greatly subdued, as can be seen in the comparison of the two types of model in Figure 7.29 and the comparison of β1 *f*0 means and variance in Table 7.15, .

Many varieties of English tend to exploit a distinction in nuclear PAs contours to contrast MDC and MWH on the one hand and MYN and MDQ on the other, i.e., falling H\*L or H\* L% as opposed to rising L\*H H% or H\* H%. It does look as if the DCE speakers here—except for F16—employ low and high register to make the same distinction.

### Phonetic parameters of nuclear pitch accents

While the mode-and-phonology models are useful in teasing out the extent of paralinguistic effects of *f*0 scaling (and temporal alignment) on mode, it is also important to evaluate the phonetic parameters of tonal targets in terms of acc\_phon itself. That is, we need to see if there is empirical evidence validating the differences proposed for the PA/register tier combinations. For this purpose, the estimated means of each tonal target parameter were extracted from the mode-and-phonology model, this time with each level of acc\_phon as intercept. As before, slopes (β1) of each pairwise comparison across levels of acc\_phon were also estimated to help identify where and if there were statistical differences between tonal targets.

As mentioned on p. 124, ANOVAs for each model indicate that acc\_phon is a significant factor at a level of p < 0. 01 for each parameter. Intercepts for each level of acc\_phon are also significant (p < 0.001) for each target parameter (see Appendix XX).

The estimated mean temporal alignment of the low and high target—with the exception of ^[L]\*H—are very similar for each level of acc\_phon, as seen in Figure 7.28a. Mean alignment of the low target is 68 ms for L\*H (CIs = 49—89), 63 ms for L\*[H] (for CIs = 40—86), and 65 ms for ^[L\*H] (43—88). For the high target, mean estimates for L\*H and ^[L\*H] are both 268 ms, while L\*^[H] is a millisecond earlier at 267 ms, with a 95% CI range of 160-161 ms. The low target of ^[L]\*H is aligned slightly later (81 ms, CIs = 52—111 ms) while the high target is aligned noticeably earlier (218 ms, CIs = 138—299 ms).

|  |  |
| --- | --- |
| a. Estimated means for temporal alignment | b.Estimated means for f0 targets |

Figure .. **acc\_phon** mean tonal target parameter estimates (mode+acc\_phon model). Error bars show 95% CIs.

The mean *f*0 estimate of the low target of L\*H is 84 ST (CIs = 80.5—87.5) while it is slightly higher for L\*^[H], at 84.7 ST (CIs = 81.1—88.3), as shown in Figure 7.28b. The mean estimates for the raised L\* targets (^[L\*]) are much closer together, at 86.9 ST for ^[L\*]H (CIs = 83.1—90.6) and only 0.2 ST lower for ^[L\*H], at 86.7 ST (CIs = 83.2—90.3). Mean *f*0 estimates for the non-raised and raised high tonal targets are in each case almost identical. That is the non-raised H in L\*H has an estimated mean of 90.2 ST, while in ^[L\*]H it is 89.9 (CIs = 86.1—94.2 and 85.5—94.3 respectively), only 0.3 ST lower in the case of ^[L\*]H. The estimated mean *f*0 for raised H is 93.7 ST in each case (CIs = 89.6—97.8 in L\*^[H] and 89.7—97.8 in ^[L\*H]).

When we plot the target parameters together on a two-dimensional time-*f*0 plane (Figure 7.33), we see quite clearly that ^[L\*H] (the green line with the circular targets) is essentially a raised version of L\*H (the lilac blue line with the triangular targets). In fact, the target of the raised H in L\*^[H] (orange line with square targets) is the same—in terms of estimated mean *f*0 and temporal alignment at least—as the raised H of the ^[L\*H] accent. The slight difference in the height of the L in the L\*^[H] and L\*H is noticeable (and statistically significant at a level of p < 0.05, p.adj. = 0.035); however, we can still see how it is much closer to the other unraised L target than it is to the raised L targets.

Chart, line chart

Description automatically generated

Figure .. Time-f0 plot of estimated means for L\*H-like acc\_phon tonal targets (mode+acc\_phon models). Error bars indicate 95% CIs.

The targets for ^[L\*]H are less like their counterparts in other pitch accents. Despite the similarities in *f*0 height of the raised L in ^[L\*]H to that of ^[L\*H] and of its unraised H to the H in L\*H, clear differences in timing are evident. The H target is aligned much earlier in ^[L\*]H, which suggests that it might more appropriately be identified as >H\* (or even H\*). Unfortunately, there are so few >H\* and H\* tokens over all (n=2, n=5 respectively), that it is difficult to make generalisations about them beyond there sparseness. In fact, there are only even six tokens for ^[L\*]H, so any conclusions regarding ^[L\*]H should be taken with a pinch of salt.

If we leave ^[L\*]H aside and consider the pairwise analysis of slopes—illustrated in Figure 7.33—we see that the differences in mean are always close to zero. Differences in l\_t slopes range from -4 to 2 ms, while h\_t slopes hover between -1 and 0 ms. In all cases the CIs cross zero, and the differences in alignment are never statistically significant (p.adj = 0.452—0.729 for l\_t slopes and 0.938—0.984 for h\_t slopes).

The estimated mean differences in *f*0—as shown in Figure 7.34c—are much as anticipated. We expect h\_f0 to be higher in both L\*^[H] and ^[L\*H] when compared to L\*H, and we see just this. In both cases, the predicted mean difference is 3.6 ST (CIs = 2.7—4.4 ST and 2.9—4.2 ST respectively). Similarly, where we expect h\_f0 to be the same—i.e., in L\*^[H] re ^[L\*H]—there is an estimated mean difference of 0 ST (CIs = ‑0.86—0.84 ST). For l\_f0, the estimated mean difference between the low targets for L\*[^H] and L\*H is 0.7 ST (CIs = 0.11—1.32 ST), which is statistically significant (p.adj. = 0.035). However, this is still much lower than the estimated mean differences for the raised L targets. The low target in ^[L\*H] is an estimated 2.7 ST higher than that of L\*H, while it is 2.0 ST higher in ^[L\*H] than L\*^[H] (CIs = 2.26—3.20 and 1.43—2.60 ST respectively)

|  |  |  |
| --- | --- | --- |
| a. l\_t | b. h\_t | c. l\_f0 and h\_f0 |

Figure .. Pairwise comparison of estimated means slopes and CIs for L\*H, L\*^[H], and ^[L\*H].

#### Summary

There is little difference across acc\_phon in terms of the temporal alignment of H and L targets for the three most common levels of acc\_phon—i.e., L\*H, L\*^[H], and ^[L\*H]—and no statistical differences between them. However, the main concern in this section is *f*0 scaling, and the essential question here is, “Does a statistical analysis of the tonal targets of different levels of acc\_phon support evidence for the phonological distinction between them?” When it comes to the three common levels of acc\_phon, the answer is clearly yes. The estimated *f*0 mean differences between identical pairs of tonal targets are remarkably similar, while those which we expect to be different are consistently different. At the same time, the differences and similarities in H targets are more precise than those for the L targets, suggesting more speaker pressure to achieve H targets more precisely. This forms an interesting corroboration in the previous chapter, where speakers appear to be under more pressure to retain H targets in PN pitch accents than L targets. Again, this speaks to the special status of H targets, even in a variety of English which has L\*H pitch accents as the unmarked norm.

The results for ^[L\*]H are less satisfying. The earlier alignment of the H target suggests that it might be more akin to >H\* than L\*H, although the *f*0 height of the raised L and the unraised H are consistent with their counterparts in other L\*H-like nuclear pitch accents. Unfortunately, a viable comparison could not be made between ^[L\*]H and >H\*pitch accents, given the sparse number of tokens for either.

### Mode and global phonetic parameters

Two utterance-wide parameters were evaluated, utterance mean *f*0 (utt\_mean\_f0) and *f*0 slope across the slope (utt\_slope). utt\_mean\_f0 is the mean of the contour measured in semitones re 1 Hz while utt\_slope is the slope of the linear regression of the *f*0 contour measured in semitones per second. As with the analysis of tonal targets parameters for mode, utterance-wide parameters were estimated using two types of models, i.e., a mode-only model and a mode-and-phonology model. Both models treat mode as a fixed factor with random slopes of mode per speaker, and random intercepts of speaker and prompt; however, the mode-and-phonology model adds random intercepts for phonological parameters, namely, acc\_phon, fin\_phon, and h\_start. h\_start is binary logical parametetr which is true when an utterance begins with a H boundary or an initial pre-nuclear H\* or >H\* pitch accent associated with the first foot. Again, the aim was to assess how the addition of phonological features to the model might change the apparent paralinguistic effects of mode on the model.

Table .. Optimal working models for LME analysis of global phonetic parameters, where x is the target parameter.

|  |  |  |
| --- | --- | --- |
|  | **Shared factors** | **mode-and-phonology model only** |
| x~ | mode + (1+mode|speaker) + (1|prompt) |  |
| mode + (1+mode|speaker) + (1|prompt) | + (1|h\_start) + (1|acc\_phon) + (1|fin\_phon) |
|  |  |  |

An ANOVA of each model indicate that mode is significant at a level of p < 0.01 for utterance mean *f*0 (p.adj. = 2.4×10-04) and utterance slope (p.adj. = 0.002).

#### Comparison of mode-only and mode-and-phonology models

The mode-only model for utterance mean *f*0 has a marginal r2 of 0.03 and a conditional r2 of 0.95, while the corresponding mode-and-phonology model has a marginal r2 of 0.02 and a conditional r2 of 0.95. The mode-only model for utterance slope has a marginal r2 of 0.5 and a conditional r2 of 0.83. The corresponding mode-and-phonology model has a marginal r2 of 0.22 and a conditional r2 of 0.91.

Because the conditional r2 for each mean *f*0 model is the same (0.95), this suggests that the addition of random phonological effects in the mode-and-phonology model does not add contribute much in terms of the explanation of variance than the mode-only model. The slight decrease 0.01 in the marginal r2 for the mode-and-phonology model suggests, however, that 1% of the explanatory values of the fixed effect of mode in the mode-only model is effectively shifted from the fixed effect of mode-only model to the random effects of the mode-and-phonology model. However, a difference of 0.01 in the marginal r2  of the two models is very minor.

The decrease of 0.28 in the marginal r2 from the mode-only to the mode-and-phonology models is slightly perplexing. However, this again may indicate that once phonological affects have been added to the model, effects which were previously most likely attributable to mode become more appropriately attributable to random phonological effects. It should also be noted that the apparently small marginal r2 *f*0 models (both here and in general, as noted in section 7.6.1) are not because the effect of mode is unimportant, rather it is because *f*0 means are particularly susceptible to random speaker intercepts, within which gender effects are also embedded. This is illustrated in Figure 7.35, which shows the per-speaker random intercepts for the mode-and-phonology models. Panel a. shows the random per-speaker intercepts for mean *f*0 and panel b. shows the same for contour slope. Note that deviations from the mean are much larger for mean *f*0. This is particularly salient in the difference between M4, who has by far the lowest mean *f*0 (85 Hz), and F5, who has the highest (205 Hz).

|  |  |
| --- | --- |
| a. Random speaker intercepts for utt\_mean\_f0 | b. Random speaker intercepts for utt\_slope |

Figure .. Random speaker intercepts for mode-and-phonology models of utterance-wide phonetic parameters.

A comparison of the mode-only and mode-and-phonology models indicates that there is very little difference between the two models for mean *f*0 estimates. Looking at Figure 7.34a, one can see that the estimated means for mean *f*0 for MDC and MWH are the same for both models (86.1 ST for MDC and 86.6 ST for MWH), with only small differences in MYN and MDQ. The estimated mean *f*0 for MYN is 86.7 in the mode-only model and 86.6 ST in the mode-and-phonology model, while for MDQ, it is 88.3 in the mode-only model and 88.0 ST in mode-and-phonology model. The β1 estimates across pairwise comparisons of levels of mode in the two models are also very similar. Figure 7.34b shows that, as with the analyses of the nuclear pitch accent, the difference between levels of mode is less pronounced in the mode-and-phonology model, but the differences are quite small. The change between two types of model ranges from a mean estimated difference of 0.01 ST between MWH and MDC, and 0.35 ST between MWH and MDQ.

|  |  |
| --- | --- |
| a. Estimated intercepts. | b. Estimated slopes for pairwise comparisons. |

Figure .. Estimated means and 95% CIs for utterance mean f0 per level of mode in the mode-only and mode-and-phonology models.

There is a more noticeable change in the *f*0 contour slopes between the two models. For contour slope, we see that the estimated means per level of mode are lower in the mode-and-phonology model than their mode-only counterparts (Figure 7.35 and Figure 7.36a). The difference becomes slightly more pronounced from MDC to MWH to MYN to MDQ, with the mode-and-phonology models being 1.4, 1.5, 1.8. and 2.8 ST/sec lower respectively than their mode-only counterparts. This is most likely because of the acc\_phon random slope compensates for the overall rise in the nuclear pitch accent region of the contour. In both models, MDC and MWH both have a negative slope (falling) while MYN and MDQ have positive slopes (rising). However, in the mode-only mode, it is only with MDC (CIs = -2—1.2 ST) that we cannot be confident the MDC does not have a zero slope (p.adj. = 0.585). In the mode-and-phonology model, however, the CIs are much wider (all cross zero), so we actually lose confidence that overall trajectory is either positive or negative (p.adj. ranges from 0.06 in MWH to 0.785 in MYN).

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

Figure .. Visual comparison of utterance-wide slope parameter in both models. Dotted lines indicate the mode-only model and solid lines the mode-and-phonology model.

The estimated slopes (β1) between each level of mode, however, do not change quite so much, as shown in Figure 7.36b. This is largely because the overall lowering of β0 in the mode-and-phonology models is largely irrelevant when considering the *relative* difference between each level. That is, we see that the β1 estimates for the mode-and-phonology model are generally only slightly closer to zero than in the mode only model. The one exception to this is in the MDC-MWH slope, where the mode β1 estimate is closer to zero. However, this is also the only estimate for which the CIs cross zero and we cannot be confident that there is any difference between the slope of MDC and MWH (p.adj. = 0.115 and 0.117 for the mode-only and mode-and-phonology models respectively). Otherwise, the mode-and-phonology models have β1 values which are between 0.6 (MDC, MYN) and 1.69 ST/sec (MWH, MDQ) closer to zero.

|  |  |
| --- | --- |
| a. Estimated means (β0) per level of mode. | b. Estimated slopes (β1) for pairwise comparisons. |

Figure .. Estimated means and 95% CIs for utterance-wide slope (ST/sec) per level of mode in the mode-only and mode-and-phonology models.

As with the previous comparison of models, we can quantify the differences between the two types of model by looking at the mean and standard deviation of β1 estimates across pairwise comparisons. The mean difference between modes in the mode-only model is 1.17 ST with a standard deviation of 0.78. In the mode-and-phonology model, the mean difference is 0.2 ST lower at 0.97 ST, with a standard deviation of 0.67 (0.1 ST lower). This is a relatively minor change. The mean difference in slopes in the mode-only model is 4.3 ST/sec, while it is 0.85 ST/sec lower in the mode-and-phonology model at 3.45 ST/sec. Thus, while there is an overall drop in slope in the mode-and-phonology model (as shown in Figure 7.36a), there is not much relative overall change in the differences between modes, which is what we are most interested in.

Table .. β1 mean and variance of all pairwise comparisons of mode for global parameters in both models.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| β1 | mode-only | |  | mode-and-phonology | |  | change | |
|  | mean | std dev. |  | mean | std dev. |  | mean | std dev. |
| mean f0 (ST re 1 Hz) | 1.17 | 0.78 |  | 0.97 | 0.67 |  | -0.20 | -0.10 |
| slope (ST/sec) | 4.30 | 3.95 |  | 3.45 | 3.29 |  | -0.85 | -0.66 |

What is clear from the comparison of the two kinds of model is that the effects of the intonational phonology on utterance-wide parameters are much less pronounced than their effects on nuclear pitch accent parameters. In fact, the lowering of the mean difference between levels of modes for utterance-wide parameters in the mode-and-phonology model is essentially negligible. It is only 0.2 ST for mean *f*0 and 0.85 ST/sec for slope. Thus, while we do see an average lowering in β1 estimates across levels of mode, we can conclude that addition of phonology has a very limited effect on the global parameters evaluated here. In fact, the only noticeable change between models is the overall lowering of the mean estimates (β0) of utt\_slope, although because this occurs for each level of mode, its relative effect is quite small.

#### Changes in global parameters across levels of mode

As the effects of the intonational phonology on the global parameters are limited, we will only consider mean *f*0 and slope in terms of the mode-only models here.

There is a very gradual increase in mean *f*0 (utt\_mean\_f0)across modes (grey bars in Figure 7.34a). MDC has the lowest estimated mean, at 86.1 ST (CIs = 82.7—89.5 ST), followed by MWH, 0.5 ST higher at 86.6 ST (CIs = 83.5-89.6). MYN is only marginally higher, with a mean estimate of 86.8 ST (CIs = 83.8—89.9), while the estimated mean for MDQ is 1.5 ST higher again, at 88.3 ST (CIs = 84.8—91.8). Looking at the pairwise comparisons (grey diamonds in Figure 7.34b), we see that there is no significant difference between MDC and MWH (p.adj. = 0.16) or between MYN and MWH (p.adj. = 0.28). This is clear from the CIs, which stretch across the zero boundary. All other slopes in the pairwise comparisons are statistically significant at a level of p < 0.05, with the greatest differences being between MDQ and the other levels of mode. (Note, however, that in the mode-and-phonology model, only the slopes of MYN and MDQ against MDC as intercept achieved significance at a level of p < 0.05.)

For utt\_slope (grey diamonds Figure 7.36a), MDC is closest to zero, at -1.0 ST/sec (CIs = -0.25—0.57 ST/sec). MWH is lower again, at -3.0 ST/sec (CIs = -5.0—-1.0 ST/sec), and it is the only level of mode which is reliably negative, even though there is always a rise in the nuclear pitch accent. MYN has an estimated mean positive slope of 2.4 ST/sec (CIs = 0.8—4.0 ST/sec) while MDQ has the highest at 5.4 ST/sec (CIs = 3.0—7.9). Turning to the pairwise comparisons of levels of mode (grey diamonds Figure 7.36b), we see that only the MDC-MWH β1 of fails to reach significance at a level of p < 0.05 (p.adj. = 0.115). This is reflected in CIs which cross zero (Figure 7.36b). All other pairwise comparisons achieve significance at p < 0.01. Given the very low utterance-wide slope of MWH, and the very high slope of MDQ, it is unsurprising that the greatest estimated mean difference is between these two these, i.e., 9.04 ST/sec (CIs = 5.11—12.96).

#### Summary

The phonological effect on utterance-wide *f*0 contour parameters is much weaker than on the nuclear pitch accent region of the utterance. This may simply reflect the fact that paralinguistic effects tend to stretch across the whole utterance / IP and that phonological effects are concentrated in the nuclear pitch accent. There is, however, an overall effect of mode on both utterance-wide mean *f*0 and slope. The changes across mean *f*0 by level of mode can be summarised as: MDC <= MWH <= MYN < MDQ, while changes across slope can be summarised as MWH < MDC ≈ 0 < MYN < MDQ. Note that MDC is clearly distinguished from MDQ, most notably in terms of slope. The mode-only utterance mean *f*0 and contour slope are plotted in Figure 7.37, with the mode-and-phonology results in grey behind them, We see that, in terms of global *f*0 parameters, there are clearly paralinguistic effects based on levels of mode which are not mitigated by the addition of phonology to the model.

Figure .. Estimated means and 95% CIs for utterance-wide mean f0 and slope for sentence mode in the phonology-only model. (Mode-and-phonology model estimates are plotted in grey in the background.)

## Summary, discussion, and conclusions

The first aim of this chapter was to describe the phonological and phonetic characteristics of intonation as a function of sentence mode. The second, and arguable more important aim, was to assess the register tier hypothesis. This hypothesis states that there is a phonological register tier which DCE speakers can utilize to help distinguish statements and wh-questions from yes-no and wh-questions. This contrasts with a more widely held view that changes in *f*0 scaling across question forms are not a matter of phonological choices but are rather part of paralinguistic effect which sees a gradual increase in *f*0 scaling from MWH to MYN to DCQ. It was proposed that such paralinguistic effects on tonal targets in the nuclear pitch accent would largely disappear once *f*0 parameters were estimated using a model which accounted for phonological choices re the implementation of the register tier. This issue was considered particularly important for northern Irish English since it is well known that the vast majority of nuclear contours in nIE are L\*H % regardless of sentence mode. As it stands, if the majority of pitch accents are L\*H regardless of mode, this would imply that NI speakers typically exploit only gradient paralinguistic effect rather than categorical phonological features. However, if there is a register tier, it is likely that nIE speakers—in this case speakers from Derry City—exploit this in absence of L\*H / H\*L contrast. It was also observed that this does not mean that nIE speakers only use the register tier, rather that it might be more apparent if a register tier is in use when the pitch accents are otherwise identical.

## Phonological analysis

A methodological difficulty arose as soon the labelling of pitch accent began (using the IViE labelling system), because it was clear that it would be very difficult to label or describe discretely different nuclear pitch contours without recourse to the additional register tier labelling. Thus, even before the analysis proper of the data began, it was necessary to make a decision regarding labelling which favoured the register tier hypothesis. Section 7.4 presented examples of the kinds of contours and contrasts which made it difficult not to use a register-tier analysis *a priori*. From an auditory perspective as was as from visual inspection of the *f*0 contour, there were clear cases where speakers made a categorical utterance-wide shift upward shift in register in the realisation of MDQ in contrast to MDC. There were also cases where the upward shift only affected the domain of the nuclear pitch accent but not the final boundary or pre-nuclear components of the IP. This strongly suggested that speakers did have access to a phonological register tier and that could implement it across different domains in the IP, i.e., either across the whole IP, in the nuclear pitch accent only, and in some cases on individual tones. This last observation is probably only true of trailing H tones, but this only became apparent through more detailed statistical analysis.

Another component element of the labelling challenge was that—while there were potentially non-register-tier methods available for labelling the pitch accents and final boundary tones—any alternative labelling would require relocating pitch events which were clearly a component of the pitch accent to the boundary, and reinterpreting rises which began from a local *f*0 minimum in the stressed syllable as H\* rather than L\*. Both the re-allocation and inversion of targets felt methodologically unsound as workarounds to maintain a non-register-tier analysis. Therefore, register tier marking was included in the labelling, with the caret (^) symbol used to identify the onset of high register and square brackets ([]) to identify its extent. As such, the expectation that patterns which are adequately explained only with reference to both the register and tonal tiers was already fulfilled before the labelling could be complete (expectation 4, section 7.1.1).

Once register-tier labelling was complete, the register-tier labelling was converted to non-register tier labelling. To avoid completely hobbling the non-register-tier analysis, this did not simply remove all register tier labelling. It also converted L\*^[H] (L)% nuclear contours to L\*H H(L)%. This was deemed to be the only phonological contour which was a viable non-register tier label, i.e., it did not require relocation or inversion of tonal targets. The non-register tier labelling was used to contrast it against the register-tier labelling to compare the descriptive effectiveness of the register-tier labelling. Admittedly, the non-register tier labelled data acts somewhat of a strawman, as it is derived from and less well specified than the register-tier labelled data. However, I felt that could still serve to demonstrate the strength of register-tier analysis of the phonology which did not relocate tonal targets from the pitch accent to the boundary or interpret identifiably low targets as high.

The non-register tier analysis indicated that the pitch accent itself was incapable of identifying contrasts as almost 99% of utterances has an L\*H nucleus. This reflects the proposition of the first uncontroversial expectation that L\*H would dominate across modes (section 7.1.1). However, this also meant that only the boundary would be associated with mode. A statistical analysis of the likelihood of H% as an effect of mode showed that there is a small possibility of H% in MYN (estimated mean 2% probability) and a slightly higher probability in MDQ (13%). To some extent, this confirmed the second (also uncontroversial) expectation that H% would be more closely associated with MYN and MDQ. However, the probabilities were much lower than expected, and they reflect an overall weak relationship between H% and MYN or MDQ in the non-register analysis. (Note also, that the third expectation of composite boundaries in a register tier analysis was also met. However, this is ultimately uninteresting, as it was a requirement of the conversion from register-tier to non-register-tier labelling to move the H of the L\*^[H] to the boundary. Therefore, if the boundary ended with an L%, it became HL% by default in the relabelling process.)

The register-tier analysis identified four different effects of high register on nuclear L\*H pitch accents. They were L\*H (no effect), ^[L\*]H (raised L target only), L\*^[H] (raised H target only), and ^[L\*H] (pitch accent fully raised). A statistical analysis of the likelihood of high register in the nuclear pitch accent as an effect of gender and mode indicated there was roughly a 10% probability of high register in MYN and a 50% chance in MDQ (MDC and MWH were both estimated at 0%). The results also indicated a strong effect of gender (p.adj. = 0.049), with male speakers more likely to use H% than female speakers, with an estimated mean probability of 39% in MYN and of 89% in MDQ.

When the two models labelling techniques were compared, mode and gender—as fixed factors—in the register-tier model accounted for 69% of the variance, while the whole model accounted for 79% of the variance. In the in the non-register tier model, mode and gender accounted for 41% of the variance, while the whole model accounted for 75%. Given this, and the fact that the higher register tier much more readily associated with MDQ than the H% of the non-register tier analysis, the register tier analysis proved an explanatorily more effective approach. As noted previously, this does not demonstrate that the register tier hypothesis is correct, and as also noted previously, the non-register tier analysis is to some extent a foil against which to demonstrate the efficacy of assuming that there is a phonological register tier. In this way the fifth expectation, that the register tier-analysis would be more indicative of phonological change as an effect of mode than a non-register-tier analysis.

The phonological analysis of IP-wide phonology suggested that speakers were more likely to avoid prenuclear pitch accents in the MYN and MDQ than in MDC and MWH. A statistical analysis of the likelihood of nuclear-PA-only IPs as an effect of gender and mode indicated a 27.4% mean predicted probability of nuclear-PA-only IPs in MDC, but around 48.5% for both MYN and MDQ (although the confidence intervals for this were large). However, there was again an effect of gender, with female speakers much more likely than male speakers to use nuclear-PA-only IPs, with male speakers having only an estimated mean probability of nuclear-PA-only IPs of only 7% in MYN and MDQ.

The two strategies—use of high register and nuclear-PA-only IP—were pooled together, and a statistical analysis was conducted to assess the use of either strategy as an effect of mode and gender. This time, there was no effect of gender. This was unsurprising, as it incorporated one strategy with a male bias and the other with a female bias. However, the mean estimated probability of one of the two strategies in MYN was 59% and 87% for MDQ. While there was still a predicted probability of 16% for MDC (due to the presence of nuclear-PA-only IPs), the fact that both phonological strategies combined—one utterance-wide, the other associated with high register in the nuclear PA—were highly associated with MYN and especially MDQ was a very satisfying outcome. Moreover, it is worth noting that the register-tier analysis of the nuclear PA identified a strategy more closely with male speakers, while an IP-wide analysis (which didn’t require the register tier) identified a different strategy more closely associated with female speakers. Once again, this highlights the explanatory value of the register tier hypothesis. That is, while it doesn’t demonstrate that it *exists*, such felicitous results indicate that by assuming it exists, we can gain insights into different ways in which the intonational phonology is leveraged to help distinguish between sentence modes rather than assuming that mode must be largely expressed through gradient changes in the *f*0 contour.

The final component of the phonological analysis related to the distribution of L%. As predicted—expectation 6, section 7.1.1—instances of L% did occur in each sentence mode, and most commonly in MDQ. A statistical analysis of the likelihood of L% as a function of mode and gender showed that there was little effect of gender (p.adj. = 0.821) while the mean predicted probability of occurrence in MDQ was noticeably higher than other modes, 23% as opposed to 2%—5% for the other modes. However, there were very wide confidence intervals for each level of the model, indicating that the association with L% with mode was quite a weak one. In fact, the marginal r-squared of the model indicated that mode and gender together explained only 7% of the variance. This suggests that the previously stated view that L% is more likely in MDQ but is not itself a marker of interrogativity, holds quite well under statistical scrutiny. Of course, L% is associated with falls at the end of an IP and is therefore also associated with statements rather than questions in other varieties of English. In terms of form, there were a few nuclear contours which resembled the kind associated with declaratives on other varieties. They were H\* L% and >H\* L%, although they were few and all from the same speaker. What was most interesting about >H\* L% was that the speaker used it in DCQ, YNQ, and DCQ. This reinforces the fact that L% is not associated with statement/question contrasts for these Derry speakers, and its frequent occurrence in MDQ—roughly two fifth of the adjusted count as opposed a fifths overall—reinforces the view, outlined in section 7.1.1, that L% serves a surprise-redundancy function rather than an interrogativity marking function. It is, however, worth remembering that tasks developed for these recordings aimed explicitly at eliciting colloquial speech. If one were to analyse more formal speech, such as by a news reader, one may well fine L% functioning as the end of a declarative fall.

## Phonetic analysis

In a similar vein to the phonological analysis, which compared two systems of labelling, the phonetic analysis compared to kinds of model for assessing the extent to which a phonological register tier might account for apparent gradient changes in *f*0 scaling in the nuclear L\*H pitch accent across sentences modes. Two *f*0 parameters were measured and evaluated, the *f*0 minimum and maximum in the nuclear pitch accent. The time points the *f*0 minimum and maximum were also evaluated (temporal alignment from the onset of the vowel in the stressed syllable). Together each time-*f*0 pairing was treated as a phonetic tonal target in the realisation of the underlying phonological tone of the pitch accent.

A statistical analysis of the tonal targets with mode-only as a fixed factor suggested that there was no difference in *f*0 scaling or alignment for MDC and MWH; however, there was a clear gradient increase from MWH to MYN and from MYN to MDQ, with the temporal alignment for MDQ alone being slightly earlier. As such, the mode-only analysis conformed with expectation about the paralinguistic effects of mode, summarised in expectation 7 (section 7.1.2) as:

1. *f0(*MDC) <= *f0(*MWH) < *f0(*MYN) < *f0(*MDQ)

The statistical analysis including mode *and* nuclear pitch accent phonology—i.e., the four different register-tier / L\*H combinations—greatly reduced the apparent scaling of *f*0 targets as an effect of mode. That is, the paralinguistic effects of mode did not disappear completely. Therefore, the expectation that, “apparent paralinguistic effects… will disappear in a model incorporating the effects of the register tier” (p. 93 above) were clearly overstated. In fact, what this suggests is that use of register tier is not unlike what we see in other varieties of English with PA contrasts between declarative falls and question rises. That is, phonological contrasts are available, but they are not always implemented by speakers. (In the IViE data, for example, it was only the Cambridge speakers who consistently exploited contrasting PAs to differentiate MDC and MWH from MYN and MDQ [REF]). Similarly, the register tier may be used to signal a contrast between MDC/MWH and YNQ/MDQ. At the same time, however, just as in the other varieties (Grabe, Kochanski and Coleman, 2003), there is still a gradient effect of question mode. In fact, when two global parameters were analysed—utterance mean *f*0 and linear regression of the slope of *f*0(t) of the utterance—the effects of phonology were still evident; however, changes in the differences between parameters across levels of mode were not greatly affected by the addition of phonological features to the model. This is most likely because the bulk of the phonological effect on mode occur, unsurprisingly, in the nuclear pitch accent. Therefore, when we look at the utterance as a whole, the phonological effects will be muted, and the gradient paralinguistic effects of mode will come to the fore. Therefore, we have to conclude that both gradient and categorical effects operate on the *f*0 contour, even if we can account for the phonological effects more fully.

The final phonetic evaluation was of the tonal targets as the targets of the pitch accents themselves because it was important to establish if the proposed register tier effects on tonal targets were statistically identifiable. In this analysis it was found that the *f*0 scaling of non-raised H targets—those in L\*H and ^[L\*]H—were not only *not* statistically different from each other (9.adj = 0.907) but that their estimated means were only 0.3 ST different from one another. Similar results were found for the *f*0 scaling raised H targets—those in L\*^[H] and ^[L\*H]—but the mean estimates were identical (p.adj = 0.984). Again, raised L targets—those in ^[L\*]H and ^[L\*H]—were very similar, with a mean estimated difference of 0.1 ST (p.adj. = 0.929). The only case where *f*0 scaling produced a statistically significant difference between targets was in the unraised L—those in L\*H and L\*^[H]—where the means estimate for the L of the L\*^[H] was 0.7 ST higher than that of L\*H (p.adj. = 0.035). However, the mean estimated difference between each raised and unraised L target was always much higher (between 2 and 2.9 semitones, p.adj. < 0.005 in each case). These results suggest that the interpretation of phonological raised/unraised register is overall very consistent, and the prediction that there would be significant differences in scaling due to register tier effects (exception 9, section 7.1.2) appear to have been correct.

The results for the timing, however, suggest that the ^[L\*]H target may be better interpreted as >H\* or H\*. This is because the H is aligned much earlier than any of the other H targets (an estimated 49-50 ms earlier), in a manner akin to the H\* and >H\* PAs of chapter six. However, as there very few tokens to compare on this matter, it was not possible to carry out a detailed analysis of this. That said, the ^[L\*]H label did raise the most doubts during labelling, so, in retrospect, it might better have been labelled as >H\* to let it lie in the ambiguous in-betweenness between L\*H and H\*.

## Conclusion

Overall, the analyses of mode in this chapter suggest that the register tier hypothesis provides an effective means of describing the phonological and phonetic data. This was true, first of all, in the labelling process, where it was deemed that a non-register tier labelling approach would have misrepresented the data and ignored salient distinctions. In the phonological analysis, the inclusion of the register tier helped identify two different (gendered) phonological strategies which speakers use to help signal the difference between sentence modes, most importantly between MDC and MDQ. The inclusion of register tier phonology in the modelling of tonal targets greatly reduced the apparent paralinguistic effects of mode although it did not erase them completely. In fact, the phonetic analysis of utterance-wide parameters showed that outside the nuclear pitch accent, the gradient effects of mode on *f*0 are more readily apparent. Therefore, in offering a description of the phonological and phonetic characteristics of nuclear pitch contours in DCE across sentence modes (RQ2), we have also answered research question four, “Does a register tier provide a plausible phonological explanation for variation across sentence modes in DCE?” The answer to this question, I believe, has to be yes. It does not explain away all the paralinguistic / gradient effects of mode, but it does combine with other phonological features to explain how speakers sometimes signal sentence mode. It also allows us to see, perhaps more accurately, the extent of the paralinguistic effect of mode, as we can add phonological register tier effects to our models to account for / remove residual phonological effects from the phonetic contour.

# From Phonology-First Analysis to Phonetics-First Analysis

## Critical reflection of approaches to formal and functional analysis.

### Summary of insights

### Highlight weaknesses

## Turning Points and Secondary Tones

### Suggested IP Structure

### Identification of Turning Points

### Potential benefits

## Tonal Centre of Gravity versus TP/STH approach

# A whirlwind tour of STH analysis

## Revisiting Metrical and lexical effects

## Revisiting Mode

## Focus

# Summary and Conclusion

Table .. Summary of model output phonetic parameters for each level of mode as intercept in the mode-only analysis. f0 is measured in ST re 1 Hz.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MDC** | **est.** | **2.5% CI** | **97.5% CI** | **std.error** | **z.value** | **df** | **p.adj** | **signif.** |
| l\_f0 (ST) | 84.3 | 80.6 | 87.9 | 1.66 | 50.84 | 11.54 | 2.0E-14 | p < 0.001 |
| h\_f0 (ST) | 90.1 | 86.1 | 94.1 | 1.78 | 50.61 | 9.13 | 3.8E-12 | p < 0.001 |
| l\_t (ms) | 66 | 44 | 89 | 10.5 | 6.34 | 12.38 | 5.3E-05 | p < 0.001 |
| h\_t (ms) | 267 | 187 | 348 | 31.2 | 8.56 | 4.95 | 0.0005 | p < 0.001 |
| MWH | est. | 2.5% CI | 97.5% CI | std.error | z.value | df | p.adj | signif. |
| l\_f0 (ST) | 84.4 | 80.8 | 88.0 | 1.66 | 50.91 | 11.54 | 2.0E-14 | p < 0.001 |
| h\_f0 (ST) | 90.5 | 86.5 | 94.5 | 1.78 | 50.84 | 9.14 | 3.7E-12 | p < 0.001 |
| l\_t (ms) | 67 | 44 | 89 | 10.5 | 6.37 | 12.38 | 5.0E-05 | p < 0.001 |
| h\_t (ms) | 267 | 186 | 347 | 31.2 | 8.55 | 4.95 | 0.0005 | p < 0.001 |
| MYN | est. | 2.5% CI | 97.5% CI | std.error | z.value | df | p.adj | signif. |
| l\_f0 (ST) | 85.9 | 82.3 | 89.5 | 1.66 | 51.84 | 11.52 | 1.8E-14 | p < 0.001 |
| h\_f0 (ST) | 91.8 | 87.8 | 95.8 | 1.78 | 51.62 | 9.10 | 3.6E-12 | p < 0.001 |
| l\_t (ms) | 64 | 41 | 87 | 10.4 | 6.13 | 12.30 | 7.3E-05 | p < 0.001 |
| h\_t (ms) | 45 | 23 | 68 | 10.4 | 4.35 | 12.14 | 1.0E-03 | p < 0.01 |
| MDQ | est. | 2.5% CI | 97.5% CI | std.error | z.value | df | p.adj | signif. |
| l\_f0 (ST) | 86.7 | 83.1 | 90.4 | 1.66 | 52.41 | 11.47 | 1.8E-14 | p < 0.001 |
| h\_f0 (ST) | 94.6 | 90.6 | 98.6 | 1.78 | 53.28 | 9.04 | 3.2E-12 | p < 0.001 |
| l\_t (ms) | 45 | 23 | 68 | 10.4 | 4.35 | 12.14 | 1.0E-03 | p < 0.01 |
| h\_t (ms) | 250 | 170 | 331 | 31.1 | 8.04 | 4.92 | 0.0007 | p < 0.001 |

|  |
| --- |
| Table .. Summary of phonetic parameters for pairwise comparison of slopes in the mode-only analysis. f0 is measured in ST re 1 Hz. The first term is always the intercept, the second the slope. |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **MDC** ,  **MWH** | **β1** | **2.5% CI** | **97.5% CI** | **SE** | **t** | **df** | **p.adj** | **sig.** | | l\_f0 (ST) | 0.1 | -0.2 | 0.4 | 0.16 | 0.79 | 610.99 | 0.513 |  | | h\_f0 (ST) | 0.4 | 0.0 | 0.8 | 0.22 | 1.83 | 615.98 | 0.097 |  | | l\_t (ms) | 0 | -4 | 4 | 2.0 | 0.20 | 610.88 | 0.906 |  | | h\_t (ms) | 0 | -6 | 5 | 3.0 | -0.13 | 613.04 | 0.931 |  | | MDC , MYN | β1 | 2.5% CI | 97.5% CI | SE | t | df | p.adj | sig. | | l\_f0 (ST) | 1.6 | 1.3 | 1.9 | 0.16 | 10.20 | 611.13 | 1.4E-21 | p < 0.001 | | h\_f0 (ST) | 1.7 | 1.3 | 2.2 | 0.22 | 7.73 | 616.12 | 2.9E-13 | p < 0.001 | | l\_t (ms) | -2 | -6 | 2 | 2.02 | -1.14 | 611.71 | 0.3220 |  | | h\_t (ms) | -2 | -8 | 3 | 2.97 | -0.83 | 613.15 | 0.4900 |  | | MDC , MDQ | β1 | 2.5% CI | 97.5% CI | SE | t | df | p.adj | sig. | | l\_f0 (ST) | 2.5 | 2.1 | 2.8 | 0.17 | 14.48 | 612.81 | 1.9E-39 | p < 0.001 | | h\_f0 (ST) | 4.5 | 4.0 | 5.0 | 0.24 | 18.94 | 617.96 | 2.6E-61 | p < 0.001 | | l\_t (ms) | -21 | -25 | -17 | 2.15 | -9.81 | 597.23 | 3.8E-20 | p < 0.001 | | h\_t (ms) | -17 | -23 | -11 | 3.18 | -5.30 | 614.88 | 5.9E-07 | p < 0.001 | | MWH , MYN | β1 | 2.5% CI | 97.5% CI | SE | t | df | p.adj | sig. | | l\_f0 (ST) | 1.5 | 1.2 | 1.8 | 0.16 | 9.44 | 611.16 | 7.3E-19 | p < 0.001 | | h\_f0 (ST) | 1.3 | 0.9 | 1.7 | 0.22 | 5.91 | 616.17 | 2.1E-08 | p < 0.001 | | l\_t (ms) | -3 | -7 | 1 | 2.01 | -1.34 | 611.91 | 0.234 |  | | h\_t (ms) | -2 | -8 | 4 | 2.97 | -0.71 | 613.21 | 0.563 |  | | MWH , MDQ | β1 | 2.5% CI | 97.5% CI | SE | t | df | p.adj | sig. | | l\_f0 (ST) | 2.4 | 2.0 | 2.7 | 0.17 | 13.73 | 612.86 | 4.4E-36 | p < 0.001 | | h\_f0 (ST) | 4.1 | 3.6 | 4.6 | 0.24 | 17.20 | 618.02 | 1.1E-52 | p < 0.001 | | l\_t (ms) | -21 | -26 | -17 | 2.15 | -10.00 | 595.95 | 8.3E-21 | p < 0.001 | | h\_t (ms) | -16 | -23 | -10 | 3.19 | -5.17 | 615.00 | 1.1E-06 | p < 0.001 | | MYN , MDQ | β1 | 2.5% CI | 97.5% CI | SE | t | df | p.adj | sig. | | l\_f0 (ST) | 0.9 | 0.5 | 1.2 | 0.17 | 5.14 | 612.36 | 1.2E-06 | p < 0.001 | | h\_f0 (ST) | 2.8 | 2.3 | 3.3 | 0.23 | 12.00 | 617.52 | 1.2E-28 | p < 0.0001 | | l\_t (ms) | -19 | -23 | -15 | 2.10 | -8.92 | 609.03 | 4.5E-17 | p < 0.0001 | | h\_t (ms) | -14 | -20 | -8 | 3.11 | -4.61 | 614.38 | 1.4E-05 | p < 0.0001 | |  |  |  |  |  |  |  |  |  | |

Table .. Summary of phonetic parameters for each level of mode as intercept in the mode and acc\_phon analysis. f0 is measured in ST re 1 Hz.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MDC** | **est.** | **2.5% CI** | **07.5% CI** | **std.error** | **t** | **df** | **p.adj** | **sig.** |
| l\_f0 (ST) | 84.0 | 80.45 | 87.53 | 1.62 | 51.75 | 11.97 | 0.000 | p < 0.001 |
| h\_f0 (ST) | 90.2 | 86.12 | 94.20 | 1.80 | 50.01 | 9.65 | 0.000 | p < 0.001 |
| l\_t (ms) | 67 | 45.9 | 88.9 | 9.9 | 6.82 | 12.22 | 0.000 | p < 0.001 |
| h\_t (ms) | 268 | 187.8 | 348.7 | 30.7 | 8.73 | 4.71 | 0.0006 | p < 0.001 |
| MWH | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj | sig. |
| l\_f0 (ST) | 84.08 | 80.54 | 87.61 | 1.62 | 51.80 | 11.97 | 7.6E-15 | p < 0.001 |
| h\_f0 (ST) | 90.6 | 86.54 | 94.62 | 1.80 | 50.23 | 9.65 | 1.4E-12 | p < 0.001 |
| l\_t (ms) | 68 | 46.23 | 89.24 | 9.89 | 6.85 | 12.22 | 2.9E-05 | p < 0.001 |
| h\_t (ms) | 268 | 187.32 | 348.20 | 30.71 | 8.72 | 4.71 | 0.0006 | p < 0.001 |
| MYN | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj | sig. |
| l\_f0 (ST) | 85.2 | 81.64 | 88.72 | 1.62 | 52.45 | 12.00 | 7.0E-15 | p < 0.001 |
| h\_f0 (ST) | 91.3 | 87.25 | 95.33 | 1.81 | 50.57 | 9.69 | 1.3E-12 | p < 0.001 |
| l\_t (ms) | 69 | 47.93 | 91.03 | 9.93 | 7.00 | 12.42 | 0.0000 | p < 0.001 |
| h\_t (ms) | 268 | 187.71 | 348.56 | 30.74 | 8.72 | 4.72 | 0.0006 | p < 0.001 |
| MDQ | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj | sig. |
| l\_f0 (ST) | 85.03 | 81.49 | 88.57 | 1.63 | 52.24 | 12.12 | 0.0000 | p < 0.001 |
| h\_f0 (ST) | 92.46 | 88.42 | 96.50 | 1.81 | 51.09 | 9.79 | 0.0000 | p < 0.001 |
| l\_t (ms) | 47.04 | 25.38 | 68.70 | 10.02 | 4.69 | 12.95 | 0.0006 | p < 0.001 |
| h\_t (ms) | 252.32 | 171.95 | 332.69 | 30.80 | 8.19 | 4.76 | 0.0007 | p < 0.001 |

|  |
| --- |
| Table .. Summary of phonetic parameters for pairwise comparison of slopes in the mode and acc\_phon analysis. f0 is measured in ST re 1 Hz. The first term is always the intercept, the second the slope. |
| |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **MDC, MWH** | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj. | | sig. | | l\_f0 (ST) | 0.09 | -0.18 | 0.35 | 0.13 | 0.64 | 599.0 | 0.602 | |  | | h\_f0 (ST) | 0.42 | 0.03 | 0.80 | 0.20 | 2.13 | 610.0 | 0.051 | |  | | l\_t (ms) | 0.3 | -3.5 | 4.2 | 2.0 | 0.17 | 604.9 | 0.928 | |  | | h\_t (ms) | -0.5 | -6.2 | 5.3 | 2.9 | -0.15 | 608.0 | 0.929 |  | | | MDC, MYN | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj. | | sig. | | l\_f0 (ST) | 1.2 | 0.9 | 1.5 | 0.15 | 8.05 | 599.0 | 3.4E-14 | | p < 0.001 | | h\_f0 (ST) | 1.1 | 0.7 | 1.6 | 0.22 | 5.21 | 610.1 | 9.1E-07 | | p < 0.001 | | l\_t (ms) | 2 | -2 | 6 | 2.2 | 0.95 | 605.2 | 0.425 | |  | | h\_t (ms) | 0 | -6 | 6 | 3.2 | -0.03 | 608.3 | 0.984 | |  | | MDC, MDQ | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj. | | sig. | | l\_f0 (ST) | 1.0 | 0.6 | 1.4 | 0.21 | 5.06 | 600.5 | 1.8E-06 | | p < 0.001 | | h\_f0 (ST) | 2.3 | 1.7 | 2.9 | 0.29 | 7.87 | 611.8 | 1.1E-13 | | p < 0.001 | | l\_t (ms) | -20 | -26 | -15 | 2.9 | -6.99 | 569.4 | 3.7E-11 | | p < 0.001 | | h\_t (ms) | -16 | -24 | -7 | 4.4 | -3.64 | 610.4 | 0.001 | | p < 0.001 | | MWH, MYN | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj. | | sig. | | l\_f0 (ST) | 1.1 | 0.8 | 1.4 | 0.15 | 7.43 | 599.0 | 2.0E-12 | | p < 0.001 | | h\_f0 (ST) | 0.7 | 0.3 | 1.1 | 0.22 | 3.28 | 610.1 | 0.002 | | p < 0.01 | | l\_t (ms) | 2 | -3 | 6 | 2.2 | 0.80 | 605.3 | 0.510 | |  | | h\_t (ms) | 0 | -6 | 7 | 3.2 | 0.12 | 608.4 | 0.931 | |  | | MWH, MDQ | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj. | | sig. | | l\_f0 (ST) | 1.0 | 0.5 | 1.4 | 0.21 | 4.62 | 600.5 | 1.3E-05 | | p < 0.001 | | h\_f0 (ST) | 1.9 | 1.3 | 2.5 | 0.29 | 6.43 | 611.8 | 1.1E-09 | | p < 0.001 | | l\_t (ms) | -21 | -26 | -14.96 | 2.9 | -7.09 | 568.0 | 2.0E-11 | | p < 0.001 | | h\_t (ms) | -15 | -24 | -6.86 | 4.4 | -3.53 | 610.4 | 0.001 | | p < 0.001 | | MYN, MDQ | est. | 2.5% CI | 07.5% CI | std.error | t | df | p.adj. | | sig. | | l\_f0 (ST) | -0.1 | -0.6 | 0.3 | 0.21 | -0.69 | 600.4 | 0.573 | |  | | h\_f0 (ST) | 1.2 | 0.6 | 1.8 | 0.30 | 3.88 | 611.6 | 0.000 | | p < 0.001 | | l\_t (ms) | -22 | -28 | -17 | 3.0 | -7.44 | 578.4 | 2.0E-12 | | p < 0.001 | | h\_t (ms) | -16 | -25 | -7 | 4.5 | -3.51 | 610.0 | 0.001 | | p < 0.01 | |  |  |  |  |  |  |  |  | |  | |

1. The term tonal event will be used throughout to refer to pitch accents, phrase accents, and boundary tones, which are all landmark events within the intonational phrase. [↑](#footnote-ref-2)
2. I am not totally convinced that global downstep can wholly account for declination. I think that declination effects are frequently seen in plateaux and valleys in the contour unless overridden by a phonological imperative. [↑](#footnote-ref-3)
3. I do not wish to suggest that this is not possible in other approaches, particularly in the British tradition; however, I do think the AM approach is more amenable to this kind of study. [↑](#footnote-ref-4)
4. Gussenhoven makes the relationship between tone and the metrical hierarchy explicit by using subscript Greek letters to indicate tiers within the metrical structure. Thus, IP is represented as ι and boundary tones associated with the IP are identified using a subscript ι, as can be seen in the formal summary of the grammar in (1). Gussenhoven also avoids the + sign for trailing accents, as he feels it wrongly implies that the two tones must be realised close together, when in fact they are subject to rightward displacement (2004, p. 134). [↑](#footnote-ref-5)
5. In order to highlight how the effort code may not be reflected in or may contradict the phonology, Gussenhoven points to a study of German [REF] in which opposite results were found. [↑](#footnote-ref-6)
6. MAE\_ToBI (Mainstream American English ToBI) is the current iteration of ToBI analysis of General American English. However, for the sake of convenience—and because there are not several variants of ToBI within English and across other languages—ToBI is used here. It should also be noted that the “Break Index” component of ToBI is not covered here, as it is more closely connected to the identification of *ip* and IP boundaries, which are less of a concern for the research reported in the subsequent chapters. [↑](#footnote-ref-7)
7. For convenience, the term *anacrusis* will generally be used both for unstressed syllables before the first lexical stress in the foot and before the lexically stressed syllable associated with the nuclear pitch accent. It is simply too clumsy to refer to the latter as “unstressed syllables preceding the lexical stress associated with nuclear pitch accents.” [↑](#footnote-ref-8)
8. Note that in this study, the IP is always the same as the utterance in this data, so the terms *utterance* and *IP* are used interchangeably here. This does not imply that the IP and the utterance are coequal in general. [↑](#footnote-ref-9)
9. Note that the term *accent phonology* is used hereon in to indicated phonological events associated with the pitch accent, which also includes register tier effects in the register-tier analysis. In statistical analysis, as previously, this is abbreviated to acc\_phon. [↑](#footnote-ref-10)
10. A nuclear contour such as L\*H ^[L%] is nonsensical, since would be indistinguishable from L\*H %, and L\*H ^[%] is equally nonsensical since there is no tone to be raised by high register when the boundary is unspecified. [↑](#footnote-ref-11)
11. For the sake of convenience and to save space, ST is shorthand for ST re 1 Hz unless stated otherwise. [↑](#footnote-ref-12)