Up-Rising: A Phonological and Phonetic Study of Intonation in Derry City English

Antoin Eoin Rodgers

Thesis for the Degree of Doctor of Philosophy

Phonetics and Speech Laboratory
School of linguistic, Speech, and Communication Sciences
Trinity College Dublin

2023

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signature of author:

Antoin Eoin Rodgers

September 2022

“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

Summary

This thesis provides an account of the phonology and phonetics of intonation in Derry City English within the Autosegmental-Metrical (AM) framework (Ladd, 2008). Derry City is the second largest urban area in Northern Ireland, and the variety of English spoken there is known as northern Irish English (nIE). Within the AM framework, intonation is understood here as the post-lexical linguistically structured use of pitch to convey meaning (Ladd, 2008).

The first aim of this thesis is to offer a description of Derry City English within the AM framework so that it is amenable to comparison with other AM studies of nIE, other varieties of English, and of other languages. It focuses on the phonological inventory and phonetic features of pitch events in different metrical and lexical contexts and in different sentence modes, i.e., declarative statements, wh-questions, yes-no questions, and declarative questions. The second aim of the thesis is theoretical in concern. There are two parts to this. Firstly, it aims to investigate if—despite the dominance of nuclear rises (L*H) as opposed to the high pitch accents (H*) found in other varieties of English—there is evidence for a special status of the H tone in DCE, and, by association, nIE. The second is to test for evidence of a phonological register tier which nIE speakers are able to exploit to aid in the kind of phonological contrast that speakers of other varieties indicate through falling and rising nuclear contours. This is the register tier hypothesis.

In order to achieve these aims, a corpus of read speech was produced by recording 11 speakers of Derry City English (6F, 5M, mean age = 40, SD 9.9). The stimuli for read speech controlled for the number of syllables in anacrusis (syllables preceding the first stress in the IP), in the first foot, and in the final foot. They also controlled for the word boundary placement, as well as for sentence mode. All utterances (n = 1427) were analysed and annotated in Praat (Boersma and Weenink, 2002) following the IViE labelling guide (Grabe 2001), albeit with a few adjustments. The f_0 contour and the annotations were processed to generate two data tables which were used for all subsequent phonological and phonetic analyses. Statistical analysis was conducted in R (R Core team, 2022), using linear mixed-effects (LME) and Bayesian generalised linear mixed-effects (BGLMM) models (Bates et al., 2015; Chung & Rabe-Hesketh, 2013).

Two major analyses were conducted, each taking the phonology of the IP as the starting points (a Phonology-first approach). The first looked at the phonology and phonetics of intonation in unmarked declaratives utterances under changes in anacrusis, foot size, and lexical boundaries. The second analysed the phonology and phonetics of intonation in different sentence modes. A third shorter analysis was also conducted. This adopted an approach which involved identifying the minimal number of turning points in each f_0 contour required produce a perceptually identical copy of the original. From these turning points, the phonological structure was analysed (a Phonetics-first approach). A Praat Plugin, K-Max (Rodgers, 2020) facilitated this analysis. A large-scale analysis was not conducted, but a number of phenomena found in the first two analyses were investigated.

The analysis of metrical and lexical effects found that L*H was more likely in prenuclear (PN) position when the foot was longer or when the word-final syllable was later in the foot. H* was more common in shorter feet and when the stressed syllable was early in the foot. This was interpreted to mean that the L target was more prone to deletion than the H target, suggesting that the H tone has a privileged status in the PN position. Nuclear pitch accents were resistant to lexical and metrical effects, and all were L*H. The phonetic analysis indicated that tonal targets in PN pitch accents were more vulnerable to metrical and lexical effects than those in nuclear pitch accents. Nuclear peak alignment was strongly affected by foot size; however, when analysed as a proportion of voicing in the final foot, it was remarkably stable. Stress clash effects were found, with the PN L*H rise being truncated in order to make way for the upcoming nuclear L, while the nuclear L target was aligned slightly later. The nuclear L*H rise was subject both to compression and truncation depending on the size of the foot, with the two strategies working in tandem to maintain the proportional alignment of the peak relative to the voiced material in the foot. The fact that a truncation strategy was available to maintain the prenuclear H target but that the prenuclear L target was subject to deletion offered more evidence of the special status of the H tone.

The phonological analysis of intonation and mode used two labelling systems, one using a register tier and one not. BGLMM analyses using both labelling systems indicated that a register-tier analysis was superior to a non-register tier analysis, with high register in the nuclear pitch accent most likely in declarative questions, especially among male speakers. IPs with a nuclear pitch-accent only (i.e., without prenuclear pitch accents) were increasingly likely in yes-no questions and declarative questions, especially among female speakers. When treated as a single parameter, the likelihood of one or other strategy in declarative questions was very high. This was further evidence of a register tier. A dual strategy was employed for the phonetic analysis, using LME models with and without register tier effects. Differences in scaling as an effect of mode were noticeably lessened in the models with register tier effects. Thus, even with the register tier, paralinguistic effects of mode are still evident even though the effect of the register tier was also clear. The f_0 scaling of low register L and H targets and high register L and H targets was well distinguished, providing more support for the register tier hypothesis.

The Phonetics-first analysis provided some evidence that the >H* PA found mostly in prenuclear position can be either an H* with a long plateau or an L*H in which the L target has been suppressed in the rise, and that downstepped !H* pitch accents may require anticipatory lowering (L_). It also indicated how the apparently earlier tonal alignment of nuclear L targets in declarative questions may in fact be a side effect of physiological constraints on the maximum rate of change of f_0.

The research reported here contributes both to the body of knowledge on the intonation of northern Irish English and to the description of intonation as conceptualised within the AM approach. It is hoped that the investigation into the possibility of a phonological register tier, even if not confirmatory, can help shed light on some otherwise challenging intonational phenomena.

Acknowledgments

I would first of all like to thank my supervisor, Prof. Ailbhe Ní Chasaide, for all her guidance and support during this long and gruelling process, and for all the enlightening, entertaining, and stimulating conversations about speech, Phonetics, childhood in Gaoth Dobhair, and the miraculous effects of a dip in the sea. I would especially like to thank Dr Irena Yanushevskaya for her endless encouragement, support, patience, and incisive feedback; Dr Maria O'Reilly for lending me her ears and a copious amount of her time to help listen to the lovely tunes of Derry English, and to talk about all things phonetic and otherwise; and to Dr Andy Murphy for his all-round support and consideration, especially in checking in to see how my work was progressing (not to mention being crossworder-in-chief).

I would also like to thank my fellow researchers in the Phonetics and Speech Laboratory for their good company, guidance, support, and encouragement.

Thanks too to Glen Doherty, his niece Aoife, and my own nieces Anne and Affric, all of whom helped pilot early versions of the recording stimuli.

I will be eternally grateful for the incredible hospitality of Kevin and Áine Abbott, who were fantastic hosts whenever I was up in Derry and were also great craic to spend time with. I don't think I can thank you enough for how welcome you made me feel.

Special thanks must go to everyone at the Verbal Arts Centre. Without James Kerr's generous offer of free use of the recording studio, and without all the support from the staff both in recruiting and being recruited as participants, it would not have been possible to even begin the work in this thesis. I am especially grateful to Rory McCarron, who was incredibly generous with his time and talent, and who made the recording process go much more smoothly than I ever could have imagined.

To Miche, Jennifer, and especially to proof-readers extraordinaire and all-round wonder-people Kerry and Nora, thank you so much for reading pages upon pages of what I know must largely have read like gobbledygook. Some of it most certain was before you got your hands, keen eyes, and (virtual) red pens on it. For all which remains so, the fault is all my own.

Finally, the biggest thank you to Mammy and Daddy Rodgers, for all your love and faith in me. Thank you so much.

Table of contents

List of Tables

List of Figures

1 Introduction

This dissertation provides an account of the phonology and phonetics of intonation in Derry City within the Autosegmental Metrical (AM) framework (Ladd, 2008). 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)[1]. The majority of AM research on northern Irish English has focused on Belfast English, with a much smaller body of research on Donegal English (Grabe et al., 2000; Grabe & Post, 2002; Kalaldeh et al., 2009; Lowry, 2002, 2011; O'Reilly et al., 2010 inter alia). Only one study of intonation in Derry City has been carried out in the last forty years (McElholm, 1986), and that presented an analysis of two speakers within the framework of the British Tradition of intonational analysis (Cruttenden, 1997). 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 different metrical and lexical contexts and in sentences with different grammatical functions. More precisely, it analyses effects of variation in foot size, anacrusis (unstressed content before the first stressed syllable), and lexical boundaries on the intonational form of unmarked declarative statements. On the functional side, it analyses and compares variation in intonation as a function of sentence mode, i.e., in declaratives, binary questions, wh-questions, and declarative questions.

The second aim of the research is more theoretical in nature. It originates in questions raised by the pervasive use of rising intonation in unmarked contexts in northern Irish English, whereas most other varieties of English (and other languages) tend to have a falling pitch. 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

---

[1] Pitch is used here to refer to a subjective percept which correlates very closely but not exactly with the objective measurement of the fundamental rate of oscillation in the vocal folds. As such, it is the fundamental frequency (f_0) of the (quasi)periodic acoustic signal in voiced speech. In the introduction, the term pitch is used, but for most of the thesis, f_0 is used, as this is the key parameter in the acoustic phonetic analysis.

distinction may be easy to maintain as long as unmarked declaratives and binary questions typically employ pitch accents with 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, as a rising contour dominates both in declaratives (L*H %) and in 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 intonational structure.

A standard AM explanation would be that there is a phonetic, gradient component of the pitch contour which is a manifestation of attitudinal shifts across sentence modes (Gussenhoven, 2004; Ohala, 1983). Because of this, question forms exploit progressively higher pitch registers as the amount of grammatical or semantic question marking decreases (Haan, 2002). Such changes would be identified as paralinguistic, i.e., they communicate meaning but without using linguistic—and therefore categorical or discrete—structures. However, this would imply that nIE speakers use pitch paralinguistically (for question forms at least) while speakers of other varieties use pitch linguistically. This in turn seems to imply a potentially major typological distinction, which feels intuitively implausible. A more plausible explanation may be that there exists a separate (linguistic) register tier with two discrete morphological units (an unmarked low and a marked high) which the speaker employs to shift overall pitch range upward under certain grammatical conditions, most obviously in binary and declarative question forms. Do note that evidence for a register tier would not rule out the presence of paralinguistic pitch effects nor would it suggest that phonological register 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 (Knight, 2008; Knight & Nolan, 2006; Nolan & Farrar, 1999; Xu, 1999b inter alia). 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, due to the use of an asterisk to identify a tone associated with lexically stressed syllable (Pierrehumbert, 1980). In nIE, however, L*H dominates in the nuclear contour of unmarked declaratives, with L being the starred tone (e.g., Grabe et al., 2000). 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.

The main approach adopted for the bulk of the analysis in this thesis can be described—on reflection—as a Phonology-first approach, which might also be viewed as a top-down approach. It assumes that there is well-recognised set of phonological pitch events and pitch accents which can be labelled using an 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 adopted here involves 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.

As a response to the reflection on the underlying Phonology-first assumptions, an alternative is also proposed, here called a Phonetics-first approach, which might also be described as a bottom-up approach. It takes a more rigid view of the tonal target, identifying it explicitly with an ideal covert target which can never be fully manifested in the pitch contour due to physiological constraints on the extent to which the laryngeal apparatus can control rate of change in the vibration of the vocal folds. While tonal targets are not fully realised in the pitch contour, they are closely associated with turning points (or elbows) in the contour, which represent moments where the f_0 is shifting most rapidly from one turning point trajectory to another. This analysis also requires that the concepts of tone and tonal target be clearly distinguished, with the tonal target representing the static ideal target of the pitch trajectory and the tone a feature with inherent duration[2].

The first pass of the Phonetics-first analysis involves identifying the minimal number of H and L turning points required to produce a resynthesis of the contour that is perceptually identical to the original. Having identified the turning points, they are then associated with intonational events and structural units, such as pitch accents and boundary tones. A brief analysis using the Phonetics-first approach revisits some of the data from the main body of research in the thesis to identify ways in which this it might help account for some of the findings in relation to intonational form and function in Derry City English.

The issues raised above have been distilled into four research questions (RQs) outlined in Table 1.1. Note that the first two RQs focus on descriptive concerns and the second two on theoretical concerns. Overall, it is hoped 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 as understood within an AM framework.

[2] I have to admit that a failure to make this distinction may well simply be a result of my own flawed understanding and not a systematic issue in AM analysis, but it still seems worth highlighting.

Table 1.1 Research Questions

Descriptive Concerns

RQ1. What are the formal phonological and phonetic characteristics of pitch accents in DCE in unmarked speech as an effect of changes in metre (anacrusis and foot size) and lexical word boundaries?

RQ2. What are the phonological and phonetic characteristics of the intonational phrase and the nuclear pitch contour in DCE as an effect of different sentence modes?

Theoretical concerns

RQ3. Is there evidence in the DCE for the special status of H tones?

RQ4. Does a register-tier analysis provide a plausible explanation for phonetic variation across sentence modes in DCE?

In order to answer the research questions, the dissertation has been organized as follows. Chapters 2 and 3 provide an overview of the literature while Chapter 4 shows how the research questions lead from previous research and how, in answering them, this thesis might contribute in an albeit small way towards a greater understanding of nIE intonation and of intonation in general. 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 intonational form (RQ1) and also offers insights on the status of H targets (RQ3). Chapter 7 analyses the function of intonation in sentence modes (RQ2) and engages with the issue of a phonological register tier (RQ4). Chapter 8 offers a critical reflection on the two main pieces of research, introducing and contrasting the Phonetics-first approach with the Phonology-first approach. As such, this chapter does not answer a separate research question but, rather, considers how the alternative approach might contribute towards answering the questions listed in Table 1.1. The final chapter provides a short summary of the research, reflecting on its value in the context of intonation and nIE studies, and suggests directions for future research.

2 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 in Intonation studies in English, namely the British Tradition (Halliday, 1967; O'Connor & Arnold, 1973) and Autosegmental Metrical Phonology (Goldsmith, 1976; Gussenhoven, 2004; Ladd, 2008; Liberman, 1975; Pierrehumbert, 1980). These are outlined in Sections 2.2.1 and 2.2.2 respectively. It then delves into more detail on the Autosegmental Metrical approach (Sections 2.3 and 2.4), as this is the framework adopted in the thesis. The discussion of the Autosegmental Metrical approach (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 of tone and intonation, and its phonetic implementation.

2.1 What is Intonation?

Following Ladd (2008, p. 4), intonation is understood here as the post-lexical linguistically structured use of pitch to convey meaning. 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 tone (e.g., in Chinese or Ewe), 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) which are in turn incorporated into meaningful intonational events referred to as *pitch accents* (Bolinger, 1986; Cruttenden, 1997; Crystal, 1972; 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). Finally, intonation is meaningful in the sense that it forms part of a language's Grammar[3], helps signal affect, and contributes towards information structure in discourse (Cruttenden, 1997; Halliday, 1967; Tench, 1996).

Intonation needs to be considered independently of other prosodic features with which it is sometimes confused, particularly 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, vowel quality, or voice quality. Stress refers exclusively to lexical stress, which is viewed here as a component of the lexical word encoded into 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 (Ladd, 2008, Chapter 2).

[3] Note that Grammar, writ large, is used in the broad sense, referring to all linguistically structured components of a language and of language in general, not in the narrow sense of the morphosyntax alone.

Like stress, pitch accents are associated with prominence, but they are not simply manifestations of prominence. As Ladd (2008, pp. 50–54) puts it, pitch accents provide *cues to prominence*. That is, they do not constitute the prominence itself, but they often coincide with prominent syllables and thus help the listener identify prominence. Ladd contrasts this with analysis conducted by the Institute of Perception Research (IPO) in which pitch is viewed as *lending* prominence to a syllable, i.e., it is constitutive of prominence (’t Hart et al., 1990, p. 96). Effectively, this understanding allows the linguistic function of the pitch accent to be viewed independently of prominence signalling.

2.1.1 Functions of Intonation

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

In English, the grammatical function is often 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 (Jarman & Cruttenden, 1976; Lowry, 2002; Sullivan, 2012).

The discourse functions can be understood in terms of information structure and speaker interaction. For example, 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 (Brazil, 1995; Cruttenden, 1997).

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 not be accented. 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 to signal focus, this does not preclude the use of prosodic focus marking (Tench, 1996).

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 speaker inferences about listener knowledge and the extent to which they share the same information. As such, functions of intonation have also been generalised into abstractions, so that distinctions between grammatical and discourse functions may be collapsed into more abstract categories.

While intonation can be analysed in terms of an expansive range of functions (e.g., Bolinger, 1989), others—such as Brazil, Gussenhoven, and Cruttenden—offer broader 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 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 a (quasi-)universal link between the pitch movement and meaning. Gussenhoven (2004), however, observes 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.

2.1.2 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 voiced speech), thus explaining why it is the main parameter used in the acoustic analysis of pitch. Even though f_0 is very strongly correlated with pitch, it is only a measurement of the first harmonic in the rich spectrum of voiced speech, with harmonics repeated throughout the spectrum at multiples of the fundamental frequency. 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 (Ladefoged, 1996).

Alongside intentional variation associated with intonation, f_0 is also subject to segmental (microprosodic) effects from consonants and vowels, which are generally viewed as a non-intentional component of the f_0 contour (Di Cristo & Hirst, 1986; Ohala, 1976; Reichel & Winkelmann, 2010), an effect which is likely universal (Whalen & Levitt, 1995). However, the non-intentional view of microprosodic effects is not universally accepted. Kingston and Diehl (1994), for example, argue that microprosodic effects may be controlled for linguistic enhancement purposes. It is also known that tones in some tonal languages evolved from the loss of syllable-onset voicing contrasts while the associated f_0 effects remained (Michaud & Sands, 2020). However, here we are concerned with microprosodic effects in relation to languages without lexical uses of pitch.

Because they are not typically viewed as intentional, microprosodic variations in f_0 are typically disregarded in the synthetic modelling of f_0 intonation contours. However, when they are integrated into synthetic models, they have been found improve the perceived naturalness of the resynthesized speech although the microprosodic effects themselves may be barely perceptible (Krug et al., 2021; Peter Birkholz, 2020). Thus, while microprosodic f_0 perturbations may be salient at some level, they are not, for the purposes of the current research, considered relevant to the analysis of intonation.

The SI unit for frequency is Hertz, which measures the number of oscillations per second of a periodic waveform. However, Hertz may not be ideal for representing pitch variation in a way that reflects human pitch perception. For example, in music, each doubling of f_0 represents an increase of one single octave, which—in most Western music traditions—is subdivided into twelve semitones, with an equally proportional difference between each semitone. As such, a semitone scale is a logarithmic (log 2) rather than a linear one. The Mel scale, proposed by Stevens at al. (1937), offers a non-linear perceptual scale based on experiments testing the perceived differences between (to the best of my understanding) different sinusoidal tones, while the Bark scale, introduced by Zwicker (1961), also offers another non-linear psychoacoustic scale based on perceptually equal differences across critical bands in the frequency spectrum. The use of Equivalent Rectangular Bandwidths (ERBs) provides an updated version of the critical bands approach (Moore & Glasberg, 1996).

Research into which scale is most appropriate for representing pitch perception (rather than bands within the frequency spectrum) has not been conclusive. For example, Rietveld and Gussenhoven (1985) found that prominence judgments may be more appropriately represented in Hertz than in semitones, while Hermes and van Gestel (1991) found that ERBs best reflected perceptual equivalence of rise-fall contours at different pitch (f_0) heights. In contrast to each of these, Nolan (2003) found that logarithmic (semitone) or quasi-logarithmic (ERB) frequency scales better reflect intuitions about pitch equivalence in an imitation experiment.

Of course, f_0 is only one—albeit a very important one—of a rich combination of harmonic frequencies contributing to our perception of pitch (Houtsma, 1995), so the perceptual equivalence between spectral frequencies reflected in the Mel scale, Bark scale, and in ERBs may still not reflect perceived differences in *pitch*. Nor too, for that matter, might the logarithmic semitone scale, even though it is fundamental to the understanding of differences in pitch in Western musical scales. Despite these caveats, the semitone (ST) scale is preferred here, based, in part, on the assumption that even a slight familiarity with Western musical conventions will facilitate the interpretation of f_0 in the analyses which follow.

2.2 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 (Cruttenden, 1997; Halliday, 1967; O'Connor & Arnold, 1973; Tench, 1996) and the

Autosegmental Metrical (AM) approach (Gussenhoven, 2004; Ladd, 2008; Pierrehumbert, 1980). After a brief overview of both approaches in Sections 2.2.1 and 2.2.2, Sections 2.3 and 2.4 focus on the AM approach in more detail since it is the one adopted in this thesis.

2.2.1 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 unstressed 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 2.1 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 in the British Tradition fuses prominence, pitch, and lexical stress features and analyses them on a syllable-by-syllable basis. 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 the 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. This approach is illustrated by the interlinear tonetic transcription, also known as a tadpole diagram, shown in Figure 2.1. The figure represents a fall-rise nuclear contour, where the high falling nucleus occurs in *thought* and is identified by the large black dot with a falling tail, followed by 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.

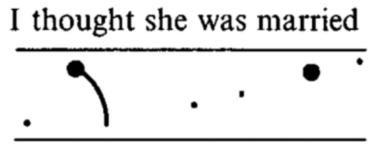


Figure 2.1 Example of interlinear tonetic transcription of intonation. (Cruttenden, 1997)

One criticism of the British tradition is that by dividing the intonation unit into a distinct hierarchical structure, it creates a level of 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 (see Section 2.4).

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.

2.2.2 The Autosegmental Metrical Approach to Intonation

The Autosegmental Metrical (AM) approach to intonation has its origins in work by Liberman (1975), Goldsmith (1976), 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[4]. The string of tones occurs within an intonational phrase (IP) and may also be marked on the initial or final edge tones at prosodic boundaries. Pitch accents are “associated with prominent syllables [typically in lexically stressed words] in the segmental string” (Ladd, 2008, p. 44). This means that there is a separate tone tier which is independent from but still coordinated with the segmental string. Moreover, there can be a

[4] The term tonal event is used to refer to pitch accents, phrase accents, and boundary tones, which are all landmark events within the intonational phrase.

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 phenomena independent from stress and prominence; however, as their location is associated with stressed syllables, they may serve as cues to prominence, as outlined Section 2.1. Again, this facilitates an analysis of the intonational structure of a phrase independently of the segmental string and stress.
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 mentioned above, 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 tones 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)[5]. 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 other AM-based studies, especially those on nIE, such as Lowry (2001, 2002, 2011), Dorn (2006) 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 an experimental approach in which target variables can be isolated and manipulated[6]. Finally, viewing an intonation contour as a string of phonological tones helps reduce much of the noise in the signal 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

[5] I am not totally convinced that global downstep can wholly account for declination. I think that a trace of declination effects may be seen in the slight downward slopes of plateaux and valleys in the contour unless overridden by a phonological imperative (see also Section 8.6.3, p. 196). Liberman and Pierrehumbert, however, make the point that analyses of declination effects are "uninterpretable" unless events such as downstep and final lowering (in IP-final pitch accents) are also taken into account (Liberman & Pierrehumbert, 1984, p. 224).

[6] 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.

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.

It has been observed that inventories of the British tradition can largely, if not completely, be converted to AM-style labels and vice-versa, as noted in Roach's (1994) discussion of the (British Tradition) annotation of the Spoken English Corpus and in Ladd's comparison of Pierrehumbert's 1980 inventory with potential corresponding British Tradition alternatives (Ladd, 2008, pp. 90–92). Despite a certain degree of inter-changeability in descriptive inventories, however, it is important to bear in mind that the two approaches reflect different assumptions about the underlying nature of intonation.

2.3 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 monographs by Ladd (2008), Cruttenden (1997), and Gussenhoven (2004).

2.3.1 Pierrehumbert (1980)

Pierrehumbert's (1980) doctoral thesis is a seminal work in the AM approach to 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 that 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 and 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 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. A plus sign is used to link leading and trailing tones to the pitch accent with which they are associated. The macron is also used for such 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 at all, and phrase accents are indicated with a hyphen instead. The plus sign is still in use, but not in all AM labelling systems or analyses (see Section 2.3.7).

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

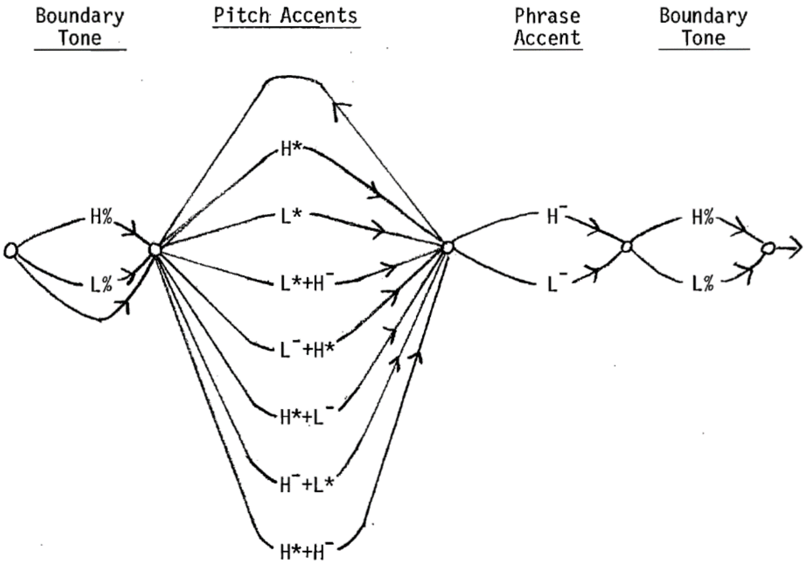
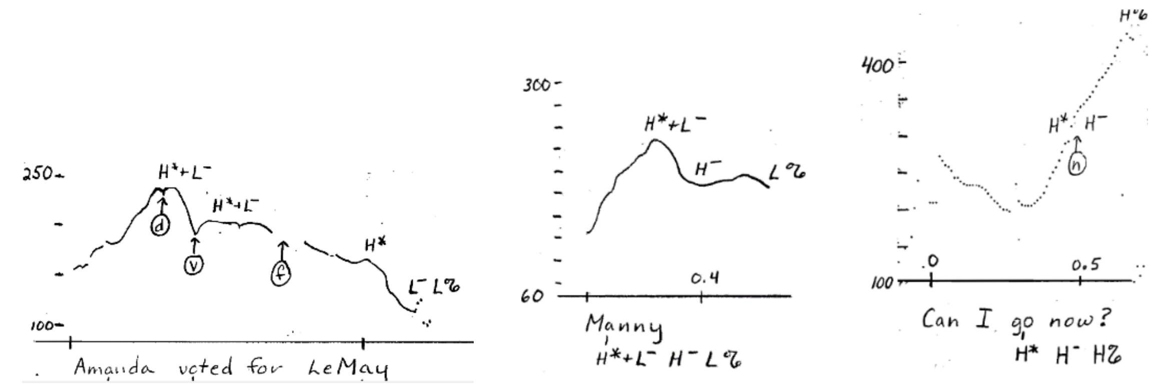


Figure 2.2 Pierrehumbert's (Pierrehumbert, 1980) finite state grammar of the intonational phrase.

In addition to the structural component of the PA and its tune-to-text associations, Pierrehumbert (1980) also posits rules that govern 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 systematic lowering of the phonetic realisation of an H tone as an automatic consequence of an intervening L tone, 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 upstepped; however, the L% essentially becomes null in such cases, representing no or only slight lowering of f_0 in the contour. The contrast between H¯L% and H¯H% can be seen clearly in Panels B and C of Figure 2.3.



A. B. C.

Figure 2.3 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 three different phenomena. That is, because there is no lowering associated with L% boundary tone, it is the L¯ phrase accent which causes the lowering of f_0 shown in Panel A.

Finally, Pierrehumbert accounts for stretches of contour between boundary tones and pitch accents by viewing them as linear interpolations between pitch events. Thus, pitch glides are a side-effect of the phonetic implementation of tones in the f_0 contour.

2.3.2 Critique of Pierrehumbert's (1980) Upstep and Downstep

Several problems 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 (Ladd, 1983), 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 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, not as an obligatory rule, but as a speaker-motivated event (labelled as !H); i.e., something which the speaker choses to do rather than the deterministic consequence of an implementational rule.

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 (1998b, 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, and L% always indicates a relative lowering of the f_0 contour at the boundary, 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 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.

2.3.3 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 nucleus 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 those preceding as prenuclear pitch accents (Ladd, 2008; Nolan, 1997). This also 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 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 organized within a metrical hierarchy. After all, this is still the underlying view that led to the Autosegmental Metrical approach.

2.3.4 Gussenhoven

A different approach to AM analysis is presented by Gussenhoven (1983, 2004, 2016). 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[7], 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 (L_ι, H_ι) is obligatory but—similar to the Grabe approach—the final IP boundary is optional (L_ι, H_ι, Ø_ι) and the height of the boundary tone 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 obviates the needs for a phrase-accent interpretation of the contour.

[7] 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 Figure 2.4. 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 (Gussenhoven, 2004).

4. **Double alignment**. Initial boundary and monotonal pitch accents extend rightwards towards the onset of the next pitch event, so that each L_ι, H_ι, 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 T_ι. Final T_ι is not double aligned, and only represents a single event (where T refers to tone, either high or low).
5. **Intra-accentual interpolation.** Interpolation of the f_0 contour only occurs inside the pitch accent, not across pitch accents. This is 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 accommodates both Ladd's speaker-motivated optional downstep and Pierrehumbert's obligatory downstep.
8. **Tri-tonal prenuclear pitch accents**. Prenuclear 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 particularly significant, what is note-worthy is that prenuclear 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 the scooped rise, sometimes labelled L+H*, and the unscooped H* (Ladd, 1983). 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 2.3.7 below).
10. **NOSLUMP**. This is an obligatory feature of SSBE, which prevents the final L_ι 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 Figure 2.4. 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" prenuclear

H*LH, Gussenhoven argues, is likely an artefact of an earlier IP form H*L H_ι (2004, pp. 302, 305–306).

$$\text{(1)}\quad \underset{\text{NOSLUMP}}{([\text{DOWNSTEP}])} \begin{Bmatrix} H_\iota \\ L_\iota \end{Bmatrix} (L) \begin{Bmatrix} H^*(L(H)) \\ L^*(H) \end{Bmatrix}_0^n (H+)(L) \begin{Bmatrix} H^*(L) \\ L^*(H) \end{Bmatrix} \begin{Bmatrix} H_\iota \\ L_\iota \\ \emptyset \end{Bmatrix}$$

Figure 2.4 Gussenhoven's (2004, p. 313) intonational grammar of Standard British English

While the formalization of the intonational 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 realised.

2.3.5 Linguistic and Paralinguistic Uses of pitch

The matter of linguistic versus paralinguistic use of pitch is important in the study of intonation. This is largely due to the fact that while intonation appears to have an internally structured phonology, it is manifested in the f_0 contour, which is affected by a range of other non-linguistic components, such as individual physiology, attitude, and psychological state. 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, paralinguistic use of pitch 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 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 aspects of pitch and its paralinguistic use. 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 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 (2004) argues that there are three biological codes which do indeed motivate the apparent aforementioned universality of pitch movements. The two additional codes 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 over time 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 and which 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 phonological units may change their associations and thus subvert the putative universality of the 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 is 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.

Alongside 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 type (L+H* v H*) to indicate different kinds 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 can 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. (Gussenhoven, 2004, pp. 60–60, in reference to Smiljaníc and Hualde, 2000.)

2.3.6 Phonological Explanations for Apparent Phonetic Scaling Effects

Several alternatives have been proposed which incorporate changes in scaling into the phonology. Looking at intonation and question forms in Spanish, Sosa (1999) argues that there is an optional initial H boundary event (%H) which raises the first stressed syllable of binary questions raises. This initial-high raising, in turn, permeates across the whole utterance, leading to an overall increase in f_0 across the question. This approach is still compatible with the view of intonation in terms of the tonal string alone. Alternatives, which call upon additional components have also been proposed.

Ladd (1990) observes that prenuclear pitch accents in English may be nested within an utterance in such a way as to give rise to apparent resetting of downstep with a single utterance, while differences in pitch accent scaling might be explained phonologically in terms of high-low and low-high relations at a higher level in a metrical tree. Elaborating on this more broadly (Ladd, 2008), he argues that variation in pitch height of semantic constituents in an utterance—such as those in the Manny/Anna study (Liberman & Pierrehumbert, 1984)—may be explained phonologically if we accept that intonational constituents can be compounded and so can also be embedded (and thus subordinated) hierarchically within larger constituents. This is particularly relevant to the role of intonation in information packaging (e.g., identifying new versus given information). The argument for a phonological explanation of apparent phonetic effects is further developed in Ladd and Arvaniti (2023), specifically with regard to prominence relations.

Another approach aimed at explaining changes in register involves the inclusion of a register tier as well as a tonal tier. This argument relates specifically to the scaling of high and low tones in African tone languages (c.f. Snider, 1988, for a summary of different approaches). This register-tier approach aims specifically to provide an explanation for observed downstep and upstep effects and to explain why phonologically low tones may in some cases be realised at the same pitch as phonologically high tones, and vice versa. The following paragraphs summarise two different views on how the register tier is understood to operate.

Snider (1988, 1990) presents a three-dimensional Autosegmental model to explain scaling effects phonologically. He views the Register Tier[8] as a sequence of highs and lows, with each successive high or low associated with a relative increase or decrease in the overall scaling of tones in the Modal [Tone] Tier. As an analogy, the Modal [Tone] Tier can be viewed as like standing or sitting on a platform to raise or lower one's height, while the Register Tier is akin to a progressive lowering or raising of the platform itself. Events in both the modal and register tiers are linked to the Tone-Bearing Unit (TBU) Tier via a Tonal Node (or Architonal) Tier. In Snider's representation, each high or low tone in the Modal [Tone] Tier is, by default, associated with a High or Low in the Register Tier (see Figure 2.5). In such a scenario, H tones are realised high in the speaker's pitch

[8] In the Africanist literature, tier names are capitalized. This convention is retained for the current discussion.

range and L tones low. However, the application of a sequence of phonological rules leads to variation in the realisation of Tones in the pitch contour, including Tone Spreading, Tone Insertion, and Stray Erasure (i.e., phonological deletion of tonal features de-associated with the TBU tier as an effect of previous processes). Within Snider's approach, since the register tier is relational, the association of the initial high or low in the register tier is essentially an arbitrary speaker decision.

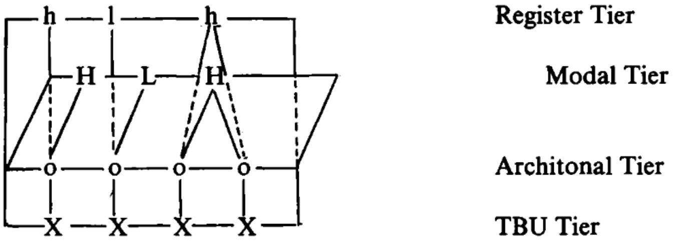


Figure 2.5 Three-dimensional representation of tone in Snider (1988, p. 245).

A different kind of Register Tier is proposed by Leben, Inkelas, and Cobler (Inkelas & Leben, 1990; Leben et al., 1989). Again, the register tier is understood to consist of high and low values, and—similar to Snider's approach—the register tier and the Tone Tier (called the Primary or Lexical tier in this case) are linked via a tonal node. In this case, however, the low of the register tier is an unmarked form while the high of the register tier is marked. Moreover, unlike the Snider's approach, changes in register do not lead to the relative raising or lower of the overall register, and there is no default association of High tones with High register and Low tones with Low register. Rather, the presence of the Low register is associated with the sequential downstepping of tones in the Tone Tier, while the presence of High register is associated with the suspension of downstep and the raising of the realisation of High tones in a High-High sequence. This means that the Highs and Lows of the register tier are expected to spread across a sequence of tones. Using this approach, Inkelas and Leben (1990) argue that, in Hausa, it is the presence of High register in binary questions which leads both to the suspension of downstep and to the raising of the final H tone which often occurs in binary questions. They also observe that, in emphatic structures, High register explains the use of an upstepped realisation of the High tone.

Snider's approach to the description of Register Tier effects is concerned with phonological rules governing the realisation of lexical tone sequences independently of intonational features. However, the description provided by Leben, Inkelas, and Cobler was developed to take into consideration both lexical and intonational uses of pitch. In an intonation-only language, such as English, therefore, it is possible that a Register Tier of the kind proposed by Leben, Inkelas, and Cobler could provide a phonological account of apparently paralinguistic-only variation in f_0 across sentence modes. This point is recognised but ultimately rejected in Haan's (2002) account of Dutch question intonation. Of course, in Dutch, the Tone Tier provides enough phonological material—i.e., the final H% boundary in questions— to help discriminate between statements and questions.

2.3.7 Labelling in Contemporary AM Analyses of English Intonation: ToBI[9] and IViE

It is clear that the labelling of PAs and boundary tones within an IP should 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) puts it—whether the analysis of the pitch accent follows an on-ramp or an off-ramp approach. Broadly, the phrase accent on-ramp approach is mostly followed in ToBI (Tone and Breaks Indices) labelling and labelling systems derived from it (Beckman et al., 2005; Beckman & Ayers Elam, 1997). 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 in Northern Ireland (Grabe et al., 1998; Gussenhoven, 2004), and the labelling system shares some principles 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 theory-neutral approach to downstep triggers, and uses downstepped H tones (!H* and H+!H*) in pitch accents (Beckman et al., 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 movement of 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 (2004), although he permits tritonal pitch accents only in prenuclear

[9] 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 now 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 not directly of concern for the approach used in this thesis.

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 (Grabe et al., 1998). (See also Chapter 3, Sections 3.2 and 3.3.1). Moreover, it allows for a more parsimonious labelling system which can facilitate the identification and comparison of structurally identical pitch accents and boundary structures across different varieties of English.

2.3.8 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 f_0 scaling and its temporal alignment in relation to a segmental landmark, frequently the onset of the vowel in the lexically stressed syllable (Arvaniti, 2022).

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 structural features are also understood to affect f_0 scaling and temporal alignment of tonal targets. These include compression and truncation, tone drift, and tonal crowding.

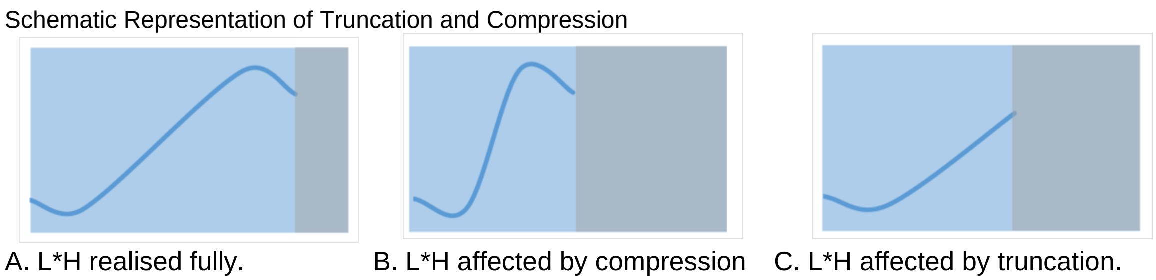


*Figure 2.6 Three Stylisations of L*H pitch accents shown truncation and compression. The grey area on the right of each box represents unvoiced segmental material or a phrase-final pause.*

It has been argued that a pitch contour may be truncated or compressed as a function of available voiced material leading to different surface realisations of underlyingly identical pitch events, a phenomenon first noted in Grønnum (1990) and taken up by others (Grabe, 1998a, 1998b; Ladd, 2008 inter alia). In the case of compression, the tonal target may be fully realised but within a shorter time window to compensate for the reduced amount of voiced segmental material, as shown in Figure 2.6, Panel B. With truncation, however, the pitch movement may be foreshortened, as represented in Panel C, so the tonal target is not realised fully. It is also possible for pitch contours to be subject to truncation and compression 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 occur in one variety will be found in another (Grabe, et al., 2000).

As regards timing, it appears that tonal targets may be fixed or subject to drift. When a target is fixed, it is typically anchored to a specific landmark in the phrase, such as a stressed syllable, the foot, or sometimes with lexical boundaries (Prieto, 2011; Prieto et al., 2005). However, it has been found that, given enough segmental material, and conditioned by foot size and anacrusis, a tonal target may drift rightwards or leftwards(Gussenhoven, 2004; Pierrehumbert & Steele, 1989; Silverman & Pierrehumbert, 1990). Analysis of variation in peak alignment strategies has shown that they can be a marker of dialect differentiation, both in English (Arvaniti & Garding, 2007; Kalaldeh et al., 2009) and other in languages (Atterer & Ladd, 2004; Bruce & Thelander, 2001; Dalton & Ní Chasaide, 2007).

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

If one were to analyse an f_0 contour without consideration of truncation or compression, tonal drift, and tonal crowding, it might be possible to misidentify fundamentally identical pitch accents as different. Through systematic analysis of pitch accent alignment 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 crowding affects the realisation of tonal targets. Such issues strongly inform the analysis in Chapter 6.

2.4 Issues for AM Analysis of the Phonology and Phonetics of Intonation

So far, the main difficulties discussed in relation to AM analysis have focused on theoretical divisions motivating differences in 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.

2.4.1 Tonal Targets and Implementational Domains

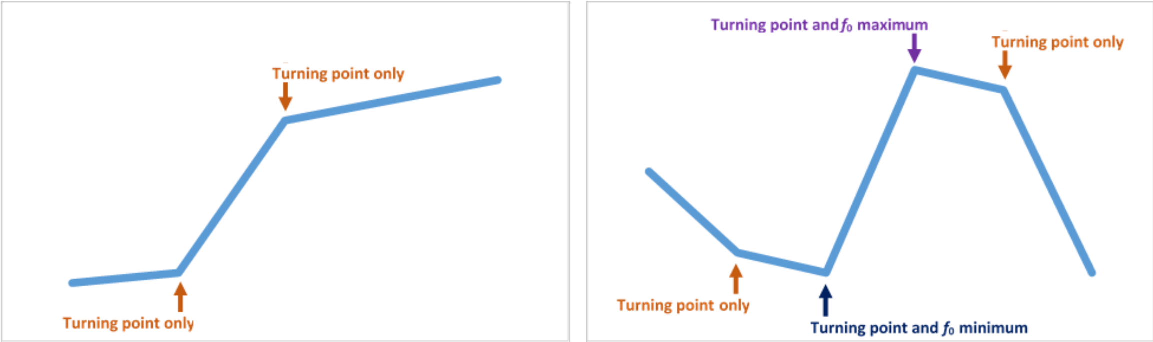
Within the AM approach, the domain of the starred tone is, by definition, the metrically stressed syllable since the star simply represents the stressed syllable which links the segmental string to the tonal sequence. There are, however, several problems with this assumption, one of the most significant 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 prenuclear 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. 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. Finally, it could be pragmatic, in that 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 extent of the perceived need to realise the pitch accent precisely.

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 the lexical word in which foot stress occurs rather than the foot alone (Silverman & 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 & Schepman, 2003; Welby, 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 stylised example of this is shown in Figure 2.7. These points are referred to as *elbows* or *turning points*, and have been used to identify L targets in several studies (D'Imperio, 2000; Shosted et al., 2006; Welby, 2003). In fact, the use of turning points can be generalised to interpret all tonal targets in this manner. Frequently, as in the Figure 2.7B, 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 8, Section 8.5, which outlines a technique for identifying turning points.

Schematic Representation of Turning Points



A. Turning points only. B. Turning points and f_0 extrema

Figure 2.7 Stylised representations of an f_0 contour, indicating turning points (in orange), and f_0 peak (purple) and f_0 minimum (blue). Note, the f_0 extrema in these stylized examples are also turning points.

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

2.4.2 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, and in analysis of question types in Dutch, German, and Greek, to name just a few (Arvaniti & Ladd, 2009; Grabe, 1998b; Grice et al., 2000; Haan, 2002). In English, this can be seen around the discussion of question modes, calling contours, and focus (Ladd, 2008, has an excellent overview of this). However, as noted above in Section 2.3.5, pitch contours sometimes vary according to function in a manner which cannot easily be 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. Unfortunately, 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 prioritised 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 a 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 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 perform a phonological analysis which dismisses some phenomena as paralinguistic 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. This is central to the analysis of sentence mode in Chapter 7.

2.4.3 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 2.8 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, as in Panel B, TCoG-f_0 is higher and TCoG-t is earlier. Conversely, whenever the glide to the peak is more scooped, as in Panel C, TCoG-f_0 is lower and TCoG-t is earlier. These effects are reflected by the different positions of the dots in Figure 2.8.

Schematic Representation of Pitch Curves and Tonal Centre of Gravity

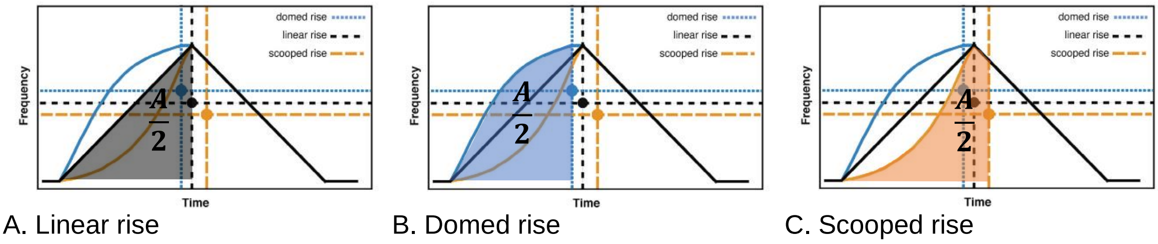


Figure 2.8 Schematic representations of different pitch curves with different TCoG. Adapted from Barnes et al. (2021) to illustrate how TCoG-t reflects timing of the balance of the area under the curve.

In their 2012 paper, the authors were able to demonstrate that the TCoG-t was a better predictor of annotators' PA categorisation 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.

This train of thought is not merely panicked catastrophizing, since the TCoG analysis does indeed appear to reflect listener's ability to categorise pitch contours without recourse to tonal targets or the underlying tonal primitives 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. Retrospective calculation 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* occur in 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 way 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 appear so effective simply because it reflects the way in which tonal targets are realised, i.e., 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 concern persists in a more recent perception study on TCoG (Barnes et al., 2021). This paper reports the results of experiments which used synthesized f_0 contours in which low f_0 targets were measured both at the point where the rise began and at the point at which it was complete. However, considering that the scooped rise begins subtly and gradually (unlike the domed rise where the f_0 change is sudden and dramatic), there is no reason to assume that this mirrors how tonal targets are manifested in real speech. It still seems plausible that it is in fact the elbow which should be identified as representative of tonal targets.

The conclusion one might draw from the 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 is given fuller consideration in the critique in Chapter 8.

2.5 Summary and Conclusion

The dominant mode of intonation analysis in English is the Autosegmental Metrical (AM) approach, and it has been employed for this thesis. In essence, there are two main schools of thought in AM research. The first is a direct descendent of Pierrehumbert's original (1980) approach, and has been propagated though ToBI and its offshoots (Beckman et al., 2005). The second is more closely aligned with the work of Gussenhoven (1983, 2004), Grabe (1998a), and the IViE project (Grabe et al., 1998; Grabe & Post, 2002). The distinguishing features of 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, they are both still influenced by competing theoretical interests.

Table 2.2 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 via metrical structure. The key difference lies in how the events in the tonal string are associated with structural elements of 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.

2.5.1 Labelling Choices for Research

ToBI-style analysis requires the labelling of PA-final edges in terms of both phrase-accents and boundary tones, even in cases where there is little evidence for the phrase accent. The solution adopted by Grabe, Gussenhoven, and the developers of the IViE project feels intuitively more appealing, largely because it does away with the phrase accent and allows for an optional final boundary (L, H, 0) in its stead. 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 is preferred in this thesis. After all, IViE was designed for labelling different 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 developers of IViE also argue that its more flexible approach to final boundaries makes it more amenable to the analysis of northern Irish English (Grabe et al., 1998). There is also no phrase accent in IViE, which means that there is no need to include labels for tones for which there may not be sufficient empirical evidence. However, it should be noted that it was still deemed necessary to make some adjustments to the IViE labelling system in order to adequately account for some of the features observed in the data. These adjustments are discussed in the Phonological Labelling sections of Chapters 6 and 7, Sections 6.4 and 7.4 respectively.

2.5.2 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. There were two core reasons for this. Firstly, minima and maxima are intuitively easier to reconcile with concepts of High and Low than turning points. Secondly, they are much easier to identify and catalogue 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.

2.5.3 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—i.e., lexical boundary and metrical—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 boundary effects on pitch accent inventories and the alignment and scaling of tonal targets in DCE.

Section 2.4.2 pointed 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 was suggested that one must consider the limitations of the theory itself and see if reasonable accommodations can be made to adapt the theory so that it can explain systematic patterns in the data more completely. Chapter 3 will indicate how a strict AM analysis of sentence modes in nIE might lead to phonological distinctions being identified as paralinguistic. Potential adjustments to the phonology are considered in the analysis of sentence mode in Chapter 7.

3 Local Context: northern Irish English and Derry City

This chapter focuses on Derry City and northern Irish English, outlining geopolitical and linguistic contexts, and summarising 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 and moves on to discuss British Tradition analyses of Belfast and Derry City intonation before turning to AM studies of nIE intonation. It ends by considering how the issues within AM discussed in the previous chapter (Section 2.4) relate to the current study of DCE intonation.

3.1 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 (Northern Ireland Statistics and Research Agency, 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 royal charter of James VI, the site along the West Bank of the river Foyle was part of Donegal (Lacy et al., 1983), and there still continues to be 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 to identify as British (Zwickl, 2002, pp. 72–101). The city has a large Roman Catholic majority (78% in the 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 allowed 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 (Grabe et al., 2005)—is contrasted with English in the Republic of Ireland, in such a way as to imply that the political border and the linguistic borders coincide. To give just one example of this, Folley, Gibbon, and Peppé write that "[a]lthough statements have a high terminal in Northern Ireland […], this is not the case in the Irish Republic" (2011, 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 one volume on Northern Ireland and one on the Republic of Ireland, but both books exclude the English of Donegal, which is in the geographical the north but is part of the Republic of Ireland.

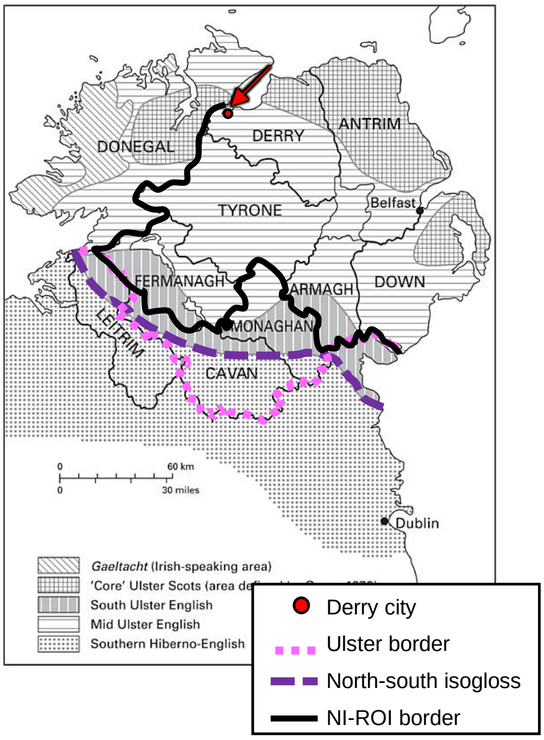


Figure 3.1 Approximate boundaries of northern Hiberno-English dialects (Hickey, 2007, p. 442 after Harris, 1985, p. 16). Derry City is highlighted in red.

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 actually 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 northern parts of the island.

As can be seen from Figure 3.1, Derry city is in the MUE region. Thus, one expects DCE to be broadly similar to the English of this region in general. In fact, one well recognised similarity across all nIE varieties is the use of rising nuclear pitch accents in unmarked declarative sentences (see Chapter 6). It is this specific feature of nIE which motivated this research, since, as shall be explained in Section 3.4, the prevalence of rising intonation patterns raises questions about the relationship between the form and function of intonation and about the linguistic/paralinguistic distinction in the AM approach to intonation analysis.

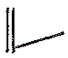
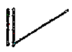
3.2 British Tradition Analyses of northern Irish English Intonation

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 a study of Belfast English, Jarman and Cruttenden (1976) found that 70% of all nuclei contained rising tones, including semantically unmarked phrases, i.e., neutral declaratives. 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 that the post-tonic section formed a plateau. Although the rise in 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 to that in Jarman and Cruttenden’s Belfast study, especially in the use of rising tones in unmarked forms. However, McElholm found no examples of nuclear falls and also found that a low-rising tone was used for wh-questions. McElholm suggests that such differences may be due to data sparsity or a difference 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 the Belfast study.

Table 3.1 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 & Cruttenden, 1976).

Tone	Stylisation	tonic movement	general meaning
A (1)		low rising	natural for all major speech functions except YNQs; also used to introduce secondary information[10]
B (2)		high rising	neutral for YNQs
C (3)		rising-falling	contrastive [preceding contrast or suggesting reservation]
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

[10] McElholm uses the terms secondary information to refer to information which is already given in the discourse. Secondary and new information are often described as topic and comment, or theme and rheme.

Like Jarman and Cruttenden, McElholm conducted his analysis in the British Tradition, largely via impressionistic analysis of two speakers (although McElholm did also include some instrumental analysis). The differences in theoretical approach and the limited amount of quantitative data make it is difficult to compare it with AM studies beyond a few broad phonological comparisons.

3.3 AM Studies of northern Irish English Intonation

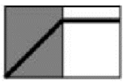
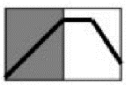
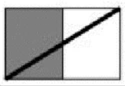
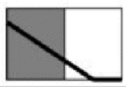
Within the AM approach, the majority of studies on nIE are based on IViE data, and thus focus on Belfast English. A few studies of Donegal English have been conducted based on data collected by O'Reilly and Dorn (Dorn, 2006; Kalaldeh et al., 2009; O'Reilly et al., 2010). The only other AM study of nIE 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 both Britain and Ireland (Grabe & Post, 2002), including Belfast. The Belfast corpus includes speech from 12 speakers (6 female, 6 male, 17 years old) who attended one of two schools near the city centre, all of whom had been born and had grown up in Northern Ireland. Five tasks were used to elicit a range of speaking styles, speech functions, and interactions.

3.3.1 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 notable of these was ToBI's inability to provide a label for the rise-plateau-slump of Belfast English. Later, using the IViE system, she identified four different nuclear contours in Belfast English, labelled L*H %, L*H L%, L*H H%, and H*L % respectively. These are shown in Table 3.2 (Lowry, 2002). These patterns have been attested in a number of analyses of the IViE Belfast data (Grabe, 2004; Grabe et al., 2005; Lowry, 2002) as well as in more recent studies (Jespersen, 2018; Sullivan, 2010, 2012). L*H % is by far the most common nuclear pitch contour across sentence modes. However, differences in the distribution of nuclear contours across speech style, gender, and sentence mode have also been found.

Table 3.2 Nuclear patterns of Belfast English, adapted from Lowry (2002).

Schematic representation				
Impressionistic description	rise-plateau	rise-plateau-slump	high rise	fall
IViE labelling	L*H %	L*H L%	L*H H%	H*L %

Lowry's (2002) study of style shift found that rising nuclei dominate all styles of speech and that 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 (2010) found L*H H% in DECs as well as YNQs. She also found that L*H L% was more likely to occur in question forms than in DECs. In a more recent study, Jespersen (2018) has also found occurrences of simple rises in declaratives.

Another study of the IViE data (Grabe et al., 2005) analysed PAs across utterances in both prenuclear and nuclear position. They found four different prenuclear accent types (H*, L*H, H*L, and L*), with H* being by far the most common, accounting for 78% of all PNs in declarative 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 Jarman and Cruttenden and in McElholm, one should expect a similar distribution of pitch accents, though with fewer—if any—instances of H*L Derry City English.

Table 3.3 Summary of pitch contours attested in Belfast English across sentence modes.

Contour	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 this phenomenon has been described as Urban Northern British intonation (UNBI) (Cruttenden, 1995, 2001, 2007). In the IViE corpus, there are four sets of 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 so, Newcastle only had 17% L*H in declaratives compared with 96% in Belfast (Grabe, 2004). Cruttenden (2007) also notes the evidence for the greater dominance of unmarked rises in Belfast English when compared to other varieties.

3.3.2 Alignment, Compression and Truncation in nIE

Nolan and Farrar (1999) used the IViE corpus to analyse peak lag in prenuclear pitch accents, which in their study 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[11]. Gender was also a significant factor, but sentence mode had only a limited effect.

A comparative study of nuclear and prenuclear alignment across varieties of Irish English (Kalaldeh et al., 2009) examined L*H PNs in Donegal English (as PNs were found to be exclusively L*H). The study found that anacrusis caused rightward drift of the prenuclear 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 prenuclear accents, L targets were stable.

*3.3.3 Proposed Source of Declarative L*H Dominance in nIE and Elsewhere*

Cruttenden (2007) speculates 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 does 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 more reasonable to suggest that the Scandinavian hypothesis is plausible for the declarative rise in Tyneside English but that the northern Irish migration hypothesis is more reasonable in 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 & Ní Chasaide, 2007; O'Reilly et al., 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 pitch 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.

[11] Sullivan appears to use the tern anacrusis to describe any sequence unstressed syllables preceding a stressed syllable not just those preceding the first stress in the IP. Thus, for example, trailing syllables in the first foot are termed as anacrusis in the relation to the stress in the second foot. However, here, I define anacrusis strictly as the sequence of unstressed syllables before the first stressed syllable in the phrase.

3.3.4 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 f_0-normalised data to facilitate the comparison of contours. The study found that, as with many other studies (including Sullivan, 2010, 2012), mean f_0 increased in question forms, with DCQs having the highest average f_0 in all dialects (see Figure 3.2). The authors were also able to show that the f_0 average and f_0 slope contribute to the distinction between sentence modes across dialects. This could be interpreted as a paralinguistic 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 differences in intonation across different functions.

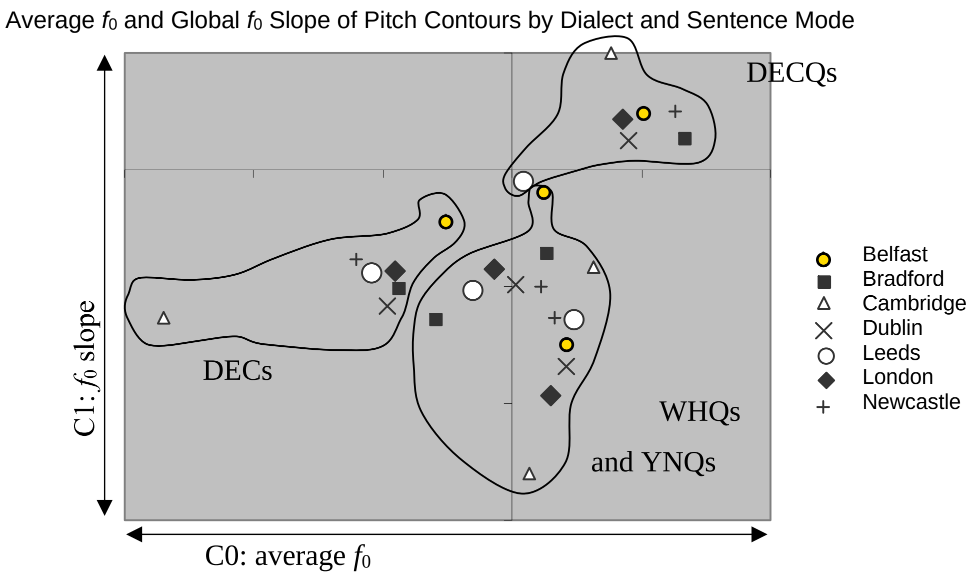


Figure 3.2 Average f_0 (x-axis) plotted against the global slope of f_0 (y-axis) for four utterance types and seven dialects. Adapted from Grabe, Kochanski and Coleman (2003)

3.4 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 with 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 how the issues raised in Chapter 2 might inform the research aims of a phonetic and phonological analysis of DCE.

3.4.1 Does nIE Provide Evidence for the Special Status of H Targets?

Aside from the study by Sullivan (2007), the vast majority of studies on tonal alignment in English deal with 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 overwhelmingly 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 that there was a greater amount of variation in prenuclear (PN) pitch accents.

Increased variation in PN pitch accents could indicate that prenuclear PAs also signal variation in meaning; however, given that nuclear pitch accents are more closely associated with communicative function than prenuclear accents, it seems unlikely that they would somehow contain more variation than the nuclear PA in order to signal a greater variety of meanings. It is more likely 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 this is indeed the case, it is very interesting to note that, in the IViE Belfast data, where they are attested, PNs are very likely to be H* and very unlikely to be L*. It 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 form of L*H, and that the H matters more than the L.

In the study of the Donegal English data discussed in Section 3.3.2 (Kalaldeh et al., 2009), 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-prioritisation 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*[L] 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 a 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 which has been discussed previously in Chapter 2, Sections 2.4.1 and 2.5.3, and which is the topic of Chapter 6.

3.4.2 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 marked YNQs. AM analyses of nIE varieties, however, have 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 discussed in Section 3.3.4, 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. 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, such an effect is generally accompanied by a phonological contrast between H*L % in DECs and L*H H% or L*H % in DCQs. In nIE, however, no such parity between continuous and discrete features has been identified.

The AM argument championed by Gussenhoven (see Section s 2.3.4 and 2.3.5) maintains that parasitic traces of pre- or paralinguistic uses of pitch remain in the pitch contour 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) must use pitch paralinguistically to distinguish between sentence modes—particularly in DECs and DCQs which are otherwise morphosyntactically identical—whereas speakers of other English varieties have access to categorical intonational contrasts as well. In other words, it implies that, functionally, nIE may lack a chunk of the intonational phonology available in 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, it 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 this premise, then we might wonder if the theory underpinning the inventory of available labels has not successfully managed to capture all the linguistic/categorical contrasts which exist.

Most AM approaches to intonation, especially those originating from Pierrehumbert (1980), employ a single tonal tier to explain the phonology of intonation and tone; however, as noted in Chapter 2 (Section 2.3.6), other proposals have been made which look beyond the tonal tier and which aim to account for variation in pitch scaling in phonological terms without recourse to the paralinguistic argument. Most relevant to current considerations is the argument for a register tier as proposed by Leben, Inkelas, and Cobler (Inkelas & Leben, 1990; 1989). While this was proposed as phonological explanation for variation in the f_0 height of tones in African tone languages, it takes into consideration both the lexical and intonational components of pitch, specifically in relation to question forms. While Haan (2002) recognises the possibility of a register-tier explanation in her discussion of pitch register shifts in Dutch question intonation, she ultimately rejects the register-tier hypothesis in favour of the paralinguistic argument.

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). 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 but *does* have control over which ones to use, much in the way a diner at an à la carte restaurant is free to choose any item from the menu but cannot order something that is not on the menu in the first place. Furthermore, all speech serves a communicative function, and that function is largely determined by speaker intent. If a set of features occurs consistently and repeatedly in association with another linguistic function (e.g., pitch-register raising in question forms), it therefore seems reasonable to assume that those features form part of the Grammar (whether phonological, syntactic, or otherwise), i.e., they might be understood as linguistic rather than paralinguistic. As such, the register tier seems to offer a reasonable means of explaining such features in the phonology.

If we allow for the possible existence of a register tier, this might well explain how nIE maintains a phonological distinction between DECs and DCQs. This would contrast with the view that nIE has access only to gradient features and context cues to aid in the discrimination between the two sentence modes. Therefore, one key object of this research is to assess the validity of a register tier hypothesis (Chapter 7).

3.5 Conclusions

This chapter has provided information on Derry City, northern Irish English (nIE), and looked at previous research into intonation in nIE varieties. It then considered how theoretical concerns regarding AM raised in Chapter 2 might influence the current study. One concern is formal and asks if there might be evidence for the special status of H tones, particularly in the realisation of PN pitch accents in Derry City English (DCE). The second is a functional concern. It leads to a question asking if an analysis of sentence modes in DCE might provide evidence for a phonological register tier. This could offer a phonological explanation for changes in pitch scaling across sentence modes which might otherwise be treated as purely paralinguistic (or gradient). Such a consideration is important, since—following on from arguments 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 do not fit the phonological model neatly.

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

4 Prospectus

The primary purpose of this chapter is to outline the research questions (RQs) this thesis aims to answer. However, a research project of this size is a long-term endeavour, and over time, one's understanding of the topic develops. This is especially true after having lived with the data for a long time. Therefore, the secondary aim of this chapter is to outline how my thinking developed.

4.1 Research Questions

The research questions are divided into two categories, descriptive and theoretical. The research was originally motivated by the descriptive aims, while the 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 intonation, the dominance L*H nuclear pitch accents as an unmarked form, as outlined in Chapter 3, Section 3.4. A tangential concern, however, arose after reflecting on the approach taken in the main body of analysis in Chapters 6 and 7. This approach, as mentioned in Chapter 1, I have characterised as a Phonology-first approach. An alternative approach, which I refer to as the Phonetics-first approach, is considered in the critique in Chapter 8.

4.1.1 Descriptive Concerns: Form and Function

In the description of Derry City English (DCE), 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 syllables in anacrusis and syllables in the foot. (Anacrusis is used strictly to refer only to the sequence of phrase-initial unstressed syllables which do not form part of a foot). The analysis of lexical boundary effects[12] is limited mostly to prenuclear (PN) pitch accents, partly because such effects on PN tonal targets have been observed previously (e.g., Prieto et al., 2005, 2010; Silverman & Pierrehumbert, 1990), and partly because lexical boundary effects on PN accents were observed during the analysis of pilot data during the corpus development phase (Chapter 5).

The functional role of intonation is assessed in relation to sentence modes, i.e., declarative statements, wh-questions, yes/no questions, and declarative questions. The description of the phonology and phonetics of intonation across sentence modes aims to establish a foundation for considering the theoretical question of a register tier outlined in Chapter 3, Section 3.4.2.

Following standard AM procedures, all analyses include a description both of phonological inventories and their phonetic implementation. The descriptive concerns of this project are articulated in the first two research questions:

[12] Lexical boundary effects are often described in this thesis simply as "lexical effects". This term should not be read as referring to anything other than the lexical boundary effects.

RQ1. What are the formal phonological and phonetic characteristics of pitch accents in DCE in unmarked speech as an effect of changes in metre (anacrusis and foot size) and lexical word boundaries?

RQ2. What are the phonological and phonetic characteristics of the intonational phrase and the nuclear pitch contour in DCE as an effect of different sentence modes?

RQ1 is answered by and large in Chapter 6, while Chapter 7 responds to RQ2.

As outlined in Chapter 3, Section 3.3, some of these descriptive concerns have been addressed in studies of 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 also noted in 3.1, Derry City is physically and socioculturally close to Donegal, but is an urban area in Northern Ireland, and so shares with Belfast the same governmental, political, educational, and social infrastructure. In answering the first two research questions, this research hopes adds to the body of knowledge of nIE and also indicate the extent to which DCE patterns with Belfast English and the extent to which it appears to pattern with Donegal English.

4.1.2 Theoretical Concerns: H Tones and Register Tiers

Two theoretical concerns were outlined in the final section of Chapter 3. The first relates to the status of the H tone. It was noted that L*H dominates as the unmarked pitch accent 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 in which the L tone has been deleted. 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 more segmental material becomes available, and occurrences of H* to increase when there is less segmental material available to realise both tones. If there is a tendency for H to be retained and for 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 that prevents them from being deleted. Thus, the third research question, which constitutes the main theoretical concern of Chapter 6, is as follows:

RQ3. Is there evidence in the 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 have access to a phonological contrast to distinguish between the two. i.e., DECs can be realised with an H*L % and DCQs with an L*H H% (or similar) nuclear pattern. In Chapter 3, Section 3.4.2, it was suggested that nIE may also have access to a phonological contrast, but that it may be manifested though changes in a register tier (Inkelas & Leben, 1990; Leben et al., 1989). However, the argument for a register tier is not commonly invoked. In fact, Gussenhoven (2004) 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 across sentence modes (Chapter 7) presents evidence for a register tier as the best explanation for functional changes in the phonology of DCE. Thus, the final research question is as follows:

RQ4. Does a register-tier analysis provide a plausible explanation for phonetic variation across sentence modes in DCE?

4.1.3 Phonology-first and Phonetics-first Approaches

Chapter 1 described how the main body of research in this thesis can be construed as taking a top-down Phonology-first approach, and that this contrasted with an alternative bottom-up Phonetics-first approach. The research questions are largely answered via the Phonology-first approach adopted in Chapters 6 and 7. The main approach is described as Phonology-first since it begins with assumptions about potential phonological inventories available and works down towards a phonetic analysis of tonal targets associated with those inventories. It also takes the view that realisations of L and H tones can largely be associated with f_0 minima and maxima. While I believe that the analytical approach adopted and the findings presented in Chapters 6 and 7 do indeed work towards answering the research questions, on reflection, some aspects of the analysis seemed to reflect a naïve view of the relationship between underlying tones and their implementation in the pitch contour. Chapter 2, Section 2.4, (which was partly revised with the benefit of hindsight) has already drawn attention to some these issues, specifically the fact that a turning-point rather than an f_0-extrema approach to the phonetic analysis of tonal targets might offer a more nuanced analysis.

Chapters 8 provides a rationale for the Phonetics-first approach in light of a critique of Chapters 6 and 7. It outlines the principles on which the Phonetics-first approach is based and describes the tools developed to facilitate this kind of analysis (Rodgers, 2020). It then considers several examples from the data analysed using a Phonetics-first approach to offer a number of insights into unresolved issues raised in Chapters 6 and 7. As such Chapter 8 helps answer the research questions more thoroughly.

5 Methodologies

This chapter summarises the development of the corpus used for this research. It describes the recording environment and the cohort of volunteers who participated in this study, the annotation and data processing procedures, and the statistical methods adopted in the analysis chapters.

5.1 Corpus Development

Central to the research in this thesis is the fact that L*H pitch accents dominate in casual or colloquial speech regardless of communicative function. Therefore, it was decided to elicit colloquial speech patterns as much as possible and to minimize style-shifting to careful or formal speech. It was also important to ensure that a representative set of tokens for each target variable was collected without placing too great a time burden on participants, all of whom volunteered their time. Furthermore, it was necessary to ensure that the data collected would be amenable to phonetic analysis.

For the data collection, four tasks were produced, a read-speech task, an interactive goal-oriented task, a storytelling task, and a contour imitation task. However, only the read-speech task is used in the analysis in Chapters 6 and 7, so only this task is outlined here.

Read speech is very convenient as it allows key variables to be controlled systematically, ensuring sufficient coverage of each variable in an efficient manner. However, it is also likely to lead to style-shifting to a more formal style. To mitigate 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 maximize the chance of approximating casual speech styles. Secondly, it was decided to record volunteers in pairs and to make sure 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*.

Three sets of stimuli were produced and are listed in Table 5.1. The sets are described in more detail in the associated chapter and can also be found in Appendix B. The corpus generated from each set is referred to using same initial, i.e., the A-Corpus, H-Corpus, and M-Corpus.

Table 5.1 Stimulus sets for the read-sentence corpus

Set	Purpose	Chapter
A	Analysis of alignment in prenuclear (PN) and nuclear pitch accents.	6
H	Subset of A to test alignment effects of word boundaries in PNs	6
M	Analysis of pitch accents in sentence modes.	7

5.1.1 Recruitment and Participants

Derry City English speakers were recruited initially by talking to staff and volunteers at the Verbal Arts Centre in the city centre on Bishop Street Within, and then through word of mouth. All potential participants were provided with an information leaflet outlining the tasks involved, time commitment required, how the data would be used, and participant anonymity (Appendix A2). Participants who agreed to take part signed an informed consent form, which clarified this information and highlighted their right to withdraw from participation at any time (Appendix A3). All data were collected and stored following the procedure outlined in the Speech and Phonetics Laboratory Research Ethics Application, 2016, which was approved by the Ethics Research Committee of the School of Linguistic, Speech, and Communication Science (Appendix A1).

In total, there were 23 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 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, while three persistently style-shifted during the recording.

Of the eleven remaining participants (6 female, 5 male, mean age 40, SD 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 background, noting that 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 is presented in Table 5.2, while the map in Figure 5.1 shows the area each was from along with the location of the Verbal Arts Centre, where the recordings were conducted.

5.1.2 Recording

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

Before the recordings, roughly 10-20 minutes were spent chatting with participants, explaining the recording procedure, and collecting biodata. Each participant also read and signed the consent form and was provided with a copy for reference (Appendix A3). 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.

Table 5.2 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

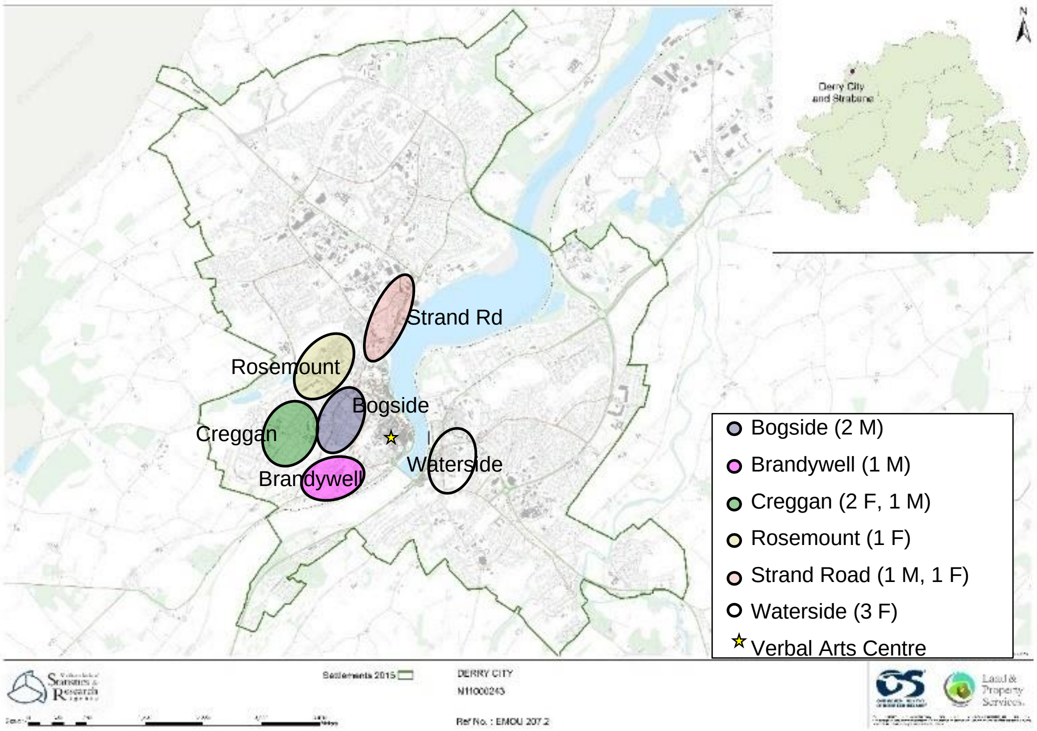


Figure 5.1 Map of Derry City. Ovals show the areas the speakers analysed come from. Yellow star marks the location of the Verbal Arts Centre. Map from Northern Ireland Statistics and Research Agency (2016).

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 imitation 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. Ten repetitions were recorded for each prompt, with 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 silently to five 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.

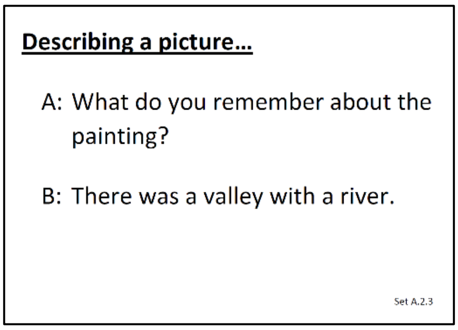


Figure 5.2 Example of read-speech corpus prompt for target phrase A3422.

5.2 Annotation and Data Processing

All data was annotated in Praat (Boersma & Weenink, 2022) using several scripts to help speed up the process[13]. The annotation tiers used for analysis are shown in Table 5.3 and an example of the annotation window in Figure 5.3.

Annotation was carried out 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

[13] All data, Microsoft Excel workbooks, Praat code, and R markdown and code used in this thesis are available in the repository github.com/AERodgers/PhD along with all output from the statistical analyses. Due to ethical restrictions (Appendix A), original recordings are not available publicly. However, resynthesized .wav files of the source waveform of each utterance are. They were generated using Praat’s Klatt synthesizer (Klatt & Klatt, 1990).

generated automatically from the syllable tier using the script `create_more_tiers`, which also generated blank phonological and vowel tiers.

Table 5.3 Annotation tiers used to facilitate phonological and phonetic analysis of IPs.

	Tier name	Type	Content Marking
1	ortho	Interval	Orthography and word boundaries.
2	syllable	Interval	Syllables and syllable boundaries.
3	rhythmic	Point	IP boundaries, metrical stress, and boundary tones.
4	phono	Interval	Pitch accents, absence of expected PA marked as (*).
5	vowel	Interval	Stressed vowels and vowels in syllables with tonal targets.
6	tone	Point	H and L tonal targets, onset (S) and offset (E) of voicing in phrase.
7	comments	Interval	Additional comments and observations.

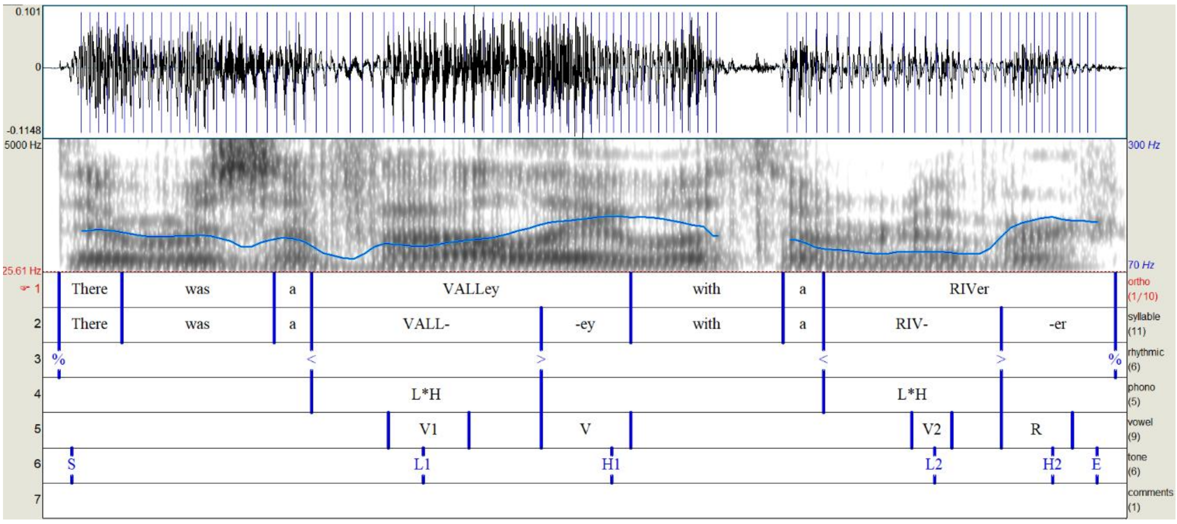


Figure 5.3 Example of annotation window in Praat with annotation tiers.

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 subcorpus, adjustments were made to the IViE labelling system. Typically, the need for such adjustments only became apparent during the labelling process itself, so there was a degree of circularity in the methodology. Adjustments to the phonological labelling are discussed fully in the relevant chapters (Sections 6.4 and 7.4).

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

After the phonological analysis stage, the vowel and tone tiers were annotated. A purpose-written script, `fix_pitch`, was used to facilitate the manual correction of errors in the f_0 contour such as pitch halving or pitch doubling. A screenshot of the script in use is shown in Figure 5.4. This script also allows the user 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, 1999a), which was found sometimes to overcorrect the contour and remove pitch points which appeared to be part of the intended intonation contour. This was especially the case when there were sharp rises or falls in f_0. To further facilitate analysis, the f_0 curve was interpolated to replace missing f_0 points. Finally, the contour was smoothed using Praat's `Smooth` function, with the bandwidth parameters set at 19 Hz. This represents 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 5.5.

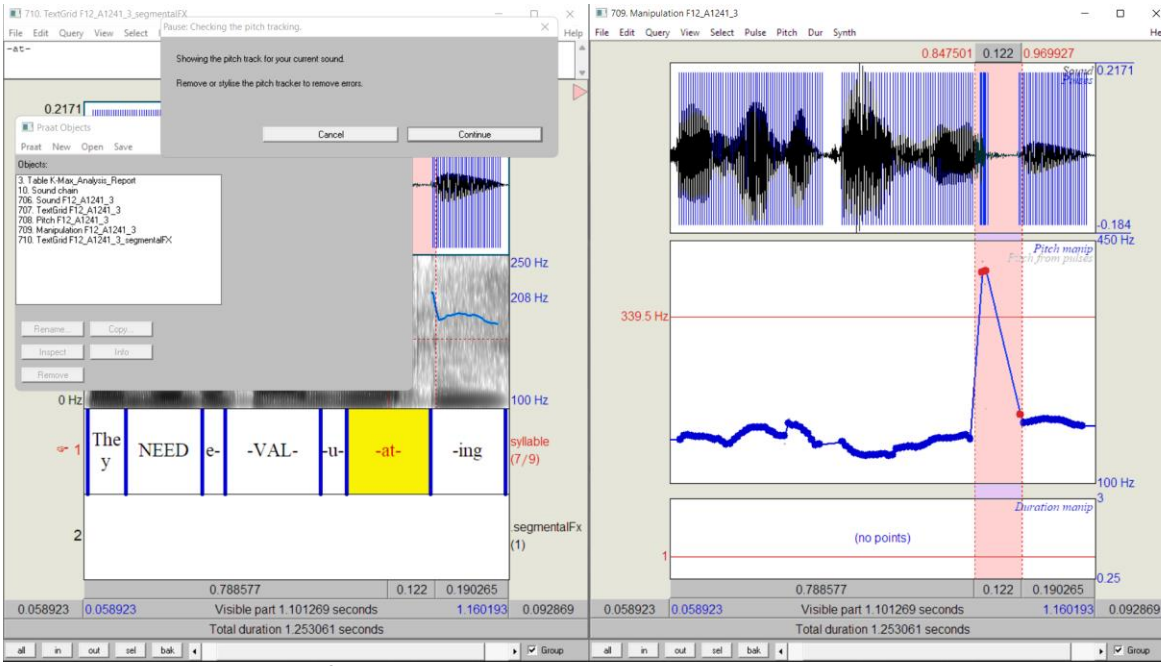


Figure 5.4 Screenshot of the `fix_pitch` *script being used to correct errors in the f_0 contour.*

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 tone tier according to the local minimum and maximum f_0. 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 damping of f_0 caused by a slight increase in intraoral pressure during the realisation of the approximant.

A Praat script called `process_texgrids` was used to produce a data table by extracting data from the text grids and corrected pitch contours for each Corpus. This table is used for the quantitative analysis in Chapters 6 and 7.

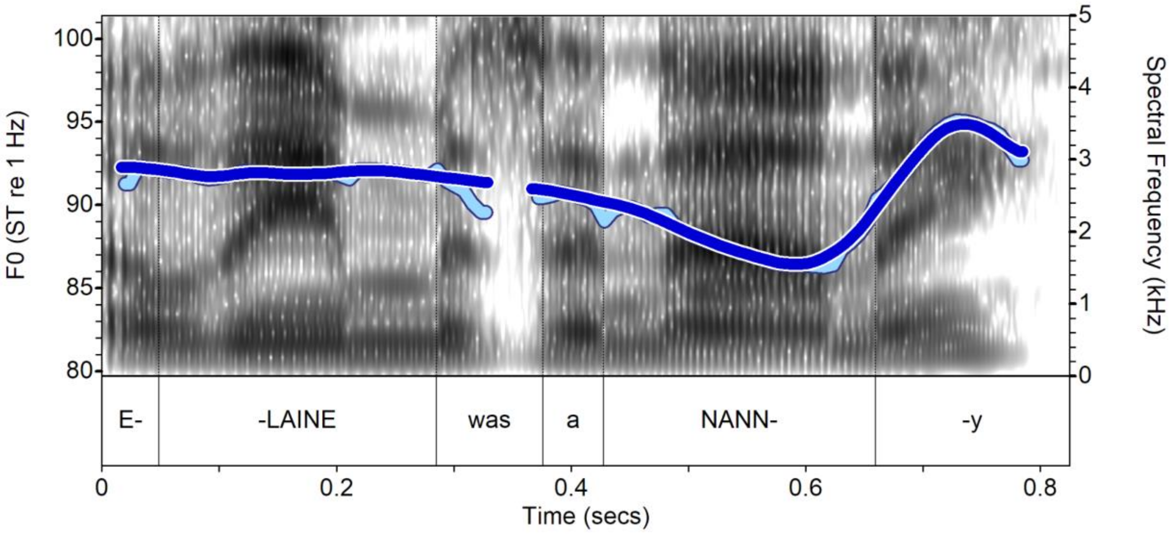


Figure 5.5 Example of corrected f_0 contour (dark blue) after running `fix_pitch` script. The original contour is shown in light blue.

5.3 Data Wrangling Methods for Visual Representation of Count Data

Each corpus contains missing data points, so there is a risk that tabulated raw phonological data could over- or under-represent specific speakers and target utterances. To ensure that visualizations of count data are more representative, raw values have been adjusted to project a more balanced predicted distribution of counts per speaker and per stimulus. The adjustment accounts for the number of repetitions per speaker, the number of speakers per target, and, in some cases, the number of tokens per gender. The output from this process is a table projecting an ideal count of five utterances per speaker per target (with an equal number of male and female speakers). Do note, however, that the data adjustment process has been used only to facilitate data visualisation and not for any of the inferential statistical analyses.

5.4 Inferential Statistics: Linear Mixed-Effects Models

All inferential statistical analysis was conducted in R (R Core Team, 2022) using the packages outlined below alongside several purpose-written functions. Microsoft Office Excel (Microsoft Corporation, 2020) was also used for some data visualisation and basic calculations. The summary which follows is largely a synthesis of Baayen (2008), Baayen et al. (2008), Bates et al. (Bates et al., 2015), and Winter (2019).

Mixed-effects models (Baayen et al., 2008; Bates et al., 2015) have been used for the bulk of the inferential statistical analysis in Chapters 6 and 7. They 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, even when 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. This is because mixed-effects models are designed to compensate for variance caused by multiple random effects and thus compensate for the amount of error or noise the random effects add to the estimates.

A random effect represents a factor which is known to influence the result but is not of interest in the analysis. More precisely, it represents a factor within the data set, the levels of which are taken from a sample of the whole population. For example, Chapter 6 analyses 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, who represent a random sampling of the population of Derry City English speakers, and each has their own idiosyncrasies. Thus, some speaker-specific variation will add 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, and two issues will be outlined here. 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 and so the estimates may not be reliable. The second problem with mixed-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. This means that the model is overly adapted to the current sample but is unlikely to be suited for generalising over the population. Therefore, similar results are unlikely to be found if the model is tested on another random sample from the population.

Two types of Mixed-effects model are used here, linear mixed-effects models (LMEMs) for continuous phonetic variables and Bayesian generalised linear mixed-effects models (BGLMMs) 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) but which does not occur in the continuous data.

5.4.1 Analysing Continuous Parameters with Linear Mixed-effects Models

Linear Mixed-effects Models (LMEMs) were chosen to evaluate the continuous phonetic parameters using 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 fixed factors(s), random factors, and the target response variable (i.e., the target phonetic parameter). If the effect of a specific factor was to be analysed, it was included as a fixed effect; however, if the factor was not to be analysed but was still believed to affect the response parameters, it was included as a random factor. However, these maximal models were often prone to convergence and singularity problems.

To mitigate convergence and singularity issues, the advice included in the `lme4` package was followed. This included lowering the tolerance threshold for singularity and using the `allFit()`

function to check for alternative optimizers. Additional approaches suggested by Nugent (2022) were also implemented. A function, `optimizeModel()`, was written to further automate this process and identify a set of model optimization parameters which most effectively avoid convergence and singularity issues.

If convergence errors or singularity issues remained, the model was simplified, and this was generally necessary. Each model was simplified in two stages. 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. However, sometimes, non-significant results are of interest (i.e., sometimes we want to check that a factor has very little effect), so they were not always removed from the model. In fact, fixed factors were only removed as a last resort, and the model was always tested by removing random effects first. In cases where the backward reduction of the model still led to convergence errors, the models were retested via the manual stepwise removal of random slopes followed by random intercepts in order to find the most complete model which did not generate convergence or singularity errors.

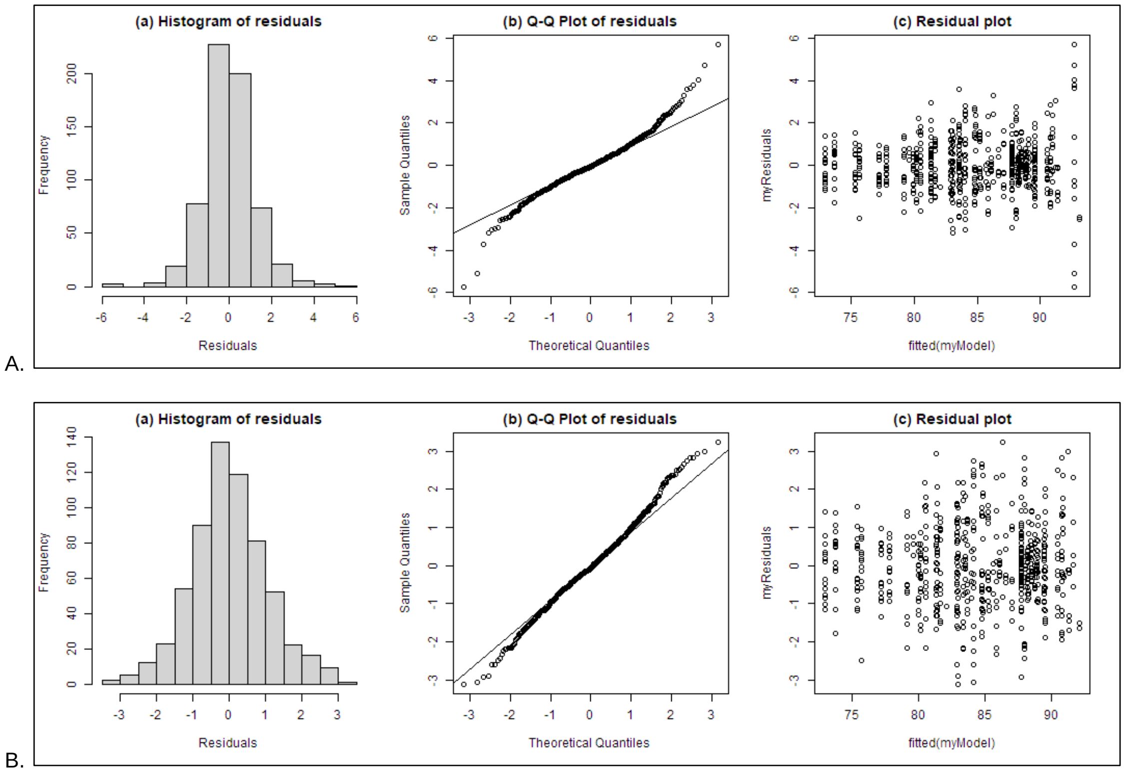


Figure 5.6 Residual plots of the model `l_f0 ~ mode + fin_phon + gender + (1 + mode | speaker)`*. Panel A shows the model residual plots before trimming while Panel B shows the model residual plots after trimming the data, with a loss of 9 out of 632 observations.*

Once the final model was established, the residuals were examined visually using a histogram of residuals, Q-Q plots of residuals, and a residual plot, as shown in Figure 5.6. 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 residuals did not look randomly scattered across 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 minimize the number of observations trimmed while producing a more reasonable distribution of residuals. Panel A in Figure 5.6 shows the model residuals before trimming and Panel B the residuals after observations were trimmed using a standard deviation threshold of 2.5 (though do note, in the second plot, heteroskedasticity has not been completely removed).

To further check each model, an Analysis of Variance test (ANOVA) was conducted to test the statistical significance of each fixed effect in each model.

5.4.2 BGLMMs and the Analysis of Discrete Categories

As with all generalised linear models, generalised 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 and consistency, the original 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 known as complete separation. This occurs when one level (or more) of a predictor variable perfectly predicts the outcome. For example, in one analysis of sentence modes (see Chapter 7), declarative statements are never associated with an H% boundary, so in a binomial analysis, the odds of H% in a declarative statement is 0:1 against. While the probability of H% in a declarative statement is in fact zero, the model generates log-odds Confidence Intervals (CIs) tending towards plus and minus infinity, giving the erroneous impression that the likelihood of H% cannot be predicted.

This issue was resolved by using Bayesian generalised mixed-effects models (BGLMMs) from the `bglme` package (Chung & Rabe-Hesketh, 2013), which is based on `lme4`. A BGLMM allows the user to impose priors on the model, which reflect expectations based on prior experience and knowledge. As we know that the confidence intervals of the model are not zero or infinity, we can impose zero-mean normal priors on the model (Bolker, 2018), i.e., inform the model that the priors predict a model with a normal distribution around the mean. This prevents the estimation of confidence intervals tending towards infinity.

As with the LMEMs, a maximal model with random slopes and intercepts was tested first and 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. Significance tests for fixed factors were tested using R's in-built `drop1()` function with a Chi-Squared (χ^2) test, as suggested in `lme4`'s R documentation. This function drops each fixed effect from the model one at a time, compares the reduced model against the full model in a Likelihood Ratio Test (LRT) and summarises the differences in fit between the two models.

5.4.3 Procedures Used in Both LMEM and BGLMM Analysis

Once the most informative working model was established for either the LMEM or BGLMM analysis and outliers were removed, the purpose-written functions `analyseModel()` and `getModelFixedFX()` 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. For LMEMs, this includes mean estimated values for each level of the independent factors, while in BGLMMs, it includes the predicted probability of each level of the independent factors. It also calculates marginal and conditional *R*-squared (R^2) values for each model using the `r2()` function from the `performance` package (Lüdecke et al., 2021; Nakagawa et al., 2017; Nakagawa & Schielzeth, 2013). Marginal R^2 (R_m^2) indicates the amount of variance explained by the fixed effects only, while conditional R^2 (R_c^2) reflects the amount of variance explained by the whole model, i.e., it includes both random and fixed effects. Thus, for example, a marginal R^2 of .05 indicates that 5% of the variance in the response variable is explained by the fixed effects, while a conditional R^2 of .95 indicates that 95% of the variance is explained by the whole model.

`analyseModel()` also outputs effect size parameters using the partial omega-squared statistic (ω_p^2) (Ben-Shachar et al., 2020). As with R^2, ω_p^2 is a ratio; however, unlike marginal R^2, which indicates the amount of variance explained by all of the fixed effects in the model, ω_p^2 indicates the amount of variance explained by each individual fixed effect in a model. It is considered a less biased estimate than alternatives such as eta-squared (Albers & Lakens, 2018; Richardson, 2011).

`getModelFixedFX()`calculates the intercepts for each level of the categorical fixed factor(s) in the model along with slopes between each level. This facilitates the pairwise comparison of effects across all levels of a fixed factor. For example, we are able to compare the estimated difference in nuclear peak alignment between a declarative statement and a declarative question, or between a declarative question and a yes-no question, or between a yes-no question and a declarative question.

Finally, after all analyses for a specific topic are complete—e.g., the analysis of lexical and metrical effects on prenuclear tonal targets—*p* values of the relevant ANOVAs for that topic are adjusted *post hoc*. This is done to compensate for the ever increasing likelihood of getting a spurious false positive result—or Type I error—as more and more statistical tests are conducted (Field et al., 2012, pp. 428–432; Roettger, 2019; Winter, 2019, pp. 175–177). To achieve this, the Benjamini and Hochberg (BH) method was used, since it minimizes the number of Type I errors while also mitigating against Type II errors (false negatives) by controlling the false discovery rate (FDR) (Benjamini & Hochberg, 1995). It was implemented via a purpose-written function `adjustP_posthoc()` which pools together all tests within the same group of analyses and adjusts the *p* values using the base R function `p.adjust()`. For example, it is used to adjust all the *p* values output by every ANOVA of each model analysing metrical and lexical effects on the nuclear pitch accent tonal targets, coming to a total of 38 values. Both original and adjusted *p* values are included in the results of the statistical tests in the appendices.

6 Analysis of Form: Metrical and Lexical Effects

This chapter focuses on the effects of meter and lexical boundaries on the phonology and phonetic implementation of the tonal tier in declarative statements in Derry City English (DCE). In other words, the first broad aim is to answer the descriptive question regarding form raised in RQ1 (4.1.1). 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. The second broad and more theoretically oriented aim of this chapter is to answer RQ3, "Is there evidence in the DCE for the special status of H tones?"

6.1 Hypotheses and Expectations

It is clear from previous research (see Sections 3.3.1 and 3.4.1 above) that nuclear pitch accent inventories are often 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. That is, 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 prenuclear pitch accent, so it is potentially more prone to non-linguistically motivated variation. Variation in pitch accents thus may have at least two competing sources:

a. Communicative intent
b. Metrical and lexical structure

If variation is motivated by communicative intent, we should expect to see little correlation between the intonational phonology and anacrusis, foot size, and lexical boundaries; however, if variation is conditioned by the metrical or lexical structure, we expect to see pitch accents vary as a function of anacrusis, foot size, and word boundaries. There is a third source of non-linguistic variation, that of speech rate. However, speech rate may also be associated with communicative intent, as speakers may increase or decrease speech rate depending on the extent to which they want the listener to attend to different components of the speech stream.

Broadly, 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 prenuclear pitch accents, where there is less communicative pressure. It is proposed that in PN pitch accents, less anacrusis and shorter foot size will be associated with increased occurrences of H*, while L*H will be more common once anacrusis and foot size increase. Overall, however, L*H is expected to dominate both in prenuclear and nuclear pitch accents. Thus, in answering the phonological component of RQ1, we have the following hypotheses:

1. L*H is the dominant pitch accent in nuclear and prenuclear 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 prenuclear 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, if prenuclear pitch accent types are found to vary as a function of metrical and lexical conditions, it follows that the labelled inventory of pitch accents reflects 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 [t^{h}] in <time> [t^{h}aɪm] but as [ɾ] in <butter> [bʌɾɚ] and [$\text{t}^{=}$] in <stay> [s$\text{t}^{=}$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 realisation.

Looking at previous research on tonal alignment in nIE, as discussed in 3.3.2, one can expect metrical and lexical context to condition the timing of PN tonal targets. An early analysis of temporal alignment of PN accents from six speakers—presented at BAAP 2018 (Rodgers, 2018, Appendix C)—indicated that this was the case for the DCE data too. Moreover, it suggested two strategies for H alignment. In one, the right boundary of the lexically stressed word acts as an anchor while in the other the foot boundary acts as the anchor. Thus, in the analysis of the full dataset, we expect H targets to be aligned later when the foot contains more syllables or when the right word boundary occurs in a later syllable. As a corollary of this, it is also possible that there is an effect of the right word boundary on the distribution of PN pitch accents. That is, given that it is hypothesized that a longer foot is associated with an increase in L*H pitch accents, it is possible that a later word boundary should also be associated with an increase in L*H. Therefore, a fourth prediction re the phonology might be added:

4. The later in the foot the initial word boundary occurs, the greater the likelihood of a prenuclear L*H pitch accent.

When it comes to anacrusis effects on PNs, it is much less clear what the effects will be. For example, a study by Nolan and Farrar (1999) found that the addition of anacrusis was associated with earlier peaks in Belfast English. However, a study of Donegal English by Kalaldeh, Dorn, and Ní Chasaide (Kalaldeh et al., 2009) found the opposite, i.e., that increased anacrusis is associated with later peak alignment in PN pitch accents.

In nuclear PAs, it is expected that H targets will align later as foot size increases, while it is presumed that an increase in prenuclear unstressed syllables 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. 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 clearly. As such, the speaker may crowd the tones into a shorter time period, regardless of foot size, realigning tonal targets where necessary. Furthermore, if there is a specified L% boundary, crowding effect should also be observed, with H targets in L*H being aligned earlier. (Remember that, taken together, the nuclear pitch accent and boundary tone are viewed here as the nuclear contour.) In other words, it is assumed that communicative pressure will be associated with compression effects, and that these will manifest most clearly in shorter feet and in L*H L% contours. 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). For PN pitch accents, however, it is unclear if, or how, truncation and compression will manifest.

To answer the phonetic component of RQ1, therefore, we have five interconnected hypotheses:

5. Tonal alignment in PN accents will be more vulnerable to metrical and lexical effects compared to tonal alignment in nuclear pitch accents.
6. Anacrusis will affect the alignment of PN H targets; however, the direction of this effect is not predictable in advance.
7. There are competing strategies for H target anchoring in PN pitch accents, one using the right lexical boundary, the other using the right foot boundary as an anchor point.
8. Compression effects in the guise of tonal crowding will be observed in nuclear pitch accents. This will be an effect of reduced foot size.
9. Truncation effects may be observed in nuclear pitch accents.

6.2 Materials

The A and H subcorpora are used for analysis in this chapter. The A-Corpus was designed primarily to investigate metrical effects on the inventory of nuclear and prenuclear 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 prenuclear 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 prenuclear 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” in Table 6.1).
4. Size of the foot associated with the nuclear pitch accent (NUC foot size), from one to four syllables.

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 different variables.

Table 6.1 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	VALerie's is valid.	PN	Anacrusis	0	*... *.
A1422	The VALLey's by the river.	PN	Anacrusis	1	.*... *.
A2422	There's a VALLey with a river.	PN	Anacrusis	2	..*... *.
A3422	There was a VALLey 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	VALerie's is valid.	PN	Foot size	4	*... *.
A1111	They know VAL.	NUC	Preceding	0	.* *
A0221	Val's is VALid.	NUC	Preceding	1	*. *.
A0321	Val's is inVAlid.	NUC	Preceding	2	*.. *.
A0423	Valerie's is VALid.	NUC	Preceding	3	*... *.
A1211	He lives with VAL.	NUC	Foot size	1	.*. *
A0221	Val's is VALid.	NUC	Foot size	2	*. *.
A1231	I live with VALerie.	NUC	Foot size	3	.*. *..
A1241	They need eVALuating.	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 the same anacrusis and foot size parameters within each pair but different word boundaries. Thus, the variable of interest—i.e., the lexically stressed PN word boundary—changes. Such variation within and across pairs will help assess the extent to which variation in H alignment is influenced by word boundary effects.

Table 6.2 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	. * .|. * .

6.2.1 Annotation and Data Extraction

The utterances were annotated as outlined in 5.2 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 made to reflect this. These adjustments are discussed below in 6.4.

6.2.2 Data Extraction, Pruning, and Preparation

Once annotation was complete and the phonology had been agreed upon, data from the A- and H-Corpora were tabulated according to the processes outlined in 5.2. Ineligible utterances were automatically pruned using a Praat script `corpus_audit`. These included utterances where speakers added special stress to a typically unstressed syllable or where they deleted a target syllable. 45 utterances were removed in this way, so the total number of valid utterances fell from 833 to 788.

Table 6.3 Summary of valid A- and H- Corpora 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				✓					✓	✓	✓					**209**
pn_foot	✓	✓	✓	✓												**203**
nuc_pre		✓	✓	✓	✓											**202**
nuc_foot		✓				✓	✓	✓								**196**
pn_lex			✓	✓								✓	✓	✓	✓	**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

Table 6.3 summarises the final distribution of valid utterances by stimulus and speaker. 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.

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. (Note that Table 6.3 also indicates to which subset or subsets each target phrase belongs.)

Table 6.4 List of datasets used for the analysis of lexical and metrical effects on pitch accent phonology and phonetic implementation of tonal targets. The abbreviation on the left of the underscore refers to the target pitch accent and the abbreviation to the right refers to the treatment variable.

Data	Target pitch event	Treatment variable(s)	Function
`nuc_foot`	Nuclear pitch contour	Foot size (syllable count)	Analysis of foot size effects on nuclear pitch accents and pitch contours
`nuc_pre`	Nuclear pitch contour	Unstressed syllables preceding the stressed syllable	Analysis of effects of unstressed syllables preceding nuclear pitch accents and pitch contours
`pn_foot`	Prenuclear pitch accent	Foot size (syllable count)	Analysis of foot size effects on prenuclear nuclear pitch accents
`pn_ana`	Prenuclear pitch accent	Anacrusis (syllable count)	Analysis of anacrusis effects on prenuclear nuclear pitch accents
`pn_lex`	Prenuclear pitch accent	Anacrusis, foot size, location of word-end syllable	Analysis of word boundaries effects on prenuclear pitch accents

6.3 Methods

Using the methods outlined in Section 5.3, the phonological data was tabulated to provide an overview both of interspeaker trends in the distribution of pitch accents and pitch contours as a function of foot size and anacrusis. To remove excessive noise in the PN data, downstep was ignored, especially since this was a reflex of an initial high boundary tone, and not a unique PN pitch accent.

To assist in the assessment of word-boundary effects—especially regarding the possibility of competing alignment strategies mentioned in Hypothesis 7—peak alignment of PNs was calculated as a function of syllable-normalised time and grand syllable-mean normalised time.

Syllable-normalised time converts time measured in milliseconds to time measured in syllables. Here, each syllable counts as one unit and each time point is measured as a ratio of the syllable in which it occurs. As such, for example, a target which occurs in the middle of the first syllable has a syllable-normalised time of 0.5, while a target that occurs three-quarters of the way into the third syllable has a syllable-normalised time of 2.75. This allows tonal alignment in different target phrases to be evaluated within the same figure.

Grand syllable-mean normalised time uses the grand mean duration of each syllable in identical target utterances to convert syllable-normalised time back into milliseconds. This allows comparison of multiple repetitions of the same target to be compared in the same figure. It also provides an accurate representation of the true duration of syllables. Each process is described more fully in Appendix E.

For the inferential statistical analysis of the phonological data, Bayesian generalised linear mixed-effects models (BGLMMs) were employed while linear mixed-effects models (LMEMs) were used for the analysis of phonetic parameters (see Section 5.4). f_0 is measured in semitones (ST) re speaker median f_0. This facilitates comparisons between speakers and across gender. The median was chosen rather than the pitch floor because pitch floor estimates are likely to be overly affected by segmental effects which dampen f_0 and are also more likely to be subject to measurement errors. The median estimate does not suffer from this problem; moreover, it will not be adversely skewed by outliers. Finally, for the sake of convenience, ST re speaker median is abbreviated to ST throughout.

The complete A- and H-Corpora data were used for the mixed-effects models analysis. There were two reasons for this. Firstly, mixed-effects models can cope with datasets where all levels of independent factors are not equally represented (see Section 5.4 on LMEMs), thus allowing the use of much more data in the analyses (see Table 6.5 and Table 6.6). However, for the current analysis, this also required that extra random effects be included to account for greater segmental variation in the data. The second reason was that it allowed for the assessment of a lexical effect which had initially been overlooked, namely the effect of the left word boundary. Further details on the pruning of data and selection of parameters for modelling (including extra random effects) are presented in each results section below.

Table 6.5 Comparison of PN observations available for analysis in A- and H-Corpora and subcorpora.

corpus / subcorpus	L target analysis	H target analysis	right word-boundary effects	f0 excursion and slope	total
pn_ana	180	205	N/A	180	205
pn_foot	124	174	N/A	117	203
pn_lex	N/A	N/A	202	N/A	260
A- and H-Corpora	498	737	489	489	788

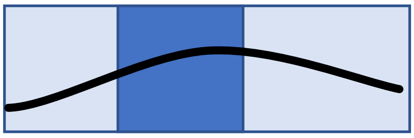
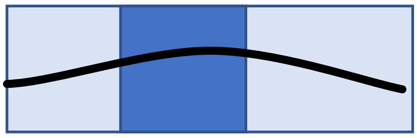
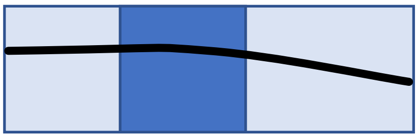
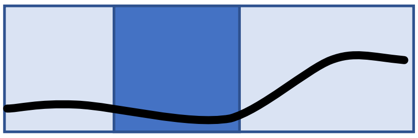
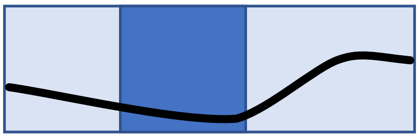
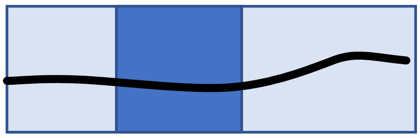
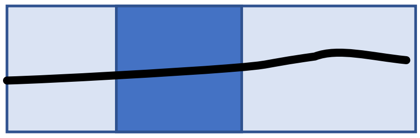
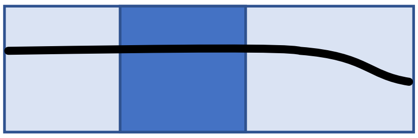
Table 6.6 Comparison of nuclear PA observations available for analysis in A- and H-Corpora and subcorpora.

corpus / subcorpus	observations for analysis
nuc_pre	202
nuc_foot	196
A- and H-Corpora	788

6.4 Phonological Labelling

For the nuclear contours, labelling was very straightforward; however, prenuclear pitch accents presented more of a challenge. This was mostly due to the loss of a clear L target in some pitch accents. To help illustrate this, Table 6.7 presents stylisations of PN contours. It begins with the most easily identifiable PAs at the top and moves down to the more challenging ones towards the bottom.

Table 6.7 Stylised representation of contours typically found in PN position. (Dark blue indicates the lexically stressed syllable.)

	Stylisation	**Comment**	**Label**
A.		Phonetically low initial boundary rising to a peak in or near the end of the lexically stressed syllable. Auditorily salient peak prominence in the stressed syllable.	%L H*
B.		Phonetically mid-range initial boundary rising to a peak in or near the end of the lexically stressed syllable. Auditorily salient peak prominence in the stressed syllable.	% H*
C.		Phonetically high boundary, leading to a small peak or plateau-like prominence in the stressed syllable before falling slightly	%H H*
D.		Phonetically low or mid-range initial boundary with a clearly identifiable L*H.	% L*H
E.			
F.		Low or mid-range initial boundary. L target is not as clearly defined as in D and E; however, it is easily perceived as an L*H	
G.		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*
H.		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.	

As shown in Panels A to C, H* was visually and auditorily salient in many cases. If there was a noticeable rise from an initial low boundary to the peak (Table 6.7a), 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 C). These were labelled as %H H* to reflect the plateau structure.

L*H PN pitch accents were often auditorily salient (Panels D to F), despite some variation in excursion size and the height of the L target. For example, sometimes the L target was visually less prominent (Panel F) but was nonetheless perceived as an L*H, both by the author and by MOR, the intonation specialist who was consulted on 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 effects 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 G. In other cases—as illustrated in Panel H—there was a phonetically high initial f_0 followed by a plateau which ended after the stressed syllable. In neither case was there an 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 the L*H PAs described. Overall, these contours sounded neither quite like L*H nor like 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.

6.4.1 Provisional Phonetic Comparison of H, >H*, and L*H Peak Alignment*

It should be borne in mind that the decision to include the intermediate >H* category was largely a practical decision, based on the lack of a salient rise-from a low (i.e., an L target) and the apparent earlier peak of the H in some PN pitch accents. Moreover, as shall be seen Section 6.5.2 below, H* and >H* are both relatively uncommon in the data, with L*H dominating in PN position.

In the Phonology-first approach taken in this chapter, a phonetic analysis of L targets was only conducted in cases where an L target was salient, so—unfortunately—L*H, >H*, and H* are only comparable in terms of their H targets (c.f. Section 6.6.2.3). The overall weaknesses of this Phonology-first approach are considered in more detail in Chapter 8, where it is also observed (Section 8.6.1) that the apparent >H* pitch accent may well be analysed as surface realisation of either an L*H or an H*. However, this is not the approach taken in this chapter.

With these caveats in mind, visualisation of the data suggests that there is a split in peak alignment between H* and >H* at least. Figure 6.1 shows the distribution of PN H alignment taken from three metrically similar phrases from the A-Corpus (A1422, A2422, A3422). These target utterances were chosen because they are evenly represented across all speakers and because each contains four syllables in the first foot (although they do have different anacrusis sizes). To make the phrases comparable, timing was normalised using grand syllable-mean normalised time to facilitate comparability. The figure suggests a bimodal distribution, with all >H* peaks aligned later than H* peaks.

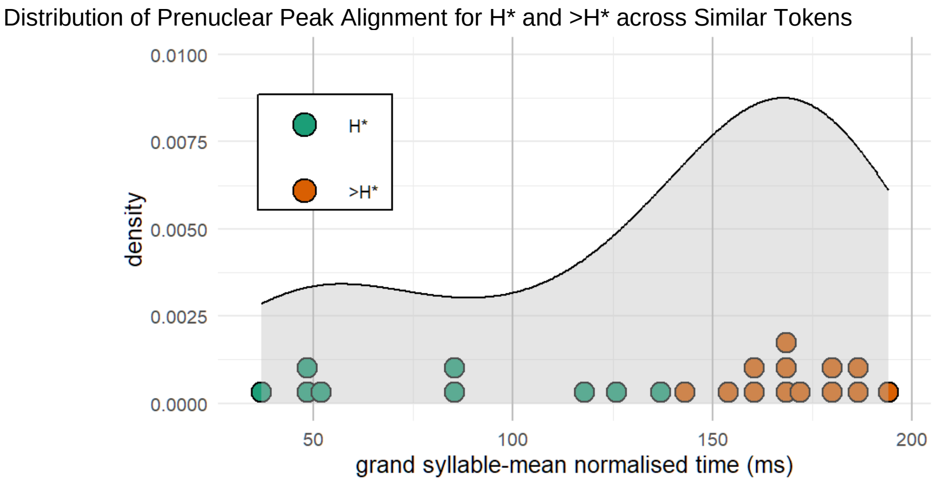


Figure 6.1 Dot plot and distribution density (in grey) of PN peak alignment of A1422, A2422, and A3422 tokens (H and >H* only).*

The picture becomes less clear when we include the PN L*H pitch accents from the same pool of data, as shown in Figure 6.2, where there is some overlap in the distribution of L*H and >H* peaks, even though the vast majority of L*H peaks are aligned much later. However, this visualisation does not account for the other element which helped distinguish >H* from L*H, namely the lack of the rise away from a low target. Moreover, as is considered in much more detail in section 6.6.1, it is possible that speakers adopt different strategies for peak alignment in L*H, which further complicates the interpretation of the data once L*H is included.

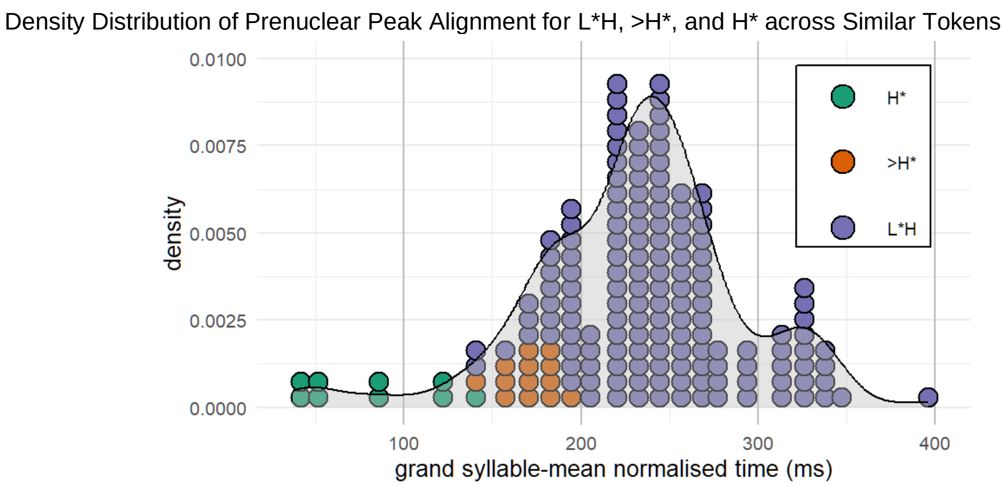


*Figure 6.2 Dot plot and density distribution (in grey) of PN peak alignment of A1422, A2422, and A3422 tokens (L*H, >H*, and H*).*

6.4.2 Provisional Phonetic Comparison of % and L% boundaries

As shall be seen in Section 6.6.1, L% boundaries were quite rare, and the majority of them were produced by F12 (15 out of 24 L% tokens in the across the A- and H- Corpora with a total of 788 tokens). If we check the plausibility of L% using a density distribution plot of boundary f_0 across all speakers (Figure 6.3), it tells us very little about the distribution of boundary tones, except that the boundary tones tend to be scaled slightly above the speaker mean and that L% tends to be scaled much lower than the %. Overall, it suggests an almost normal distribution of f_0 at the final boundary, albeit one with a noticeably long tail in the lower f_0 range as an effect of L%. However, when we look at the distribution of boundary f_0 for F12 alone (Figure 6.4), the only speaker who produced a sizeable number of L% tones, we see that there is indeed a distinct bimodal distribution of the boundary tones, with L% accounting for the peak in the distribution at lower f_0 values. This suggests that the phonological distinction in boundary tones, analysed as % and L%, is valid. The statistical analysis of boundary tone effects on boundary f_0 reinforces this analysis (Section 6.6.3.3, p. 114).

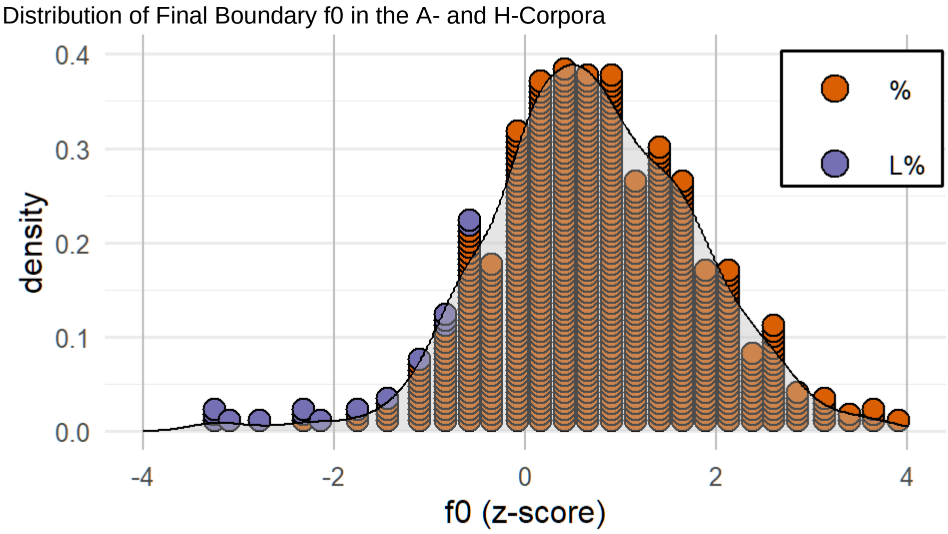


Figure 6.3 Dot plot and distribution of final boundary f_0 (z-scored to speaker mean) in A- and H- Corpora (excluding target phrases with unbalanced representation.)

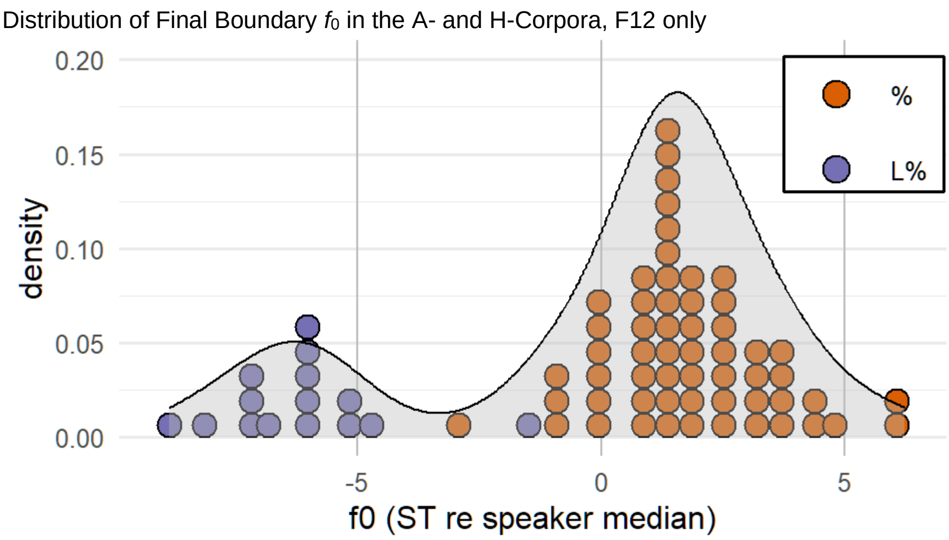


Figure 6.4 Dot plot and distribution of all final boundary f_0 tokens from F12 in A- and H- Corpora.

6.5 Phonological Analysis and Results

The phonology of nuclear pitch accents is largely uninteresting and will be considered first before moving on to prenuclear pitch accents, which are more varied and offer more challenging data for analysis.

6.5.1 Phonology of Nuclear Pitch Contours

The distribution of nuclear pitch contours as a function of foot size (`foot_syls`) and preceding unstressed syllables (`pre_syls`) are summarised in Table 6.8 while per-speaker summaries are presented in Table 6.9 (next page). The tables present adjusted values only (see Chapter 5, Section 5.3), while raw counts can be found in Appendix D (Table D1.4–0).

Table 6.8 shows that L*H was used exclusively as the nuclear pitch accent for all speakers across all conditions. This confirms L*H as the dominant pitch accent. It also confirms that variation in metrical context has no effect on the inventory of nuclear pitch accents. The table also shows that there were both L% and unspecified boundary tones. However, L% accounts for only 4.3% of the adjusted data (3.5% of raw data, *n* = 12/343). Previous AM studies of nIE (c.f. 3.3.1) also found L*H L% contours in declaratives, so its presence is unsurprising. L% does not appear to be an effect of `foot_syls` or `pre_syls`, as there is no pattern associated with either factor, as evidenced by the distributions shown in the final column of each panel in Table 6.8. In fact, Table 6.9 reveals that the majority of L% boundaries come from a single speaker (F12), accounting for 30% of her utterances, while the remaining few (*n* = 4/343, *n.adj*. = 5/385) come from only two other speakers, M8 and M9.

Table 6.8 Nuclear contours by foot size (adjusted) in the A-Corpus subsets.

A. Distribution of pitch contours in `nuc_foot`.

foot size (syls)	L*H %	L*H L%
1	50	5
2	55	0
3	49	6
4	52	3
Total	206	14

B. Distribution of pitch contours in `nuc_pre`.

preceding syllables	L*H %	L*H L%
0	51	4
1	55	0
2	55	0
3	53	2
Total	214	6

Table 6.9 Nuclear contours by speaker (adjusted) in the `nuc_foot` and `nuc_pre` datasets.

speaker	L*H %	L*H L%
F5	35	0
F6	35	0
F12	24	11
F15	35	0
F16	35	0
F17	35	0
M4	35	0
M5	35	0
M8	31	4
M9	34	1
M10	35	0
Total	369	16

L% may convey a meaning based on the speaker's interpretation of the stimuli or simply reflect idiosyncratic speaker preference. However, based on my own impression of the L*H L% contour—as well as on intuitions elicited from DCE speakers and others familiar with DCE—it 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.

L*H L% seems similar in form to the rising-falling nuclear contour (tone C) described in McElholm's (McElholm, 1986) two-speaker study of DCE. However, McElholm states that the fall in tone C may only be slight, which makes it more akin to L*H % than L*H L%. Formally, it is somewhat similar to the extra-high rise-fall (tone D), which he describes as having a steep fall which returns to a low pitch. Of course, the difference here is that L*H L% is not extra-high. Functionally, L*H L% also seems closer to tone D, which McElholm suggests indicates surprise or assertion. 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. (For more on this, however, see Section 6.7.3 later in this chapter and the discussion of L% in Chapter 7, Section 7.7.1)

6.5.2 Foot-size and Anacrusis Effects on Prenuclear Pitch Accent Phonology

The distribution of prenuclear pitch accents as a function of foot size (`foot_syls`) and anacrusis (`ana_syls`) are shown in the left- and right-hand panels of Table 6.10 respectively. Again, only adjusted counts are shown, while the raw data is available in Appendix D (Table D1.1–Table D1.3).

The most noticeable feature of the distribution of PN pitch accent types is simply that there are more of them, namely L*, H*, >H*, and L*H. There are also with several cases of non-accentuation, represented by (*). This is in sharp contrast to the nuclear pitch accents, which were exclusively L*H. While five PA 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. The larger inventory of prenuclear pitch accents is the first hint to the validity of Hypothesis 3, that variation in metrical context has a strong effect on the (surface) phonology of prenuclear pitch accents.

Table 6.10 Prenuclear pitch accents by foot size and anacrusis in the A-Corpus subsets (adjusted count).

A. Distribution of pitch accents in `pn_foot`.

foot size (syls)	(*)	L*	H*	>H*	L*H
1	9	5	19	2	20
2	2	2	21	3	27
3	1	0	13	6	35
4	0	0	0	5	50
Total	12	7	53	16	132

B. Distribution of pitch accents in `pn_ana`.

anacrusis (syls)	H*	>H*	L*H
0	0	5	50
1	9	5	41
2	0	4	51
3	0	4	51
Total	9	18	193

A second striking feature in the `pn_foot` data is that, as foot size increases, there is an overall proportional increase in occurrences of L*H while instances of non-accentuation and L* decrease. This can be seen clearly in Figure 6.5, which shows the adjusted PA distributions across PNs across foot-size conditions, where we see non-accentuation and L* occurrences drop off sharply while L*H becomes more dominant as foot size increases. This lends more support to Hypothesis 3, as it does appear that increased foot size is associated with increased occurrence of L*H.

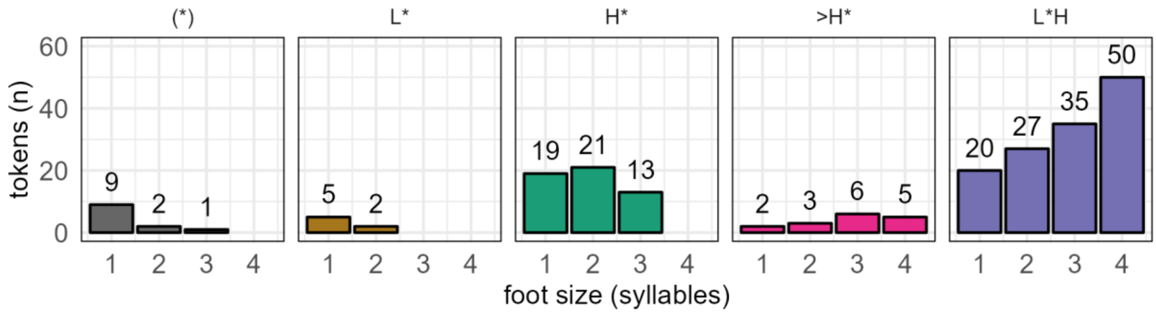


Figure 6.5 PNs across foot-size conditions (adjusted).

This trend is not mirrored in the `pn_ana` data, where L*H dominates throughout. This is reflected in Figure 6.6, which shows the adjusted distributions of PA tokens across anacrusis conditions. It should, however, be noted that all target phrases in the `pn_ana` data contain an initial four-syllable foot. This suggests that, once foot-size conditions are sufficient, anacrusis has little effect on the phonology. Unfortunately, the extent of the foot size effect had not been anticipated and was not incorporated into the corpus design.

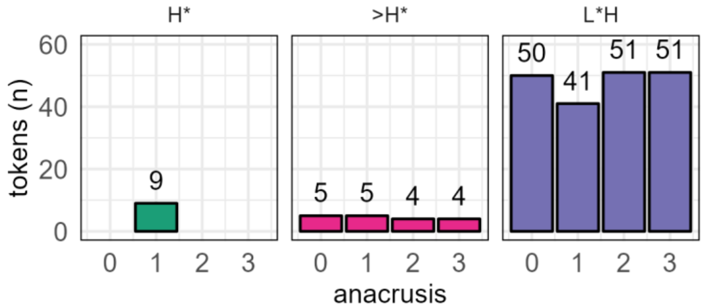


Figure 6.6 PNs across anacrusis conditions (adjusted)

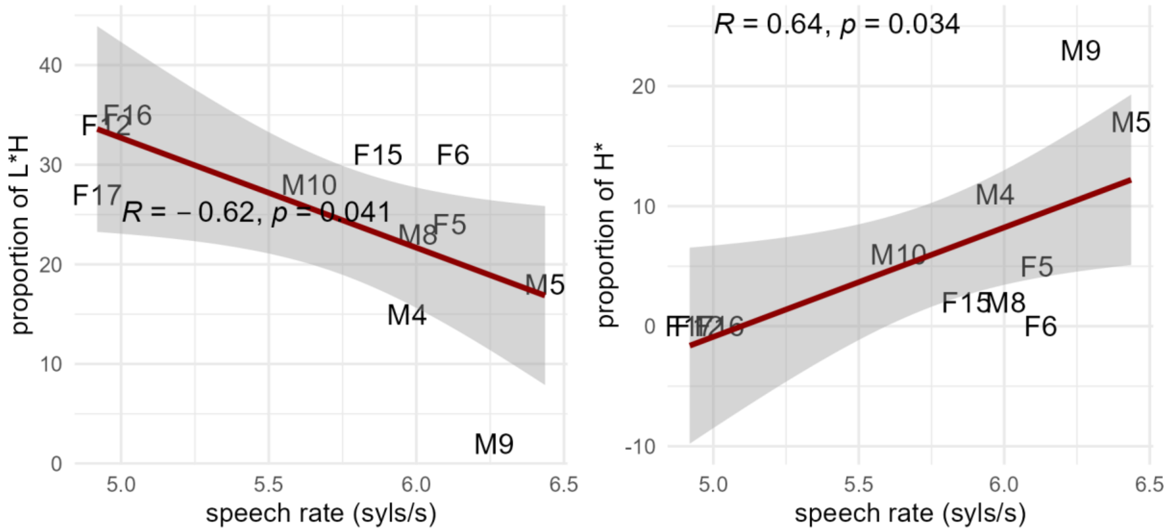
When we consider the distribution of prenuclear pitch accents by speaker across the `pn_ana` and `pn_foot` data combined—as presented in Table 6.11 below—we can see that L*H dominates, accounting for 70% the adjusted count. This further confirms hypothesis (1), L*H is the dominant pitch accent in both nuclear and prenuclear position. While this hypothesis holds true for all the female participants, it is not the same for the male speakers. While M8 and M10 both show a clear preference for L*H, M5 has an almost even split between L*H and H* (51% and 49% respectively), while M4 shows a weaker preference for L*H (42%), followed closely by H* (32%). Unlike all the other speakers, M9 has a clear preference for H* (67%) and uses L*H only occasionally (7%). In fact, M9 also uses the ambiguous >H* much more frequently than the other speakers (27%**)**. Overall, these results give the impression that there is an effect of gender on the distribution of pitch accent types.

Table 6.11 Prenuclear PAs by speaker in `pn_ana` and `pn_foot` subcorpora (adjusted).

speaker	mean speech rate (syls/s)	(*)	L*	H*	>H*	L*H
F5	6.11	4	0	5	2	24
F6	6.12	3	0	0	1	31
F12	4.95	0	1	0	0	34
F15	5.87	0	0	2	2	31
F16	5.02	0	0	0	0	35
F17	4.92	1	7	0	0	27
M4	5.97	5	0	11	5	15
M5	6.44	0	0	17	0	18
M8	6.01	0	0	2	10	23
M9	6.26	0	0	23	9	2
M10	5.64	0	0	6	1	28
Total	63.31	13	8	66	30	268

Interspeaker variation in the realisation of pitch demonstrates that the correlation between foot size and L*H realisation suggested by the summary data (Figure 6.5) does not reflect all individual trends. The aggregated data do appear to reflect a general trend, but speakers such as M9 (and to a lesser extent M5 and M4) do not follow it. Perhaps these speakers continue to produce H* because, for them, it is functionally contrastive with L*H, or, perhaps functional contrast is not intended, but they are affected by metrical conditions.

Proportional use of L*H and H* by Speaker as an Effect of Mean Speech Rate



A. Correlation between L*H and speech rate B. Correlation between H* and speech rate

*Figure 6.7 Linear regression of per-speaker proportional use of L*H and H* and mean speech rate in prenuclear pitch accents in the* `pn_ana` *and* `pn_foot` *subcorpora.*

Based on auditory impressions, it also seemed that realisations of L*H versus H* were partially determined by speech rate. This was confirmed by plots (Figure 6.7) correlating the proportion both of H* and of L*H per speaker against each speaker's average speech rate (syllables per second) for `pn_ana` and `pn_foot` data combined. This suggests a negative correlation between speaker average speech rate and use of L*H, and a positive correlation between speaker average speech rate and H* use. The r-squared value for `L*H ~ mean_speech_rate` is .38, p = .041, while it is .41, p = .034, for `H* ~ mean_speech_rate`, indicating a significant effect of speech rate on the realisation of each PA. However, it also seemed that male speakers tended to speak faster than the female speakers, so there may be an interaction of gender and speech rate. Therefore, to assess the general effect of speech rate and gender, two linear models were tested, as shown in Equations 6.1 and 6.2.

`L*H ~ gender * mean_speech_rate` (6.1)

`H* ~ gender * mean_speech_rate` (6.2)

`L*H` and `H*` here refer to the proportion of the target pitch accent per speaker in the adjusted data. An ANOVA of each model indicates that the only significant effect is `gender`, $F(1,7)$ = 10.72, p = .014 in the `L*H` model and F(1,7) = 14.11, p =.007 in the `H*` model. Therefore, we cannot reject the null hypothesis for `mean_speech_rate` or the `gender:mean_speech_rate` interaction, i.e., there does not appear to be an effect of either.

6.5.3 Word Boundary Effects in Prenuclear Pitch Accents

Figure 6.8 shows the adjusted counts of prenuclear PAs for each pair of utterances in the H-Corpus. In each pair, the target with the earlier word boundary is shown on the left and the target with the later boundary on the right. For both the first and the third pairs, there is a noticeable increase in L*H occurrences when the right word boundary is later, i.e., at the end of the second syllable and not the first, stressed syllable. In the first pair, L*H occurrences increase by 11, from 35 to 46. In the third pair, L*H occurrences increase dramatically from 34 to 50, an increase of 16. In each case, there is a corresponding decrease in H* occurrences. In the second pair, where the right word boundary occurs at the end of the second and third syllable, we see only a very minor increase, from 48 to 50.

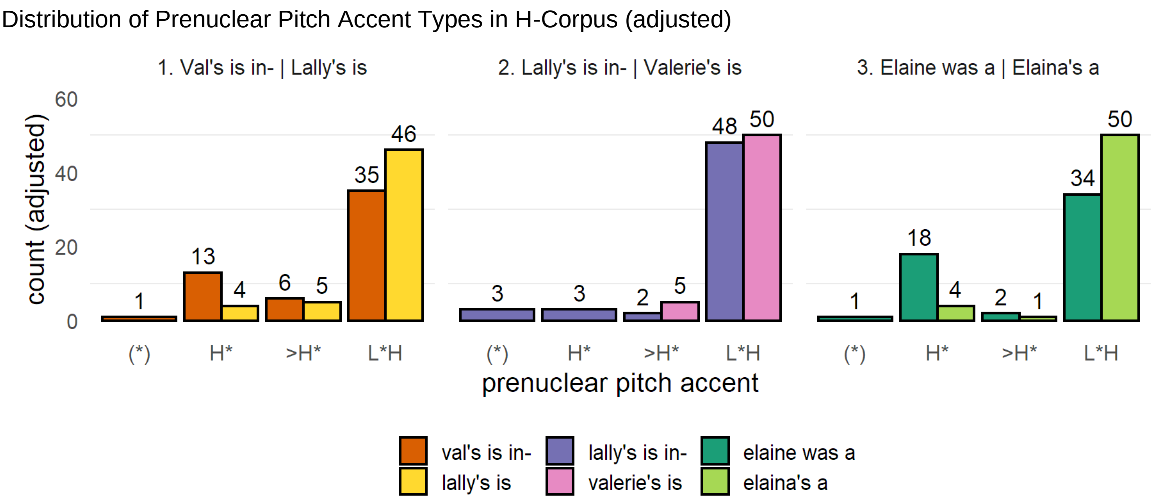


Figure 6.8 Pitch accent counts (adjusted) per target phrase in H-Corpus.

These results suggest that prenuclear pitch accent realisation is indeed subject to the effects of the right word boundary. That is, if the right word boundary is later, then L*H is more likely. However, it also seems that the distinction might only matter in terms of whether or not the boundary occurs at the end of the stressed syllable, since there is little difference in the distribution of pitch accents in pair two, where the word boundary is either one or two syllables after the stressed syllable.

6.5.4 Mixed-effects Modelling of Effects on PN Pitch Accents

Based on the initial results, it appears that anacrusis (`ana_syls`) has no effect on prenuclear pitch accents, but that foot size (`foot_syls`) and the location of the right word boundary (`wrd_end_syl`) do. Speech rate (`speech_rate`) appeared to have an effect when considered in terms of the proportional use of pitch accent per speaker, but this disappeared when `gender` was also included in the analysis. To test the effects of each of these parameters using mixed-effects models, all these parameters, along with `gender`, were treated as fixed factors. `ana_syls` was also included because it was thought that the larger data set might reveal an effect obscured in the smaller `pn_ana` dataset. `speaker` was treated as a random intercept with random slopes for `foot_syls` and `wrd_end_syl`.

Since it was decided to run the mixed-effects models on the whole A and H-Corpora rather than on the five subsets individually, three extra random effects were included to account for variation in segmental content across utterances. These were `ana_text`, `pn_str_syl`, and `nuc_pre_text`. These refer respectively to the text in anacrusis, the text of the stressed syllable, and the unstressed text of the foot, with each effect acting as a proxy for segmental variation in anacrusis and the first foot of the utterance.

Two models were constructed, one to test the likelihood of L*H (`isLH`) and the other to test the likelihood of H* (`isHStar`). The final model is shown in Equation 6.3, where `x` refers to the response parameter. Detailed results for each test can be found in Appendix F.

```
x ~ ana_syls + foot_syls + wrd_end_syl + speech_rate + gender +
    (1|speaker) + (1|ana_text) + (1|nuc_pre_text) + (1|pn_str_syl)
```
(6.3)

A Likelihood Ratio Test (LRT) was performed on each model using R's `drop1()` function. (See Chapter 5, Section 5.4.2.) The LRT of the `isLH` model indicates that the effects of `foot_syls`, `wrd_end_syl`, and `speech_rate` were significant, but that the effects of `ana_syls` and `gender` were not, as summarised in Table 6.12. The marginal R^2 of the model is .4 with a conditional R^2 of .8, indicating that the fixed effects account for 40% of the variance in likelihood of L*H, while the whole model accounts for 80%.

Table 6.12 Results of LRTs comparing full `isLH` model with using R's `drop1()` function.

effect	npar	Chisq	Df	Pr(>Chisq)	p.adj (BH)
`ana_syls`	3	564.92	4.33	.228	.254
`foot_syls`	3	574.56	13.97	.003	.006
`wrd_end_syl`	2	579.22	16.64	< .001	< .001
`speech_rate`	1	579.52	14.93	< .001	< .001
`gender`	1	567.53	2.94	.086	.108

The LRT of the `isHStar` model indicates significant effects of `ana_syls`, `wrd_end_syl` and `gender` but not of `foot_syls` or `speech_rate`. The initial analysis of the `pn_ana` dataset indicated no effect of anacrusis. However, in that dataset, the first foot of each target phrase always contained four syllables, and the four-syllable foot was already strongly associated with L*H. In the larger dataset, there is more variation in foot size across anacrusis conditions, so this biasing of the results disappears.

Table 6.13 Results of LRTs comparing full `isHStar` model with using R's `drop1()` function.

effect	npar	Chisq	Df	Pr(>Chisq)	p.adj (BH)
`ana_syls`	3	513.02	14.64	.002	.005
`foot_syls`	3	506.79	8.41	.038	.055
`wrd_end_syl`	2	518.4	18.02	< .001	< .001
`speech_rate`	1	502.15	-0.23	1	1
`gender`	1	509.7	7.32	.007	.011

6.5.4.1 Effects on the Likelihood of L*H. Looking at Figure 6.9A, there does appear to be an effect of anacrusis in the two- and three-syllable conditions, at .64 and .69 respectively, 95% CIs [.04, .99] and [.05, .99] respectively. Unfortunately, the confidence intervals are very large, so it is impossible to assume this effect would be found in a larger sample or in the population as a whole.

The initial analysis suggested that the likelihood of L*H increases with foot size, and the mixed-effects model indicates this also. However, unlike the initial analysis, the mixed-effects model suggests that the effect of foot size reaches saturation point at three syllables. This is clear from Figure 6.9B, which shows the predicted probability of L*H as a function of foot size alone in the `isLH` model. We see that the probability rises steadily from .29, 95% CI [.02, .89], in the one-syllable-foot condition up to .81 in both the three- and four-syllable conditions, 95% CIs [.18, .99] and [.18, 1] respectively. In the pairwise comparison of foot-size conditions, L*H is 10.5 [2.8, 49.4] times more likely in the three-syllable foot and 10.7 [1.89, 60.4] times more likely in the four-syllable foot compared to the one-syllable foot, *p.* < .001 and *p* = .007 respectively.

Looking at the right word boundary effect, the probability of L*H increases dramatically when the right word boundary is not coterminous with the right edge of the stressed syllable. This can be seen in Figure 6.9C, which shows the predicted probability of L*H as a function of the word-final syllable. When the lexically stressed syllable and word-final syllable are the same (the intercept), the probability of L*H is .29, but this rises sharply to .48 when it occurs in the second syllable and is only slightly higher again in the third, at .88, 95% CIs [.02, .89], [.18, .99], and [.18, 1] in turn. The likelihood of L*H is 12.7 [4, 2.44] times greater than the intercept in the two-syllable condition and 18.7 [2.44, 129.6] times greater in three-syllable condition, *p* < .001 and *p* = .005 respectively.

Speech rate was statistically significant, with the expected odds of L*H falling by 62% with each additional syllable per second, odds ratio (OR) = 0.38 [0.23, 0.63], *p.adj* < .001. The red line in Figure 6.9D indicates that the mean predicted likelihood of L*H exceeds 50% when lower than 5 syls/s. (For reference, the actual speech rate in the A- and H- ranges between 2.87 and 9.82 syls/sec.)

Turning finally to gender, the odds of a male speaker using L*H are 79% lower than those of a female speaker, OR = 0.21, 95% CI [0.03, 1.56], *p* = .125. However, when we look at the 95% confidence interval, we see that it stretches from 33:1 against to 3:2 in favour, so we cannot reject the null hypothesis here.

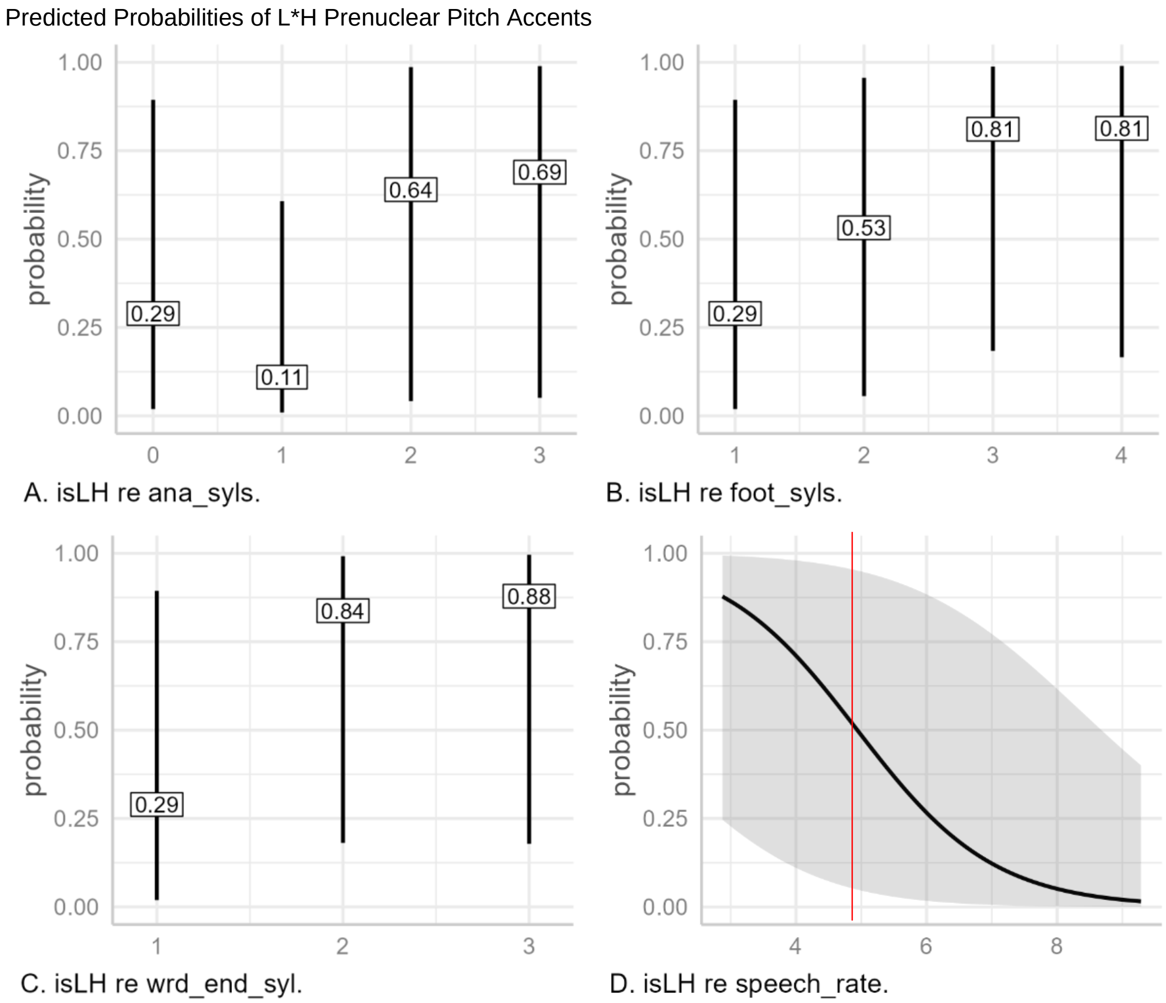


*Figure 6.9 Predicted probability of L*H as a for each fixed factor (excluding gender) in the `isLH` model[14]. Bars indicate 95% CIs. Red line Panel C indicates where the upper 95% CI crosses 50%.*

6.5.4.2 Effects on the likelihood of H*. The predicted probabilities of `isHStar`, as shown in Figure 6.10, are essentially the inverse of those of `isLH`. The probability of H* is never greater than .01 when there is more than one syllable of anacrusis, .04 when foot size is greater than two syllables, and .01 when the end of the word is later than the end of the stressed syllable, 95% CIs [0, .18], [0, .24], and [0, .15] respectively. As was noted in Section 6.5.3, H* is much less likely than L*H in prenuclear pitch accents. Therefore, it is unsurprising that the predicted probability of H* does not exceed .28 [.06, .7] as an effect of `ana_syls` alone, .12 [.02, .54] as an effect of `foot_syls` alone, or .09 [.01, .48] as an effect of `wrd_end_syl`. While the likelihood of H* increases with increased speech rate, OR = 1.16 [0.73, 1.86], *p* = .524, it achieves a predicted

[14] For each prediction plot, the first level of a categorical factor indicates the intercept of the model. Each prediction for subsequent levels shows a value which is only true while effects of other factors are held constant. For example, therefore, in the plots in panels a, b, and d of the `isLH` model above, the prediction for the first term is the same but differs for subsequent term. This may seem redundant, but I believe that plots of predicted values are often easier to understand than alternatives, especially for factors with multiple levels.

probability of only .16 [0, .84] at 10 syls/s. The greatest effect on the likelihood of H* is gender. The male speakers are estimated to be 11.4 [1.84, 70.7] times more likely to produce H* than female speakers, $p < .001$.

The results of the two mixed-effects models reflect the much greater likelihood of L*H over H* even in PN position. Increased foot size and post-stress word-end syllables are both closely associated with greatly increased likelihood of L*H. For foot size, the increased likelihood of L*H appears to reach saturation point once there are at least three syllables in the foot, while for the word-end syllable, it appears that L*H is more likely as long as the word-final syllable occurs after the stressed syllable. Conversely, H* is less likely in longer feet or when the stressed syllable boundary is not coterminous with the end of the word. The effect of speech rate is slightly more pronounced than expected as a predictor of L*H, while the effect of gender on the likelihood of H* is much stronger.

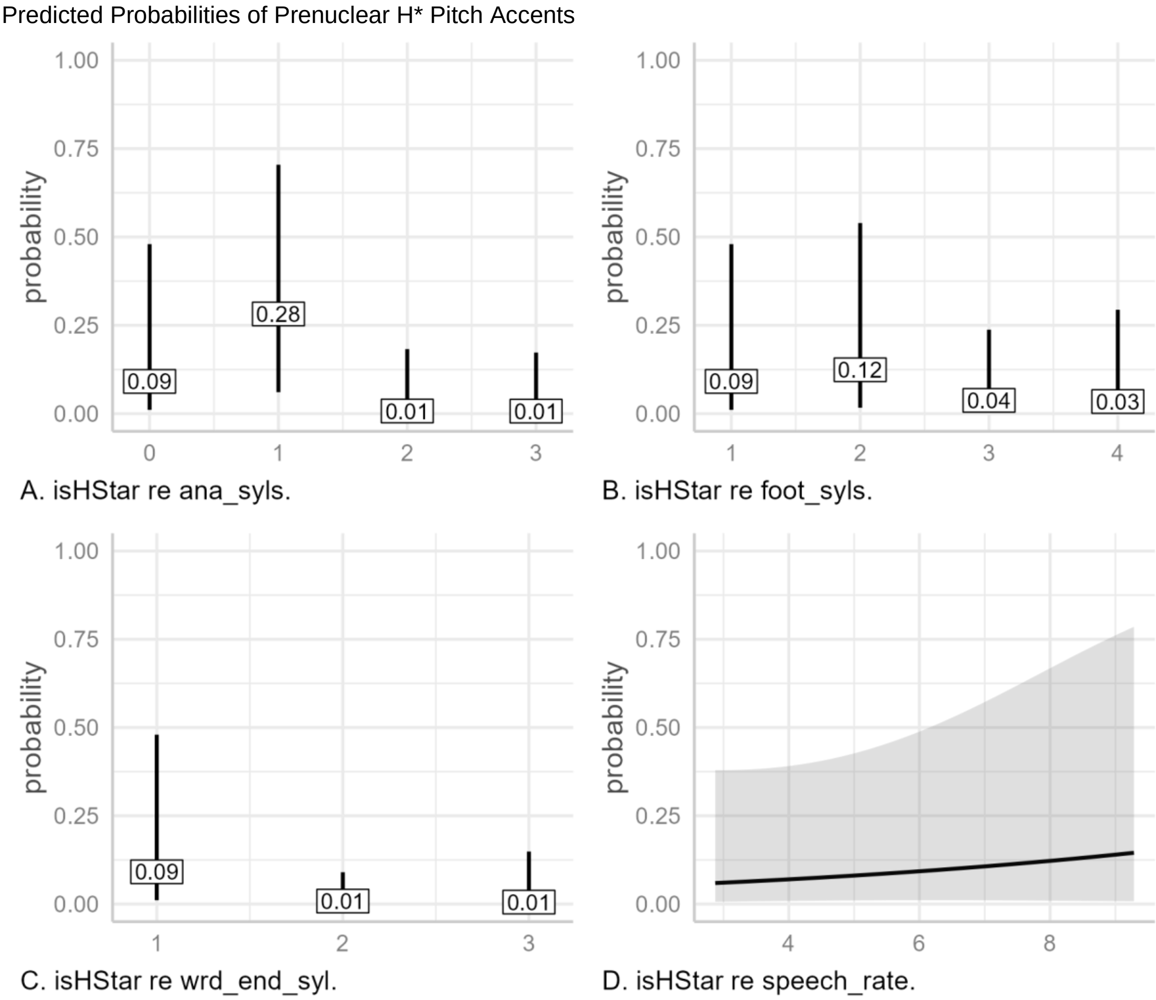


*Figure 6.10 Predicted probability of L*H as a for each fixed factor in the* `isHStar` *model (excluding* `gender`*). Bars indicate 95% CIs. Red line Panel C indicates where the upper 95% CI crosses 50%*

6.5.5 Potential Interpretations of the Phonological Analysis PN Pitch accents

Arvaniti et al. (2006b) warn that one needs to be careful in assigning a phonological representation to represent differences in what might in fact simply be variation in phonetic detail. This caveat must be borne in mind when considering the status of the >H* pitch accent, which was, after all, included in part to avoid making a categorical labelling decision on pitch events which were categorically ambiguous between L*H and H*. As such, it is possible that the >H* label is nothing more than the phonological representation of phonetic detail.

More importantly, it should also be noted that the results presented here show that prenuclear L*H becomes more common as foot size increases while H* becomes proportionally less common, to the extent that H* was entirely absent from the data described in Section 6.5.2. Similarly, L*H is more likely when the stressed syllable is not also the final syllable of the foot. As mentioned in Section 6.1, one interpretation of these results is that PN pitch accents labelled as H* are themselves surface realisations of an underlying L*H which has “succumbed” to metrical or lexical pressure.

One explanation for this is that there is some kind of anticipatory truncation in which the rise from the L tone is truncated. Such truncation has been noted in utterance-initial nuclear pitch accents of Greek wh-questions (Arvaniti & Ladd, 2009), in which a leading L tone is salient when there are unstressed syllables preceding the stressed syllable, but the rise is pre-emptively truncated when the phrase begins with the stressed syllable associated with the nuclear pitch accent. That is, superficially, the pitch accent appears to be H* when the sentence begins with a stressed syllable, but the phonological representation is seen to be L+H* or possibly L*+H when unstressed syllables precede the stressed syllable. (The authors opt for the latter based on the analysis proffered previously in Arvaniti et al. 2006b.)

Similarly, in the current case, prenuclear accents analysed as H* may in fact be truncated L*H pitch accents. If this is the case, the phonetic analysis should show that the temporal alignment of L targets is affected by the tonal crowding, with L being aligned increasingly earlier as the number of syllables in the foot decreases, thus forcing the H target to be aligned earlier. If this is indeed the case, it would suggest that the apparent H* pitch accents may indeed be a phonological representation of what is essentially a matter of phonetic implementation. The phonetic analysis of PN nuclear targets is presented in Sections 6.6.2.1 and 6.6.2.2.

An alternative explanation for the apparent disappearance of L tones might be that the L tone associated with the stressed syllable is simply deleted when tonal crowding effects are too strong. This latter alternative suggests a phonological strategy for dealing with tonal crowding effects.

6.6 Phonetic Analysis and Results

This section begins with an analysis of prenuclear peak alignment as an effect of the lexical boundary before moving on to the LMEM analysis of tonal target parameters. The separate analysis of lexical boundary effects does not use linear mixed-effects models (LMEMs), as it is concerned with potential bimodality in PN peak alignment. It is believed that this issue is best assessed using density plots to analyse peak alignment because these show if there is a modal, bimodal, or multimodal distribution of the data. However, lexical boundary effects will also still be included in the subsequent LMEM analyses.

6.6.1 Distribution and Temporal Alignment of PN Peaks in the H-Corpus

In the H-Corpus, L*H is the most common pitch accent (n = 250) while H* and >H* are much sparser (n = 42 and n = 19 respectively). To avoid adding pitch accent type as an additional confounding factor, L*H pitch accents alone are considered here. Alignment effects of pitch accents are still considered in mixed-effects modelling in subsequent sections.

An important caveat must be mentioned before continuing. The H-Corpus was designed so that there were three pairs of target phrases, with the key variable in each phrase being the syllable count in the word with lexical stress in the first foot. However, sets one and two also vary in segmental content. The effect that this variation might have on the results was (quite naively) not considered until after the data had been collected. Therefore, one must be careful when trying to interpret the results from these two sets. Fortunately, the third set contains two target sentences which are segmentally and morphosyntactically very similar, namely *Elaine was a nanny* and *Elaina's a nanny*. The only segmental difference between these is /wəz ə/ in the first target and /əz ə/ in the second, while the only morphosyntactic difference is in the tense of the verb *be*. Therefore, any differences in peak alignment between these two sentences can be attributed more confidently to word boundary effects.

6.6.1.1 Analysis of Peak Timing using Syllable-normalised Time. Dot plots of PN peak temporal alignment of L*H accents were generated in R using `ggplot2` (Wickham, 2016), with the bin width set to the default (1/30 of total range). This facilitates the assessment of distribution density of peak timing, much like a histogram, but with each dot representing an individual token. Figure 6.11 shows a dot plot of peak alignment for each pair of target phrases in the H-Corpus per speaker using syllable normalised time (Section 6.3). Each panel begins with the stressed syllable in the foot. The first six rows show the distribution density for the female speakers and the final five rows for the males, with speaker IDs listed on the right.

At a glance, we can see that the male speakers (with the exception of M10) typically align the peaks earlier than the female speakers. We can also see that for each pair, the phrase with the earlier word boundary is aligned earlier than its counterpart with the later word boundary (dark and light circles respectively). The only exception to this is F5, who only produced one PN L*H token in *Lally's is* (late word boundary), the peak of which is aligned in the second syllable of the foot, while each peak of *Val's is in-* (early word boundary) is aligned in the middle of the third syllable (Figure 6.11, first row, first column).

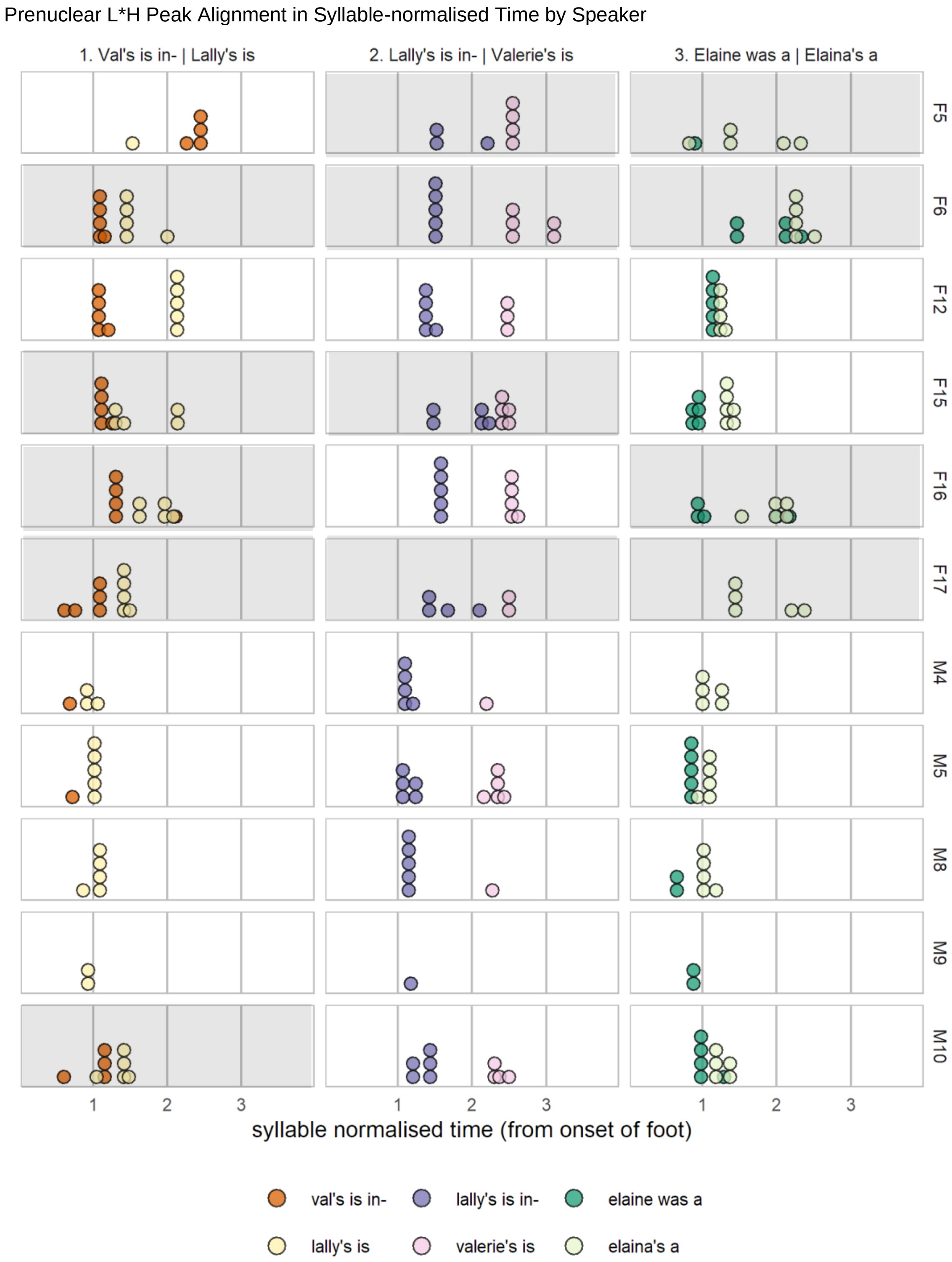


*Figure 6.11 Dot plot of syllable-normalised peak timing in L*H PNs in the H-Corpus. Each panel begins with the stressed syllable, and each dot represents an individual utterance, with bin width set to 1/30 of the total range (`geom_dotplot` default). Panels highlighted in grey indicate instances with noticeable variation in peak alignment across repetitions of the same utterance.*

With the exception of M10, the male speakers are consistent with each other in their alignment of targets. In feet where the word-end syllable is also the stressed syllable (*Val's is in-* and *Elaine was a*), peaks are aligned somewhere between the final third and the end of the first syllable. Each of these represents a case where the word boundary is at the end of the stressed syllable. In phrases where the right word boundary is in the second syllable of the foot (*Lally's is*, *Lally's is in-*, and *Elaina's a*), peaks are aligned near the boundary of the first and second syllable. In the phrase where the right-word boundary is at the end of the third syllable (*Valerie's is*), peaks occur in the first half or towards the middle of the third syllable, i.e., in the word-final syllable of *Valerie*. Thus, we see that as the number of syllables before the end of the word increase, the male speakers—with the exception of M10—tend to align the targets later in a consistent manner. However, this represents a sample of four, so it is unwise to make generalisations about male DCE speakers in general from this data. M10 aligns the peaks of *Lally's is in-* and *Valerie's is* (column 2) in a similar manner to the other male speakers, but the alignment patterns are different for the other utterances and also suggest speaker-internal variation in alignment strategies.

Among the female speakers—with the exception of F12—what is most striking is the greater intraspeaker variation in peak alignment. Instances where intraspeaker variation is evident are indicated by panels shaded in grey. We see that the right edge of the first syllable is the most common location of peak alignment in phrases where the right word boundary is the end of the stressed syllable (*Val's is in-* and *Elaine was a)*, but this is not consistent speaker-internally. For example, F6 aligns the peak in *Val's is in-* (first column, second row) just after the boundary, but the peaks in *Elaine was a* are all aligned much later (third column), in either the second or third syllables. Unlike the other female speakers, F5 consistently aligns the peak of *Val's is in-* in the middle of the third syllable, and F16 aligns the peak with the start of the third syllable on one occasion.

There is similar inter- and intraspeaker variation in the cases where the word ends in the second syllable. Of the female speakers, only F12 and F17 are consistent in the peak alignment in *Lally's is-* (first column), where it is aligned at the start of the third syllable for F12 and in the middle of the second syllable for F17. The other female speakers show a tendency to use both alignment locations. A similar pattern occurs both in *Lally's is in-* and *Elaina's a*, although the second anchoring point is a little later in these cases.

For the phrase with the word-end syllable in the third syllable (*Valerie's is*), most of the female speakers align the peak in the centre of the third syllable, much like the males. However, there are two occasions where F6 aligns with the beginning of the following word (second row).

In summary, there is a general and consistent tendency to align peaks earlier when there are fewer syllables before the word boundary, with males being most consistent on this front. However, it also appears that there are alternative locations for anchoring peaks.

6.6.1.2 Analysis of Peak Timing using Grand Syllable-mean Normalised Time. In order to further assess variation in peak alignment, data were analysed in terms of grand syllable-mean normalised time. A combined density and dot plot for each utterance was produced (Figure 6.12), which shows peak alignment trends over time across all repetitions of each target utterance. To avoid an imbalance in per-speaker token representation for each target, each plot only includes tokens from speakers who produced at least four L*H tokens of the target phrase. (The selection is shown in Table 6.14.) Unlike syllable-normalised time, grand syllable-mean normalised time is excellent for reflecting syllable duration in a realistic manner. However, also unlike syllable-normalised time, it is not good for comparing different phrases. Therefore, each target phrase is plotted separately.

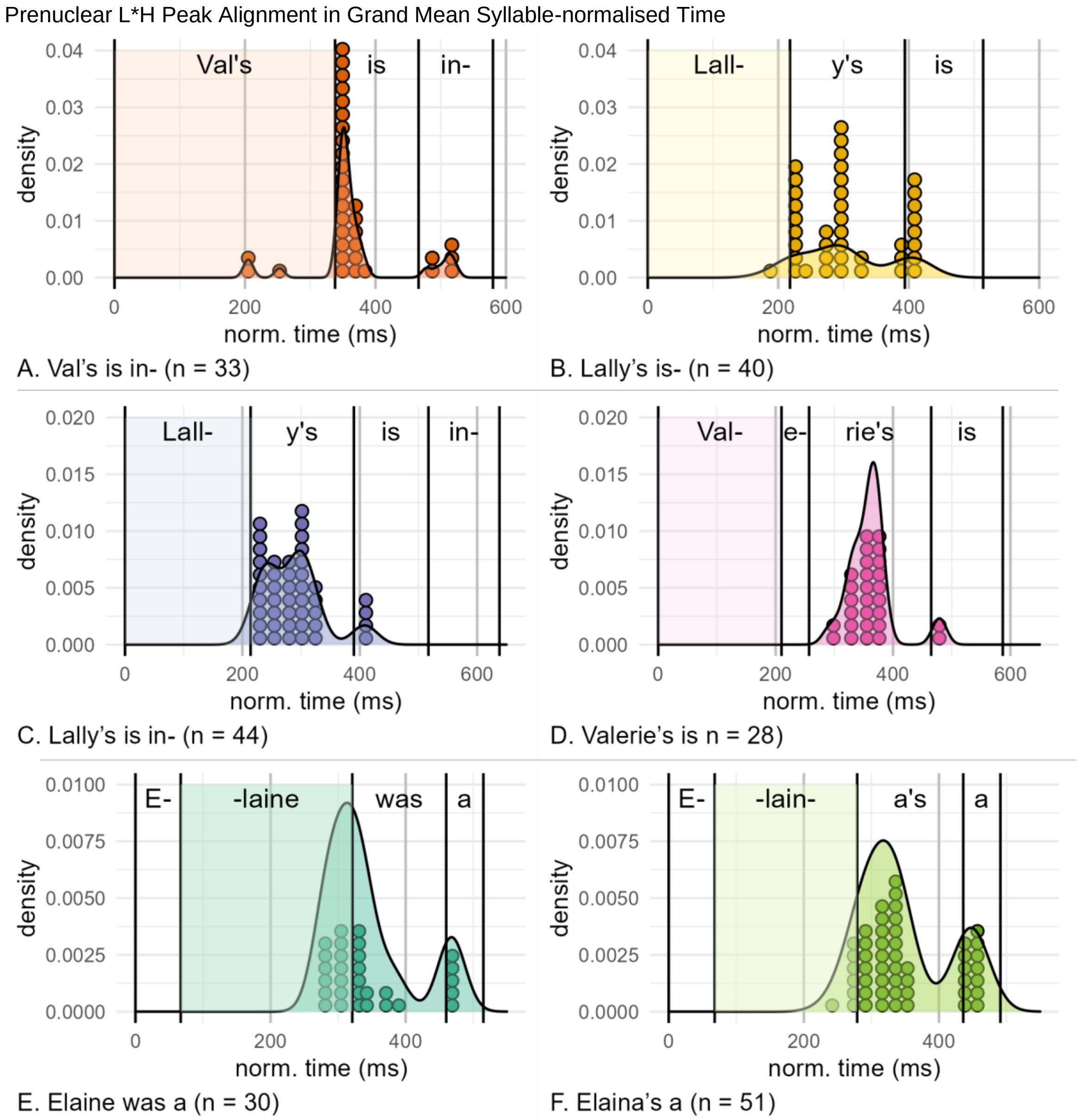


Figure 6.12 Density and dot plots for the prenuclear portion of each phrase in the H-Corpus plotted by grand mean syllable-normalised time. The lightly shaded rectangle indicates the stressed syllable while vertical black lines indicate syllable boundaries. Each panel states the number of tokens after filtering out underrepresented speakers.

*Table 6.14 Raw counts for L*H per speaker and target utterance in the H-Corpus. Cells highlighted in grey represent data used for the generation of density / dot plots using grand mean syllable timing.*

pn_wrds	F5	F6	F12	F15	F16	F17	M4	M5	M8	M9	M10	Total
val's is in-	4	5	5	5	5	5	1	1	0	0	4	35
lally's is	1	5	5	5	5	5	3	5	5	2	5	46
lally's is in-	3	5	5	5	5	4	5	5	5	1	5	48
valerie's is	4	5	3	5	5	2	1	5	1	0	4	35
elaine was a	1	5	5	5	5	0	0	5	2	2	5	35
elaina's a	5	5	5	6	5	5	5	5	5	0	5	51

In *Val's is in-* (Figure 6.12A), we see that the majority of peaks are aligned at roughly 350 ms, just after the edge of the stressed syllable, which is also the end of the word. In *Lally's is* (Panel B), there appears to be a three-way split. One cluster is at the edge of the stressed syllable (c. 215 ms). There is a second one at around 300 ms in the middle of the second (word-final) syllable, and a final one just after the end of the word (c. 400 ms), at the boundary of the second and third syllables. Thus, the first and third clusters resemble the syllable and lexical alignment of *Val's is* because one of the peaks clusters at the right edge of the stressed syllable and the other at the word boundary; in *Val's is in*, however, the right-word boundary and stressed-syllable boundaries happen to be the same location.

Lally's is in- (Panel C) is very similar to *Lally's is* (Panel C), in that there are three clusters, one near the right edge of the stressed syllable, one in the middle of the second syllable and one just after the right edge of the word boundary. The absolute timing of these is also almost identical to *Lally's is*, which is unsurprising as they are also segmentally the same.

In *Valerie's is* (Panel D), there is one large cluster, peaking at around 380 ms, which is similar to the absolute timing of the peak density in *Val's is in*. However, in this case, the cluster is in the middle of the third (word-final) syllable. This makes the main region of peak alignment similar to the second clusters in both *Lally's is* and *Lally's is in-* in that they all occur in the middle of the word-final syllable of the word with lexical stress.

In the third pair (Panels E and F), we see an apparent bimodal distribution. In *Elaine was a* (Panel E), there is a cluster which peaks at the end of the stressed syllable (also the end of the word) at around 310 ms, and a second small cluster at the start of the third syllable in the foot (c. 460 ms), a full syllable after the end of the word. In *Elaina's a* (Panel F), there is a large cluster which peaks near the centre of the last syllable in *Elaina's*, at a similar time point to the one in *Elaine was a*. This is followed by a second smaller cluster at the start of the third syllable, just after the right edge of the word, with a density peak around 450 ms. Thus, we see that the bimodal peaks are similarly timed. However, the first density peak in *Elaine was a* is at the end of the word/stressed syllable, just as in *Val's is in-*. This is also similar to the first cluster in each of the *Lally* phrases in that it is also at the right edge of the stressed syllable. The second density peak in *Elaine was a* is not associated with the boundary of the stressed syllable or the final syllable of the word with lexical stress but occurs one syllable after the end of the word. In *Elaina's a*, the first large cluster peaks towards the middle of the word-final syllable, much like the second clusters in both of the *Lally's* phrases (Panels B and C) as well as in *Valerie's is* (Panel D). In short, with the final two utterances, we see a pattern which could be interpreted in terms of absolute timing or in terms of anchoring to the word-final syllable of the word with lexical stress (with the exception of the second peak in *Elaine was a)*.

When we consider the six target utterances together, the main clusters of peak alignment occur anywhere between 300 and 380 ms. Of course, this is time-normalised data, so the actual timing would vary from speaker to speaker. In general, however, peak alignment seems to be associated with three lexical and syllable locations. These are:

1. The right edge of (or just after) the stressed syllable.
2. The middle of the last syllable of the word with lexical stress.
3. The right edge of (or just after) the word with lexical stress.

These three options appear to create bimodal or multimodal distributions of peak alignment in four of the six target phrases.

One way to test for unimodal/multimodal distributions is via the Hartigan dip test of unimodality (Hartigan & Hartigan, 1985). In this test, the null hypothesis is that the distribution is unimodal, so the alternative hypothesis ($p < .05$) means that the distribution is at least bimodal. Unfortunately, the data here may not be ideally suited to such a test since the observations are not independent, i.e., they are grouped by speaker, and (to the best of my understanding) the Hartigan dip test is not designed for nested observations. However, it can still be used to get a general impression of the presence or absence of multimodality. Bearing these issues in mind, the test was conducted on the four sets in which there appeared to be bi- or multimodal distributions (Figure 6.12, Panels B, C, E, F) using the `diptest` module in R (Maechler, 2021), with the results summarised in Table 6.15. After adjustment for multiple testing using the Benjamini and Hochberg method (see Chapter 5, Section 5.4.3), three out of the four results were found to be significant. *Elaine was a* (Panel E) is non-significant (*p.adj* =.366), but this is unsurprising, as there are only five tokens in the second smaller peak-alignment cluster.

*Table 6.15 Hartigan dip test for unimodality / multimodality in peak alignment of L*H in the H-Corpus for potentially bi-/multimodal phrases. N.B. These results should be considered with caution since the observations are not independent (i.e., they are grouped by speaker).*

set	observations	D.value	p.value	p.adj (BH)	signif.
Lally's is	40	.09	.006	.011	p < .05
Lally's is in-	44	.08	.036	.048	p < .05
Elaine was a	30	.07	.366	.366	
Elaina's a	51	.09	.004	.011	p < .05

These results suggest that temporal alignment of peaks is not simply an effect of the number of syllables in the foot (although it does seem to play some role in skewing the peaks slightly later), but that there is a range of peak alignment options for the speaker, which are largely a matter of preference. Thus, in the sections which follow, we need to be aware of the fact that the distribution of peak alignment may not simply be accounted for by the lexical and metrical structure of the phrase, nor by the segmental string, but that an amount of error in the results may be inevitable as we cannot account for variation in speaker preference across tokens.

6.6.2 Mixed-effects Modelling of Prenuclear Pitch Accents

Data was trimmed to exclude utterances without prenuclear pitch accents. Only L* and L*H PNs were included in the analysis of L targets, and only L*H, >H*, and H* were used for the analysis of H targets. Slope and excursion size were measured for L*H targets only.

Six phonetic response parameters were tested. These were the temporal alignment and f_0 of each L and H target alongside the excursion size and slope of contours in L*H pitch accents. As elsewhere, f_0 minima and maxima are used to represent tonal targets, while temporal alignment is measured in milliseconds from the onset of the vowel in the stressed syllable. f_0 is measured in semitones centred around speaker median. f_0 excursion size is calculated as the difference between L and the H target f_0. f_0 slope is calculated as the slope of the linear regression of f_0(t) between the L and H time points. However, in the modelling, slope was logged since it resulted in a better model fit. The response parameters are summarised in Table 6.16.

Table 6.16 Response parameters for PN and nuclear pitch accent analysis.

Parameter type	Parameter code	Description
L target alignment	`l_t`	Time in milliseconds from vowel onset to f_0 minimum
L target f_0	`l_f0`	f_0 minimum measured in ST, centred on speaker median f_0.
H target alignment	`h_t`	Time in milliseconds from vowel onset to f_0 maximum
H target f_0	`h_f0`	f_0 maximum measured in ST, centred on speaker median f_0.
f_0 excursion	`f0_exc`	f_0 maximum - f_0 minimum (ST)
slope	`log_lh_slope`	Log slope of linear regression of f_0(t) between L and H

For each response parameter associated with tonal targets, six fixed effects were included in the model. These were pitch accent, anacrusis, foot size, right-word boundary, presence/absence of word break at the foot onset, and gender (see Table 6.17). For the analysis of compression and truncation effects, only `foot_syls` was of interest, so it was the only fixed factor included in the models, with the other factors included as random intercepts. As there were a large number of random effects, resulting in convergence and singularity issues, the `step()` function (see Chapter 5, Section 5.4.1) was used to help reduce the number of factors by eliminating non-significant random effects in the models.

Table 6.17 Fixed Effects for PN pitch accent analysis.

Parameter type	Parameter code	Comments / levels
pitch accent	acc_phon	Not included in slope and excursion size analysis
Anacrusis	ana_syls	Zero to three syllables
foot size	foot_syls	One to four syllables
right-word boundary	wrd_end_syl	One to three syllables
word break at foot onset	pn_new_word	True/False
Gender	gender	Female/Male

After an optimal model was established, each model typically had random intercepts of speaker (speaker) and stressed syllable (pn_str_syl). Stressed syllable had been added to the models to mitigate random effects introduced once the dataset was enlarged. (Other random effects were tested, but they either caused convergence issues or had no meaningful effect on the output.) A range of random slopes were tested, most of which led to convergence or singularity issues. In the final models, foot_syls was the only random per-speaker slope used. For the sake of clarity, the random effects for each model are listed in Table 6.18.

Table 6.18 Random effects for PN LMEMs in analysis of A- and H-Corpora.

Response	Random effects in model
l_t	(1 | speaker) + (1 | pn_str_syl)
l_f0	(1 + foot_syls | speaker)
h_t	(1 + foot_syls | speaker) + (1 | pn_str_syl)
h_f0	(1 + foot_syls | speaker) + (1 | pn_str_syl)
f0_exc	(1 | speaker) + (1 | gender) + (1 | ana_syls) + (1 | pn_str_syl)
log_lh_slope	(1 | speaker) + (1 | gender) + (1 | ana_syls) + (1 | pn_str_syl) + (1 | wrd_end_syl)

The following subsections present the analysis of the L target models, followed by those of the H targets, and finally by the analysis of the f_0 excursion and slope. In some cases, the results warranted further investigation, so additional models were tested. These are presented separately. Full LMEM results for PN analysis can be found in Appendix G.

6.6.2.1 L Targets. An ANOVA of the l_t model indicates three statistically significant effects, namely acc_phon, $F(1, 471.2) = 6.37$, $p.adj = .024$, wrd_end_syl , $F\ (2, 350.7) = 8.84$, $p.adj < .001$, and gender, $F(1, 9.8) = 100.26$, $p.adj < .001$. The model has a marginal R^2 of .42 and a conditional R^2 of .83, indicating that the fixed effects count for 42% of the variance in l_t while the whole model accounts for 83% of the variance. An ANOVA of the l_f0 model also indicates three significant effects. In this case, they are ana_syls, $F(3,\ 451) = 12.97$, $p.adj < .001$, pn_new_word, $F(1,\ 460) = 8.72$, $p.adj = .007$, and gender, $F(1,\ 6.2) = 12.88$, $p.adj = .022$. gender is the only effect which is significant for both l_t and l_f0. The l_f0 model has a marginal R^2 of .19 and a conditional R^2 of .41, indicating that the fixed effects count for 19% of the variance in `l_f0`, while the complete model accounts for 41%.

Pitch Accent Effects. Beginning with the effects of pitch accent (`acc_phon`), shown in Figure 6.13, we see that the mean estimated timing of L*H is 46 ms, 95% CI [11, 82]. L* is an estimated 24 [5, 43] ms later, p = .012. The estimated f_0 of the L target is -0.7 [-1.6, 0.1] ST, with the estimated mean f_0 in L* 0.6 [-1.6, -0.36] ST lower, p = .211, sitting at -1.3 [-2.5, -0.1] ST. The 95% CI for the L* `l_f0` is large, but this is unsurprising since L* accounts for only nine out of 498 prenuclear PA tokens. As such, it is unwise to read too much into the estimated differences between L* and L*H, beyond the fact that it appears to be aligned slightly later, which may be due to the fact that there is no backwards pressure on it to accommodate an upcoming H target.

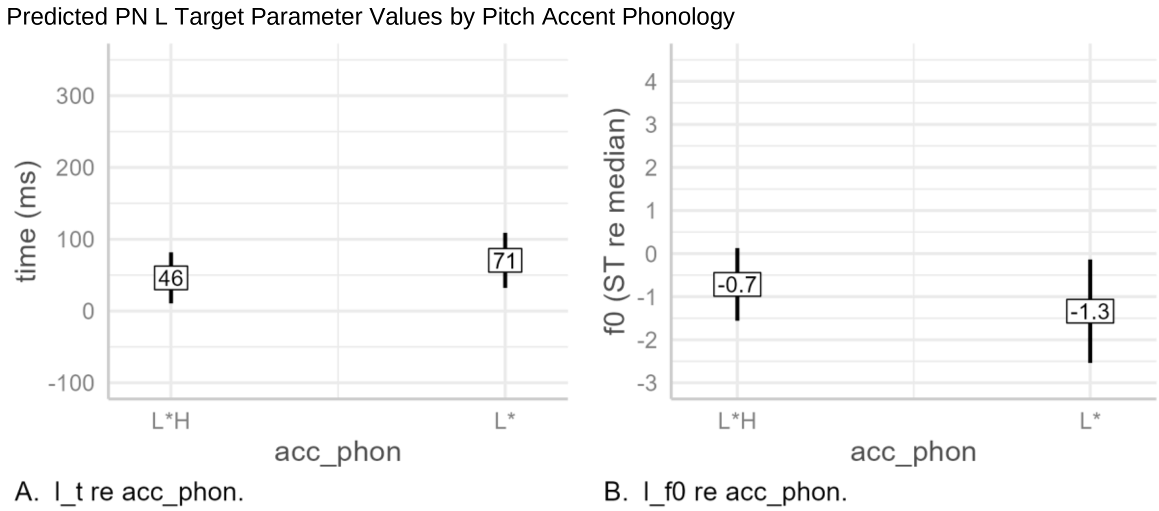


Figure 6.13 Predicted values of L targets based on pitch accent effects alone .

Anacrusis Effects. Turning to anacrusis effects (Figure 6.14), it appears, superficially at least, that the addition of a single syllable of anacrusis causes the L target to align much earlier (Panel A). For example, when there is no anacrusis, the estimated mean alignment of the target is 46 ms, 95% CI [11, 82], while for the one syllable condition it drops to 14 [-62, 89] ms, after which it rises slightly until it reaches 26 [-51, 102] ms. This implies that it is simply the presence or absence of anacrusis which affects the alignment of the L target. However, the CI for `ana_syls` conditions one to three are exceedingly large, as can be seen in Panel A, and there is no significant difference between the zero-syllable condition and other levels of `ana_syls`. A summary of this is shown in Table 6.19. Note again, the exceedingly large 95% CIs make the estimated mean differences unreliable. This suggests that the apparent effect of the presence or absence of anacrusis on the timing of L targets is too diffuse to describe as a trend.

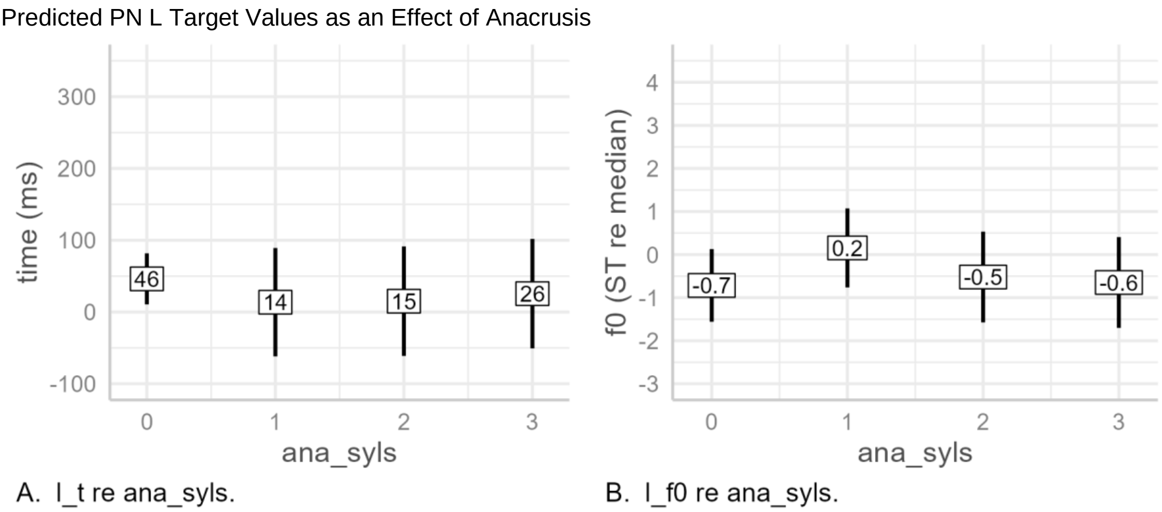


Figure 6.14 Predicted values of L targets based on anacrusis effects alone .

Table 6.19 Summary of effect of levels 1-3 of `ana_syl` *as slope and* `ana_syls0` *as intercept in the model* `l_t ~ acc_phon + ana_syls + foot_syls + wrd_end_syl + pn_new_word + gender + (1 | speaker) + (1 | pn_str_syl)` *for PN pitch accents.*

intercept	slope	est. (ms)	2.5% CI	97.5% CI	SE	|*t*|	df	p
ana_syls0	ana_syls1	-32	-136	701	41.2	0.79	5.46	.463
ana_syls0	ana_syls2	-31	-134	72	41.26	0.75	5.52	.482
ana_syls0	ana_syls3	-21	-124	82	41.3	0.5	5.54	.637

When we consider anacrusis and `l_f0` (Figure 6.14B), there is an odd effect. The estimated means show very little difference in the effects of `ana_syls0`, `ana_syls2`, and `ana_syls3`. However, there is a large effect of `ana_syls1`, in which the `l_f0` estimate is much higher than the other levels of `ana_syls`, at -0.7 ST, 95% CI [-1.6, 0.1]. That is, the estimated difference between `ana_syls1` and the other levels of `ana_syls` ranges from 0.68 [-1.1, -0.2] ST in `ana_syls2` to 0.87 [0.52, 1.2] ST in `ana_syls0`, p =.002 and p < .001 respectively.

It is difficult to explain this phenomenon, beyond the possibility that there are two conflicting underlying causes, one which causes f_0 lowering at the onset of the IP, and one associated with f_0 lowering when the extra anacrusis provides more planning time to realise a lower f_0 target. However, this still does not adequately explain the anomalously higher mean f_0 in the single syllable of anacrusis.

Foot-size Effects. The number of syllables in the foot (`foot_syls`) has almost no effect on the scaling of L target and only a very minor effect on its alignment (Figure 6.15). In one regard, this is unsurprising as we expect foot size to have a stronger effect on the alignment of H targets than on L targets, especially in L*H.

It was noted in Section 6.5.5 that L targets might align earlier as an effect of foot size due to tonal crowding effects. That is, whenever the H target is aligned earlier due to a decrease in foot size, the L target might be aligned earlier as well. As such, this would indicate that prenuclear pitch accents analysed as H* may simply be surface realisations of L*H in which the earlier portion of the rise from the L tone has been truncated to such an extent that it is no longer salient. The results indicated no significant difference in alignment between any of the one-, two-, and three-syllable foot conditions (Table 6.20). However, there is a statistically significant difference between the four-syllable foot condition and all the other conditions. That is, the L target is estimated to be aligned on average between 18 and 12 ms later in the four-syllable foot than other sizes, highlighted in yellow in Table 6.20. If this is a meaningful difference, it suggests that tonal crowding effects are absent in the four-syllable foot but present in the other foot-size conditions. Further, given that the statistically significant differences are in fact rather small, it is still only very weak evidence suggesting that tonal crowding affects the alignment of the L target. That is, if the difference between the H* and L*H was truly one of truncation of the earlier part of the rise, one would expect to see a more striking backwards shift in L target alignment in L*H pitch accents as the foot became shorter. Thus, these results lend credence to the alternative view that tonal crowding pressures from the upcoming H tone leads to the L target being deleted altogether. In other words, they favour a phonological explanation over the phonetic one (an issue discussed again in Section 6.6.2.6).

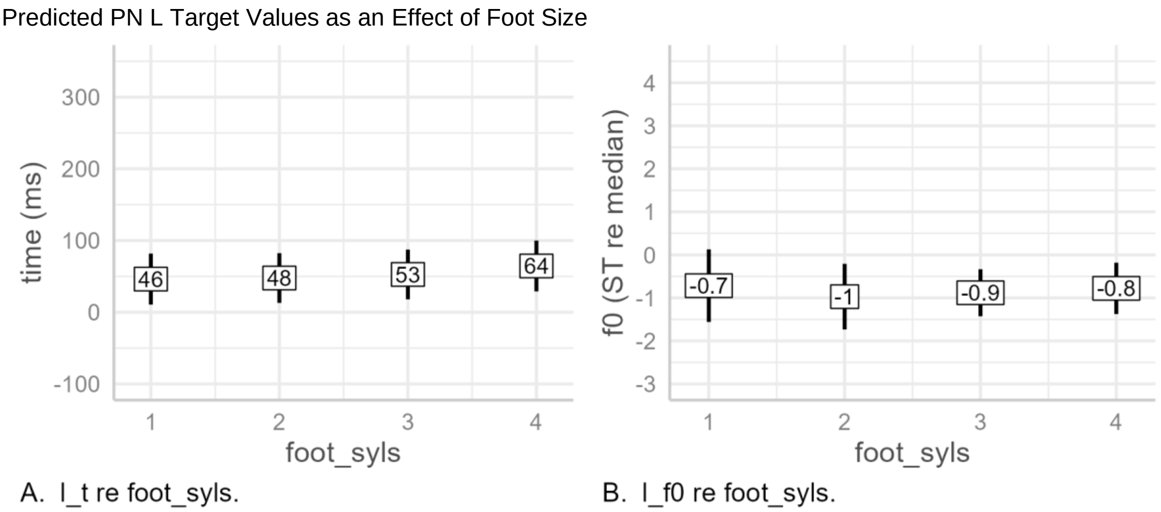


Figure 6.15 Predicted values of L targets based on foot size effects alone .

Table 6.20 Results of LMEM comparison of L target alignment (l_t) effects between the foot-size conditions. (Significant differences highlighted in yellow.)

intercept	Slope	est. (ms)	2.5% CI	97.5% CI	SE	|t|	df	p
foot_syls1	foot_syls2	2	-12	15	6.94	0.22	457.23	.827
foot_syls1	foot_syls3	6	-7	20	6.84	0.95	465.65	.342
foot_syls1	foot_syls4	18	2	35	8.53	2.14	453.5	.033
foot_syls2	foot_syls3	5	-7	17	6.32	0.79	458.11	.431
foot_syls2	foot_syls4	17	1	33	8.11	2.07	437.12	.039
foot_syls3	foot_syls4	12	2	22	5.17	2.28	466.1	.023

Word boundary Effects. It is surprising—at least superficially—that there is a word boundary effect (`wrd_end_syl`) on the timing of the L target (Figure 6.16A). When the word boundary is at the end of the stressed syllable (`wrd_end_syl1`), the predicted timing of the L boundary is quite early at 46 ms, 95% CI [11, 82], but it is later when the word boundary is at the end of the second or third syllable, at 70 [34, 106] ms and 60 [27, 95] ms respectively. The estimated difference between the first- and second-syllable boundary is 24 [12, 35] ms while it is 15 [-10, 39] ms between the first and third syllable, $p < .001$ and $p = .249$ respectively. The estimated difference between the second- and third-syllable boundaries is even smaller again, at 9 [-31, 13] ms, $p = .42$. While the difference between the first- and third-syllable boundary is not statistically significant, we do still see a general trend towards later alignment once the word boundary is no longer coterminous with the stressed syllable. This suggests that the true difference may be between words which end in a stressed syllable and those which do not.

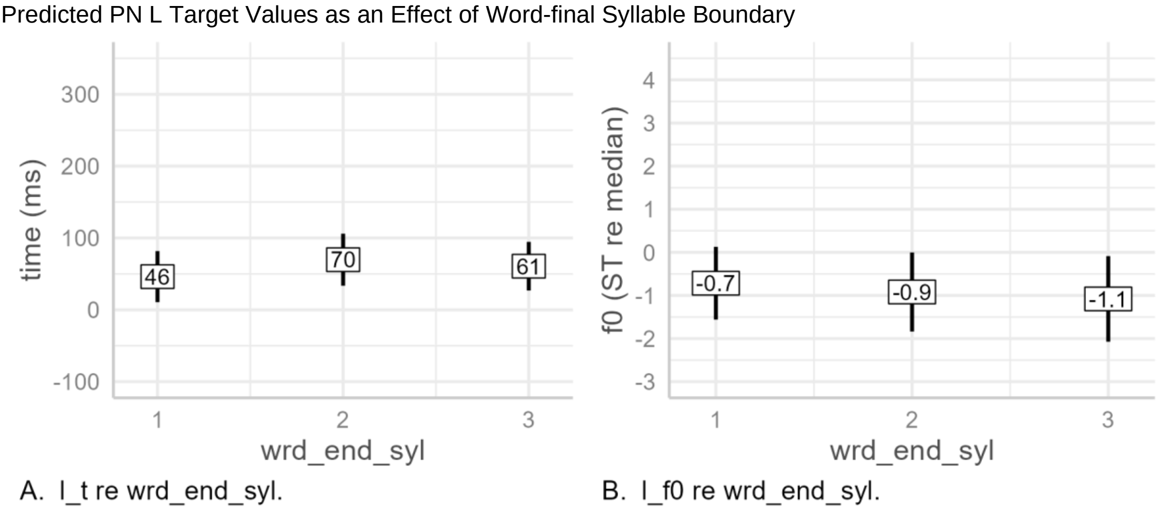


Figure 6.16 Predicted values of L targets based on word-end syllable effects alone .

In contrast to the word-end syllable, there is no significant alignment effect on `l_t` when the foot begins with a new word (`pn_new_word`) (Figure 6.17A). Even though the effect of the `True` condition appears be associated with later alignment of the L target—that is 69 ms as opposed to 46 ms in the `False` condition, 95% CIs [-25, 164], [11, 82]—the extremely large confidence interval for the true condition renders any claim about this effect spurious.

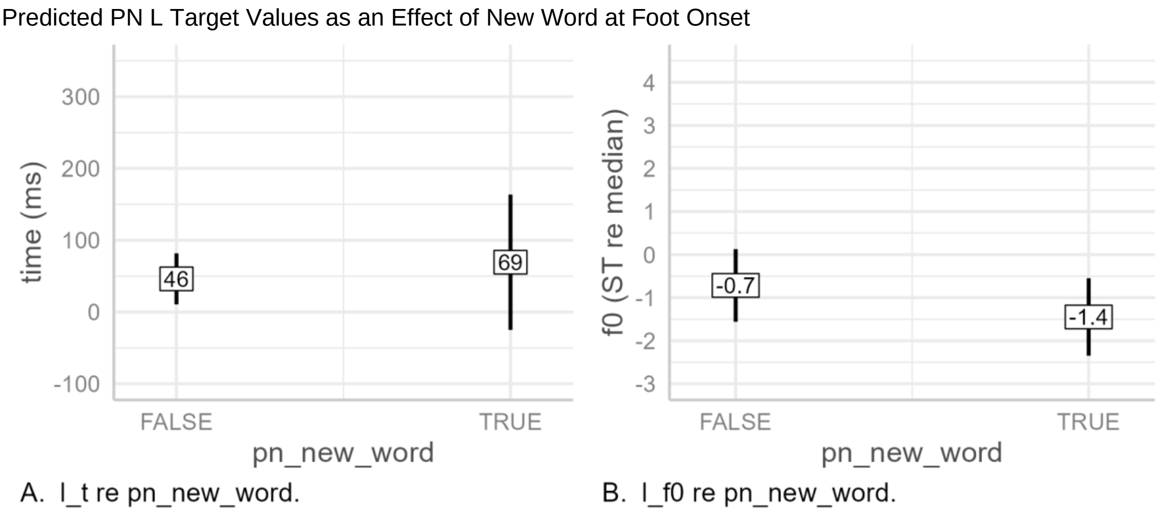


Figure 6.17 Predicted values of L targets based on word boundary effects at foot onset alone .

There is a significant effect of `pn_new_word` on `l_f0`, as can be seen in Figure 6.17B. A foot beginning with a new word is associated with an estimated 0.7 ST lowering f_0 in the L target, 95% CI [-1.2, 0.2], p = .003. While such an effect was not originally considered, and the reason for it is not completely clear, the lower f_0 may help to indicate juncture between the previous word and the current word.

Gender effects. Finally, the effect of gender on alignment is striking (Figure 6.18), with the male speakers aligning the L target an estimated 70 ms earlier than females, 95% CI [-86, -55], p. < .001. There is also a significant effect of gender on the height of the L target, with an estimated group mean increase of 1.0 [0.3, 1.6] ST, p = .011 for the `genderM` (male speakers) condition compared with the intercept (female speakers). This effect will be seen again in PN `h_f0`, but not in the nuclear pitch accents.

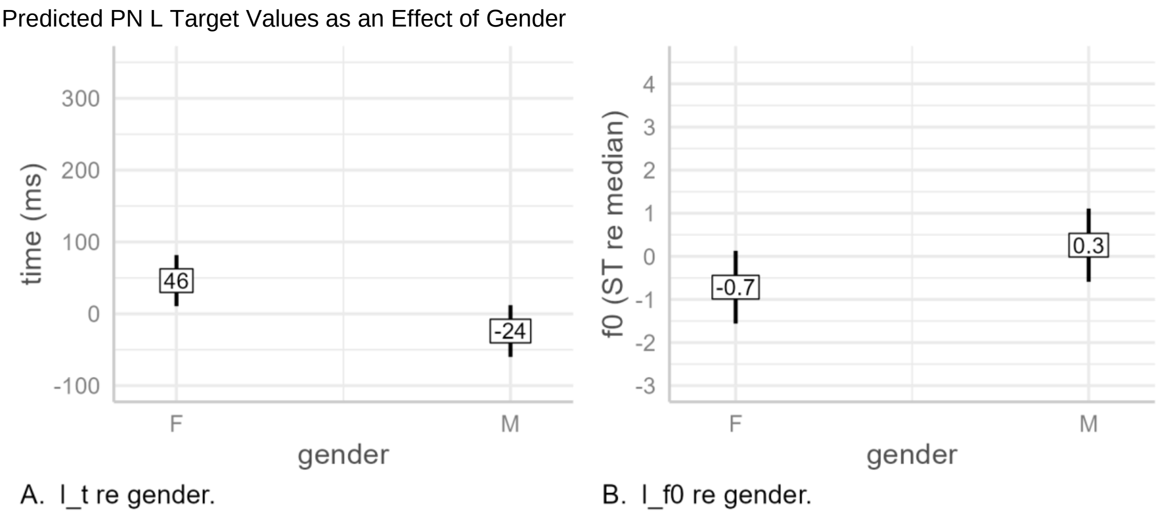


Figure 6.18 Predicted values of L targets based on `gender` alone .

6.6.2.2 L Targets with Anacrusis and Late Word Boundaries. Having analysed effects on PN `l_t` and `l_f0`, it seems that the effect of anacrusis on the timing of the L target may more accurately be viewed in terms of the presence or absence of anacrusis. Similarly, the effect of the word-end syllable on the L target may be a matter of whether or not the word and the stressed syllable are coterminous. To test this, an additional model was constructed. Non-significant factors were removed from the first `l_t` model. `ana_syls` was replaced with `has_anacrusis`, a logical parameter which is true when the phrase includes anacrusis. `wrd_end_syl` was replaced with a logical parameter called `wrd_end_syl_late`, which is true when the word-end syllable is not the stressed syllable, i.e., when it is the second or third syllable of the foot, as suggested by the analysis above. The model, including random effects, is shown in Equation 6.4.

```
l_t ~ acc_phon + has_ana_syls + wrd_end_syl_late + gender
    + (1 + foot_syls | speaker) + (1 | pn_str_syl)
```
(6.4)

An ANOVA of the model indicates that `acc_phon` (as before) is significant and so too is `gender`, with $F(1, 171.95) = 15.66$, $p.adj < .001$ and $F(1, 9.38) = 139.6$, $p.adj < .001$ respectively. While the effect of `has_ana_syls`, is not significant, `word_end_syl_late` is, with $F(1, 454.9) = 24.75$, $p.adj = .547$ and $F(1, 454.9) = 139.6$, $p.adj < .001$ respectively. The model has a marginal R^2 of .29 and a conditional R^2 of .86. Thus, while the amount of variance explained by fixed effects of the model is lower than the original `l_t` model (29% as opposed to 42%), the overall variance is roughly the same (86% as opposed to 83%).

When we look at the estimated slopes of the fixed effects (see Figure 6.19), we see that the presence of anacrusis (`has_ana_syls`) has almost no effect anyway, with an estimated mean effect of 4 ms, 95% CI [-5, 12]. As such, the earlier impression that it is simply the addition of anacrusis which causes the earlier alignment of the L target is not supported. The late word-end syllable (`word_end_syl_late`) is associated with an estimated later alignment of 27 [16, 37] ms. This appears to confirm the interpretation of the original `l_t` model, namely that the more appropriate interpretation of word boundary effects on the alignment of the L target relates to whether or not the stressed syllable and the word-end syllable are coterminous. When they are coterminous, the L target is aligned earlier.

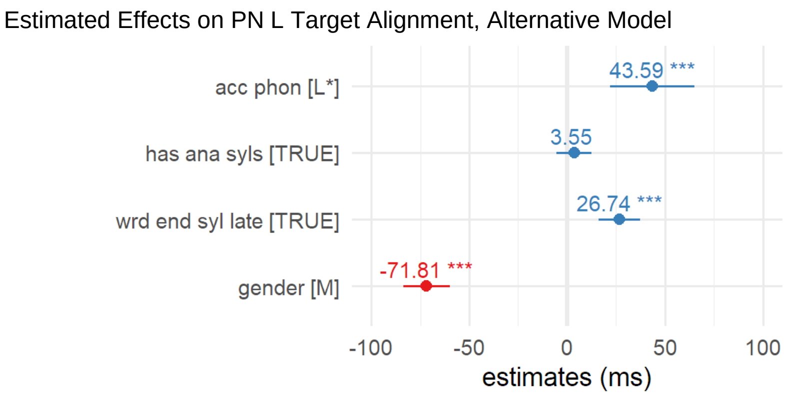


Figure 6.19 Estimated effects on `l_t` using model from Equation 6.4

6.6.2.3 H targets. An ANOVA of the `h_t` model indicates statistically significant effects for `acc_phon`, `ana_syls`, and `foot_syls`, at $F(2, 381.5) = 147.4$, $p.adj < .001$, $F(3, 14.2) = 19.84$, $p.adj < .001$, and $F(3, 28.44) = 12.5$, $p.adj < .001$ respectively. The model has a marginal R^2 of .59

and a conditional R^2 of .83. This indicates that the fixed effects explain the majority of the variance in `h_t` (59%), a larger amount than was explained by the `l_t` model (42%). An ANOVA of the `h_f0` model also indicates two significant effects, `foot_syls`, and `gender`, $F(3, 26.42)$, *p.adj* = .024 and $F(1, 9.42)$, *p.adj* <.001 respectively. The model has a marginal R^2 of .23 and a conditional R^2 of .46. The fact that the fixed factors in the H target models explain a greater deal of the variance than they did in the L target models suggests that the H target is generally more susceptible to metrical, lexical, and gender effects. (Complete model analysis and tables can be found in Appendices G5 and G6.)

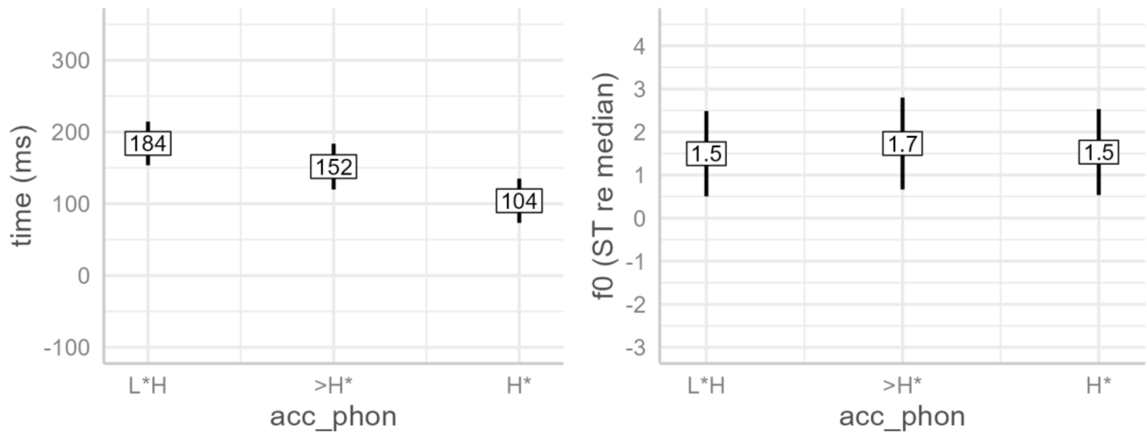
Pitch Accents Effects. Figure 6.20 shows the predicted values of `h_t` as an effect of pitch accent alone (L*H, >H*, and H*). We see a noticeable and steady decrease in the mean estimated timing of the H targets, from 184 ms in L*H, down to 152 ms in >H*, with H* the earliest, at 104 ms, 95% CIs [154, 214], [120, 184], and [73, 135] respectively. These results reflect the intuited distinctions in timing across the three pitch accent types. Moreover, the estimated mean difference between the timing of each peak is statistically significant, as shown in Table 6.21, indicating that the timing between each is distinct.

Table 6.21 Results of LMEM comparison of effects of peak alignment (`h_t`) across each PN pitch accents (H, >H*, and L*H).*

intercept	Slope	est. (ST)	2.5% CI	97.5% CI	SE	|t|	df	p
L*H	>H*	-32	-44	-20	5.99	-5.39	614.9	< .001
L*H	H*	-80	-89	-71	4.65	-17.17	287.66	< .001
>H*	H*	-48	-60	-35	6.3	-7.56	399.87	< .001

In contrast, there is almost no difference between the estimated mean f_0 of each pitch accent type, at 1.5 ST for both L*H [0.5, 2.5] and H* [0.7, 2.8], and a mere 0.2 ST higher at 1.7 [0.54, 2.53] ST for >H*. This is reflected in the absence of any statistically significant estimated mean differences in `h_f0` between pitch accent types (c.f. Table G6.10 in Appendix G6).

Predicted PN H Target Values as an Effect of Pitch Accent Phonology



A. h_t re acc_phon. B. h_f0 re acc_phon.

Figure 6.20 Predicted values of H targets based on pitch accent effects alone .

Anacrusis Effects. The predicted effects of anacrusis on the alignment of H targets look odd (Figure 6.21A). It appears that, with the addition of a single syllable of anacrusis, the mean estimated time of `h_t` falls from 184 ms to 147 ms, 95% CIs [154, 214] and [100, 194], giving the impression that the addition of anacrusis leads to earlier alignment of `h_t`. However, once a second is added, estimated mean alignment returns to roughly the same as the zero-anacrusis condition and then increases only slightly in the three-syllable condition, (i.e., `ana_syls2` is an estimated 1 [-60, 62] ms later than `ana_syls0`, and `ana_syls3` an estimated 12 [-49, 73] ms later, p = .976 and .645 respectively). Thus, the current study suggests that the addition of anacrusis does lead to earlier H alignment, but that these effects are lost with further syllables of anacrusis.

When it comes to the effects of anacrusis on `h_f0`, the addition of anacrusis is associated with an increase in f_0, but the effect dissipates beyond the two-syllable condition (Figure 6.21B). The only significant differences, however, are between the one- and three- and the two- and three-syllable condition, at -0.6 and -0.7 ST, 95% CIs [-1, 1] and [-1.2, -0. 2], p = .017 and .009 respectively.

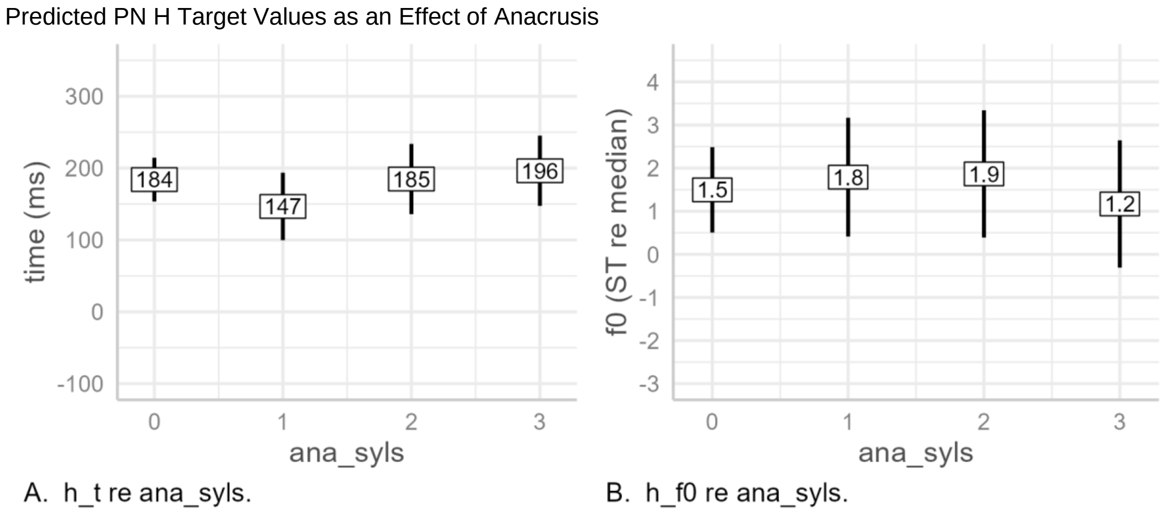


Figure 6.21 Predicted values of H targets based on anacrusis effects alone .

Foot-size Effects. The effect of foot size on peak alignment is clear, as seen in Figure 6.22A. That is, as foot size increases, the alignment of the peak is later, but the effect only extends up to three syllables, after which the timing does not change. This is interesting, as it matches the effect of `foot_syls` on the likelihood of L*H (see Section 6.5.4 above), wherein the estimated likelihood of L*H increases until it reaches saturation point in the three-syllable condition. Together, these results suggest that foot size effects plateau at three syllables. This is very clear when we compare the estimated differences between neighbouring foot-size conditions. The estimated difference between each condition dissipates from 40 ms between the one- and two-syllable conditions, to 27 ms between the two- and three-syllable condition, to a negligible 1 ms between the three- and four-syllable conditions, 95% CIs [22, 58], [1, 53], [-18, 20] and p < .001, p = .043, p = .899 respectively.

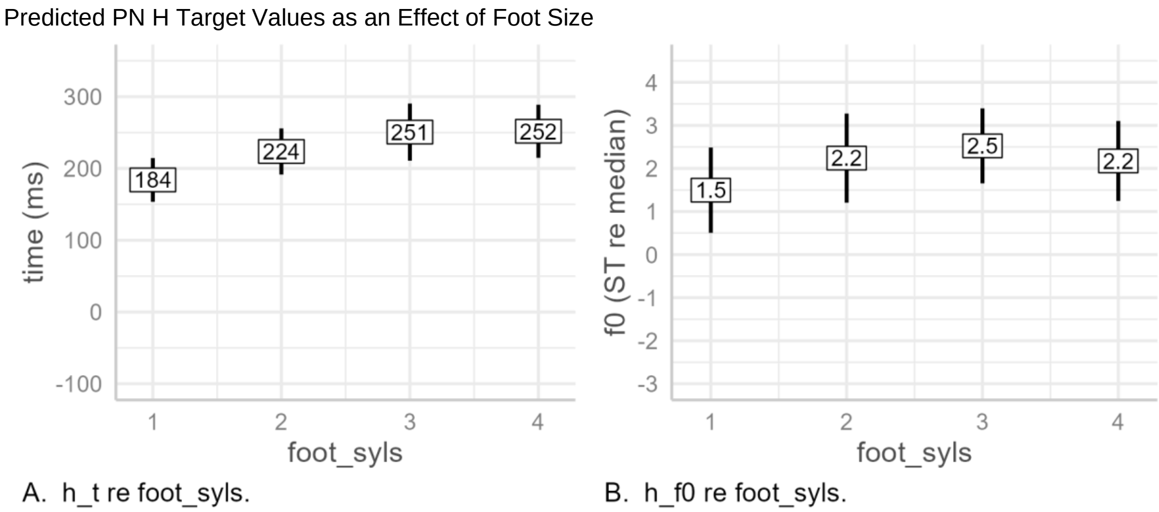


Figure 6.22 Predicted values of H targets based on foot-size effects alone .

The estimated effects of foot size on `h_f0` (Figure 6.22B) are quite similar to those on `h_t`. That is, as foot size increases, f_0 also increases. However, in this case, it seems that rather than stabilizing at the three-syllable condition, the effect seems to wear off, so to speak, in the four-syllable condition. That is, after rising to an estimated 2.5 ST, 95% CI [1.7, 2.4] in the three-syllable condition, it falls back down to 2.2 [1.2, 3.1] ST, roughly where it was in the two-syllable condition, i.e., 2.2 [1.2, 3.3] ST. However, these estimates are still all very similar, differing by at most 0.3 ST. The greatest estimated differences are between the one-syllable condition and the others. The two-syllable condition is an estimated 0.7 [0, 1.4] ST higher, the three-syllable conditions an estimated 1 [0.4, 1.6] ST higher, and the four-syllable condition an estimated 0.8 [-0.7, 1.6] ST higher, p = .039, .002, and .155 respectively. Despite the fact that the difference between the one- and four- syllable conditions is not statistically significant, there is still a general effect on f_0, whereby an increase in syllable count is associated with higher f_0 scaling.

The one-syllable condition is an instance of stress clash, and it appears that anticipation of the upcoming low of the next pitch accent (which is always L*H) leads to a damping of the f_0 peak. Given that both `h_t` and `h_f0` increase and then become stable at a point beyond the one-syllable condition, it seems that the PN pitch accent rise is truncated in the stress-clash condition. When the predicted means of the L target and H target parameters are plotted together for each level of `foot_syls`, we see that truncation does indeed appear to occur as a function of stress-clash in the one-syllable foot conditions, as shown in Figure 6.23 below. Truncation and compression are considered further in Section 6.6.3.4

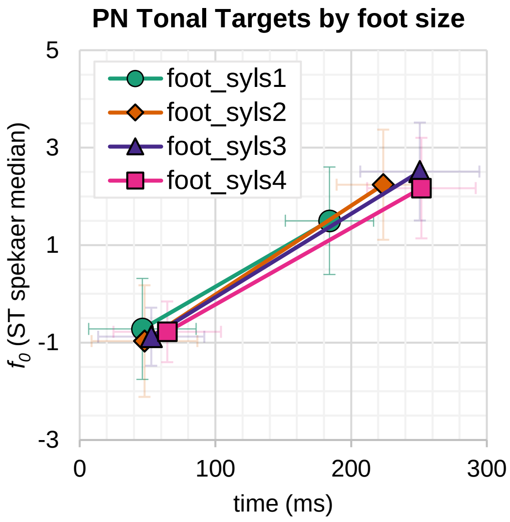


Figure 6.23 Predicted means and 95% CIs of L and H PN targets from LMEMs based on foot-size effects.

Word-boundary Effects. As can be seen in Figure 6.24, there is essentially no effect of the word-end syllable on either the alignment or scaling of the PN peak. When compared to the first-syllable word-end condition, there is an estimated mean difference of 6 ms and 16 ms in the alignment of the second- and third-syllable word-end conditions respectively, 95% CIs [-6, 19] and [-13, 46], p = .34 and .273 in turn. The difference in scaling is even less pronounced, at 0.1 [-0.4, 0.6], [-0.9, 1.2] ST in each case, p = .665, .787 for the second- and third-syllable word-end conditions respectively when compared with the one-syllable condition.

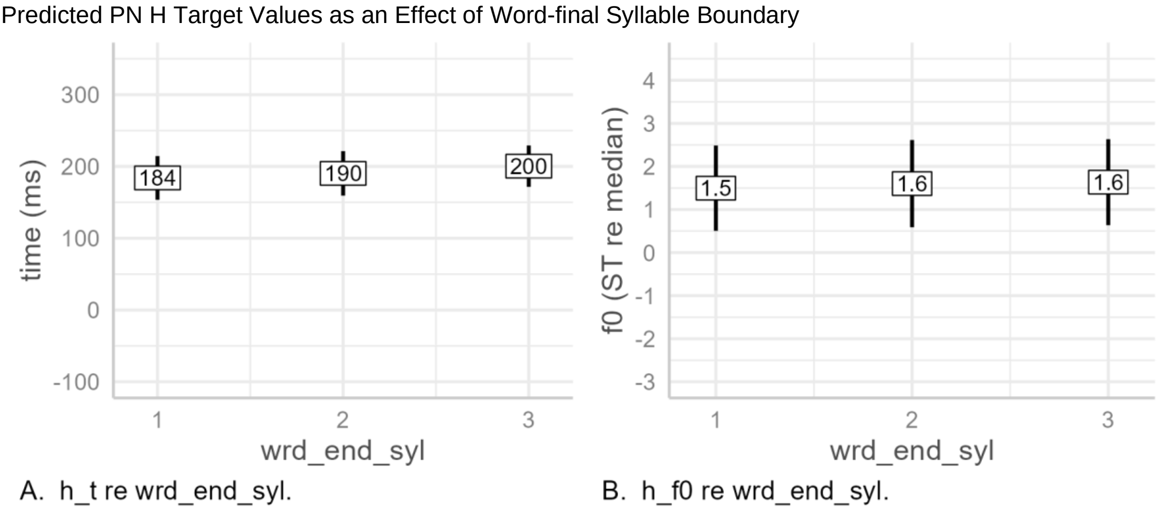


Figure 6.24 Predicted values of H targets based on word-end syllable effects alone .

Superficially, it appears that when the foot begins with a new word (Figure 6.25), the H target is aligned slightly earlier, and the peak is slightly lower, with estimated difference in the `true` condition of -22 ms and -0.8 ST in each case, 95% CIs [-80, 38], [-2.8, 1.1] and p = .403, .288 respectively. However, the CIs for each true condition are large, with each CI stretching well into a positive effect, indicating that the mean estimates are far from reliable (and thus the high p values).

Predicted PN H Target Values as an Effect of New Word at Foot Onset

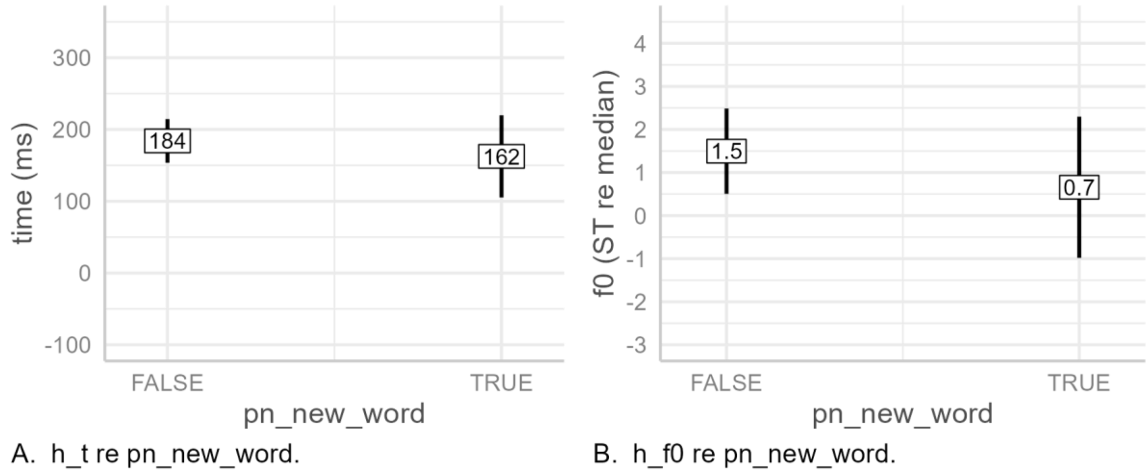


Figure 6.25 Predicted values of H targets based on word boundary effects at foot onset alone .

Gender Effects. As with the L targets, the effect of gender in H targets is quite striking (Figure 6.26), with the male speakers once again aligning the peak 20 ms earlier than the female speakers, 95% CI [-43, 3.4], p = .086, but in this case the CI is larger, and the difference is statistically non-significant. On average, male speakers realise the f_0 peak at 2.9 [1.9, 3.9] ST, which is an estimated 1.4 [0.9, 1.9] ST higher than the estimated female peak, p < .001. Peak scaling reflects the trend seen in the L f_0 target.

Predicted PN H Target Values as an Effect of Gender

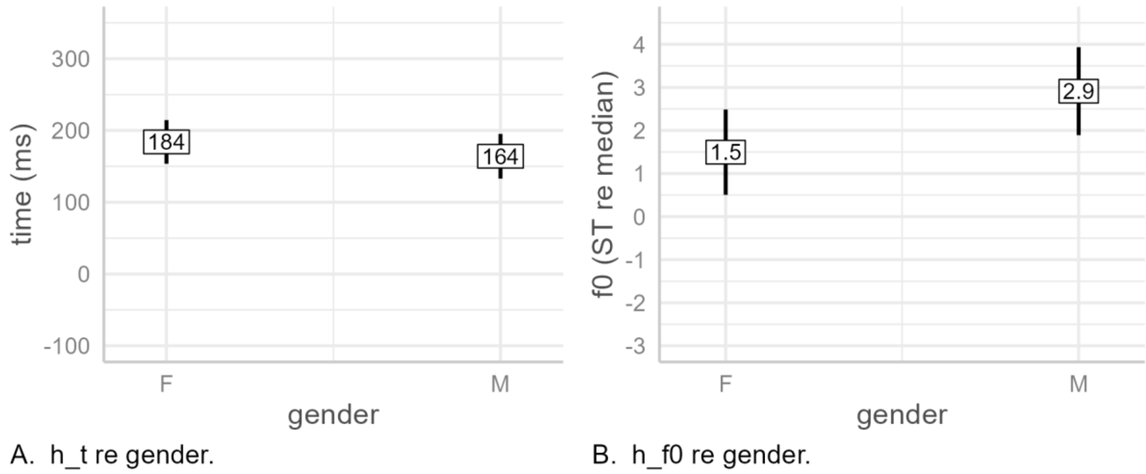


Figure 6.26 Predicted values of H targets based on `gender` effects alone .

6.6.2.4 PN Peak Alignment and Anacrusis Effects in L*H and H*. Given the strange results of the effects of anacrusis on `h_t`, it seemed worth further investigation. Based on the results above, we know that the peak alignment in >H* and H* is progressively earlier compared to L*H. However, when we look at the distribution of PN pitch accents by `ana_syls` in the PN `h_t` data, we see that there are only four examples each of >H* for `ana_syls2` and `ana_syls3`, with no examples of H* in either condition (see Table 6.22).

Table 6.22 Number of pitch accents per `ana_syls` condition in the data set for the `h_t` model (n = 737).

ana_syls	**>H***	**H***	**L*H**
0	21	60	211
1	20	139	175
2	4	0	52
3	4	0	51

`h_t` was retested using two subsets of the data, one containing only L*H and another only H* pitch accents.[15] Each new dataset was tested using the model in Equation 6.5.

```
h_t ~ ana_syls + (1 | speaker) + (1 | gender)
    + (1 | foot_syls) + (1 | pn_str_syl)
```
(6.5)

An ANOVA of the L*H-only model showed a significant effect of `ana_syls`, $F(3, 364) = 23.4$, $p.adj < .001$, but an ANOVA of the H*-only model did not, $F(1, 5.76) = 2.3$, $p.adj < .257$. The marginal R^2 of the L*H-only was .1, with a conditional R^2 of .82. In the H*-only model they were .13 and .61 respectively, so `ana_syls` accounts for a similar amount of variance in each model.

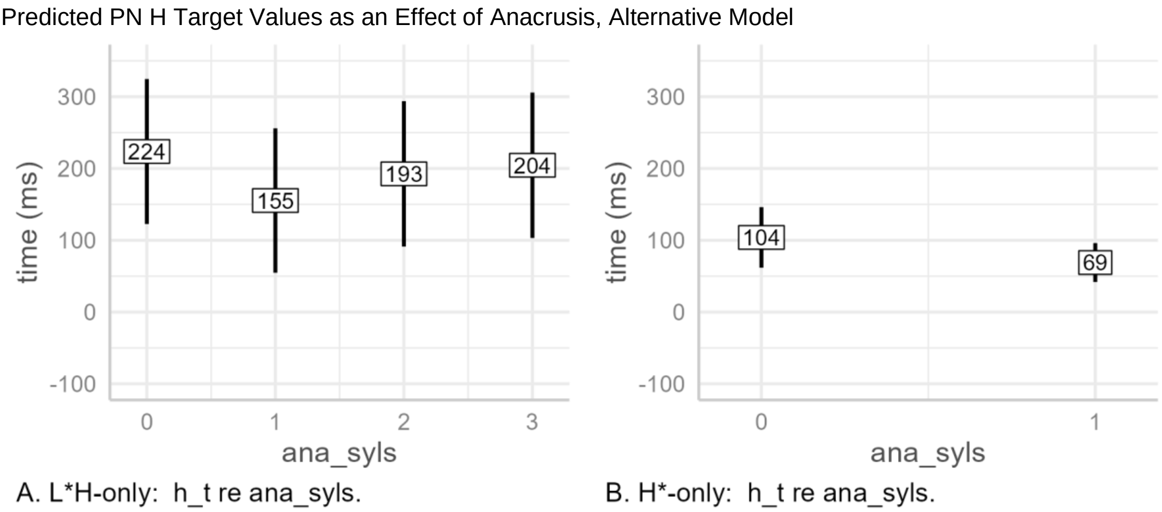


*Figure 6.27 `h_t` modelled re the effects of `ana_syls` from L*H-only and H*-only data sets.*

Figure 6.27 shows predicted peak alignment as an effect of `ana_syls` in the L*H-only and H*-only models, in Panels A and B respectively. We see that, as expected, predicted peak alignment in the H* model is much lower than in the L*H model.

The L*H model looks quite similar to the original model, with peak alignment falling in the one-syllable condition and then rising progressively in the two- and three-syllable conditions. However, in the zero-anacrusis condition, the peak is aligned noticeably later. That is, in the L*H-

[15] A model with an `ana_syls:acc_phon` interaction was tested but would not converge, so it was removed.

only model, the predicted alignment of h_t is 224 ms, 95% CI [123, 325] compared to 184 [154, 214] ms in the original model. Thus, when there is no anacrusis at all, peak alignment is noticeably late, but as more anacrusis is added, the peak slowly drifts rightward again. Moreover, unlike the original model, estimated h_t in the L*H model is always earlier than the zero-anacrusis condition once anacrusis is added to the beginning of the phrase. In fact, all three conditions with anacrusis are significantly earlier then the zero-anacrusis condition, by 68 [-85, -51], 31 [-48, -14], and 19 [-36, -2] ms in the one-, two-, and three-syllable conditions, $p < .001$, $p < .001$, $p = .028$ respectively. The estimated peak alignment in the one-syllable condition of the H*-only model is also earlier than the zero-anacrusis condition, by 35 [-92, 22] ms, $p = .183$, but this is not a statistically significant difference.

Overall, this reanalysis helps shed light on the effect of anacrusis on PN peaks. Essentially, we see that a lack of anacrusis causes a significant delay in peak alignment, while the addition of anacrusis allows it to be aligned earlier. However, with the addition of further syllables of anacrusis, the peak begins to drift rightwards again.

6.6.2.5 Stress-clash Effects on PN H Targets. Because of the apparent effect of stress clash on h_t and h_f0, two additional models were tested. Non-significant fixed effects were removed from the model, while significant effects which were not of interest were treated as random intercepts. A new factor, stress_clash, was added, which is true if there is only one syllable in the foot. The model was further reduced by running the step() function (see Section 5.4.1) and removing non-significant random effects. The h_t model, however, would not converge unless acc_phon was retained as a fixed effect. The final two models are shown in Equations 6.6 and 6.7.

```
h_t ~ stress_clash + acc_phon + (1 + stress_clash | speaker)
    + (1 | gender) + (1 | pn_str_syl)
```
(6.6)

```
h_f0 ~ stress_clash + (1 + stress_clash | speaker)
   +  (1 | gender) + (1 | pn_str_syl)
```
(6.7)

ANOVAs of the models indicate that stress_clash is statistically significant, $F(1, 20.57) = 28.86$, $p.adj < .001$ in the h_t model, and $F(1, 17.29) = 8.74$, $p.adj = .21$ in the h_f0 model. (acc_phon is also significant in the h_t model, $F(2, 527) = 145.26$, $p < .001$.) The marginal R^2 of the stress-clash h_t model is .26, with a conditional R^2 of .74, indicating that the overall variance explained by the fixed factors in the alternative model is much less than that of the original—which was .59—but the variance explained by the whole model is only slightly less than the original, which was .83. The marginal R^2 of stress-clash h_f0 model is .02, very low compared to the original (.23), while there is little difference in the conditional R^2 of the original and alternative models, at .44 and .46 respectively. The similarity of the conditional R^2 in each original and alternative model reflects the fact that the alternative models have simply shifted most of the fixed effects to random effects. Full reporting of the stress-clash h_t and h_f0 models can be found in Appendices G7 and G8.

The presence of stress_clash in h_t was associated with an earlier estimated alignment of 54 ms, 95% CI [-75, -33], $p.adj < .001$. The effect of stress_clash on h_f0 was also significant, as the presence of stress clash was associated with a lowering of peak f_0 by an estimated 0.88 [-1.5, -0.3] ST, *p* = .006. This suggests that the effect of foot size on the H target may best be interpreted by the absence or presence of stress clash rather than the number of syllables in the foot.

6.6.2.6 Truncation and Compression. An ANOVA of the `f0_exc` model indicates a significant effect of `foot_syls` on f_0 excursion, *F*(3, 306.15) = 4.95, *p.adj* = .006. However, an ANOVA of the `log_lh_slope` model indicates that there is no significant effect of `foot_syls` on the slope of the L*H rise, *F*(3, 184.88) = 1.19, *p.adj.* = .418. The marginal R^2 of the `f0_exc` is .03, while the conditional R^2 is .56, indicating that foot size only accounts for 3% of the variance f_0 excursion. For the `log_lh_slope` model, the marginal R^2 is .01 with a conditional R^2 of .46, indicating that foot size accounts for merely 1% of the variance in f_0 slope.

Estimated Effects Foot Size on Prenuclear L*H f_0 Excursion and Slope

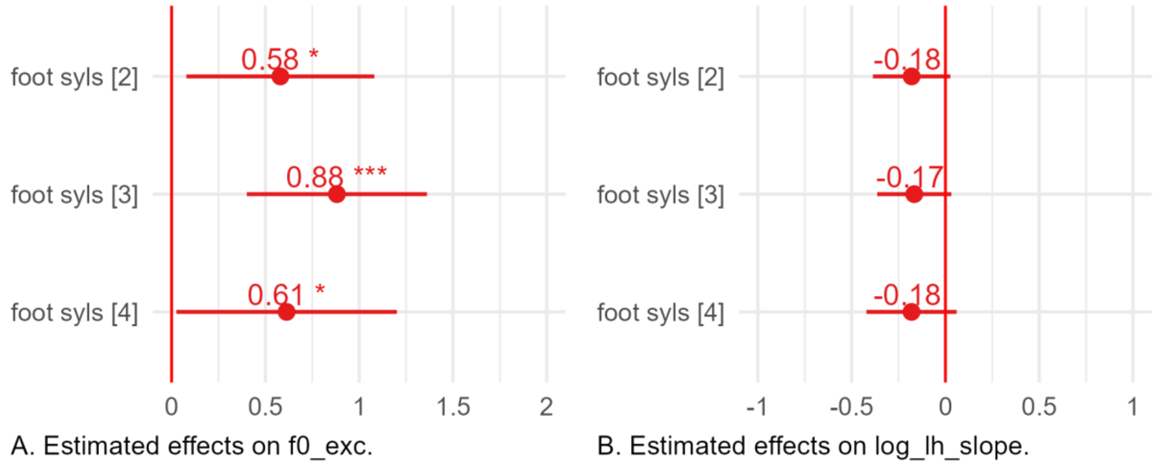


*Figure 6.28 Estimated effects of foot size on f_0 excursion and log slope of PN L*H from LMEMs. The vertical red line indicates the intercept,* `foot_syls[1]`.

*Table 6.23 Results of LMEM comparison of effects of two-, three-, and four-syllable feet on f_0 excursion in PN L*H pitch accents.*

intercept	Slope	est. (ST)	2.5% CI	97.5% CI	SE	|t|	df	p
foot_syls2	foot_syls3	0.3	-0.1	0.7	0.22	1.39	274.89	.166
foot_syls2	foot_syls4	0.0	-0.5	0.6	0.28	0.12	216.89	.904
foot_syls3	foot_syls4	-0.3	-0.6	0.1	0.18	1.48	403.29	.139

Turning to slope, it is clear—as shown in Figure 6.28B—that there is little effect on `log_lh_slope` across `foot_syls` conditions. The estimated differences between one-syllable condition (the intercept) and the two-, three-, and four- syllable conditions are -0.18, -0.17, and -0.18 log(ST/s) respectively, 95% CIs [-0.39, 0.27], [-0.36, 0.32], and [-0.42, 0.06], *p* = .088, .1, and .142 in turn. The mean estimates of the effects of the non-stress clash conditions are almost identical, with each slightly lower than the stress-clash condition. However, even this difference is quite small, and it is not statistically significant (rows one to three of Table 6.24).

*Table 6.24 Results of LMEM comparison of effects of foot-size effects on log L*H slope in PN pitch accents.*

intercept	slope	est. (ST)	2.5% CI	97.5% CI	SE	|t|	df	p
foot_syls1	foot_syls2	-0.18	-0.387	0.027	0.11	-1.72	191.38	0.088
foot_syls1	foot_syls3	-0.17	-0.365	0.032	0.10	-1.65	177.07	0.1
foot_syls1	foot_syls4	-0.18	-0.423	0.061	0.12	-1.48	130.36	0.142
foot_syls2	foot_syls3	0.07	-0.25	0.37	0.14	0.52	12.75	.609
foot_syls2	foot_syls4	0.05	-0.24	0.34	0.14	0.37	16.38	.719
foot_syls3	foot_syls4	-0.02	-0.25	0.21	0.11	-0.19	21.04	.848

From the point of view of perception, these differences in slope are also most likely indistinguishable. For example, the seminal study of pitch perception by ’t Hart et al. found that the differential threshold of pitch change was best measured as g_1/g_2, where g refers to the glide or slope measured in ST/s. For listeners to distinguish between two glides, there needed to be a difference of at least a factor of two between glides (’t Hart et al., 1990, pp. 33–35). In the current analysis, when the mean estimated slopes are converted back into their non-logged counterparts (Figure 6.29) and compared, we see that there is likely almost no perceptual difference between each glide. As indicated by Table 6.25, the differential is only 1.2 even between the one-syllable condition and the

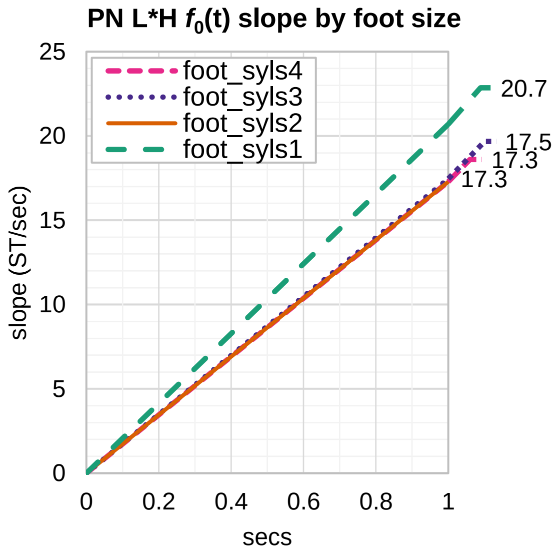


Figure 6.29 Visual representation of slope in ST/s as an effect of foot-size in PN pitch accents.

others, greatly below the threshold for differentiation between slopes. (It should be noted that in the original experiments outlined by ’t Hart et al., this threshold was established based on perception experiments of non-speech pitch glides. It should also be noted that the slopes mentioned here are not actual slopes, but rather estimates abstracted from a model. Therefore, the analysis using the differential threshold should be seen as suggesting rather than confirming the extent of similarity and difference between slopes.)

Table 6.25 Differentials of pitch change in foot_syl conditions, i.e., ratio between slopes measured in ST/s.

	foot_syls1	**foot_syls2**	**foot_syls3**
foot_syls2	1.2		
foot_syls3	1.2	1.0	
foot_syls4	1.2	1.0	1.0

The phonetic analysis of prenuclear pitch accents indicates effects of stress clash on both f_0 excursion and slope, but the effects on slope are relatively small. However, for current purposes, we note that slope barely changes as a function of foot size even though the size of the excursion does change as a function of foot size. It was also noted that there is a strong effect of foot size on the alignment of the H target, with longer feet being associated with later alignment up to the three-syllable condition. The earlier inference that truncation effects on the rise to the H target are in operation in the PN condition is reinforced by the fact that, while both the timing and f_0 of the peak do change, the slope of the L*H rise barely changes at all. The overall effect is that stress-clash between the PN stressed syllable and the nuclear pitch accent's stressed syllable causes truncation of the rise in PN L*H. This strongly implies that speakers effectively sacrifice the full rise of the PN contour to ensure that they can realise the L of the nuclear L*H pitch accent more fully.

Together, this suggests a hierarchy of importance. Firstly, as just noted, in the stress clash condition, the prenuclear PA rise is less important than the nuclear PA, and so the end of the rise is truncated. Secondly, within the prenuclear pitch accent itself, the L target is less important than the H target and is therefore prone to deletion, and this is what leads to the increased observations of H* under more constrained lexical and metrical conditions. The phonological deletion strategy was preferred here over the pre-emptive truncation of a "virtual [L] target" suggested by Arvaniti and Ladd re Greek wh-questions (2009)[16]—discussed previously in Section 6.5.5—because L-target alignment is otherwise relatively stable in PN L*H pitch accents under varying lexical and metrical conditions (Section 6.6.2.1, p. 90), and thus shows no hint of truncation. Finally, the truncation and deletion strategies used in the PN pitch accent reinforce the sense that the phonological identity of

[16] The critique of this point feels somewhat overstated in Xu et al. (2015). The original paper (Arvaniti & Ladd, 2009) suggests that there could be a virtual target before the start of the phrase, but that the interpolation from the L to the H is not manifested in the contour itself. However, Xu et al. insists that an AM analysis *must* include interpolation from a fully realised L target at the left edge of the utterance. Otherwise, the explanation that Xu et al. propose is not terribly different from that of Arvaniti and Ladd, which shows that tones appear to originate from an earlier (virtual) common f_0 midpoint which not manifested in the contour. The 2015 paper highlights the fact that the PENTA model can cope with the utterance-initial phenomenon described in Ladd and Arvaniti (2009), but this alone does not invalidate the AM explanation proffered in the 2009 paper. That said, the current study does favour a phonological tone deletion strategy over a truncation strategy. This too does not discredit the possibility of truncation of a rise from a virtual target.

the prenuclear pitch accent is less important than that of the nuclear pitch accent, which is never realised as H* in this dataset.

6.6.3 Mixed-effects Modelling of Nuclear Pitch Accents and Contours

The six phonetic response parameters assessed for the prenuclear pitch accent were also tested for the nuclear pitch accent, i.e., L target alignment, L target scaling, H target alignment, H target scaling, f_0 excursion, and slope (see Table 6.16). In addition, temporal alignment and scaling of the final boundary were evaluated (using parameter codes `e_t` and `e_f0`). The timing of the final boundary is not of interest, but it was included so that stylised f_0 contours could be correctly plotted (see Figure 6.43, p. 118 below).

Several changes were made to the fixed effects of the nuclear models. `acc_phon` was removed since nuclear pitch accents were exclusively L*H. For similar reasons, `wrd_end_syl` was not used since there is only one word in the final foot of each target utterance. `ana_syls` was replaced with the number of unstressed syllables in the preceding foot (`pre_syls`). The effect of the final boundary tone (`fin_phon`) was also included in each model. The fixed parameters for nuclear PA and boundary tone analysis are summarised in Table 6.26.

Table 6.26 Fixed effects for nuclear pitch accent analysis.

Parameter type	**Parameter code**	**Comments / levels**
Foot size	`foot_syls`	One to four syllables
Preceding syllables	`pre_syls`	Zero to three syllables
Final boundary	`fin_phon`	Phonological tone of the final boundary, % or L%
Word break at foot onset	`nuc_new_word`	True/False
Gender	`gender`	Female/Male

As with PN analysis, all fixed effects were included in the models for the tonal target response parameters. For the f_0 excursion and slope models, `foot_syls` and `fin_phon` were the only fixed-effect parameters, while the other parameters were treated as random intercepts. Random effects also were reduced until a maximal model without convergence or singularity issues was found. Because no effect of foot size on the phonological choice of the final boundary was found in the phonological analysis (See Section 6.5.1), no interaction between `foot_syls` and `fin_phon` was included in the model as it could be assumed that these did not interact.

Table 6.27 Random effects for nuclear PA and boundary tone LMEMs in analysis of A- and H-Corpora.

Response	Random effects in model
l_t	(1 | speaker) + (1 | nuc_str_syl)
h_t	
e_t	
l_f0	
h_f0	
e_f0	
f0_exc	(1 + foot_syls | speaker) + (1 | nuc_str_syl) + (1 | pre_syls)
log_lh_slope	

As with the previous analysis, the analysis of tonal targets is presented first (L, then H, and then the boundary tone), followed by the analysis of the f_0 excursion and slope. Again, in some cases, additional models were tested to refine the analyses. These are presented as separate subsections after the initial analysis of the associated target parameter. Full results are listed in Appendix H

6.6.3.1 L Targets. An ANOVA of the nuclear `l_t` model indicates two significant effects, namely `pre_syls`, $F(3, 758) = 4.03$, $p.adj = .014$, and `gender`, $F(1, 8.96) = 58.7$, $p.adj < .001$. The model has a marginal R^2 of .37 and conditional R^2 of .71, suggesting that the fixed effects account for more than a third of the variance in the response parameter. An ANOVA of nuclear `l_f0`, however, indicates that only the effect of `foot_syls` is significant, $F(3, 543) = 8.07$, $p.adj = <$ .001. The `l_f0` model has a marginal R^2 of .05 with a conditional R^2 of .36. This suggests that neither the fixed effects of the `l_f0` model nor the whole model itself capture a large amount of the variance in f_0 scaling. This may be due to the fact that, with the exception of `foot_syls` effects, there is little variance in the L target f_0 at all. (Full details in appendices H2 and H3).

Effects of Preceding Syllables. Beginning with `pre_syls` effects, we see that there is a gradual decrease in the predicted temporal alignment of `l_t` as the number of preceding syllables increases (Figure 6.30A), from 94 ms in the zero-syllable condition down to 80 ms in the three-syllable condition, 95% CIs [63, 125] and [50, 111] respectively. When we consider these differences in terms of pairwise comparison of effects between `pre_syls` conditions, we see that both the one-syllable and three-syllable conditions are statistically significantly different from the zero-syllable condition, at 8 [-15, -1] and 14 [-25, -2] ms earlier in turn, $p = .026$ and .019 respectively. While these are not particularly large shifts in alignment, they do suggest that there is a residual effect of stress-clash, in that the `l_t` is aligned slightly later in the zero-condition. The effect, however, is not as large as the effect of stress clash on `h_t` in the PN pitch accent, where `h_t` is aligned an estimated 30 ms earlier in the one-syllable condition than in the two-syllable foot condition (see Section 6.6.2, p. 88). Thus, the stress-class effect on nuclear `l_t` mirrors the findings in the previous section, i.e., while anticipation of a nuclear L target is associated with the earlier alignment of the PN H target in stress clash conditions, the nuclear L target is aligned slightly later. It also reinforces the view that nuclear pitch accent targets are less subject to metrical effect than prenuclear PA targets, given that the effect is noticeably smaller in the nuclear pitch accent.

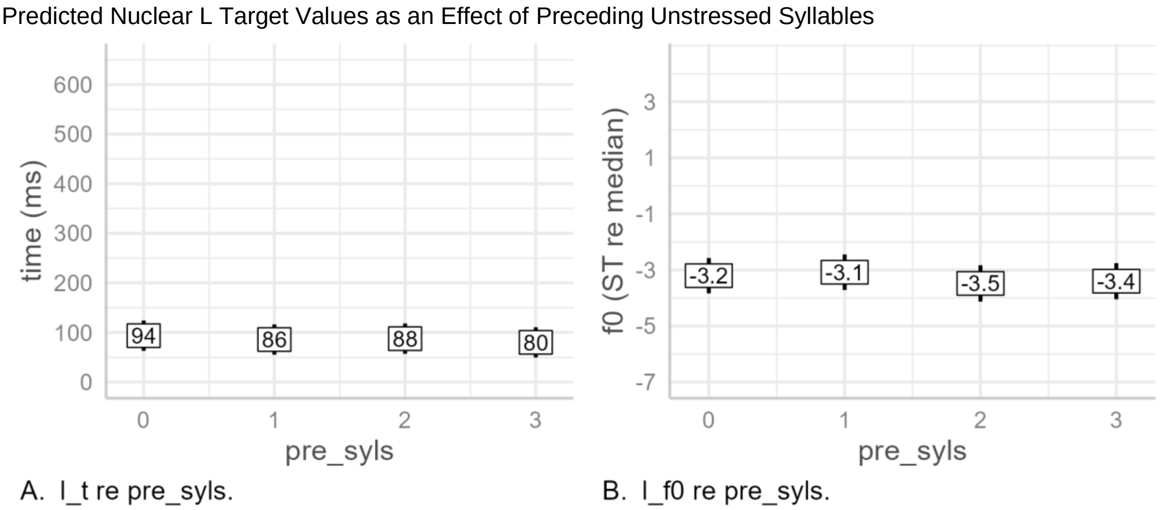


Figure 6.30 Predicted values of nuclear L targets based on `pre_syls` *effects alone .*

The ANOVA indicates that we cannot reject the null hypothesis that `pre_syls` has no effect on `l_f0`. There does appear to be very slight division in `l_f0` between the zero- and one-syllable conditions on the one hand, to the two- and three-syllable conditions on the other (Figure 6.30b), but this is essentially a visual effect based on the order of presentation of levels in the plot. If we sort the estimated means by f_0 value, we see that there is little difference between the two groups. For example, the zero-syllable condition is only slightly lower than the two-syllable condition, (0.3 ST, 95% CI [-0.7, 0.1], p = .167). The largest estimated differences are between `pre_syls1` and `pre_syls2` at 0.4 [0.7, 0.1] ST and between `pre_syls1` and `pre_syls3` at 0.3 [-0.6, 0] ST, p = .009 and .045 respectively. While these are statistically significant, and even though such differences may be salient when played as discrete tones, they are still small and may not be salient in continuous speech.

Effects of Foot Size and Foot-initial Word Boundary. There is almost no effect of either `foot_syls` or `nuc_new_word` on `l_t`, as shown in Panel A of Figure 6.31 and Figure 6.32 respectively. There is barely any deviation from the intercept of 94 ms, 95% CI [61, 124], across different levels of either factor. This is also true of the effect of `nuc_new_word` alone on `l_f0` (Figure 6.32b). In this case, a new word at the onset of the nuclear foot is only associated with a 0.2 [0, 0.4] ST rise in the L target, p = .52.

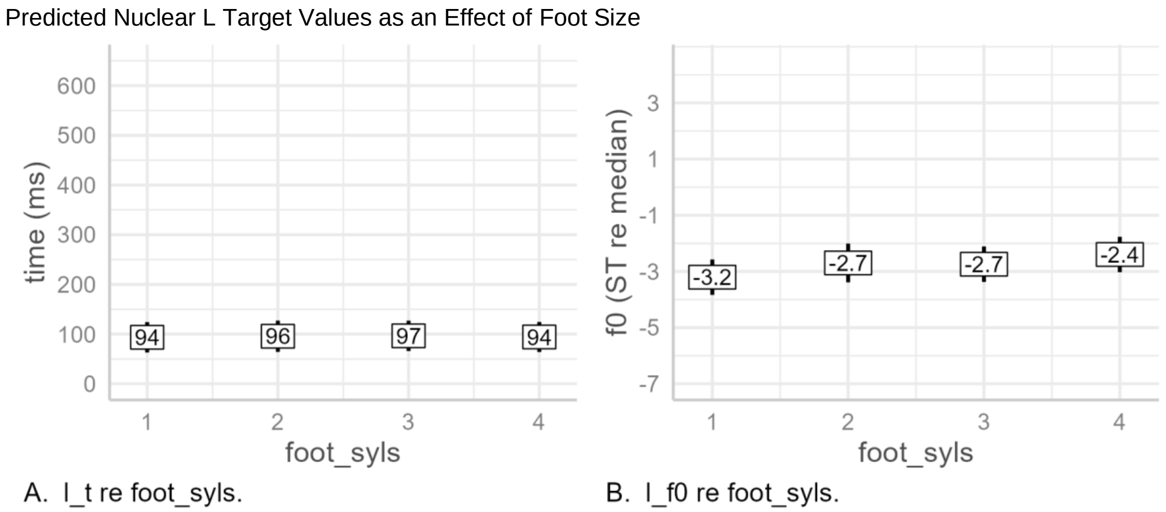


Figure 6.31 Predicted values of nuclear L targets based on `foot_syls` *effects alone* .

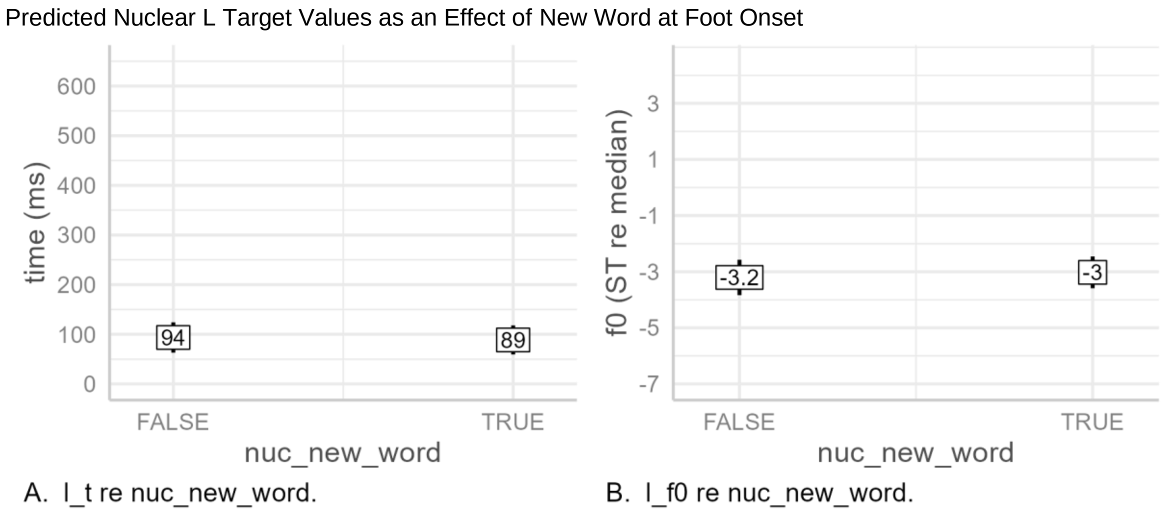


Figure 6.32 Predicted values of nuclear L targets based on `nuc_new_word` *effects alone* .

There is, however, a clear effect of foot size on `l_f0`, which increases steadily with the addition of more syllables to the foot. In fact, the main effect appears to be related to the absence or presence of trailing syllables in the foot. That is, the two-, three-, and four- syllable feet conditions are at least 0.5 ST higher than the one-syllable condition, at 0.5 ST, 0.5 ST, and 0.8 ST, CIs [0.2, 0.8], [0.2, 0.7], and [0.4, 1.2], $p = .002$, $p < .001$, $p < .001$ respectively (see also Table 6.28 rows one to three). However, the two-, three-, and four-syllable feet conditions do not differ from each other by more than 0.3 ST, and there is no significant difference in estimated `l_f0` between any of the three, as indicated by the last three rows in Table 6.28

Table 6.28 Table H2.8. Pairwise comparison (b_1) of effects of levels of `foot_syls` on nuclear `l_f0`.

Intercept	slope	estimate (ST)	2.5% CI	95.5 % CI	SE	|t|	df	p
foot_syls1	foot_syls2	0.5	0.2	0.8	0.16	3.08	749.68	.002
foot_syls1	foot_syls3	0.5	0.2	0.7	0.13	3.58	749.7	< .001
foot_syls1	foot_syls4	0.8	0.4	1.2	0.2	4.04	279.87	< .001
foot_syls2	foot_syls3	0.0	-0.4	0.3	0.17	-0.26	749.95	.797
foot_syls2	foot_syls4	0.3	-0.1	0.7	0.21	1.46	303.32	.144
foot_syls3	foot_syls4	0.3	-0.1	0.8	0.21	1.67	305.73	.095

Boundary Tone Effects. Again, as indicated by the ANOVAs for the `l_t` and `l_f0` models, there is barely any effect of `fin_phon` on either `l_t` or `l_f0` (Figure 6.33), with only a 4 ms and -0.1 ST difference between L% and %, 95% CIs [-7, 17], [-0.5, 0.31], and p = .682, .492 respectively. It is unsurprising that the boundary tone has little effect on L target parameters, given that the two are separated by the H target.

Predicted Nuclear L Target Values as an Effect of Boundary Tone

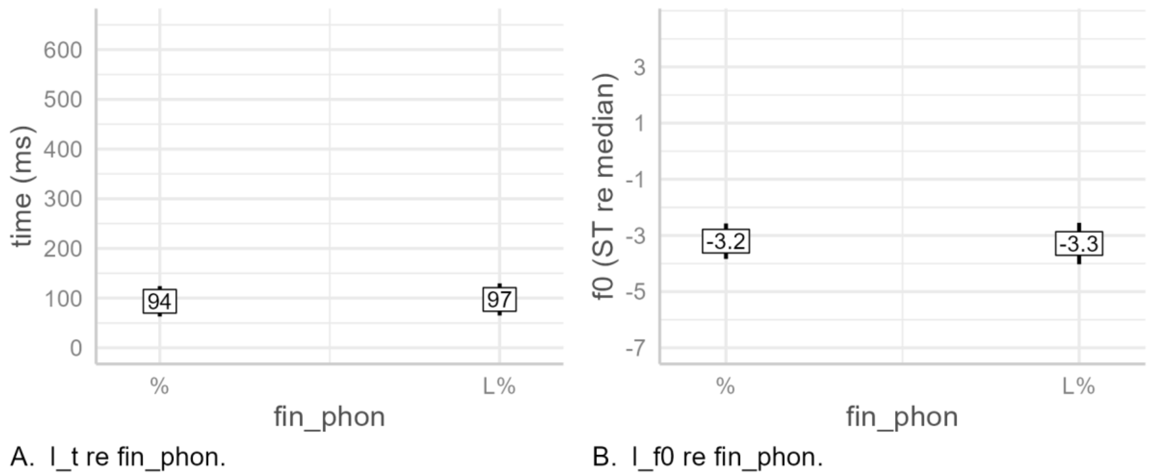


Figure 6.33 Predicted values of nuclear L targets based on `fin_phon` effects alone .

Gender Effects. Finally, as indicated by the ANOVAs, there is a significant effect of `gender` on the alignment of `l_t` (Figure 6.34), with male speakers realising the target an estimated 54 ms earlier than the female speakers, 95% CI [-74, -12], p < .001. The effect of `gender` on nuclear `l_t` reflects a similar trend in prenuclear alignment, and this pattern will be seen again in `h_t` below. While the `l_f0` model indicates that the male speakers realise the L target an estimated 0.3 [-1.1, 0.7] ST lower than female speakers, this is a small difference and is not statistically significant, p = .52.

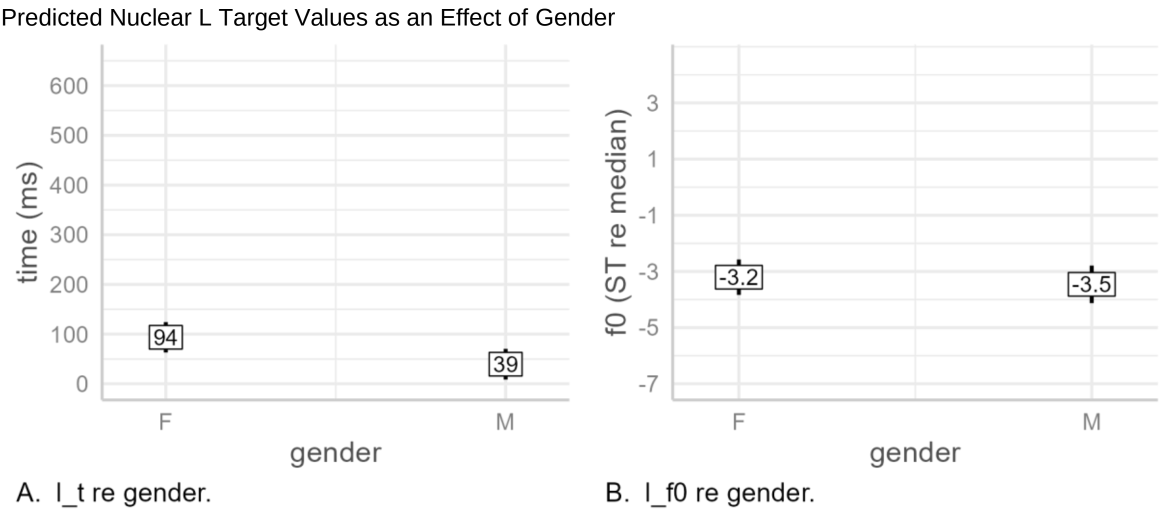


Figure 6.34 Predicted values of nuclear L targets based on `gender` effects alone .

6.6.3.2 H Targets. An ANOVA of the nuclear `h_t` model indicates that all fixed effects are significant, as shown in Table 6.29. The model has a marginal R^2 of .71 and conditional R^2 of .9, indicating that the fixed effects account for a large proportion of the variance in the response parameter while the whole model accounts for 90% of the variance. An ANOVA of nuclear `h_f0`, however, indicates that only the effect of `foot_syls` and `pre_syls` are significant, $F(3, 767) = 41.23$, *p.adj* = < .001, and $F(3, 607) = 5.98$, *p.adj* = .001 respectively. The `h_f0` model has a marginal R^2 of .11 with a conditional R^2 of .44. This indicates that the fixed effects of the `h_f0` model accounts for a noticeably larger amount of variance in the response parameter than its `l_f0` counterpart ($R_m^2 = .05$) while the complete model accounts for a slightly larger amount of variance than the `l_f0` model ($R_c^2 = .36$). Overall, the differences in R^2 between the L target and H target parameters indicate that the H target is more susceptible to changes in the metrical and lexical structure of the IP. (Complete tables and charts for the models can be found in Appendices H4 and H5.)

Table 6.29 ANOVA of nuclear `h_t` model.

term	sumsq	meansq	NumDF	DenDF	*F* value	p	*p.adj* (BH)
`foot_syls`	1.2×10^6	4.0×10^5	3	765.88	586.86	<.001	<.001
`pre_syls`	56659	18886	3	763.69	27.43	<.001	<.001
`fin_phon`	41246	41246	1	768.38	59.9	<.001	<.001
`nuc_new_word`	8036	8036	1	766.46	11.67	<.001	.001
`gender`	12918	12918	1	9.01	18.76	.002	.004

Effects of Preceding Unstressed Syllables. There is a gradual, slight, decrease in peak alignment as the number of preceding syllables increases (Figure 6.35a). However, the estimated alignment of the stress-clash condition (`pre_syls0`) is noticeably higher than the non-stress-clash conditions, at 292 ms, 95% CI [255, 230]. There is at least a 34 [-41, -26] ms estimated difference between the stress-clash condition and the other conditions, the smallest difference being between `pre_syls0` and `pre_syls1`, $p < .001$. However, there is only a slight difference between each non-stress-clash condition (`pre_syls2-4`). The largest estimated difference found here is between `pre_syls1` and `pre_syls2`, at 11 [-21, -1] ms, $p = .024$, which is still only small (albeit statistically significant) difference.

Predicted Nuclear H Target Values as an Effect of Preceding Unstressed Syllables

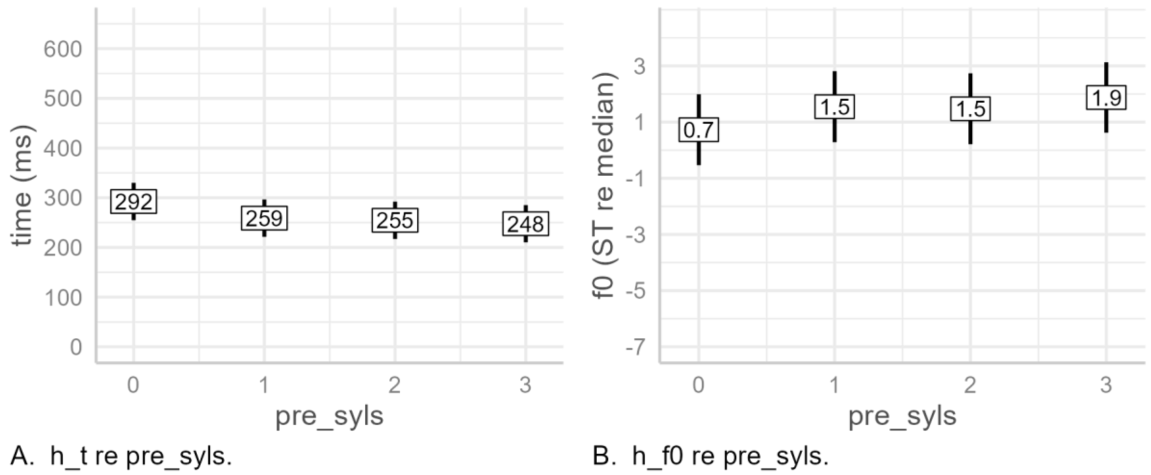


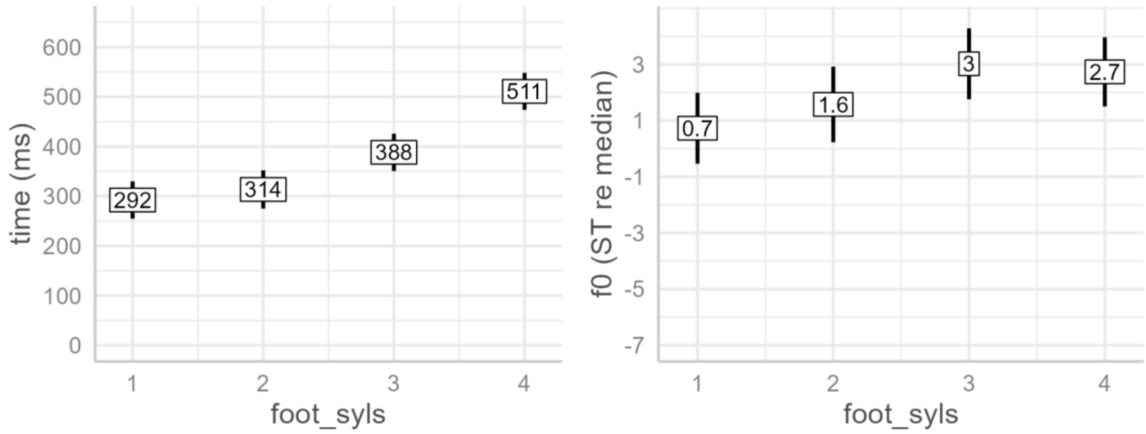
Figure 6.35 Predicted values of nuclear H targets based on `pre_syls` *effects alone .*

A similar effect of stress clash is seen in the predicted values of `h_f0` as an effect of `pre_syls`, where the predicted scaling of the H target ranges between 1.5 ST in both the one- and two-syllable conditions and 1.9 ST in the three-syllable conditions, 95% CIs [0.28, 2.83], [0.21, 2.74], [0.62, 3.13]. Each of the non-stress-clash conditions is associated with noticeably higher f_0 scaling than the stress-clash condition, an estimated 0.8 ST higher in both `pre_syls1` [0.4, 1.2] and `pre_syls2` [0.1, 1.4], and 1.1 [0.4, 1.8] ST higher in `pre_syls3`, p < .001, $p = .31$, and $p = .001$ respectively. The difference in f_0 peak scaling is much smaller between non-stress-clash conditions, ranging from 0.1 [0, 0.8] ST between `pre_syls1` and `pre_syls2` and 0.4 [-0.6, 0.4] ST between `pre_syls2` and `pre_syls3`, $p = .53$ and .781 respectively.

It is interesting that both the timing and height of the H target are more vulnerable to the effects of `pre_syls` than the L Target, even though the L target is closer to the preceding foot. As indicated by the ANOVAs, this suggests H targets are more susceptible to metrical and lexical effects.

Effects of Foot Size. The effect of `foot_syls` on the H target is very strong, especially in comparison to the more muted effect on the L target. As the number of syllables in the foot increases, there is a corresponding increase in the alignment and timing of the peak (Figure 6.36). The largest change in `h_t` occurs between the one-syllable and four-syllable conditions, with peak alignment an estimated 219 ms later, 95% CIs [207, 230], p < .001. For `h_f0`, it is between the one-syllable and three-syllable conditions, with the latter being an estimated 2.3 [1.9, 2.7] ST higher than the former, p < .001. However, the pattern of changes between foot size conditions is not the same for alignment and f_0 height.

Predicted Nuclear H Target Values as an Effect of Foot Size



A. h_t re foot_syls.   B. h_f0 re foot_syls.

Figure 6.36 Predicted values of nuclear H targets based on `foot_syls` *effects alone .*

In `h_t`, there is only a small increase in peak alignment between the one- and two- syllable conditions of 21 ms, but a much larger increase of 75 ms from the two- to the three-syllable condition, and an even larger increase again of 123 ms from the three- to the four-syllable condition, 95% CIs [12, 2], [65, 85], and [110, 135] respectively, p < .001 in each case. This suggests that as foot size increases, the peak is aligned increasingly later. For example, the change between the one- and two-syllable foot conditions is very small, at 21 [12, 31] ms, but the difference between the three- and four-syllable foot conditions is much greater, at 123 [185, 210] ms, p < .001 in each case. This is similar to the findings for nuclear peak alignment in Donegal English (Kalaldeh et al., 2009), but the exponential increase is more extreme in the current case.

In contrast, it appears that there is a more linear correspondence between `foot_syls` and the scaling of the f_0 peak, i.e., `h_f0`. However, the effect of `foot_syls` on f_0 peak scaling appears to reach saturation point by the three-syllable condition, after which f_0 drops slightly. Still, there is only an estimated fall of 0.3 ST between the three- and four-syllable conditions, 95% CIs [-1, 0.4], p = .422, which is both small and statistically non-significant.

The fact that the alignment of the peak continues shifting rightward as foot size increases but f_0 peak stops rising after three syllables suggests that there is an element of truncation in the one- and two- syllable conditions but an element of expansion between the three- and four- syllable conditions. This issue is considered in more detail in Section 6.6.3.4 below.

Foot-initial Word-boundary Effects. A word boundary at the syllable onset appears to have a minor effect on the timing and height of the H target (Figure 6.37). When there is a word boundary, the peak is aligned an estimated 13 ms earlier, 95% CI [-20, -5], $p < .001$. The presence of a foot-initial word boundary is associated with an estimated increase in f_0 peak of 0.4 [0, 0.9] ST, $p = 0.036$. Each effect is statistically significant, but the difference is small, especially compared to the effect of `foot_syls` or even `pre_syls`.

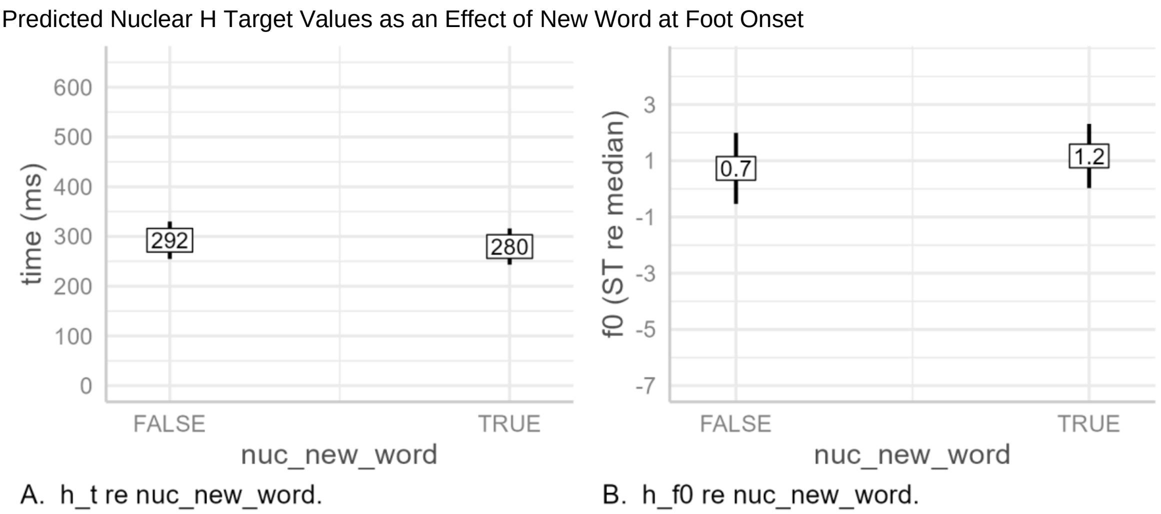


Figure 6.37 Predicted values of nuclear H targets based on `nuc_new_word` *effects alone .*

Boundary Tone Effects. As can be seen in Figure 6.38, the presence of an L% is associated with the earlier alignment and lower f_0 scaling of the peak, i.e., the H target is aligned an estimated 46 ms earlier as an effect of L% and is realised an estimated 0.6 ST lower, CIs [-57, -34] and [-1.3, 0], $p < .001$ and $p = .058$. However, because of the relatively large CIs in `h_f0`, the effect of L% is not statistically significant.

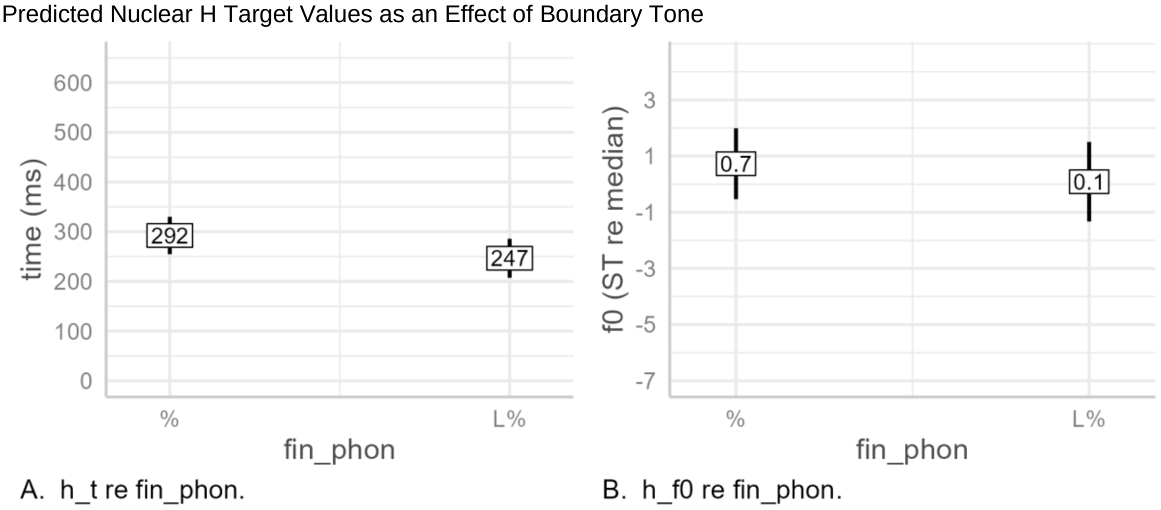


Figure 6.38 Predicted values of nuclear H targets based on `fin_phon` *effects alone .*

Gender Effects. Finally, we see once again that there is a noticeable effect of `gender` on the alignment of the tonal target (Figure 6.39), with male speakers realising the H target an estimated 64 ms earlier than female speakers, 95% CI [-98, -31], p = .002. However, there is essentially no effect of `gender` on the f_0 height of the peak, with males realising it only an estimated average of 0.1 [-1.2, 1.4] ST higher in their register than female speakers, p = .872.

Predicted Nuclear H Target Values as an Effect of Gender

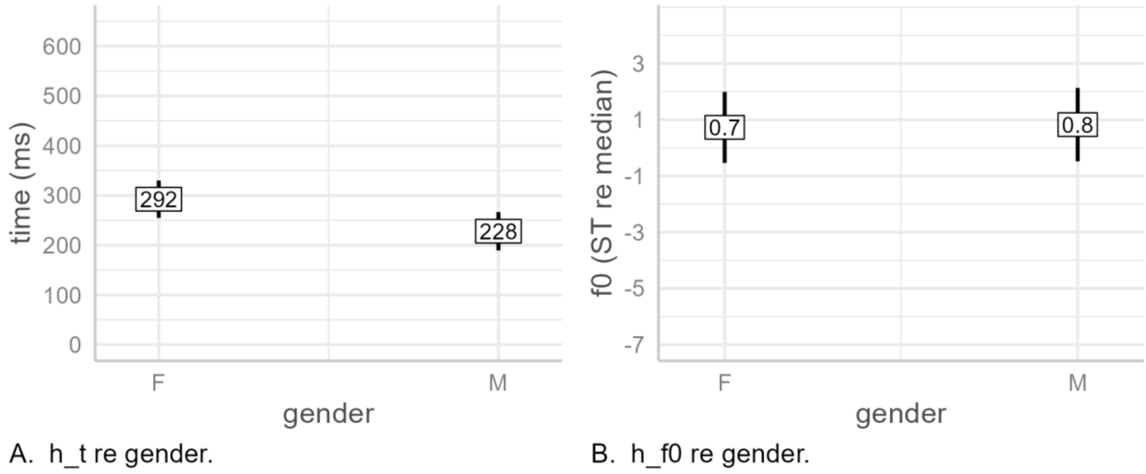


Figure 6.39 Predicted values of nuclear H targets based on `gender` *effects alone .*

6.6.3.3 Final Boundary. The timing of the boundary always refers to last frame of voiced speech in the IP. Therefore, it is really a measurement of the duration of voicing in the foot re the onset of the vowel in the stressed syllable. This is not of interest here, so it is not discussed further.

An ANOVA of nuclear `e_f0`, indicates significant effects of `foot_syls` and `fin_phon`, $F(3, 754) = 19.6$, $p.adj < .001$, and $F(1, 768) = 228$, $p.adj < .001$ respectively. The model has a marginal R^2 of .22 and a conditional R^2 of .53.

Looking at the parameters most closely associated with the beginning of the foot, i.e., `pre_syls` and `nuc_new_word` (Figure 6.40, Panels B and D), we see that there is only a very slight effect of either, with the predicted f_0 values barely veering from the intercept (i.e., 0.2 ST, 95% CI [-1.2, 1.6]). The only exception is `pre_syls1`, which is an estimated 0.5 [0, 1] ST higher than the intercept, p = .038. While this may be a residual effect of stress clash, it is still clear that the overall effect of `pre_syls` is quite small.

There seems to be a slight effect of `gender` (Panel E), with male speakers realising the final boundary an average estimated 0.7 ST lower than female speakers, 95% CI [-2.6, 1.2], p = .408. However, given the large CIs, it is not possible to make any claims about the significance of this difference, as reflected in the p value.

Looking at the height of the boundary as an effect of foot size (Panel A), it is clear that the essential difference is between a single-syllable (monosyllabic) foot and any polysyllabic foot. That is, in the one-syllable (monosyllabic) condition, the mean estimated f_0 is between 1.4 and 1.7 ST lower than the polysyllabic conditions. Each of these differences is statistically significant, as shown in rows one to three of Table 6.30. In contrast, there is only a small (non-significant) effect of foot_syls between each polysyllabic foot condition, as shown in rows four to six of the same table.

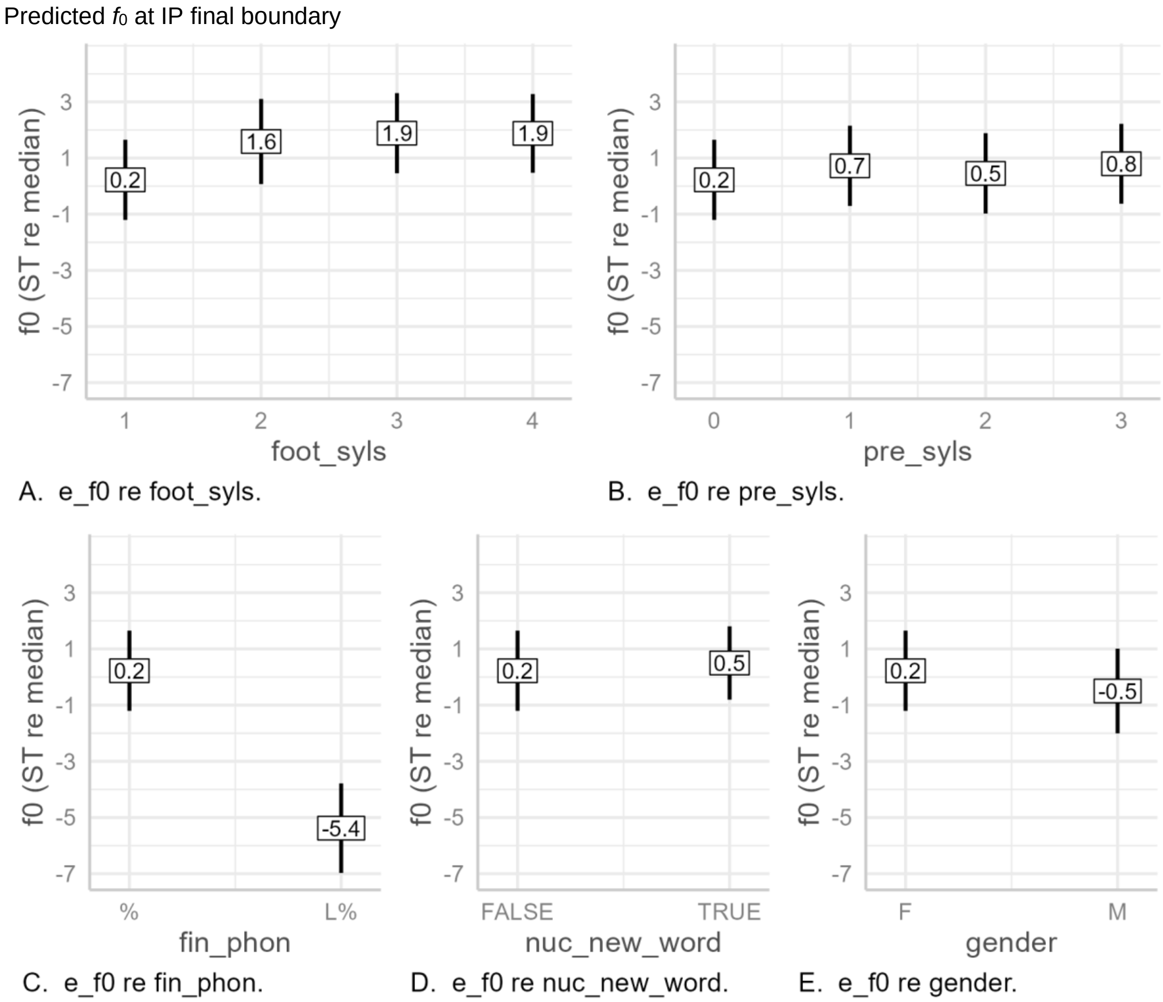


Figure 6.40 Predicted f0 values of final boundary in e_f0 *model.*

Table 6.30 Pairwise comparison of levels of foot_syls *in the* e_f0 *model.*

intercept	slope	est. (ST)	2.5% CI	95.5% CI	SE	|*t*|	df	p
foot_syls1	foot_syls2	1.4	0.8	2	0.3	4.53	764	<.001
foot_syls1	foot_syls3	1.7	1.2	2.1	0.24	6.97	764	<.001
foot_syls1	foot_syls4	1.6	0.9	2.4	0.38	4.37	705	<.001
foot_syls2	foot_syls3	0.3	-0.3	0.9	0.32	0.93	764	.355
foot_syls2	foot_syls4	0.3	-0.5	1.1	0.39	0.74	713	.459
foot_syls3	foot_syls4	0	-0.8	0.8	0.39	0.02	715	.985

Finally, we see that the presence of an L% boundary has a very strong effect on f_0 scaling of the final boundary (Figure 6.40E). The L% is associated with an estimated lowering of f_0 by 5.6 ST, 95% CI [-6.3, -4.9], $p < .001$. This effect is unsurprising, given that the L% should be associated with a low f_0. (However, it does serve to reinforce the phonological boundary tone judgments.)

6.6.3.4 Truncation and Compression. An ANOVA of the f_0 excursion model (`f0_exc`) indicates significant effects of `foot_syls` and `fin_phon`, *F*(3, 13.72) = 26.13, *p.adj* < .001, and *F*(1, 623) = 9.75, *p* = .002 respectively. This model has a marginal R^2 of only .08 and a conditional R^2 of .62. The low R_m^2 is to be expected given that there are only two fixed effects. An ANOVA of f_0 slope (`log_lh_slope`) also indicates a significant effect of `foot_syls`, *F(*3, 16.45) = 55.6, *p.adj* < .001, but not for `fin_phon`, *F*(1, 722) = 4.65, *p.adj* = .052. The model has a marginal R^2 of .21 and a conditional R^2 of .69. The marginal R^2 is quite high compared with the f_0 excursion model.

Looking at the effects of foot size on the nuclear L*H f_0 excursion (Figure 6.41A), we see that there is a small increase from the one-syllable condition (5 ST) to the two-syllable condition (5.5 ST), 95% CIs [4, 6] and [4.5, 6.5] respectively, and the estimated difference between them of 0.4 [-0.2, 1.1] ST is not statistically significant, *p* = .164. This is followed by a large jump to 7 [5.8, 8.1] ST in the three-syllable foot condition, and a distinct fall-off in the final condition to 6.1 [4.9, 7.3] ST. This lowering of the excursion is essentially a reflex of the tapering off of foot-size effects on `h_f0` in the four-syllable foot condition (see p. 94 above).

Predicted Nuclear L*H f_0 Excursion and Slope as an Effect of Foot Size

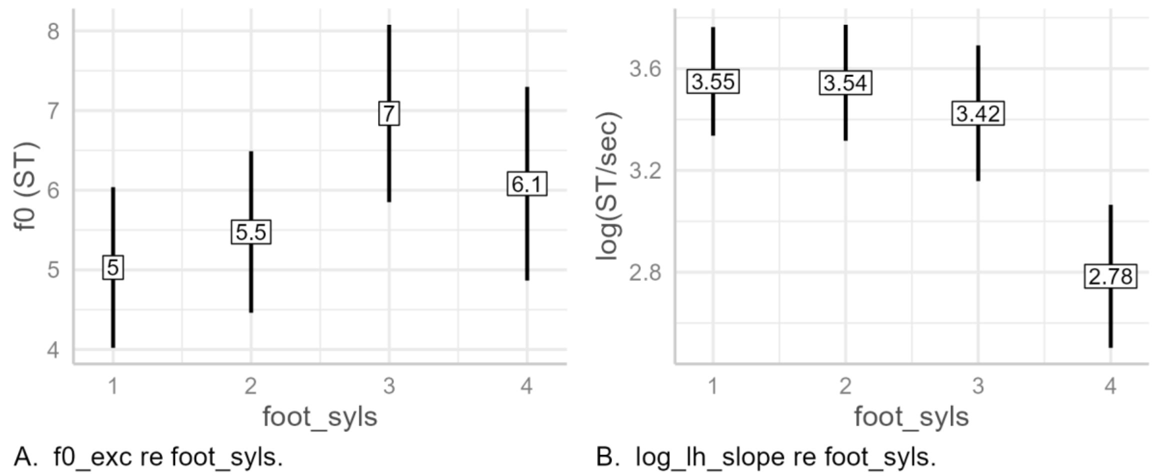


*Figure 6.41 Predicted f_0 excursion and log slope in nuclear L*H as an effect of foot size.*

When it comes to the effect of `foot_syls` on f_0 slope, there is very little change in slope until the four-syllable foot condition, where upon the slope becomes noticeably less steep, falling from 3.42 to 2.78 log(ST/s), 95% CIs [3.16, 3.69], [2.5, 3.06]. A pairwise comparison of levels of the `foot_syls` condition suggests that there is only the slightest of differences between the one-, two-, and three-syllable foot conditions, none of which are statistically significant, with estimated mean differences ranging between 0.01 and 0.13 log(ST/s) (see Table 6.31 for full statistics). In contrast, there is a difference of at least 0.64 log(ST/s) between the four-syllable foot condition and the other three conditions, each of which is statistically significant (rows three to six in Table 6.31).

Table 6.31 Pairwise comparison of levels of `foot_syls` *on nuclear* `log_lh_slope`*.*

intercept	slope	estimate	2.5% CI	95.5% CI	SE	|*t*|	df	p
foot_syls1	foot_syls2	-0.01	-0.12	0.10	0.05	-0.1	30.22	.922
foot_syls1	foot_syls3	-0.13	-0.27	0.02	0.06	-1.96	10.23	.078
foot_syls1	foot_syls4	-0.77	-0.92	-0.61	0.07	-10.75	12.48	< .001
foot_syls2	foot_syls3	-0.12	-0.26	0.02	0.07	-1.82	21	.082
foot_syls2	foot_syls4	-0.76	-0.92	-0.60	0.07	-10.3	14.45	< .001
foot_syls3	foot_syls4	-0.64	-0.76	-0.52	0.06	-11.43	15.74	< .001
intercept	fin_phonL%	0.12	0.01	0.23	0.06	2.16	721.69	.031

When the effect of foot size on f_0 slope is considered in terms of ST/s (Figure 6.42), there is very little difference among the one-, two-, and three-syllable foot conditions, but we see a fall from 32.1 ST/s in three-syllable foot condition to 16.9 ST/s in the four-syllable condition. Viewed in terms of the differential threshold of pitch change ('t Hart et al., 1990), the difference between the three- and four-syllable foot conditions is just shy of the factor-of-two perceptual threshold, while the one- and two-syllable conditions both exceed it, being at least twice as steep as the four-foot condition, at a factor 2.2 and 2.1 respectively (Table 6.32). In sharp contrast, there is either no difference or a negligible one between the slopes of the one-, two-, and three-syllable conditions.

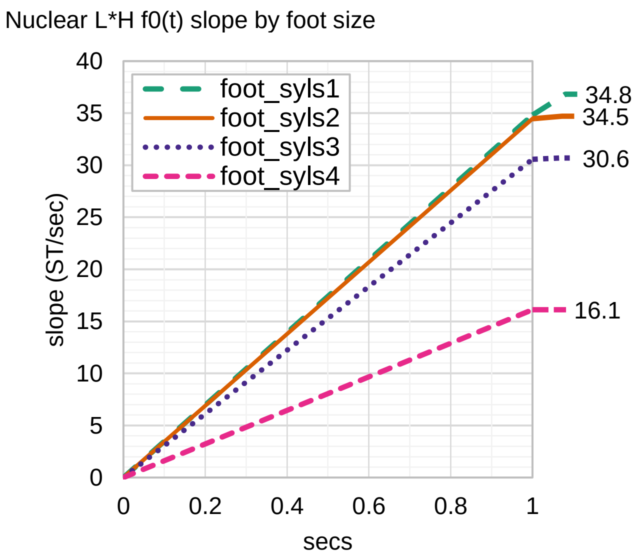


*Figure 6.42 predicted mean $f_0(t)$ slopes of nuclear L*H PA as an effect of* `foot_syls` *shown in ST/s.*

Table 6.32 Differentials of pitch change in `foot_syls` *conditions (ratio between slopes measured in ST/s).*

	foot_syls1	foot_syls2	foot_syls3
foot_syls2	1.0		
foot_syls3	1.1	1.1	
foot_syls4	2.2	2.1	1.9

It is easiest to assess the importance of these effects in terms of compression and truncation by working backwards from the four-syllable foot conditions. There is a very large increase in slope from the four- to the three-syllable foot, followed by only very slight increases in slope from the three- to the two-syllable foot, and from the two- to one-syllable foot conditions. This is highly indicative of a compression effect between the four- and three-syllable foot conditions, with little to no compression as the foot size decreases further. In contrast, there is a noticeable decrease in f_0 excursion from the three-syllable to the two-syllable foot-size condition. This is followed by a less dramatic decrease in the one-syllable foot. Given that there is little change in slope across these three conditions, this indicates truncation effects as the foot decreases in size from three syllables to one syllable. This is also what was inferred from the analysis of the H target in Section 6.6.3.2 above.

The tonal targets of the nuclear contour are plotted together as a function of foot size effects in Figure 6.43. This shows the combination of truncation and compression more clearly. Again, working backwards from the four-syllable foot condition, one can see clearly that the contour is compressed in the three-syllable condition, but is then truncated in both the two- and one-syllable conditions. This suggests that there is a limit to the amount of compression which can occur—at least for these speakers of DCE—leading to the replacement of compression by truncation as the foot becomes shorter.

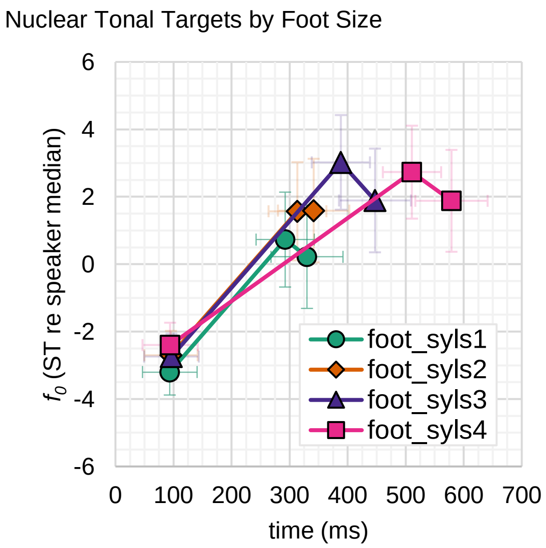


*Figure 6.43 Nuclear L*H % pitch contours plotted as a function of* `foot_syls` *alone.*

Boundary tone effects on f_0 excursion and slope are shown in Figure 6.44. A low boundary is associated with a lowering in the f_0 excursion of L*H by 0.8 ST and a slight rise in slope of 0.12 log(ST/s), 95% CIs [-1.3, -0.3] and [0.01, 0.23], p = .031 and .002 respectively. This combination of a lower excursion and little change in slope is indicative of truncation. This is reflected in Figure 6.45, which plots the estimated coordinates of the L*H % and L*H L% tonal targets. Thus, it seems that the addition of an L boundary is not just associated with earlier peak alignment but also with truncation of the rise.

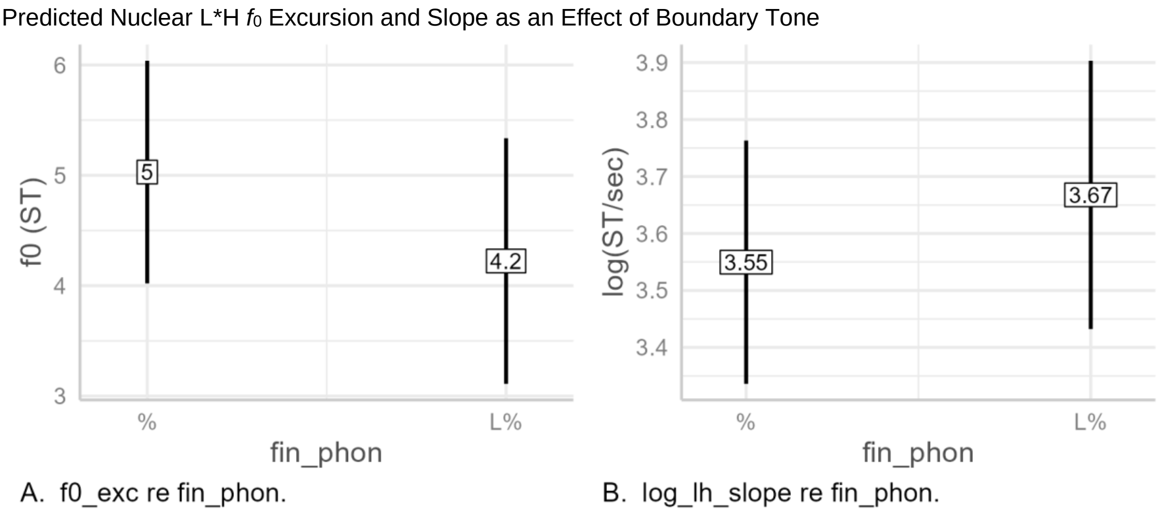


*Figure 6.44 Predicted f_0 excursion and log slope in nuclear L*H as an effect of boundary tone.*

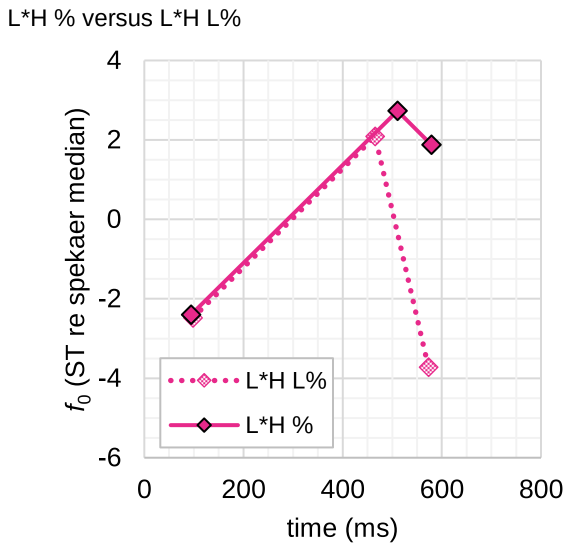


*Figure 6.45 Coordinates and 95% CIs of L*H % and L*H L% tonal targets in the 4-syllable foot condition.*

6.6.4 Comparing Metrical and Lexical Effects in Prenuclear and Nuclear PAs.

The prediction that metrical and lexical effects are stronger in the prenuclear than the nuclear position is not transparently supported by the preceding analyses. To assess the prediction more clearly, the effect size of lexical and metrical effects in each model has been calculated using the partial omega-squared statistic (ω_p^2) (Albers & Lakens, 2018; Ben-Shachar et al., 2020; Richardson, 2011).

Table 6.33 summarises the size of lexical and metrical fixed effects on each PN tonal target parameter, while Table 6.34 shows them in relation to nuclear PA target parameters. Each table summarises only effect sizes from the original models, not the additional models used to further refine the analyses. In both tables, the first column lists the fixed effect parameter with tonal target response parameters listed in the header row. `wrd_end_syl` is a lexical effect parameter; however, it is only used in the prenuclear PA models, so it is not included here. However, `ana_syls` in the PN models is comparable with `pre_syls` in the nuclear PA models, while `pn_new_word` and `nuc_new_word` are also equivalent parameters. Therefore, along with `foot_syls`, `ana_syls` / `pre_syls` and `pn_new_word` / `nuc_new_word` parameters are compared.

Table 6.33 Effect size (ω_p^2) of lexical and metrical fixed effects on prenuclear pitch accent tonal targets. Arrows indicate comparison with equivalent nuclear parameter. [17]

	l_t		l_f0		h_t		h_f0		
parameter		**ω_p^2 95% CI**		**ω_p^2 95% CI**		**ω_p^2 95% CI**		**ω_p^2 95% CI**	**Ave.**
`ana_syls`	▲	.2 [0, .49]	▼	.07 [.03, .12]	▲	.76 [.42, .87]	▲	.35 [0, .68]	▲.35
`foot_syls`	▲	.01 [0, .02]	▼	0 [0, 0]	▼	.52 [.21, .68]	▲	.25 [0, .46]	▼.20
`pn_new_word`		0 [0, 0]	▼	.02 [0, .05]	▼	0 [0, 0]	▲	.1 [0, .66]	▲.03
`average`	▲	.07	▼	.03	▲	.43	▲	.10	▲.19

Table 6.34 Effect size (ω_p^2) of lexical and metrical fixed effects on nuclear pitch accent tonal targets. Arrows indicate comparison with equivalent prenuclear parameter.

	l_t		l_f0		h_t		h_f0		
parameter		**ω_p^2 95% CI**		**ω_p^2 95% CI**		**ω_p^2 95% CI**		**ω_p^2 95% CI**	**Ave.**
`pre_syls`	▼	.01 [0, .03]	▲	.1 [0, .26]	▼	.09 [.06, .13]	▼	.02 [0, .05]	▼.06
`foot_syls`	▼	0 [0, 0]	▲	.04 [.01, .07]	▲	.7 [.6, .73]	▼	.14 [.09, .18]	▲.22
`nuc_new_word`		0 [0, .01]	▲	.03 [0, .18]	▲	.01 [0, .03]	▼	.01 [0, .02]	▼.01
`average`	▼	.00	▲	.06	▼	.27	▼	.06	▼.10

While `ana_syls` has an average effect size (ω_p^2) of .35, i.e., 35%, its nuclear counterpart, `pre_syls`, only has an average of .06. Similarly, while `pn_new_word` has an average of .03 in the prenuclear models, its nuclear equivalent, `nuc_new_word`, is even lower, .01. In contrast, while `foot_syls` has an average effect size of .2 in prenuclear tonal target parameters, it has a slightly higher effect size of .22 on nuclear tonal target parameters. The overall average effect size on prenuclear tonal targets is .19 but is noticeably lower for nuclear tonal targets, at .1 (nine percentage points lower). This gives the overall impression that nuclear pitch accents are indeed less susceptible to metrical and lexical effects than prenuclear pitch accents.

When it comes to lexical and metrical effects on alignment parameters alone, `l_t` is only minimally affected, whether in prenuclear or nuclear pitch accent models, with average effect sizes (ω_p^2) of .07 and .003 respectively. Lexical and metrical effects on `h_t`, however, are much greater, with an average of .43 in the prenuclear model and of .27 in the nuclear model. The difference here

[17] It is possible for ω_p^2 to be negative; however, the proportion of variance should be between zero and one. For the tables here and for the calculation of averages, negative values have been replaced with zero. The unmodified tables, including negative values, can be found in Appendices G1and H1.

is largely because `ana_syls` has an effect size of .76 in the prenuclear model while `pre_syls` has an effect size of only .09 in the nuclear model. Still, we can see that once combined, lexical and metrical effects have a greater effect on the alignment of prenuclear targets than on nuclear targets.

In terms of f_0 scaling, lexical and metrical parameters have a marginally larger average effect size in prenuclear models than nuclear tonal target models, averaging .065 in prenuclear f_0 models, and .06 in nuclear f_0 models. The effect sizes of `ana_syls` and `foot_syls` on PN `h_f0` are noticeably higher than their nuclear equivalents (`pre_syls` and `foot_syls`) are on nuclear `h_f0`. That is, `ana_syls` has an effect size of .35 and `foot_syls` one of .25 in prenuclear model, while in the nuclear model, `pre_syls` has an effect size of .02 and `foot_syls` one of .14.

The large effect of `ana_syls` may be due to planning, in that the speaker may have more opportunity to make physiological adjustments to achieve the high tonal target based on the amount of time gained through additional anacrusis. However, this is purely speculative. The effect size of foot size on `h_f0` is more readily explained by the fact that in the one-syllable foot-size (and stress-clash) condition, there is noticeable truncation of the PN rise in anticipation of the nuclear L target, an issue discussed previously in 6.6.2.5 above.

In summary, nuclear and prenuclear pitch accents are both affected by lexical and metrical effects, but with a greater average effect on PN pitch accents (.19 vs .10.) Added to this, we must remember that PN L targets are also prone to deletion as an effect of foot size, meaning that L*H is sometimes replaced by or realised as >H* or H*. Given the strong phonetic and phonological effects of foot size and word boundary, we can conclude that prenuclear tonal targets are more prone to lexical and metrical effects than their nuclear counterparts.

6.6.5 Analysis of H Alignment as a Proportion

The analysis of nuclear PAs indicated `h_t` is highly susceptible to changes in foot size, more so than the PN peak (.7 as opposed to .52). In nuclear PAs, the effect becomes increasingly larger as foot size increases, while it weakens in the PN position after the three-syllable foot conditions. These effects give the impression that the H target is a floating target which drifts rightwards as the foot gets longer, especially in the nuclear position. However, it is possible that the H target is aligned in such a way as to remain proportional to the foot, or possibly even to the amount of voiced material and so may be more stable than the measurements in milliseconds suggest.

For this reason, two additional types of model were generated to assess the stability of peak alignment in L*H. The first calculates H alignment proportionally to the duration of the foot. The second type measures H alignment proportionally to the amount of voiced material, measured from the vowel onset to the last voiced frame. However, this second type was not tested on prenuclear peak alignment since, with very few exceptions, voicing continues to the very right edge of the foot.

To ensure each type of model was comparable, each model was kept as similar as possible, using `foot_syls` alone as the fixed factor, with all other factors treated as random intercepts. The model for proportional prenuclear peak alignment is shown in Equation 6.8, while the model for proportional nuclear peak alignment is shown in Equation 6.9.

```
PN response ~ foot_syls + (1 | speaker) + (1 | gender)
    + (1 | ana_syls) + (1 | pn_str_syl) + (1 | wrd_end_syl)
```
(6.8)

```
nuclear response ~ foot_syls + (1 | speaker)
    + (1 | gender) + (1 | pre_syls) + (1 | nuc_new_word)
    + (1 | nuc_str_syl) + (1 | fin_phon)
```
(6.9)

6.6.5.1 Prenuclear Peak Alignment as a Proportion of Foot Size. An ANOVA of the model testing `h_t` as a proportion of the foot in the prenuclear PA indicates that the effect of `foot_syls` is significant, $F(3, 435) = 43.92$, $p < .001$. The marginal R^2 for this model is .23, while the conditional R^2 is .88. This suggests that, when considered as a proportion of the foot, `foot_syls` accounts for 23% of the variance in prenuclear peak alignment, only two percentage points higher than the results of the equivalent nuclear model.

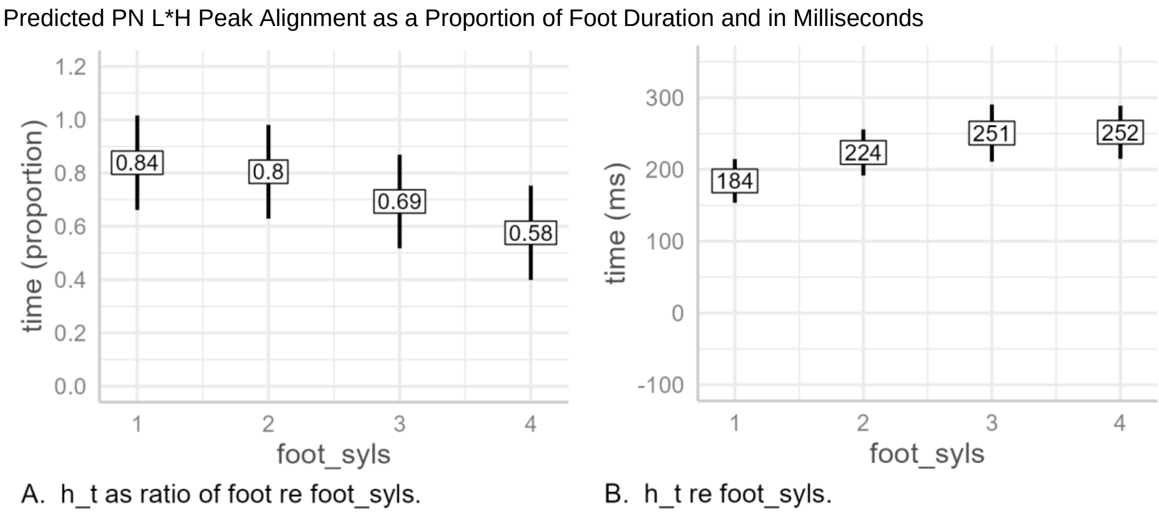


*Figure 6.46 Prenuclear L*H peak alignment as an effect of `foot_syls` when calculated proportionally to the duration of the foot (A) and when calculated in milliseconds (B).*

Table 6.35 Pairwise comparison of `foot_syls` effects when `h_t` is measured proportionally to foot size.

intercept	slope	estimate	2.5% CI	97.5% CI	SE	|*t*|	df	p
foot_syls1	foot_syls2	-0.03	-0.07	0.01	0.02	-1.62	441.02	.106
foot_syls1	foot_syls3	-0.15	-0.18	-0.11	0.02	-7.25	444.16	< .001
foot_syls1	foot_syls4	-0.26	-0.31	-0.22	0.02	-10.69	430.27	< .001
foot_syls2	foot_syls3	-0.11	-0.15	-0.07	0.02	-6.06	421.86	< .001
foot_syls2	foot_syls4	-0.23	-0.28	-0.18	0.02	-9.86	405.55	< .001
foot_syls3	foot_syls4	-0.12	-0.15	-0.09	0.01	-8.1	452.18	< .001

When we look at predicted peak alignment as an effect of `foot_syls`, we see that as the number of syllables in the foot increases, the alignment of the peak as a proportion of the foot becomes lower (Figure 6.46A), and the difference between each foot size condition is statistically significant, except for `foot_syls1` and `foot_syls2` (see Table 6.35). At first, this is odd since, when compared to the earlier analysis (repeated in Figure 6.46B for comparison), PN peak alignment actually increases with foot size and the affect becoming less pronounced as foot size increases. However, if we assume that foot duration increases roughly linearly with foot size, then the contrast disappears. This is because, if the duration of the foot increases but peak alignment does not keep pace with this change, the peak will be aligned increasingly earlier proportionally to foot. This is illustrated in the Figure 6.47, which shows a theoretical plot of alignment as an effect of foot duration.

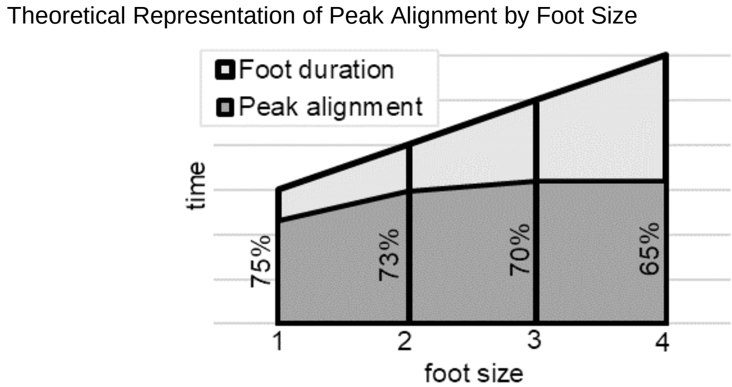


Figure 6.47 Theoretical plot of foot size and peak alignment. Foot duration increases linearly. Peak alignment also increases but less rapidly with each additional foot. Conversely, peak alignment as a proportion of the foot decreases more rapidly.

6.6.5.2 Nuclear Peak Alignment Measured Proportionally to Foot Size and Voicing Duration. An ANOVA of the model testing `h_t` as a proportion of the foot indicates a significant effect of `foot_syls`, $F(3, 19.06) = 117.37$, $p < .001$. The marginal R^2 for this model is .21, while the conditional R^2 is .76. Thus, `foot_syls`, as the only fixed effect, accounts for 21% of the variance in peak alignment. An ANOVA of the model testing `h_t` as a proportion of voiced material in the nuclear PA also indicates that the effect of `foot_syls` is significant, $F(3, 15.38) = 4.45$, $p = .035$. The marginal R^2 for this model is .01, while the conditional R^2 is .78. This suggests that, when, considered as a proportion of voicing, that `foot_syls` only accounts for 1% of the variance in nuclear peak alignment, substantially lower than 21% for the model measuring peak alignment as a proportion of the foot.

When we look at peak alignment as a proportion of the foot, we see that it ranges between an estimated 70% and 85% of the foot (Figure 6.48A), but varies very little when measured proportionally to the duration of voicing, hovering around 75% (Figure 6.48B). The largest difference in peak alignment measured proportionally to the foot is 15%, between `foot_syls4` and `foot_syls2`, 95% CI [0.12, 0.18], $p < .001$. In contrast, we see that the largest estimated difference in the voicing model is only 3% [-0.05, -0.01], between `foot_syls2` and `foot_syls3`, $p < .005$.

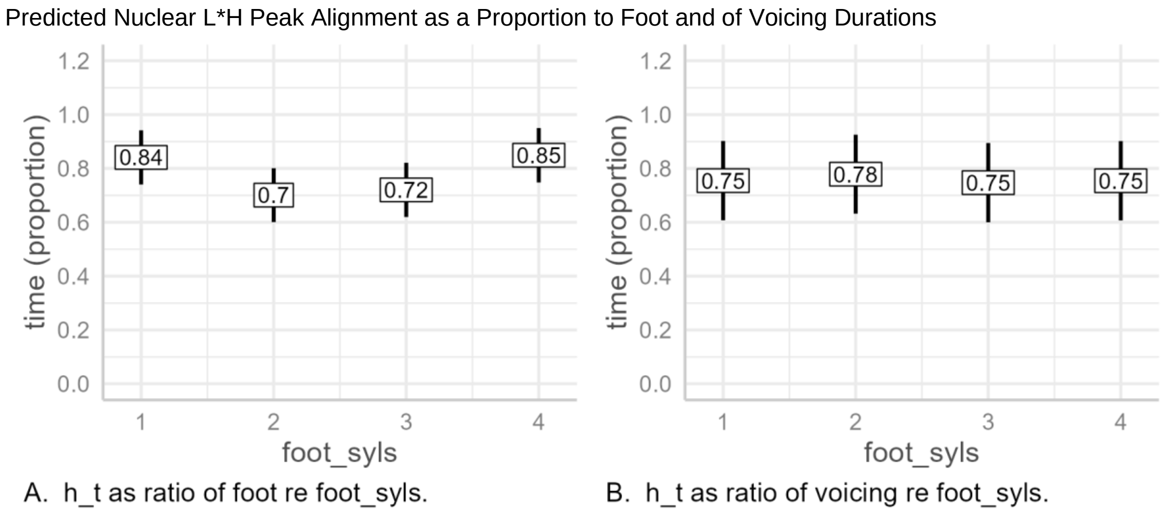


Figure 6.48 Nuclear peak alignment as an effect of `foot_syls` *when calculated proportionally to foot duration (A) and the duration of voicing (B).*

The implication of these two types of analysis, especially in the light of the initial findings that peak alignment increases with foot size, is that peak alignment in the nuclear PA is quite stably aligned as a proportion of the voiced material in the final foot. This further implies that compression and truncation actually work in tandem to maintain this proportionality. Furthermore, at least based on this data, it seems that it is not the duration of the foot which matters so much as the duration of voicing.

6.6.5.3 Summary of Proportional Peak Alignment Analysis. The results of the PN model measuring peak alignment proportionally to foot duration simply mirror the results of the original model, which measured it in milliseconds. This tells us that—*unlike* nuclear peak alignment—peak alignment in the PN pitch accent is *not* less susceptible to the effects of foot size when measured proportionally. In the nuclear analysis, however, while peak alignment increases dramatically in absolute terms (ms) with foot size, this is not the case when it is measured relative to the foot. In fact, nuclear peak alignment, when measured relative to the foot, vacillates between 70% and 85% of the duration of the foot. When measured proportionally to the duration of voicing, it remains stable at roughly 75%. Therefore, peak alignment of the nuclear pitch accent—unlike that of its prenuclear counterpart—is only susceptible to metrical effects in so much as it is adjusted to maintain a stable position proportional to the voiced material in the foot.

These results are similar to those found by Lickley et al. (2005) in relation to L- phrase accents in Dutch question forms. In that study, the authors found no evidence that the phrase accent was aligned at a fixed distance relative to the end of the utterance. Rather, they concluded that the low target attempts to align, where possible, with a post-nuclear syllable containing secondary stress and only with the final weakly stressed syllable when the former is not an option. In the current case, it seems that the alignment of the peaks is proportional to the voiced content of the phrases, and thus—as in the paper by Lickley et al.— not at a fixed duration from the phrase boundary, nor indeed at a fixed duration from the vowel with primary stress.

6.7 Summary and Discussion

This chapter has focused on the phonological form and phonetic implementation of pitch accents in prenuclear and nuclear position in neutral declaratives. The main motivation for this was to establish the extent to which lexical boundaries and meter affect both the distribution of pitch accents and their phonetic implementation, thus dealing with the descriptive concern raised by research question one:

RQ1. What are the phonological and phonetic characteristics of pitch accents in DCE in unmarked speech under variation in metrical and lexical structure?

In working towards an answer to this question, it was also possible to evaluate the importance of underlying phonological tones by assessing the extent to which they are realised in the surface phonology. This helped answer research question three:

RQ2. Is there evidence in the realisation of pitch accents in DCE for the special status of H tones?

To further guide the analysis, a series of hypotheses were generated based on the literature, provisional analysis of the data, and, to a lesser extent, intuition.

To answer the research questions and test the hypotheses, prenuclear and nuclear pitch accents in the A- and H-Corpora were analysed. Firstly, the distribution of pitch accents was evaluated in subcorpora designed to test the effects of anacrusis, foot size, and word boundaries on prenuclear pitch accents, as well as the effects of preceding unstressed syllables and foot size on the nuclear PA. After this, Bayesian generalised linear mixed-effects models (BGLMMs) were used to evaluate these effects across all utterances in the A- and H-Corpora. This was followed by a phonetic analysis of peak alignment in PN pitch accents in the H-Corpus. Linear mixed-effects models (LMEMs) were used to evaluate phonetic parameters associated with tonal targets along with compression and truncation effects on pitch accents. Moreover, wherever the results hinted at previously unexplored issues, these too were assessed using updated models. For the mixed-effects model analyses, the complete A- and H-Corpora were used, which created a much larger dataset for the analyses (Section 6.3, Table 6.5 and Table 6.6)

The summary of the analysis which follows is organized in terms of the hypotheses and expectations laid out previously in Section 6.1.

6.7.1 Pitch Accent Labelling

One problem became apparent early in the analysis and labelling of PN pitch accents (Section 6.4). Namely, there was an ambiguous quality to some PN pitch accents, wherein a distinct pitch event with a noticeable peak or a high plateau which ended after the stressed syllable was observed (much like in an L*H) but without any salient L-like quality in or near the stressed syllable. As such, these neither looked like nor sounded like L*H *or* H*. In the early analysis, these had been categorised either L*H and H* categories (Rodgers, 2019). However, on reflection it was decided that it was wiser to introduce an intermediate category for these, namely >H*, which is used in this chapter. Admittedly, this adds a degree of gradience to the range of PAs available and introduces what might be viewed as over-specified phonetic marking in the phonological description. However, it felt more honest to use the intermediate category for pitch accents which were neither clearly H* nor L*H rather than maintaining the strict two-way categorical distinction and forcing them into categories to which they did not clearly belong.

6.7.2 Metrical and Lexical Effects on Prenuclear and Nuclear PA Phonology

The first hypothesis proposed that L*H would be the dominant pitch accent in nuclear and prenuclear position. The second, third, and fourth hypotheses proposed that there would be no metrical effect on nuclear pitch accent phonology but that both metrical and lexical effects would be apparent in the distribution of prenuclear pitch accents. All four hypothesis were confirmed in the phonological analysis, as summarised below.

L*H was the only nuclear pitch accent observed in A- and H-Corpora (Section 6.5.1), while L*H accounted for 71% of the PN pitch accents in the prenuclear subcorpora (`pn_foot` and `pn_ana`). In the nuclear contour, L*H L% boundaries were also observed, but these were less common than L*H %, i.e., nuclear pitch accents with unmarked boundaries, with L*H L% accounting for only 4% of the nuclear subcorpora (`nuc_foot` and `nuc_pre`). However, L% was not associated with any lexical or metrical effect and was largely a matter of speaker preference. Intuitively, the low final boundary seems to imply a separate discourse function, one which indicates the speaker's surprise that their interlocutor was previously unaware of the propositional content of the utterance. Essentially, L% conveys the message, "…and I thought you already knew that!"

The third expectation was that an increase in foot size and anacrusis would be associated with increased occurrences of L*H in the prenuclear position. Initial analysis of the `pn_ana` and `pn_foot` data (Section 6.5.2) indicated that there was no effect of anacrusis, but that increased foot size was strongly associated with increased occurrences of L*H. Further, initial analyses of the subcorpora suggested that there was an effect of speech rate and gender. A subsequent linear model analysis of per-speaker proportional use of L*H and H* as an effect of gender and average speech rate suggested that only gender had a significant effect.

The BGLMMs indicated a significant effect of foot size on PN L*H pitch accents, but the effect reached saturation by the three-syllable foot condition (Section 6.5.4.1). That is, as foot size increases from one to three syllables so too does the likelihood of L*H, after which there is no further change in likelihood. Speech rate was significant, with an increase in speech rate associated with a decrease in the likelihood of L*H. While male speakers were less likely to produce L*H than female speakers, these results were not significant, so we cannot be sure if there is truly any effect of gender.

For H*, the results were slightly different, mostly due to the overall lower likelihood of H* (Section 6.5.4.2). The likelihood of H* generally decreases with foot size and remains below 10% for the three- and four-syllable conditions while the likelihood increases slightly with speech rate. Neither of these effects were found to be significant, however. There was a strong effect of gender on the likelihood of H*, which indicated that male speakers are much more likely to produce H* than females.

The fourth prediction was that there would be a greater likelihood of L*H when the right boundary of the word with the lexical stress occurred later in the foot. This was indeed suggested by the initial analysis of the H corpus (Section 6.5.3), where, in each case, the foot with the later word boundary had a greater number of L*H pitch accents. However, there was only a marginal difference in the "Lally's is in- | Valerie's is" pairing, suggesting that the effect of the word boundary depended on whether or not the word boundary was coterminous with the stressed syllable. The BGLMM models of L*H and H* likelihood confirmed this (Section 6.5.4). That is, L*H is much more likely when the word boundary occurs later than the right edge of the stressed syllable while conversely H* is much less likely. This effect was very strong in both cases.

6.7.3 A comment on L%

While the aim of this chapter is not an analysis of function, it is still worth revisiting the L% boundary and considering it in the light of McElholm's (1986) study of Derry City English (see also 3.2 above). The nuclear L*H L% label seems similar to the rising-falling nuclear contour (contour C) described by McElholm. However, he describes the fall in contour C as "slight", which is in sharp contrast to the size of the fall in the L*H L% phrases, where the mean fall in f_0 at the boundary was 5.5 ST. In contrast, it was only 0.5 ST in L*H %. Thus, contour C, as described in McElholm is in fact more similar to L*H % than L*H L%. Functionally, however, he describes contour C as suggesting reservation or signalling an upcoming contrast, which is quite different to the neutral declarative effect of the L*H %. Functionally more similar to the L*H L% contour in McElholm is contour D (extra-high-rising-falling), which is described as signalling surprise or assertiveness. However, the L*H L% contour—while containing a steep fall—is not extra high, and has, as noted above, a slightly lower peak than the L*H %.

6.7.4 Metrical and Lexical effects on Phonetic PA Parameters

The final set of predictions dealt with metrical and lexical effects on phonetic parameters. Hypothesis number five was that tonal alignment of PN accents would be more susceptible to metrical and lexical effects than nuclear pitch in general, while number six predicted specifically that anacrusis would affect the alignment of PN H targets. However, given conflicting findings in previous studies (Kalaldeh et al., 2009; Nolan & Farrar, 1999), it was unclear what direction this effect might be in. The seventh prediction was that there would also be competing strategies for the anchoring of PN H targets, one associated with the right word boundary and one with the right edge of the foot. Finally, in relation to compression and truncation, it was predicted that compression was likely in the nuclear pitch accent, but that truncation might also be observed. As to truncation and compression in prenuclear accents, no prediction was made.

A comparison of effect sizes in prenuclear and nuclear tonal target parameters indicated that prenuclear tonal targets were indeed more susceptible to lexical and metrical effects (Section 6.6.4).

It was also noted that there is already a clear foot-size and word-boundary effect on the phonology of the prenuclear pitch accent which leads to the deletion of the prenuclear L target.

As for the specific prediction that word boundary and meter would have a greater effect on the alignment of prenuclear pitch accent targets, this does indeed appear to be the case. The effect sizes of comparable lexical and metrical effects in the prenuclear and nuclear positions were compared using the partial omega-squared (ω_p^2) statistic (Section 6.6.4). In PN pitch accents, L target alignment was only mildly affected by metrical and lexical effects, with an average ω_p^2 of .07, while in nuclear pitch accents, the average effect size of the metrical and lexical effects was almost zero (mean $\omega_p^2 =$ .003). Effects on peak alignment were much larger, both in prenuclear and nuclear contexts but still with a noticeably lower mean effect size in the nuclear position, ω_p^2 = .43 and .27 respectively. The extremely low average effect size of metrical and lexical effects on L target alignment reflects its especially strong stability in the nuclear pitch accent.

The issue of peak tonal alignment was investigated further by assessing peak timing proportionally to foot duration in prenuclear and nuclear pitch accents, and also as a proportion of voicing in the nuclear pitch accent (Section 6.6.5). The analysis of prenuclear peak alignment as a proportion of the foot (6.6.5.1) provided results which simply mirrored the results of the original analysis in milliseconds, with foot size accounting for 22% of the variance in peak alignment (R^2 = .22). Nuclear peak alignment proved slightly more stable when measured as a proportion of the foot (Section 6.6.5.2), with the peak occurring at between 70% to 87% of the foot, CIs [.61,.79], [.73, .92], and with foot size accounting for 21% of the variance in peak alignment (R_m^2 = .21). When measured as a proportion of voicing, however, nuclear peak timing proved remarkably stable, occurring mostly at around 75% under each foot-size condition, and with foot size accounting for only 1% of the variance in peak alignment (R_m^2 = .01). The extremely low marginal R^2 of the model (in which foot size was the only fixed effect), indicates the very small influence foot size has on peak nuclear alignment when measured proportionally to the duration of voicing.

In short, while peak alignment appears generally to be much more vulnerable to the effects of foot size than the alignment of the L target, the effect size lessens in the nuclear position when the H target is viewed as a proportion of the foot, and almost disappears when analysed as a proportion of the voiced material.

In the prenuclear pitch accent, the effect size of anacrusis on peak alignment was the largest of all fixed effects in the model, ω_p^2 = .76, 95% CI [.42, .87] (6.6.4). However, the effect of anacrusis appeared largely as the difference between one syllable of anacrusis and all other conditions, including the zero-anacrusis condition. The special status of one syllable of anacrusis was hard to interpret, so the effects of anacrusis on peak alignment in H* and L*H pitch accents were tested separately, after which the pattern became clearer. In conditions with no anacrusis, PN peak alignment is noticeably delayed, whether in H* or L*H. However, with the addition of anacrusis, it shifts noticeably leftward, leading to earlier alignment. With the addition of further anacrusis, the peak begins to drift rightwards again, although never becoming as late as in the zero-anacrusis condition. The effect of *additional* anacrusis could only be observed in L*H since there were no instances of H* beyond one syllable of anacrusis.

These results go some way to reconciling conflicting findings in previous studies of nIE. Nolan and Farrar's (1999) study found that in Belfast English the addition of anacrusis led to earlier PN peak alignment, while Kalaldeh et al.'s (Kalaldeh et al., 2009) study of Donegal English found that increasing amounts of anacrusis were associated with later alignment of the PN peaks. In the current study, Nolan and Farrar's findings are borne out by the results of the analysis of anacrusis on PN H* accents and on the zero- and one-syllable anacrusis condition effects on L*H. The rightward drift effects found in the study by Kalaldeh et al., however, are borne out by the one-, two-, and three-syllable anacrusis conditions.

Two analytical approaches were adopted to assess if there were competing anchor points for H targets in prenuclear pitch accents (Section 6.6.1). First, PN peak alignment of L*H in the H-Corpus was analysed in terms of syllable-normalised time, with each syllable counting as one unit of time and peaks measured proportionally to the syllable (Section 6.6.1.1). This facilitated the comparison of peak alignment across different pairs of target utterances. The syllable-normalised time analysis indicated that the peak was typically aligned early when the syllable boundary was in an earlier syllable. Furthermore, six of the nine speakers appeared to alternate between different metrical and lexical landmarks when aligning the peaks of L*H pitch accents. However, the landmarks were not as predicted. The alignment targets were the right edge of the stressed syllable (or start of the second syllable), the right edge of the word boundary (one of the predictions), the middle of the final syllable of the word, or the middle of the third syllable. In no case was the boundary the right edge of the foot, which was the other prediction.

The H-Corpus PN L*H data were then analysed in terms of grand-mean syllable time to evaluate peak alignment of each individual target phrase in terms both of lexical and metrical anchors and of absolute timing (Section 6.6.1.2). This led to a refinement of the analysis of peak alignment anchor points, with each potential *location* of the peak as follows:

1. The right edge of (or just after) the stressed syllable.
2. The middle of the last syllable of the word with lexical stress.
3. The right edge of (or just after) the word with lexical stress.

Furthermore, this analysis indicated that there did indeed seem to be a bimodal or multimodal distribution of peak alignment in three of the six target phrases in the H corpus, further supported by a Hartigan dip test. However, the Hartigan dip test results should be treated with caution, as each observation in the data was not independent (i.e., there were multiple observations per target per speaker).

6.7.5 Truncation and Compression Effects

Hypothesis eight predicted that compression would be observed in the nuclear PA but with the H peak shifted leftward in order to accommodate cases of L%, and also in cases where the foot had fewer syllables. That is, it was assumed that the rise of the nuclear L*H pitch accent would be preserved via compression in the light of phrase-final tonal crowding and reduced syllabic content. Given that truncation had also been observed in Belfast English, it was also noted that truncation *might* be observed. There were, however, no expectations regarding how truncation or compression might manifest in prenuclear pitch accents.

In the prenuclear pitch accent, truncation appears to be the main strategy employed to maintain the L*H rise in cases of stress clash (Section 6.6.2.6). In the nuclear L*H pitch accents, both compression and truncation were observed (6.6.3.4). In the shift from a four-syllable to a three-syllable foot, peak alignment was shifted earlier, without any significant effect on peak height or f_0 excursion but with an increase in f_0 slope. This indicates compression. However, once the foot was only one or two syllables long, there was little to no change in f_0 slope while both f_0 scaling and excursion decreased systematically. This indicates truncation.

The addition of an L% boundary was associated with earlier peak alignment and a lower excursion size but only a very slight rise in slope (Section 6.6.3.4). Together, these indicate that the addition of an L% tone to the boundary does not lead to compression, as predicted, but to truncation of the rise.

Differences in slope were also evaluated using ’t Hart *et al*’s differential threshold of pitch change (’t Hart et al., 1990). This indicated that there was unlikely to be any salient perceptual difference between slopes as an effect of foot size in prenuclear pitch accents. However, in the case of nuclear PAs, the f_0 slope in the one-syllable foot condition was likely to be perceptually distinct from slopes in the three- and four-syllable foot conditions, and possibly also the two-syllable foot condition. A comparison of slopes as an effect of the final boundary tone indicated that they were unlikely to be perceptually distinct.

These results can be put in context via comparison with other studies. Xu and Sun (2000, 2002) examined the rate of change of pitch movements among a cohort of General American English speakers and Mandarin speakers studying in the US. They found, on average, that a rise of 4.8 ST was associated with a mean slope of 36.5 ST/s rise from L to H and a duration of 132 ms (Xu & Sun, 2002, p. 1403). They also found that the English speakers tended to have a longer duration of the rise and steeper slope than the Mandarin speakers. In the current study, we see that the intercept for f_0 excursion and slope (the one-syllable condition) is associated with an average rise of 5 ST, with a slope of 34.8 ST/s, and a duration of 162 ms[18], CIs [4, 6], [146, 179], and [3.43, 3.9][19] respectively, although slope rises to 39.25 [3.11, 5.34] ST/s in the one-syllable foot L*H L% condition. Thus, when compared with Xu and Sun's study, there is little difference in the slope or rise (in the intercept at least, which is the most readily comparable), but the duration is longer by 30 ms. This may seem inconsistent, because longer duration should be associated either with a similar slope but greater excursion or a shallower slope but similar excursion. However, it should be noted that slopes in the Xu and Sun study were calculated differently. They divided the height of the excursion by its duration, while here slope was calculated as the slope of the linear regression of f_0(t) between the valley and the peak. We also need to bear in mind that Xu and Sun's data were collected in a different context. That is, they were not recorded in a communicative context but via a contour repetition task using non-meaningful contours and phoneme sequences (e.g., /malamalama/). When it comes to the prenuclear pitch accents, a study of prenuclear rises in Dutch by Ladd et al. Schepman (Ladd et al., 2000) found rises less steep than Xu and Sun[20], averaging 23 ST/s and 31 ST/sec across two experiments. In the data analysed here, the average slopes of PN L*H rises ranged between 17.3 and 20.7 ST/s, which is less steep than the prenuclear rises found Ladd *et al*. However, the slopes of the nuclear rises in this current study were also on average greater than the PN rises the Ladd et al. study.

We can infer from this that nuclear rises are typically found to be steeper than prenuclear rises, which is also the case in in the current study. Unfortunately, there is not enough comparable data for a thorough cross-study comparison. However, it does seem that the speakers in this analysis tend to use rises which are less steep than those in the other studies.

6.7.6 An Alternative Phrase-accent Interpretation of Nuclear Contours

The current study has assumed that there are no phrase accents in the phonology (c.f., 2.5.1). However, a case for phrase accents is still very much present in the AM approach. Grice et al. (2000) propose that phrase accents should be identified as tones which are not simply associated with the edge of the phrase, but they can have a secondary association with post-nuclear stressed syllables. This stress-seeking behaviour is attested in the study by Lickley et al. (2005), previously discussed in Section 6.6.5.3, as well as in several other studies (Arvaniti et al., 2006a, 2006b; Arvaniti & Ladd, 2009).

It has been pointed out to me that part of the analysis presented here may fit with this notion of secondary alignment of the phrase accent, specifically the analysis of peak alignment as a function of foot size (see Section 6.6.3.2, p. 112). It appeared that the alignment effects of foot-size tapered

---

[18] The duration of the rise is only reported here for comparison with the Xu and Sun (2000, 2002) study. Appendix L contains full results of the duration of the nuclear rise.
[19] log values for slope have not been exponentiated.
[20] Xu and Sun (2002) convert the findings of Ladd, Schepman and Mennen's (2000) study to ST/s for comparison.

off after the three-syllable foot condition, with no meaningful change in peak alignment in the four-syllable foot. However, the only instance of a four-syllable foot in the nuclear position in the current dataset comes from the word *evaluating*, which has a secondary stress in the penultimate syllable, i.e., /əˈvaljəˌeɪtɪn/, which is the third syllable in the foot. Therefore, it is possible that the weakening of foot-size effects after the three-syllable condition may simply reflect the stress-seeking nature of the H target, and so the H target still aligns with the third syllable in the foot. If this is the case, and if the H target is more appropriately interpreted as a H- phrase accent, the nuclear contour might well be more appropriately represented phonologically as L* H-0% or L* H-L%.

This is entirely reasonable; however, one might argue that trailing syllables could also have secondary alignment features, once again obviating the need for the phrase accent (at least in this case), and thus retaining the L*H [L]% interpretation. Another reason for retaining the nuclear L*H pitch accent interpretation is that, when given enough segmental material, the prenuclear pitch accent is wont to be realised as L*H. The overall impression in the data as it is analysed here—much as in previous studies of nIE (see Section 3.3)— is that L*H is dominant both in prenuclear and nuclear position. If the unmarked nuclear declarative contour were to be analysed as L* with an H- phrase accent, this would fail to reflect the apparent underlying tonal similarities between the nuclear and prenuclear pitch accents.

Regardless of the affiliation of the H tone in the nuclear contour, it is worth investigating further if it does truly exhibit stress-seeking behaviour. Unfortunately, the corpus analysed here was not designed with that goal in mind, so is ill-suited for such an analysis.

6.8 Conclusion

To return to the key research questions which this chapter aimed to answer, there was a descriptive and a theoretical question. The descriptive question dealt the phonological and phonetic characteristics of declarative pitch accents under variation in metrical and lexical structure, while the theoretical question dealt with the possible special status of H tones more generally.

6.8.1 Descriptive Concerns

To begin with the phonological description, the most common pitch accent is L*H, both in nuclear and prenuclear positions, while H* occurs less commonly, but only in PN position. An additional pitch accent category was used, >H*, for PN pitch accents which were intermediate between L*H and H*. This was a practical decision to avoid a forced categorisation of ambiguous pitch accents. L* was also observed in PN position but was exceedingly rare (n = 9/788), with seven of the tokens coming from a single speaker (F17). Moreover, L* only occurs in one- and two-syllable feet, suggesting that it is an instantiation of H-deletion of L*H targets in shorter feet. There is also evidence that the increased occurrence of prenuclear >H* and H* is an effect of metrical and lexical effects, and that typically L*H is the underlying pitch accent. That is, when the stressed syllable associated with the L* is also the final syllable of the word to which the H tone may be anchored, the L is more likely to be deleted. In such cases, only the H tone remains, which is reinterpreted as

H*. A similar effect occurs when foot size decreases, leading to a greater likelihood of H* in shorter feet, while conversely L*H becomes increasingly likely as foot size increases.

Although not central to current research goals, two other effects were observed on the likelihood of L*H and H* in prenuclear pitch accents. Firstly, an increase in speech rate was strongly associated with a decrease in the likelihood of L*H. Conversely, increased speech rate was associated with a slight increase in the likelihood of H*. This latter effect was statistically non-significant, most likely due to the sparsity of H* in general. Secondly, there was also a gender effect on H*, with the male speakers being an estimated eleven times more likely to use H* than the female speakers.

When it comes to phonetic effects, PN pitch accents were found to be generally more susceptible to lexical and metrical effects than nuclear pitch accents. However, in both PN and nuclear position, the alignment and scaling of L targets was very stable. Alignment of L*H peaks in both prenuclear and nuclear position appear to be strongly affected by foot size. Further analysis indicated that, while this is indeed true for the PN peak, the nuclear peak is also stable when viewed as a proportion of the duration of voicing in the final foot, at around 75%. Thus, there is evidence of pressure to maintain peak stability in the nuclear pitch accent which is not found in the prenuclear pitch accent.

Prenuclear PA rises were subject to truncation, which was especially pronounced in stress-clash conditions, indicating that the rise of the L*H is sacrificed to ensure the stable realisation of the nuclear L target in L*H. In the nuclear pitch accent, it was found the addition of an L% boundary led to the truncation of the rise and not, as expected, simply compression. Compression was observed in between the four-syllable and three-syllable foot conditions, but once the foot was only two- or one-syllable in length, the rise was truncated. This suggests that there is a limit to the amount of compression which can be accommodated by these speakers and that the nuclear rise is not deleted, just truncated, after the limit to compression is reached.

6.8.2 Theoretical Concerns

The theoretically oriented research question for this chapter was, “Is there evidence in the realisation of PN pitch accents in DCE for the special status of H tones?” The answer to this question is *yes*, for both for phonological and phonetic reasons. Firstly, in the phonology, by far the most common strategy for dealing with earlier peak alignment (as an effect of earlier word boundaries or shorter feet) is to delete the L target while retaining the H target. Secondly, where the L target is not deleted and the L*H pitch accent remains, the H is truncated as an effect of stress clash in anticipation of the upcoming L of the nuclear L*H. This may seem, on the surface, to indicate that the L target takes precedence over the H target. A better interpretation is that the nuclear target takes precedence over the prenuclear targets. However, rather than a prenuclear H target deletion strategy to make way for the upcoming nuclear PA, a phonetic truncation strategy is available to help retain the prenuclear H tone. Such a strategy is not available for the L tone, which *is* more likely to be deleted. Thus, in the prenuclear position, we see a mechanism in place to help preserve the H target but not the L tone.

Taken together, these two phenomena—one phonological and one phonetic—indicate a special status for the H tone, even in this variety of English known for its starred L targets.

7 Analysis of Function: Sentence Modes

Chapter 3 observed that AM analyses of nIE attest to the dominance of L*H nuclear pitch accents across declarative and question forms. This includes declarative questions, which are lexically and grammatically identical to declarative statements. However, in other (standard) varieties of English, speakers tend to employ a falling nuclear contour for statements (H* L% or H*L %) and a rising contour for binary questions (L*H % or L*H H%). Even in varieties which use declarative rises (e.g., in the North of England), the declarative rise does not dominate to the same extent as in nIE.

It was also observed that, if one follows the AM approach championed by Gussenhoven (Gussenhoven, 2004) and illustrated in Haan's study of Dutch question intonation (2002), 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 the possibility that the correlation reflects a phonological event, not simply a paralinguistic one. Not to do this may lead to a corollary danger. Namely, by partitioning off uncooperative data as paralinguistic, we might simply exclude data which challenge the theory's ability to adequately describe our observations. This would 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, 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 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. It should also be noted that the register tier is used here in the sense of that described by Leben, Inkelas, and Cobler (Inkelas & Leben, 1990; Leben et al., 1989) and outlined in Chapter 2 (Section 2.3.6). In this view, the register tier has an unmarked low and a marked high. The marked high is responsible for the suspension of downstep, and more importantly for the current case, the raising of the pitch register in questions.

This chapter assesses the viability of a register-tier hypothesis in the phonology and phonetics of sentence mode in Derry City English. That is, it aims to answer the following research questions:
Descriptive: What are the phonological and phonetic characteristics of nuclear pitch contours in DCE across sentence modes?

1. **Theoretical:** Does a register-tier analysis provide a plausible explanation for phonetic variation across sentence modes in DCE?

Although utterance-wide intonational features are not central to the core questions, the chapter will also consider them since the nuclear contour is still situated within the domain of the IP[21].

7.1 Expectations

There will be two sets of analyses of the phonetics and phonology of intonation in relation to sentence 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 employing 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 twice.

Throughout this chapter, abbreviations are used for each sentence mode type. These follow the conventions used for naming target utterances during data collection. That is, they begin with a corpus reference (M) followed by a code identifying the sentence mode type, as shown in Table 7.1.

Table 7.1 Abbreviations for different types of sentence mode.

Sentence mode	Abbreviation
Declarative Statement	MDC
Wh-Question	MWH
Yes-No Question	MYN
Declarative Question	MDQ

[21] 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.

7.1.1 Phonological Expectations

Setting aside the register-tier component, L*H % is expected to dominate as the nuclear contour across sentence modes. However, as both L*H % and L*H L% were observed in the analysis of formal effects on declaratives, L*H L% is also expected. Given that there were no H* % or H*L% contours in the H- and A-Corpora, they are unlikely to occur in MDCs and even less likely in MYNs or MDQs, sentence modes with which they are not typically associated in other varieties of English.

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. Like before, we should also expect to see L% in boundary tones; however, we should also expect a difference in distribution across sentence modes.

In Chapter 6 (Section 6.5.1), it was suggested that L% is used for discourse functional purposes, namely that the speaker uses 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 they 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 function 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 General American English (GenAm), which also involves a final fall (1975). However, it is not because surprise and redundancy share the same intonational contour, or—as Sag an Liberman suggest it might (ibid. p. 497)—that redundancy is a secondary effect. Here, I would argue that use of L% instantiates the same discourse mechanism. The L% is interpretable as surprise in the echo question—which, as an echo, is inherently also redundant—since the speaker uses it to show surprise that the propositional content in the sentence conflicts with their assumptions about shared knowledge within the discourse. In the declarative statement responding to the question, “Where do you live?”, the speaker is indicating surprise that the information was not already established as shared knowledge, and also therefore—from the speaker’s perspective—the response is (or should be) redundant.

In associating L% and H% with different functions, this reflects an 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% to signal interrogativity while also indicating 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.

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 more parsimonious account of phonological change across sentence modes than the non-register-tier account. The expectations from the register-tier phonological analysis are as follows:

4. Patterns will occur which are adequately explained only with reference to both a register tier and the tonal tier.
5. The register-tier analysis will account for 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 hypothesized NOT to be primarily a function of sentence 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 of L% in either approach. The expectation from L% is:

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

7.1.2 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. This is summarised in Equation 7.1 below.

$$f_{0_{MDC}} \leq f_{0_{MWH}} < f_{0_{MYN}} < f_{0_{MDQ}} \quad (7.1)$$

If, on the other hand, the register-tier hypothesis is valid, we should expect to see the differences in 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 with changes in pitch accent and register tier[22] rather than sentence 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:

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

7.1.3 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 are thus 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 prenuclear H* PAs. H* is less likely to occur in declaratives, where PN accentuation may be less common. If we follow Haan's hierarchy (2002, and 2.3.5 above), 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 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 number of syntactic and lexical markers of interrogativity.

7.2 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, i.e., declarative statements (MDC), wh-questions (MWH), yes-no questions (MYN), and declarative questions (MDQ) (see Table 7.2). There are three phrases per sentence mode, ending with the word *valley*, *vases*, and *valuables* in turn. These

[22] Note that the term *accent phonology* is used hereon in to indicate phonological events associated with the pitch accent, including register-tier effects in the register-tier analysis. In statistical analysis, as previously, this is abbreviated to `acc_phon` as in the previous chapter.

were chosen to have both two- and three-syllable final feet. *Valley* was chosen as it ends in a fully voiced syllable, while the final /z/ in *vases* is likely to be devoiced, i.e., realised as [z̥]. By varying syllable count and the potential amount of phrase-final voiced material, this data can be used in further future research on truncation and compression. As with all the read speech stimuli, each target utterance is a response to a stimulus read by a speaking partner and is embedded in a short dialogue (see Appendix B).

Table 7.2 Target phrases for sentence mode analysis.

Code	Sentence 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**?

Table 7.3 Summary of valid utterances in M-Corpus. Red shows 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**

Ideally, there would be 660 utterances in the M-corpus. As previously, if there were speaker or recording errors, speakers were asked to record one or two extra repetitions, and only the good repetitions were retained. Unfortunately, some errors were not 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 7.3. (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.)

7.3 Methods

As with the other corpora, utterances were annotated in Praat, and a data table was extracted using the `process_texgrids` script, as described in Chapter 5, Section 5.2. IViE labelling conventions were used as the basis for the phonological labelling (Grabe, 2001). However, during the annotation process, IViE labelling proved inadequate for labelling distinctive pitch patterns—which is to be expected given the register-tier hypothesis—so modifications were made which accommodated the register-tier hypothesis while preserving the underlying IViE labelling system. These adjustments and the rationale behind them are outlined below in Section 7.4.

For the representative visualisation of count data, raw counts were adjusted to be proportionally representative, using the process outlined in Chapter 5, Section 5.3. For inferential statistical analyses, phonetic parameters are analysed using Linear Mixed-effects models (LMEMs) while phonological categories are analysed using Bayesian Generalised Linear Mixed-effects models (BGLMMs), as outlined in Chapter 5, Section 5.4. As before, inferential statistical analyses were conducted in R (R Core Team, 2022). As with the previous chapter, all code and markdown for this chapter, along with the data sets, and results of analysis can be found in the GitHub repository (https://github.com/AERodgers/PhD).

7.4 Phonological Labelling

As with the A- and H-Corpora, L*H dominated the nuclear position. However, there were many cases where using the L*H label alone would have been misleading, and an alternative which incorporated the register tier into the labelling was developed. Firstly, there were often cases in which the L*H simply occurred 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 by 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 labelling alternatives. After this, the new approach to labelling is outlined.

Figure 7.1 presents all the pitch contours for F5 for MDC2 and MDQ2. In the MDQ utterances there is a distinct shift to a register high in the speaker's range. This indicates that the speaker is making a categorical distinction between an unmarked low register and a marked higher register. Following Sosa (1999), this overall increase in pitch register could perhaps be explained by the presence of an utterance-initial high tone (%H) in the question, leading to the higher scaling of all subsequent tones throughout the utterance. While this is appealing and would mitigate the need to call upon a register tier, register shifts are not always manifested as utterance-wide pitch raising, as will be shown below.

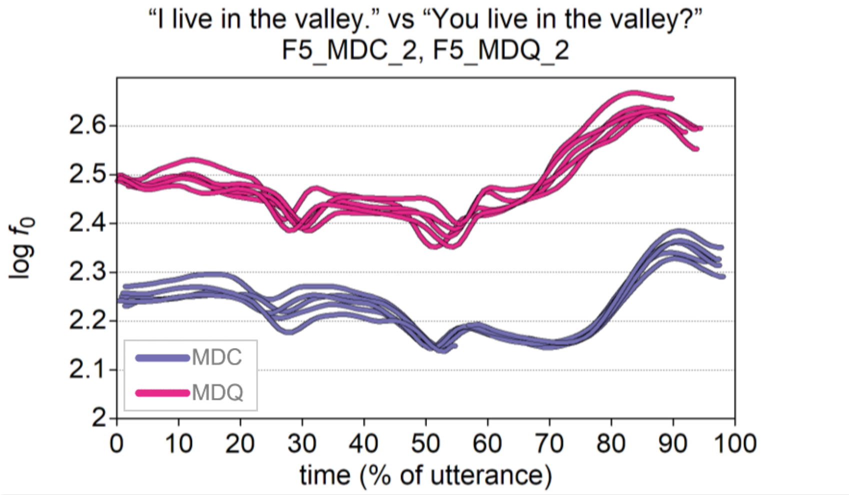


Figure 7.1 An illustration of IP-wide raised register across all repetitions. f_0 contours from F5, MDQ2 and MDC2, "You live in the valley?" and "I live in the valley." respectively.

Figure 7.2 illustrates cases where raised register occurs in the pitch accent (PA) but not in the prenuclear stretch or on the boundary tone. Each phrase has an L*H nuclear PA, which we see in the rise out of the stressed syllable in the final foot at roughly 65-75% into the utterance. One MDC has an unspecified boundary while the other ends with L%. This is also true of the two MDQ contours. During the prenuclear stretch (up to roughly 45%), each contour contains an H* pitch accent with similar pitch scaling. However, in the nuclear pitch accent of the MDQs, f_0 drops only slightly in the stressed syllable before beginning to rise. This is distinct from the MDCs, where the fall is much greater. Despite differences in the initial scaling of the nuclear PA, the rise in each MDQ is essentially a copy of the MDC pitch accent but just at a higher register. For the MDQ and MDC with the unspecified boundaries, there is a slight f_0 drop 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 $\log_{10}$ Hz, or 130 Hz). This suggests that the process causing the upward shift in f_0 during the MDQ nuclear PA 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.

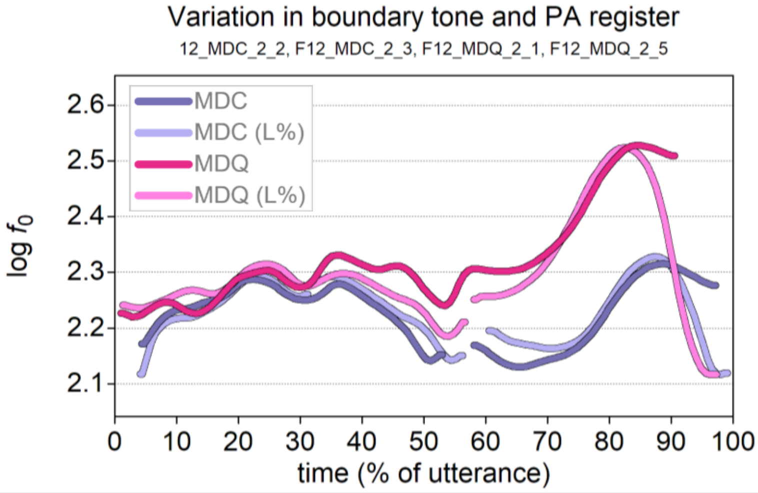


Figure 7.2 An illustration of register shift in nuclear PA which does not affect the PN stretch or the final boundary. f_0 contours from F12_MDC2_2, F12_MDC2_3, F12_MDQ2_1, and F12_MDQ2_6, "I live in the valley." and "You live in the valley?" respectively.

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 changes in f_0 due to the register shift. Alternatively, one might label them in a different way. As such, one might 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 7.2 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 earlier. It would seem odd, therefore, that while the MDC would be L*H %, the MDQ with a similar contour shape would be interpreted as H* H%. In fact, such an analysis would require that the H target at the end of each rise be ascribed to a different structural unit of the IP, i.e., to the pitch accent in the case of MDC but to the boundary in the case of MDQ. However, there is no evidence for such an analysis. In fact, in these particular cases, the peaks of the MDQs are aligned slightly earlier than 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 MDQ counterpart would become H* HL%, with the H peak again reassigned to the boundary. This would, once more, also suggest a shift in the structural association which is not borne out by the alignment of the targets. It would also require a compound boundary, although this is not an unreasonable possibility, as mentioned in Section 7.1.1. However, the compound boundary does not account for the distinctiveness of the boundary fall with its large negative excursion, which suggests that there is more happening at the boundary than the relative low of the L tone. 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 a desire to reject the possibility of a register tier while simultaneously using the boundary sequence simply to provide a more detailed *phonetic* representation of the large fall.

Finally, the shapes of the f_0 contours in the MDQ pitch accents in Figure 7.2 do not look anything like prototypical H* 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 the nuclear PA was H*, we might expect to see a sagging transition between the prenuclear H* and the nuclear H*. Rather, what we see is quite different. After the PN H*, the f_0 contour keeps falling right up until the nucleus of the lexically stressed syllable and only then does it begin to rise. All in all, it is hard to interpret either MDQ in the figure as H* H% or H* HL%.

The most sensible reading of the sample 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 register. 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, i.e., high register, is indicated by ^[...] in the labelling, while the normal (low) register is unlabelled. In this way, the four example contours in Figure 7.2 are represented symbolically as follows:

(1) MDC 1: % H* L*H 0%
(2) MDC 2: % H* L*H L%
(3) MDQ 1: % H* ^[L*H] 0%
(4) MDQ 2: % H* ^[L*H] L%

This approach preserves the clear similarity in contour shape across the four examples while also 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 7.1:

(5) MDC: % H* L*H %
(6) MDQ: ^[% H* L*H %]

An example of the labelling convention is schematised in Figure 7.3 to show that only the marked high of the register tier is represented in the labelling.

Meter	%	*	*		%
Tone		H	L	H	L
Register	L		H		L
Labelling	%	H*	^[L*	H]	L%

Figure 7.3 Schematic representation of labelling in example (4).

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 encountered 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 a self-fulfilling hypothesis trap.

The first strategy was to generate a set of alternative labels that excludes the register tier. This was done by replacing register-tier/PA combinations with alternatives which can be expressed adequately in terms of PA plus boundary tone, 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 7.2) 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 misleading to use the H* H(L)% labels. Table 7.4 summarises register-tier labelling, the non-register-tier alternative, and the rejected alternatives.

Table 7.4 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 non-register-tier alternative
L*^[H] %	L*H H%	
L*^[H] L%	L*H HL%	
^[L]*H L%	L*H L%	
^[L*H] %	L*H %	H* H%
^[L*H] L%	L*H L%	H* HL%
^[L*H L%]	L*H L%	H* HL%, H* HLL%

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

7.4.1 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, it was prenuclear pitch accents which gave rise to the greatest difficulty, specifically the second prenuclear pitch accent of wh-questions. Primarily for the sake of transparency, these difficulties are outlined below, but will be considered again in Chapter 8.

In most cases, the second prenuclear pitch accent was unproblematic, especially in cases where it was clearly L*H or 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 Chapter 5, Section 5.2). Figure 7.5 superimposes each contour on top of the other, illustrating the differences between each PN pitch accent.

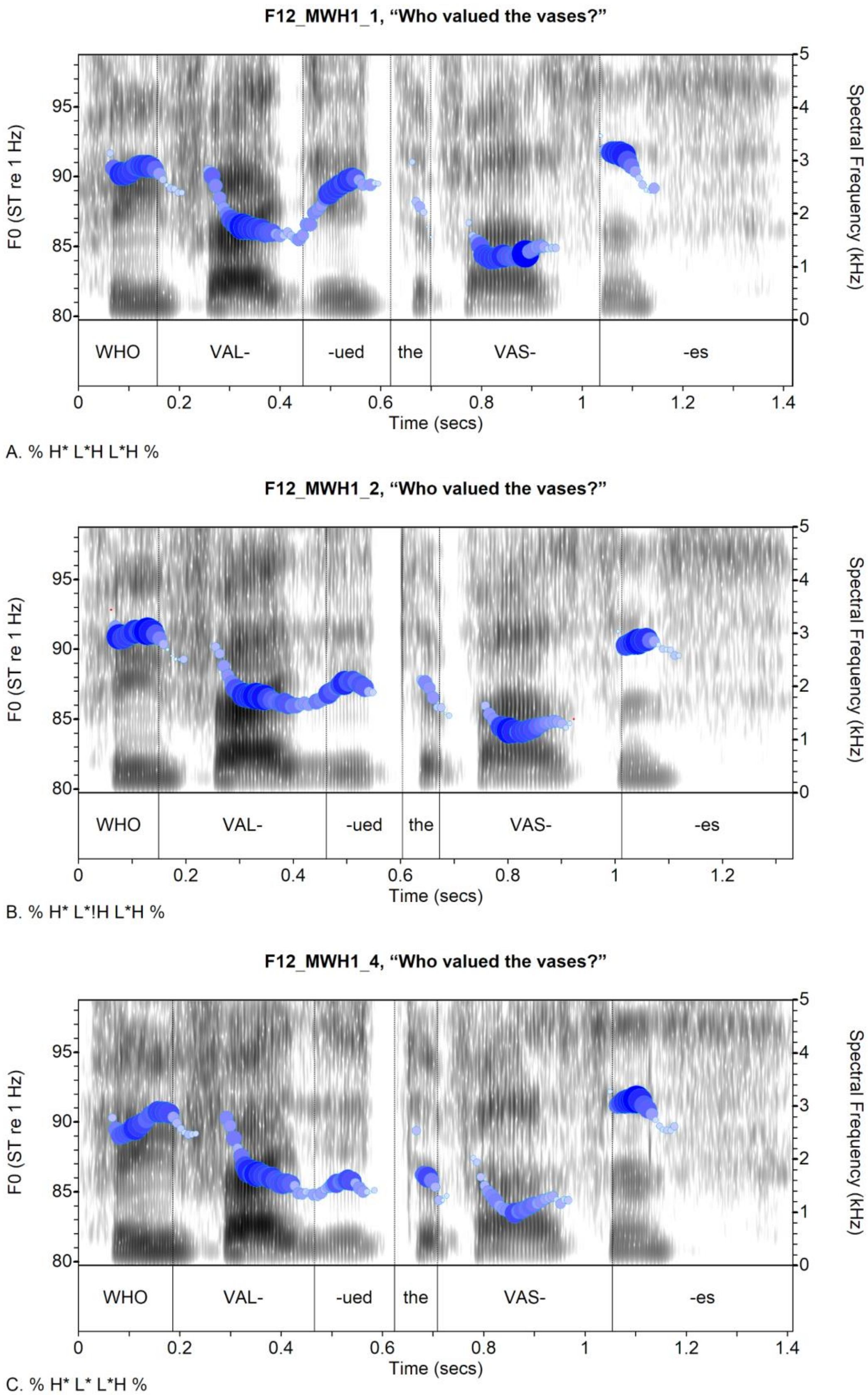


Figure 7.4 Spectrograms and f_0 contours of three repetitions of the same target with different prenuclear pitch accents in the second foot. f_0 contours show Cepstral Peak Prominence (CPP) shown by the size and intensity of the circles. The combination of f_0 and CPP is used throughout to help visually de-emphasise parts of the contour with low periodicity, often due to microprosodic segmental effects. Adapted from Albert et al. (2019).

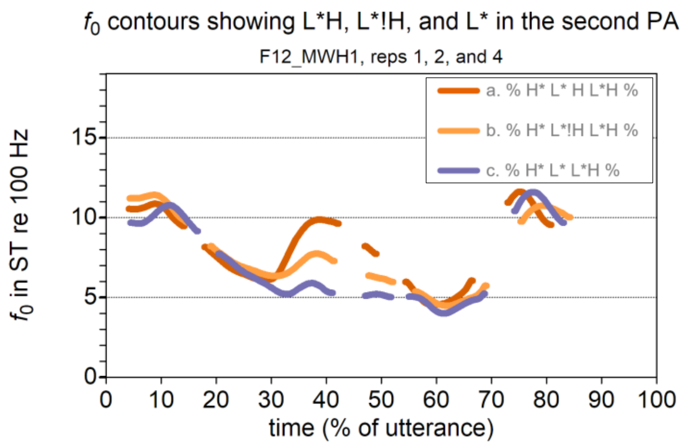


Figure 7.5 The contours in Figure 7.4 superimposed on each other (time normalised).

In terms of function, there is little apparent difference between the L*H and the L*!H, and in each case, the speaker appears to be packaging the semantic content of the utterance into three units: WHO, VALUED, and VASES. However, in the utterance with L*, there is an impression that the speaker is dividing the information between the question word WHO and its complement, VALUED THE VASES.

Another issue with the prenuclear PA pertains to the difficulty in distinguishing between accentuation and lack thereof, especially in the second foot. Sometimes, there is a slight perturbation in f_0, but it does not trigger the percept of a phonological pitch event. This issue is illustrated in Figure 7.6, where each line shows an f_0 contour from F15's repetitions of MYN2, "Do you live in the valley?"

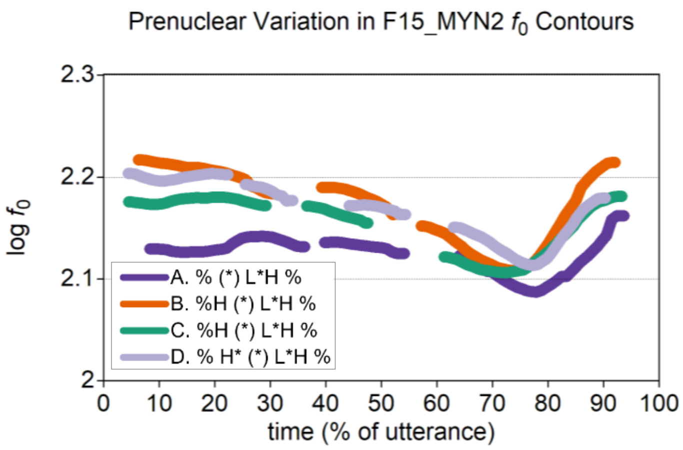


Figure 7.6 Contours for M-corpus repetitions of MYN2 by F15, "Do you live in the valley?"

In Figure 7.6, contour A begins relatively low in the speaker's range and there is no salient pitch event until the nuclear pitch accent despite a slight f_0 rise around the word "live". Contour B has an initial high boundary but no pitch accent until the nuclear pitch accent. Even though there are some minor f_0 perturbations around "live", they do not trigger a percept of prominence. Only in D is there a salient pitch event, on "you", transcribed as H*. While C also has a bump in f_0 around "you", it is auditorily less salient and was not identified as a pitch accent.

The final issue for annotation relates to whether a pitch accent in the second foot should be interpreted as L* or !H*. L* is perceptually different from !H* and visually different too, as each is typically accompanied by a different contour shape. When the L* occurs after a preceding H, the L* often has a concave elbow, with the most curved portion of the elbow occurring in the stressed syllable. This is exemplified in Figure 7.7, Panel A. This shows a concave elbow of the f_0 contour in the /a/ vowel of *val-*. In contrast, with !H*, a convex elbow is more likely in the stressed syllable as illustrated in Panel B. Figure 7.8 highlights the difference between the concavity of the L* and the convexity of the !H* by superimposing the f_0 from Figure 7.7 on top of each other. As with all labelling, 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.

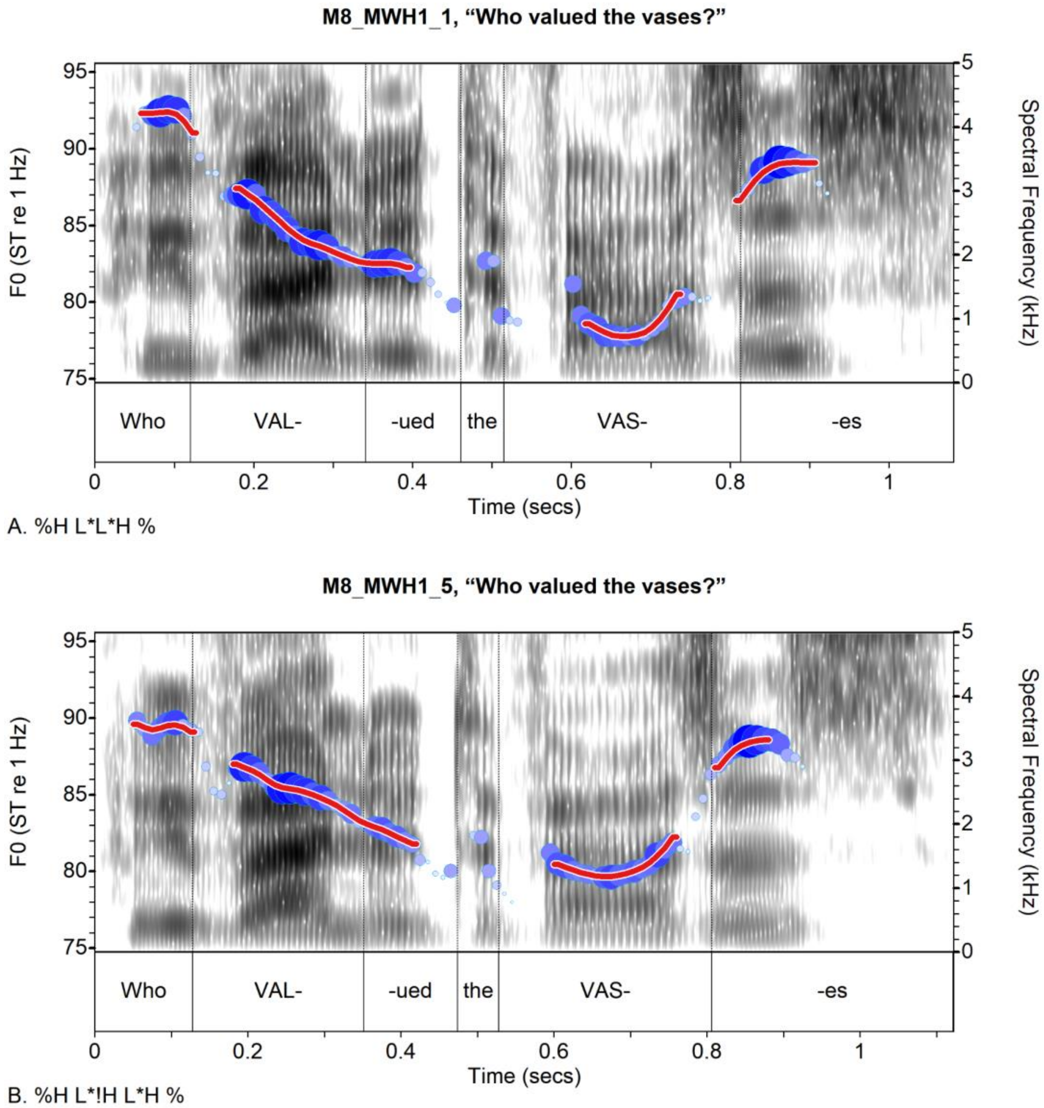


Figure 7.7 Spectrogram and f_0 contour of two repetitions of MWH1 (“Who valued the vases?”) from M5. Red lines indicate a slightly smoothed contour (bandwidth = 19) drawn with Praat settings set to ignore f_0 at time points with low amplitude and low periodicity (silence threshold = 0.15, voicing threshold = 0.6).

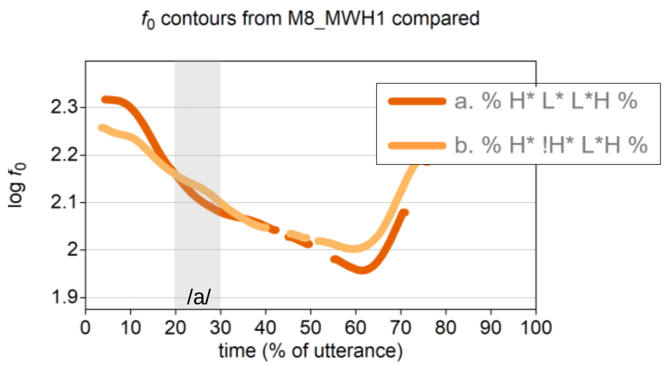


Figure 7.8 f_0 Stylised contours from Figure 7.7 compared, with time normalised to utterance duration. The grey bar indicates the vowel in "val-". (Speaker M8, two repetitions of MWH1, "Who valued the vases?")

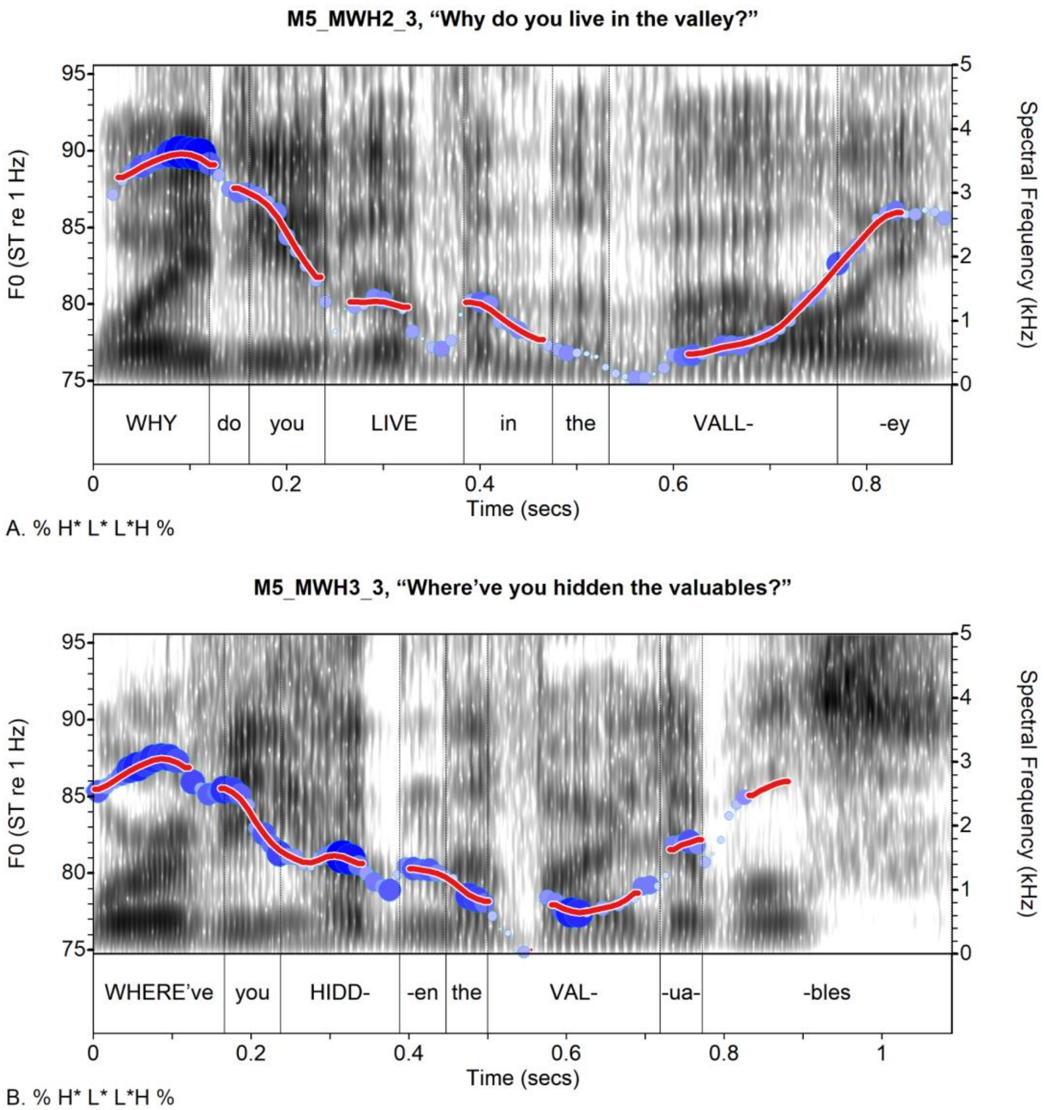


Figure 7.9 Spectrogram and f_0 contour of two utterances labelled as % H L* L*H % from M5. Red lines indicate a slightly smoothed contour (bandwidth = 19) drawn with Praat settings set to ignore f_0 at time points with low amplitude and low periodicity (silence threshold = 0.26, voicing threshold = 0.28).*

The reasoning for the L* choice is as follows. Overall, the contours curve concavely toward the stressed syllable, and the low continues throughout the lexical word. (Note, however, that the dip in f_0 before the syllable nucleus of *hidd-* is an effect of the voiced glottal fricative [ɦ] in the onset.) f_0

then drops slightly again at the onset 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 PA. This anticipatory lowering is quite common and will be revisited in Chapter 8.

All the examples discussed above reflect edge cases in categorical judgments, and most utterances did not pose such problems. Other labellers might have made slightly different judgments.

7.5 Phonological Analysis and Results

Statistical analysis of the relationship between sentence mode and intonational phonology was carried out using the non-register-tier analysis and the register-tier analysis. Non-register-tier analysis results are presented first, followed by the register-tier analysis. They are then compared, before turning to IP-wide phonology. (Note, sentence mode is shortened to *mode* hereon.)

7.5.1 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 I.) The only cases without nuclear L*H are H* (*n* = 2) and >H*(*n* = 5), both of which were 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, much as expected.

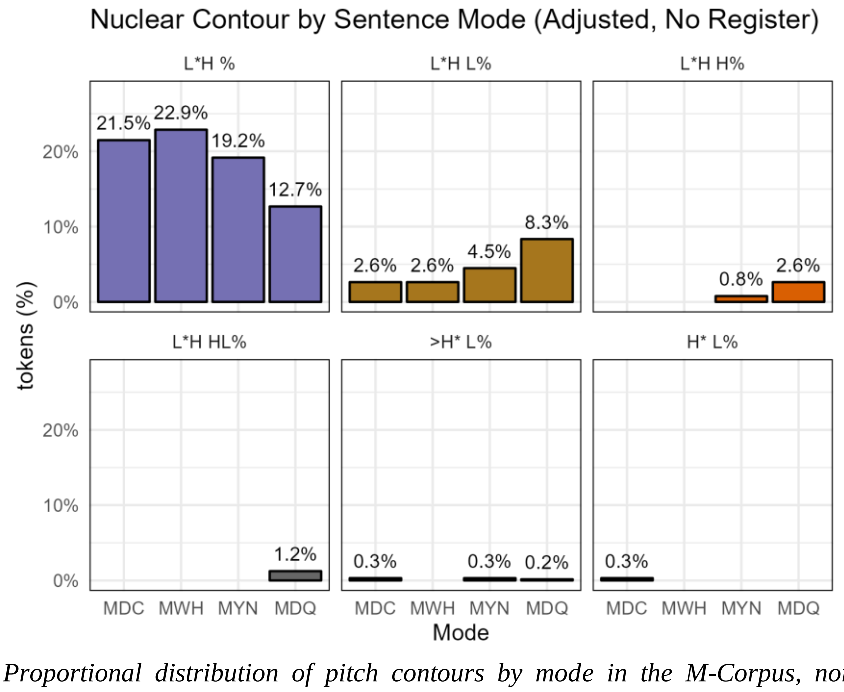


Figure 7.10 Proportional distribution of pitch contours by mode in the M-Corpus, non-register-tier analysis.

When we look at the distribution of nuclear contours (pitch accent plus boundary), we see a more interesting distribution, as illustrated in Figure 7.10, which shows the adjusted distribution of nuclear contours by mode. Percentages indicate the percentage across all tokens in the adjusted data. 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 also 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, as expected from a non-register-tier analysis of intonational phonology.

Figure 7.11 shows the boundary tone distribution by mode alone. It is almost identical to Figure 7.10, 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 7.12) That is, the male speakers tend to use L% more frequently in MDQ than female speakers, at 6.2% of all tokens in the adjusted data as opposed to 2.6% among females. Conversely, they are slightly less likely than female speakers to use the unspecified % boundary, at 5.6% as opposed to 6.9% in MDQ tokens. Female speakers, on the other hand, appear slightly more likely to use the high boundary in YNQ and DCQ than men, accounting for 3.5% of all tokens in the adjusted data as opposed to 0.8% for the male speakers. Therefore, we can speculate that gender may have some effect on the choice of boundary tone.

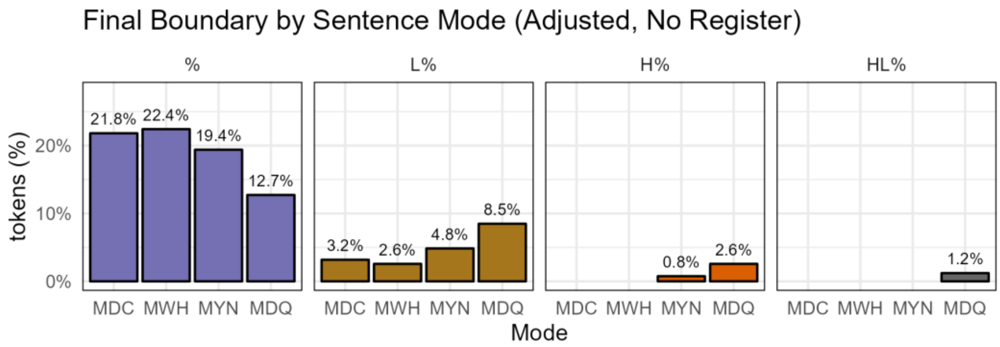


Figure 7.11 Proportional distribution of boundary tones by mode in the M-Corpus, non- tier analysis.

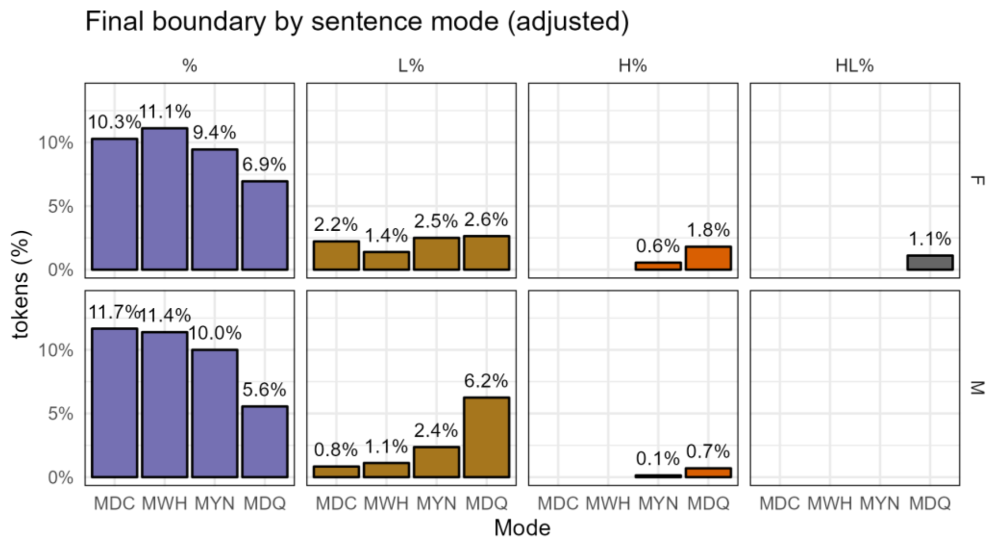


Figure 7.12 Boundary tones by mode and gender (adjusted), M-Corpus, non-register-tier analysis.

7.5.1.1 BGLMM Analysis of H% and Mode. A model was constructed to test the likelihood of H% as an effect of mode and gender, with speaker and prompt as random intercepts, as shown in Equation 7.2.

```
`H%` ~ mode + gender + (1|speaker) + (1|prompt)
```
(7.2)

A Likelihood Ratio Test (LRT) was performed on the model using the `drop1()` function in R. It indicates significant effects of `mode` but not `gender`, $\chi^2(3) = 65$, *p.adj* < .001 and $\chi^2(1) = 3.45$, *p.adj* = .076. The model has a marginal R^2 of .41 and conditional R^2 of .75. This indicates that mode and gender account for an estimated 41% of the variance in the likelihood of a high boundary tone.

Figure 7.13 illustrates two sets of statistics from the BGLMM analysis. This mode of graphical representation of the results will be used throughout the rest of the chapter, so it is worth explaining what each panel represents. Panel A shows the predicted probabilities of H% as an effect of each level `mode` while the other fixed effect (`gender`) does not change from the reference level (intercept), i.e., female speakers in this case. Thus, for example, there is an almost zero probability of H% in MDC among female speakers. Panel B shows the estimated odds ratio (OR) difference between levels of `mode` and of `gender`. Going left to right in Panel B, the first item show the estimated difference between MDC and MWH, followed by MDC and MYN, and so on, until the final comparison between female and male speakers. In each pair, the first item listed is the reference level (or intercept) while the second item is the level being compared to it. Thus, each value indicates the estimated OR difference of the second level in the pair from the first. An OR of one (the red line) indicates no predicted difference between the two. A value greater than one indicates that the outcome (H% in this case) is more likely in the second item of the pair while a value lower than one indicates that the outcome is less likely in the second item. The vertical bars in each panel indicate the 95% confidence interval. In Panel B, if the 95% CI crosses the red line at one (1:1 odds), the probability of the null hypothesis (i.e., no difference between levels or a 1:1 OR) is greater than .05, i.e., $p > .05$.

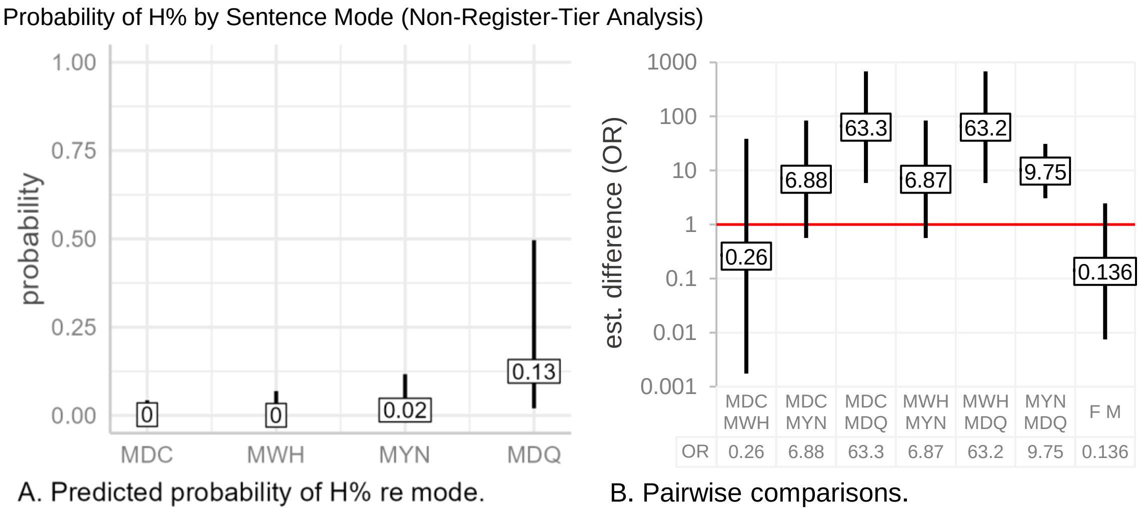


Figure 7.13 Graphical summary of predicted probabilities of H% by mode and pairwise comparison of likelihood of H% across levels of mode and gender. (Data from the M-corpus and labelling from the non-register-tier analysis.)

As can be seen in Panel A of Figure 7.13, the predicted probability of H% as an effect of mode is almost zero for MDC and MWH, 95% CIs [0, .04] and [0, .07] respectively. While MYN has a slightly higher probability of being associated with H% at .02 [0, .12], it is still very low. MDQ is the most likely to be associated with H%, but still only with an estimated probability of .13 [.02, .5].

Turning to the pairwise comparison of slopes in Panel B, we see that only MDQ is significantly more likely to have a H% when compared to the other levels of mode. That is, it is 63 times more likely to be associated with H% than both MDC and MWH, and 9.75 times more likely than MYN, 95% CIs [5.88, 682], [5.86, 682] and [3.06, 31.1] respectively, $p < .001$ in each case. The male speakers are approximately 85% [0, 2.46] less likely than the females to use H%, $p = .177$, but with the 95% CIs crossing through 1:1 odds, we cannot be confident that there truly is an effect of gender.

7.5.2 Register-tier Analysis

Looking at the nuclear contour in the register-tier analysis labelling—shown in Table 7.5—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 7.5 Nuclear contour / register-tier tokens in the M-Corpus (raw data).

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

7.5.2.1 Nuclear Pitch Accents. There are seven tokens without an L*H-like nuclear PA. As noted in the non-register-tier analysis, these are >H* L% (5 tokens), and H* L% (2 tokens). These are the only tokens similar to the falling nuclear contour of standard Southern British English and General American English; however, of the >H* L% tokens, only two 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 a fall to the low boundary does not serve the same function in DCE as in standard varieties, or at least does not 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 pitch 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 7.6 shows 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 occur more frequently again for MDQ, with 25 tokens (adj.) for L*^[H] and 74 for ^[L*H].

Table 7.6 Distribution of pitch accent by mode (adjusted) in the M-Corpus.

mode	H*	>H*	^[L*]H	L*H	L*^[H]	^[L*H]
MDC	2	2	0	162	0	0
MWH	0	0	0	165	0	0
MYN	0	2	3	117	5	38
MDQ	0	1	3	62	25	74
Total	2	5	6	506	30	112

There is also considerable interspeaker variation in the use of pitch accents, which Figure 7.14 demonstrates (next page). 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, F16, and F17) 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, therefore, is that high register is associated with YNQs and DCQs, but that it is optional rather than obligatory.

The relationship between mode and high register was tested using the model in Equation 7.3, with `mode` and `gender` as fixed factors, and `speaker` and `prompt` as random intercepts.

```
nuc_H_reg ~ mode + gender + (1 | speaker) + (1 | prompt)
```
(7.3)

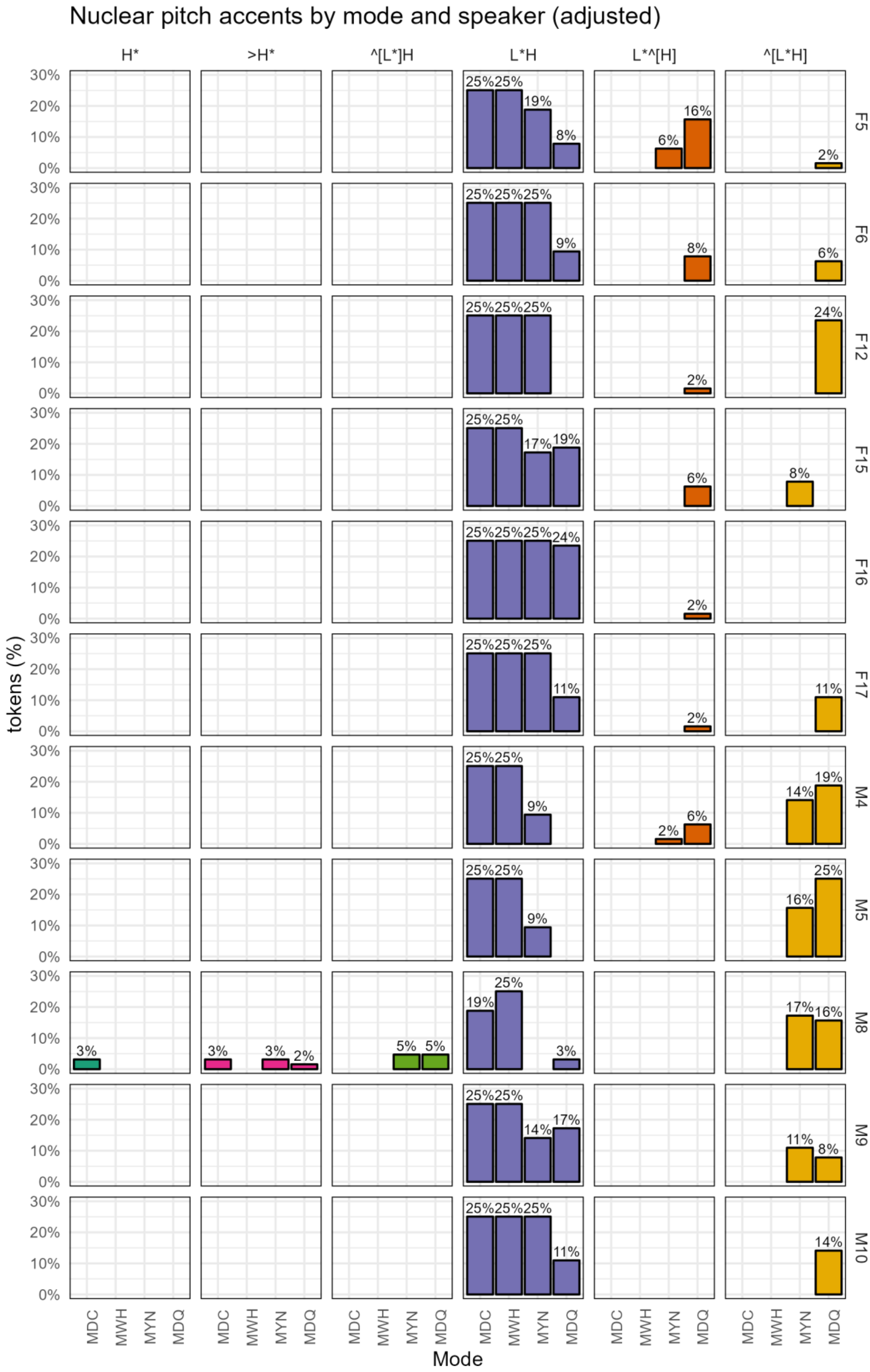


Figure 7.14 Proportional distribution of pitch accents by speaker and mode (adjusted data) in the M-Corpus. Percentages refer to the individual speakers (listed on the right y-axis).

A Likelihood Ratio Test (LRT) was performed on the model using the `drop1()` function in R. It indicates significant effects of both `mode` and `gender`, $\chi^2(3) = 320$, *p.adj* < .001 and $\chi^2(1) = 5.38$, *p.adj* = .027. The model has a marginal R^2 of .69 and conditional R^2 of .79. This indicates that mode and gender alone account for an estimated 69% of the variance in the likelihood of high register.

As both fixed factors are significant, Figure 7.15A shows the predicted probabilities of high register as an effect of `gender` and `mode`. For the female speakers, the probability of high register in MDC and MWH as an effect of mode alone is zero, 95% CIs [0, .01], [0, .01] (Figure 7.15A). There is a slight rise in probability to .2 [.03, .28] in MYN followed by a large rise to .5 [.237, .762] in MDQ. The predicted probabilities of high register for the male speakers are much greater in both MYN and MDQ, with the probability of high register at .39 [.15, .69] in MYN and .85 [.6, .95] in MDQ. The lower limit of the 95% CI for MDQ is .6, so we can conclude that male speakers are more likely than not to use high register in declarative questions.

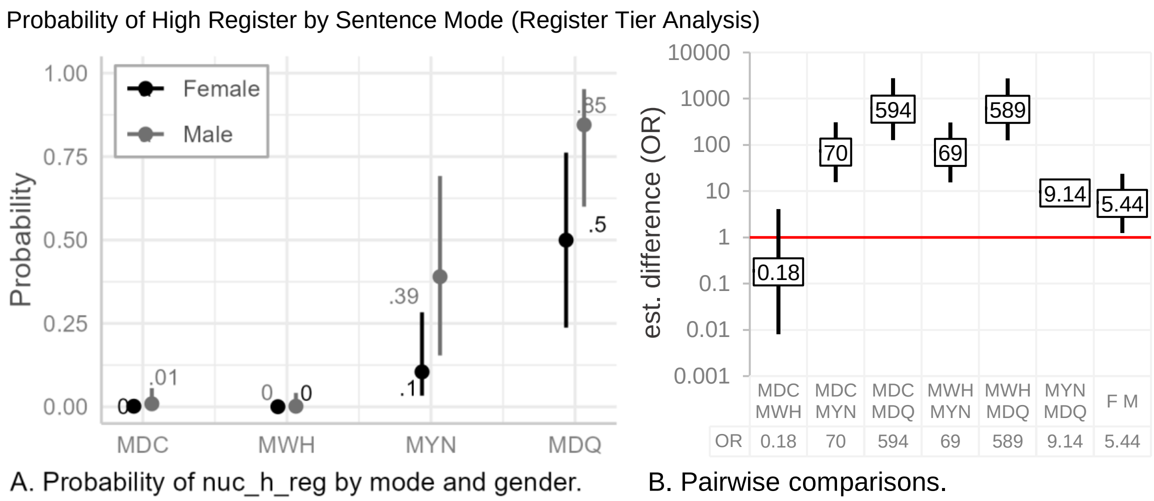


Figure 7.15 Graphical summary of predicted probabilities of high register by mode and gender, and pairwise comparison of likelihood of high register across levels of mode and gender. (Data from the M-corpus and labelling from the register-tier analysis.)

Looking at the estimated differences between each level of mode (Figure 7.15B), we see that MYN and MDQ are both more likely be associated with high register in the nuclear pitch accent than either MDC or MWH. That is, high register is an estimated 70 [15.7, 309] times more likely in MYN than MDC, and 69 [15.5, 306] times more likely in MYN than MWH, $p < .001$ in each case. High register is 594 [127, 2777] times more likely in MDQ than in MDC and 587 [126, 2754] times more likely in MDQ than in MWH. The difference in gender is reflected in the fact that a high boundary is 5.4 times more likely among the male than the female speakers, CI [1.24, 23.8], $p = .025$.

7.5.2.2 Boundary Tones. There are only two boundary conditions in the register-tier analysis, L% and the unspecified boundary (%). The low boundary tone may also be affected by high register, e.g., there are instances of ^[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. That is, a nuclear contour such as L*H ^[L%] would be nonsensical, since it would be indistinguishable from L*H %. L*H ^[%] would be equally nonsensical since the boundary is unspecified, and therefore, there would be no tone to raise via high register. Therefore, the occurrence of raised register at the boundary is essentially an extension of raised tone across the whole IP, in the nuclear pitch accent, or in the final tone in the pitch accent.

Since it has been established that high register in the nuclear contour is already exclusively associated with MYN and MDQ sentence modes, and since raised register at the boundary is an extension of raised register preceding the boundary, the inclusion of raised register in the analysis boundary tones as an effect of mode would be uninformative. Therefore, high register is not included in the analysis boundary tone effects on mode.

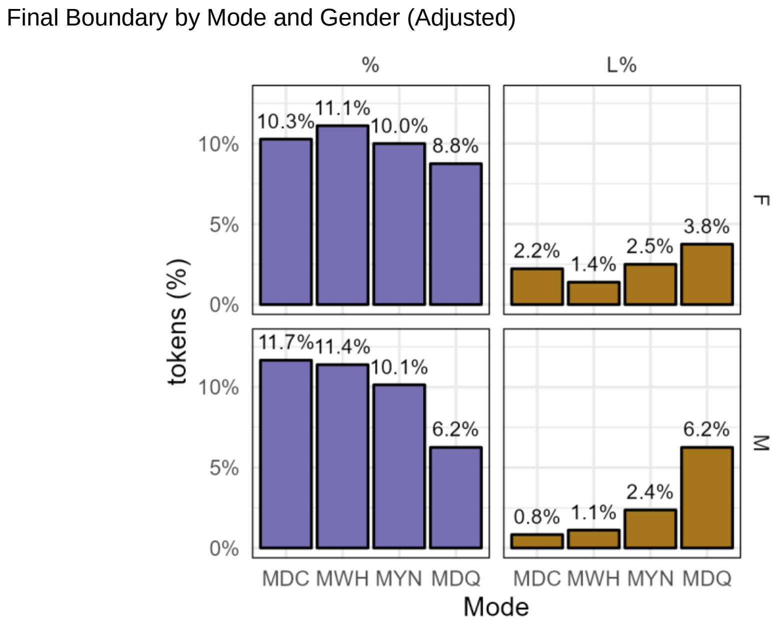


Figure 7.16 Final boundary by mode and gender (adjusted data, register-tier analysis) in the M-Corpus. High and low register have been collapsed into a single category.

Figure 7.16 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 of H%. Among the female speakers there is a slight drop in L% between MDC and MWH (from 2.2% to 1.4%) before a small but steady rise in MYN and MDQ (2.5% and 3.8%). Among the male speakers, however, while the use of L% in MDC, MWH, and MYN is less common than for female speakers, at 0.8%, 1.1%, and 2.4% respectively, there is a noticeably large increase of L% use in MDQ, at 6.2%. However, L% accounts for only 20.4% of all final boundaries, with almost half of those—i.e., 10% of all tokens—occurring in MDQ.

To test the effects of mode and gender on the boundary in the register-tier analysis, a BGLMM analysis of the model was conducted, as shown in Equation 7.4.

```
`L%` ~ mode + gender + (1 | speaker) + (1 | prompt)
```
(7.4)

A Likelihood Ratio Test (LRT) was performed on the model using the `drop1()` function in R. It indicates a significant of `mode` but not of `gender`, $\chi^2(3) = 62.8$, *p.adj* < .001 and $\chi^2(1) = 0.04$, *p.adj* = .914. The model has a marginal R^2 of .07 and conditional R^2 of .79. The conditional R^2 of the model is the same as that of the high register model outlined in the previous section (7.5.2.1), indicating that, overall, both models explain the same amount of variance in the outcome parameter. However, the fixed effects only explain 7% of the variance in L% likelihood, which is much lower than the high register model (69%).

The results of the statistical analysis reflect the findings of the initial analysis of the adjusted data. As shown in Figure 7.17A, the predicted probability of L% in either MDC or MWH is very low, at .03 and .02, 95% CIs [0, .39] and [0, .3] respectively, while MYN is only slightly more likely to be associated with L%, with a predicted probability of .05 [0, .54]. As with high register, L% is most probable in MDQ, at .23 [.01, .86]; however, this is still much lower than the probability of high register in MDQ, which was .5 [.24, .76].

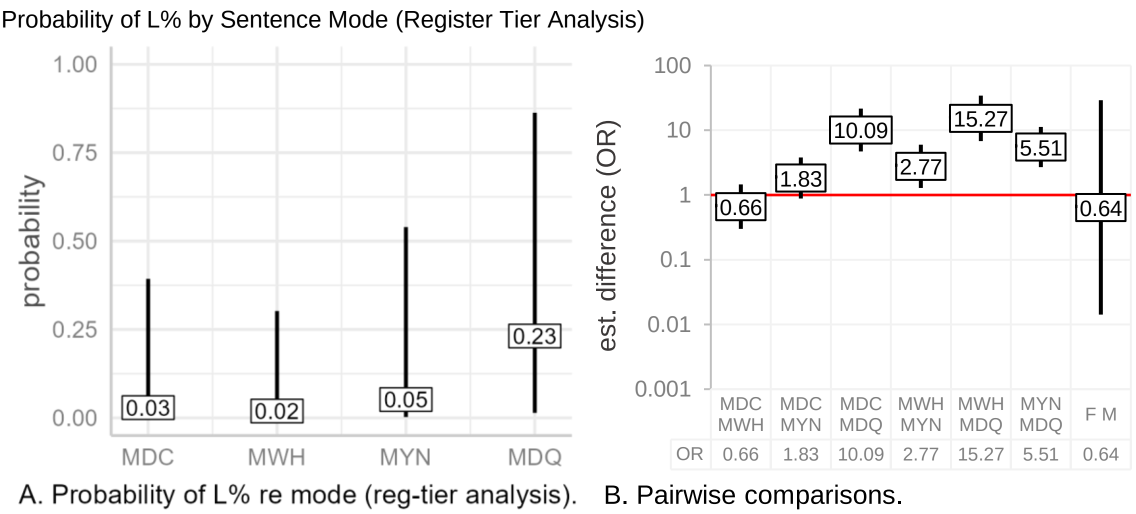


Figure 7.17 Graphical summary of predicted probabilities of L% by mode and pairwise comparison of likelihood of L% across levels of mode and gender. Note that register tier effects at the boundary are excluded. (Data from the M-corpus and labelling from the register-tier analysis.)

Looking at the pairwise comparisons (Figure 7.17B), we see that there is little difference in the likelihood of L% between MDC and either MWH or MYN, with odds ratios of 0.66 and 1.83 between each pair, 95% CIs [0.3, 1.46] and [0.88, 3.8], p = .304 and .104 respectively. The likelihood of L% in MYN compared with MWH is slightly higher, with a statistically different OR of 2.77 [1.28, 5.98], p = .009. The difference in the estimated likelihood of L% between MDQ and the other modes are all noticeably larger (and statistically significant), with OR differences of 10.1 [4.7, 21.6], 15.3 [6.7, 34.3], and 5.5 [2.7, 11.3] respectively, $p < .001$ in each case. However, although the likelihood of L% in MDQ is greater when compared to other sentence modes, the differences between MDQ and both MDC and MWH are several orders of magnitude lower than those found in the high register model. In that model, high register was roughly 590 times more likely in MDQ than in MDC or MWH.

Finally, when we compare the effect of gender on the likelihood of L%, there is little difference between the male and female speakers, with male speakers an estimated 36% less likely than the female speakers to use L%, OR = 0.64, 95% CI [0.01, 29]. The exceedingly wide CIs suggest that there is no meaningful effect of gender at all. In fact, differences which appeared to be gender-specific in the initial analysis—in which the male speakers appeared slightly more likely to use L%—are much more likely simply to be speaker-specific.

7.5.2.3 Summary of register-tier analysis. 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. Given the overall low marginal R^2 of the L% model (.07) compared the high register model (.69) and given also that L% is slightly more probable in MDC and MWH than high register (which is approximately zero), the view that L% is not in itself a marker of interrogativity is upheld. That is, in MDQ, L% is more likely to indicate surprise in response to the interlocutor's previous statement.

7.5.3 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 of H% as a marker of interrogativity, while they explain 69% of the variance in the register-tier model. To assess the effect of `mode` alone, each model was retested with `gender` as a random effect, but this created convergence issues, so `gender` was removed as a factor (see Equation 7.5).

```
response ~ mode + (1 | speaker) + (1 | prompt)
```
(7.5)

ANOVAs[23] of the non-register-tier H% and the high-register-tier models indicated that both were significant, $\chi^2(3) = 65.3$, $p < .001$ and $\chi^2(3) = 318.6$, $p < .001$ respectively.

The marginal R^2 of the non-register-tier H% model is .33 with a conditional R^2 of .74, while the marginal R^2 of the nuclear high-register model is .63 with a conditional R^2 of .8 (Table 7.7). These results indicate that mode accounts for 33% of the variance in the non-register-tier H% model but that it accounts for a much larger 63% in the register-tier analysis model. The amount of variance explained by each model in total, on the other hand, is quite similar, at 74% for the non-register-tier model and 80% for the register-tier model.

Table 7.7 Marginal and conditional R^2 (R^2_m and R^2_c) of BGLMM models for H% and high register analysis.

Type of analysis	Model	r^2_m	R^2_c
Non-register-tier	\`H%\` ~ mode + (1 | speaker) + (1 | prompt)	0.33	0.74
Register-tier	nuc_H_reg ~ mode + (1 | speaker) + (1 | prompt)	0.63	0.80

Of course, the non-register-tier analysis is bound to be less informative since only L*^[H] (L)% was relabelled with a H boundary, i.e., it was interpreted as L*H H(L)%. Other instances of

[23] As there is only one fixed effect in these models, the `drop1()` function will not work. Therefore, an ANOVA was conducted against the null model in each case.

L*H with high register were not relabelled with H(L)% because that would have required the pitch accent also be relabelled as H* despite the fact that, in terms both of contour shape and auditory percept, they were still clearly L*H. As observed previously (Section 7.4), such relabelling would have required an improbable re-association of tones with different structural elements of the IP and would have represented a description of the data in service to the theory more than a true representation of the data. However, had ^[L*H] (L)% contours been relabelled as H* H(L)%, the results for the non-register-tier analysis would have been almost identical.

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.

7.5.4 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 7.8 shows the total number of tokens for IP-wide phonology accounting for at least 1.25% of the corpus (see Appendix I, Table I1.3 for the full list). Overall, the most common IP-level intonation pattern is % L*H % (n = 78), which indicates no prenuclear accentuation. In fact, 23.3% (n = 147) of all IPs contain no prenuclear accents, a phenomenon which was not expected, and which warranted further investigation.

Table 7.8 IP-level intonation by mode for tokens accounting for at least 1.25% of the M-corpus (raw counts).

#	Phonology	MDC	MWH	MYN	MDQ	total
1	% H* L* L*H %	4	49	0	0	53
2	% H* L*H %	33	0	4	0	37
3	% L*H %	32	0	31	15	78
4	% >H* L*H %	30	0	24	7	61
5	% L*H L*H %	25	0	26	13	64
6	% H* L*!H L*H %	0	21	0	0	21
7	% >H* ^[L*H] %	0	0	14	19	33
8	% H* L*H L*H %	0	17	0	0	17
9	% H* !H* L*H %	5	16	0	0	21
10	% L*H L%	1	0	12	0	13
11	% >H* L*H L%	10	0	2	1	13
12	% L*H ^[L*H] %	0	0	3	10	13
13	% L*^[H] %	0	0	4	9	13
14	% >H* L*!H L*H %	0	8	0	0	8
15	% H* L*H %	3	8	1	0	12
16	% H* ^[L*H] %	0	0	7	5	12
17	% H* L*H L%	6	0	2	0	8
18	% ^[L*H] %	0	0	6	6	12

Table 7.9 summarises the raw counts for nuclear-PA-only utterances. MYN and MDQ show the highest amount of nuclear-PA-only use, at 34.6% and 35.3% respectively, while MDC is much lower, at 22.7%. There is only one utterance where MWH lacks prenuclear accentuation. It is unsurprising that all but one MWH utterances have prenuclear 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 not accentuating the prenuclear content of the IP, greater salience is lent to the rise in the nucleus.

Table 7.9 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 prenuclear accentuation. To test this, the model shown in Equation 7.6 was used 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.

```
nuc_PA_only ~ mode + gender + (1 | speaker) + (1 | prompt)
```
(7.6)

An LRT of the model indicates that both `mode` and `gender` are significant, $\chi^2(3) = 114.1$, $p < .001$ and $\chi^2(1)=5.47$, $p = .027$ respectively. It also has a marginal R^2 of .39 and conditional R^2 of .75. (See Appendix J for complete output from the model).

Because both fixed factors are significant, Figure 7.18A shows the predicted probabilities of high register as an effect of `gender` and `mode` combined. Unsurprisingly, the lowest predicted probability of a nuclear-PA-only IP for both male and female speakers is with MWH, at 0 for female speakers and at .01 for male speakers, CIs [0,.09] and [0, .01] respectively. This is because almost every MWH also has a PA on the question word at the start of the sentence. Setting aside MWH, we can see that the predicted probability of female speakers using nuclear-PA-only IPs is much higher than for the male speakers, at .24 [.04, .73] in MDC, .44 [.9, .87] in MYN, and .45 [.004, .27] in MDQ. For the male speakers, the probability only reaches .09 [.01, .48] in MYN and MDQ. In fact, male speakers are an estimated 8.35 times less likely than the female speakers to produce nuclear-PA-only IPs, OR = 0.12 [0.015, 0.947], $p = .044$. This is shown in the last column in Figure 7.18B.

The fact that MWH is much less likely to have only one single PA compared to the other modes is reflected in the pairwise comparisons of mode in Figure 7.18B, in the first, fourth, and fifth elements in the figure. Similarly, as we already saw in the predicted probability, there is almost no difference between MYN and MDQ, with a nuclear-PA-only IP a mere 1.06 times more likely in MYN than MDQ, 95% CI [0.6, 1.86], $p = .847$. However, both MYN and MDQ are reliably more likely to be associated with a nuclear-PA-only, each with odds ratios of 2.5, 95% CIs [1.37, 4. 5] and [1.39, 4.61], $p = .003$ and .002 respectively.

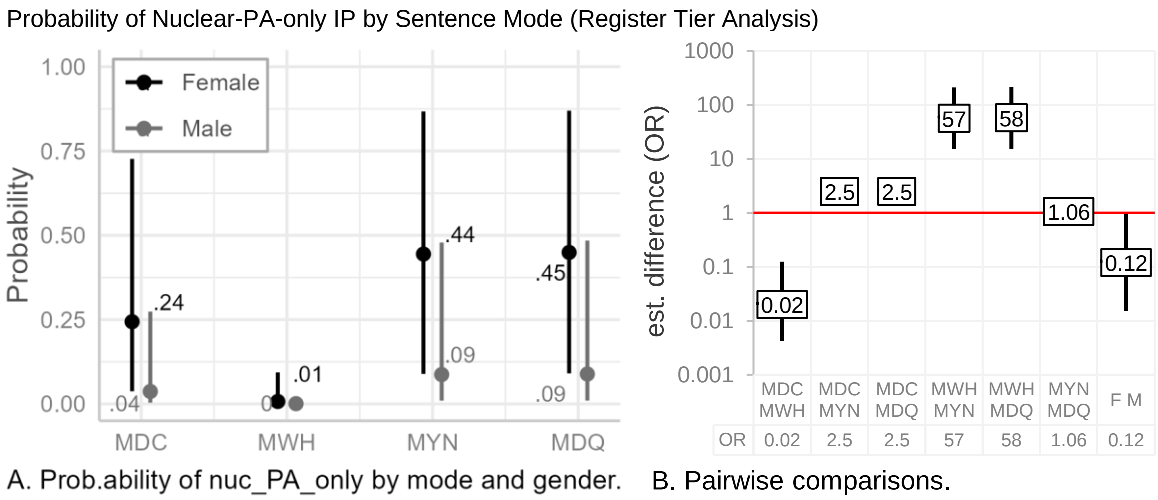


Figure 7.18 Graphical summary of predicted probabilities of Nuclear-PA only IP by mode and gender, and pairwise comparison of likelihood of nuclear-PA-only IP. (Data from the M-corpus and labelling from the register-tier analysis.)

7.5.5 Combining IP-wide and Nuclear PA Strategies

Based on the analysis of the data so far, females are more likely to use the nuclear-PA-only strategy (Section 7.5.4) while males are more likely to use the high-register strategy to distinguish MYN and

MDQ from other modes (Section 7.5.2.1). Thus, it appears that two, albeit gendered, phonological strategies are available to help identify question types and to distinguish MDC from MDQ. A count of the raw data shows that more than 56% of MYN tokens and 80% of MDQ tokens include at least one of the two strategies. Conversely, both are almost absent in MWH (0.6%) and not even very common in MDC (22.1%)

Table 7.10 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%

To see if these two strategies combined are indeed indicative of MYN and MDQ, the BGLMM model in Equation 7.7 was tested, where `at_least_1_strat` is `true` if at least one of the two strategies occurs in the utterance.

```
`at_least_1_strat` ~ mode + gender + (1 | speaker)
    + (1 | prompt)
```
(7.7)

An LRT of the model indicates a significant effect of mode but not of gender, $\chi^2(3) = 339$, *p.adj* <.001 and $\chi^2(1) = 0.01$, *p.adj* = .932 respectively. The model has a marginal R^2 of .57 and a condition R^2 of .78. The marginal R^2 is exceptionally high considering that gender is not statistically significant, and the model conflates two separate response parameters.

In the combined model, as implied by the LRT, we see that gender has essentially no effect. There is only a very slight difference in the estimated effects of gender on the likelihood of at least one strategy being employed. As indicated in Figure 7.19B, male speakers are only 6% less likely to use at least one strategy compared to the female speakers, OR = 0.94, CIs [0.19, 4.65], $p = .939$.

The estimated probability of at least one strategy (Figure 7.19A) is only .16 in MDC and 0 in MWH, 95% CIs [.03, .54] and [0, .04] respectively. However, the estimated probability is noticeably higher in MYN, at .59 [.19, .9], and extremely high in MDQ, at .87 [.51, .98], with even the lower limit of the 95% CI above .5 in MDQ. The strength of the effects is also evident in the pairwise comparisons (Figure 7.19B). If we set aside the difference between MWH and the other sentence modes (given that it almost always has at least two PAs because of the phrase-initial question word), we see that both MYN and MDQ are noticeably more likely to be associated with at least one of the two strategies than MDC. That is, 7.9 [4.39, 14.2] times more likely in MYN and 36 [17.8, 71] times more likely in MDQ, $p < .001$ in each case. MDQ is also 4.5 [2.48, 8.2] times more likely than MYN to be associated with one of the two strategies, $p < .001$.

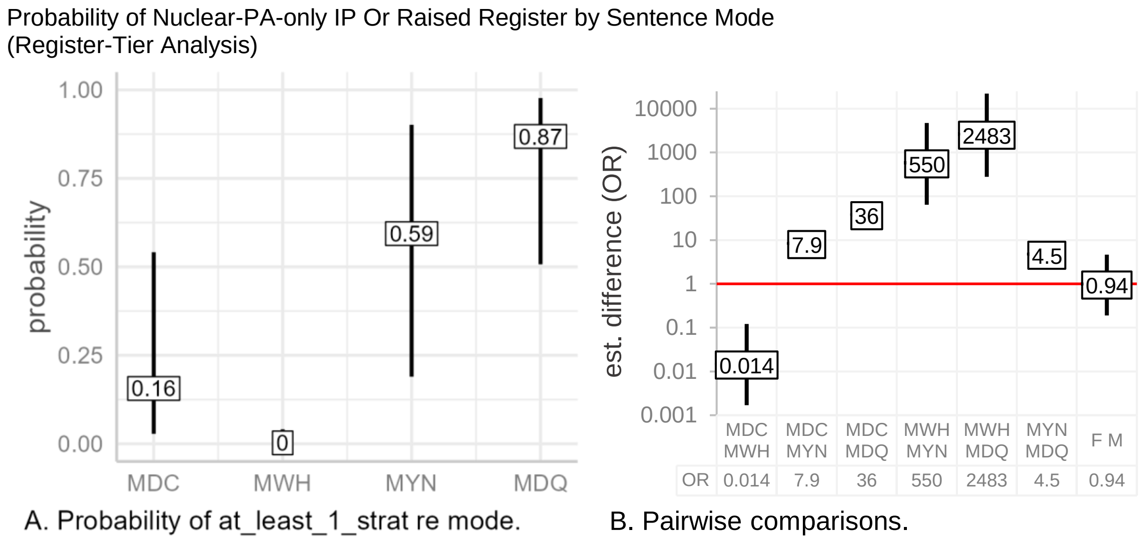


Figure 7.19 Graphical summary of model testing likelihood of nuclear-PA-only ***or*** *raised register in the nuclear. Panel A shows predicted probability by gender while Panel B shows pairwise comparisons of likelihood of across levels of* `mode` *and* `gender`*. (M-corpus data with register-tier-analysis labelling.)*

7.5.6 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 but 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 strong likelihood of high register in MDQ for male speakers (85%). The register-tier analysis demonstrated superior explanatory value for phonological variation across modes than the non-register-tier analysis (r_m^2 = .63 versus .33).

In the utterance-wide analysis, there were a large number of nuclear-PA-only IPs, which was unexpected. An analysis of the phenomenon in relation to sentence mode suggests that it is a strategy to help make the rise of the nuclear pitch accent more salient, thus reinforcing its function as a question. It was also more common among female than male speakers.

It appears, therefore, that speakers (in this study at least) have two phonological strategies available to help reinforce interrogativity in YNQs and DCQs, namely, register raising in the pitch accent or pitch-accent avoidance before the nuclear PA, with the former preferred by male speakers and the latter by female speakers. To assess the extent to which at least one of these two strategies is used in distinguishing between modes, they were combined into a single parameter (`at_least_1_strat`). It was found that the presence of at least one of the two strategies was associated with a predicted 16% probability of MDC and an 87% probability of MDQ. The effect of gender also disappeared once the two strategies were combined. Thus, in a register-tier analysis of the intonational phonology, we can see that speakers are likely to employ one of two phonological strategies to distinguish between the two modes (MDC and MDQ) which are otherwise syntactically and semantically identical. Moreover, the analysis suggests that male and female speakers are likely to adopt different strategies.

7.6 Phonetic Analysis and Results

Four phonetic parameters were analysed. These were the f_0 and temporal alignment of the f_0 minimum in the nuclear PA and the f_0 and temporal alignment of the f_0 maximum, as summarised in Table 7.11. The combined alignment and f_0 parameters were treated as the phonetic implementation of tonal targets.

Table 7.11 Phonetic parameters used in analysis of nuclear accent phonology.

Parameter type	Parameter code	Parameter description
f_0 parameters (ST re speaker median f_0)	`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 was the utterance-mean f_0 (`utt_mean_f0`) measured in ST re speaker median f_0[24]. The second was the slope of the f_0 contour (`utt_slope`), measured as 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/s.

Because the focus here is on paralinguistic and phonological changes to the rise, only utterances with L*H-like PAs are included. 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 if there is any meaningful—as opposed to purely formal—phonological distinction between the unmarked L*H and ^[L]*H. However, as it was analysed as an L*H-like pitch accent, it was retained.

7.6.1 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 sentence mode (`mode`) as the sole fixed factor affecting the scaling and timing of tonal targets. This type of model represents the effect of

[24] As before, ST re speaker median is abbreviated to ST for the sake of convenience. (See Section 6.3, p. 62.)

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 second type is a mode-plus-phonology model, which treats both pitch accent phonology (`acc_phon`) and `mode` as fixed factors, i.e., it assumes that there are phonological processes which affect the alignment and scaling of f_0, specifically via the register tier. If the register-tier hypothesis is valid, we should expect to see the apparent paralinguistic effects of `mode` in the mode-only models largely disappear in the mode-plus-phonology models. Finally, while it is not the main focus of the study, `gender` was also included as a fixed factor.

For each type of model, per-speaker random slopes and intercepts of `mode` (and `acc_phon` in the mode-plus-phonology model) were included in the ideal maximal model. Boundary tone phonology (`fin_phon`) was included as a random intercept. This is because we know from Chapter 6 that boundary tone affects the alignment and scaling of tonal targets; however, these effects are not of direct interest in this chapter, so it was not treated as a fixed factor. The target phrase in the final foot (`prompt`) was also treated as a random intercept, given that each phrase is essentially one of an infinite range of potential phrases. However, as before, due to convergence and singularity issues, the models had to be reduced to intercepts only models. `prompt` was also removed from the f_0 models as it led to convergence issues. The final working models are those shown in Table 7.12. Note that for each comparable response parameter, the only difference between models is the absence or presence of `acc_phon`. This ensures that the results of each model can be compared fairly across similar response parameters in the two different types of model.

Table 7.12 optimal working models for LMEM analysis of nuclear PA tonal targets.

<table>
<tr><th></th><th>fixed factors</th><th>random factors</th></tr>
<tr><td rowspan="2">l_t ~</td><td>mode + gender</td><td rowspan="4">+ (1 | speaker) + (1 | fin_phon) + (1 | prompt)</td></tr>
<tr><td>mode + acc_phon + gender</td></tr>
<tr><td rowspan="2">h_t ~</td><td>mode + gender</td></tr>
<tr><td>mode + acc_phon + gender</td></tr>
<tr><td rowspan="2">l_f0 ~</td><td>mode + gender</td><td rowspan="4">+ (1 | speaker) + (1 | fin_phon)</td></tr>
<tr><td>mode + acc_phon + gender</td></tr>
<tr><td rowspan="2">h_f0 ~</td><td>mode + gender</td></tr>
<tr><td>mode + acc_phon + gender</td></tr>
</table>

7.6.1.1 Mode-only Models. ANOVAs of each mode-only model indicate a significant effect of `mode`, i.e., $F(3, 608) = 43.71$, $p.adj < .001$ in the `l_t` model, $F(3, 613) = 98.8$, $p.adj < .001$ in the `l_f0` model, $F(3, 614) = 12.03$, $p.adj < .001$ in the `h_t` model, and $F(3, 617) = 140$, $p.adj < .001$ in the `h_f0` models. The ANOVAs indicate a significant effect of `gender` only in the temporal alignment models, i.e., $F(1, 8.97) = 36.1$, $p.adj < .001$ in the `l_t` model, and $F(1, 8.99) = 14.14$, $p.adj = .006$ in the `h_t` model, but $F(1, 9) = 0.03$, $p.adj = .859$ in the `l_f0` model, and $F(1, 9.01) = 0.65$, $p.adj = .442$ in the `h_f0` model. The effects of gender on the nuclear tonal targets mirror those found in Chapter 6 (see Sections 6.6.3.1 and 6.6.3.2).

Table 7.13 presents the R^2 and omega-squared partials (ω_p^2) for each mode-only model. As can be seen from the conditional R^2 values, the temporal alignment models explain more of the variance in the response parameters `l_t` and `h_t` (R_c^2 = .76 and .85) than the f_0 scaling models do for `l_f0` and `h_f0` (R_c^2 = .63 and .85). However, when we consider the marginal R^2, it is only in the `l_t` model that the fixed effects explain a large proportion of the variance in the response (R_m^2 = .57). The marginal R^2 of the other models ranges from .19 in the `l_t` model to .25 in `h_f0` model. In terms of the omega-squared partials, gender explains a sizeable proportion of the variance in both alignment models, ω_p^2 = .76 in the `l_t` model and ω_p^2 = .54 in the `h_t` model. `mode` explains less of the variance in the alignment models than it does in the f_0 scaling models, i.e., ω_p^2 = .17 and .05 in the `l_t` and `h_t` models respectively but ω_p^2 = .32 and .4 in the `l_f0` and `h_f0` models in turn. This is unsurprising, as we expect changes in mode to be associated most strongly with variation in f_0 scaling (i.e., as a paralinguistic effect) not with changes in tonal alignment.

Table 7.13 Effect size parameters for mode-only models. (Note, negative ω_p^2 are reported here as zero but reported fully in Appendix L1.)

		R^2		Omega-squared partials (ω_p^2)			
	parameter	**marginal**	**conditional**	**mode**	**95% CI**	**gender**	**95% CI**
L Target	l_t	.57	.76	.17	[.12, .22]	.76	[.34, .89]
	l_f0	.2	.63	.32	[.26, .38]	0	[0, 0]
H Target	h_t	.19	.85	.05	[.02, .09]	.54	[.06, .78]
	h_f0	.25	.68	.4	[.35, .45]	0	[0, 0]

As gender effects are not the central focus of this study and have already been covered in Chapter 6, they will not be discussed further here; however, a full breakdown of all output from the mode-only models—including gender effects—can be found in Appendix L1.

Figure 7.20 summarises the estimated means of the tonal target alignment parameters as an effect of `mode`. There is almost no difference in either the low or high targets across MDC, MWH, and MYN. That is, `l_t` estimates range between only 88 ms in MWN and 91 ms in MYN, 95% CIs [73, 103] and [76, 106], while `h_t` ranges from 292 [232, 252] ms in MYN to 294 [342, 354] ms in both MDC and MWH. In contrast, the low and high tonal targets in MDQ are noticeably earlier, with an estimated `l_t` of 69 [55, 84] ms and an estimated `h_t` of 277 [217, 338] ms. This means that the only significant differences are between MDQ and the other sentence modes. `l_t` for MDQ is aligned an estimated 21 [-25, -17], 21 [-26, -17], and 19 [-23, -15] ms earlier than MDC, MWH, and MYN respectively, $p < .001$ in each case. Similarly, the MDQ `h_t` estimate is 17 [-23, -10], 16 [-23, -10], and 14 [-20, -8] ms earlier than each in turn, again $p < .001$.

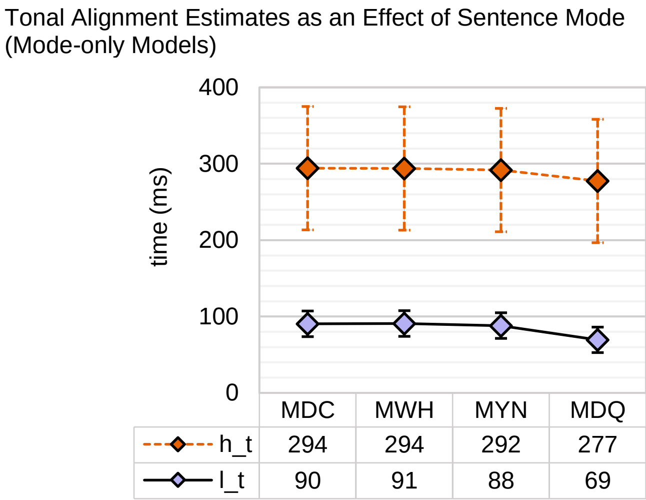
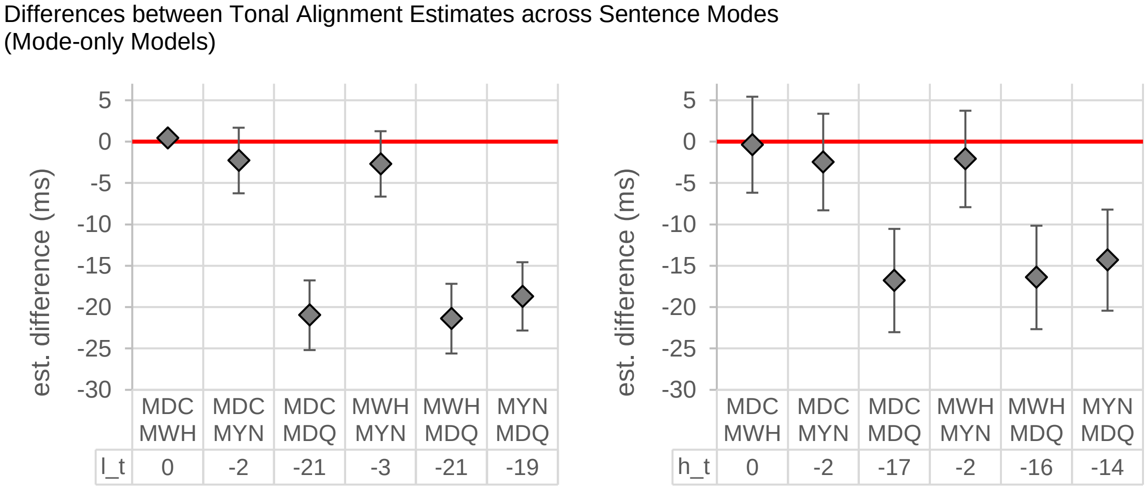


Figure 7.20 Estimated tonal alignment parameters as an effect of `mode` *in mode-only models (M-Corpus). Error bars indicate 95% CIs.*

These differences can be seen clearly in Figure 7.21, which visualises the estimated differences in effects of mode on `l_t` and `h_t` (Panels A and B respectively). As with the previous plots of pairwise comparisons, the marker indicates the estimated mean deviation of the second sentence mode listed from the first, while the vertical lines indicate 95% CIs. In cases where the upper or lower limit of the CI crosses through zero (i.e., no estimated difference), we cannot reject the null hypothesis. That said, the estimated differences in tonal alignment in the non-significant comparisons are miniscule anyway, ranging between zero and three milliseconds. (Full reports of pairwise comparisons can be found in Appendix L1.)



A. Comparison of effects of `mode` on `l_t`. B. Comparison of effects of `mode` on `h_t`.

Figure 7.21 Pairwise comparisons of estimated difference between effects of mode on tonal alignment parameters in mode-only models. First term in each pair is the intercept. Error bars indicate 95% CIs.

The results for the f_0 parameters are quite different (Figure 7.22). There is barely any difference between the `l_f0` estimates of MDC and MWH, with a very slight rise from -2.3 ST in MDC to -2.1 ST in MWH, 95% CIs [-3.9, -0.6] and [-3.8, -0.5] respectively. This is followed by a jump to -0.6 [-2.3, 1] ST in MYN, and another rise again to 0.2 [-1.4, 1.9] ST in MDQ. This pattern is reflected in the h_f0 estimates, although the changes in f_0 become increasingly large from MDC to MWH to MYN and finally to MDQ, with estimates of 3.2 [0.8, 5.7], 3.6 [1.2, 6.1], 5 [2.6, 7.4], and 7.8 [5.4, 10.2] respectively.

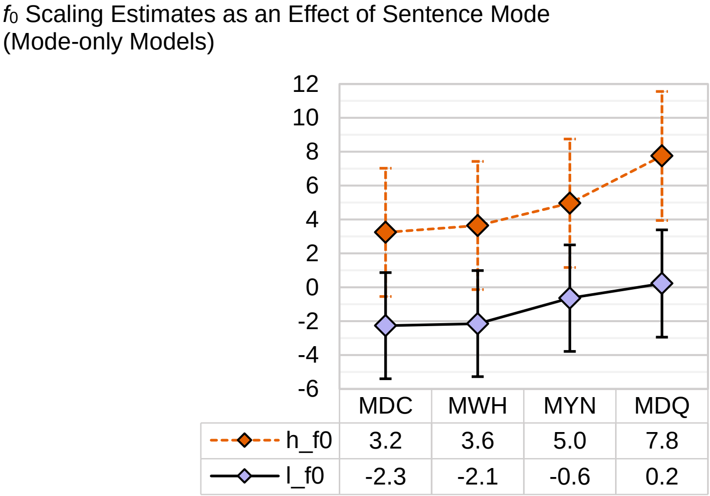
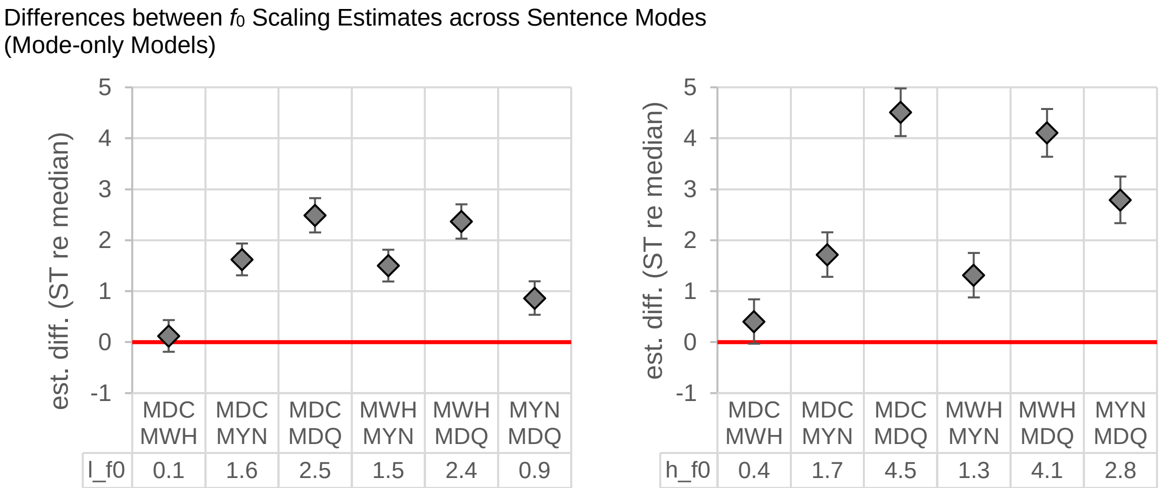


Figure 7.22 Estimated f_0 scaling parameters as an effect of mode in mode-only models (M-Corpus). Error bars indicate 95% CIs.

In terms of the pairwise comparisons the effects of levels of mode on f_0 scaling parameters, the only non-significant difference for both l_f0 and h_f0 is between MDC and MWH, at 0.1 ST and 0.4 ST respectively, 95% CIs [-0.2, 0.4] and [0, 0.8], p = .442 and .068 in turn. The greatest differences in l_f0 estimates are found in those between MDQ and both MDC and MWH, with MDQ being an estimated 2.5 [2.2, 2.8] ST higher than MDC and 2.4 [2, 2.7] ST higher than MWH, p < .001 in each case. MYN l_f0 estimates follow a similar pattern, although the differences are not as large, as can be seen in Figure 7.23A. As with the l_f0 estimates, the greatest difference in h_f0 as an effect of mode is between MDQ and both MDC and MWH, where MDQ is an estimated 4.5 [4, 5] and 4.1 [3.6, 4.6] ST higher than each in turn, p < .001. The estimated relative increase in h_f0 as an effect of MYN when compared to MDC and MWH, however, is much lower, at 1.7 [1.3, 2.2] and 1.3 [0.9, 1.8] ST in turn, p < .001. These differences are more or less the same as those in l_f0. This suggests that it is more important for the H target to be scaled higher in MDQ than in the other sentence modes.



A. Comparison of effects of `mode` on `l_f0`. B. Comparison of effects of `mode` on `h_f0`.

Figure 7.23 Estimated differences in effect of sentence modes on f_0 scaling of tonal targets in mode-only models. First term in each pair is the intercept. Error bars indicate 95% CIs.

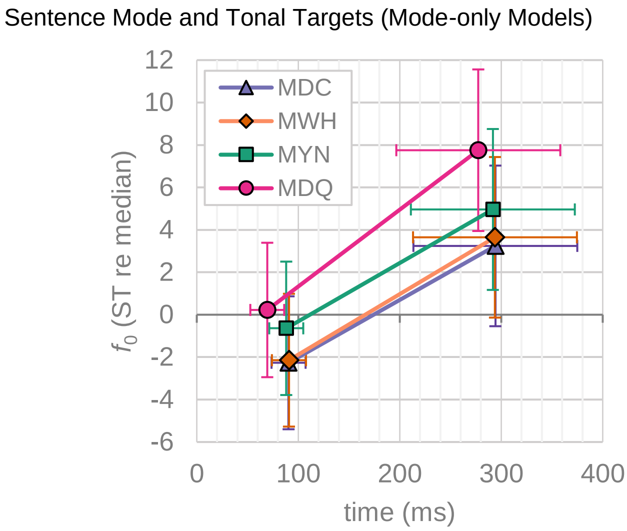


Figure 7.24 Estimated means of tonal target parameters as an effect of `mode` in mode-only LMEMs. Error bars indicate 95% CIs.

All told, the mode-only model analysis gives the impression that there is indeed a difference in f_0 scaling from MWH/MDC to MYN to MDQ, much in the vein of Haan's work on Dutch (2002). In fact, there is almost no difference between MDC and MWH scaling, while the largest change is found in the H target of MDQ. As such, the mode-only models support the view that there is a gradient increase in f_0 across sentence modes as the number of syntactic and lexical signals of interrogativity decreases. This is displayed very clearly in the visualisation of the tonal targets in Figure 7.27. Here, we see that the estimated means of MDC and MWH are practically identical, but that MYN and MDQ have increasingly higher f_0 scaling in both the low and high targets, with MDQ being scaled noticeably higher than all the others.

It was suggested to me that these data might indicate a bimodal distribution. To check if this was the case, the distribution of nuclear f_0 peaks was plotted with f_0 calculated as z-scores based on individual speaker means. However, because this is a less refined analytical instrument than a linear-mixed model and does not compensate for missing data points, some data had to be culled to ensure that the data was balanced. For this reason, the plot does not incorporate data from the *valley* prompts (see Table 7.2) and excludes M9 altogether. The resulting plot, shown in Figure 7.25, does not quite have a bimodal distribution. It has a peak at around one standard deviation (+1 SD) above the mean, and there is a small bump in the curve (but not a peak) at just above +3 SDs. MDQ accounts for the vast bulk of the bump along with MYN also contributing slightly. Of course, each sentence mode represents only 25% of the data, so even if MDQ were responsible for a bimodal distribution, it would not generate a peak as large as the other modes combined. Thus, while we do not see a bimodal distribution in this more modest dataset, we do see that the distribution is skewed slightly to the right as an effect of MDQ and, to a much lesser extent, MYN.

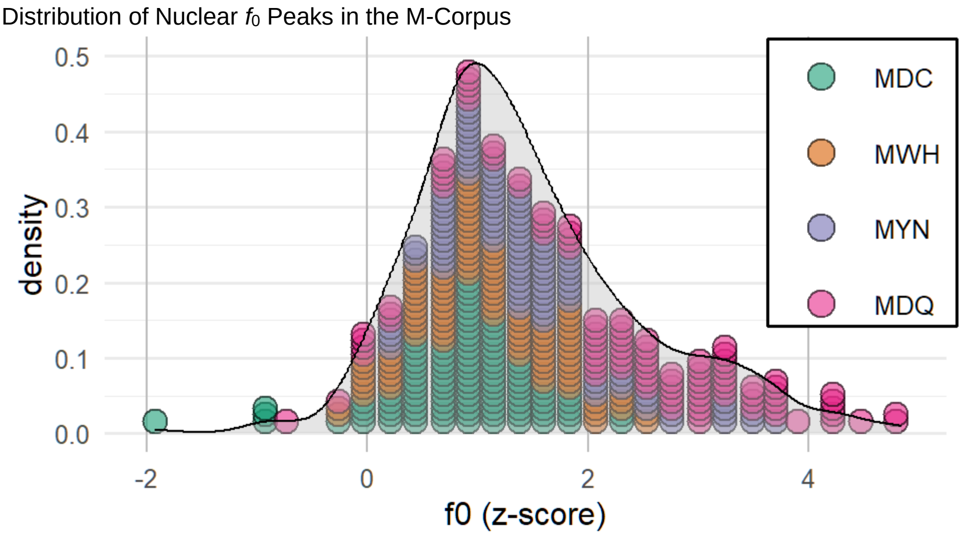


Figure 7.25 Dot plot and distribution of nuclear peak f_0 (z-scored to speaker mean) from the M-Corpus for "vases" and "valuables" prompts but excluding M9 data.

7.6.1.2 Mode-plus-phonology Models. ANOVAs of the mode-plus-phonology models indicate a significant effect of both `mode` and `acc_phon` on all tonal target parameters while, once again, the effect of `gender` is significant only in relation to temporal alignment. The results of the ANOVAs are summarised in Table 7.14. Full results from all mode-plus-phonology models can be found in Appendix L2.

Table 7.14 Summary of ANOVAs of mode-plus-phonology models used in analysis of tonal target parameters of nuclear pitch accents in M-Corpus.

	mode			acc_phon			gender		
model	**df**	**F**	**p.adj**	**df**	**F**	**p.adj**	**df**	**F**	**p.adj**
`l_t`	(3, 610)	18.4	< .001	(3, 613)	4.67	.004	(1, 9)	35.2	< .001
`l_f0`	(3, 6.8)	57.4	< .001	(3, 612)	57.4	< .001	(1, 9)	1.19	.359
`h_t`	(3, 612)	11.8	.001	(3, 614)	11.8	< .001	(1, 9)	13.3	.007
`h_f0`	(3, 615)	37.8	< .001	(3, 617)	56.3	< .001	(1, 9)	0.15	.737

Table 7.15 presents the R^2 and omega-squared partials (ω_p^2) for each mode-plus-phonology model. The arrows on the right of each parameter indicate the direction of change from the mode-only model while the values in grey indicate the magnitude of the change. With the exception of the conditional R^2 of the `l_t` model, each mode-plus-phonology model has a greater marginal and conditional R^2 than that of the mode-only model. This indicates that, with the addition of pitch accent phonology (`acc_phon`), each model—except for `l_t`—explains a greater degree of variance in the response parameter.

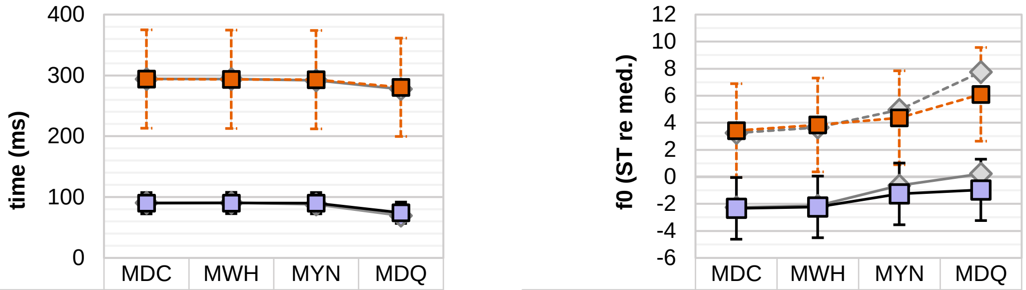
Table 7.15 Effect size parameters for mode-plus-phonology models. Arrows and numbers in grey text indicate change from mode-only models. (Negative ω_p^2 are reported here as zero but reported fully in Appendix L2.)

	R^2		Omega-squared partials (ω_p^2)				
param.	**marginal**	**conditional**	**mode [95% CI]**		**gender [95% CI]**		**acc_phon [95% CI]**
`l_t`	.58 ▲.01	.77 ▼.01	.08 [.04, .12]	▼.09	.76 [.34, .89]		.02 [0, .04]
`l_f0`	.33 ▲.13	.68 ▲.05	.12 [.07, .16]	▼.2	.02 [0, .36]	▲.02	.22 [.16, .27]
`h_t`	.21 ▲.02	.86 ▲.01	.02 [0, .05]	▼.03	.53 [.05, .77]	▼.01	.05 [.02, .08]
`h_f0`	.35 ▲.1	.73 ▲.05	.15 [.1, .2]	▼.25	0 [0, 0]		.4 [.16, .26]

Looking at the omega-squared partials (ω_p^2), we see that mode holds less explanatory value with the addition of `acc_phon` to the models. The magnitude of the change is greater in the f_0 scaling parameters, with ω_p^2 .2 lower in the `l_f0` model and .25 lower in the `h_f0` model, ω_p^2 = .12 and .15, 95% CIs [.07, .16] and [.1, .2] respectively. The bulk of the reduction in the effect size of `mode` can be accounted for by the effect size of `acc_phon`. That is, while the effect size of `acc_phon` is quite small in both the `l_t` and `h_t` models, at .02 [0, .04] and .05 [.02, .08] respectively, it is noticeably larger in `l_f0` and `h_f0`, at .22 [.16, .27] and .4 [.16, .26] each. This provides the first piece of evidence that apparent paralinguistic effects of sentence mode on f_0 scaling can (in part) be explained by phonological effects. Unlike that of mode, there is little to no estimated change in the effect size of `gender` in any of the models.

Figure 7.26 summarises the estimated means of the tonal target parameters as an effect of mode in the mode-plus-phonology models. The estimates from the mode-only models are shown in grey for comparative purposes. From Panel A, it is clear that there is little to no meaningful change from the mode-only models in the estimated alignment of targets, including no change in the MDC estimates at all. The only change in the MWH estimates is that `l_t` is an estimated one millisecond earlier than in the mode-only model, at 90 ms, CI [76, 105]. In MYN, both `l_t` and `h_t` are estimated to be ever so slightly later (by 2 and 1 ms respectively) at 90 [75, 105] and 293 [234, 352] ms in turn. The largest difference in temporal alignment between the two types of model is for MDQ, with `l_t` an estimated 5 ms later at 74 [60, 89] ms and `h_t` an estimated 8 ms later at 280 [221, 340] ms. However, even this is a very small change.

Tonal Alignment and f0 Scaling Estimates as an Effect of Sentence Mode
(Mode-and-phonology Models)



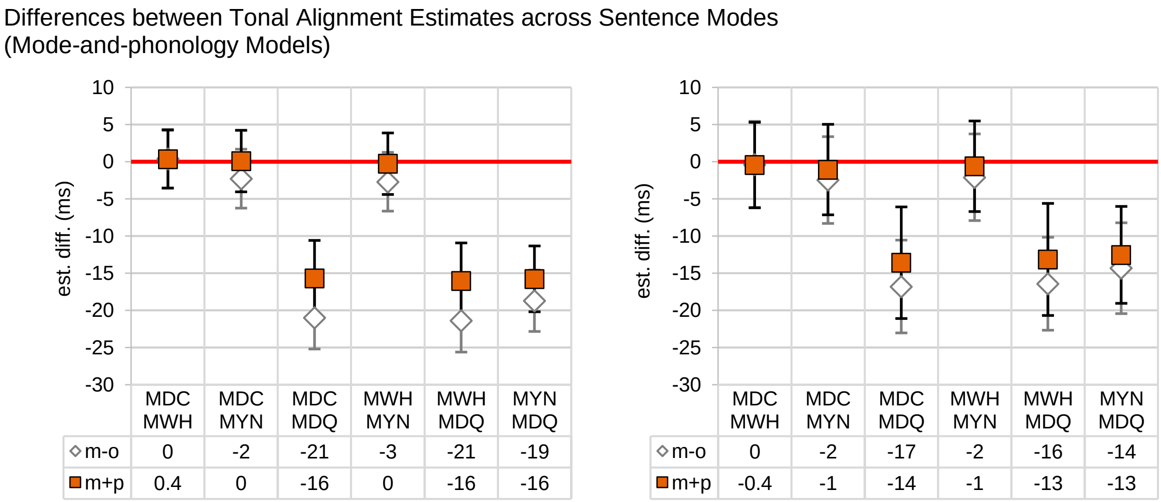
	MDC	MWH	MYN	MDQ
H (m-o)	294	294	292	277
L (m-o)	90	91	88	69
H (m+p)	294	294	293	280
L (m+p)	90	90	90	74

	MDC	MWH	MYN	MDQ
H (m-o)	3.2	3.6	5.0	7.8
L (m-o)	-2.3	-2.1	-0.6	0.2
H (m+p)	3.4	3.8	4.4	6.1
L (m+p)	-2.3	-2.2	-1.3	-1.0

A. Temporal alignment parameter estimates. B. f_0 scaling parameter estimates.

Figure 7.26 Estimated means of tonal target parameters as an effect of mode in mode-plus-phonology (m+p) models (M-Corpus). Mode-only models are shown in grey (m-o). Error bars indicate 95% CIs.

What is clear in these models, just as in the mode-only models, is that the only noticeable change in temporal alignment is in MDQ, where both the L target and the H target are aligned slightly earlier. This is made clear in Figure 7.27. The two panels here show the estimated differences in the effect of mode on temporal alignment of the low and high targets in the mode-plus-phonology models (with the mode-only model backgrounded in grey for comparison). In each of these panels, we see that there is essentially no difference in the alignment of tonal targets as an effect of MDC, MWH, or MYN, but that MDQ is associated with a significantly earlier alignment of each target.



A. Comparison of effects of mode on l_t. B. Comparison of effects of mode on h_t.

Figure 7.27 Estimated differences in effect of sentence modes on temporal alignment of tonal targets in mode-plus-phonology (m+p) models. Estimates for the mode-only (m-o) models are shown in grey. First term in each pair is the intercept. Error bars indicate 95% CIs.

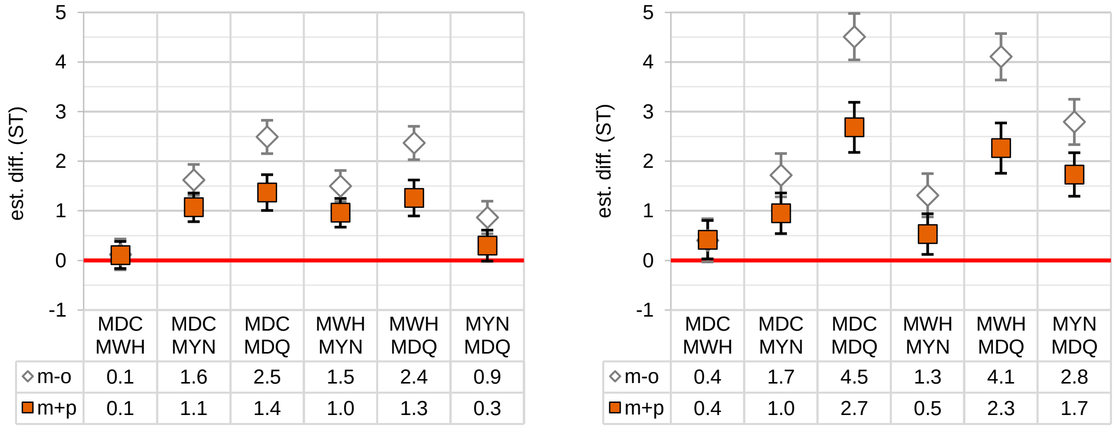
Turning to f_0 scaling in the mode-plus-phonology models (Figure 7.26B), we the see same general trend that was found in the mode-only model, namely that there is little difference between l_f0 and h_f0 in MDC or MWH, but that estimated f_0 scaling is higher in MYN and highest in MDQ. However, it is also clear that increases in f_0 scaling associated with MYN and MDQ are much less pronounced than in the mode-only model. For example, estimated l_f0 in MDC is -2.3 ST in both the mode-only and mode-plus-phonology models, 95% CIs [-3.9, -0.6] and [-3.7, -0.9] respectively for MDC effects in each model. When it comes to MDQ, on the other hand, l_f0 is an estimated -1 [-2.4, 0.4] ST in the mode-plus-phonology model but 0.2 [-1.4, 1.9] ST in the mode-only mode, i.e., there is an estimated difference of 2.5 [2.2, 2.8] ST between MDC and MDQ in the mode-only model, but the difference in the mode-plus-phonology model is 1.1 ST lower at an estimated 1.4 [1, 1.7] ST, $p < .001$ in each case.

The dampening of the effect of mode is even more pronounced in the h_f0 mode-plus-phonology model. Here, h_f0 is an estimated 3.4 ST in MDC but 6.1 ST in MDQ, 95% CIs [1.2, 5.6] and [3.9, 8.3]. This represents a difference of 2.7 [2.2, 3.2] ST, $p < .001$. In the mode-only model, however, h_f0 was an estimated 3.2 [0.8, 5.7] ST in MDC and 7.8 [5.4, 10.2] ST in MDQ, representing a change of 4.5 [4, 5] ST, $p < .001$. In other words, when register effects are incorporated into the model via the inclusion of acc_phon, the estimated mean difference in the effect of mode on h_f0 between MDC and MDQ drops by 1.8 ST, from 4.5 to 2.7 ST.

In the pairwise comparison of tonal alignment parameters shown Figure 7.27 above, we can see how the addition of acc_phon to the model has led to a dampening of the differences between levels of mode in the mode-plus-phonology models when compared to the mode-only model. That is, for each comparison, the difference in the mode-plus-phonology model estimates (orange squares) is closer to zero (the red line) than in the mode-only model estimates (grey diamonds). This pattern is also reflected in the Figure 7.28, which shows the pairwise comparison of effects of mode on f_0

scaling parameters. However, we see that for `h_f0` (Panel B), the dampening of the differences in f_0 peak between MDQ and other levels of `mode` alone is very distinct. That is, in MDC-MDQ, MWH-MDQ, and MYN-MDQ (columns three, five, and six), the difference between the mode-only estimates (grey diamonds) and the mode-plus-phonology estimates (orange boxes) are larger than elsewhere, with the mode-plus-phonology estimates being much closer to zero than their mode-only counterparts.

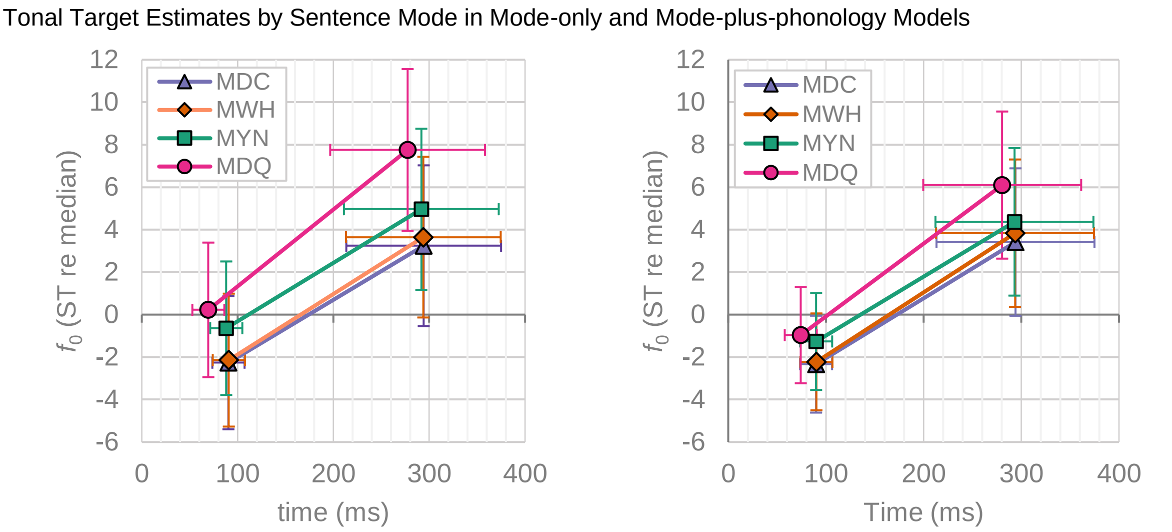
Differences between f_0 Scaling Estimates across Sentence Modes
(Mode-and-phonology Models)



A. Comparison of effects of `mode` on `l_f0`. B. Comparison of effects of `mode` on `h_f0`.

Figure 7.28 Estimated differences in effect of sentence modes on f_0 scaling of tonal targets in mode-plus-phonology (m+p) models. Estimates for the mode-only (m-o) models are shown in grey. First term in each pair is the intercept. Error bars indicate 95% CIs.

7.6.1.3 Summary. The overall effect of modelling the low and high tonal targets using a mode-plus-phonology model is that we see the estimates for mode coming considerably closer together compared to the mode-only model. The collective effect of this is indicated very clearly in Figure 7.29. Here, Panel A shows the tonal targets plotted as an effect of `mode` alone in the mode-only models, while Panel B shows the same plots, but with estimated tonal targets from the mode-plus-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 most important point 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. Mode-only tonal target estimates. B. Mode-plus-phonology tonal target estimates.

Figure 7.29 Comparing the estimated means nuclear tonal target parameters as an effect of mode in the mode-only and mode-plus-phonology LMEMs. Error bars indicate 95% CIs.

A simple but effective way of quantifying the extent to which mode-plus-phonology models reduce the apparent effects of mode on f_0 scaling is to compare the average differences in scaling across levels of mode for each f_0 parameters. That is, we can calculate the means of the pairwise comparisons of the effects of mode on l_f0 and h_f0 (the values summarised in Figure 7.23 and Figure 7.28) in each type of model. The mean and standard deviations for these calculations are shown in Table 7.16. The results show that, once pitch accent phonology has been taken into account (i.e., PA plus register tier), the apparent mean differences in f_0 as an effect of mode alone are much lower. That is, the average estimated difference between effects of mode on f_0 falls by 0.7 ST (*SD* = 0.4) in l_f0 and by 1 ST (SD – 0.7) in h_f0. This mirrors the standardised omega-squared partial estimates of effect size, in which the effect size of mode on the f_0 scaling parameters was lower in each of the mode-plus-phonology models than in the mode-only models (see Table 7.15 above).

Table 7.16 Mean and standard deviation (SD) of pairwise comparisons of effects of mode on f_0 parameters in nuclear PAs (in ST). Estimates are from mode-only and mode-plus-phonology LMEMs of M-Corpus data.

param.	**mode-only model**		**mode-plus-phonology**		**Change**	
	mean	**SD**	**mean**	**SD**	**mean**	**SD**
l_f0	1.5	0.9	0.8	0.5	▼0.7	▼0.4
h_f0	2.5	1.6	1.4	0.9	▼1.0	▼0.7

These results do still indicate that the paralinguistic effect of sentence mode on f_0 scaling exists in both the mode-only model and in the mode-plus-phonology models. However, the apparent effect is quantifiably weaker once the pitch accents and register tier have been incorporated into the model.

Again, as with the phonological analysis, the comparison of mode-only and mode-plus-phonology model analyses does not on its own demonstrate the existence of the register tier.

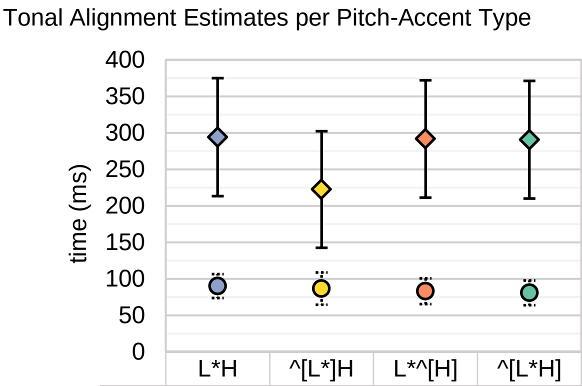
However, it does demonstrate how a register-tier analysis helps separate apparent paralinguistic effects from phonological effects. It shows that apparently strong paralinguistic effects are greatly dampened when accent phonology is introduced to the model. However, paralinguistic effects do not disappear completely, as initially suggested in Section 7.1.2. They are still in operation but are greatly subdued, as indicated by the omega-squared partial effect-size estimates (Table 7.15), the stylized comparison of nuclear tonal targets by model (Figure 7.29), and mean estimated differences across sentence modes (Table 7.16).

Many varieties of English exploit a categorical distinction in nuclear PA contours to help 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%. The current analysis indicates that DCE speakers—with the exception of F16—exploit the difference between low and high register to make the same distinction.

7.6.2 Nuclear Pitch Accent Targets and their Phonetic Parameters

While the mode-plus-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 pitch accent phonology (`acc_phon`) itself. That is because we need to see if there is empirical evidence validating the differences proposed for the PA/register-tier combinations. For this purpose, an assessment was conducted of the estimated means of each tonal target parameter from the mode-plus-phonology model as an effect of `acc_phon`. As noted in Section 7.6.1.2, ANOVAs of each mode-plus-phonology model indicate a significant effect of `acc_phon`. (See also Table 7.14 above and Appendices K5 to K8.)

Figure 7.30 shows the estimated mean temporal alignment of the low and high targets, `l_t` and `h_t`. There is only a small change in `l_t` across PAs, from 81 ms in ^[L*H] (fully raised) to 90 ms in L*H (unraised), CIs [65, 96] and [75, 104]. This is the only statistically significant, albeit small, difference in L target alignment, with ^[L*H] aligned 9 [-14, -4] ms earlier than L*H, $p < .001$.

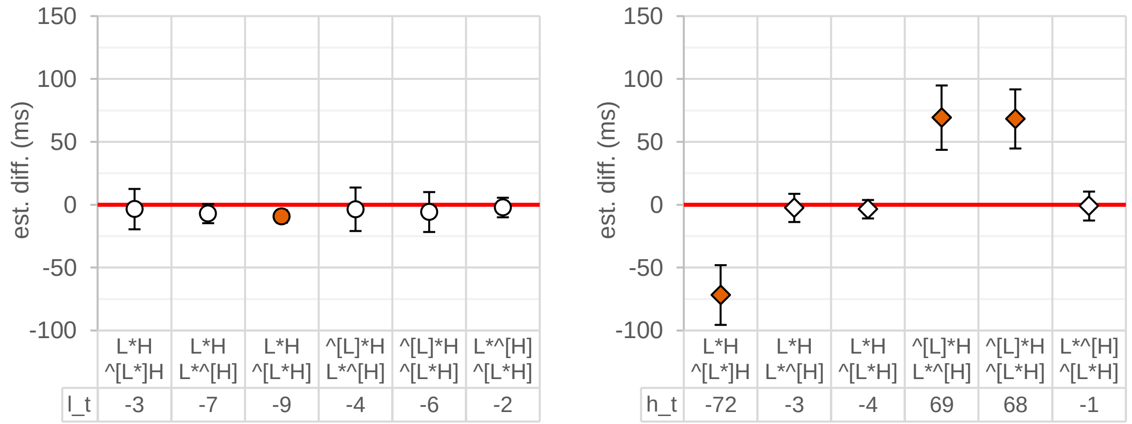


	L*H	^[L*]H	L*^[H]	^[L*H]
○ l_t	90	86	83	81
◇ h_t	294	222	292	290

Figure 7.30 Estimated tonal alignment of nuclear PA tonal targets as an effect of `acc_phon` in mode-plus-phonology LMEMs of M-Corpus data. Error bars show 95% CIs.

The overall small differences between the effects of `acc_phon` on `l_t` are illustrated clearly in Figure 7.31, Panel A. It shows the pairwise comparison of effects levels of `acc_phon` on `l_t`. The statistically significant difference between L*H and ^[L*H] is highlighted in orange, as it might be hard to identify otherwise.

Differences between Tonal Alignment Estimates across Pitch Accent Types
(Mode-and-phonology Models)



A. Comparison of `acc_phon` effects on `l_t`. B. Comparison of `acc_phon` effects on `h_t`.

Figure 7.31 Pairwise comparisons of estimated differences between effects of `acc_phon` on tonal alignment parameters in mode-plus-phonology LMEMs. First term in each pair is the intercept. Error bars indicate 95% CIs. Orange markers indicate statistically significant differences.

There are similarly small variations in H alignment across L*H, L*^[H], and ^[L*H], with estimates of 294 [235, 354], 292 [231, 352], and 290 [231, 350] ms in turn, with no significant differences between any of the estimates. However, the estimated peak alignment of H in ^[L*]H is noticeably earlier, at 222 [128, 286] ms, which is also statistically significantly different from the other estimated `h_t` values as an effect of `acc_phon`. As with `l_t`, the difference in the effect of `acc_phon` on `h_t` is visualised in Figure 7.31B, with the orange markers indicating the statistically significant differences (as the CIs do not cross through zero). ^[L*]H is aligned an estimated 72 [-96, -48] ms earlier than L*H, 79 [44, 95] ms earlier than L*^[H], and 68 [45, 92] ms earlier than ^[L*H], $p < .001$ in each case.

The estimated tonal target f_0 scaling parameters (`l_f0` and `h_f0`) are plotted in Figure 7.32, while the pairwise comparisons across levels of `acc_phon` are shown in Figure 7.33. There is little difference in L targets when they are not raised in L*H and L*^[H], with f_0 estimates of -2.3 and -2 ST respectively, 95% CIs [-3.7, -0.9] and [-3.5, -0.5], a difference of 0.3 [-0.2, 0.9] ST, $p = .221$. The difference between the raised L targets in ^[L*]H and ^[L*H] is slightly larger, with the first estimated at 0.5 [-1.2, 2.3] ST and the second at 0 [-1.5, 1.3] ST, with an estimated difference of 0.6 [-1.7, 0.5] ST, $p = .27$. Note that in neither case is the difference statistically significant. (These are identified in Figure 7.32A by the markers with white circles. The 95% CIs stretching across zero indicate that we cannot be confident that there truly is a difference between each pair.)

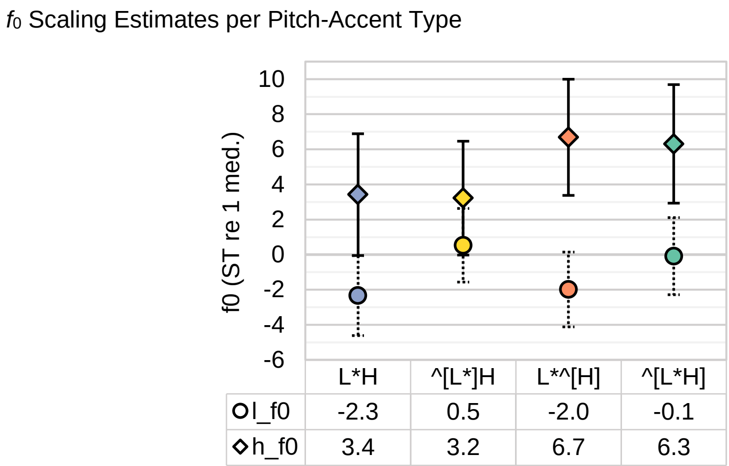
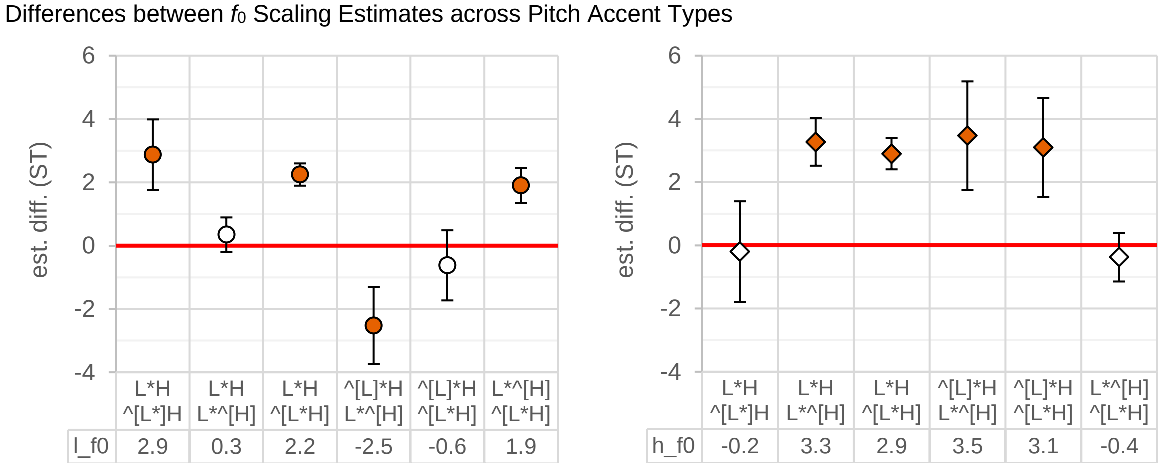


Figure 7.32 Estimated f_0 scaling of nuclear PA tonal targets as an effect of acc_phon *in mode-plus-phonology LMEMs of M-Corpus data . Error bars show 95% CIs.*



A. Comparison of `acc_phon` effects on `l_f0`. B. Comparison of `acc_phon` effects on `h_f0`.

Figure 7.33 Pairwise comparisons of estimated differences between effects of acc_phon *on f_0 tonal scaling parameters in mode-plus-phonology LMEMs. First term in each pair is the intercept. Error bars indicate 95% CIs. Orange markers indicate statistically significant differences.*

For the `h_f0` estimates, we see, once again, that the unraised targets are very similar, with H an estimated 3.4 ST in L*H and 3.2 in ^[L*]H, CIs [1.2, 5.6] and [0.5, 5.9] respectively, an estimated difference of only 0.2 [-1.8, 1.4] ST, p = .806. For the raised H tones, there is a slightly larger difference, with H estimated at 6.7 [4.4, 9] ST in L*^[H] and at 6.3 [4.1, 8.6] ST in ^[L*H], lower by 0.4 [-1.1, 0.45] ST, p = .339. Again, there is no statistically significant difference between the two pairs, as shown by the first and last markers in Figure 7.33B. There is, however, a clear difference between the raised and unraised H targets, ranging from 2.9 [2.4, 3.4] ST between L*H and ^[L*H] and 3.5 [1.8, 5.2] ST between ^[L*]H and L*^[H], p < .001 in all cases. These differences are very apparent in Figure 7.33B, as indicated by the orange diamonds.

When we plot the target parameters together on a two-dimensional plane (Figure 7.34), we see that, with the exception of the H target in ^[L*]H (orange squares), there is a great similarity between tonal targets with the same tone and register, both in scaling and alignment. We also see that ^[L*H] (green circles) is indeed simply a raised version of L*H (light blue triangles), and that L*^[H] begins at roughly the same point as L*H and rises to approximately the same peak as ^[L*H].

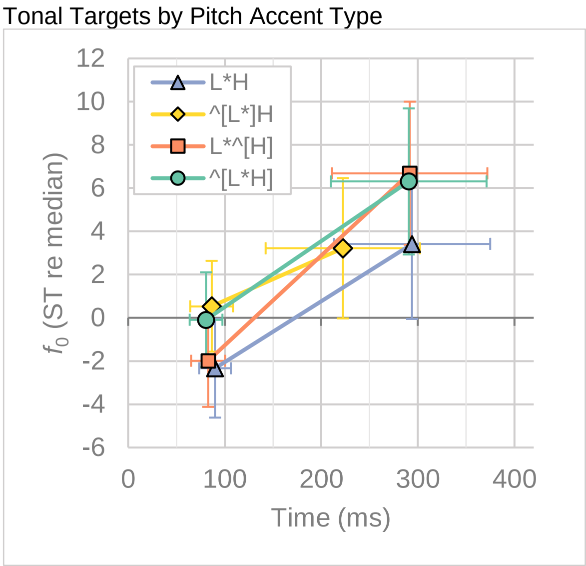


*Figure 7.34 Plot of tonal L*H-like targets of in mode-plus-phonology models. Error bars indicate 95% CIs.*

The only outlier in the tonal targets is the H target of ^[L*]H. Despite the similarity of the L target in ^[L*]H to its ^[L*H] equivalent in both alignment and scaling, the unraised H target is aligned noticeably earlier than its H counterpart in L*H. This suggests that it might be identified more appropriately as >H* (or even H*). Unfortunately, there are so few nuclear >H* and H* tokens, *n* = 2 and 5 in turn, that it is difficult to make generalisations about them, aside from observing their sparseness. In fact, there are only six nuclear ^[L*]H tokens, so the findings regarding ^[L*]H should be taken with a pinch of salt.

7.6.2.1 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]. However, the main concern in this section is f_0 scaling, and the essential question is, "Does a statistical analysis of the tonal targets in different levels of `acc_phon` support evidence for the proposed phonological distinction between the raised and non-raised permutations of L*H?" When it comes to L*H, L*^[H], and ^[L*H], the answer is clearly yes. The estimated f_0 mean differences between phonologically identical pairs of tonal targets are very small (and statistically non-significant), while those which we expect to be different are indeed consistently different.

The results for ^[L*]H are less satisfying. Here, the earlier alignment of the H target suggests that it might be more akin to >H* than L*H. Unfortunately, a viable comparison could not be made between ^[L*]H and >H*pitch accents (or even H*), given the sparsity of tokens.

7.6.3 Mode and Global Phonetic Parameters

Two utterance-wide parameters were evaluated, utterance mean f_0 (`utt_mean_f0`) and utterance-wide f_0 slope (`utt_slope`). `utt_mean_f0` is the mean of the contour measured in semitones re speaker median f_0 while `utt_slope` is the slope of the linear regression of the f_0(t) contour measured in semitones per second. In the previous analysis of tonal targets, utterance-wide parameters were estimated using two kinds of model, i.e., a mode-only model and a mode-plus-phonology model. The reason for constructing two different models was to extract and compare tonal target parameters as an effect of sentence mode along with those which estimate tonal target parameters as an effect of mode and pitch accent phonology (although gender was also included as an effect in each type of model). Here the aims are similar, so two types of model are compared again; however, there are several small changes in the equations themselves. Firstly, in the mode-plus-phonology models, `fin_phon` was treated as a fixed effect rather than a random effect since it is a component the IP phonology. A new fixed factor, `h_start`, was also added to account for phonological events early in the IP which might raise the mean f_0 or tilt the slope downwards. As such `h_start` is a logical parameter which is `true` when an utterance begins with a H boundary or has prenuclear H* or >H* PA associated in the first foot. There are also some differences in the random effects. Neither `prompt` as a random intercept nor `mode` as a random by-speaker slope caused convergence issues in any of the models, so—unlike in the previous models—they were retained in the models. The mode-only mode is shown in Equation 7.8 and the mode-plus-phonology model in Equation 7.9.

```
response ~ mode + gender
             + (1 + mode | speaker) + (1 | prompt)
```
(7.8)

```
response ~ mode + gender
             + h_start + acc_phon + fin_phon
             + (1 + mode | speaker) + (1 | prompt)
```
(7.9)

7.6.3.1 Comparison of mode-only and mode-plus-phonology models. ANOVAs of the mode-only models indicate a significant effect of `mode` on both `utt_mean_f0` and `utt_slope`, $F(3, 10) = 20.6$, $p.adj < .001$, and $F(3, 10) = 11.8$, $p.adj = .002$ respectively. ANOVAs of the mode-plus-phonology models also indicate significant effects of `mode` on `utt_mean_f0` and `utt_slope`, $F(3, 11.8) = 14.8$, $p.adj < .001$, and $F(3, 10.1) = 10.5$, $p.adj = .003$. There was a significant effect of `gender` on `utt_slope` in both the mode-only and mode-plus-phonology models, $F(1, 9.1) = 8.84$, $p.adj = .019$, and $F(1, 9.1) = 8.8$, $p.adj = .019$; however, there was no significant effects of `gender` on `mean_utt_f0` in either type of model, $F(1, 9.1) = 0.9$, $p.adj = .409$, and $F(1, 9.1) = 10.1$, $p.adj = .726$ respectively.

When it comes to phonological effects in the mode-plus-phonology models, there is a significant effect of `acc_phon` and `fin_phon` on `utt_mean_f0`, $F(5, 350) = 7$, $p.adj < .001$, and $F(1, 220.4) = 9.3$, $p.adj = .004$; however, there was no significant effect of `h_start`, $F(1, 270) = 1.15$, $p.adj = .345$. In the `utt_slope` model, there are significant effects of all three phonological parameters (i.e., `h_start`, `acc_phon`, and `fin_phon`), $F(1, 416) = 19$, $F(5, 460) = 42.7$, and $F(1, 9.9) = 32.4$ respectively, $p.adj < .001$ in each case. The fact that `h_start` is significant in the `utt_slope` model but not in the `utt_mean_f0` model, is most likely down to the fact that the presence of an IP initial (or near-initial) H tone is likely to skew the slope slightly downwards.

Effect size parameters which are shared across the mode-only and mode-plus-phonology models are summarised in Table 7.17. For the R^2 results, it is clear that the addition of phonological effects to the models increases the amount of variance they explain in the response parameters, with the fixed effects of mode-and-phonology models explaining an extra 5% of the variance in `utt_mean_f0` and an additional 7% percent of the variance in `utt_slope`. (Changes in R_m^2 are shown in the table by the arrows and numbers in grey text next to the R_m^2 values.) Similarly, the fixed and random effects combined in the mode-and-phonology models explain an additional 1% of the variance in `utt_mean_f0` and an additional 4% of the variance in `utt_slope`, as indicated by the green arrows and numbers in grey text next to the r_c^2 values in the table. Admittedly, these are small changes with the addition of the phonological effects.

Table 7.17 Effect size parameters for mode-only (m-o) and mode-plus-phonology (m+p) utterance-wide LMEMs. ω_p^2 values are shown only for fixed effects which are found in each type of model. Arrows and numbers in grey indicate change from the mode-only model. (Negative ω_p^2 values are reported here as zero but reported fully in Appendix L.)

Model		R^2		ω_p^2			
Type	**Response**	R_m^2	R_c^2	`mode` [95% CI]		`gender` [95% CI]	
m-o	`utt_mean_f0`	.25	.54	.81 [.42, .9]		.0 [0, 0]	
	`utt_slope`	.5	.83	.7 [.17, .85]		.41 [,0 .71]	
m+p	`utt_mean_f0`	.3 ▲.05	.55 ▲.01	.71 [.27, .85]	▼.10	0 [0, .1]	
	`utt_slope`	.57 ▲.07	.87 ▲.04	.67 [.11, .83]	▼.03	.72 [.11, .87]	▲.31

Similarly, the omega-squared partial statistic (ω_p^2) indicates that the apparent effect of `mode` on `utt_mean_f0` falls by .1, from .81 in the mode-only model to .71 in the mode-plus-phonology model, CIs [.42, .9] and [.27, .85]. ω_p^2 for `mode` in the `utt_slope` model falls by .03, from .7 [.17, .85] to .67 [.11, .87]. The change in the effect size of `mode` on `utt_mean_f0` is notable, but the change in its effect size on `utt_slope` is very slight. Oddly, the effect size of `gender` in the `utt_slope` model increases by .31 to .72 [.11, .87]. Unfortunately, I have no reasonable explanation as to why this is the case.

Turning to the effect sizes of the three phonological fixed effects in the mode-plus-phonology models (Table 7.18), we see that `h_start` has an ω_p^2 of 0, 95% CI [0, 0] in `utt_mean_f0`, and only .04 [.01, .09] in `utt_slope`, which indicates its effect on f_0 slope is quite small. The effect size of `fin_phon` is equally small in both the `utt_mean_f0` and `utt_slope` models, ω_p^2 = .04 [0, .1] and .04 [.01, .08] respectively. The effect size of `acc_phon` is slightly larger in the `utt_mean_f0` model, ω_p^2 = .08 [.02, .13], but is much larger in the `utt_slope` model, ω_p^2 = .31 [.24, .37]. This tells us that nuclear PA phonology has the largest effect on global f_0 parameters.

None of the ω_p^2 effect size estimates come close to the effect size of `mode`, which is considerably larger in both the `utt_mean_f0` and `utt_slope` mode-plus-phonology models (ω_p^2. = 71 [.27, .85] and .67 [.12, .83] respectively). We can infer that pitch accent phonology has a lesser effect on the global f_0 parameters and that the paralinguistic effects of sentence mode are stronger. This is quite distinct from the comparable nuclear pitch accents models in which the effect size of `acc_phon` was always larger than that of `mode` (see Section 7.6.1.2 and Table 7.15 above)

Table 7.18 ω_p^2 for fixed effects exclusive to mode-plus-phonology utterance-wide LMEMs. (Negative ω_p^2 values are reported here as zero but reported fully in Appendix L2)

Parameter	h_start	95% CI	acc_phon	[95% CI]	fin_phon	[95% CI]
utt_mean_f0	0	[0, 0]	.08	[.02, .13]	.04	[.0, .1]
utt_slope	.04	[.01, .09]	.31	[.24, .37]	.04	[.01, .08]

The third means of comparing the difference between the mode-only and mode-plus-phonology models is to calculate the means and standard deviations of the pairwise differences between each level of `mode` (see also Section 7.6.3.3 above). This is useful because it also compares the models in terms of the units in which the measurements are taken, i.e., ST and ST/s. The results of these calculations are shown in Table 7.19. We see that, once pitch accent phonology has been accounted for in the mode-plus-phonology models, the apparent effects of `mode` are lower. That is, the average estimated difference in the effect of `mode` on `utt_mean_f0` falls by 0.2 ST (*SD* = 0.11) while it falls by 0.8 ST/s in `utt_slope`. Again, these changes reflect those found in ω_p^2 (Table 7.18). The average lowering of the difference between mean f_0 values by 0.2 ST is small, as too is a change in slope of 0.9 ST/sec. Therefore, as with the analysis of ω_p^2, it is reasonable to conclude that the overall effect of mode on utterance wide parameters is not as large as those on IP parameters.

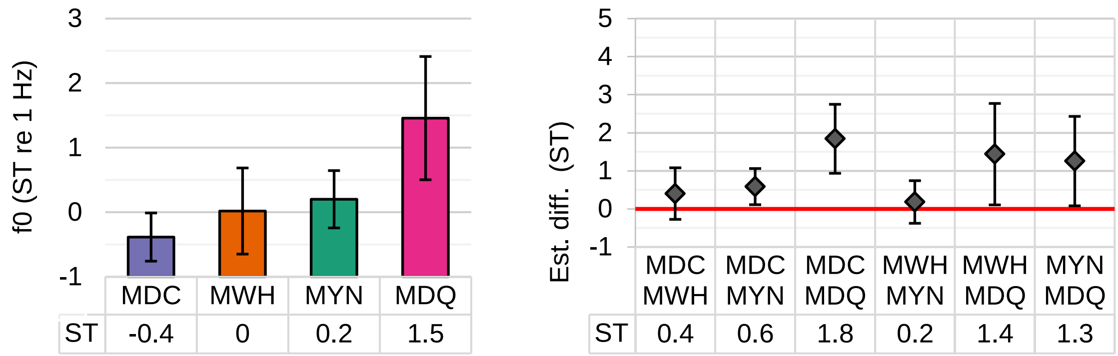
Table 7.19 Mean and standard deviation (SD) of pairwise comparisons of effects of `mode` on utterance-wide f_0 parameters Estimates are from mode-only and mode-plus-phonology LMEMs of M-Corpus data.

		mode-only model		mode-plus-phonology		Change	
response variable	unit	mean difference	SD	mean difference	SD	mean	SD
utt_mean_f0	ST	1.1	0.76	1.0	0.66	▼0.2	▼0.11
utt_slope	ST/s	4.3	3.95	3.5	3.27	▼0.8	▼0.68

7.6.3.2 Effects of mode on global parameters. The mode-plus-phonology models are more comprehensive, and we can be more confident that the effects of sentence mode which they predict on utterance-wide f_0 parameters are not in fact indirect effects of the intonational phonology. This is because phonological effects are accounted for by other fixed factors in this type of model. In other words, the effects of mode in the mode-plus-phonology models can be construed more readily as paralinguistic effects of sentence mode alone. For this reason, only the mode-plus-phonology model results are discussed in this section. (Complete tables of mode-only LMEMs of utterance-wide parameters can be found in Appendix L1.)

Panel A in Figure 7.35 shows utterance-wide mean f_0 estimates as an effect of sentence mode alone. We see that, as sentence mode moves from MDC to MWH to MYN and finally to MDQ, there is a concomitant increase in mean f_0 estimates. However, the largest rise is found in MDQ, which is considerably higher than the others. That is, MDC is associated with an estimated mean of -0.4 ST, with MYN only slightly higher at 0 ST, and MYN only slightly higher again at .02 ST, CIs [-0.7, 0], [-0.6, 0.6], and [-0.2, 0.6] respectively. MDQ, however is associated with an estimated mean f_0 of 1.5 [0.6, 2.3] ST.

Estimated Effects of Sentence Mode on Utterance-wide Mean f_0



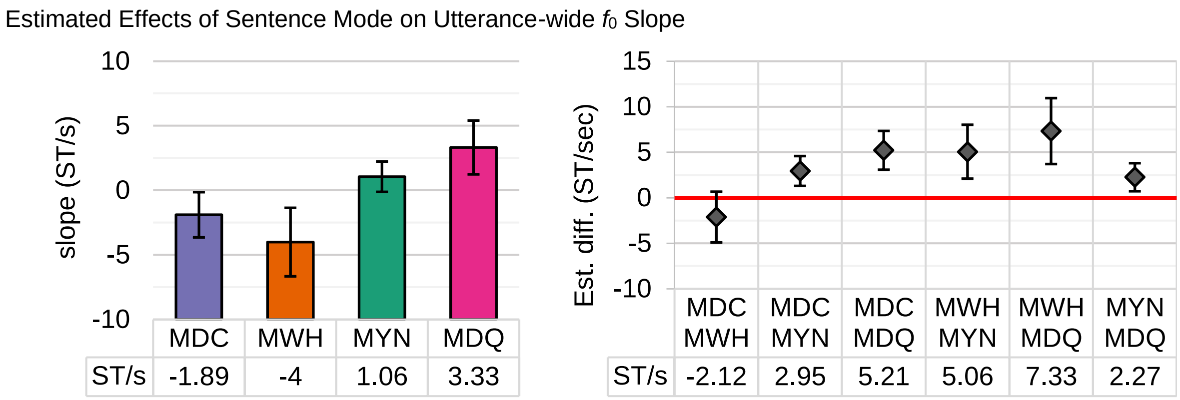
A. Estimated effects of `mode` on mean f_0. B. Comparisons of effects of `mode` on mean f_0.

Figure 7.35 Estimated effects of `mode` *on utterance-wide mean f_0 in mode-plus-phonology LMEMs of M-Corpus data. Error bars show 95% CIs.*

Panel B of the same figure shows the pairwise comparisons of estimated differences in mean utterance-wide f_0 between sentence modes. As in the previous plots of pairwise comparisons, whenever the 95% CIs cross through zero, this indicates that we cannot be confident that there is a difference in the effect of the two sentence-mode types being compared. Here we see that the slight difference 0.4 ST between MDC and MWH is not statistically significant, 95% CI [-0.3, 1.1], p = .216, nor is the slight difference between MWH and MYN, 0.2 [-0.4, 0.7] ST, p = .495. In fact, while the difference between MDC and MYN, 0.6 [0.1, 1.1] ST, p = .2, is statistically significant, the estimated mean f_0 of MYN is still not a great deal higher than that of MDC. The greatest difference between sentence-mode types is—as we saw in Panel A—the difference between MDC and MDQ, at 1.8 [0.9, 2.7] ST, $p < .001$. Given that there is the greatest amount of communicative pressure to use f_0 to distinguish between these two sentence modes (there are no lexical or syntactical markers of interrogatively), it makes sense that they are the most clearly distinguished in terms of average utterance-wide f_0 scaling.

The estimated differences in mean f_0 between MDQ and both MWH and MYN are also both statistically significant, at 1.4 and 1.3 ST, 95% CIs [0.1, 2.8] and [0.1, 2.4], p = .037 and .038 respectively. However, the 95% CIs are very wide and the lower limit of each almost reaches zero. This mostly likely reflects the fact that there is less consistency across speakers in realising MWH and MYN differently from MDC as there is less communicative pressure to do so. Therefore, only some speakers may use an elevated f_0 in MYN, thus generating a greater deal of (between-speaker) variance between MYN and MDQ.

Panel A in Figure 7.36 shows utterance-wide estimates of f_0 slope as an effect of sentence mode alone. The pattern observed for utterance-wide mean f_0, is not repeated. This is largely because of the noticeably lower estimated slope associated with MWH, at -4 ST/s, 95% CI [-6.33, -1.68], and this is probably an effect of increased f_0 excitation at the initial boundary or syllable associated with the question word. If we set aside the syntactic anomaly associated with MWH, we see that, otherwise, the same pattern is repeated as seen with utterance mean f_0. That is, f_0 slope increases from MDC to MYN to MDQ, at -1.89 [-3.46, -0.31], 1.06 [0.04, 2.08], and 3.33 [1.42, 5.23] ST/s respectively.



A. Estimated effects of mode on f_0 slope B. Comparisons of effects of mode on f_0 slope

Figure 7.36 Estimated effects of ***mode*** *on utterance-wide f_0 slope in mode-plus-phonology LMEMs of M-Corpus data. Error bars show 95% CIs.*

Panel B of the Figure 7.36 shows the pairwise comparisons of estimated differences in utterance-wide f_0 slopes across sentence modes. We see that there is a statistically significant estimated difference in f_0 slope between each pair of sentence modes with the exception of MDC and MWH, at -2.12 ST/s, 95% CI [-4.9, 0.67], $p = .122$.

7.6.3.3 Summary. The phonological effect on utterance-wide f_0 contour parameters is weaker than on the nuclear PA parameters analysed. This may indicate that paralinguistic effects tend to stretch across the whole utterance / IP and that phonological effects are concentrated in the nuclear pitch accent.

There is an overall effect of sentence mode both on utterance-wide mean f_0 and slope. The change in effect size of levels of mode on mean f_0 can be summarised in terms of the expectations for paralinguistic effects of mode that we might assume in a non-register-tier analysis, as proposed in Section 7.1.2 and summarised by equation 7.1, repeated below.

$$f_{0_{MDC}} \leq f_{0_{MWH}} < f_{0_{MYN}} < f_{0_{MDQ}}$$

This neat pattern is upset, however, in the analysis of utterance-wide f_0 slope, since there is a typically a large f_0 excursion at the start of the MWH, meaning that f_0 slope of MWH is lower or equal to that of MDC. However, the order of question form modes remains unchanged as does the relationship between MDC, MYN, and MDQ. Therefore, despite the change in order due to MWH, we might show how the general pattern is still broadly observed by bracketing MWH and MDC, as shown in Equation 7.10.

$$[\text{slope}_{MWH} \leq \text{slope}_{MDC}] < \text{slope}_{MYN} < \text{slope}_{MDQ} \quad (7.10)$$

The utterance-wide mean f_0 and slope estimates as an effect of mode are plotted in Figure 7.37 using the mode-plus-phonology models. This facilitates several summative observations. Firstly, MDQ is clearly separated from all the other sentence modes in terms of both mean average f_0 scaling and slope. Secondly, MDQ and— to a lesser extent—MYN are associated with a global f_0 rising trend and is also realised with a mean f_0 above the speaker median f_0. Finally, we see that MDC is realised with mean f_0 below speaker median and with an overall utterance wide f_0 falling trend. This downward trend in the f_0 slope—i.e., the linear regression of f_0(t)—cannot be put down to a change in the direction of the nuclear pitch contour (i.e., L*H % and L*H H% versus H* L% and H*L %), since there are only seven non-L*H nuclear PAs in the M-corpus (N = 639), or possibly twelve, if we recategorize ^[L*]H as H* or >H*, and these are not even always associated with MDC.

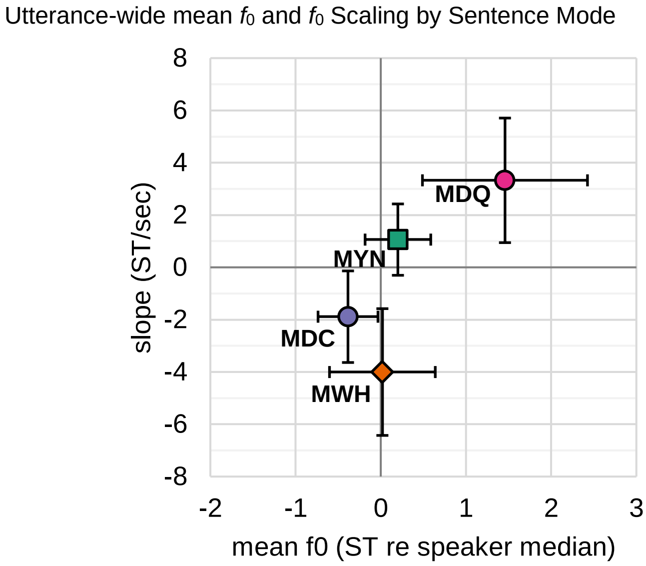


Figure 7.37 Estimated means and 95% CIs for utterance-wide mean f_0 and slope for sentence mode in the phonology-only model. (Mode-plus-phonology model estimates are plotted in grey in the background.)

7.7 Summary and Discussion

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 exists 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 choice but are rather a matter of paralinguistic effects which see a gradual increase in f_0 scaling from MWH to MYN to MDQ. Moreover, it has been argued that such changes in scaling can be explained through the tonal tier alone without recourse to a register tier (Gussenhoven, 2004). It was proposed that 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 it in lieu of an L*H / H*L contrast. It was also noted that this proposal does not mean that only nIE speakers use a register tier, rather it suggests that the presence of a register tier might simply be more salient in this variety of English, where pitch accents are often otherwise identical.

7.7.1 Phonological Analysis

A methodological difficulty arose as soon as 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 distinctly different nuclear pitch contours without a priori recourse to 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. From an auditory perspective as well as from visual inspection of the f_0 contour, there were clear cases where speakers made a categorical utterance-wide 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 prenuclear components of the IP. This strongly suggested that speakers did have access to a phonological register tier. There was also evidence that register tier changes could be implemented across different domains in the IP, i.e., across the whole IP, in the nuclear pitch accent only, or—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 of the labelling challenge was that—despite the potential option to use non-register-tier methods for labelling PA and final boundary tones—any alternative labelling would have required either the relocation of pitch events from the pitch accent to the boundary or—in some cases—the relabelling of L as H. That is, one could have used L*H H% in place of L*^[H] % or H* H% in place of ^[L*H] %. However, the data suggest that such peaks were not associated with the boundary but with the pitch accents. Moreover, the shape of ^[L*H] % contours were structurally more similar to other L*H-type than to H*-type PAs. Specifically, both types of L*H (raised and unraised) comprised a local valley in the stressed syllable of the final foot followed by a rise to a peak toward the end of the foot, with a slight fall off in f_0 near the end of voicing in unspecified boundaries or a dramatical fall in f_0 in L% boundaries. Therefore, it was decided that it would be methodologically unsound to employ non-register-tier labelling, which would demand a reinterpretation of these pitch accents by relocating the tones to different structural units or by switching tonal identity from L to H. Such an approach, it was argued, would simply serve to reflect an orthodox interpretation of the data rather than reflect the evidence found in those data. (In other words, it would represent an instance of fitting the data to support the theory rather than adapting the theory to explain the data.) Therefore, register-tier marking was employed in the labelling, with the caret (^) symbol to identify the onset of high register and square brackets ([]) to identify its extent. As such, the expectation that nuclear pitch events would best be explained via both a tonal and a register tier was already fulfilled before the labelling could even be properly completed (see Expectation 4, Section 7.1.1).

Once register-tier labelling was complete, it could easily be converted to non-register-tier labelling as it was more highly specified than the non-register tier labelling system would have been. To avoid completely hobbling the non-register-tier analysis, however, this did not entail simply removing all register-tier labelling. It also converted L*^[H] (L)% contours to L*H H(L)%. This was viewed as the only phonological contour with a viable non-register-tier labelling alternative, i.e., it did not require the relocation or identity-switching of tonal targets. The non-register-tier labelling was compared with the register-tier labelling to assess the descriptive effectiveness of the register-tier labelling. Admittedly, the non-register-tier labelled data acts as somewhat of a strawman, as it is derived from and less well specified than the register-tier labelled data. However, I felt that this approach could still serve to demonstrate the strength of a register-tier analysis of the intonational phonology.

The non-register-tier analysis indicated that the pitch accent alone was incapable of identifying contrasts as almost 99% of utterances had 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, it also meant that, in practice, only the boundary could be associated with sentence mode. A statistical Bayesian general mixed-effect model analysis of the likelihood of H% as an effect of mode (Section 7.5.1.1) showed that there is a small possibility of H% in MYN (estimated mean .02, 95% CIs [0, .04]) and a slightly higher probability in MDQ, at .13 [.02, .5]. 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 expectation that there would be composite boundaries in a non-register-tier analysis was also met, i.e., HL%. However, this is ultimately uninteresting, as it was required by the conversion from register-tier to non-register-tier labelling that the ^[H] of L*^[H] be moved 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 raised register), ^[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 a probability of .1, 95% CI [.03, .28] of high register in MYN and of .5 [.24, .76] in MDQ. (MDC and MWH were both associated with predicted probabilities of almost zero, at .002 [0, .01] and 0 [0, .008]). A linear regression test

(LRT) of the model also indicated a strong effect of `gender` (*p.adj* = .027), with male speakers more than five times more likely to use H% than female speakers, OR = 5.43 [1.24, 23.8], *p* = .025.

In the register-tier model, the fixed factors of `mode` and `gender` explained 69% of the variance in the response parameter, R_m^2 = .69, while the complete model explained 79%, R_c^2 = .79. In the non-register-tier model, `mode` and `gender` explained 41% of the variance, R_m^2 = .41, while the complete model explained 75%, R_c^2 = .75. Given this, and the fact that the higher register tier (register-tier analysis) is much more readily associated with MDQ than H% (non-register-tier analysis), the register-tier analysis proved a more effective approach in explaining the relationship between intonational phonology and mode. As noted previously, this does not demonstrate that the register-tier hypothesis is true, 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 the first place. However, it does provide evidence to support the register-tier hypothesis.

The phonological analysis of IP-wide phonology suggested that speakers were more likely to avoid prenuclear pitch accents in 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 (Section 7.5.4) indicated a predicted probability of .27 of nuclear-PA-only IPs in MDC, but of .45 for both MYN and MDQ, 95% CIs [.04, .73] and [.09, .87] respectively. There was again an effect of gender, with male speakers being more than 8 times *less* likely than female speakers to use nuclear-PA-only IPs, OR = 0.15 [0.01, 0.95], *p* = .044.]. This meant that the predicted probabilities of a male speaker using a nuclear-PA-only IP for MYN and MDQ were low, at .09 [.01, .48] in each case.

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` (Section 7.5.5). This time, no effect of gender was found. This was unsurprising, as the model combined the strategy preferred by the female speakers and the strategy preferred by the male speakers into a single parameter. However, the effect of `mode` was found to be significant. When at least one of the two strategies was employed, the predicted probabilities of MYN and MDQ were .59 and .87 respectively, 95% CIs [.19, .9] and [.51, .98] in turn. 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 gratifying outcome. Moreover, it is worth noting that it was only through the use of a register-tier analysis of nuclear PAs that it was possible to identify a phonological strategy more closely with male speakers while the IP-wide analysis (which did not require the register tier) identified a strategy more closely associated with female speakers. Once again, this highlights the explanatory value of the register-tier hypothesis. That is, while it does not demonstrate that the register tier *exists*, such felicitous results indicate that it offers an effective and meaningful account of the data. It permits insights into the different ways in which the intonational phonology is leveraged to help distinguish between sentence modes without depending on the assumption that question-form sentence mode must be largely signalled via gradient changes in the f_0 contour.

The final component of the phonological analysis related to the distribution of L% (Section 7.5.2.2). As predicted in Expectation 6 (Section 7.1.1), instances of L% did occur in each sentence mode, 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`, LRT(1) = 0.04, *p.adj* = .914, but that there was a clear effect of mode, LRT(3) = 62.8, *p.adj* <.001. The predicted probability of L% in MDQ was noticeably higher than other modes, at .23 as opposed to the next highest in MYN of .05, 95% CIs [.01, .86] and [.003, .54] respectively. However, the very wide confidence intervals for each level of mode in the model indicated that the association between L% and mode was quite a weak one. In fact, `mode` and `gender` together appear to explain only 7% of the variance in the response parameter (R^2_m = .07).

The results of the analysis of final boundary effects suggest that the previously stated view that L% is more likely in MDQ but not a marker of interrogativity holds quite well under statistical scrutiny. In many varieties of English, L% is associated with statements rather than questions. In terms of form alone, there were a few nuclear contours in the data which resembled those associated with declaratives in other varieties, specifically H* L% and >H* L%; however, 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 provides further (through an admittedly small amount of) evidence that L% is not associated with statement/question contrasts for the Derry speakers in this study. Once the data were adjusted, L% accounted for 39% of all MDQ boundaries occurrences (31% of the raw count) as opposed to an average of 14% in the other sentence modes (14% also in raw data). The frequent occurrence in MDQ 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 were aimed explicitly at eliciting colloquial speech. If one were to analyse more formal speech, such as by a newsreader, one may well find L% functioning as the end of a declarative fall as an effect of accommodation to institutional norms.

These results suggest that the role of the register tier is similar to the role of pitch contour and boundary tone contrasts in other varieties of English, i.e., between L% and H% or between H*L % and L*H H% in contrasting MDC with MDQ. First of all, the phonological contrast is available but is not always implemented by the speaker, and secondly MDQ is typically phonologically more distinct from the other question forms. For example, if we look at the analysis of the distribution of boundary tones and nuclear contours in the Cambridge data from the IViE corpus (Grabe et al., 2005), we see that high boundary tones (H%) were not found at all in MDC, but in 47% of MWH, 44.4% of YNQs, and 77.7% of DCQ tokens. Full nuclear rises (L*H H%) accounted for 29.4% of MWH, 11.1%, of MYN, and 77.7% of MDQ. While the overall distributions differ, (especially when it comes to MWH), we see that there is no fully categorical distinction between sentence modes in terms of pitch accent used. Rather, it is probabilistic. Secondly, we see that the greatest distinction between sentence modes in terms of pitch contour is between MDC (no instances of L*H H%) and MDQ (78% L*H H%).

7.7.2 Phonetic Analysis

The phonetic analysis compared two different models to assess the extent to which a phonological register-tier analysis might account for apparent gradient changes in f_0 scaling in the nuclear L*H PA across sentences modes. One was a mode-only model, which used mode and gender as fixed effects, while the other was a mode-plus-phonology model, which added fixed phonological effects to the model. L and H tonal target parameters were analysed using these models. The tonal target parameters were f_0 scaling (measured in terms of minimum and maximum f_0) and temporal alignment (measured from the onset of the vowel in the stressed syllable to the associated maximum or minimum f_0).

A mode-only analysis of the tonal targets 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 Equation 7.1 (Section 7.1.2) as:

$$f_{0_{MDC}} \leq f_{0_{MWH}} < f_{0_{MYN}} < f_{0_{MDQ}}$$

The mode-plus-phonology model analyses reduced the apparent scaling of f_0 targets as an effect of mode. However, 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" (Section 7.1.2, p. 139) were clearly overstated. Just as in the other varieties (Grabe et al., 2003), therefore, there is still a gradient effect of question mode.

Two global f_0 parameters were also analysed (Section 7.6.3), utterance mean f_0 and the slope of the linear regression of $f_0(t)$ across the utterance (or more simply, utterance-wide f_0 slope). Phonological effects on these parameters were also evident; however, the effect of mode was found to be greater. This is most likely because the bulk of the phonological effects of sentence mode are concentrated in the nuclear pitch accent. When we look at the utterance as a whole, the phonological effects are diluted, and the gradient paralinguistic effects of mode come to the fore. Therefore, we have to conclude that both gradient and categorical effects operate on the f_0 contour, with gradient effects being an effect of sentence mode, and categorical effects resulting from the use of phonological features which help signal different sentence modes.

Because a register tier is not commonly evoked in AM analyses, it was deemed necessary to evaluate the extent of support for the register-tier analysis in terms of the distinctiveness of each tone variant, i.e., L* versus ^[L*] and H versus ^[H] in L*H, ^[L*]H, L*^[H]. and ^[L*H] (Section 7.6.2). It was found that the f_0 scaling and temporal alignment of L* were very similar in both L*H and L*[H], as were those of ^[L*] in ^[L*]H and ^[L*H]. Moreover, the f_0 scaling of the non-raised and raised-register L* tones were statistically categorically distinct from each other. The f_0 scaling of the trailing H tones in L*H and ^[L*]H were also found to be almost identical, and both were statistically categorically distinct from their raise-register counter parts. However, the H tone of ^[L*]H was aligned noticeably earlier, leading to the suggesting that perhaps the few instances of ^[L*]H found in the data may more appropriately be considered variants of H* or even >H*. The high-register H tonal targets, i.e., those in L*^[H] and ^[L*H] were also found to be remarkably similar both in f_0 scaling and temporal alignment, and statistically categorically distinct from their low-register counterparts.

These results suggest that the interpretation of tonal targets in terms of low and high register tier is overall very consistent, and the prediction that there would be significant differences in scaling due to register-tier effects (Expectation 8, Section 7.1.2) is borne out. The consistency in f_0 scaling of high and low register-tier instances of L and H tonal targets exceeded all expectations and the results were remarkably satisfying as this provides further evidence for the register-tier hypothesis.

7.8 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 beyond the domain of the nuclear pitch accent, gradient effects of sentence mode on f_0 are more readily apparent.

The first aim of this chapter was to answer the second research question, that is, it aimed to provide a description of the phonological and phonetic characteristics of nuclear pitch contours in DCE across sentence modes. In doing so, it has also the fourth research question, “Does a register-tier analysis provide a plausible explanation for phonetic variation across sentence modes in DCE?” The answer to this question, I believe, has to be yes. The register-tier analysis does not explain away all paralinguistic/gradient effects of sentence mode, but it does explain how speakers sometimes signal sentence mode in conjunction with other phonological features, i.e., the use of high register appears to be one of two phonological strategies employed to help signal MDQ, the other being the avoidance of prenuclear pitch accents. It also allows us to evaluate the extent of the paralinguistic effects of sentence mode more accurately, since, by including register-tier along with other phonological effects in our models, we can be more confident that we are not inadvertently ascribing categorical phonological effects to gradient effects of sentence mode.

8 Critique and Alternative Approach

The previous two chapters have provided a detailed description of the phonology and phonetics of the neutral declaratives in Derry City English under different metrical and lexical boundary conditions, as well as a detailed analysis of the phonology and phonetic implementation of nuclear pitch contours and IP-wide intonation across different sentence modes. These served the descriptive and theoretical aims set out in the research questions, and I believe answered them adequately. However, the analyses also raised questions, particularly about the implementation of the intonational phonology and features observed in the f_0 contours which were not directly connected to answering the research questions, but which it would be irresponsible to overlook. This is because they relate to the methodological and theoretical issues which informed the approach to the analysis to begin with. The aim of this chapter, therefore, is to engage with these issues, highlight areas where future work might take a different methodological approach, and provide some examples of how such an approach might be of service to the field of intonational phonology.

8.1 Prenuclear Pitch Accent Labelling

In Chapter 6, a labelling issue arose in pitch accents where the H target was aligned later in the foot than the typical H* and earlier than the typical L*H. In such cases the auditory impression was somewhere between an L*H and an H*, especially as these pitch accents lacked the distinctive percept of a low pitch rising up out of the stressed syllable. Thus, such pitch accents were labelled as >H*. As noted in 6.7.1, this could be construed as a case of overspecification of the phonology to incorporate an element of gradience; however, it seemed more reasonable to adopt a strategy that acknowledged the inherent ambiguity that can otherwise be erased in sticking doggedly to a strict approach to categorisation.

Ironically, there was a second element in some of these categorically ambiguous pitch accents which was largely glossed over. That is, in some cases, >H* pitch accents were clearly some sort of rise, but—unlike L*H—there was no salient rise from a low pitch target on the stressed syllable (Figure 8.1B). In other cases, >H* represented the end of a plateau which began at the boundary (Figure 8.1A). For the latter, these were labelled as %H >H*. Similar pitch events occurred in a small number of PN H* pitch accents as well, as outlined in Table 6.7 (Chapter 6, p. 66). Arguably, the rise-to-H and plateau-like >H* and H* PAs, much like >H* itself, also represent a degree of gradient variation (this case in terms of the f_0 slope at the IP onset) which could have been analysed as categorically distinct. However, at the time, it seemed that this would have been a much clearer instance of adding too much phonetic detail to a phonological labelling system.

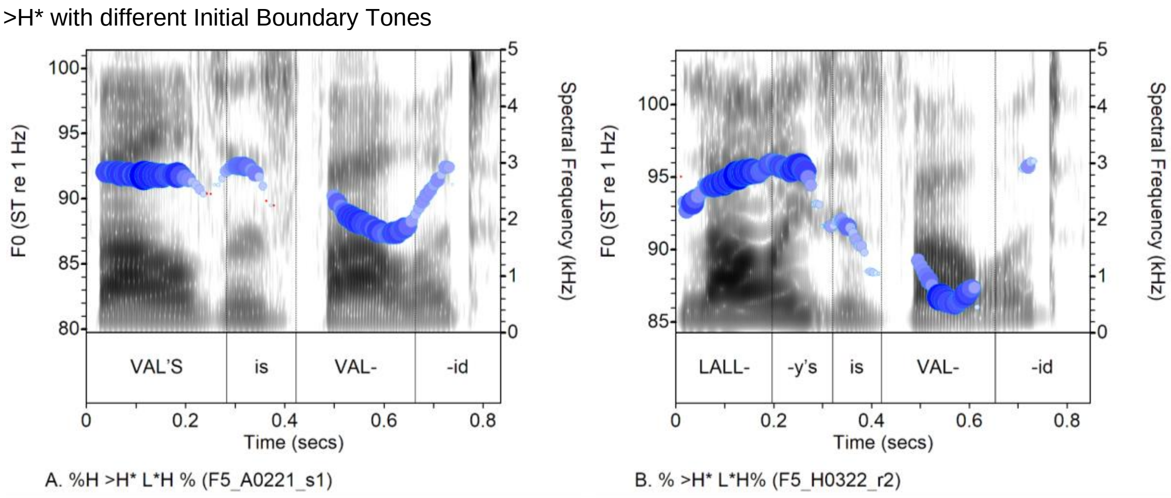


Figure 8.1 Spectrogram and f_0 contour of two utterances labelled with prenuclear >H but with different initial boundaries. The f_0 contours incorporates Cepstral Peak Prominence (CPP), which is shown by the size and intensity of the circles.*

In Chapter 7 (Section 7.4.1), a similar issue with labelling was noted in relation to PN pitch accents which might be interpreted either as !H* or L*. Sometimes there was both an auditory and visual distinction between the two types of the PA, notably in the shape of the curve around the stressed syllable. This was specifically a difference between a convex contour in the !H* and a concave contour in the L*. However, in some cases the distinction was not so easily explained. The problem was that there was a concave turning point before the stressed syllable followed by a convex turning point, meaning that the visual distinction appeared ambiguous. Ultimately, auditory impressions must win out in such analyses, and in the examples provided, the impression was that they were L* rather than !H*. It was also observed that after such pitch accents, there was a tendency for an anticipatory lowering of f_0 before that nuclear L*H. In other words, there was a valley preceding the L target of the L*H pitch accent. The second PN pitch accents were not terribly important for the analysis in Chapter 7, which focussed primarily on nuclear pitch accents and contours, while PN pitch accents were considered important only in so much as they were either present or absent in the IP. As a result, further thought was not given to them at the time.

While neither the best labelling strategy for the second PN nor the anticipatory lowering of f_0 was considered very crucial to the analysis, on reflection, this points to a weakness in the analysis. This is because it hints at a potential flaw in the general approach, which might most aptly be described as a Phonology-first approach. That is, the approach adopted in the main analyses involves first identifying the intonational phonology and then extracting phonetic parameters which are viewed as implementations of that phonology. However, perhaps a better (or at least alternative) approach is to focus on the structure of the contour itself first without much attention to the ultimate goal of describing the intonational phonology. In a sense, this is what the Institute for Perception Research (IPO) did in its close copy synthesis of speech contours ('t Hart et al., 1990). That is, it involved stylizing and resynthesizing f_0 contours until they were auditorily indistinguishable from the original. After this process, the stylizations were evaluated and described in terms of phonological form and function.

8.2 Identification of tonal targets

In the analyses, tonal targets were identified in terms of the scaling and temporal alignment of f_0 maxima and minima. This was in part because minima and maxima make sense intuitively as the best candidates for high or low targets. It was also, in part, because the alternative was more difficult to identify reliably, i.e., turning points or elbows in the contour (Chapter 2, Section 2.4.1). However, as should be clear from the above description of prenuclear PAs in terms of concave and convex contours, it is quite difficult to avoid using elbows as a reference for the identification of tonal targets. As was also noted in Chapter 2, an f_0 maximum or minimum can be one special type of turning point.

It should, of course, be noted that treating f_0 maxima and minima served its purpose in the analyses and led to several valuable insights. For example, it helped establish the fact that nuclear peak alignment can be understood (in the corpora analysed here at least) to be stable when measured as a proportion of voicing in the nuclear contour. It also helped demonstrate that both truncation and compression strategies are used in the nuclear rises but that only truncation strategies appear to be available in prenuclear rises.

For the assessment of bimodality in prenuclear peaks, however, it is possible that by associating f_0 peaks with H targets, the results of the analysis may have been flawed. The findings that PN f_0 peaks may be associated with different anchor points (word-final syllables or the right edge of the stressed syllable) are not in question here. However, what might be questionable is the assumption that these peaks are also the tonal targets and that there is a bifurcation in the timing of the targets. If, instead, the analysis had treated elbows rather than peaks as stand-ins for tonal targets, the results may have been quite different.

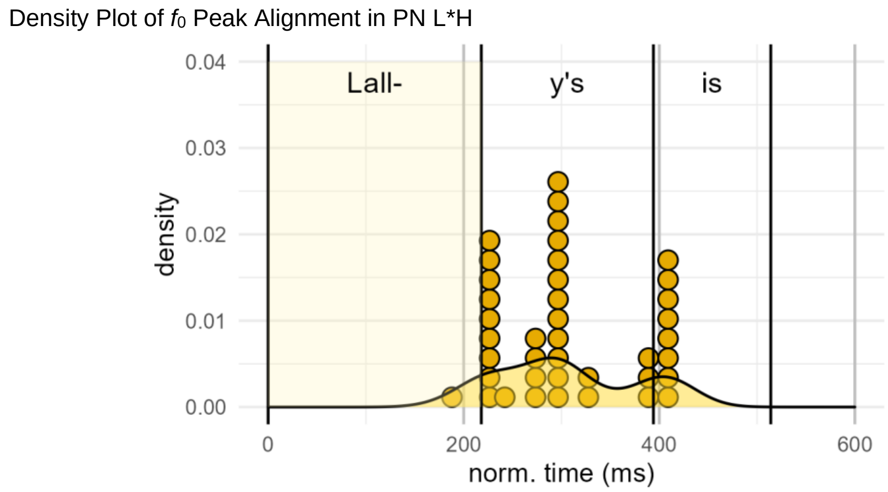
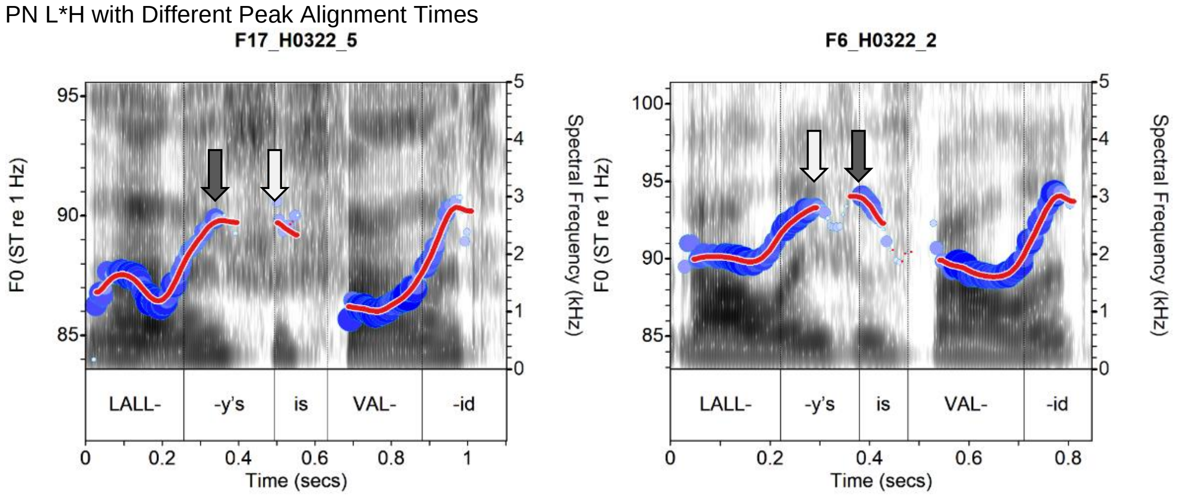


*Figure 8.2 Density and dot plot of prenuclear L*H f_0 peaks in the phrase "Lally's is valid" from the H-Corpus (code H0322) in grand mean syllable-normalised time. The lightly shaded rectangle indicates the stressed syllable and vertical black lines syllable boundaries. Each circle represents a single utterance (n = 40).*

For example, consider the H-Corpus target phrase, *Lally's is valid*. The analysis of density distribution of PN L*H f_0 peaks in that phrase suggested a possible bimodal—or even multimodal—distribution (Section 6.6.1.2). The distribution is shown here again in Figure 8.2, where we see possible density peaks at the right edge of the stressed word, the middle of the word-final syllable, and at the end of the word-final syllable. However, if we think of the H tonal target not as an f_0 peak but as an elbow, we might find a more apt explanation of the phenomenon.

Figure 8.3 illustrates why this might be the case via two examples of PN L*H in *Lally's is valid*. In the example in Panel A, the PN f_0 peak occurs in the middle of the second syllable at around 0.35 seconds, indicated by the dark grey arrow. This is also the onset of a plateau which continues until the end of the stressed word at 0.5 seconds. In Panel B, the peak of the contour occurs is at the right edge of the word at around 0.37 seconds, again shown with a dark grey arrow. However, in this case, this is end of a plateau (or of a very shallow incline) which begins in the middle of the word-final syllable at 0.28 seconds, as indicated by the white arrow. Thus, in both examples, we see that the start and end points of each plateau are in the same location, i.e., the middle and end of the word-final syllable. If the analysis treats the f_0 peak as a tonal target, then both each phrase does appear to align the tonal target with a different anchor point. If, however, we take the onset of the plateau (or even its offset) as the tonal target, then each phrase has a tonal target in the same location.



A.% L*H L*H % (F17_H0322_5) B. % L*H L*H % (F6_H0322_2)

Figure 8.3 Spectrogram and f0 / CPP contour of two repetitions of Lally's is valid (H0322) from different speakers (F17 and F6). Arrows indicate high turning points, red lines a smoothed contour (bandwidth = 12).

In the turning-point oriented analysis, it makes sense intuitively to treat the onset of the plateau as the tonal target because this is the next time point after the beginning of the rise where the trajectory of the contour changes direction rapidly, and it is therefore likely to be auditorily salient.

8.3 Tonal targets versus Tones

If tonal targets are to be considered in terms of turning points rather than peaks, this opens up a new range of possibilities for the phonetic description of intonation in an AM context. Most notably are those related to the extension of the tone, a phenomenon which has been considered under a range of guises, such as secondary association (Pierrehumbert & Beckman, 1988) or double alignment (Gussenhoven, 2004). Essentially, we can view one point in a valley or plateau as the tonal targets and the other as the point to which (or from which) the tone extends. In other words, a tone inherently has a duration, but to achieve that tone, one has to hit the f_0 targets. It also possibly provides an alternative means of resolving the disputed existence of the phrase accent[25].

The idea of secondary association has been considered as a means of explaining the apparent spreading of phrase accents to other locations in the IP, such as from the right edge of the word to the right edge of an intermediate phrase (ip), which is marked by a phrase accent but lacks a boundary tone since boundary tones are associated with the IP boundaries only and not ip boundaries (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988). Grice et al. (2000) argue that there is evidence for the phrase accent in several European languages (English, German, standard and Cypriot Greek, Hungarian, and Romanian). They view the phrase accent as an edge-tone phenomenon, and as such, see it as a (literally) peripheral event which does not signal prominence. Prieto et al. (2005) argue that secondary association of phonological tones is needed to account for the structural features found in their analysis of some Romance languages (Central Catalan, Neapolitan Italian, and Pisa Italian). That is, they see the need for the secondary association not simply to be a feature of a peripheral tone but as a constituent of the pitch accent without which categorically distinct pitch events are difficult to describe formally. Whether or not one views the secondary association as a constituent of the pitch accent or the boundary tone within this approach, it is to a large extent a manifestation of the of on-ramp / off-ramp debate (see Chapter 2, Sections 2.3.4, 2.3.7, and 2.5.1), and a matter of whether or not an L tone is linked to the pitch accent as a trailing tone or is linked to the right IP boundary as a phrase tone.

Gussenhoven's view is slightly different. He argues that tones can have both right and left alignment (Gussenhoven, 2004, 2016). That is, in English, for example, he argues that the initial boundary and monotonal pitch accents extend rightward until the edge of the next pitch event, as too can the nuclear PA trailing tone before the phrase boundary. Thus, this kind of secondary association is *not* related to a different tone type but is just an extension of a boundary tone, a starred tone, or a trailing tone.

The idea proposed here is that we can view all of these as manifestations of the same underlying phenomenon if we distinguish the tonal target from the tone itself. When we adopt a turning point analysis of tonal targets, this becomes much more transparent. For example, in the two f_0 contours shown in Figure 8.3 above, there are two pitch accents, the prenuclear L*H and the nuclear L*H. The prenuclear L*H is physically different from the nuclear L*H because it has a plateau.

[25] I still believe that the ToBI interpretation of the -L phrase accents as non-lowering tonal event is misguided as it requires faith in the existence of a tone for which there is little empirical evidence, but this does not mean that there is no evidence for phrase accents at all.

However, there are still only two tonal targets in each pitch accent, an L and H. With the turning points analysis, the real difference between the two is that PN tone is sustained after it achieves its target. The H tone is not doubly aligned, it simply has extent, as all tones inherently must (i.e., a tone is not a transient event). The end of the plateau might be better construed, therefore, not as a secondary association or double alignment, but simply as the right edge of the tone.

Note that this does not conform to Gussenhoven's description of English intonation, since it represents a non-nuclear trailing tone which extends, not until the next tonal event, but to the end of the lexical word. In Gussenhoven's description, if I have understood correctly, the extent of the double alignment is determined by phonological rules. Here, however I suspect that the selection of tonal alignment target and the extent of the tone itself are pragmatically motivated and can be used to extend or foreshorten the scope of the semantic content which the pitch accent covers. This will be illustrated with another example from *Lally's is valid* (Figure 8.4).

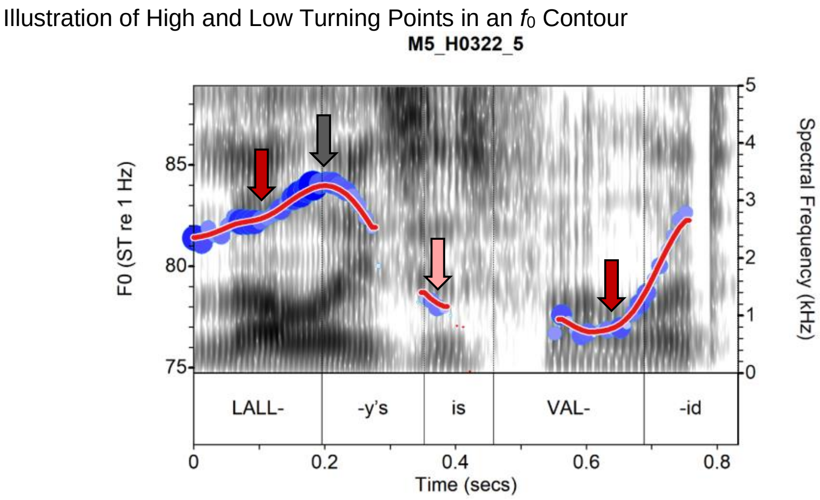


Figure 8.4 Spectrogram and f0 / CPP contour of a repetition of Lally's is valid (H0322) from different speakers M5. The grey arrow indicates the high turning points, pink and red arrows low turning points, and red lines a smoothed contour (bandwidth = 12).

Figure 8.4 shows an example of f_0 contour of *Lally's is valid* from one of the male speakers, M5. The PN in this case is also L*H, as in the previous examples, but this may not be immediately apparent if one looks at the smoothed (red line) contour, which slightly obscures the low turning point, making the L target auditorily salient (highlighted here by the first red arrow). The PN peak can be identified at the right edge of the stressed syllable.[26] In this case, the peak is also the turning point. As such, there is no extent of the tone, and f_0 falls to another turning point in the syllable nuclear of *is*. I would argue that this is an example of L* anticipation, in that it represents the initiation of the L* target, the main turning point of which is at the end of the vowel in *VAL-*. In short, I am

[26] One could argue that there are two turning points, one just before and one just after the boundary; however, this seems phonetically over specified, especially since peaks can themselves be a special class of turning point.

suggesting that the tone of the nuclear L* begins at the start of *is* but the true target (the one that is intended as salient) is located in the stressed syllable of *valid*. The purpose of the anticipatory lowering is to prepare the listener to attend more carefully to the semantic content associated with the nuclear pitch accent.

The effect of this switching up of tonal targets and the duration of the tones, therefore, is to change the degree to which the listener will attend to the different elements of the semantic content. That is, I contend that in the phrases illustrated in Figure 8.3, more attention is drawn to *Lally's* as the trailing H tone extends into the last syllable of the word. In the example in Figure 8.4, however, attention is cut short as the H tone is not extended, and in fact the listener is encouraged to get ready to attend to the semantic content associated with the nuclear pitch accent via the anticipatory lowering at the start of the complement, i.e., in *is*. This may be a subtle distinction, but it explains why there are different locations for turning points and facilitates a move away from a formal explanation for the location of tonal targets, which was the aim of—and thus the approach taken in—Chapter 6, towards a functional pragmatic explanation.

Implicit here is the fact that H targets are more likely to be associated with the first (and potentially only) turning point while L targets are more likely to be associated with the second turning point (if there are two). This might also go some way to explaining why L targets in MDQs appeared to be aligned earlier (Chapter 7, Section 7.6.1). That is, when analysing tonal targets in terms of turning points, it appears that some speakers, when they use a raised register, scale the anticipatory lowering before the starred tone lower than the starred tone itself. This is likely a side effect of raised register, in that the actual phonetic raising of register may not be fully implemented until the L* target itself. This is illustrated in Figure 8.5.

Figure 8.5 shows a repetition of MDQ2 from speaker M5 in which there was clear evidence of raised register in the nuclear pitch accents. The IP was labelled %L >H* ^[L*H] %. If we treat the tonal target as the f_0 minimum, we see that it occurs at 0.52 seconds (the pink arrow). However, there is a second turning point at 0.6 seconds (the red arrow), where the f_0 begins to increase even more rapidly. This second turning point occurs at roughly the same location in the syllable as the typical non-raised f_0 minimum found in the data, i.e., at the end of the vowel in the stressed syllable. However, because of the general upward trajectory of register raising, its identity is less salient than that of the f_0 minimum. There are two ways of reading this. One is that, as an effect of register raising, the L* has been relocated to a time point more typically associated with anticipatory lowering. The other is that this is still an instance of anticipatory lowering and that the L* is simply less salient as an effect of raised register. In fact, if we were to specify the domain of the raised register more precisely, we might argue that it begins, in this example, on the L* itself, which further masks the identity of the L* targets, as it is incorporated into the rise more generally.

Illustration of Two Low Turning Points

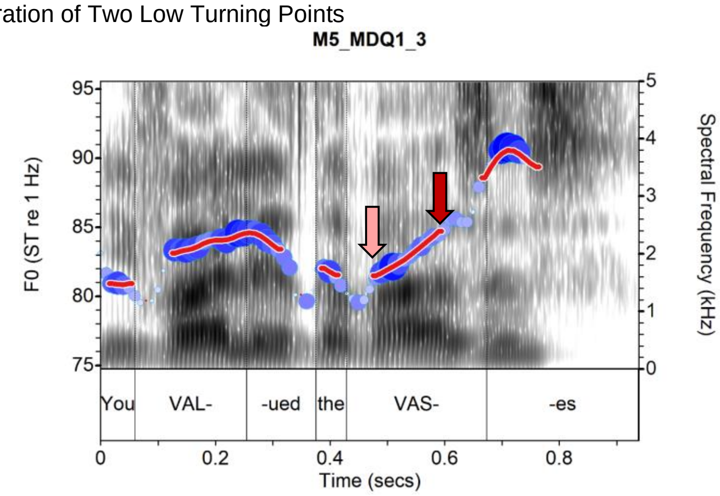


Figure 8.5 Spectrogram and f0 / CPP contour of a repetitions of You valued the vases? (MDQ1) from M5. Pink and red arrows indicate low turning points, and red lines a smoothed contour (bandwidth = 12).

8.4 A Phonetics-first Approach

The critique of the approach adopted in the two core analytical chapters of the thesis has focussed on two things. One is that tonal targets might have been analysed more effectively in terms of turning points in the contour rather than f_0 extrema. The second is that the contour could have been analysed in a more nuanced way had the focus not been on the tonal targets of the pitch accents alone (regardless of whether extreme or elbow) but on both the tonal target and the extent (or duration) of the tone. As noted above, the flaw here was in taking what I have described as a Phonology-first approach to the phonetic analysis. However, I would like to outline the alternative Phonetics-first approach to the phonology based on the above discussion on the advantages of a turning-point analysis which takes into consideration the extent of the tone.

The Phonetics-first approach begins with the AM assumption that an underlying sequence of tonal primitives are realised in the f_0 contour. As such, the f_0 contour can be conceptualised as a series of interpolations from one tonal target to the next. The tonal target is essentially a coordinate in the contour towards which the speaker's f_0 trajectory is oriented, but the target is not the only aspect of the tone. A tone, by definition is a temporal phenomenon. That is, it is either high or low, and must be realised via the vibration of the vocal folds. In fact, f_0 is a measurement of the fundamental oscillation of the vocal folds and can be construed as the angular velocity of the vocal folds as they vibrate back and forth (Hardcastle, 1976). Since velocity is defined as displacement over time, this is why tone is a phenomenon which has inherent duration.

If we view tonal targets and tones in this manner, the Phonetics-first approach involves identifying the least number of turning points in the contour which can be used to generate a synthesized close copy which is indistinguishable from the original when the waveform is resynthesized with the copied contour. This element of the Phonetics-first approach is very much inspired by research conducted for the IPO ('t Hart et al., 1990); however, while they used a resynthesis technique, it was not based on turning points and tonal targets.

Once the turning points have been established as phonetic events in the contour through this process, they can be analysed through the AM lens. That is, they can be categorised as L or H and they can be linked to structural units, i.e., boundaries, pitch accents, phrase accents, or any constituent in the phonological structure of the IP. Of course, there may be more or fewer turning points than expected, but as the only determining factor in the initial selection of turning points is the identification of the smallest number required to reproduce the contour, this is where phonological analysis comes in. What matters most here is that the phonological analysis must be based on phonetic evidence defined in terms of turning points. This may appear to be a strictly literal approach to the implementation of phonological tones, but that is the point. (Thus, the Phonetics-first approach.)

8.5 Implementation of a Phonetics-first Approach

The Phonetics-first approach was implemented though the development of an interactive semi-automated Praat plugin, K-Max, which uses the second time derivative of the f_0 contour to identify tonal targets and allows the user to select from a range of candidate turning points until the least number of tonal targets have been selected which still can be used to generate a close copy f_0 resynthesis of the original contour. K-Max was originally reported in a paper presented at Speech Prosody 2000 (Rodgers, 2020) and is also reproduced in Appendix N.[27] As mentioned above, the resynthesis component of the procedure is inspired by the IPO approach, but in the identification of tonal targets, it takes some of its cues from the PENTA model of prosody (Xu, 2005, 2004; Xu et al., 2022). However, it is still based on and designed for AM analysis. The description which follows focuses on the aspects of the modelling which are directly pertinent to the description of and identification of tonal targets which informed the algorithm used in K-Max.

If we begin with the view that the f_0 contour is the angular velocity of vocal fold vibration over time, i.e., f_0(t), then its first time derivative is rate of change of velocity, i.e., acceleration, while its second time derivative, f_0''(t), is the rate of change of acceleration, also known as jolt or jerk. The effect of jolt is what one feels while sitting in a car which begins to accelerate more rapidly or breaks suddenly. In fact, the unpleasant experience one feels when a car breaks suddenly (the literally jolt) is caused by a peak in jolt.

$f_0''(t)$ extrema align temporally with points of maximum curvature in the f_0 contour, both concave and convex. It should be noted that $f_0''(t)$ is not a direct measure of curvature, but for the sake of convenience the term K_{max} (maximum curvature) will be used instead of the more

[27] K-Max is freely available at github.com/AERodgers/Praat-K-Max under the GNU general public license.

cumbersome $|f_0''(t)|_{max}$ (absolute peak in magnitude of the second time derivative of f_0). In the f_0 contour, these points of maximum curvature are evident as turning points, of which—as noted previously—some f_0 minima and maxima are a special subset (they are roots of both the first and third time derivative).

The nature of the relationship between $f_0''(t)$ and f_0 turning points is shown in the stylised contour in Figure 8.6. In this figure, f_0 is shown in black and $f_0''(t)$ in red. $f_0''(t)$ spikes at turning points in the f_0 contour (black), regardless of whether they are elbows or f_0 extrema. Negative spikes coincide with the most convex points (including f_0 maxima) and positive spikes with the most concave points (including f_0 minima), so polarity of f_0'' indicates whether the turning point is more H-like (negative) or L-like (positive). It is worth noting that in real-world cases f_0 extrema often occur near rather than at K_{max}, so f_0 maxima and minima may even sometimes be viewed as symptoms of or epiphenomena associated with K_{max}. Understood in this way, however, we do still have—in principle—a method of identifying turning points in the f_0 contour. Of course, in practice, the stylization is a long way from resembling a real contour, in part due to its unnaturally perfectly angular turning points and in part due to the fact that real f_0 contours are much messier than this.

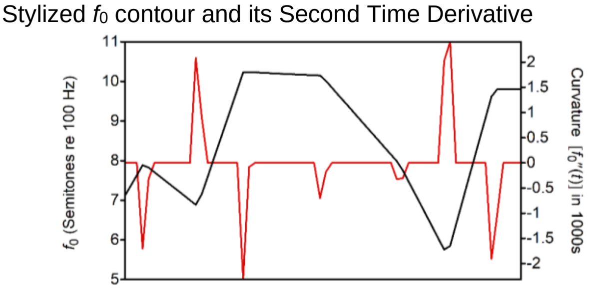


Figure 8.6 Stylised f_0 contour (black line) and its second time derivative (red line).

The PENTA model of f_0 prosody (Xu et al., 2022) makes a distinction between the underlying tonal targets and how they are manifested in the f_0 contour, or—more accurately—how they fail to be manifested. That is, in the constant dynamic realisation of f_0, tonal targets are never fully realised but are only approximated due to physiological constraints on the larynx which limit the maximum rate of change of f_0. This results in phenomena such as tonal undershoot, which occurs because speakers must adjust their f_0 trajectory over time to approximate subsequent tonal targets (Xu, 2005; Xu & Sun, 2002). Xu and Xu (2005) have described tonal targets in this sense as *covert* targets, which they contrast with *overt* targets, or direct measurement of f_0 in the contour itself. They criticise the AM approach for relying on *overt* tonal targets, such as f_0 peaks and minima. (As noted at the head of this chapter, such an approach can provide useful insights into intonational phenomena.) However, this critique is somewhat of a strawman, as there is no principled reason for AM analyses to demand the use of *overt* targets. In fact, the AM approach itself works on the assumption that the underlying abstract tones may be fully or partially realised in the f_0 contour, so one may well expect a mismatch between tonal targets as idealisations and their realization in the contour.

Therefore, in working on a means of identifying tonal targets in the shape (so to speak) of turning points, the view adopted in the Phonetics-first approach is similar to that of the PENTA mode; specifically, that tonal targets are ideal targets of a trajectory which is never fully realised because physiological constraints require that the speaker must turn the f_0 trajectory away from one target in order to begin the trajectory towards the next.

Given that the turning point is not the target itself but the point at which the f_0 contour is moving most rapidly away from one tonal target to adjust its trajectory toward another, the best way—perhaps—to identify the never-realised covert target is to identify the points where the trajectory of the contour is most stable, i.e., the points at which it is most likely moving toward the ideal target. Bearing in mind that $f_0''(t)$ minima and maxima indicate points of maximum convexity and concavity respectively in the f_0 contour, the points of zero curvature between these two points—mathematical inflexion points identifiable using the roots of the second derivative—represent just such moments of stability. Thus, they can be used to identify points when the f_0 contour is most 'on course' towards the ideal target. Consequently, the f_0 slope (or tangent) at the inflexion point can be viewed as an *ideal* slope towards an *ideal* (covert) target. By plotting the intersections between the tangents of mathematical inflexion points, therefore, we can estimate the timing and scaling of the covert tonal target. This process is exemplified in Figure 8.7, which shows f_0 and $f_0''(t)$ contours of a model curve. The blue lines indicate tangents projected from inflexion points between maximum f_0 curvature. The intersection of the two tangents (around a time of 0.35) can be understood as the best approximation of the (covert) tonal target.

Schematic Representation of Tangents Projected from f_0 Inflexion Points

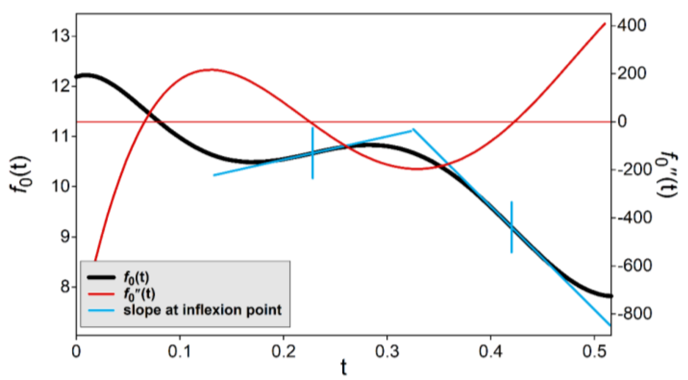


Figure 8.7 Example of f_0 contour (black line), its second derivative (red line), and tangents (slopes) projected from inflexion points (light blue line).

K-Max employs the `fix_pitch` function (outlined in Chapter 5, Section 5.2) to allow the user to manually correct f_0 errors in a target utterance. It also smooths the contour to facilitate the identification of mathematical inflexion points. K-Max then calculates $f_0''(t)$ and identifies a sequence of candidate tonal targets based on the intersections of tangents projected from the mathematical inflexion points. These are then used to populate a TextGrid which identifies the time points of candidate turning points, the script is paused, and the user selects the minimum number of tonal targets they think will be needed to adequately generate a convincing resynthesis of the contour.

After this, the utterance is resynthesized with a new f_0 contour based on the time points selected by the user. The user listens to test the adequacy of the resynthesized copy and, if necessary, adjusts the selection of time points until an acceptable copy is generated. The user can adjust the degree of smoothing both for turning point estimation and for the simulation of physiological constraints. (A more detailed description of the process can be found in the paper in Appendix N.)

K-Max Tonal Target Selection Interface in Praat

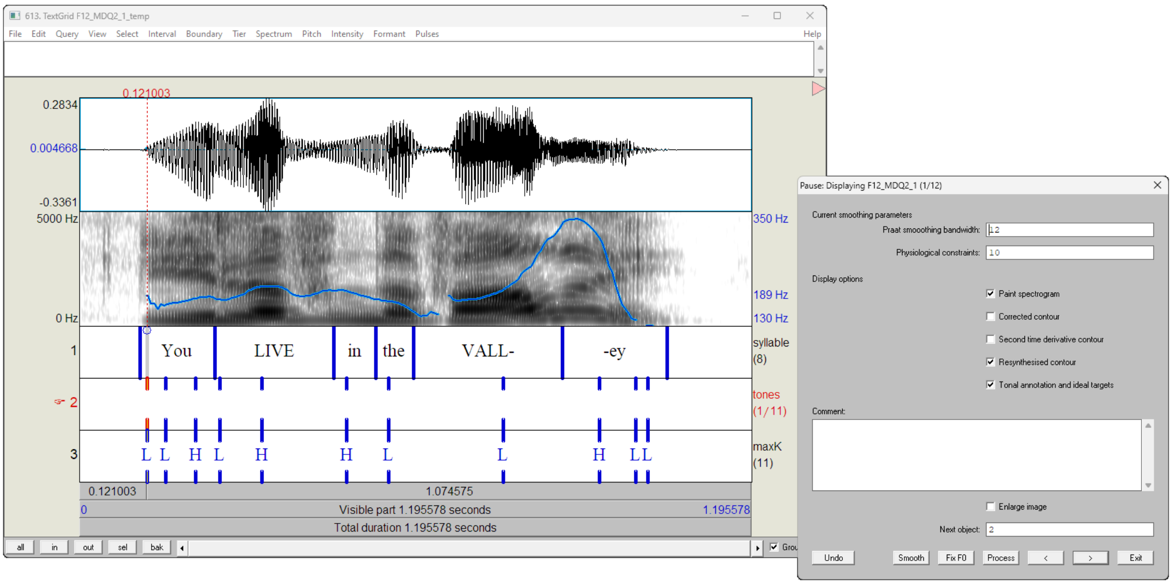


Figure 8.8 Example of K-Max tonal target selection process.

8.6 Examples of the Phonetics-first approach

Unfortunately, a complete reassessment of the data to answer the research questions using the Phonetics-first approach is beyond the scope of this thesis. Therefore, in its place, a small sampling of contours analysed using K-Max is presented here to demonstrate how the Phonetics-first approach may provide insights into phenomena which could not be fully explained or were not fully explored in the Phonology-first analysis. Admittedly, this is just a taster of the possibilities for an alternative approach and should not be seen in any way as an attempt at a comprehensive analysis. However, it does permit some insight into issues noted during the main body of the analysis in Chapters 6 and 7, specifically in relation to the status of >H*, downstepped prenuclear PAs, anticipatory lowering, and the apparent earlier alignment of L targets in declarative questions.

In the turning point selection process, turning points are labelled as L, H or 0. An L label is only used if the K-Max analysis indicated that the turning point is L-like, and likewise for H labels. 0 is used at the boundaries as K-Max is not clever enough to draw the contour unless the start and end points of the contour are identified. If there is a clear H or L boundary tone, however, it is labelled appropriately. Note that lowercase L—e.g., l*+H—was used in cases where there was no auditorily salient L pitch event but when the contour could not be resynthesized adequately without it. The significance of this discussed below in Section 8.6.1.

Additional tags are used to identify the possible role of each turning point in the intonational phonology. Starred tones are labelled as normal, i.e., as H* or L*. Trailing tones in bitonal pitch accents are identified with a plus sign to indicate the starred tone with which they are associated; therefore, L*H becomes L*+H. If the H tone has a clear extent, its right edge is labelled as _0, and in cases of anticipatory lowering, the tonal event is labelled as L_. The reason for the use of these conventions is to facilitate machine readability. For example, an L*H pitch accent with both anticipatory lowering and a high plateau can be rendered as a single unit, L_L*+H_0. This detailed specification of the pitch accent can then be processed to return a minimal specification, i.e., L*H. Therefore, from the more highly specified phonetic analysis of turning points we can arrive at a sparse phonological description.

*8.6.1 The status of >H**

Figure 8.9 shows a K-Max analysis of an example of *Lally's is valid* from the H-Corpus (F5_H0322_r2). The blue dots represent the original contour with the intensity and size of each dot indicating the magnitude of cepstral peak prominence, as has been the practice until now. The yellow line shows the resynthesized contour generated using the minimal number of (covert) turning points. The resynthesized contour has also been smoothed to simulate physiological constraints. The ideal location of each turning point is marked by an X with the identifying label above it.

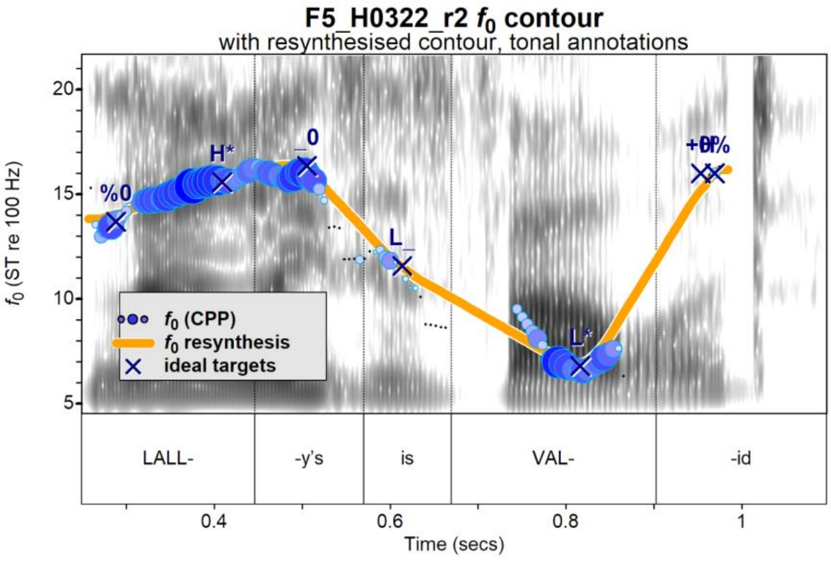


Figure 8.9 K-Max analysis of a contour labelled in the original analysis with PN >H which may be better labelled as H*.*

In the original analysis, this IP was labelled as % >H* L*H %. However, when we use the phonetics first approach, we have to include a turning point in the first syllable near the right edge of the vowel (0.39 secs), while the peak at the end of the second syllable is identified as the end of the plateau. As such, the >H* is reassessed as H*, with the tonal target of the starred tonal aligned in a prototypical location, i.e., the right edge of the vowel. This phenomenon was noted in several pitch accents, suggesting that in some (but not all) cases, >H*, as analysed in Chapter 6, is a variant of H*.

There are other cases where there is a slight concave rise in the stressed syllable of the PN position but no audibly salient L-like event. Such cases were also identified as >H* in the original analysis. However, the concave element of the contour may be the phonetic trace of an L tone which has been deleted. In some cases, it was necessary to include this trace as a turning point in the K-Max analysis so that the resynthesis would render correctly. An example of this is shown in Figure 8.10.

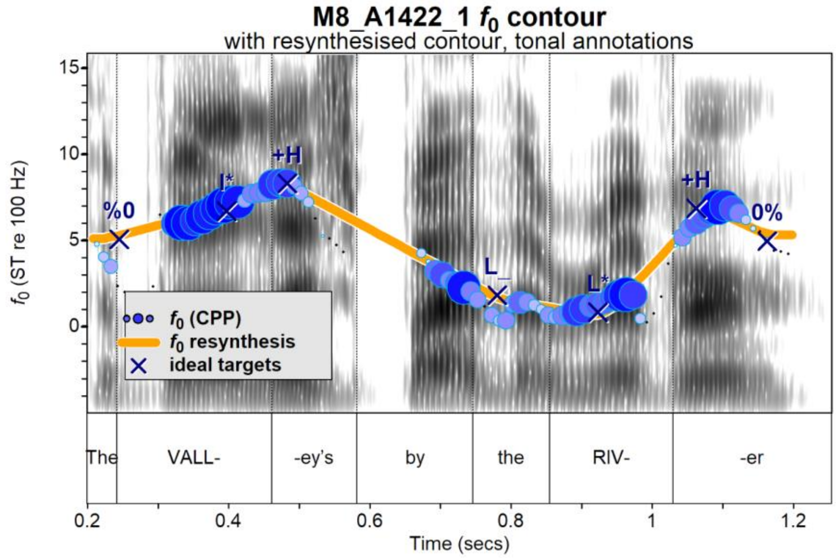


Figure 8.10 K-Max analysis of a contour labelled in the original analysis with PN >H which may be a true example of L deletion.*

In Figure 8.10, there is a rise to a PN peak at the beginning of the second syllable in the first foot. In the original analysis, this was identified as the starred tone, but because it was delayed until the second syllable and did not create the auditory impression of a typical H* (or L*H), it was interpreted as >H*. Here, however, it is interpreted as the trailing tone of an L*H in which the L tone has effectively been deleted but where a trace remains in the contour, without which an adequate resynthesis cannot be produced (thus the use of the lowercase-L notation in l*+H notation). When K-Max estimates the location of the deleted tonal target, it is also at the right edge of the stressed syllable, a prototypical location for the tonal target of the starred tone. Therefore, this l*+H might be interpreted as >H*, or alternatively as a (weakened) realisation of L*H in which the L* has (almost) been deleted.

8.6.2 Downstep

Figure 8.11 illustrates where the question word has a H* pitch accent and is followed by a downstepped H* on *val-*. K-Max identifies the turning point which has been interpreted as the tonal target of the downstepped H* at the right edge of the vowel in the stressed syllable, as one might expect. It is preceded by an L (concave) turning point. In fact, an intervening concave L target in the form of anticipatory lowering seems logically necessary for a downstepped H* to be realised at all. Therefore, fully specified sequence L_H* may simply be part of a phonetic realisation of !H*. This is not to suggest, as in Pierrehumbert's original (1980) analysis, that an intervening L tone triggers downstep, rather than when there is a downstepped H*, it requires preceding anticipatory lowering.

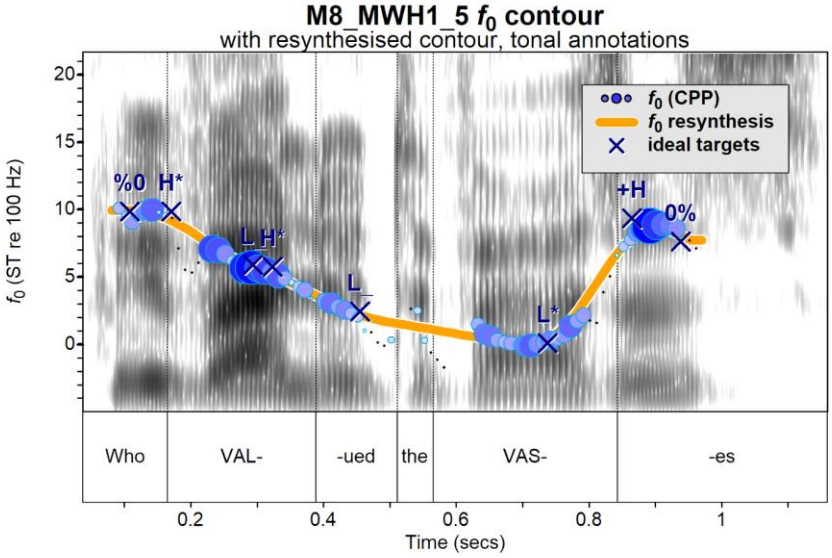


Figure 8.11 K-Max analysis of a contour with downstepped PN !H.*

8.6.3 Anticipatory Lowering

Anticipatory lowering seems to be a common feature in almost all contours before the nuclear L*H as seen in every example in this section. It was suggested in 8.3 that the purpose of anticipatory lowering is pragmatic, in that it signals listener to attend to the upcoming nuclear pitch accent and the semantic content associated with it. However, Figure 8.12 indicates an alternative function. In the original analysis, the turning point in *HIDD-* was interpreted as an L*. This was largely because it seemed to signal some kind of prominence, and there was clearly a tonal event associated with it. L* may well still be an appropriate interpretation. That said, an alternative analysis is that the scope of the nuclear pitch accent extends back to the start of the second foot, so that the speaker intends the listener to parcel [HIDDEN THE VALUABLES] as a single unit, which is the topic of the sentence.

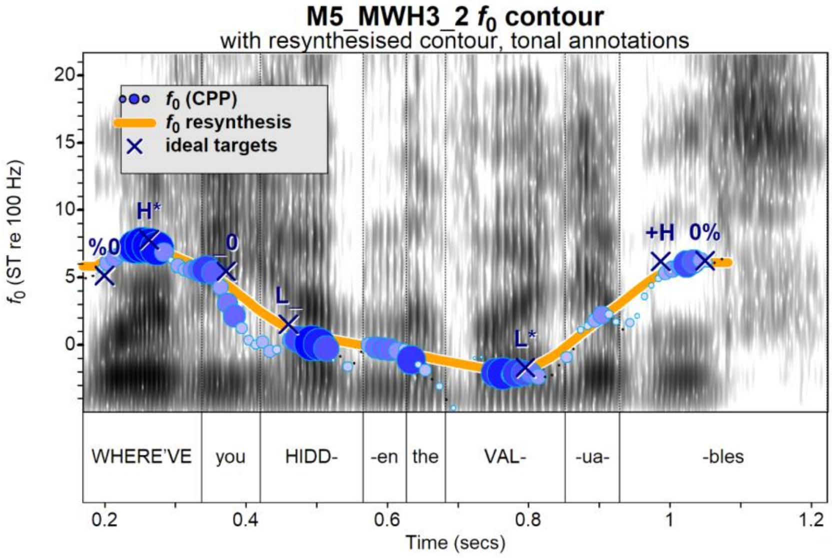


Figure 8.12 K-Max analysis of a contour with anticipatory lowering before the nuclear pitch accent.

8.6.4 Tonal Alignment L target in Declarative questions

The final illustration of the Phonetics-first approach relates to high register in the nuclear pitch accent of declarative questions. Section 8.3 presented an example of raised register where the raised register was not fully in effect at the site of anticipatory lowering (L_) but was at the site of the tonal target of the starred tone. This meant that L* was scaled higher than L_, leading to the likely incorrect impression that the tonal target itself was aligned earlier (see Figure 8.5).

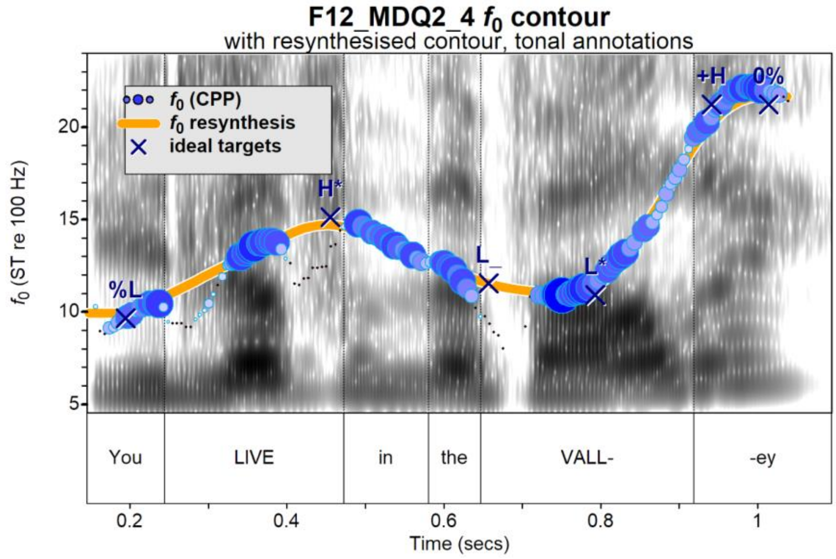


*Figure 8.13 K-Max analysis of illustrating L tonal target of nuclear L*H in a declarative question.*

There is another effect of raised register which becomes apparent in the Phonetics-first approach when we conceptualise tonal targets as covert targets which are never fully present in the contour. Figure 8.13 illustrates this. Unlike the previous example of raised register discussed in this chapter, we see that in this example, the L_ turning point is scaled slightly higher than the L*, and that the f_0 minimum is located at the start of the vowel in the stressed syllable (around 0.57 seconds). The L* target, however, is located (once again!) at the right edge of the vowel in the stressed syllable. It seems that due to the increased size of the f_0 excursion, the trajectory of the contour is redirected earlier than normal in order to begin approximating the next target. As such, while the predicted location of the covert L* tonal target remains in its prototypical location at the right edge of the vowel, this is not immediately transparent when one does not employ a turning-point analysis of tonal targets.

8.7 Summary

This chapter drew attention to what I consider a weakness in the analyses conducted in Chapters 6 and 7, namely that they involved a Phonology-first approach which worked backwards from assumptions about the identity of pitch accents and boundary tones. This led to the use of f_0 minima and maxima as the main identifiers of tonal targets in the f_0 contour. This critique does not invalidate the findings, rather, it leads to consideration of an alternative approach which may clarify or offer further insight into the main analytical approach.

Some of the issues which were unresolved in the analyses in Chapters 6 and 7 were highlighted, and it was demonstrated how an analysis of tonal targets in terms of turning points may be more fruitful than an analysis of f_0 extrema as tonal targets. It was also observed that, if one recognises the distinction between tonal targets as (static) locations in the f_0 contour and tones as pitch events with inherent duration, it becomes easier to explain plateaux and valleys as continuations of the H tone or anticipatory lowering before an L tone. Finally, it was also noted that this interpretation shares some common ground with secondary association and double alignment analyses but is not the same.

This alternative approach is called a Phonetics-first approach, because the analytical procedure involves an initial analysis of the contour as a minimal sequence of turning points which facilitate the resynthesis of a perceptually equivalent f_0 contour. It was noted that the Phonetics-first approach borrows from the IPO analytical approach in its use of perceptually equivalent contour resynthesis. It was also noted that it takes its understanding of tonal targets from the PENTA models, in which the tonal target is a covert target which is only approximated but never achieved because physiological constraints demand that the f_0 trajectory must be adjusted to prepare for the subsequent tonal target. However, despite these sources of inspiration, it was pointed out that the Phonetics-first approach adopted here is still an AM approach, most notably since it assumes that the intonation, at heart, involves the phonetic implementation of a sequence of underlying H and L tones.

The rationale for why turning points can be associated with L and H tones (and tonal targets) was outlined along with the method for identifying tonal targets in the contour. This included a brief description of the Praat plugin (K-Max) which was written specifically to facilitate this analysis.

Examples of the Phonetics-first analytical approach conducted using K-Max were presented. The first demonstrated how the approach might be used to help resolve the ambiguity of the >H* label adopted in Chapter 6, wherein the turning point analysis can help identify H* and L*H pitch accents which have been supressed or obscured during phonetic realisation. It also presented an example to demonstrate that downstepped H* can be analysed as an L_H* sequence in which the L_ represents anticipatory the lowering required for downstep to occur, but which is not a trigger for downstep.

The concept and function of anticipatory lowering of f_0 before a nuclear L*H was also considered. It was argued that it serves a pragmatic function which aims to signal to the listener when to shift attention to the semantic content associated with the nuclear pitch accent and also to identify the scope of the nuclear pitch accent. That is, it is used to help the listener package the informational content of the IP as the speaker intends. This view of anticipatory lowering, it should be noted, may be specific to northern Irish English, where nuclear L*H is the unmarked form.

The final example illustrated how an understanding of the tonal target as a covert rather than overt tonal target may explain the apparent earlier tonal alignment of nuclear pitch accents in MDQs. In fact, it was noted in all the examples that the tonal target of the starred tone, when viewed as a covert target, is nearly always aligned to the right edge of the vowel. This may suggest that the starred tone is in fact even more stable than suggested by the analyses in Chapter 6.

9 Summary and Conclusion

This thesis set out to analyse the Phonology and Phonetics of intonation in Derry City English, a city in the Northwest of Northern Ireland close to the border with Donegal in the Republic of Ireland.

9.1 Background and motivation

The English spoken in Derry City belongs to the northern Irish English variety, and one of the most characteristic features of which is the dominance of low nuclear rises in unmarked declarative statements (Jarman & Cruttenden, 1976; Lowry, 2001; McElholm, 1986). This is considered typologically interesting, because nuclear falls are dominant across most varieties of English and indeed other languages, and biological evolutionary processes have even been proposed to explain the quasi-universal phenomenon in which unmarked phrases in speech typically end with a lowering of pitch (Gussenhoven, 2004; Ohala, 1983). Even in other varieties of English, where unmarked nuclear rises are attested, they do not appear to dominate to the same extent as they do in nIE (Grabe et al., 2000). The dominance of low rises, therefore, is of intrinsic interest in intonation studies because it is unusual and because it raises questions regarding assumptions about the function of different kinds of contour, particularly in the contrast between rising and falling nuclear contours.

Nuclear rising intonation in northern Irish English has been studied before. However, the vast majority of work has focused on Belfast English (Grabe, Post, Nolan, et al., 2000; Jarman & Cruttenden, 1976; Jespersen, 2018; Lowry, 2002; Sullivan, 2012 inter alia), with a smaller number of studies on Donegal English (Kalaldeh et al., 2009; O'Reilly et al., 2010), and only one on Derry English (McElholm, 1986).

Currently, the dominant approach to the intonational analysis of English is the Autosegmental Metrical (AM) framework, which is adopted here (Gussenhoven, 2004; Ladd, 2008; Pierrehumbert, 1980). To the best of this researcher's knowledge, no previous research on Derry City English has been conducted using the AM approach until now. Therefore, one central aim of this thesis was to provide a phonological and phonetic description of Derry City intonation within this framework in order to contribute the current literature on nIE and intonation.

A second aim of the thesis was theoretical in bent. Within the AM framework, intonation is viewed as a linguistically structured phenomenon which is represented in the underlying phonology as a sequence of low and high tonal primitives (L and H). These tones form part of an autonomous string of tones, which are then linked to the segmental string through association with different metrical features (such as the stressed syllable or the edge of an intonational phrase). The underlying tones are then realised in the pitch contour through a series of implementational rules. At the same time, there are competing factors which contribute to changes in pitch. For the current study, the most important of these was the role that the paralinguistic component of communication plays on the realisation of pitch. This refers to non-linguistically structured effects which cause gradient changes in the scaling of f_0 (the acoustic parameter most closely associated with pitch). Such effects have been described in terms of biological codes, and it has been proposed that they are the evolutionary source of linguistically structured, intonational phonology and of the categorical tonal primitives which it employs (Gussenhoven, 2004; Ohala, 1983).

It has been argued that, in the realisation of different kinds of sentence modes, the phonology of intonation typically offers at least two meaningfully contrastive pitch accents (H*L vs L*H) which can be realised in the pitch contour to distinguish between statements on the one hand and binary questions on the other, while at the same time the paralinguistic use of pitch leads to a gradient increase in scaling as the number of linguistic—i.e., lexical, morphosyntactic, and phonologic—markers of interrogativity decreases. This raises a question about the linguistic and paralinguistic use of pitch in nIE. If nIE speakers typically lack an H*L/L*H contrast, does this mean that they must rely solely on non-linguistic pitch cues to help signal and distinguish between statements and interrogatives? It struck this researcher (who is a speaker of northern Irish English) as odd that nIE and nIE speakers would appear to lack access to a large chunk of the Grammatical system which is used to signal a fundamental contrast in speech, i.e., the difference between a statement and a question. Yes? Yes.

It was proposed that nIE speakers exploit phonological contrasts marked on another phonological tier (the register tier) to signal the same kind of contrast which speakers of other varieties of English typically signal using the tonal tier. This view is uncommon in AM intonation studies, however, and it has been argued that all changes in register can be accounted for via the events in the tonal tier (Gussenhoven, 2004).

9.2 Research Aims and Research Questions

The description and analysis of Derry City English was divided into two parts. The first part focused on the phonology and phonetics of intonation in unmarked declarative statement under different metrical conditions and with variation in the location of lexical boundary. The second focussed on the role of intonation in signalling different kinds of sentence mode, i.e., statements, wh-questions, yes-no questions, and declarative questions. In this way, there was a focus on some formal aspects of the intonation and also on a functional component of the intonation. Chapter 6 dealt with the formal aspect, while Chapter 7 looked at the functional aspect.

Each chapter has a theoretic aim as well as a descriptive aim. In Chapter 6, it was proposed that the analysis of variation in pitch accent types and their phonetic realisation might indicate if there was something “special” about H tones. In Chapter 7, the theoretical aim was to assess the evidence for a register tier.

In this way, there were four core research questions to answer:

RQ1. What are the formal phonological and phonetic characteristics of pitch accents in DCE in unmarked speech as an effect of changes in metre (anacrusis and foot size) and lexical word boundaries?

RQ2. What are the phonological and phonetic characteristics of the intonational phrase and the nuclear pitch contour in DCE as an effect of different sentence modes?

RQ3. Is there evidence in DCE for the special status of H tones?

RQ4. Does a register-tier analysis provide a plausible explanation for phonetic variation across sentence modes in DCE?

9.3 Materials and Methods

To facilitate answering these questions, a set of read speech stimuli was developed, recorded, and analysed. Utterances (*n* = 1427) from eleven speakers (6F, 5M, mean age = 40, SD = 9.9) of Derry City English were used for the analysis. Each utterance was annotated in Praat (Boersma & Weenink, 2022), and a phonological analysis was conducted using (with a few adjustment) the IViE labelling system (Grabe, 2001; Grabe & Post, 2002). Count data were analysed with the aid of data visualization, while Bayesian generalised linear mixed-effects models (BGLMMs) were used to for inferential statistical analysis of the phonological data, and linear mixed-effects models (LMEMs) were used for inferential statistical analysis of the phonetic data.

9.4 Analysis of Metrical and Lexical Effects on Intonation

A phonological analysis of prenuclear (PN) and nuclear pitch accents (PAs) under different metrical and lexical effects yielded several findings (Chapter 6).

First of all, L*H was the only nuclear PA found in the corpora (A and H) analysed. In fact, if we include the M-Corpus, which was used for the analysis of sentence mode, there were only 13 out of 1427 nuclear PAs (0.9%) which might be classified as something other than L*H, and only two instances of a distinct H* (0.14%). The almost complete absence of nuclear falls (H*L % or H* L%) mirrors one of the findings of McElholm in his 1986 study following the British Tradition. That is, he compared his two-speaker analysis to the results a single speaker analysis by Jarman and Cruttenden (1977), noting that while nuclear falls were completely absent in his data, they were not so in the Belfast data. In this analysis, L*H was found to be similarly much more dominant in declaratives than in the comparable IViE analysis of Belfast English (Grabe, 2004).

A small number of low final boundaries (L%) were also attested, creating an L*H L% nuclear contour. The vast majority of these were produced by a single speaker. While the analysis of metrical and lexical effects was formal and not functional in scope, it was noted that the L% seems to add a pragmatic discourse function to the utterance. Namely, it appears to add the sense that the speaker is implying the idea, "…and I thought you knew that already." In other words, they seem to be suggesting that the answer to the question they are responding to should already be obvious to the interlocutor. To put it in a more technical way, the speaker is implying that the propositional content of their response should already be understood as *given* within the discourse context.

The analysis of PN pitch accents found some variation in PA type, although L*H was by far the most common (71%), followed by H* (17%), and finally by L* (7%). There were also a few instances of non-accentuation in syllables where accentuation was possible (3%). A fourth type of pitch accent was also identified, labelled >H* (7%) (not included in the IViE labelling system). This was a pitch accent with an audibly later peak than the typical H* which also lacked the distinct percept of a rise from the lexically stressed syllable associated with an L*H. As such they were somewhat ambiguous between L*H and H* but were not labelled as one or the other since that felt like it would be shoehorning ambiguous data into categories which they did not readily fit.

BGLMMs were conducted using the A- and H-Corpora data (n = 788). These indicated a significant effect of foot size, the location of the right word boundary, and speech rate on the likelihood of L*H in a prenuclear pitch accent. That is, as foot size increases, so too does the probability of L*H, although this appears to reach saturation point at 81% probability once the foot is three syllables long, OR = 10.5, 95% CI [2.8, 39.4], p < .001. Similarly, as long as the word-final syllable is later than the stressed syllable, the probability of PN L*H jumps from around 29% [2, 89] to 84% [18, 99]. As speech rate increases, the likelihood of L*H decreases, OR = 0.38 [0.24, 0.63], p < .001. The inverse of these results was found for the likelihood of H*, although the probability of H* was lower, at 9% [1, 48] at the intercept compared to L*H's 29% [2, 89]. Moreover, foot size was found not to be a significant predictor of H* nor was speech rate, although the general trend was still the opposite of that found for L*H, i.e., likelihood of H* increases with speech rate and decreases in longer feet. Anacrusis was also a significant factor on the likelihood of H*, with H* becoming extremely unlikely once there were at least two syllables of anacrusis before the first stressed syllable, OR = 0.04 [0.002, 0.68], p = .026.

The statistical analysis of PN phonology indicated two things central to the research questions. Firstly, it reinforced the dominance of L*H nuclear position of neutral declaratives. This confirmed that no L*H / H*L (or H* L%) phonological contrast would be available to distinguish between different sentence modes, so if a phonological contrast were to be found it would likely originate from a register-tier contrast. Secondly, it showed that, given a large enough foot (or planning time during anacrusis), speakers were much more likely to use L*H pitch accents. When they deleted a tone, it was vastly more likely to be the L tone than the H tone, suggesting that there is greater pressure to retain the H tone, and thus suggesting that the H tone has a privileged status in the phonology. As such, it provided the first piece of evidence for an answer to research question three, in that it offered phonological evidence for the special status of H tones.

The phonetic analysis of metrical and lexical effects on pitch accents comprised two parts. The first was an analysis of peak alignment in prenuclear L*H pitch accents in relation to the right-word boundary of the lexically stressed syllable using both syllable-normalised and grand syllable-mean normalised time. The second was a set of LMEM analyses of phonetic parameters associated with tonal targets—i.e., f_0 scaling and alignment—and truncation/compression of L*H slopes—i.e., f_0 excursion size and the slope of the linear regression of f_0(t).

The analysis of PN L*H peaks looked at utterances from the H-Corpus (n = 312, n_{L*H} = 250). It indicated that when the word boundary was later, the peak also tended to be aligned later. Both the syllable-normalised and grand syllable-mean normalised time analyses indicated three typical anchor points for the peak, namely, the end of the stressed syllable, the middle of the word-final syllable, and the end of the word-final syllable. The analysis also indicated a bimodal (or possibly multi) distribution for the alignment of the f_0 peaks, which was noted both as an intraspeaker and interspeaker phenomenon. This suggests that peak alignment is a matter of speaker-choice rather than an obligatory outcome of an implementational rule.

Through the LMEM analysis of PN and nuclear tonal targets, it was found that L targets were remarkably stable under lexical and metrical variation, both in terms of f_0 scaling and alignment. In PN position, there did appear to be a very slight effect of anacrusis on L alignment in the prenuclear pitch accent, where the appearance of anacrusis led to marginally earlier alignment of the target; however, when the absence/presence of anacrusis was tested as a factor, it had a very minor effect which was not found to be statistically significant. That is, in the presence of anacrusis, the L target was aligned an estimated 4 ms later, 95% CI [-5, 12.4], $p = .432$. The only significant formal effect on the timing of the PN L target was the location of the word-end syllable. In other words, if the word-end syllable is not the same as the stressed syllable, then the L target is aligned later, by an estimated 27 ms 95% CI [16, 37], $p < .001$.

The timing of prenuclear peaks was strongly affected by changes in anacrusis and foot size, although this was found only to be true of H targets in L*H pitch accents. Here, in the absence of anacrusis, the H target is aligned much later than in the presence of a single syllable of anacrusis. The addition of anacrusis, it seems causes the peak to be aligned significantly earlier, -68 ms, 95% CI [-85, -51], $p < .001$, but then as more anacrusis is added, the peak begins to drift slowly rightwards again. With foot size, the main effect appears to be that as the foot gains more syllables, the peak drifts rightwards until it stabilizes at around 251 [211, 290] ms once the foot is three syllables long. However, as the alignment in the one-syllable foot was noticeably earlier than any of the other conditions, it was suspected that this was really an effect of stress clash, with backward pressure on the prenuclear pitch accent, causing it to be aligned earlier to make way for the nuclear L target. Two supplementary models indicated that this was the case, with the presence of stress clash leading to an estimated 54 [-75, -33] ms, $p < .001$, earlier alignment of the prenuclear H target, with f_0 lower by an estimated 0.9 [-1.49, 0.28] ST, $p = .006$. In fact, stress clash and, by association, foot size were the only significant metrical or lexical effects on the scaling of the PN H target.

The temporal alignment of the nuclear peak was strongly affected by metrical effects, i.e., by foot size, the number of preceding unstressed syllables, and the presence of a word boundary at the start of the foot. However, while all of these were statistically significant, the actual magnitude of the effect was often quite small. The presence of a new word seems to cause an average 12 ms earlier alignment of the H target, 95 % CI [-20, -5], $p < .001$, which is relatively small and also quite hard to explain. The effect of the preceding unstressed syllables is basically the effect of stress clash, with the H peak being aligned at an estimated minimum of 34 [-41, -26] ms earlier once the stress class condition is removed, p. < .001. The addition of preceding unstressed syllables causes the peak to be aligned slightly earlier again, but the main effect relates to the presence or absence of stress clash.

As with the L target, peak f_0 is also scaled slightly higher in the absence of stress clash, by at least 0.8 [0.4, 12] ST, p <.001, so the apparent significant effects of preceding syllables in the nuclear pitch accent are really just stress clash effects in disguise.

The most dramatic effect on the nuclear H target, however, is that of foot size. Once the foot is more than two syllables long, the timing of the peak increases dramatically (by an estimated 106 ms followed by another rise of 132 ms, 95% CIs [94, 117] and [118, 146], p. < .001 in each case). When there is only one syllable in the nuclear foot, f_0 is scaled an estimated 1.4 [0.8,.2] ST lower than the two-syllable condition, p < .001, although there is almost no difference in scaling across the two,- three-, and four-syllable conditions.

The strong effect of foot size on nuclear pitch accent parameters appeared to go against the prediction that nuclear tonal targets would be less vulnerable to metrical effects than those in the prenuclear pitch accent. However, it occurred to me that perhaps H targets were more stable if measured as a proportion of the foot. This would be the case, especially if the domain of the pitch accent was the foot itself, and so the H target changes in proportion to the size of the foot. When this was tested in the PN and nuclear pitch accents, nuclear pitch accents were found to be more stable, but the results for prenuclear pitch accents were essentially the same as the initial results, but with scales measured in percentages rather than milliseconds. The alignment of the nuclear peak was then measured as a proportion of the voiced material in the foot, and in this case, the proportional alignment of the peak was remarkably stable, hovering at an estimated 75% of the voiced material in the one-, three-, and four-syllable syllable conditions, 95% CI [.60,.91]. The two-syllable condition was aligned at 78% [.63, .93] of the voiced material, which was a significant but small difference from the others. Therefore, it seems that, when considered proportionally, the peak alignment of the nuclear H target has a degree of stability absent in the prenuclear H peak alignment.

The effect sizes of metrical and lexical effects on prenuclear tonal target parameters were compared. That is, the effect sizes of foot-initial word boundaries and foot size could be compared between the nuclear and prenuclear target, as too could the effect size of anacrusis on prenuclear tonal targets and of preceding unstressed syllables on nuclear tonal targets. Excluding foot size, the metrical and lexical effects were greater on PN pitch accents than on nuclear pitch accents, and even when foot size was included, the average effect size on PN pitch accents was greater. Therefore, overall, in terms of the alignment and scaling of tonal targets, PN pitch accents do appear to be more susceptible to metrical and lexical effects than nuclear pitch accents.

Truncation and compression were evaluated by plotting the tonal target estimates on a two-dimensional plot, and also via LMEMs of the effect of foot size on f_0 excursion and the slope of the linear regression of f_0(t) between L and H targets of L*H. These analyses indicated very strongly that in the one-syllable / stress-clash condition, the prenuclear L*H rise is heavily truncated, but that in the nuclear pitch accent there was evidence of both truncation and compression. That is, in the nuclear pitch accent, the L*H rise in the three-syllable foot was compressed relative to the four-syllable foot. The rise in the two-syllable foot was then truncated in comparison with the three-syllable foot (but not compressed), and so again was the one-syllable foot in comparison to the two-syllable foot. If we assume that the compression here is really just an effect of the concertinaing of the rise while H is being kept at a proportionally stable location in the voiced portion of the foot, we might interpret the compression as a strategy aimed to maintain this proportionality. However, when it becomes physiologically too challenging to compress the rise any further, compression is replaced by a truncation strategy to retain the relative position of the peak in the foot.

It was observed in Chapter 4 that an AM description of the phonology and phonetics of pitch accents in DCE under different lexical and metrical effects would facilitate a comparison with other varieties of nIE, specifically with Donegal English and Belfast English. The phonological analysis suggests that there is a much lower likelihood of H* or H*L in nuclear pitch accents in Derry City English when compared with the IViE study of Belfast English, which parallels the McElholm's (1986) findings compared to those of Jarman and Cruttenden (1976). This makes the almost exclusively L*H nuclear PA in DCE very similar to the findings on Donegal English (Kalaldeh et al., 2009; O'Reilly et al., 2010). However, in the Donegal studies, prenuclear pitch accents were also exclusively L*H. In the current study, however, while L*H dominates, there are also H* and L* pitch accents, along with the >H*, which is not comparable with the Donegal or Belfast studies. However, it is clear that the larger inventory of prenuclear pitch accent in the DCE data is more similar to that found in the Belfast studies (Grabe, 2004). Thus, DCE shares obvious similarities with both Belfast and Donegal in its distribution and use of L*H, but there is some overlap in DCE with Belfast English in the variety of PN pitch accents, and some overlap with Donegal in the much greater use of L*H in the nuclear position. This is apt for a city which shares a border and historical ties with Donegal, and a political and governmental infrastructure with Belfast.

Another similarity found between this study of DCE and previous research on Belfast English (Sullivan, 2007) is that the alignment on the nuclear L target is affected by both gender and the number of preceding syllables. In contrast to this similarity, there was a small difference between the findings of the current study regarding gender and a study by Lowry (2011). She found that men were much more likely to use L*H nuclear pitch accent than women across different styles of speech in her Belfast data. In the current study, the only speaker who does *not* exclusively use nuclear L*H, is male. However, these non-L*H tokens are so rare that it is difficult to generalise about them.

Overall, the results of the phonetic analysis tell us that the H target in nuclear L*H is more vulnerable to variation from metrical and lexical effects than the starred L target, but that a combination of compression and truncation strategies is available to ensure that alignment of the H target remains relatively stable in proportion to the duration of voicing in the foot. In the prenuclear pitch accent, while the starred L target is phonetically more stable than its trailing H counterpart, it is also more apt to be deleted in shorter feet, when the stressed syllable is also the word-final syllable, or when there is less anacrusis preceding the first foot. At the same time, in order to preserve the H tone in PN L*H pitch accents, there appears to be a truncation strategy available. That means that, even though the rise of the L*H is shortened in a PN before a starred nuclear L as an effect of stress clash, there is still a strategy available that allows it to remain rather than having it deleted altogether. Therefore, we see that on top of the preferential phonology status of the PN H target (i.e., it is less likely to be deleted than the L target), there is also a phonetic strategy (i.e., truncation) available to ensure that the H tone can be retained even when there is pressure for it also to be deleted.

Taken together, these two features, one phonological, one phonetic, suggest that even in Derry City English, where L dominates as the starred tone, privileged status is accorded to the H tone. Therefore, in answering research question one by describing the phonological and phonetic characteristics of pitch accents in unmarked declaratives under the differing metrical and lexical (boundary) effects, we also have an answer for research question two. Is there evidence in DCE for the special status of H tones? Yes, there is.

9.5 Analysis of Intonation and Sentence Modes

One of the central research aims of the analysis of intonation across sentence modes was to test the plausibility of a phonological register-tier. To achieve this, the central strategy for the phonological component was to compare a non-register-tier analysis of the data with a register-tier analysis. This led to a methodological dilemma early in the labelling process because it became clear that it was practically impossible to label the pitch accents using a non-register tier analysis without ignoring all the auditory and visual cues which strongly indicated that there was indeed a register tier. Therefore, a register tier component was added to the labelling system, wherein high register was labelled using the caret (^) symbol and the extent of high register indicated by the use of square brackets. The reason for the bracketing was due to the fact the high register appeared to apply over a range of constituent domains, from individual tones, to pitch accents, to the whole IP. In order to generate the non-register tier analysis, a script was written which converted register-tier labelling back into its most reasonable non-register tier alternative. It is acknowledged that this swung the odds in favour of the register-tier analysis, but it was also judged that it would have been more misleading to use non-register-tier labels requiring the reassignment of tones to structural units to which they in no way appeared to be associated with (e.g. reassigning trailing tones to the boundary), or even worse, to re-interpret L tones as H tones to make the labelling "fit" better.

9.5.1 Phonological Analysis of Mode and Intonation

Based on intuitions on the function of L% from the analysis of the A- and H-Corpora, it was predicted that the use of L% would be more common in question forms, particularly MYN and MDQ, but not because it signalled interrogativity. Rather, it was proposed that it would serve the same function as before. That is, it would indicate a conflict between what the interlocutor has just said and what the speaker has believed to be shared knowledge until that point (or assumed was *given* within the discourse context). For example, in response to the statement from their interlocuter, "I live in the valley," the speaker may show surprise by saying, "You live in the valley [L%]?" In this case, the L% would imply, "I'm surprised. That's not what I thought."

A larger proportion of L% was found the M-Corpus than in A- and H-Corpora used for the analysis of lexical and metrical effects. L% was least common in MWH questions, with increased use in MYN, and was most common in MDQ (roughly 39% of all MDQs). This confirmed the expectations about the use of L%. A BGLMM analysis of the likelihood of L% as an effect of gender and mode did find a significant effect of mode, LRT(3) = 62.8, *p.adj* < .001, with L% being slightly over 10 times more likely in MDQ than MDC, OR = 10.1 [4.7, 21.6], *p* < .001. The predicted probability of L% in MDQ among female speakers was 23% [.01, .86], and only 16% [.01, .84] among male speakers. So even though L% was statistically more probable in MDQ, the estimated mean probability was still less than 25%.

With very few exceptions (two H* and five >H* tokens, all from the same speaker) all nuclear pitch accents were of an L*H type (raised or unraised). There were two declaratives with H* pitch accents ending with L%. These were the only two H* L% tokens found in 951 declaratives analysed in this thesis.

Because of the lack of phonological contrast in pitch-accent type, the only phonological means available to help signal interrogativity in a non-register understanding of the phonology is via the boundary tone H%. A BGLMM analysis of the likelihood of H% as an effect of mode and gender indicated that mode was a statistically significant effect (LRT(3) = 65, *p.adj* < .001). While H% was over sixty times more likely in MDQ than in MDC (OR = 63.3, 95% CI [509, 682]m *p* < .001]), the predicted probability of H% in MDQ was still only 13% [0.02, 0.5]. In the register-tier analysis, there were no H% boundaries, and all boundaries were interpreted either as unspecified (%) or as L%. In the analysis of the count data, however, occurrences of raised register in nuclear pitch accents—^[L]*H, L*^[H], or ^[L*H]—increased noticeably in MYN and even more so in MDQ. ^[L]*H was quite rare (*n.adj* = 6) and was only used by one person, the same speaker who produced the >H* and H* nuclear pitch accents. ^[L*H] was by far the most common raised register type, accounting for almost 46% of MDQs (adjusted).

A BGLMM analysis of the likelihood of high register as an effect of sentence mode found that high register is extremely unlikely to occur in the nuclear pitch accent of either MDC or MWH, but was more likely to occur in MYN, and most likely to occur in MDQ. There was also a strong effect of gender, with male speakers much more likely to use high register.

The relationship between mode and high register in the register-tier analysis was compared with the relationship between mode and H% in the non-register-tier analysis using two new models which removed gender as an effect. These found that mode explained a greater amount of the variance in the register-tier model than in the non-register tier model (r_m^2 = .63 and .33 respectively). This indicates that the register-tier approach provides a superior explanation of the relationship between mode and the phonology.

An utterance-wide analysis of the phonology revealed a large number of nuclear-PA-only IPs, and a BGLMM indicated that this is strongly associated with sentence mode, most likely as a strategy to help make the rise of the nuclear pitch accent in MYN and MDQ more salient, thus reinforcing the interrogativity of the utterance. While register-raising was more common among the male speakers, the nuclear-PA-only strategy was more common among females.

When these two phonological strategies were combined into a single parameter, a BGLMM analysis indicated that the likelihood of at least one of the two strategies occurring in MDQ was 550 times more likely than in MDC, 95% CI [64, 2723], $p < .001$, with an overall probability of either strategy occurring in MDQ of 87% [.5, .98]. Therefore, it seems that the DCE speakers in this study have two phonological strategies for reinforcing interrogativity in YNQs and DCQs, namely, register raising in the pitch accent and pitch-accent avoidance before the nuclear PA. The former is preferred by male speakers and the latter by female speakers. Without a register-tier analysis the gender-differentiated nature of these phonological strategies would have gone unobserved.

9.5.2 Phonetic Analysis of Intonation and Mode

The phonetic analysis compared two kinds of model to evaluate the extent to which a phonological register-tier analysis might account for gradient changes in f_0 scaling in nuclear L*H across sentence modes. The first was a mode-only model, using mode and gender as fixed effects, while the second was a mode-plus-phonology model, which added phonological fixed effects. L and H tonal target parameters (tonal alignment and f_0 scaling) were then analysed in both models.

A mode-only analysis of the tonal targets of the nuclear pitch accent indicated that there was no difference in the f_0 scaling or alignment of MDC and MWH; however, it indicated a clear gradient increase in scaling from MWH to MYN and from MYN to MDQ. The only change in tonal alignment was that MDQ was aligned slightly earlier. As such, the mode-only analysis conformed with expectations about the paralinguistic effects of mode, i.e., that as the number of morphosyntactic markers of interrogativity in the utterance decreases there is a gradient increase in the scaling of f_0 targets. In the mode-plus-phonology models, gradient changes in f_0 scaling as an effect of mode were reduced; however, they did not disappear completely.

Two global f_0 parameters were analysed, utterance mean f_0 and utterance-wide f_0 slope. Again, two types of model were tested, one with mode-only fixed effects, and one which added phonological fixed effects. As in the analysis of the nuclear tonal targets, phonological effects were evident in the second type of model, but here the effect of mode was greater. This was considered to be due to the fact that the majority of the phonological effects of sentence mode are concentrated in the nuclear pitch accent as it carries most of the communicative weight of the IP, so when we look across the whole utterance, phonological effects will inevitably become diluted, and gradient paralinguistic effects of mode will be comparatively greater. The analysis of the global contour parameters reinforced the conclusion that both gradient and categorical effects operate on the f_0 contour in the signalling different sentence modes.

The phonetic characteristics of the four variations of L*H were also analysed (i.e., L*H, ^[L]*H, L*^[H] and ^[L*H]). It was found that the f_0 scaling and temporal alignment of L* was very similar in both L*H and L*[H], as too were those of ^[L*] in ^[L*]H and ^[L*H]. Furthermore, f_0

scaling was categorically distinct between L* and ^[L*]. f_0 scaling was also found to be almost identical in the trailing H tones of L*H and ^[L*]H and was categorically distinct from their raised-register counterparts. However, the H tone of ^[L*]H was aligned earlier, giving the impression that ^[L*]H may more appropriately have been analysed as a variant of H* or even >H*. High-register H targets (those in L*^[H] and ^[L*H]) were also remarkably similar both in alignment and scaling, and both were categorically distinct from their low-register counterparts. These results provided further evidence that the interpretation of the tonal target as a combined manifestation of tone and register tier is appropriate.

The consistency in the scaling of high and low register alternatives of L and H tonal targets was very gratifying as it lent further support to the register-tier hypothesis. As such, in relation to research question four—"Does a register-tier analysis provide a plausible explanation for phonetic variation across sentence modes in DCE?"—it seems that there is both good phonological and phonetic evidence for a register-tier analysis. In adopting several different strategies to evaluate the potential effects (or non-effects) of the register tier, I believe that the second research question has also been answered. i.e., the analysis has provided a comprehensive description of the phonological and phonetic characteristics of intonation across sentence modes in DCE.

9.6 Critique and Alternative Approach

In reflecting on the approach of the main body of research in this thesis, it seemed that, although it was a study of both the phonology and the phonetics of intonation in Derry City English, it adopted a broadly Phonology-first approach. That is, it worked from assumptions about the phonological identity of pitch accents and boundary tones.

An alternative approach was proposed, a Phonetics-first approach. This involves a first pass analysis of the contour in terms of a minimal sequence of turning points which facilitate the resynthesis of a perceptually equivalent f_0 contour. In this approach, the turning points identified are viewed more appropriate indicators of tonal targets. The importance of the distinction between tonal targets as (static) locations in the f_0 contour and phonetic tones as pitch events with inherent duration was noted, and it was also noted that not recognising the importance of this distinction was a weakness on part of this researcher. Through the use of turning points and a recognition of the distinction between phonetic tones and targets, it becomes easier to explain a plateau as a continuing H tone or and a valley as the anticipatory lowering of f_0 in lead up to an L target. Similarly, secondary association and double alignment (Beckman & Pierrehumbert, 1986; Gussenhoven, 2004) were recognised.

A system (Rodgers, 2020) was developed to work as a plugin in Praat (Boersma & Weenink, 2022) to facilitate the Phonetics-first approach. It adopted the IPO analytical approach of generating perceptually equivalent contour resynthesis ('t Hart et al., 1990), although with an emphasis on turning points rather than glides. It used an understanding of tonal targets adopted from the PENTA model (Xu, 2004; Xu et al., 2022), in which the tonal target is viewed as a covert target which is only approximated in the f_0 contour because physiological constraints demand that the f_0 trajectory must be adjusted to prepare for the subsequent target. Despite these sources of inspiration, I believe that the Phonetics-first approach is still fundamentally an AM approach, most notably since it assumes that intonation, at heart, involves the phonetic implementation of a sequence of underlying H and L tonal primitives.

Several examples of the Phonetics-first approach were provided which, it was hoped, would indicate its potential benefits and also shed light on issues raised during the two central analyses. Looking at two examples of prenuclear >H* pitch accents, it was demonstrated that the turning point analysis can help identify a distinction between H* and L*H pitch accents which have been obscured during phonetic realisation. It also presented an example showing how downstepped H* can be analysed as an L_H* sequence, where the L_ represents anticipatory lowering before the downstepped !H*. It was noted that, unlike Pierrehumbert's original (1980) interpretation of downstep, the view here is that an anticipatory L is required for downstep but does not trigger it.

The function of anticipatory lowering of f_0 before a nuclear L*H was also considered, with an argument that it serves a pragmatic information packaging function. That is, it may signal to the listener to shift attention from the preceding semantic content to the semantic content associated with the nuclear pitch accent. As such, anticipatory lowering (L_) helps identify the scope of the nuclear pitch accent. Because L*H is the unmarked nuclear pitch accent in nIE but not the norm in other varieties of English, it was observed that anticipatory lowering of this kind may well be specific to northern Irish English.

The final example proffered indicated how the approach to identifying tonal targets can help explain why the alignment of L targets in nuclear L*H pitch accents appears to be earlier in MDQs. That is, when viewed as a covert target, L was nearly always aligned to the right edge of the vowel. In the case of register raising and in cases where there is a sharp f_0 rise in the nuclear pitch accent, the trajectory may turn away from the covert target earlier, giving the surface impression of an earlier target.

9.7 Relevance and Contribution to the Field

In AM studies, it is common to evaluate differences in phonetic realisation of different pitch accents and how they are affected by changes in metrical, lexical, and segmental structure. However, in Chapter 6, it was shown that there are also phonological effects, with a shorter foot and early word-boundary being associated with L deletion and PN L*H being reinterpreted as H*, or the ambiguous >H* (although this last type, as indicated in Chapter 8, may be a phonetically disguised H* or L*H).

Secondly, there is evidence of a pitch accent hierarchy, in which the nuclear pitch accent is prioritised over the prenuclear pitch accent. For example, deletion effects were seen to operate on the prenuclear pitch accent but not on the nuclear pitch accent. It was also seen that the rise of PN L*H is likely to be truncated in order to make way for the more important L target of the L*H. While evidence of a pitch accent hierarchy is not new, the results presented here indicate how the hierarchy is likely to affect the stability of tones in the nuclear and prenuclear positions.

Finally, it was found that peak alignment of the nuclear L*H is realised stably as a proportion of the voiced material in the foot (and slightly less stably as a proportion of the foot itself). Such stability is not found in the prenuclear L*H. At the same time, the general 'universal' tendency toward H* nuclei and nuclear falls in declaratives is not reflected in the overwhelming dominance of L*H % in the nuclear pitch contour. Neither of these observations is proposed as novel. However, we see that the H target is prioritised over the L target in prenuclear pitch accent (i.e., the H is much less likely to be phonologically deleted) and that a phonetic truncation strategy is available in the L*H rise which helps preserve the H target, but such a strategy is not available to the L target. Taken together, these indicate, even in this variety of English dominated by unmarked nuclear L*H %, the H target itself appears to have a special status. I believe this in itself is an interesting finding which contributes to the discussion on intonational universals.

The analysis of sentence modes followed an unorthodox path for the analysis of intonation, at least in terms of typical AM studies. That is, it looked for evidence for a register tier. The fact that nIE appears to lack clear phonological binary contrast between falling and rising pitch accents is what primarily facilitated this, with the view that, in the absence of such a contrast, the presence of register tier effects would be more transparent. I believe the analysis proffered in Chapter 7 provided good evidence for the existence of a phonological register tier, and that it demonstrated the plausibility of the register tier as an efficient means of explaining perceptually salient intonational events. Even if one rejects the register-tier hypothesis, the irony of the approach taken is that it actually provides robust support for the view that gradient paralinguistic uses of pitch interact with categorical phonological uses of intonation in communicating meaning. That is, even when the register-tier hypothesis was assumed to be true (in the mode-plus-phonology models analysing the effects of mode), there was still evidence of gradient paralinguistic effects on the f_0 contour.

In terms of the description of the intonation in varieties of English, this thesis has offered a detailed AM account of two key aspects of nIE from a corpus of almost 1500 utterances from eleven speakers from the second largest urban area in Northern Ireland, Derry City. To the best of my knowledge, an account of intonation in Derry City at this scale has never been carried out before. As such, I believe this thesis provides an ultimately small but useful descriptive contribution to the study of intonation.

9.8 Future work

It was mentioned in Chapter 5 that more data was recorded than could be analysed within the scope of this thesis. This includes a corpus of read speech with list-form intonation, a corpus of read speech with broad and narrow focus, and recordings of a personal story-telling task and an interactive spot-the-difference task. These all need attention, and the non-read speech tasks will be particularly useful to analyse in the light of the research described here as they contain spontaneous non-directed speech which will help refine and improve upon the current findings.

Moreover, this thesis has worked within the AM framework, which has allowed the analysis here to be compared with other AM analyses. However, it was not a replication study, so the comparisons were still limited. I would like very much to work on producing a larger corpus of speech aimed at the comparative intonational analysis of different varieties of Irish English, and particularly of Northern Irish English, to see if the findings in this study can be replicated or expanded upon.

Finally, in order to further assess the efficacy of the Phonetics-first approach, I plan to develop the research tool to make it more accessible, flexible, and user-friendly. In this way, it might offer a useful mode of analysis that others can exploit. It will also facilitate a comparison between the turning points and peak/valley alignment approaches to the evaluation of tonal targets. An additional benefit of the Phonetics-first approach is that, because it works to generate parsimonious resynthesized f_0 contours, it is very amenable to the production of perception tests using analysis-by-synthesis. This includes the possibility of user-directed resynthesis, such as in Murphy et al. (2019, 2020). This kind of research will be of great benefit for investigating issues such as the function of L% in nIE or the possible scope-marking effects of high plateaux and anticipatory lowering discussed in Chapter 8.

It seems, therefore, the small number of research questions set out at the beginning of this thesis have generated many more questions begging to be answered, providing a foundation for future work.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*(April 2016), 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Albert, A., Cangemi, F., & Grice, M. (2019). Can you draw me a question? *ICPhS 19*. https://doi.org/10.13140/RG.2.2.15700.14729

O'Connor, G. F., & Arnold, J. D. (1973). *Intonation of colloquial English* (2nd ed.). Longman.

Arvaniti, A. (2022). Autosegmental-Metrical Model of Intonational Phonology. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Prosodic Theory and Practice* (pp. 25–83). MIT Press. https://doi.org/10.13140/RG.2.1.3737.0728

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. *Papers in Laboratory Phonology IX*, 547–576.

Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, *26*(1), 43–74. https://doi.org/10.1017/S0952675709001717

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, *26*(1), 3–25. https://doi.org/10.1006/jpho.1997.0063

Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 119–131.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006a). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*(6), 667–696. https://doi.org/10.1016/j.specom.2005.09.012

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006b). Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech*, *49*(4), 421–450. https://doi.org/10.1177/00238309060490040101

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of '"segmental anchoring"' of F0: evidence from German. *Journal of Phonetics*, *32*(2), 177–197. https://doi.org/10.1016/S0095-4470(03)00039-1

Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge University Press. https://doi.org/10.1558/sols.v2i3.471

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). *Memory and Language Mixed-effects modeling with crossed random effects for subjects and items*. *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barnes, J., Brugos, A., Veilleux, N., & Shattuck-hufnagel, S. (2021). On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics*, *85*, 101020. https://doi.org/10.1016/j.wocn.2020.101020

Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology*, *3*(2), 337–383. https://doi.org/10.1515/lp-2012-0017

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beckman, M. E., & Ayers Elam, G. (1997). *Guidelines for ToBI labelling (version 3.0)*. The Ohio State University Research Foundation.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic models and transcription: Towards prosodic typology* (pp. 9–54). https://doi.org/10.1093/acprof:oso/9780199249633.003.0002

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonation structure in Japanese and English. In *Phonology* (Vol. 3, Issue 01, pp. 255–309). https://doi.org/10.1017/S095267570000066X

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boersma, P., & Weenink, D. (2022). *Praat: doing phonetics by computer (v. 6.2.14)*. www.praat.org

Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford University Press.

Bolinger, D. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse*. Stanford University Press. https://doi.org/10.1515/9781503623125

Bolker, B. (2018). *GLMM worked examples*. http://bbolker.github.io/mixedmodels-misc/ecostats_chap.html

Brazil, D. (1995). *A Grammar of Speech*. Oxford University Press.

Bruce, G., & Thelander, I. (2001). A pitch accent journey in southern Sweden. *Working Papers, Lund University, Department of Lingustics*, *49*, 14–17.

Chung, Y., & Rabe-Hesketh, S. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *1*(19), 331.

Corrigan, K. (2010). *Irish English, volume 1 - Northern Ireland*. Edinburgh University Press.

Cruttenden, A. (1995). Rises in English. In J. Windsor Lewis (Ed.), *Studies in General and English Phonetics* (pp. 155–173). Routledge.

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Cruttenden, A. (2001). Mancunian intonation and intonational representation. *Phonetica*, *58*, 53–80. https://doi.org/10.1159/000028488

Cruttenden, A. (2007). Intonational diglossia: A case study of Glasgow. *Journal of the International Phonetic Association*, *37*(3), 257–274. https://doi.org/10.1017/S0025100307002915

Crystal, D. (1972). The Intonation Systems of English. In D. Bolinger (Ed.), *Intonation* (pp. 110–136). Penguin.

D’Imperio, M. (2000). *The Role of Perception in Defining Tonal Targets and Their Alignment*. Ohio

State University.

Dalton, M., & Ní Chasaide, A. (2005). Tonal alignment in Irish dialects. *Language and Speech*, *48*(Pt 4), 441–464. https://doi.org/10.1177/00238309050480040501

Dalton, M., & Ní Chasaide, A. (2007). Nuclear accents in four Irish (Gaelic) dialects. *ICPhS 16, August*, 965–968.

Di Cristo, A., & Hirst, D. J. (1986). Modelling French Micromelody: Analysis and Synthesis. *Phonetica*, *43*(1–3), 11–30. https://doi.org/10.1159/000261758

Dorn, A. (2006). *An Intonational Analysis of Donegal English (M.Phil disseration)*. Trinity College, Dublin.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage. https://doi.org/10.7326/M16-0968

Foley, M., Gibbon, F. E., & Peppé, S. (2011). Benchmarking Typically Developing Children's Prosodic Performance on the Irish-English Version of the Profiling Elements of Prosody in Speech-Communication (PEPS-C). *Journal of Clinical Speech and Language Studies*, *18*(1), 19–40. https://doi.org/10.3233/acs-2011-18105

Fujisaki, H. (2004). Information, prosody, and modeling - with emphasis on tonal features of speech. *Speech Prosody 2004*, 1–10.

Goldsmith, J. A. (1976). *Autosegmental Phonology*. Massachusetts Institute of Technology.

Grabe, E. (1998a). *Comparative Intonational Phonology: English and German* [Radboud University Nijmegen]. https://doi.org/doi:10.17617/2.2057683

Grabe, E. (1998b). Pitch accent realization in English and German. *Journal of Phonetics*, *26*(2), 129–143. https://doi.org/10.1006/jpho.1997.0072

Grabe, E. (2001). *The IViE Labelling Guide (Version 3)*. http://www.phon.ox.ac.uk/files/apps/IViE/guide.html

Grabe, E. (2004). Intonational Variation in Urban Dialects of English Spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional Variation in Intonation* (pp. 9–31). Niemeyer. http://www.phon.ox.ac.uk/files/people/grabe/Grabe_Niemeyer.pdf

Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles: potential for miscommunication? In K. Dziubalska-Kolaczyk & J. Przedlacka (Eds.), *English pronunciation models: a changing scene* (pp. 311–337). Peter Lang.

Grabe, E., Kochanski, G., & Coleman, J. (2003). Quantitative modelling of intonational variation. *Speech Analysis and Recognition in Technology, Linguistics and Medicine*, 45–57. https://doi.org/10.1.1.97.4270

Grabe, E., Nolan, F., & Farrar, K. (1998). IViE - A Comparative Transcription System for Intonational Variation in English. *ICSLP 1998*. https://doi.org/10.21437/ICSLP.1998-583

Grabe, E., & Post, B. (2002). Intonational Variation in the British Isles. *Speech Prosody 2002*, 343–346.

Grabe, E., Post, B., Farrar, K., & Nolan, F. (2000). Pitch accent realization in four varieties of British

English. *Journal of Phonetics*, *28*(2), 161–185. https://doi.org/10.1006/jpho.2000.0111

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*(2), 143–185. https://doi.org/10.1017/S0952675700003924

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107. https://doi.org/10.1016/j.wocn.2017.03.003

Grønnum, N. (1990). Prosodic Parameters in a Variety of Regional Danish Standard Languages, with a View towards Swedish and German. *Phonetica*, *47*(3–4), 182–214. https://doi.org/10.1159/000261862

Gussenhoven, C. (1983). A semantic analysis of the nuclear tones of English. In *On the Grammar and Semantics of Sentence Accents*. Indiana University Linguistics Club. https://doi.org/10.1515/9783110859263.193

Gussenhoven, C. (1999). Discreteness and Gradience in Intonational Contrasts. *Language and Speech*, *42*(August), 283–305. https://doi.org/10.1177/00238309990420020701

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

Gussenhoven, C. (2016). Analysis of Intonation: the Case of MAE_ToBI. *Laboratory Phonology*, *7*(1). https://doi.org/10.5334/labphon.30

Haan, J. (2002). *Speaking of Questions - An Exploration of Dutch Question Intonation*. LOT.

Halliday, M. A. (1967). *Intonation and Grammar In British English*. Mouton.

Hardcastle, W. J. (1976). *Physiology of Speech Production*. Academic Press.

Harris, J. (1985). *Phonological Variation and Change: Studies in Hiberno-English*. Cambridge University Press.

Harris, J. (1994). *English Sound Structure*. Blackwell.

’t Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation*. Cambridge University Press.

Hartigan, J. A., & Hartigan, P. M. (1985). The Dip Test of Unimodality. *Annals of Statistics*, *13*(1), 70–84. https://doi.org/doi:10.1214/aos/1176346577

Hayes, B. (2009). Syllabification in English (unpublished chapter). In *Introductory Phonology*. Blackwell.

Hermes, D. J., & van Gestel, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, *90*(1), 97–102. https://doi.org/10.1121/1.402397

Hickey, R. (2007). *Irish English: History and present-day forms*. Cambridge University Press. https://doi.org/10.1017/CBO9780511551048

Hirst, D. J. (1998). Intonation in British English. In D. J. Hirst & A. Di Cristo (Eds.), *Intonation systems. A survey of twenty languages* (pp. 56–77). Cambridge University Press.

Hirst, D. J. (2013). UK declarative rises and the frequency code. In S. Hancil & D. Hirst (Eds.), *Prosody and Iconicity* (Issue April, pp. 149–160). John Benjamins.

https://doi.org/10.1075/ill.13.08hir

Houtsma, A. J. M. (1995). Pitch Perception. In B. C. J. Moore (Ed.), *Hearing* (pp. 267–295). Academic Press. https://doi.org/https://doi.org/10.1016/B978-012505626-7/50010-8

Inkelas, S., & Leben, W. R. (1990). Where phonology and phonetics intersect: the case of Hausa intonation. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 17–34). Cambridge University Press. https://doi.org/10.1017/cbo9780511627736.002

Jarman, E., & Cruttenden, A. (1976). Belfast intonation and the myth of the fall. *Journal of the International Phonetic Association*, *6*(1), 4–12. https://doi.org/10.1017/S0025100300001432

Jespersen, A. B. (2018). Innovations in the stylistic variation of nuclear tunes in Belfast English. *9th International Conference on Speech Prosody*, 527–531.

Kalaldeh, R., Dorn, A., Ní Chasaide, A., Kalaldah, R., Dorn, A., & Ní Chasaide, A. (2009). Tonal alignment in three varieties of Hiberno-English. *Proceedings of Interspeech 2009*, 2443–2446.

Kallen, J. L. (2013). *Irish English Volume 2: The Republic of Ireland*. De Gruyter Mouton.

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language*, *70*(3), 419–454.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*(2), 820–857. https://doi.org/10.1121/1.398894

Knight, R. A. (2008). The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux. *Lang Speech*, *51*(3), 223–244.

Knight, R. A., & Nolan, F. (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association*, *36*(2004), 21–38. https://doi.org/10.1017/S0025100306002349

Krug, P. K., Gerazov, B., van Niekerk, D. R., Xu, A., Xu, Y., & Birkholz, P. (2021). Modelling microprosodic effects can lead to an audible improvement in articulatory synthesis. *The Journal of the Acoustical Society of America*, *150*(2), 1209–1217. https://doi.org/10.1121/10.0005876

Kuznetsova, A. (2017). *lmerTest Package: Tests in Linear Mixed Effects Models*. *82*(13). https://doi.org/10.18637/jss.v082.i13

Lacy, B., Cody, E., Cotter, C., Cuppage, J., Dunne, N., Hurley, V., O'Rahilly, C., Walsh, P., & Ó Nualláin, S. (1983). *Archaeological Survey of County Donegal*. Donegal Couny Council.

Ladd, D. R. (1983). Phonological Features of Intonational Peaks. *Linguistic Society of America*, *59*(4), 721–759.

Ladd, D. R. (1990). Metrical representation of pitch register. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 35–57). Cambridge University Press. https://doi.org/10.1017/cbo9780511627736.003

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press.

Ladd, D. R., & Arvaniti, A. (2023). Prosodic Prominence Across Languages. *Annual Review of Linguistics*, *9*, 171–193. https://doi.org/10.1146/annurev-linguistics-031120-101954

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *The Journal of the Acoustical Society of America*, *107*(5), 2685–2696. https://doi.org/10.1121/1.428654

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. In *Journal of Phonetics* (Vol. 31, Issue 1). https://doi.org/10.1016/S0095-4470(02)00073-6

Ladd, D. R., Schepman, A., White, L., Quarmby, L. M., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, *37*(2), 145–161. https://doi.org/10.1016/j.wocn.2008.11.001

Ladefoged, P. (1996). *Elements of Acoustic Phonetics*. University of Chicago Press.

Leben, W. R., Inkelas, S., & Cobler, M. (1989). Phrases and Phrase Tones in Hausa. In P. Newman & R. D. Botne (Eds.), *Current Approaches to African Linguistics (vol. 5)* (pp. 45–62). https://doi.org/10.1515/9783112420089-005

Liberman, M. (1975). The Intonational System of English. In *Department of Foreign Literature and Linguistics: Vol. Ph.D.* MIT.

Liberman, M., & Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff, R. Oehrle, F. Kelley, & B. Wilker Stephens (Eds.), *Language Sound Structure: Studies in Phonology* (pp. 157–233). The MIT Press.

Lickley, R. J., Schepman, A., & Ladd, D. R. (2005). Alignment of "phrase accent" lows in Dutch falling rising questions: Theoretical and methodological implications. *Language and Speech*, *48*(2), 157–183. https://doi.org/10.1177/00238309050480020201

Lowry, O. (2001). *Belfast intonation: testing the ToBI framework of intonational analysis*. University of Ulster.

Lowry, O. (2002). The stylistic variation of nuclear patterns of Belfast English. *Journal of the International Phonetic Association*, *32*(32), 33–42. https://doi.org/10.1017/S0025100302000130

Lowry, O. (2011). Belfast intonation and speaker gender. *Journal of English Linguistics*, *39*(3), 209–232. https://doi.org/10.1177/0075424210380053

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Maechler, M. (2021). *diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected*. https://cran.r-project.org/package=diptest

McCafferty, K. (2001). *Ethnicity and Language Change. English in (London)Derry, Northern Ireland*. Benjamins.

McElholm, D. D. (1986). Intonation in Derry English. In H. Kirkwood (Ed.), *Studies in intonation. Occasional Papers in Linguistics and Language Learning* (pp. 1–58). New University of Ulster.

Michaud, A., & Sands, B. (2020). Tonogenesis. *Oxford Research Encyclopedia of Linguistics*, *September*, 1–28. https://doi.org/10.1093/acrefore/9780199384655.013.748

Microsoft Corporation. (2020). *Microsoft Excel*. https://office.microsoft.com/excel

Moore, B. C. J., & Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acta Acustica United with Acustica*, *82*(2), 335–345.

Moritz, N. (2016). Uptalk variation in three varieties of Northern Irish English. *Speech Prosody*, 119–122. https://doi.org/10.21437/SpeechProsody.2016-25

Murphy, A., Yanushevskaya, I., Chasaide, A. N., & Gobl, C. (2020). Testing the GlórCáíl system in a speaker and affect voice transformation task. *Speech Prosody 2020*, 950–954. https://doi.org/10.21437/SpeechProsody.2020-194

Murphy, A., Yanushevskaya, I., Chasaide, A. N., & Gobl, C. (2019). The role of voice quality in the perception of prominence in synthetic speech. *Interspeech 2019*, 2543–2547. https://doi.org/10.21437/Interspeech.2019-2761

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nolan, F. (1997). Intonation. In B. Aarts & A. McMahon (Eds.), *Handbook of English Linguistics* (pp. 432–457). Cambridge University Press.

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. *ICPhS 15*, 1249–1252.

Nolan, F., & Farrar, K. (1999). Timing of F0 Peaks and Peak Lag. *ICPhS 14*, 961–964.

Northern Ireland Statistics and Research Agency. (2015). *Review of the Statistical Classification and Delineation of Settlements The Northern Ireland Statistics and Research Agency* (Issue March). https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/review-of-the-statistical-classification-and-delineation-of-settlements-march-2015_0.pdf

Northern Ireland Statistics and Research Agency. (2016). *Derry City (N11000243)*. http://www.ninis2.nisra.gov.uk/public/StaticMapViewer.aspx?geoCode=N11000243&geoLevel=SETT2015

Nugent, J. (2022). *Using allFit() with (g)lmer*. https://joshua-nugent.github.io/allFit/

O'Reilly, M., Dorn, A., & Ní Chasaide, A. (2010). Focus in Donegal Irish (Gaelic) and Donegal English bilinguals. *Speech Prosody*, 1–4.

Ohala, J. J. (1976). Explaining the intrinsic pitch of vowels. In *The Journal of the Acoustical Society of America* (Vol. 60, p. S44). https://doi.org/10.1121/1.2003351

Ohala, J. J. (1983). Cross-Language Use of Pitch: An Ethological View. In *Phonetica* (Vol. 40, pp.

1–18).

Peter Birkholz, X. Z. (2020). Accounting for Microprosody in Modeling Intonation. *ICASSP 2020*, 8099–8103.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. MIT.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese Tone Structure*. MIT Press.

Pierrehumbert, J. B., & Steele, S. A. (1989). Categories of tonal alignment in english. In *Phonetica* (Vol. 46, Issue 4, pp. 181–196). https://doi.org/10.1159/000261842

Prieto, P. (2011). Tonal Alignment. In M. von Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1–19). John Wiley & Sons. https://doi.org/10.1002/9781444335262.wbctp0050

Prieto, P., D'Imperio, M., & Gili Fivela, B. (2005). Pitch Accent Alignment in Romance: Primary and Secondary Associations with Metrical Structure. *Language and Speech*, *48*(4), 359–396. https://doi.org/10.1177/00238309050480040301

Prieto, P., Estebas-Vilaplana, E., & Mar Vanrell, M. del. (2010). The relevance of prosodic structure in tonal articulation Edge effects at the prosodic word level in Catalan and Spanish. *Journal of Phonetics*, *38*(4), 687–705. https://doi.org/10.1016/j.wocn.2010.10.004

R Core Team. (2022). *R: A language and environment for statistical computing* (4.2.0). R Foundation for Statistical Computing.

Reichel, U. D., & Winkelmann, R. (2010). Removing micromelody from fundamental frequency contours. *Proceedings of the International Conference on Speech Prosody*, 1–4.

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rietveld, A. C. M., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, *13*(3), 299–308. https://doi.org/10.1016/s0095-4470(19)30761-2

Roach, P. (1994). Conversion between prosodic transcription systems: "Standard British" and ToBI. *Speech Communication*, *15*(1–2), 91–99. https://doi.org/10.1016/0167-6393(94)90044-2

Rodgers, A. E. (2018). Prenuclear pitch accents and peak alignment in Derry~Londonderry English. *Colloquium of the British Association of Academic Phoneticians*.

Rodgers, A. E. (2020). K-Max: a tool for estimating, analysing, and evaluating tonal targets. *Speech Prosody 2020*, *May*, 225–229. https://doi.org/10.21437/SpeechProsody.2020-46

Rodgers, A. E. (2019). The effects of anacrusis and foot size on prenuclear pitch accents in northern Irish English (Derry City). *ICPhS 2019*, 1307–1311.

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology*, *10*(1), 1–27. https://doi.org/10.5334/labphon.147

Sag, I., & Liberman, M. (1975). The Intonational Disambiguation of Indirect Speech Acts. *Papers from the Eleventh Regional Meeting*, *April 18-20*, 487–497.

Shosted, R., Giudice, A. Del, & Arvaniti, A. (2006). Comparing Methods for Measuring Pitch "Elbows." *152nd Meeting of the ASA*.

Silverman, K., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Betweeen the grammar and physics of speech* (pp. 72–108). Cambridge University Press.

Smiljaníc, R., & Hualde, J. I. (2000). Lexical and pragmatic functions of tonal alignments in two Serbo-Croation dialects. *Proceedings from the Main Session of the 36th Regional Meeting of the Chicago Linguistics Society*, *36*(1), 469–482.

Snider, K. L. (1988). Towards the Representation of Tone: A Three-Dimensional Approach. In H. von der Hulst & N. Smith (Eds.), *Features, Segmental Structure and Harmony Processes (Part I)* (pp. 237–267). De Gruyter Mouton. https://doi.org/10.1515/9783110875812-009

Snider, K. L. (1990). Tonal Upstep in Krachi: Evidence for a Register Tier. *Language*, *66*(3), 453–474. https://www.jstor.org/stable/414608

Sosa, J. M. (1999). *La entonación del español*. Cátedra.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. *Journal of the Acoustical Society of America*, *19*(1936), 14–19.

Sullivan, J. N. (2007). Variability of F0 Valleys: The Case of Belfast English. In N. Hilton, R. Arscott, K. Barden, A. Krishna, S. Shah, & M. Zellers (Eds.), *CamLing 2007: Proceedings of the Fifth University of Cambridge Postgraduate Conference in Language Research* (Issue L, pp. 245–252). Cambridge Institute of Language Research.

Sullivan, J. N. (2010). *Approaching intonational distance and change* (Issue September).

Sullivan, J. N. (2012). The why of Belfast rises. In *New Perspectives on Irish English* (pp. 68–73). John Benjamins.

Taylor, P. A. (1992). *A Phonetic Model of English Intonation*. University of Edinburgh.

Tench, P. (1996). *The intonation systems of English*. Cassell Academic.

Welby, P. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. Ohio State University.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*(3), 349–366. https://doi.org/10.1016/S0095-4470(95)80165-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R* (1st ed.). Routledge. https://doi.org/https://doi.org/10.4324/9781315165547

Xu, Y. (1999a). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *5*, 55–105.

Xu, Y. (1999b). F0 peak delay: When, where and why it occurs. *ICPhS 14*, 1881–1884.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech*

Communication, *46*(3–4), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Xu, Y. (2004). the PENTA Model of Speech Melody: Transmitting Multiple Communicative Functions in Parallel. *From Sound to Sense, MIT, April*, 1–6. papers3://publication/uuid/4681AF2D-6DD1-49EC-88DA-CABD35AC3183

Xu, Y., Lee, A., Prom-On, S., & Liu, F. (2015). Explaining the PENTA model: A reply to Arvaniti and Ladd. *Phonology*, *32*(3), 505–535. https://doi.org/10.1017/S0952675715000299

Xu, Y., Prom-on, S., & Liu, F. (2022). The PENTA Model: Concepts, Use, and Implications. In J. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic Theory and Practice* (Issue September). The MIT Press. https://doi.org/10.7551/mitpress/10413.003.0014

Xu, Y., & Sun, X. (2000). How fast can we really change pitch? Maximum speed of pitch change revisited. *Proceedings of ICSLP 2000*, 666–669. http://www.haskins.yale.edu/yixu/00571.pdf

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, *111*(3), 1399–1413. https://doi.org/10.1121/1.1445789

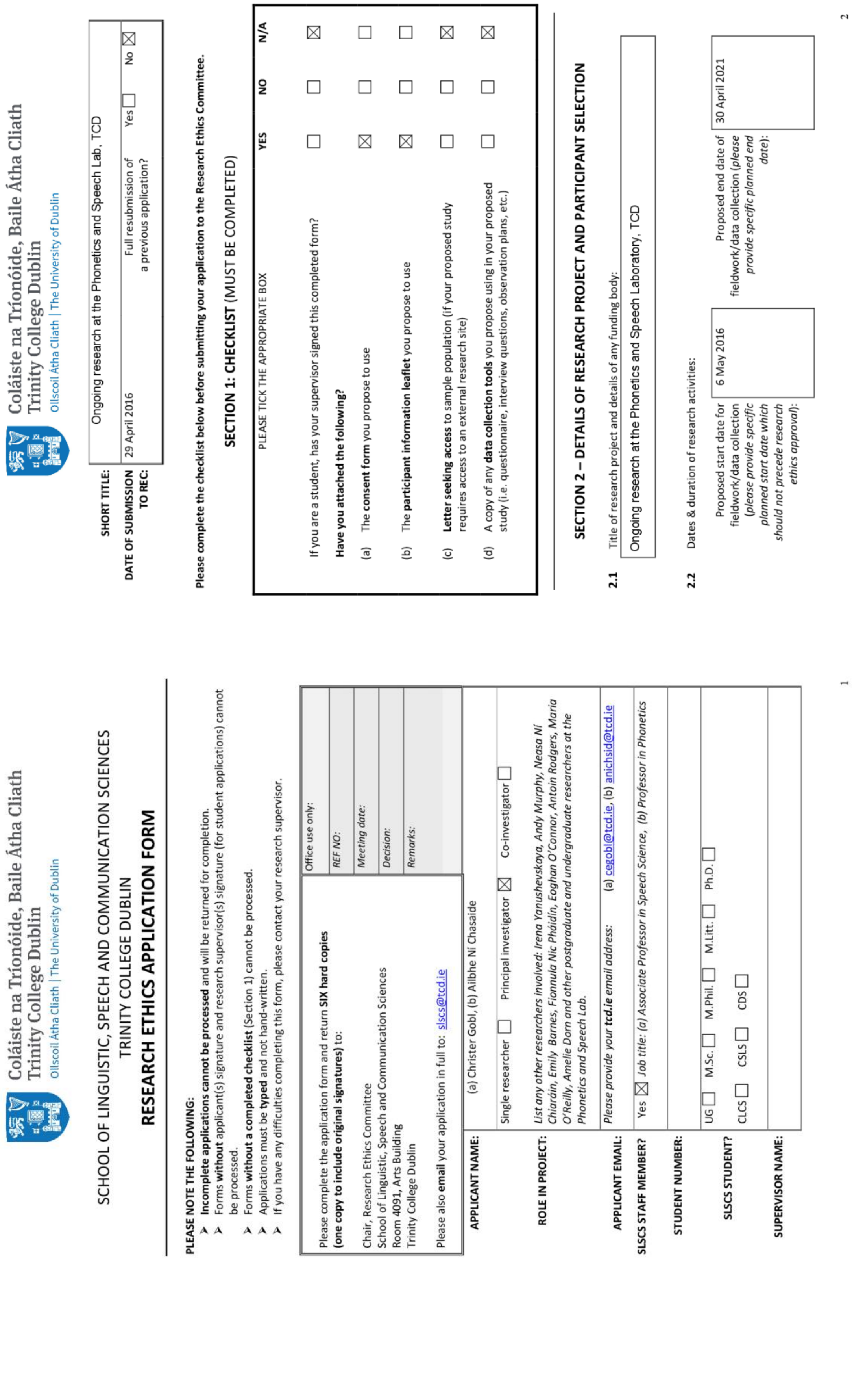
Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

Zwicker, E. (1961). Subdivision of the Audible Frequency Range into Critical Bands. *Journal of the Acoustical Society of America*, *33*(2), 248.

Zwickl, S. (2002). *Language Attitudes, Ethnic Identity and Dialect Use across the Northern Ireland Border: Armagh and Monaghan*. Cló Ollscoil na Banríona.

Appendix A. Ethics and Consent Documentation

A1. Phonetics and Speech Laboratory Ethics Application, 2016



Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

SCHOOL OF LINGUISTIC, SPEECH AND COMMUNICATION SCIENCES
TRINITY COLLEGE DUBLIN
RESEARCH ETHICS APPLICATION FORM

PLEASE NOTE THE FOLLOWING:

- **Incomplete applications cannot be processed** and will be returned for completion.
- Forms **without** applicant(s) signature and research supervisor(s) signature (for student applications) cannot be processed.
- Forms **without a completed checklist** (Section 1) cannot be processed.
- Applications must be **typed** and not hand-written.
- If you have any difficulties completing this form, please contact your research supervisor.

Please complete the application form and return **SIX hard copies (one copy to include original signatures)** to:

Chair, Research Ethics Committee
School of Linguistic, Speech and Communication Sciences
Room 4091, Arts Building
Trinity College Dublin

Please also **email** your application in full to: slscs@tcd.ie

Office use only:
REF NO:
Meeting date:
Decision:
Remarks:

APPLICANT NAME:	(a) Christer Gobl, (b) Ailbhe Ní Chasaide
ROLE IN PROJECT:	Single researcher ☐ Principal investigator ☒ Co-investigator ☐ *List any other researchers involved: Irena Yanushevskaya, Andy Murphy, Neasa Ní Chiaráin, Emily Barnes, Fionnula Nic Pháidín, Eoghan O'Connor, Antoin Rodgers, Maria O'Reilly, Amelie Dorn and other postgraduate and undergraduate researchers at the Phonetics and Speech Lab.*
APPLICANT EMAIL:	*Please provide your* ***tcd.ie*** *email address:* (a) cegobl@tcd.ie, (b) anichsid@tcd.ie
SLSCS STAFF MEMBER?	Yes ☒ *Job title: (a) Associate Professor in Speech Science, (b) Professor in Phonetics*
STUDENT NUMBER:	
SLSCS STUDENT?	UG ☐ M.Sc. ☐ M.Phil. ☐ M.Litt. ☐ Ph.D. ☐ CLCS ☐ CSLS ☐ CDS ☐
SUPERVISOR NAME:	

1

Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

SHORT TITLE:	Ongoing research at the Phonetics and Speech Lab, TCD		
DATE OF SUBMISSION TO REC:	29 April 2016	Full resubmission of a previous application?	Yes ☐ No ☒

Please complete the checklist below before submitting your application to the Research Ethics Committee.

SECTION 1: CHECKLIST (MUST BE COMPLETED)

PLEASE TICK THE APPROPRIATE BOX	YES	NO	N/A
If you are a student, has your supervisor signed this completed form?	☐	☐	☒
Have you attached the following?			
(a) The **consent form** you propose to use	☒	☐	☐
(b) The **participant information leaflet** you propose to use	☒	☐	☐
(c) **Letter seeking access** to sample population (if your proposed study requires access to an external research site)	☐	☐	☒
(d) A copy of any **data collection tools** you propose using in your proposed study (i.e. questionnaire, interview questions, observation plans, etc.)	☐	☐	☒

SECTION 2 – DETAILS OF RESEARCH PROJECT AND PARTICIPANT SELECTION

2.1 Title of research project and details of any funding body:

Ongoing research at the Phonetics and Speech Laboratory, TCD

2.2 Dates & duration of research activities:

Proposed start date for fieldwork/data collection *(please provide specific planned start date which should not precede research ethics approval)*:	6 May 2016	Proposed end date of fieldwork/data collection (*please provide specific planned end date*):	30 April 2021

2

2.3 What are the primary location(s) for data collection? E.g. classroom, clinic, lab, participants' home, place of convenience for participants. Researchers conducting fieldwork in participants' home should ensure they have familiarised themselves with the Lone Researcher Guidelines on the School website.

Phonetics and Speech Laboratory, Trinity College Dublin and field work at other locations.

2.4 Please provide a brief outline of the proposed project (**maximum 400 words in total**). This should include aim(s) and objective(s), background, research question(s) or hypothesis, research design, recruitment and sampling, and data collection procedures/instruments.

(a) Aims/objectives and theoretical background

The application is intended to cover ongoing day to day recordings and listening tests that are routinely used both for research and teaching at the Phonetics and Speech Laboratory, TCD.

Recordings are typically high quality audio, articulatory or aerodynamic speech data. Recordings are most frequently made in the recording studio at the Laboratory, but can also on occasion be conducted elsewhere, as the need arises.

Listening tests are often required to establish the perceptual correlates of given acoustic parameters. These can be conducted in the laboratory, or in certain cases, in other locations or online.

The recordings and listening tests would only involve adult participants. Recorded materials do not have any personal content. The recordings are always kept in anonymised files and only demographic information retained with them (e.g., dialect, age, gender and educational status). In the case of the listening tests, there is likewise no personal information elicited, and again only demographic information is retained for the listeners involved.

In all recordings of human voices, we will elicit at the time permission to use the data in teaching and research presentations. The form for this is attached.

(b) Research question(s) or hypothesis

As this is a generic application, it applies to future unspecified research questions.

(c) Research design. Please briefly describe the project research design/methodology.

Again, the details of design cannot be fully specified here, as they will vary from case to case. However, recordings are typically of read speech (this may include isolated vowels, words, sentences, short dialogues or texts), and frequently up to 10 repetitions of the same utterance may be recorded to ensure the reliability of elicited data.

A variety of acoustic analyses may be performed, e.g., inverse filtering, spectral and spectrographic analyses, as well as analyses using many other techniques available at the Phonetics and Speech Lab.

The listening experiments are typically conducted to assess the perceptual relevance of particular acoustic features. These would typically involve listening to a number of systematically differentiated stimuli and filling in an answer sheet.

(d) Recruitment and sampling. Please specify (i) **who** may be contacted by you during fieldwork/data collection, e.g. in seeking access to research population or a gatekeeper, (ii) **how** they will be contacted by you, and (iii) **expected sample size and composition.**

The details of recruitment and sampling cannot be fully specified here, as they will vary from case to case.

(e) Data collection procedures. Please tick the research instruments you intend to use (ensuring you append copies of each instrument) and (ii) provide an estimation of the time commitment involved for participants/respondents.

(i) *Please tick as many boxes as relevant to your project:*

☐ Questionnaires	☐ Video recordings
☐ Interviews	☐ Observations
☐ Focus Groups	☐ Classroom intervention
☒ Audio recordings	☐ Ethnographic research
	☒ Other (*please specify*): Articulatory and aerodynamic recordings, listening tests

(ii) *Estimation of the time commitment involved for participants:*

The time commitment for participants involved in both recording and listening tests will not exceed 1 hour.

SECTION 3 – CONSENT AND CONFIDENTIALITY (INCLUDING DATA PROTECTION)

3.1 Will informed consent be obtained from the adult research participants?

YES ☒ **NO** ☐ **N/A** ☐

If **YES**, please give details of who will obtain consent from participants and how it will be done. Please attach a copy of any letters, consent form (if required) and information leaflet (where appropriate). Please see guidelines on how to prepare these documents on the School website, and adapt the examples provided to suit your study and participants.

Typically, the researcher undertaking the study will be responsible for obtaining participants' consent. This entails ensuring that participants have read the participant information leaflet and have signed the consent form.

If **NO**, please explain what alternative approaches are being implemented.

If **Not Applicable,** please explain why.

3.2 Working with children: will assent be obtained from any children under 16?

YES ☐ **NO** ☐ **N/A** ☒

Will informed consent be obtained from parents/carers on behalf of any children under 16?

YES ☐ **NO** ☐ **N/A** ☒

If **YES**, please give details of who will obtain assent from children/consent from parents/carers, and how it will be done. Please attach a copy of any letters, assent/consent form (if required) and information leaflet (where appropriate). Please see guidelines on how to prepare these documents on the School website, and adapt the examples provided to suit your study and participants.

If **NO**, please explain what alternative approaches are being implemented. (See *Introduction to Research Ethics* document and *Frequently Asked Questions* for more information about the difference between *assent* and *consent*).

If **N/A**, please comment.

No children under 16 will be recruited as participants.

5

3.3 Please specify if you will allow for a time interval between providing your participants with information about the research and seeking their consent:
(For example, in some research methodologies, it is recommended that a period of 3 to 7 days be provided for reflection before asking individuals to participate in an experiment.)

A period of three days will be provided to the participants for reflection between informing them about the research and seeking their consent.

3.4 Will the participants be from any of the following groups (tick as appropriate):

	YES	NO
Children under 16 years of age	☐	☒
Adults with learning disabilities	☐	☒
Adults with language or communication difficulties	☐	☒
Adults with mental illness	☐	☒
Clinical population	☐	☒
Other groups who may be considered vulnerable Please specify:	☐	☒

3.5 If participants are to be recruited from any of the potentially vulnerable groups listed above, please give details of:

(a) Any special steps taken to ensure that participants from vulnerable groups are as fully informed as possible about the nature of their involvement:

(b) Who will give consent:

(c) How consent will be obtained (e.g. will it be verbal, written or visually indicated?):

(d) The arrangements that have been made to inform those responsible for the care of the research participants of their involvement in research:

(e) The use of a gatekeeper in accessing participants

6

3.6 During and after the study, what steps will you take to protect:

(a) Participant identities?

To protect participants' identities, their names will be coded/pseudonymised. The access to the code will only be available to the PI and researchers directly involved in the project.

(b) Hardcopy records?

Hardcopy records will not be kept.

(c) Digital data? Please describe measures to be taken during transfer and storage.

Digital data will be stored on a password-protected computer. Backup copies of the digital data will be stored on USB or external hard drives, at the Phonetics and Speech Laboratory. Note that names and any other identifying details will not be stored with the data.

3.7 If the data is sensitive, what other person(s) other than the researcher(s) named in this form will have access to the data collected, and what steps will be taken to protect confidentiality?

No sensitive data will be collected.

3.8 Will participants be given access to a copy or transcript of any recorded material (including audio or video files), if they so wish?

The participant's entitlement in this regard should be mentioned in the consent form and participant information leaflet (if these forms are used).

YES	NO	N/A	PLEASE EXPLAIN WHAT YOU WILL DO:
☒	☐	☐	A copy of their recording will be available to participants upon request. This will be clearly stated in the consent form and participant information leaflet.

3.9 Who will take responsibility for the secure storage of, and access to, the data generated by the research during and after the project?

The research supervisor ☐

The student researcher ☐

Research supervisor/student researcher, jointly ☒

Principal researcher ☐

Other researcher(s) (*please state their name(s) and contact details*) ☐

3.10 Who will be responsible for archiving or destroying the data?

Please provide the name of the person responsible, and details of what will eventually happen to the data. In the case of student researchers, it is generally the research supervisor who will take responsibility for storage and destruction of data after a five year period.

Most of the recordings are valuable archival materials, e.g., recordings of an endangered dialect of Irish or of a variety of Hiberno-English. It is important that these should not be destroyed, especially as they contain no sensitive information. The digitised data will be archived and stored on a password protected server at the Phonetics and Speech Laboratory indefinitely. Where appropriate, it will be used for future research and teaching purposes at the Phonetics and Speech Laboratory, TCD.

SECTION 4 - RISK, HARM AND BENEFIT

4.1 What is the potential for an adverse outcome for research participants? (For example, inconvenience, physical or emotional risk, discomfort, stress, anxiety, fatigue or embarrassment. In low risk projects, adverse outcomes are usually rare.) NOTE: for the protection of both the researcher and participants, this list must appear in full in the participant information leaflet.

No adverse outcomes are envisaged for participants other than possible fatigue.

4.2 Please indicate what steps you will take in order to minimize any potential adverse outcomes for research participants:

Participants will be given an opportunity to take breaks during recording sessions and listening tests, as required.

4.3 What is the potential for benefit, if any, for research participants?

No direct benefit for participants is associated with this research. However, in very many cases, participants have a vested interest in ensuring that their dialect/accent is recorded and studied for posterity. In some cases, there are potential indirect benefits for the speech community of the participant. For example, modelling the articulation or prosodic features of a dialect can serve for building teaching materials appropriate to the dialect.

4.4 Will payment be made to research participants?

☐ YES

☐ NO

☒ Minimal payment to cover travel costs etc.

4.5 If you answered **YES** to the previous question, please specify for what purpose the payment will be made and the amount per participant:

4.6 Are you aware of any conflicts of interest that could arise in the course of this project? If your answer is **YES**, please give full details below:

No.

4.7 Are there any other ethical considerations which you anticipate in relation to your study that have not been covered by the questions above? If so, what steps will you take to address these?

No.

SECTION 5 - DECLARATION OF APPROVAL AND SIGNATURES

APPLICANT DECLARATION: I confirm that the information provided in this form is correct, that I am not aware of any other ethical issues not addressed within this form. I understand the obligations to and the rights of participants (particularly concerning their safety and welfare, the obligation to provide information sufficient to give informed consent and the obligation to respect confidentiality).		
APPLICANT NAME:	Christer Gobl, Ailbhe Ní Chasaide	
SIGNATURE (for hard copies):		DATE: 29 April 2016

RESEARCH SUPERVISOR SIGNATURE Student applicants are required to have their Research Supervisor complete this section.
SUPERVISOR NAME :
As the student's supervisor, I have read this document, and to the best of my knowledge, this project conforms to the School's Research Ethics Guidelines. SUPERVISOR'S SIGNATURE & DATE

A2. Participant Information Leaflet

TRINITY COLLEGE DUBLIN

SCHOOL OF LINGUISTIC SPEECH AND COMMUNICATION SCIENCES

Information Leaflet

Ongoing research at the Phonetics and Speech Laboratory, TCD

Principal Investigator(s): Ailbhe Ní Chasaide, Christer Gobl

You are invited to participate in this research project, which is being carried out by Antoin Rodgers. Your participation is voluntary. Even if you agree to participate now, you can withdraw at any time without any consequences of any kind.

The study is designed to explore the phonetics and phonology of Derry-Londonderry English as part of ongoing research to document and study different varieties of English in Ireland. If you agree to participate, this will involve you:

- Taking part in short interactive tasks with another volunteer;
- Reading short lists of dialogues aloud with the other volunteer;
- Re-telling a familiar story.

The recording will take place in a location that is suitable for recording and which is you and the other volunteer are both comfortable with. The recording may take approximately one hour. During the recording, you will be given breaks as required.

You will not benefit directly from participating in this research.

The data obtained from you during this research will contain no personal information. Your name will only be known to the researchers conducting this study and in reporting of the research, your name will be coded to ensure your anonymity. The data will be kept on a password-protected server at the Phonetics and Speech Laboratory, Trinity College Dublin.

Data from this research project may be published in future and the results may be presented at conferences and public gatherings.

Portions of recordings may be played in classes or during conference presentations and written transcriptions may be made for teaching purposes or for further analysis. You can request a copy of the recordings. Should you wish to remove any portion of the recording, this will be complied with.

If you have any questions about this research you can ask Antoin Rodgers (rodgeran@tcd.ie). You are also free, however, to contact any of the other people involved in the research to seek further clarification and information: Ailbhe Ní Chasaide (ANICHSID@tcd.ie) and Christer Gobl (cegobl@tcd.).

A3. Participant Consent Form

TRINITY COLLEGE DUBLIN
SCHOOL OF LINGUISTIC SPEECH AND COMMUNICATION SCIENCES
Consent Form

Ongoing research at the Phonetics and Speech Laboratory, TCD
Principal Investigator(s): Antoin Rodgers

I am invited to participate in this research project which is being carried out by Antoin Rodgers. My participation is voluntary. Even if I agree to participate now, I can withdraw at any time without any consequences of any kind.

The study is designed to explore the Phonology and Phonetics of Derry-Londonderry English. If I agree to participate, this will involve me:

- Taking part in short interactive tasks with another volunteer;
- Reading short lists of dialogues aloud with the other volunteer;
- Re-telling a familiar story.

The recording will take place in a location that is suitable for recording and which is familiar and comfortable to me. The recording may take approximately one hour. During the recording, I will be given breaks as required.

I will not benefit directly from participating in this research.

The data obtained from me during this research will contain no personal information. My name will only be known to the researchers conducting this study and in reporting of the research, my name will be coded to ensure my anonymity. The data will be kept on a password-protected server at the Phonetics and Speech Laboratory, Trinity College Dublin.

Data from this research project may be published in future and the results may be presented at conferences and public gatherings.

Portions of recordings may be played in classes or during conference presentations and written transcriptions may be made for teaching purposes or for further analysis. I can request a copy of the recordings. Should I wish to remove any portion of the recording, this will be complied with.

If I have any questions about this research I can ask Antoin Rodgers (rodgeran@tcd.ie). I am also free, however, to contact any of the other people involved in the research to seek further clarification and information: Ailbhe NíChasaide (ANICHSID@tcd.ie) or Christer Gobl (cegobl@tcd.ie).

I understand what is involved in this research and I agree to participate in the study. I have been given a copy of the Participant Information Leaflet and a copy of this consent form to keep.

-- ----------------
Signature of participant Date

Signature of researcher
I believe the participant is giving informed consent to participate in this study.

-- ----------------------
Signature of researcher Date

Appendix B. Read Speech Stimuli

B1. Stimuli for M-Corpus

Code	Context	Stimulus (A)	target response (B)
MDC1	Talking about work	So, what did you do at work today?	I valued the vases.
MDC2	Talking about yourself	So, where do you live?	I live in the valley.
MDC3	Getting ready to go on holiday	Have you done anything on the list?	I've hidden the valuables
MYN1	Talking about work	I think everything's ready for the auction.	Have you valued the vases?
MYN2	Talking about yourself	What else did you want to know about me?	Do you live in the Valley?
MYN3	Getting ready to go on holiday	I've finished packing and I'm ready to go.	Have you hidden the valuables?
MWH1	Talking about work	Have you any questions about the auction?	Who valued the vases?
MWH2	Talking about yourself	You wanted to ask a question?	Why do you live in the Valley?
MWH3	Getting ready to go on holiday	I've finished packing and I'm ready to go.	Where have you hidden the valuables?
MDQ1	Talking about work & checking you heard correctly	I valued the vases.	You valued the vases?
MDQ2	Talking about yourself & checking you heard correctly	I live in the valley.	You live in the valley?
MDQ3	Getting ready to go on holiday & checking you heard correctly	I've hidden the valuable.	You've hidden the valuables?

B2. Stimuli for A-Corpus

Code	Context	Stimulus (A)	target response (B)
A1422	Describing a picture	What can you tell me about the valley?	The valley's by the river.
A2422	Describing a picture	What do you see in the painting?	There's a valley with a river.
A3422	Describing a picture	What do you remember about the painting?	There was a valley with a river.
A0131	chit-chat	What did the burglars steal?	Val's valuables.
A0221	Organising a group trip	What did you say about Val's travel card?	Val's is valid.
A0321	Organising a group trip	What did you say about Val's travel card?	Val's is invalid.
A0423	Organising a group trip	What did you say about Valerie's travel card?	Valerie's is valid.
A1111	Organising a group trip	Have they met any of the others yet?	They know Val.
A1211	Talking about yourself	Who does he live with?	He lives with Val.
A1231	Talking about yourself	Who do you live with?	I live with Valerie.
A1241	Giving information to a work mate	What's happening with those job applications?	They need evaluating.

B3. Stimuli for H-Corpus

Set	Context	Stimulus (A)	target response (B)
A0221	Organising a group trip	What did you say about Val's travel card?	Val's is invalid.
H0322	Organising a group trip	What did you say about Lally's travel card?	Lally's is valid.
H0422	Organising a group trip	What did you say about Lally's travel card?	Lally's is invalid.
A0423	Organising a group trip	What did you say about Valerie's travel card?	Valerie's is valid.
H1321	talking about friends and work	What was Elaine's old job?	Elaine was a nanny.
H1322	talking about friends and work	What does Elaina do?	Elaina's a nanny.

Appendix C. BAAP2018 Poster Presentation (Rodgers, 2018)

Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

Phonetics and Speech Laboratory
School of Linguistic, Speech and Communication Sciences

Prenuclear pitch accents and peak alignment in Derry~Londonderry English

Antoin Eoin Rodgers
rodgeran@tcd.ie

INTRODUCTION

Focus: rhythmic / lexical factors in PN alignment and tone type.

- **Northern Irish English**: L*H nuclear accents well documented
- **Aim:** Identify initial PN tonal target anchors & degree of stability:
  1. ***Does number of syllables in anacrusis and the foot affect:***
     a. ***'choice' of pitch-accent?***
     b. ***tonal alignment?***
  2. ***Is peak alignment affected by right word boundary*** *(more than the stressed syllable / right foot boundary)?*

MATERIALS

- 9 target phrases from short read dialogues, 3 sets of data:
  - Anacrusis: 0 to 3 syllables
  - Foot: 1 to 4 syllables (no anacrusis)
  - Same foot / different word boundaries: paired sentences
- 5 repetitions x 6 speakers (3F, 3M, 35-58 y/o) = 30 reps / phrase

PREPARING THE DATA

- f_o tracking errors corrected manually in Praat
- **Annotation:** (see *Fig. 1*) word, syllable, rhythm; vowel onset;
- f_o maxima and minima marked (or elbow if none visible)
- **Phonological analysis:** (IViE)
- Utterances with no PN accent excluded from analyses:
  - PN accent and alignment analysis (n=191)
  - Analysis of effect of lexical boundary on peak timing (n=57).

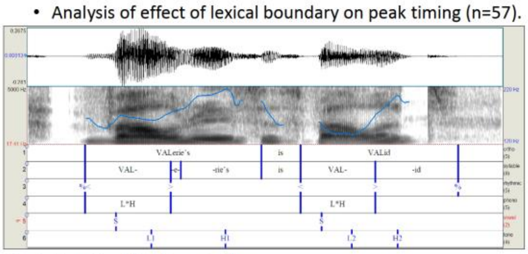


Fig. 1. *Sample of Praat annotation.*

RESULTS: 'CHOICE' OF PN ACCENT

- No anacrusis → fewer PN accents (6/30 ana 0 tokens unaccented)
- More anacrusis or more syllables in feet → higher chance of L*H
- H* more likely when less 'space' for L*H realisation (*Fig. 2*).

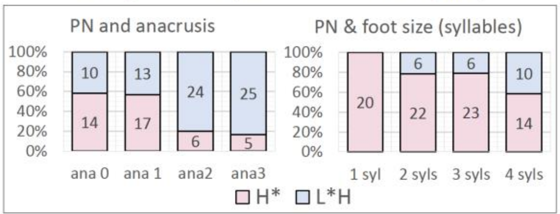


Fig. 2. *PN occurrences as function of anacrusis (left) and foot size (right)*

RESULTS: PN TONAL ALIGNMENT

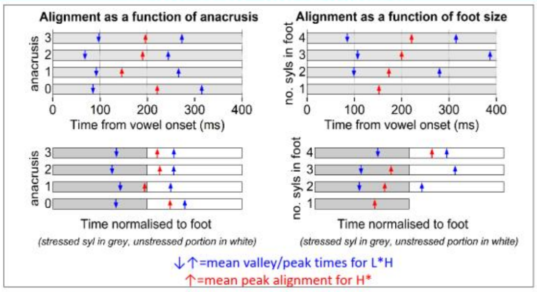


Fig. 3. *PN alignment as a function of time (means across all speakers)*

- **L targets (↓)**: relatively stable in both contexts
- **H targets**: **trailing H (↑)** aligned later than **H* target (↑)**
- More Anacrusis → all **H tones** tend to be *slightly* earlier
- More syllables in feet → **H* targets (↑)** consistently later **trailing H (↑)** : no clear trend

RESULTS: WORD BOUNDARY EFFECTS

- Exclusively H*
- 2 distinct strategies (see *Fig. 4*):
  - Lexically motivated peak-shift: 4 speakers (*upper*)
  - Rhythmically motivated peak-alignment: 2 speakers (*lower*)

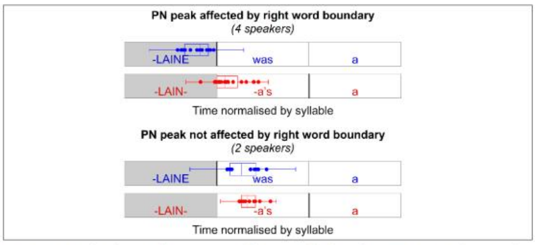


Fig. 4. *Box plot for H alignment with same rhythm but varying right word boundaries (bold line) in phrases* Elaine was a nanny *and* Elaina's a nanny.

SUMMARY & CONCLUSIONS

- L*H preferred in non-time pressure contexts
- L 'sacrificed' due to forward / backward time-pressure → H*
- PN deaccentuation more common when no anacrusis
- H alignment competition: stressed syllable, word *v* foot boundary
- **Provisional conclusions:**
  - L*H → H* → no accent
  - L*H = default PN (hypothesis: H* may be an allophone)
  - Trailing H instability a function of competing forces
- **Further work:**
  - More data: speakers, contexts, interaction types
  - f_o contour analysis

Appendix D. A-Corpus, Summary of Raw Counts

Table D1.1 PN pitch accents by number of syllables in the foot (`foot_syls`) in `pn_foot` subcorpus.

foot_syls	(*)	L*	H*	>H*	L*H
1	9	5	19	2	20
2	2	2	21	3	27
3	1	0	13	6	35
4	0	0	0	3	35

Table D1.2 PN pitch accents by syllables of anacrusis (`ana_syls`) in `pn_ana` subcorpus.

ana_syls	H*	>H*	L*H
0	0	3	35
1	9	5	42
2	0	4	52
3	0	4	51

Table D1.3 Prenuclear pitch accents by speaker.

speaker	(*)	L*	H*	>H*	L*H
F5	4	0	5	2	24
F6	3	0	0	1	31
F12	0	1	0	0	32
F15	0	0	2	2	32
F16	0	0	0	0	35
F17	1	6	0	0	25
M4	4	0	10	4	13
M5	0	0	17	0	18
M8	0	0	2	9	22
M9	0	0	20	8	2
M10	0	0	6	1	28

Table D1.4 Nuclear contours by syllables in the foot (`foot_syls`) in `nuc_foot` subcorpus.

foot_syls	L*H %	L*H L%
1	50	1
2	55	0
3	30	5
4	52	3

Table D1.5 Nuclear PAs by preceding unstressed syllables (`pre_syls`) in `nuc_pre` subcorpus.

pre_syls	L*H %	L*H L%
0	50	4
1	55	0
2	55	0
3	37	1

Table D1.6 Nuclear contours by speaker.

speaker	L*H %	L*H L%
F5	34	0
F6	35	0
F12	23	10
F15	30	0
F16	35	0
F17	28	0
M4	31	0
M5	33	0
M8	21	3
M9	25	1
M10	34	0

Appendix E. Syllable-normalised and Grand Mean-syllable Normalised Time

E1. Calculating syllable-normalised time

Syllable normalised time is calculated in two stages.

1. The timing of each target is converted to a ratio of the syllable in which it occurred. This is shown in Equation (E1.1, where *t* refers to time while *on* and *off* refer to the onset and offset of the target syllable.

$$ratio_{target} = \frac{t_{target} - t_{on}}{t_{off} - t_{on}} \tag{E1.1}$$

2. $ratio_{target}$ is converted to an utterance-wide value by adding the number of the syllable in which the target occurred and subtracting it by one, as shown in Equation E1.2. Here, *t.syl.norm* refers to the syllable-normalised time while n_{syl} refers to the number of the syllable containing the target.

$$t.syl.norm_{target} = ratio_{target} + n_{syl} - 1 \tag{E1.2}$$

In this way, the value on the left of the decimal point is the number of complete syllables while the number on its right is a proportion of the current syllable.

To calculate syllable-normalised time in the first foot, the number of syllables in anacrusis is subtracted from $t.syl.norm_{target}$.

E2. Calculating grand syllable-mean normalised time

Grand syllable-mean normalised time requires that the grand mean duration of each syllable in a target utterance must first be calculated from each repetition in the corpus. Each repetition used in the calculation must be of the same target phrase and use the same meter.

Once the grand-mean duration of each syllable in the target is calculated, each target time can be converted to grand syllable-mean normalised time. This is achieved by multiplying by grand-mean duration of the syllable in which the target occurs and adding it the sum of all preceding grand-mean-syllable durations. This is summarised in Equation E1.3. Here, $t.GM_{target}$ is the grand mean-normalised time of the target, n_{syl} is number of the syllable containing the target, and $\overline{dur}$ is mean duration of a syllable.

$$t.GM_{target} = \sum_{n=1}^{n_{syl}-1} \overline{dur}_n + ratio_{target} \times \overline{dur}_{n_{syl}} \tag{E1.3}$$

Appendix F. BGLMM Models for PN Phonology in A- and H-Corpora.

F1. Likelihood of L*H (isLH) in PNs in A- and H-Corpora.

*Table F1.1 Summary of model testing likelihood of PN L*H.*

```
Cov prior  : nuc_pre_text ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : pn_str_syl ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : ana_text ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Fixef prior: normal(sd = c(10, 2.5, ...), corr = c(0 ...), common.scale = FALSE)
Prior dev  : 51.12

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: isLH ~ ana_syls + foot_syls + wrd_end_syl + speech_rate + gender +
    (1 | speaker) + (1 | ana_text) + (1 | nuc_pre_text) + (1 |      pn_str_syl)
   Data: pn
Control: glmerControl(optimizer = "nlminbwrap", tol = 2e-05)

     AIC      BIC   logLik deviance df.resid
   566.6    636.6   -268.3    536.6      773

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7836 -0.2721  0.1015  0.3151 13.1239

Random effects:
 Groups       Name        Variance Std.Dev.
 nuc_pre_text (Intercept) 0.1139   0.3374
 speaker      (Intercept) 3.1132   1.7644
 pn_str_syl   (Intercept) 0.2511   0.5011
 ana_text     (Intercept) 3.2248   1.7958
Number of obs: 788, groups:
nuc_pre_text, 12; speaker, 11; pn_str_syl, 8; ana_text, 7

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.7173     1.9081   2.472 0.013425 *
ana_syls1     -1.1976     1.4384  -0.833 0.405064
ana_syls2      1.4729     1.7800   0.827 0.407985
ana_syls3      1.6939     1.7969   0.943 0.345840
foot_syls2     1.0260     0.6880   1.491 0.135888
foot_syls3     2.3524     0.6746   3.487 0.000489 ***
foot_syls4     2.3690     0.8835   2.681 0.007332 **
wrd_end_syl2   2.5382     0.5875   4.320 1.56e-05 ***
wrd_end_syl3   2.8775     1.0137   2.839 0.004530 **
speech_rate   -0.9549     0.2506  -3.811 0.000139 ***
genderM       -1.5585     1.0149  -1.536 0.124636
```

*Table F1.2 ANOVA of model of PN L*H likelihood.*

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
ana_syls	3	564.92	4.33	.228	.254	
foot_syls	3	574.56	13.97	.003	.006	p < .05
wrd_end_syl	2	579.22	16.64	< .001	< .001	p < .05
speech_rate	1	579.52	14.93	< .001	< .001	p < .05
gender	1	567.53	2.94	.086	.108	

*Table F1.3 R^2 of PN L*H likelihood model.*

R2_conditional	R2_marginal
.4	.8

*Table F1.4 Predicted likelihood of PN L*H as an effect of* `ana_syls`.

ana_syls	predicted	conf.low	conf.high	std.error
0	0.29	0.02	0.89	1.55
1	0.11	< 0.01	0.61	1.29
2	0.64	0.04	0.99	1.89
3	0.69	0.05	0.99	1.9

*Table F1.5 Predicted likelihood of PN L*H as an effect of* `foot_syls`.

foot_syls	predicted	conf.low	conf.high	std.error
1	0.29	0.02	0.89	1.55
2	0.53	0.06	0.96	1.51
3	0.81	0.18	0.99	1.5
4	0.81	0.17	0.99	1.57

*Table F1.6 Predicted likelihood of PN L*H as an effect of* `wrd_end_syl`.

wrd_end_syl	predicted	conf.low	conf.high	std.error
1	0.29	0.02	0.89	1.55
2	0.84	0.18	0.99	1.61
3	0.88	0.18	1	1.79

*Table F1.7 Predicted likelihood of PN L*H as an effect of* `speech_rate`.

speech_rate	predicted	conf.low	conf.high	std.error
2	0.94	0.39	1	1.65
3	0.86	0.23	0.99	1.57
4	0.71	0.11	0.98	1.52
5	0.49	0.05	0.95	1.52
6	0.27	0.02	0.88	1.55
7	0.12	< 0.01	0.77	1.63
8	0.05	< 0.01	0.62	1.73
9	0.02	< 0.01	0.45	1.87
10	< 0.01	< 0.01	0.3	2.02

*Table F1.8 Predicted likelihood of PN L*H as an effect of* `gender`.

gender	predicted	conf.low	conf.high	std.error
F	0.29	0.02	0.89	1.55
M	0.08	< 0.01	0.65	1.56

*Table F1.9 Estimated intercepts of each level of each factor (b0) re likelihood of PN L*H (odds ratio).*

term	estimate	conf.low	conf.high	std.error	z.value	p.value
(Intercept)	111.90	2.66	4708.30	213.45	2.47	.013
ana_syls1	0.30	0.02	5.06	0.43	-0.83	.405
ana_syls2	4.36	0.13	142.80	7.76	0.83	.408
ana_syls3	5.44	0.16	184.20	9.78	0.94	.346
foot_syls2	2.79	0.72	10.70	1.92	1.49	.136
foot_syls3	10.50	2.80	39.40	7.09	3.49	< .001
foot_syls4	10.70	1.89	60.40	9.44	2.68	.007
wrd_end_syl2	12.70	4.00	40.00	7.44	4.32	< .001
wrd_end_syl3	17.80	2.44	129.60	18.01	2.84	.004
speech_rate	0.38	0.23	0.63	0.1	-3.81	< .001
genderM	0.21	0.03	1.54	0.21	-1.54	.125

*Table F1.10 Pairwise comparisons of each level of each factor (b1) re likelihood of PN L*H (odds ratio).*

intercept	slope	estimate	conf.low	conf.high	std.error	z.value	p.value
ana_syls0	ana_syls1	0.30	0.02	5.06	0.43	-0.83	.405
ana_syls0	ana_syls2	4.36	0.13	142.80	7.76	0.83	.408
ana_syls0	ana_syls3	5.44	0.16	184.20	9.78	0.94	.346
ana_syls1	ana_syls2	6.61	0.21	211.60	11.69	1.07	.286
ana_syls1	ana_syls3	8.23	0.25	275.60	14.75	1.18	.239
ana_syls2	ana_syls3	3.03	0.10	90.90	5.26	0.64	.522
foot_syls1	foot_syls2	2.79	0.72	10.70	1.92	1.49	.136
foot_syls1	foot_syls3	10.50	2.80	39.40	7.09	3.49	< .001
foot_syls1	foot_syls4	10.70	1.89	60.40	9.44	2.68	.007
foot_syls2	foot_syls3	4.31	1.34	13.80	2.56	2.46	.014
foot_syls2	foot_syls4	5.94	1.03	34.30	5.31	1.99	.046
foot_syls3	foot_syls4	1.56	0.36	6.64	1.15	0.6	.55
wrd_end_syl1	wrd_end_syl2	12.70	4.00	40.00	7.44	4.32	< .001
wrd_end_syl1	wrd_end_syl3	17.80	2.44	129.60	18.01	2.84	.005
wrd_end_syl2	wrd_end_syl3	1.91	0.34	10.70	1.68	0.73	.464
intercept	speech_rate	0.39	0.24	0.63	0.1	-3.78	< .001
intercept	genderM	0.21	0.03	1.56	0.22	-1.52	.128

F2. Likelihood of H* (`isHStar`) in PNs in A- and H-Corpora.

Table F2.1 Summary of model testing likelihood of PN H.*

```
Cov prior  : nuc_pre_text ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : pn_str_syl ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : ana_text ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Fixef prior: normal(sd = c(10, 2.5, ...), corr = c(0 ...), common.scale = FALSE)
Prior dev  : 55.3384

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: isHStar ~ ana_syls + foot_syls + wrd_end_syl + speech_rate +
    gender + (1 | speaker) + (1 | ana_text) + (1 | nuc_pre_text) +
    (1 | pn_str_syl)
   Data: pn
Control: glmerControl(optimizer = "bobyqa", tol = 2e-05)

     AIC      BIC   logLik deviance df.resid
   504.4    574.4   -237.2    474.4      773

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.2716 -0.3011 -0.0812  0.0609  4.0985

Random effects:
 Groups       Name        Variance Std.Dev.
 nuc_pre_text (Intercept) 0.09527  0.3087
 speaker      (Intercept) 2.38777  1.5452
 pn_str_syl   (Intercept) 0.19002  0.4359
 ana_text     (Intercept) 0.42374  0.6510
Number of obs: 788, groups:
nuc_pre_text, 12; speaker, 11; pn_str_syl, 8; ana_text, 7

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.2046     1.5364  -2.086 0.036997 *
ana_syls1      1.3642     0.8885   1.535 0.124677
ana_syls2     -2.5842     1.5777  -1.638 0.101429
ana_syls3     -2.6283     1.5796  -1.664 0.096138 .
foot_syls2     0.3448     0.6308   0.547 0.584608
foot_syls3    -0.9431     0.6447  -1.463 0.143513
foot_syls4    -1.0354     0.9230  -1.122 0.261991
wrd_end_syl2  -2.4769     0.6400  -3.870 0.000109 ***
wrd_end_syl3  -2.9042     1.4456  -2.009 0.044539 *
speech_rate    0.1541     0.2386   0.646 0.518327
genderM        2.4351     0.9243   2.635 0.008424 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table F2.2 ANOVA of model testing likelihood of PN H.*

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
ana_syls	3	513.02	14.64	.002	.005	p < .05
foot_syls	3	506.79	8.41	.038	.055	
wrd_end_syl	2	518.4	18.02	< .001	< .001	p < .05
speech_rate	1	502.15	-0.23	1	1	
gender	1	509.96	7.58	.006	.01	p < .05

Table F2.3 R^2 of PN H likelihood model.*

R2_conditional	R2_marginal
.76	.51

Table F2.4 Predicted likelihood of PN H as an effect of* `ana_syls`.

ana_syls	predicted	conf.low	conf.high	std.error
0	0.09	0.01	0.48	1.13
1	0.28	0.06	0.7	0.92
2	< 0.01	< 0.01	0.18	1.73
3	< 0.01	< 0.01	0.17	1.71

Table F2.5 Predicted likelihood of PN H as an effect of* `foot_syls`.

foot_syls	predicted	conf.low	conf.high	std.error
1	0.09	0.01	0.48	1.13
2	0.12	0.02	0.54	1.08
3	0.04	< 0.01	0.24	1.06
4	0.03	< 0.01	0.29	1.25

Table F2.6 Predicted likelihood of PN H as an effect* `wrd_end_syl`.

wrd_end_syl	predicted	conf.low	conf.high	std.error
1	0.09	0.01	0.48	1.13
2	< 0.01	< 0.01	0.09	1.26
3	< 0.01	< 0.01	0.15	1.76

Table F2.8 Predicted likelihood of PN H as an effect of* `speech_rate`.

speech_rate	predicted	conf.low	conf.high	std.error
2	0.05	< 0.01	0.39	1.25
3	0.06	< 0.01	0.38	1.15
4	0.07	< 0.01	0.39	1.09
5	0.08	0.01	0.43	1.09
6	0.09	0.01	0.49	1.14
7	0.11	0.01	0.57	1.23
8	0.12	< 0.01	0.67	1.36
9	0.14	< 0.01	0.76	1.52
10	0.16	< 0.01	0.84	1.69

Table F2.9 Predicted likelihood of PN H as an effect of* `gender`.

gender	predicted	conf.low	conf.high	std.error
F	0.09	0.01	0.48	1.13
M	0.53	0.11	0.91	1.13

Table F2.10 Estimated intercepts of each level of each factor (b0) re likelihood of PN H (odds ratio).*

intercept	estimate	conf.low	conf.high	std.error	z.value	p.value
ana_syls0	0.04	0.00	0.82	0.06	-2.09	.037
ana_syls1	0.10	0.00	2.12	0.16	-1.47	.142
ana_syls2	0.00	0.00	0.00	0	-25794	< .001
ana_syls3	0.00	0.00	0.00	0	-5.47	< .001
foot_syls1	0.04	0.00	0.82	0.06	-2.09	.037
foot_syls2	0.07	0.07	0.07	0	-3916	< .001
foot_syls3	0.03	0.00	0.65	0.04	-2.23	.026
foot_syls4	0.02	0.00	0.66	0.03	-2.18	.029
wrd_end_syl1	0.04	0.00	0.82	0.06	-2.09	.037
wrd_end_syl2	0.00	0.00	0.08	0.01	-3.48	< .001
wrd_end_syl3	0.00	0.00	0.11	0	-3.11	.002

Table F2.11 Pairwise comparisons of each level of each factor (b1) re likelihood of PN H (odds ratio).*

intercept	slope	estimate	conf.low	conf.high	std.error	z.value	p.value
ana_syls0	ana_syls1	3.91	0.69	22.30	3.48	1.54	.125
ana_syls0	ana_syls2	0.07	0.00	1.66	0.12	-1.64	.101
ana_syls0	ana_syls3	0.07	0.00	1.60	0.11	-1.66	.096
ana_syls1	ana_syls2	0.04	0.00	0.68	0.06	-2.22	.026
ana_syls1	ana_syls3	0.04	0.00	0.64	0.05	-2.27	.023
ana_syls2	ana_syls3	0.26	0.26	0.26	0	-3089	< .001
foot_syls1	foot_syls2	1.41	0.41	4.86	0.89	0.55	.585
foot_syls1	foot_syls3	0.39	0.11	1.38	0.25	-1.46	.144
foot_syls1	foot_syls4	0.36	0.06	2.17	0.33	-1.12	.262
foot_syls2	foot_syls3	0.32	0.32	0.32	0	-1605	< .001
foot_syls2	foot_syls4	0.42	0.42	0.42	0	-1271	< .001
foot_syls3	foot_syls4	1.65	0.35	7.87	1.32	0.63	.527
wrd_end_syl1	wrd_end_syl2	0.08	0.02	0.30	0.05	-3.87	< .001
wrd_end_syl1	wrd_end_syl3	0.06	0.00	0.93	0.08	-2.01	.045
wrd_end_syl2	wrd_end_syl3	0.26	0.01	7.20	0.44	-0.79	.428
intercept	speech_rate	1.16	0.73	1.85	0.28	0.64	.524
intercept	genderM	11.40	1.84	70.70	10.61	2.61	.009

Appendix G. LMEMs of PN Phonetic Parameters in A- and H-Corpora

G1. Effect sizes (partial omega-squared) of fixed effects on PN pitch accents.

Table G1.1 Effect size (ω_p^2) of each lexical and metrical fixed effect in PN tonal *target LME models.*

	l_t		**l_f0**		**h_t**		**h_f0**	
Parameter	ω_p^2	95% CI	ω_p^2	95% CI	ω_p^2	95% CI	ω_p^2	95% CI
acc_phon	.01	[0, .04]	.00	[0, .04]	.43	[.36, .49]	0	[0, 0]
ana_syls	.2	[0, .49]	.07	[.03, .12]	.76	[.42, .87]	.35	[0, .68]
foot_syls	.01	[0, .02]	-.21	[0, 0]	.52	[.21, .68]	.25	[0, .46]
wrd_end_syl	.04	[.01, .09]	0	[0, 0]	-.01	[0, 0]	-.05	[0, 0]
pn_new_word	-.1	[0, 0]	.02	[0, .05]	-.02	[0, 0]	.1	[0, .66]
gender	.89	[.69, .95]	.59	[.01, .83]	.17	[0, .54]	.77	[.38, .89]

G2. Temporal Alignment of L Target (`l_t`) in PN Pitch Accents

Table G2.1 Summary of PN `l_t` model.

```
Formula:
l_t ~ acc_phon + ana_syls + foot_syls + wrd_end_syl + pn_new_word + gender + (1 |
speaker) + (1 | pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: l_t_equation
Data: pn_l_t_data.trimmed
Control: lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl =
list(method = "nlminb",      starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 4481.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2322 -0.5633  0.0551  0.6304  2.7388

Random effects:
 Groups     Name        Variance Std.Dev.
 speaker    (Intercept)  112.6   10.61
 pn_str_syl (Intercept) 1396.5   37.37
 Residual                627.9   25.06
Number of obs: 486, groups:  speaker, 11; pn_str_syl, 8

Fixed effects:
                Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       46.205     18.135  11.746   2.548   0.0259 *
acc_phonL*        24.379      9.659 471.180   2.524   0.0119 *
ana_syls1        -32.515     41.203   5.458  -0.789   0.4629
ana_syls2        -31.099     41.265   5.518  -0.754   0.4820
ana_syls3        -20.607     41.304   5.536  -0.499   0.6370
foot_syls2         1.519      6.944 457.231   0.219   0.8270
foot_syls3         6.501      6.841 465.647   0.950   0.3425
foot_syls4        18.287      8.530 453.503   2.144   0.0326 *
wrd_end_syl2      23.556      5.715 456.381   4.122 4.46e-05 ***
wrd_end_syl3      14.533     12.585 312.103   1.155   0.2491
pn_new_wordTRUE   23.163     40.540   5.204   0.571   0.5915
genderM          -70.145      7.005   9.774 -10.013 1.88e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table G2.2 ANOVA of PN `l_t` model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
acc_phon	4000	4000	1	471.18	6.37	.012	.023	p < .05
ana_syls	4447	1482	3	12.19	2.36	.122	.184	
foot_syls	3980	1327	3	453.9	2.11	.098	.161	
wrd_end_syl	11103	5552	2	350.73	8.84	<.001	<.001	p < .05
pn_new_word	204.99	204.99	1	5.2	0.33	.592	.630	
gender	62954	62954	1	9.77	100.26	<.001	<.001	p < .05

Table G2.3 R^2 of PN `l_t` model.

R2_conditional	R2_marginal
.83	.42

Table G2.4 Predicted values of PN `l_t` re `acc_phon` (ms).

acc_phon	predicted	conf.low	conf.high	std.error
L*H	46.2	10.66	81.75	18.14
L*	70.58	32.21	108.95	19.58

Table G2.5 Predicted values of PN `l_t` re `ana_syls` (ms).

ana_syls	predicted	conf.low	conf.high	std.error
0	46.2	10.66	81.75	18.14
1	13.69	-61.78	89.16	38.51
2	15.11	-61.12	91.33	38.89
3	25.6	-50.63	101.82	38.89

Table G2.6 Predicted values of PN `l_t` re `foot_syls` (ms).

foot_syls	predicted	conf.low	conf.high	std.error
1	46.2	10.66	81.75	18.14
2	47.72	13.05	82.4	17.69
3	52.71	18.08	87.33	17.67
4	64.49	29.08	99.9	18.07

Table G2.7 Predicted values of PN `l_t` re `wrd_end_syl` (ms).

wrd_end_syl	predicted	conf.low	conf.high	std.error
1	46.2	10.66	81.75	18.14
2	69.76	33.57	105.96	18.47
3	60.74	26.9	94.58	17.27

Table G2.8 Predicted values of PN 7_t re pn_new_wrd (ms).

pn_new_word	predicted	conf.low	conf.high	std.error
FALSE	46.2	10.66	81.75	18.14
TRUE	69.37	-24.87	163.6	48.08

Table G2.9 Predicted values of PN 7_t re gender (ms).

gender	predicted	conf.low	conf.high	std.error
F	46.2	10.66	81.75	18.14
M	-23.94	-59.95	12.07	18.37

Table G2.10 Pairwise comparison of effects of levels of fixed effects (b1) on PN 7_t (ms).

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
ana_syls0	ana_syls1	-32.5	-135.8	70.8	41.20	-0.79	5.46	.463
ana_syls0	ana_syls2	-31.1	-134.2	72	41.26	-0.75	5.52	.482
ana_syls0	ana_syls3	-20.6	-123.8	82.5	41.30	-0.5	5.54	.637
ana_syls1	ana_syls2	1.42	-8.91	11.7	5.25	0.27	466.33	.788
ana_syls1	ana_syls3	11.9	1.57	22.3	5.26	2.26	466.47	.024
ana_syls2	ana_syls3	10.5	0.78	20.2	4.94	2.12	459.49	.034
foot_syls1	foot_syls2	1.52	-12.1	15.2	6.94	0.22	457.23	.827
foot_syls1	foot_syls3	6.5	-6.94	19.9	6.84	0.95	465.65	.342
foot_syls1	foot_syls4	18.3	1.52	35.1	8.53	2.14	453.50	.033
foot_syls2	foot_syls3	4.98	-7.46	17.4	6.32	0.79	458.11	.431
foot_syls2	foot_syls4	16.8	0.833	32.7	8.11	2.07	437.12	.039
foot_syls3	foot_syls4	11.8	1.63	21.9	5.17	2.28	466.10	.023
wrd_end_syl1	wrd_end_syl2	23.6	12.3	34.8	5.71	4.12	456.38	< .001
wrd_end_syl1	wrd_end_syl3	14.5	-10.2	39.3	12.58	1.15	312.10	.249
wrd_end_syl2	wrd_end_syl3	-9.02	-31	13	11.18	-0.81	237.39	.42
intercept	acc_phonL*	24.4	5.4	43.4	9.66	2.52	471.18	.012
intercept	pn_new_wordT	23.2	-79.8	126.2	40.54	0.57	5.20	.592
intercept	genderM	-70.1	-85.8	-54.5	7	-10.01	9.77	< .001

G3. Temporal alignment of L target (`l_f0`) in PN pitch accents

Table G3.1 Summary of PN `l_f0` model.

```
Formula:
l_f0 ~ acc_phon + ana_syls + foot_syls + wrd_end_syl + pn_new_word + gender + (1
+ foot_syls | speaker)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: l_f0 ~ acc_phon + ana_syls + foot_syls + wrd_end_syl + pn_new_word +
    gender + (1 + foot_syls | speaker)
   Data: pn_l_f0.trimmedControl:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 1597.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2154 -0.6221 -0.0179  0.5683  3.3374

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 speaker  (Intercept) 1.22694  1.1077
          foot_syls2  0.06781  0.2604   -0.57
          foot_syls3  0.86240  0.9287   -0.82  0.78
          foot_syls4  0.78958  0.8886   -0.87  0.49  0.59
 Residual             1.31776  1.1479
Number of obs: 496, groups:  speaker, 11

Fixed effects:
                 Estimate Std. Error        df t value Pr(>|t|)
(Intercept)      -0.71571    0.43054   6.31720  -1.662   0.1450
acc_phonL*       -0.62192    0.49554 157.29452  -1.255   0.2113
ana_syls1         0.87102    0.17983 450.04182   4.844 1.76e-06 ***
ana_syls2         0.19498    0.28433 452.76339   0.686   0.4932
ana_syls3         0.06817    0.28547 452.80563   0.239   0.8114
foot_syls2       -0.25526    0.25462   4.58383  -1.002   0.3660
foot_syls3       -0.16476    0.40027   8.76796  -0.412   0.6905
foot_syls4       -0.06302    0.40306  13.05163  -0.156   0.8781
wrd_end_syl2     -0.20359    0.18229 458.76717  -1.117   0.2646
wrd_end_syl3     -0.36455    0.28948 457.94374  -1.259   0.2085
pn_new_wordTRUE  -0.73239    0.24798 459.69223  -2.953   0.0033 **
genderM           0.97467    0.27163   6.15808   3.588   0.0110 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table G3.2 ANOVA of PN `l_f0` model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
acc_phon	2.08	2.08	1	157.29	1.57	.211	.303	
ana_syls	51.26	17.09	3	451.06	12.97	<.001	<.001	p < .05
foot_syls	1.44	0.48	3	7.02	0.36	.781	.806	
wrd_end_syl	2.32	1.16	2	458.78	0.88	.415	.507	
pn_new_word	11.49	11.49	1	459.69	8.72	.003	.008	p < .01
gender	16.97	16.97	1	6.16	12.88	.011	.023	p < .05

Table G3.3 R^2 of PN `l_f0` model.

R2_conditional	R2_marginal
.41	.19

Table G3.4 Predicted values of PN `l_f0` re `acc_phon` (ST re speaker median).

acc_phon	predicted	conf.low	conf.high	std.error
L*H	-0.72	-1.56	0.13	0.43
L*	-1.34	-2.54	-0.14	0.61

Table G3.5 Predicted values of PN `l_f0` re `ana_syls` (ST re speaker median).

ana_syls	predicted	conf.low	conf.high	std.error
0	-0.72	-1.56	0.13	0.43
1	0.16	-0.76	1.07	0.47
2	-0.52	-1.57	0.53	0.54
3	-0.65	-1.7	0.41	0.54

Table G3.6 Predicted values of PN `l_f0` re `foot_syls` (ST re speaker median).

foot_syls	predicted	conf.low	conf.high	std.error
1	-0.72	-1.56	0.13	0.43
2	-0.97	-1.73	-0.21	0.39
3	-0.88	-1.43	-0.33	0.28
4	-0.78	-1.38	-0.18	0.3

Table G3.7 Predicted values of PN `l_f0` re `wrd_end_syl` (ST re speaker median).

wrd_end_syl	predicted	conf.low	conf.high	std.error
1	-0.72	-1.56	0.13	0.43
2	-0.92	-1.84	-2.5e-03	0.47
3	-1.08	-2.07	-0.09	0.51

Table G3.8 Predicted values of PN `l_f0` re pn_new_wrd (ST re speaker median).

pn_new_word	predicted	conf.low	conf.high	std.error
FALSE	-0.72	-1.56	0.13	0.43
TRUE	-1.45	-2.35	-0.55	0.46

Table G3.9 Predicted values of PN `l_f0` re `gender` (ST re speaker median).

gender	predicted	conf.low	conf.high	std.error
F	-0.72	-1.56	0.13	0.43
M	0.26	-0.59	1.11	0.43

Table G3.10 Pairwise comparison of effects of levels of fixed effects (b1) on PN l_f0 *(ST re speaker median).*

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
ana_syls0	ana_syls1	0.87	0.52	1.22	0.18	4.84	450.04	< .001
ana_syls0	ana_syls2	0.19	-0.36	0.75	0.28	0.69	452.76	.493
ana_syls0	ana_syls3	0.07	-0.49	0.63	0.29	0.24	452.81	.811
ana_syls1	ana_syls2	-0.68	-1.12	-0.24	0.22	-3.07	453.85	.002
ana_syls1	ana_syls3	-0.8	-1.24	-0.37	0.22	-3.62	453.71	< .001
ana_syls2	ana_syls3	-0.13	-0.57	0.32	0.23	-0.56	445.73	.576
foot_syls1	foot_syls2	-0.26	-0.93	0.42	0.25	-1.00	4.58	.366
foot_syls1	foot_syls3	-0.16	-1.07	0.74	0.40	-0.41	8.77	.691
foot_syls1	foot_syls4	-0.06	-0.93	0.81	0.40	-0.16	13.05	.878
foot_syls2	foot_syls3	0.09	-0.83	1.01	0.35	0.26	4.61	.806
foot_syls2	foot_syls4	0.19	-0.6	0.98	0.38	0.51	17.94	.616
foot_syls3	foot_syls4	0.1	-0.58	0.79	0.33	0.31	17.65	.759
wrd_end_syl1	wrd_end_syl2	-0.2	-0.56	0.16	0.18	-1.12	458.77	.265
wrd_end_syl1	wrd_end_syl3	-0.36	-0.93	0.2	0.29	-1.26	457.94	.209
wrd_end_syl2	wrd_end_syl3	-0.16	-0.60	0.28	0.23	-0.71	455.86	.475
intercept	acc_phonL*	-0.62	-1.60	0.36	0.5	-1.26	157.30	.211
intercept	pn_new_wordTR	-0.73	-1.22	-0.24	0.25	-2.95	459.69	.003
intercept	genderM	0.98	0.31	1.64	0.27	3.59	6.16	.011

G4. Alternative model of L target temporal alignment (`l_t`) in PN pitch accents

Table G4.1 Summary of alternative PN `l_t` model with `has_ana_syls` + `wrd_end_syl_late`.

```
Formula:
l_t ~ acc_phon + has_ana_syls + wrd_end_syl_late + gender + (1 + foot_syls |
speaker) + (1 | pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: l_t_equation_lateControl:
lmerControl(optimizer = "nloptwrap", optCtrl = list(algorithm =
"NLOPT_LN_NEWUOA_BOUND",
    maxfun = 1e+09, maxeval = 1e+07, xtol_abs = 1e-09, ftol_abs = 1e-09))

REML criterion at convergence: 4558.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3776 -0.5695  0.0189  0.6205  2.8647

Random effects:
 Groups     Name        Variance Std.Dev. Corr
 speaker    (Intercept)  752.5   27.43
            foot_syls2   104.9   10.24    -0.86
            foot_syls3  1270.3   35.64    -0.83  0.84
            foot_syls4   481.7   21.95    -0.95  0.97  0.85
 pn_str_syl (Intercept) 1660.3   40.75
 Residual                574.2   23.96
Number of obs: 490, groups:  speaker, 11; pn_str_syl, 8

Fixed effects:
                     Estimate Std. Error      df t value Pr(>|t|)
(Intercept)            39.640     15.788   9.127   2.511 0.032933 *
acc_phonL*             43.587     11.014 171.953   3.957 0.000111 ***
has_ana_sylsTRUE        3.552      4.513 455.902   0.787 0.431715
wrd_end_syl_lateTRUE   26.740      5.375 454.908   4.975 9.27e-07 ***
genderM               -71.810      6.078   9.375 -11.816 6.10e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) acc_L* h__TRU w___TR
acc_phonL*  -0.031
hs_n_syTRUE -0.202  0.008
wrd_n__TRUE -0.121  0.062  0.008
genderM     -0.159 -0.008 -0.027  0.000
optimizer (nloptwrap) convergence code: 0 (OK)
```

Table G4.2 ANOVA of PN `l_t` alternative model using `has_ana_syls` + `wrd_end_syl`_late.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
acc_phon	8992	8992	1	171.95	15.66	<.001	<.001	p < .05
has_ana_syls	355.6	355.6	1	455.9	0.62	.432	.506	
wrd_end_syl_late	14211	14211	1	454.91	24.75	<.001	<.001	p < .05
gender	80163	80163	1	9.38	139.61	<.001	<.001	p < .05

Table G4.3 R^2 of PN alternative PN `l_t` model.

R2_conditional	R2_marginal
.86	.29

Table G4.4 Alternative model predicted values of PN `l_t` re `acc_phon` (ms).

acc_phon	predicted	conf.low	conf.high	std.error
L*H	39.64	8.7	70.58	15.79
L*	83.23	46.05	120.4	18.97

Table G4.5 Alternative model predicted values of PN `l_t` re `has_has_syls` (ms).

ana_syls	predicted	conf.low	conf.high	std.error
FALSE	39.64	8.7	70.58	15.79
TRUE	43.19	12.77	73.61	15.52

Table G4.6 Alternative model predicted values of PN `l_t` re `foot_syls` (ms).

has_ana_syls	predicted	conf.low	conf.high	std.error
1	46.2	10.66	81.75	18.14
2	47.72	13.05	82.4	17.69
3	52.71	18.08	87.33	17.67
4	64.49	29.08	99.9	18.07

Table G4.7 Alternative model predicted values of PN `l_t` re `wrd_end_syl_late` (ms).

wrd_end_syl_late	predicted	conf.low	conf.high	std.error
FALSE	39.64	8.7	70.58	15.79
TRUE	66.38	34.92	97.84	16.05

Table G4.8 Alternative model predicted values of PN `l_t` re `gender` (ms).

gender	predicted	conf.low	conf.high	std.error
F	39.64	8.7	70.58	15.79
M	-32.17	-63.51	-0.83	15.99

Table G4.9 Summary and CIs of alternative PN `l_t` model using `wrd_end_syl_late + has_ana_syls`.

term	estimate	conf.low	conf.high	std.error	t.value	df	p.value
(Intercept)	39.64	4	75.28	15.79	2.51	9.13	0.033
acc_phonL*	43.59	21.85	65.33	11.01	3.96	171.95	< .001
has_ana_sylsT	3.55	-5.32	12.42	4.51	0.79	455.9	0.432
wrd_end_syl_lateT	26.74	16.18	37.3	5.38	4.97	454.91	< .001
genderM	-71.81	-85.47	-58.14	6.08	-11.82	9.37	< .001

G5. Temporal alignment of H target (h_t) in PN pitch accents

Table G5.1 Summary of PN h_t model.

```
Formula:
h_t ~ acc_phon + ana_syls + foot_syls + wrd_end_syl + pn_new_word + gender + (1 +
foot_syls | speaker) + (1 | pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: h_t_equation
   Data: pn_h_t.trimmedControl: lmerControl(optimizer = "nloptwrap", optCtrl =
list(algorithm = "NLOPT_LN_PRAXIS",
    maxfun = 1e+09, maxeval = 1e+07, xtol_abs = 1e-09, ftol_abs = 1e-09))

REML criterion at convergence: 7182.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.4934 -0.6360 -0.0906  0.5600  3.1807

Random effects:
 Groups     Name        Variance Std.Dev. Corr
 speaker    (Intercept)  265.5   16.29
            foot_syls2   252.3   15.88    0.41
            foot_syls3  1126.6   33.56    0.91 0.23
            foot_syls4   608.8   24.67    0.80 0.77 0.79
 pn_str_syl (Intercept)  446.7   21.13
 Residual               1298.3   36.03
Number of obs: 718, groups:  speaker, 11; pn_str_syl, 8

Fixed effects:
                Estimate Std. Error       df t value Pr(>|t|)
(Intercept)     184.0255    15.5110  18.5473  11.864 4.22e-10 ***
acc_phon>H*     -32.2362     5.9861 614.8998  -5.385 1.03e-07 ***
acc_phonH*      -79.8318     4.6497 287.6566 -17.169  < 2e-16 ***
ana_syls1       -37.3458    24.9283   5.9002  -1.498   0.1856
ana_syls2         0.7885    25.4247   6.6438   0.031   0.9762
ana_syls3        12.2865    25.4991   6.6825   0.482   0.6453
foot_syls2       39.6293     8.7665  29.0933   4.521 9.53e-05 ***
foot_syls3       66.6114    12.5780  16.7271   5.296 6.25e-05 ***
foot_syls4       67.7866    12.3095  37.3050   5.507 2.85e-06 ***
wrd_end_syl2      6.3148     6.6093 393.2917   0.955   0.3399
wrd_end_syl3     16.3784    14.7875  54.6311   1.108   0.2729
pn_new_wordTRUE -21.5277    23.8696   5.7616  -0.902   0.4032
genderM         -19.9582    10.5187  10.2622  -1.897   0.0862 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table G5.2 ANOVA of PN h_t model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
acc_phon	3.8×10^{05}	1.9×10^{05}	2	381.45	147.4	<.001	<.001	p < .05
ana_syls	77283	25761	3	14.15	19.84	<.001	<.001	p < .05
foot_syls	48696	16232	3	28.44	12.5	<.001	<.001	p < .05
wrd_end_syl	1887	943.41	2	87.34	0.73	.486	.554	
pn_new_word	1056	1056	1	5.76	0.81	.403	.507	
gender	4674	4674	1	10.26	3.6	.086	.150	

Table G5.3 R^2 of PN h_t model.

R2_conditional	R2_marginal
.83	.59

Table G5.4 Predicted values of PN h_t re acc_phon (ms).

acc_phon	predicted	conf.low	conf.high	std.error
L*H	184.03	153.62	214.43	15.51
>H*	151.79	119.89	183.69	16.27
H*	104.19	73.37	135.02	15.73

Table G5.5 Predicted values of PN h_t re ana_syls (ms).

ana_syls	predicted	conf.low	conf.high	std.error
0	184.03	153.62	214.43	15.51
1	146.68	99.79	193.57	23.92
2	184.81	135.96	233.66	24.92
3	196.31	147.46	245.17	24.93

Table G5.6 Predicted values of PN h_t re foot_syls (ms).

foot_syls	predicted	conf.low	conf.high	std.error
1	184.03	153.62	214.43	15.51
2	223.65	191.55	255.76	16.38
3	250.64	210.88	290.4	20.29
4	251.81	214.85	288.78	18.86

Table G5.7 Predicted values of PN h_t re wrd_end_syl (ms).

wrd_end_syl	predicted	conf.low	conf.high	std.error
1	184.03	153.62	214.43	15.51
2	190.34	159.38	221.3	15.8
3	200.4	171.62	229.19	14.69

Table G5.8 Predicted values of PN h_t re pn_new_wrd (ms).

pn_new_word	predicted	conf.low	conf.high	std.error
FALSE	184.03	153.62	214.43	15.51
TRUE	162.5	105.26	219.74	29.2

Table G5.9 Predicted values of PN h_t re gender (ms).

gender	predicted	conf.low	conf.high	std.error
F	184.03	153.62	214.43	15.51
M	164.07	132.96	195.17	15.87

Table G5.10 Pairwise comparison of effects of levels of fixed effects (b1) on PN h_t (ms).

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
acc_phonL*H	acc_phon>H*	-32.2	-44	-20.5	5.99	-5.39	614.90	< .001
acc_phonL*H	acc_phonH*	-79.8	-89	-70.7	4.65	-17.17	287.66	< .001
acc_phon>H*	acc_phonH*	-47.6	-60	-35.2	6.30	-7.56	399.87	< .001
ana_syls0	ana_syls1	-37.4	-98.6	23.9	24.93	-1.50	5.90	0.186
ana_syls0	ana_syls2	0.79	-60	61.6	25.42	0.03	6.64	0.976
ana_syls0	ana_syls3	12.3	-48.6	73.2	25.50	0.48	6.68	0.645
ana_syls1	ana_syls2	38.1	24.8	51.5	6.78	5.62	645.44	< .001
ana_syls1	ana_syls3	49.6	36.3	63	6.81	7.28	657.99	< .001
ana_syls2	ana_syls3	11.5	-1.94	24.9	6.84	1.68	672.00	0.093
foot_syls1	foot_syls2	39.6	21.7	57.6	8.77	4.52	29.09	< .001
foot_syls1	foot_syls3	66.6	40	93.2	12.58	5.30	16.73	< .001
foot_syls1	foot_syls4	67.8	42.9	92.7	12.31	5.51	37.31	< .001
foot_syls2	foot_syls3	27	0.89	53.1	12.36	2.18	16.94	0.043
foot_syls2	foot_syls4	28.2	7.071	49.2	10.59	2.66	75.54	0.01
foot_syls3	foot_syls4	1.18	-17.7	20	9.21	0.13	28.31	0.899
wrd_end_syl1	wrd_end_syl2	6.31	-6.68	19.3	6.61	0.96	393.29	0.34
wrd_end_syl1	wrd_end_syl3	16.4	-13.3	46	14.79	1.11	54.63	0.273
wrd_end_syl2	wrd_end_syl3	10.1	-16.1	36.3	12.93	0.78	36.15	0.441
intercept	pn_new_wordT	-21.5	-80.5	37.5	23.87	-0.90	5.76	0.403
intercept	genderM	-20	-43.3	3.398	10.52	-1.9	10.26	0.086

G6. f_0 of H target (h_f0) in PN pitch accents

Table G6.1 Summary of PN h_f0 model.

```
Formula:
h_f0 ~ acc_phon + ana_syls + foot_syls + wrd_end_syl + pn_new_word + gender + (1
+ foot_syls | speaker) + (1 | pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: h_f0_equation
   Data: pn_h_dataControl:
lmerControl(optCtrl = list(maxit = 1e+09, maxfun = 1e+09, xtol_abs = 1e-09,
    ftol_abs = 1e-09))

REML criterion at convergence: 2647.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6550 -0.6674 -0.0881  0.6258  3.6964

Random effects:
 Groups     Name        Variance Std.Dev. Corr
 speaker    (Intercept) 0.6130   0.7829
            foot_syls2  0.4397   0.6631    0.00
            foot_syls3  0.1125   0.3354   -0.78  0.30
            foot_syls4  0.9690   0.9844   -0.86 -0.12  0.48
 pn_str_syl (Intercept) 0.3203   0.5660
 Residual               1.9406   1.3931
Number of obs: 737, groups:  speaker, 11; pn_str_syl, 8

Fixed effects:
                 Estimate Std. Error        df t value Pr(>|t|)
(Intercept)       1.49572    0.50524  12.19988   2.960 0.011725 *
acc_phon>H*       0.23633    0.22898 614.72866   1.032 0.302438
acc_phonH*        0.03689    0.18036 540.08757   0.205 0.837994
ana_syls1         0.29501    0.69493   3.15757   0.425 0.698466
ana_syls2         0.36983    0.72834   4.05226   0.508 0.638006
ana_syls3        -0.32741    0.73109   4.07293  -0.448 0.677060
foot_syls2        0.74429    0.33802  21.08722   2.202 0.038940 *
foot_syls3        1.02835    0.29662  31.26625   3.467 0.001554 **
foot_syls4        0.67842    0.46586  32.14443   1.456 0.155016
wrd_end_syl2      0.10564    0.24323 171.77696   0.434 0.664614
wrd_end_syl3      0.13951    0.50846  20.25873   0.274 0.786579
pn_new_wordTRUE  -0.83767    0.66641   3.40474  -1.257 0.288074
genderM           1.41568    0.22636   9.42318   6.254 0.000122 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table G6.2 ANOVA of PN h_f0 model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
acc_phon	2.08	1.04	2	596.16	0.54	.586	.630	
ana_syls	16.6	5.53	3	6.28	2.85	.123	.185	
foot_syls	25.33	8.44	3	26.42	4.35	.013	.024	p < .05
wrd_end_syl	0.37	0.19	2	33.21	0.1	.909	.909	
pn_new_word	3.07	3.07	1	3.4	1.58	.288	.396	
gender	75.9	75.9	1	9.42	39.11	<.001	<.001	p < .05

Table G6.3 R^2 of PN h_f0 model.

R2_conditional	R2_marginal
.46	.23

Table G6.4 Predicted values of PN `h_f0` re `acc_phon` (ST re speaker median).

acc_phon	predicted	conf.low	conf.high	std.error
*H	1.5	0.51	2.49	0.51
>H*	1.73	0.67	2.8	0.54
H*	1.53	0.54	2.53	0.51

Table G6.5 Predicted values of PN `h_f0` re `ana_syls` (ST re speaker median).

ana_syls	predicted	conf.low	conf.high	std.error
0	1.5	0.51	2.49	0.51
1	1.79	0.41	3.17	0.7
2	1.87	0.39	3.34	0.75
3	1.17	-0.31	2.64	0.75

Table G6.6 Predicted values of PN `h_f0` re `foot_syls` (ST re speaker median).

foot_syls	predicted	conf.low	conf.high	std.error
1	1.5	0.51	2.49	0.51
2	2.24	1.21	3.27	0.53
3	2.52	1.65	3.39	0.44
4	2.17	1.25	3.1	0.47

Table G6.7 Predicted values of PN `h_f0` re `wrd_end_syl` (ST re speaker median).

wrd_end_syl	predicted	conf.low	conf.high	std.error
1	1.5	0.51	2.49	0.51
2	1.6	0.59	2.61	0.52
3	1.64	0.64	2.63	0.51

Table G6.8 Predicted values of PN `h_f0` re pn_new_wrd (ST re speaker median).

pn_new_word	predicted	conf.low	conf.high	std.error
FALSE	1.5	0.51	2.49	0.51
TRUE	0.66	-0.98	2.3	0.84

Table G6.9 Predicted values of PN `h_f0` re `gender` (ST re speaker median).

gender	predicted	conf.low	conf.high	std.error
F	1.5	0.51	2.49	0.51
M	2.91	1.89	3.93	0.52

Table G6.10 Pairwise comparison of levels of fixed effects (b1) on PN `h_f0` *re* `gender` *(ST re speaker median).*

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
acc_phonL*H	acc_phon>H*	0.24	-0.21	0.69	0.23	1.03	614.73	0.302
acc_phonL*H	acc_phonH*	0.04	-0.33	0.39	0.18	0.20	540.09	0.838
acc_phon>H*	acc_phonH*	-0.2	-0.69	0.29	0.25	-0.80	686.56	0.426
ana_syls0	ana_syls1	0.3	-1.86	2.44	0.69	0.42	3.16	0.698
ana_syls0	ana_syls2	0.37	-1.64	2.38	0.73	0.51	4.05	0.638
ana_syls0	ana_syls3	-0.33	-2.34	1.69	0.73	-0.45	4.07	0.677
ana_syls1	ana_syls2	0.07	-0.43	0.58	0.26	0.29	604.75	0.772
ana_syls1	ana_syls3	-0.62	-1.13	-0.11	0.26	-2.39	625.76	0.017
ana_syls2	ana_syls3	-0.7	-1.22	-0.18	0.26	-2.64	677.79	0.009
foot_syls1	foot_syls2	0.74	0.04	1.45	0.34	2.20	21.09	0.039
foot_syls1	foot_syls3	1.03	0.42	1.63	0.30	3.47	31.27	0.002
foot_syls1	foot_syls4	0.68	-0.27	1.63	0.47	1.46	32.14	0.155
foot_syls2	foot_syls3	0.28	-0.37	0.94	0.32	0.88	29.25	0.384
foot_syls2	foot_syls4	-0.07	-1.11	0.98	0.51	-0.13	24.28	0.898
foot_syls3	foot_syls4	-0.35	-1.10	0.40	0.36	-0.96	23.85	0.346
wrd_end_syl1	wrd_end_syl2	0.11	-0.37	0.59	0.24	0.43	171.78	0.665
wrd_end_syl1	wrd_end_syl3	0.14	-0.92	1.2	0.51	0.27	20.26	0.787
wrd_end_syl2	wrd_end_syl3	0.03	-0.9	0.97	0.43	0.08	14.01	0.939
intercept	pn_new_wordT	-0.84	-2.82	1.15	0.67	-1.25	3.40	0.288
intercept	genderM	1.42	0.91	1.92	0.23	6.25	9.42	1.2e-04

G7. Stress clash effects on prenuclear H alignment (h_t)

Table G7.1 Summary of alternative PN h_t model testing stress_clash effects.

```
Formula:
h_t ~ stress_clash + acc_phon + (1 + stress_clash | speaker) + (1 | gender) + (1
| pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: h_t_equation
   Data: pn_h_t_alt.trimmedControl:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 7436.7

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.53411 -0.65821 -0.08324  0.60228  2.89139

Random effects:
 Groups     Name              Variance Std.Dev. Corr
 speaker    (Intercept)       1229.1   35.06
            stress_clashTRUE   500.2   22.37    -0.99
 pn_str_syl (Intercept)       2157.6   46.45
 gender     (Intercept)        416.7   20.41
 Residual                     1626.5   40.33
Number of obs: 722, groups:  speaker, 11; pn_str_syl, 8; gender, 2

Fixed effects:
                 Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       109.653     24.552   2.996   4.466    0.021 *
stress_clashTRUE  -53.997     10.052  20.569  -5.372 2.68e-05 ***
acc_phon>H*        46.472      6.814 642.595   6.820 2.10e-11 ***
acc_phonL*H        83.608      4.907 305.713  17.038  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table G7.2 ANOVA of PN h_t alternative model testing stress_clash effects.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
stress_clash	46932	46932	1	20.57	28.86	<.001	<.001	p < .05
acc_phon	4.7×10^{05}	2.4×10^{05}	2	527.01	145.26	<.001	<.001	p < .05

Table G7.3 R^2 of PN h_t stress_clash model.

R2_conditional	R2_marginal
.74	.26

Table G7.4 Predicted values of PN h_t re acc_phon (ms) in model testing stress_clash effects.

acc_phon	predicted	conf.low	conf.high	std.error
H*	109.65	61.53	157.77	24.55
>H*	156.13	107.03	205.22	25.05
L*H	193.26	145.26	241.27	24.49

Table G7.5 Predicted values of PN `h_t` *re* `stress_clash` *(ms) in model testing* `stress_clash` *effects.*

stress_clash	predicted	conf.low	conf.high	std.error
FALSE	109.65	61.53	157.77	24.55
TRUE	55.66	10.06	101.25	23.26

Table G7.6 Pairwise comparison of fixed effects (b1) on PN `l_t` *(ms) in model testing* `stress_clash`*.*

intercept	slope	est.	conf. low	conf. high	std. error	t. value	df	p. value
acc_phonH*	acc_phon>H*	46.5	33.1	59.9	6.81	6.82	643	< .001
acc_phonH*	acc_phonL*H	83.6	74	93.3	4.91	17.04	306	< .001
acc_phon>H*	acc_phonL*H	37.1	24.4	49.9	6.49	5.72	698	< .001
intercept	stress_clashT	-54	-74.9	-33.1	10.05	-5.37	21	< .001

G8. Stress clash effects on prenuclear H f_0 (`h_f0`)

Table G8.1 Summary of alternative PN `h_f0` *model testing* `stress_clash` *effects.*

```
Formula:
h_f0 ~ stress_clash + (1 + stress_clash | speaker) + (1 | gender) + (1 |
pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: h_f0_equation
   Data: pn_h_f0_alt.trimmedControl:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 2613.9

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.76007 -0.68945 -0.04213  0.64059  3.02344

Random effects:
 Groups     Name              Variance Std.Dev. Corr
 speaker    (Intercept)       0.2015   0.4489
            stress_clashTRUE  0.4122   0.6421   0.40
 pn_str_syl (Intercept)       0.2412   0.4911
 gender     (Intercept)       1.0341   1.0169
 Residual                     1.9777   1.4063
Number of obs: 728, groups:  speaker, 11; pn_str_syl, 8; gender, 2

Fixed effects:
                 Estimate Std. Error      df t value Pr(>|t|)
(Intercept)        2.7707     0.7557  1.1197   3.666  0.14823
stress_clashTRUE  -0.9179     0.3104 17.2863  -2.957  0.00871 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table G8.2 ANOVA of PN `h_f0` *alternative model testing PN* `stress_clash` *effects.*

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
stress_clash	17.29	17.29	1	17.29	8.74	.009	.02	p < .05

Table G8.3 R^2 of PN `h_f0` `stress_clash` *model.*

R2_conditional	R2_marginal
.44	.02

Table G8.4 Predicted values of `h_f0` *re PN* `stress_clash` *(ms) in alternative model.*

stress_clash	predicted	conf.low	conf.high	std.error
FALSE	2.83	1.26	4.4	0.8
TRUE	1.94	0.28	3.61	0.85

Table G8.5 Summary and CIs of alternative PN `h_f0` *model testing* `stress_clash` *effects.*

term	estimate	conf.low	conf.high	std.error	t.value	df	p.value
(Intercept)	2.83	-5.17	10.82	0.8	3.54	1.11	0.155
stress_clashTRUE	-0.88	-1.49	-0.28	0.29	-3.01	23.57	0.006

G9. f_0 excursion in prenuclear L*H

Table G9.1 Summary of PN `f0_exc` *model.*

```
Formula:
f0_exc ~ foot_syls + (1 | ana_syls) + (1 | speaker) + (1 | gender) + (1 |
pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: f0_exc_equation
   Data: pn_f0_exc.trimmedControl:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 1321.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.7738 -0.6537 -0.0548  0.5866  3.3266

Random effects:
 Groups     Name        Variance Std.Dev.
 speaker    (Intercept) 0.1681   0.4100
 pn_str_syl (Intercept) 0.5206   0.7215
 ana_syls   (Intercept) 0.1643   0.4053
 gender     (Intercept) 0.1091   0.3302
 Residual               0.8003   0.8946
Number of obs: 484, groups:  speaker, 11; pn_str_syl, 8; ana_syls, 4; gender, 2

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)   2.3543     0.4739   7.0995   4.968 0.001558 **
foot_syls2    0.5798     0.2549 331.4416   2.274 0.023578 *
foot_syls3    0.8813     0.2445 320.7600   3.605 0.000362 ***
foot_syls4    0.6131     0.2991 256.6059   2.050 0.041413 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
           (Intr) ft_sy2 ft_sy3
foot_syls2 -0.345
foot_syls3 -0.365  0.624
foot_syls4 -0.344  0.513  0.797
```

Table G9.2 ANOVA of PN `f0_exc` model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	11.89	3.96	3	306.15	4.95	.002	.006	p < .05

Table G9.3 R^2 of PN `f0_exc`.

R2_conditional	R2_marginal
.46	.23

Table G9.4 Predicted values of PN `f0_exc` re `foot_syls` (ST re speaker median).

foot_syls	predicted	conf.low	conf.high	std.error
1	2.35	1.43	3.28	0.47
2	2.93	2.04	3.82	0.45
3	3.24	2.36	4.11	0.45
4	2.97	2.06	3.88	0.47

Table G9.5 Pairwise comparison of effects of levels of `foot_syls` (b1) on PN `f0_exc` (ST re speaker median).

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	0.58	0.078	1.081	0.25	2.27	331.44	0.024
foot_syls1	foot_syls3	0.88	0.4	1.362	0.24	3.60	320.76	< .001
foot_syls1	foot_syls4	0.61	0.024	1.202	0.30	2.05	256.61	0.041
foot_syls2	foot_syls3	0.3	-0.125	0.728	0.22	1.39	274.89	0.166
foot_syls2	foot_syls4	0.03	-0.511	0.577	0.28	0.12	216.89	0.904
foot_syls3	foot_syls4	-0.27	-0.624	0.087	0.18	-1.48	403.29	0.139

G10. Log slope between L and H targets (log_lh_slope) in prenuclear L*H

Table G10.1 Summary of PN log_lh_slope model.

```
Formula:
log_lh_slope ~ foot_syls + (1 | ana_syls) + (1 | wrd_end_syl) + (1 | speaker) +
(1 | gender) + (1 | pn_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: lh_slope_equation
   Data: pn_lh_slope.trimmedControl:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 487.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2743 -0.6077  0.0234  0.6262  3.2717

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept) 0.021199 0.14560
 pn_str_syl  (Intercept) 0.054258 0.23293
 ana_syls    (Intercept) 0.003308 0.05751
 wrd_end_syl (Intercept) 0.007067 0.08406
 gender      (Intercept) 0.032796 0.18110
 Residual                0.142094 0.37695
Number of obs: 488, groups:  speaker, 11; pn_str_syl, 8; ana_syls, 4;
wrd_end_syl, 3; gender, 2

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)   3.0281     0.1923   3.5949  15.747 0.000191 ***
foot_syls2   -0.1802     0.1050 191.3782  -1.716 0.087761 .
foot_syls3   -0.1661     0.1006 177.0656  -1.651 0.100445
foot_syls4   -0.1807     0.1222 130.3603  -1.478 0.141837
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table G10.2 ANOVA of PN log_lh_slope model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	0.5	0.17	3	184.88	1.19	.317	.418	

Table G10.3 R^2 of PN log_lh_slope model.

R2_conditional	R2_marginal
.46	.23

Table G10.4 Predicted values of PN log_lh_slope re foot_syls (log[syls/sec]).

foot_syls	predicted	conf.low	conf.high	std.error
1	3.03	2.65	3.41	0.19
2	2.85	2.48	3.21	0.19
3	2.86	2.51	3.22	0.18
4	2.85	2.48	3.21	0.19

Table G10.5 Pairwise comparison of effects of each level of `foot_syls` *(b1) on PN log_*`lh_slope` *(log[syls/sec]).*

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	-0.18	-0.387	0.027	0.11	-1.72	191.38	0.088
foot_syls1	foot_syls3	-0.17	-0.365	0.032	0.10	-1.65	177.07	0.1
foot_syls1	foot_syls4	-0.18	-0.423	0.061	0.12	-1.48	130.36	0.142
foot_syls2	foot_syls3	0.01	-0.162	0.19	0.09	0.16	179.65	0.875
foot_syls2	foot_syls4	0	-0.224	0.223	0.11	0.00	127.43	0.997
foot_syls3	foot_syls4	-0.01	-0.161	0.132	0.07	-0.19	290.30	0.846

Appendix H. LMEMs of Nuclear Phonetic Parameters in A- and H-Corpora

H1. Effect sizes (partial omega-squared) of fixed effects on nuclear pitch accents

Table H1.1 Effect size (ω_p^2) of each lexical and metrical fixed effect in nuclear PA tonal target LME models.

	l_t		**l_f0**		**h_t**		**h_f0**	
Parameter	ω_p^2	95% CI	ω_p^2	95% CI	ω_p^2	95% CI	ω_p^2	95% CI
foot_syls	0	[0, 0]	0.04	[.01, .07]	0.7	[.6, .73]	0.14	[.09, .18]
pre_syls	0.01	[0, .03]	0.1	[0, .26]	0.09	[.06, .13]	0.02	[0, .05]
fin_phon	0	[0, 0]	0	[0, 0]	0.07	[.04, .11]	0	[0, .02]
nuc_new_word	0	[0, .01]	0.03	[0, .18]	0.01	[0, .03]	0.01	[0, .02]
gender	0.84	[.52, .92]	-0.05	[0, 0]	0.62	[.13, .82]	0.14	[.09, .18]

*Table H1.2 Effect size (ω_p^2) of `foot_syls` in nuclear L*H f_0 excursion and slope LME models.*

	f0_exc		**log_lh_slope**	
Parameter	ω_p^2	95% CI	ω_p^2	95% CI
foot_syls	.81	[.53, .9]	.9	[.75, .94]

H2. Temporal alignment of L target (`l_t`) in nuclear L*H

Table H2.1 Summary of nuclear `l_t` model.

```
Formula:
l_t ~ foot_syls + pre_syls + fin_phon + nuc_new_word + gender + (1 | speaker) +
(1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: l_t_equation
   Data: nuc_l_t_data.trimmed
Control: lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl =
list(method = "nlminb",      starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 7076.9

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.07196 -0.66993  0.01925  0.63383  2.95894

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept) 132.8    11.52
 nuc_str_syl (Intercept) 533.0    23.09
 Residual                517.6    22.75
Number of obs: 780, groups:  speaker, 11; nuc_str_syl, 3

Fixed effects:
                 Estimate Std. Error       df t value Pr(>|t|)
(Intercept)       92.2265    15.1578   3.2340   6.084  0.00715 **
foot_syls2        -0.4425     4.2340 759.4160  -0.105  0.91678
foot_syls3         0.9313     3.3107 759.2965   0.281  0.77856
foot_syls4        -1.6680     5.3078 760.9706  -0.314  0.75342
pre_syls1         -4.8925     3.3465 759.9802  -1.462  0.14416
pre_syls2         -1.4337     5.3172 761.2515  -0.270  0.78751
pre_syls3        -11.6185     5.4877 761.6638  -2.117  0.03457 *
fin_phonL%         3.3729     5.1187 767.0586   0.659  0.51013
nuc_new_wordTRUE  -3.3704     3.2429 760.6453  -1.039  0.29897
genderM          -54.9250     7.1692   8.9565  -7.661 3.21e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table H2.2 ANOVA of nuclear `l_t` *model.*

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	128.29	42.76	3	760.16	0.08	.969	.969	
pre_syls	6264	2088	3	757.92	4.03	.007	.014	p<.05
fin_phon	224.72	224.72	1	767.06	0.43	.510	.569	
nuc_new_word	559.08	559.08	1	760.64	1.08	.299	.372	
gender	30378	30378	1	8.96	58.7	<.001	<.001	p<.05

Table H2.3 R^2 of nuclear `l_t` *model.*

R2_conditional	R2_marginal
.71	.37

Table H2.4 Predicted values of nuclear `l_t` *re* `foot_syls` *(ms).*

foot_syls	predicted	conf.low	conf.high	std.error
1	93.6	63.01	124.19	15.61
2	95.69	64.2	127.18	16.07
3	96.52	65.94	127.11	15.6
4	94.23	63.99	124.47	15.43

Table H2.5 Predicted values of nuclear `l_t` *re* `pre_syls` *(ms).*

pre_syls	predicted	conf.low	conf.high	std.error
0	93.6	63.01	124.19	15.61
1	85.8	55.19	116.4	15.61
2	87.55	56.98	118.12	15.6
3	80.12	49.66	110.59	15.54

Table H2.6 Predicted values of nuclear `l_t` *re* `fin_phon` *(ms).*

fin_phon	predicted	conf.low	conf.high	std.error
%	93.6	63.01	124.19	15.61
L%	97.32	65.05	129.59	16.47

Table H2.7 Predicted values of nuclear `l_t` *re* `nuc_new_wrd` *(ms).*

nuc_new_word	predicted	conf.low	conf.high	std.error
FALSE	93.6	63.01	124.19	15.61
TRUE	88.62	59.31	117.92	14.95

Table H2.8 Predicted values of nuclear `l_t` *re* `gender` *(ms).*

gender	predicted	conf.low	conf.high	std.error
F	93.6	63.01	124.19	15.61
M	39.46	8.58	70.35	15.76

Table H2.9 Pairwise comparison of effects of levels of fixed effects (b1) on nuclear `l_t` *(ms).*

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	2	-7	11	4.4	0.47	767.35	.635
foot_syls1	foot_syls3	3	-4	10	3.48	0.84	767.36	.402
foot_syls1	foot_syls4	1	-10	12	5.55	0.11	768.93	.910
foot_syls2	foot_syls3	1	-8	10	4.68	0.18	767.81	.859
foot_syls2	foot_syls4	-1	-13	10	5.72	-0.26	768.8	.799
foot_syls3	foot_syls4	-2	-14	9	5.76	-0.4	769.05	.691
pre_syls0	pre_syls1	-8	-15	-1	3.5	-2.23	767.76	.026
pre_syls0	pre_syls2	-6	-17	5	5.56	-1.09	769.22	.277
pre_syls0	pre_syls3	-14	-25	-2	5.75	-2.34	769.54	.019
pre_syls1	pre_syls2	2	-7	10	4.31	0.41	768.91	.685
pre_syls1	pre_syls3	-6	-15	3	4.56	-1.24	767.98	.214
pre_syls2	pre_syls3	-7	-14	-1	3.38	-2.2	751.77	.028
intercept	fin_phonL%	4	-7	14	5.41	0.69	775.05	.492
intercept	nuc_new_wordT	-5	-12	2	3.4	-1.47	768.4	.143
intercept	genderM	-54	-71	-37	7.61	-7.12	8.97	<.001

H3. f_0 of L target (`l_f0`) in nuclear L*H

Table H3.1 Summary of nuclear `l_f0` model.

```
Formula:
l_f0 ~ foot_syls + pre_syls + fin_phon + nuc_new_word + gender + (1 | speaker) +
(1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: l_f0_equation
   Data: nuc_l_f0.trimmed
Control:
lmerControl(optimizer = optimizer, calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2055.7

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.39360 -0.67453 -0.01464  0.63447  2.87877

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept) 0.38301  0.6189
 nuc_str_syl (Intercept) 0.01128  0.1062
 Residual                0.79008  0.8889
Number of obs: 770, groups:  speaker, 11; nuc_str_syl, 3

Fixed effects:
                  Estimate Std. Error        df t value Pr(>|t|)
(Intercept)       -3.20444    0.32159  17.41030  -9.964 1.30e-08 ***
foot_syls2         0.50679    0.16479 749.67597   3.075  0.00218 **
foot_syls3         0.46211    0.12919 749.70143   3.577  0.00037 ***
foot_syls4         0.80844    0.20004 279.87196   4.041 6.87e-05 ***
pre_syls1          0.12279    0.13027 749.90887   0.943  0.34622
pre_syls2         -0.27655    0.19945 260.78119  -1.387  0.16676
pre_syls3         -0.19659    0.20298 138.42171  -0.969  0.33447
fin_phonL%        -0.08214    0.20027 755.56247  -0.410  0.68180
nuc_new_wordTRUE   0.18113    0.11522  44.15525   1.572  0.12310
genderM           -0.25456    0.38029   8.98291  -0.669  0.52008
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1ANOVA of nuclear
l_f0 model.
```

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	19.14	6.38	3	543.12	8.07	<.001	<.001	p<.05
pre_syls	6.24	2.08	3	40.99	2.63	.063	.096	
fin_phon	0.13	0.13	1	755.56	0.17	.682	.723	
nuc_new_word	1.95	1.95	1	44.16	2.47	.123	.172	
gender	0.35	0.35	1	8.98	0.45	.520	.569	

Table H3.2 R^2 of nuclear `l_f0` model.

R2_conditional	R2_marginal
.36	.05

Table H3.3 Predicted values of nuclear `l_f0` *re* `foot_syls` *(ST re speaker median).*

foot_syls	predicted	conf.low	conf.high	std.error
1	-3.2	-3.83	-2.57	0.32
2	-2.7	-3.39	-2.01	0.35
3	-2.74	-3.37	-2.11	0.32
4	-2.4	-3.03	-1.77	0.32

Table H3.4 Predicted values of nuclear `l_f0` *re* `pre_syls` *(ST re speaker median).*

pre_syls	predicted	conf.low	conf.high	std.error
0	-3.2	-3.83	-2.57	0.32
1	-3.1	-3.71	-2.45	0.32
2	-3.5	-4.13	-2.83	0.33
3	-3.4	-4.04	-2.76	0.33

Table H3.5 Predicted values of nuclear `l_f0` *re* `fin_phon` *(ST re speaker median).*

fin_phon	predicted	conf.low	conf.high	std.error
%	-3.2	-3.83	-2.57	0.32
L%	-3.29	-4.02	-2.55	0.38

Table H3.6 Predicted values of nuclear `l_f0` *re nuc_new_wrd (ST re speaker median).*

nuc_new_word	predicted	conf.low	conf.high	std.error
FALSE	-3.2	-3.83	-2.57	0.32
TRUE	-3.02	-3.59	-2.46	0.29

Table H3.7 Predicted values of nuclear `l_f0` *re* `gender` *(ST re speaker median).*

gender	predicted	conf.low	conf.high	std.error
F	-3.2	-3.83	-2.57	0.32
M	-3.46	-4.13	-2.79	0.34

Table H3.8 Pairwise comparison of levels of fixed effects (b1) on nuclear l_f0 (ST).

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	0.5	0.2	0.8	0.16	3.08	749.68	.002
foot_syls1	foot_syls3	0.5	0.2	0.7	0.13	3.58	749.7	<.001
foot_syls1	foot_syls4	0.8	0.4	1.2	0.2	4.04	279.87	<.001
foot_syls2	foot_syls3	0	-0.4	0.3	0.17	-0.26	749.95	.797
foot_syls2	foot_syls4	0.3	-0.1	0.7	0.21	1.46	303.32	.144
foot_syls3	foot_syls4	0.3	-0.1	0.8	0.21	1.67	305.73	.095
pre_syls0	pre_syls1	0.1	-0.1	0.4	0.13	0.94	749.91	.346
pre_syls0	pre_syls2	-0.3	-0.7	0.1	0.2	-1.39	260.78	.167
pre_syls0	pre_syls3	-0.2	-0.6	0.2	0.2	-0.97	138.42	.334
pre_syls1	pre_syls2	-0.4	-0.7	-0.1	0.15	-2.65	108.97	.009
pre_syls1	pre_syls3	-0.3	-0.6	0	0.16	-2.06	53.18	.045
pre_syls2	pre_syls3	0.1	-0.1	0.3	0.1	0.76	15.53	.456
intercept	fin_phonL%	-0.1	-0.5	0.3	0.2	-0.41	755.56	.682
intercept	nuc_new_wordT	0.2	-0.1	0.4	0.12	1.57	44.16	.123
intercept	genderM	-0.3	-1.1	0.6	0.38	-0.67	8.98	.520

H4. Temporal alignment of H target (h_t) in nuclear L*H

Table H4.1 Summary of nuclear h_t model.

```
Formula:
h_t ~ foot_syls + pre_syls + fin_phon + nuc_new_word + gender + (1 | speaker) +
(1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_t_equation
   Data: nuc_data %>% filter(abs(scale(resid(nuc_h_t_mdl))) <= 3)
Control:
lmerControl(optimizer = optimizer, calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 7363.6

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.98998 -0.67401 -0.05753  0.59246  2.88175

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept) 591.7    24.32
 nuc_str_syl (Intercept) 689.1    26.25
 Residual                688.5    26.24
Number of obs: 786, groups:  speaker, 11; nuc_str_syl, 3

Fixed effects:
                 Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       292.368     19.173   4.746  15.249 3.26e-05 ***
foot_syls2         21.226      4.808 765.110   4.415 1.16e-05 ***
foot_syls3         95.949      3.810 765.105  25.185  < 2e-16 ***
foot_syls4        218.552      6.062 766.738  36.053  < 2e-16 ***
pre_syls1         -33.472      3.830 765.247  -8.740  < 2e-16 ***
pre_syls2         -37.785      6.072 766.966  -6.223 8.02e-10 ***
pre_syls3         -44.687      6.286 767.187  -7.109 2.69e-12 ***
fin_phonL%        -45.810      5.919 768.383  -7.740 3.14e-14 ***
nuc_new_wordTRUE  -12.681      3.712 766.455  -3.416 0.000668 ***
genderM           -64.322     14.850   9.006  -4.331 0.001898 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table H4.2 ANOVA of nuclear h_t model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	1.2e+06	4.0e+05	3	765.88	586.86	<.001	<.001	p<.05
pre_syls	56659	18886	3	763.69	27.43	<.001	<.001	p<.05
fin_phon	41246	41246	1	768.38	59.9	<.001	<.001	p<.05
nuc_new_word	8036	8036	1	766.46	11.67	<.001	.002	p<.05
gender	12918	12918	1	9.01	18.76	.002	.004	p<.05

Table H4.3 R^2 of nuclear h_t model.

R2_conditional	R2_marginal
.90	.71

Table H4.4 Predicted values of nuclear h_t re foot_syls (ms).

foot_syls	predicted	conf.low	conf.high	std.error
1	292.37	254.79	329.95	19.17
2	313.59	275.14	352.05	19.62
3	388.32	350.74	425.9	19.17
4	510.92	473.68	548.16	19

Table H4.5 Predicted values of nuclear h_t re pre_syls (ms).

pre_syls	predicted	conf.low	conf.high	std.error
0	292.37	254.79	329.95	19.17
1	258.9	221.3	296.49	19.18
2	254.58	217.03	292.14	19.16
3	247.68	210.22	285.15	19.11

Table H4.6 Predicted values of nuclear h_t re fin_phon (ms).

fin_phon	predicted	conf.low	conf.high	std.error
%	292.37	254.79	329.95	19.17
L%	246.56	207.33	285.79	20.01

Table H4.7 Predicted values of nuclear h_t re nuc_new_wrd (ms).

nuc_new_word	predicted	conf.low	conf.high	std.error
FALSE	292.37	254.79	329.95	19.17
TRUE	279.69	243.35	316.02	18.54

Table H4.8 Predicted values of nuclear h_t re gender (ms).

gender	predicted	conf.low	conf.high	std.error
F	292.37	254.79	329.95	19.17
M	228.05	189.49	266.6	19.67

Table H4.9 Pairwise comparison of effects of levels of fixed effects (b1) on nuclear h_t *(ms).*

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	21	12	31	4.81	4.41	765.11	<.001
foot_syls1	foot_syls3	96	88	103	3.81	25.18	765.1	<.001
foot_syls1	foot_syls4	219	207	230	6.06	36.05	766.74	<.001
foot_syls2	foot_syls3	75	65	85	5.11	14.63	765.25	<.001
foot_syls2	foot_syls4	197	185	210	6.24	31.6	766.66	<.001
foot_syls3	foot_syls4	123	110	135	6.29	19.5	766.72	<.001
pre_syls0	pre_syls1	-34	-41	-26	3.83	-8.74	765.25	<.001
pre_syls0	pre_syls2	-38	-50	-26	6.07	-6.22	766.97	<.001
pre_syls0	pre_syls3	-45	-57	-32	6.29	-7.11	767.19	<.001
pre_syls1	pre_syls2	-4	-14	5	4.71	-0.92	766.95	.360
pre_syls1	pre_syls3	-11	-21	-1	4.98	-2.25	766	.024
pre_syls2	pre_syls3	-7	-14	0	3.7	-1.86	751.16	.063
intercept	fin_phonL%	-46	-57	-34	5.92	-7.74	768.38	<.001
intercept	nuc_new_wordT	-13	-20	-5	3.71	-3.42	766.46	<.001
intercept	genderM	-64	-98	-31	14.85	-4.33	9.01	.002

H5. f_0 of H target (h_f0) in nuclear L*H

Table H5.1 Summary of nuclear h_f0 model.

```
Formula:
h_f0 ~ foot_syls + pre_syls + fin_phon + nuc_new_word + gender + (1 | speaker) +
(1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_f0_equation
   Data: nuc_data %>% filter(abs(scale(resid(nuc_h_f0_mdl))) <= 4)
Control:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 2915.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6221 -0.6663 -0.1041  0.6423  3.1530

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept) 0.9098   0.9538
 nuc_str_syl (Intercept) 0.4128   0.6425
 Residual                2.2501   1.5000
Number of obs: 787, groups:  speaker, 11; nuc_str_syl, 3

Fixed effects:
                  Estimate Std. Error        df t value Pr(>|t|)
(Intercept)        0.72839    0.64398  11.81951   1.131 0.280457
foot_syls2         0.84635    0.27474 766.16016   3.081 0.002140 **
foot_syls3         2.29420    0.21723 766.17035  10.561  < 2e-16 ***
foot_syls4         2.00642    0.34513 758.92204   5.813 9.01e-09 ***
pre_syls1          0.82007    0.21852 766.43100   3.753 0.000188 ***
pre_syls2          0.74566    0.34543 752.93859   2.159 0.031191 *
pre_syls3          1.14703    0.35690 725.27573   3.214 0.001367 **
fin_phonL%        -0.64126    0.33787 772.26023  -1.898 0.058077 .
nuc_new_wordTRUE   0.44109    0.20998 667.70898   2.101 0.036047 *
genderM            0.09731    0.58757   8.98828   0.166 0.872129
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Table H5.2 ANOVA of nuclear h_f0 model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	278.32	92.77	3	766.63	41.23	<.001	<.001	p<.05
pre_syls	40.36	13.45	3	607.56	5.98	<.001	.001	p<.05
fin_phon	8.11	8.11	1	772.26	3.6	.058	.092	
nuc_new_word	9.93	9.93	1	667.71	4.41	.036	.060	
gender	0.06	0.06	1	8.99	0.03	.872	.898	

Table H5.3 R^2 of nuclear h_f0 model.

R2_conditional	R2_marginal
.44	.11

Table H5.4 Predicted values of nuclear `h_f0` *re* `foot_syls` *(ST re speaker median).*

foot_syls	predicted	conf.low	conf.high	std.error
1	0.73	-0.53	1.99	0.64
2	1.57	0.23	2.92	0.69
3	3.02	1.76	4.28	0.64
4	2.73	1.5	3.97	0.63

Table H5.5 Predicted values of nuclear `h_f0` *re* `pre_syls` *(ST re speaker median).*

pre_syls	predicted	conf.low	conf.high	std.error
0	0.73	-0.53	1.99	0.64
1	1.55	0.28	2.81	0.64
2	1.47	0.21	2.74	0.64
3	1.88	0.62	3.13	0.64

Table H5.6 Predicted values of nuclear `h_f0` *re* `fin_phon` *(ST re speaker median).*

fin_phon	predicted	conf.low	conf.high	std.error
%	0.73	-0.53	1.99	0.64
L%	0.09	-1.33	1.5	0.72

Table H5.7 Predicted values of nuclear `h_f0` *re nuc_new_wrd (ST re speaker median).*

nuc_new_word	predicted	conf.low	conf.high	std.error
FALSE	0.73	-0.53	1.99	0.64
TRUE	1.17	0.03	2.31	0.58

Table H5.8 Predicted values of nuclear `h_f0` *re* `gender` *(ST re speaker median).*

gender	predicted	conf.low	conf.high	std.error
F	0.73	-0.53	1.99	0.64
M	0.83	-0.48	2.13	0.67

Table H5.9 Pairwise comparison of levels of fixed effects (b1) on nuclear h_f0 ST re speaker median

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	0.8	0.3	1.4	0.27	3.08	766.16	.002
foot_syls1	foot_syls3	2.3	1.9	2.7	0.22	10.56	766.17	<.001
foot_syls1	foot_syls4	2	1.3	2.7	0.35	5.81	758.92	<.001
foot_syls2	foot_syls3	1.4	0.9	2	0.29	4.96	766.46	<.001
foot_syls2	foot_syls4	1.2	0.5	1.9	0.36	3.26	760.43	.001
foot_syls3	foot_syls4	-0.3	-1	0.4	0.36	-0.8	760.9	.422
pre_syls0	pre_syls1	0.8	0.4	1.2	0.22	3.75	766.43	<.001
pre_syls0	pre_syls2	0.8	0.1	1.4	0.35	2.16	752.94	.031
pre_syls0	pre_syls3	1.1	0.4	1.8	0.36	3.21	725.28	.001
pre_syls1	pre_syls2	-0.1	-0.6	0.4	0.27	-0.28	715.34	.781
pre_syls1	pre_syls3	0.3	-0.2	0.9	0.28	1.16	652.07	.246
pre_syls2	pre_syls3	0.4	0	0.8	0.21	1.94	392.25	.053
intercept	fin_phonL%	-0.6	-1.3	0	0.34	-1.9	772.26	.058
intercept	nuc_new_wordT	0.4	0	0.9	0.21	2.1	667.71	.036
intercept	genderM	0.1	-1.2	1.4	0.59	0.17	8.99	.872

H6. Temporal alignment of L target (e_t) in nuclear L*H

Table H6.1 Summary of nuclear e_t model.

```
Formula:
e_t ~ foot_syls + pre_syls + fin_phon + nuc_new_word + gender + (1 | speaker) +
(1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: e_t_equation
   Data: nuc_data %>% filter(abs(scale(resid(nuc_e_t_mdl))) <= 4)
Control:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 7539.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1237 -0.6307 -0.0471  0.5892  3.1918

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept)  699.1   26.44
 nuc_str_syl (Intercept) 1039.9   32.25
 Residual                 853.4   29.21
Number of obs: 787, groups:  speaker, 11; nuc_str_syl, 3

Fixed effects:
                 Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       330.004     22.625   4.134  14.586 0.000104 ***
foot_syls2         11.413      5.355 766.108   2.131 0.033373 *
foot_syls3        117.127      4.247 766.110  27.582  < 2e-16 ***
foot_syls4        248.909      6.747 767.541  36.894  < 2e-16 ***
pre_syls1         -31.242      4.273 766.237  -7.312 6.65e-13 ***
pre_syls2         -31.400      6.766 767.775  -4.641 4.08e-06 ***
pre_syls3         -36.048      7.006 768.183  -5.146 3.39e-07 ***
fin_phonL%         -6.716      6.589 769.521  -1.019 0.308423
nuc_new_wordTRUE   -5.961      4.127 767.913  -1.445 0.148995
genderM           -47.861     16.147   9.000  -2.964 0.015857 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table H6.2 ANOVA of nuclear e_t model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	1.7×10^{06}	5.7×10^{05}	3	766.77	670.13	<.001	<.001	p < .05
pre_syls	46809	15603	3	766.07	18.28	<.001	<.001	p < .05
fin_phon	886.52	886.52	1	769.52	1.04	.308	.372	
nuc_new_word	1781	1781	1	767.91	2.09	.149	.201	
gender	7497	7497	1	9	8.79	.016	.029	p < .05

Table H6.3 R^2 of nuclear e_t model.

R2_conditional	R2_marginal
.90	.68

Table H6.4 Predicted values of nuclear `e_t` *re* `foot_syls` *(ms).*

foot_syls	predicted	conf.low	conf.high	std.error
1	330	285.66	374.35	22.63
2	341.42	296.15	386.69	23.1
3	447.13	402.8	491.47	22.62
4	578.91	534.92	622.91	22.45

Table H6.5 Predicted values of nuclear `e_t` *re* `pre_syls` *(ms).*

pre_syls	predicted	conf.low	conf.high	std.error
0	330	285.66	374.35	22.63
1	298.76	254.4	343.12	22.63
2	298.6	254.29	342.92	22.61
3	293.96	249.74	338.17	22.56

Table H6.6 Predicted values of nuclear `e_t` *re* `fin_phon` *(ms).*

fin_phon	predicted	conf.low	conf.high	std.error
%	330	285.66	374.35	22.63
L%	323.29	277.21	369.37	23.51

Table H6.7 Predicted values of nuclear `e_t` *re nuc_new_wrd (ms).*

nuc_new_word	predicted	conf.low	conf.high	std.error
FALSE	330	285.66	374.35	22.63
TRUE	324.04	281	367.09	21.96

Table H6.8 Predicted values of nuclear `e_t` *re* `gender` *(ms).*

gender	predicted	conf.low	conf.high	std.error
F	330	285.66	374.35	22.63
M	282.14	236.82	327.47	23.12

Table H6.9 Pairwise comparison of effects of levels of fixed effects (b1) on nuclear `e_t` *(ms).*

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	11	1	22	5.35	2.13	766.11	.033
foot_syls1	foot_syls3	117	109	126	4.25	27.58	766.11	<.001
foot_syls1	foot_syls4	249	236	262	6.75	36.89	767.54	<.001
foot_syls2	foot_syls3	106	94	117	5.71	18.51	766.25	<.001
foot_syls2	foot_syls4	238	224	251	6.95	34.18	767.46	<.001
foot_syls3	foot_syls4	132	118	146	7.02	18.78	767.51	<.001
pre_syls0	pre_syls1	-31	-40	-23	4.27	-7.31	766.24	<.001
pre_syls0	pre_syls2	-31	-45	-18	6.77	-4.64	767.78	<.001
pre_syls0	pre_syls3	-36	-50	-22	7.01	-5.15	768.18	<.001
pre_syls1	pre_syls2	0	-10	10	5.24	-0.03	768.01	.976
pre_syls1	pre_syls3	-5	-16	6	5.54	-0.87	767.74	.386
pre_syls2	pre_syls3	-5	-13	3	4.11	-1.13	758.91	.259
intercept	fin_phonL%	-7	-20	6	6.59	-1.02	769.52	.308
intercept	nuc_new_wordT	-6	-14	2	4.13	-1.45	767.91	.149
intercept	genderM	-48	-84	-11	16.15	-2.96	9	.016

H7. f_0 of H target (e_f0) in nuclear L*H

Table H7.1 Summary of nuclear e_f0 model.

```
Formula:
e_f0 ~ foot_syls + pre_syls + fin_phon + nuc_new_word + gender + (1 | speaker) +
(1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: e_f0_equation
   Data: nuc_data %>% filter(abs(scale(resid(nuc_e_f0_mdl))) <= 3.5)
Control:
lmerControl(optimizer = optimizer, calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

REML criterion at convergence: 3055.8

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.08780 -0.66934 -0.03013  0.59134  3.10446

Random effects:
 Groups      Name        Variance Std.Dev.
 speaker     (Intercept) 1.8468   1.359
 nuc_str_syl (Intercept) 0.2275   0.477
 Residual                2.7100   1.646
Number of obs: 785, groups:  speaker, 11; nuc_str_syl, 3

Fixed effects:
                 Estimate Std. Error       df t value Pr(>|t|)
(Intercept)        0.2233     0.7283  17.5324   0.307   0.7627
foot_syls2         1.3654     0.3015 763.9825   4.528 6.90e-06 ***
foot_syls3         1.6618     0.2384 763.9879   6.970 6.84e-12 ***
foot_syls4         1.6544     0.3785 705.2020   4.371 1.42e-05 ***
pre_syls1          0.4996     0.2398 764.1468   2.083   0.0376 *
pre_syls2          0.2307     0.3772 679.2374   0.612   0.5411
pre_syls3          0.5730     0.3887 579.6235   1.474   0.1410
fin_phonL%        -5.6016     0.3712 768.3282 -15.092  < 2e-16 ***
nuc_new_wordTRUE   0.2754     0.2278 420.4949   1.209   0.2273
genderM           -0.7209     0.8314   8.9850  -0.867   0.4084
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Table H7.2 ANOVA of nuclear e_f0 model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	159.34	53.12	3	754.12	19.6	<.001	<.001	$p < 0.05$
pre_syls	18.81	6.27	3	330.09	2.31	.076	.111	
fin_phon	617.26	617.26	1	768.33	227.77	<.001	<.001	$p < 0.05$
nuc_new_word	3.96	3.96	1	420.5	1.46	.227	.295	
gender	2.04	2.04	1	8.98	0.75	.408	.477	

Table H7.3 R^2 of nuclear e_f0 model.

R2_conditional	R2_marginal
.53	.22

Table H7.4 Predicted values of nuclear `e_f0` *re* `foot_syls` *(ST re speaker median).*

foot_syls	predicted	conf.low	conf.high	std.error
1	0.22	-1.2	1.65	0.73
2	1.59	0.07	3.11	0.77
3	1.89	0.46	3.31	0.73
4	1.88	0.48	3.28	0.72

Table H7.5 Predicted values of nuclear `e_f0` *re* `pre_syls` *(ST re speaker median).*

pre_syls	predicted	conf.low	conf.high	std.error
0	0.22	-1.2	1.65	0.73
1	0.72	-0.71	2.15	0.73
2	0.45	-0.98	1.88	0.73
3	0.8	-0.63	2.22	0.73

Table H7.6 Predicted values of nuclear `e_f0` *re* `fin_phon` *(ST re speaker median).*

fin_phon	predicted	conf.low	conf.high	std.error
%	0.22	-1.2	1.65	0.73
L%	-5.38	-6.97	-3.79	0.81

Table H7.7 Predicted values of nuclear `e_f0` *re nuc_new_wrd (ST re speaker median).*

nuc_new_word	predicted	conf.low	conf.high	std.error
FALSE	0.22	-1.2	1.65	0.73
TRUE	0.5	-0.81	1.8	0.67

Table H7.8 Predicted values of nuclear `e_f0` *re* `gender` *(ST re speaker median).*

gender	predicted	conf.low	conf.high	std.error
F	0.22	-1.2	1.65	0.73
M	-0.5	-2	1.01	0.77

Table H7.9 Pairwise comparison of levels of fixed effects (b1) on nuclear e_f0 ST re speaker median

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	1.4	0.8	2	0.3	4.53	763.98	<.001
foot_syls1	foot_syls3	1.7	1.2	2.1	0.24	6.97	763.99	<.001
foot_syls1	foot_syls4	1.6	0.9	2.4	0.38	4.37	705.2	<.001
foot_syls2	foot_syls3	0.3	-0.3	0.9	0.32	0.93	764.17	.355
foot_syls2	foot_syls4	0.3	-0.5	1.1	0.39	0.74	712.92	.459
foot_syls3	foot_syls4	0	-0.8	0.8	0.39	-0.02	714.71	.985
pre_syls0	pre_syls1	0.5	0	1	0.24	2.08	764.15	.038
pre_syls0	pre_syls2	0.2	-0.5	1	0.38	0.61	679.24	.541
pre_syls0	pre_syls3	0.6	-0.2	1.3	0.39	1.47	579.62	.141
pre_syls1	pre_syls2	-0.3	-0.8	0.3	0.29	-0.92	546.41	.356
pre_syls1	pre_syls3	0.1	-0.5	0.7	0.31	0.24	402.46	.810
pre_syls2	pre_syls3	0.3	-0.1	0.8	0.22	1.55	142.39	.123
intercept	fin_phonL%	-5.6	-6.3	-4.9	0.37	-15.09	768.33	<.001
intercept	nuc_new_wordT	0.3	-0.2	0.7	0.23	1.21	420.49	.227
intercept	genderM	-0.7	-2.6	1.2	0.83	-0.87	8.99	.408

H8. f_0 excursion (f0_exc) in nuclear L*H pitch accent

Table H8.1 Summary of nuclear f0_exc model.

```
Formula:
f0_exc ~ foot_syls + (1 + foot_syls | speaker) + (1 | nuc_str_syl) + (1 |
pre_syls) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: f0_exc_equation
   Data: nuc_data %>% filter(abs(scale(resid(nuc_f0_exc_mdl))) <= 2.5)
Control:
lmerControl(optCtrl = list(maxit = 1e+09, maxfun = 1e+09, xtol_abs = 1e-09,
    ftol_abs = 1e-09))

REML criterion at convergence: 2400.3

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.64645 -0.61665 -0.09058  0.62767  2.92312

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 speaker     (Intercept) 0.8088   0.8993
             foot_syls2  0.6290   0.7931   -0.32
             foot_syls3  0.3285   0.5732    0.26  0.54
             foot_syls4  0.7987   0.8937    0.26  0.75  0.89
 pre_syls    (Intercept) 0.1522   0.3901
 nuc_str_syl (Intercept) 0.3974   0.6304
 fin_phon    (Intercept) 0.2901   0.5386
 Residual                1.1757   1.0843
Number of obs: 770, groups:
speaker, 11; pre_syls, 4; nuc_str_syl, 3; fin_phon, 2

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   4.6624     0.6481  4.6392   7.194  0.00110 **
foot_syls2    0.4520     0.3046 15.9083   1.484  0.15734
foot_syls3    1.9307     0.2348  9.9915   8.223 9.29e-06 ***
foot_syls4    1.0535     0.3345 11.1244   3.149  0.00913 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Table H8.2 ANOVA of nuclear f0_exc model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	91.68	30.56	3	13.72	25.99	<.001	<.001	$p < .05$

Table H8.3 R^2 of nuclear f0_exc model.

R2_conditional	R2_marginal
.65	.08

Table H8.4 Predicted values of nuclear f0_exc re foot_syls (ST).

foot_syls	predicted	conf.low	conf.high	std.error
1	4.66	3.39	5.93	0.65
2	5.11	3.83	6.4	0.65
3	6.59	5.24	7.95	0.69
4	5.72	4.27	7.16	0.74

Table H8.5 Pairwise comparison of levels of fixed effects (b1) on nuclear `f0_exc` *ST*

Intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	0.4	-0.2	1.1	0.3	1.48	15.91	.157
foot_syls1	foot_syls3	1.9	1.4	2.5	0.23	8.22	9.99	<.001
foot_syls1	foot_syls4	1	0.3	1.8	0.33	3.15	11.12	.009
foot_syls2	foot_syls3	1.5	0.9	2.1	0.29	5.14	18.94	<.001
foot_syls2	foot_syls4	0.6	0	1.2	0.27	2.21	16.65	.041
foot_syls3	foot_syls4	-0.9	-1.4	-0.3	0.25	-3.48	14.3	.004

H9. Log of f_0(t) slope between L and H targets (`log_lh_slope`) in nuclear L*H

Table H9.1 Summary of nuclear `lh_slope` model.

```
Formula:
log_lh_slope ~ foot_syls + (1 + foot_syls | speaker) + (1 | nuc_str_syl) + (1 |
pre_syls) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: lh_slope_equation
   Data: nuc_data %>% filter(abs(scale(resid(nuc_lh_slope_mdl))) <= 3.5)
Control: ctrl

REML criterion at convergence: 73.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3482 -0.6252  0.0239  0.5796  3.1611

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 speaker     (Intercept) 0.025094 0.15841
             foot_syls2  0.011466 0.10708  0.39
             foot_syls3  0.031261 0.17681  0.70 0.49
             foot_syls4  0.035279 0.18783  0.93 0.22 0.84
 pre_syls    (Intercept) 0.026331 0.16227
 nuc_str_syl (Intercept) 0.006024 0.07761
 fin_phon    (Intercept) 0.005383 0.07337
 Residual                0.056511 0.23772
Number of obs: 785, groups:
speaker, 11; pre_syls, 4; nuc_str_syl, 3; fin_phon, 2

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  3.597522   0.122316  6.797174  29.412 2.03e-08 ***
foot_syls2  -0.007161   0.053636 30.549433  -0.134   0.8947
foot_syls3  -0.124250   0.063771 10.258383  -1.948   0.0792 .
foot_syls4  -0.766358   0.071365 12.471752 -10.739 1.16e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Table H9.2 ANOVA of nuclear `lh_slope` model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	9.37	3.12	3	14.82	55.28	<.001	<.001	$p < 0.05$

Table H9.3 R^2 of nuclear `lh_slope` model.

R2_conditional	R2_marginal
.70	.20

Table H9.4 Predicted values of nuclear `lh_slope` re `foot_syls` (log [ST/s]).

foot_syls	predicted	conf.low	conf.high	std.error
1	3.6	3.36	3.84	0.12
2	3.59	3.34	3.84	0.13
3	3.47	3.19	3.76	0.15
4	2.83	2.53	3.13	0.15

Table H9.5 Pairwise comparison of levels of fixed effects (b1) on nuclear `lh_slope` ST

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	-0.01	-0.12	0.1	0.05	-0.13	30.55	.895
foot_syls1	foot_syls3	-0.12	-0.27	0.02	0.06	-1.95	10.26	.079
foot_syls1	foot_syls4	-0.77	-0.92	-0.61	0.07	-10.74	12.47	<.001
foot_syls2	foot_syls3	-0.12	-0.25	0.02	0.07	-1.78	20.82	.09
foot_syls2	foot_syls4	-0.76	-0.92	-0.6	0.07	-10.26	14.4	<.001
foot_syls3	foot_syls4	-0.64	-0.76	-0.52	0.06	-11.41	15.54	<.001

H10. Additional model: H alignment as a proportion of voicing in nuclear L*H

Table H10.1 Summary of `h_t`_as_prop_of_voicing model.

```
Formula:
h_t_as_prop_of_voicing ~ foot_syls + (1 + foot_syls + fin_phon | speaker) + (1 |
gender) + (1 | nuc_str_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_t_as_prop_of_voicing_equation
   Data: nuc_h_t_as_prop_of_voicing.trimmed
Control: ctrl

REML criterion at convergence: -2085.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1338 -0.6030 -0.0536  0.6156  3.1765

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 speaker     (Intercept) 0.0019744 0.04443
             foot_syls2  0.0003568 0.01889  -0.24
             foot_syls3  0.0011566 0.03401  -0.10  0.84
             foot_syls4  0.0009756 0.03123  -0.73  0.45  0.41
             fin_phonL%  0.0067750 0.08231   0.87 -0.69 -0.49 -0.77
 nuc_str_syl (Intercept) 0.0024451 0.04945
 gender      (Intercept) 0.0026958 0.05192
 Residual                0.0035415 0.05951
Number of obs: 779, groups:  speaker, 11; nuc_str_syl, 3; gender, 2

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  0.801318   0.048368  1.970032  16.567  0.00387 **
foot_syls2   0.025956   0.008712 18.355541   2.980  0.00791 **
foot_syls3  -0.008524   0.013029 11.700914  -0.654  0.52563
foot_syls4   0.011007   0.012875 13.157975   0.855  0.40790
---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table H10.2 ANOVA `h_t`_as_prop_of_voicing model.

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	0.08	0.03	3	13.1	7.18	.004	.009	p<.05

Table H10.3 R^2 of nuclear `h_t`_as_prop_of_voicing model.

R2_conditional	R2_marginal
.67	.02

Table H10.4 Predicted values of nuclear `h_t`*_as_prop_of_voicing model (ratio).*

foot_syls	predicted	conf.low	conf.high	std.error
1	0.8	0.71	0.9	0.05
2	0.83	0.73	0.92	0.05
3	0.79	0.7	0.89	0.05
4	0.81	0.72	0.91	0.05

Table H10.5 Pairwise comparisons of `foot_syls` *levels (b1) on nuclear* `h_t`*_as_prop_of_voicing model.*

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	0.03	0.01	0.04	0.01	2.98	18.36	.008
foot_syls1	foot_syls3	-0.01	-0.04	0.02	0.01	-0.65	11.7	.526
foot_syls1	foot_syls4	0.01	-0.02	0.04	0.01	0.86	13.16	.408
foot_syls2	foot_syls3	-0.03	-0.06	-0.01	0.01	-3.56	12.31	.004
foot_syls2	foot_syls4	-0.01	-0.04	0.01	0.01	-1.24	10.51	.242
foot_syls3	foot_syls4	0.02	-0.01	0.05	0.01	1.32	10.03	.216

Appendix I. Summary of Sentence Mode and Pitch Accent Phonology in M-Corpus

I1. Summary based on non-register-tier analysis

Table I1.1 Distribution by speaker of nuclear pitch accents in M-corpus (raw count, non-register tier analysis)

speaker	H*	L*H	>H*
F5	0	59	0
F6	0	60	0
F12	0	60	0
F15	0	60	0
F16	0	60	0
F17	0	55	0
M4	0	57	0
M5	0	60	0
M8	2	54	5
M9	0	47	0
M10	0	60	0
Total	2	632	5

Table I1.2 Utterance level intonational phonology by mode for tokens accounting for at least 1.25% of all tokens. (Raw count, non-register tier analysis)

#	Phonology	MDC	MWH	MYN	MDQ	total
1	% H* L* L*H %	5	53	0	0	58
2	% >H* L*H %	30	0	38	27	95
3	% L*H %	32	0	37	28	97
4	% H* L*H %	33	0	11	5	49
5	% L*H L*H %	25	0	29	24	78
6	% H* L*!H L*H %	0	21	0	0	21
7	% H* L*H L*H %	0	17	0	0	17
8	% H* !H* L*H %	5	17	0	0	22
9	% L*H L%	1	0	12	12	25
10	% >H* L*H L%	10	0	8	11	29
11	% H* L*H L%	6	0	4	10	20
12	% L*H H%	0	0	4	9	13
13	% >H* L*!H L*H %	0	8	0	0	8
14	% H* L*H %	3	8	1	0	12
15	% L*H L*H L%	1	0	1	8	10

Table I1.3 Utterance level intonational phonology without Register Tier in M-corpus for all tokens(raw count, non-register tier analysis).

Phonology	count	Phonology	count
% L*H %	97	% H* >H* L%	3
% >H* L*H %	95	% L*H L%	3
% L*H L*H %	78	% L*H %	2
% H* L* L*H %	58	% H* H* L%	2
% H* L*H %	49	% L*H HL%	1
% >H* L*H L%	29	% H* L*H HL%	1
% L*H L%	25	%H H* L* L*H %	1
% H* !H* L*H %	22	% >H* L*H L*H %	1
% H* L*!H L*H %	21	% H* L*!H L*H L%	1
% H* L*H L%	20	% L*H L* L*H L%	1
% H* L*H L*H %	17	% >H* L* L*H L%	1
% L*H H%	13	%H >H* L*H %	1
% H* L*H %	12	%H >!H* L*H %	1
% L*H L*H L%	10	%H H* L*H L*H %	1
% >H* L*!H L*H %	8	%L H* L*H %	1
% L*H L*!H L*H L%	6	% H* !H*L L*H %	1
% >H* L* L*H %	5	% H* >!H* L*H %	1
% H* L* L*H L%	5	%H !H* L*H %	1
%H L*H L*H %	5	% H* L*/!H* L*H %	1
%H L*H %	5	% >H* L*H %	1
% H* L*H L%	5	% >H* >H* L%	1
% L*H L*H H%	3	% >H* L%	1
% >H* L*H H%	3	% L*H L*H %	1
%H L* L*H %	3	% L*H L*H L%	1
%H L*!H L*H %	3	%L H* L*H L%	1
%H >H* L*!H L*H %	3	% L*H L*H L*H %	1
% L* L*H %	3	% >H*> L*H %	1
% H* L*H H%	3		

Table I1.4 Distribution by speaker of nuclear pitch accents in M-corpus (adjusted, non-register tier analysis)

speaker	H*	L*H	>H*
F5	0	60	0
F6	0	60	0
F12	0	60	0
F15	0	60	0
F16	0	60	0
F17	0	60	0
M4	0	60	0
M5	0	60	0
M8	2	54	5
M9	0	60	0
M10	0	60	0
Total	2	654	5

Table I1.5 Distribution by mode of nuclear contours in M-corpus (adjusted, non-register tier analysis)

mode	L*H %	L*H L%	L*H H%	L*H HL%	>H* L%	H* L%
MDC	139	17	0	0	2	2
MWH	148	17	0	0	0	0
MYN	124	29	5	0	2	0
MDQ	82	54	17	8	1	0
Total	493	117	22	8	5	2

I2. Summary based on register-tier analysis

Table I2.1 Nuclear contours by mode in M-corpus (raw count, register-tier analysis)

mode	MDC	MWH	MYN	MDQ	Total
H* L%	2	0	0	0	2
>H* L%	2	0	2	1	5
^[L*]H L%	0	0	3	3	6
L*H %	141	144	97	37	419
L*H L%	18	17	18	6	59
L*^[H] %	0	0	5	17	22
L*^[H L%]	0	0	0	6	6
L*^[H] L%	0	0	0	2	2
^[L*H] %	0	0	30	51	81
^[L*H L%]	0	0	6	12	18
^[L*H] L%	0	0	1	18	19

Table I2.2 Distribution by speaker of nuclear pitch accents in M-corpus (raw count, register-tier analysis)

speaker	H*	^[L*]H	>H*	L*H	L*^[H]	^[L*H]
F5	0	0	0	45	13	1
F6	0	0	0	51	5	4
F12	0	0	0	45	1	14
F15	0	0	0	51	4	5
F16	0	0	0	59	1	0
F17	0	0	0	47	1	7
M4	0	0	0	36	5	16
M5	0	0	0	36	0	24
M8	2	6	5	29	0	19
M9	0	0	0	40	0	7
M10	0	0	0	52	0	8
Total	2	6	5	491	30	105

Table I2.3 Nuclear pitch contours (raw data, register-tier analysis)

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

Table I2.4 Distribution by of pitch accents by mode in M-corpus (adjusted, outliers excluded, register-tier analysis)

mode	L*H	L*^[H]	^[L*H]
MDC	162	0	0
MWH	165	0	0
MYN	117	5	38
MDQ	62	25	74
Total	506	30	112

Table I2.5 Distribution by mode of nuclear pitch accents in M-corpus (adjusted, register tier analysis)

mode	H*	>H*	^[L*]H	L*H	L*^[H]	^[L*H]
MDC	2	2	0	162	0	0
MWH	0	0	0	165	0	0
MYN	0	2	3	117	5	38
MDQ	0	1	3	62	25	74
Total	2	5	6	506	30	112

Table I2.6 Distribution by speaker of nuclear pitch accents in M-corpus (adjusted, register-tier analysis)

speaker	H*	>H*	^[L*]H	L*H	L*^[H]	^[L*H]
F5	0	0	0	46	13	1
F6	0	0	0	51	5	4
F12	0	0	0	45	1	14
F15	0	0	0	51	4	5
F16	0	0	0	59	1	0
F17	0	0	0	52	1	7
M4	0	0	0	36	5	19
M5	0	0	0	36	0	24
M8	2	5	6	29	0	19
M9	0	0	0	49	0	11
M10	0	0	0	52	0	8
Total	2	5	6	506	30	112

Appendix J. BGLMM Models of Phonology and Sentence Mode in M-Corpus.

J1. Likelihood of H%, non-register-tier analysis

Table J1.1 Summary of Likelihood of H%.

```
Cov prior  : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : prompt ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Fixef prior: normal(sd = c(10, 2.5, ...), corr = c(0 ...), common.scale = FALSE)
Prior dev  : 24.292

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: `H%` ~ mode + gender + (1 | speaker) + (1 | prompt)
   Data: m_corpus_is_H_boundary
Control:
glmerControl(optimizer = "optimx", calc.derivs = FALSE, optCtrl = list(method =
"nlminb",
    starttests = FALSE, kkt = FALSE))

     AIC      BIC   logLik deviance df.resid
   164.2    195.4    -75.1    150.2      632

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.4111 -0.1227 -0.0412 -0.0180  4.7220

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 3.7152   1.9275
 prompt  (Intercept) 0.7351   0.8574
Number of obs: 639, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.093      1.526  -3.993 6.52e-05 ***
modeMWH       -1.349      2.552  -0.529 0.596977
modeMYN        1.929      1.276   1.511 0.130706
modeMDQ        4.148      1.213   3.420 0.000626 ***
genderM       -2.037      1.497  -1.360 0.173819
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.352
modeMYN -0.710  0.421
modeMDQ -0.765  0.443  0.891
genderM -0.305  0.000 -0.003 -0.008
```

Table J1.2 Drop 1 ChiSq test of model: H% ~ mode + gender + (1 | speaker) + (1 | prompt)

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
mode	3	223.2	65.02	< .001	< .001	p < .05
gender	1	165.64	3.45	.063	.076	

Table J1.3 Conditional and marginal R2

marginal R2	conditional R2
.41	**.75**

Table J1.4 Predicted probabilities of H% (no reg tier)

mode	predicted	conf.low	conf.high	std.error
MDC	0.002	< 0.001	0.043	1.526
MWH	< 0.001	< 0.001	0.069	2.47
MYN	0.015	0.002	0.117	1.092
MDQ	0.125	0.02	0.496	0.984

Table J1.5 Predicted probabilities of H% (no reg tier)

gender	predicted	conf.low	conf.high	std.error
F	0.002	< 0.001	0.043	1.526
M	< 0.001	< 0.001	0.01	1.782

Table J1.6 b1: H% ~ mode + gender + (1 | speaker) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	z.value	p.value
modeMDC	modeMWH	0.26	0.00	38.50	0.66	-0.53	.597
modeMDC	modeMYN	6.88	0.56	84.00	8.78	1.51	.131
modeMDC	modeMDQ	63.30	5.88	682.00	76.78	3.42	< .001
modeMWH	modeMYN	6.87	0.56	84.00	8.78	1.51	.131
modeMWH	modeMDQ	63.20	5.86	682.10	76.71	3.42	< .001
modeMYN	modeMDQ	9.76	3.06	31.10	5.76	3.86	< .001
intercept	genderM	0.14	0.01	2.46	0.2	-1.35	.177

J2. Likelihood of High Register in Nuclear Pitch Accent (Register-tier Analysis)

Table J2.1 Printout of Model

```
Cov prior  : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : prompt ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Fixef prior: normal(sd = c(10, 2.5, ...), corr = c(0 ...), common.scale = FALSE)
Prior dev  : 33.7782

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: nuc_H_reg ~ mode + gender + (1 | speaker) + (1 | prompt)
   Data: m_corpus_h_reg

     AIC      BIC   logLik deviance df.resid
   338.2    369.4   -162.1    324.2      632

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8269 -0.1790 -0.0447 -0.0150  3.0599

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 1.5153   1.2310
 prompt  (Intercept) 0.1848   0.4299
Number of obs: 639, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.3898     0.9529  -6.706 2.00e-11 ***
modeMWH      -1.7054     1.5900  -1.073   0.2835
modeMYN       4.2430     0.7605   5.580 2.41e-08 ***
modeMDQ       6.3876     0.7865   8.121 4.61e-16 ***
genderM       1.6990     0.7697   2.207   0.0273 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.216
modeMYN -0.759  0.273
modeMDQ -0.783  0.261  0.915
genderM -0.406 -0.004  0.024  0.070
```

Table J2.2 Drop 1 ChiSq test of model: nuc_H_reg ~ mode + gender + (1 | speaker) + (1 | prompt)

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
mode	3	652.71	320.5	< .001	< .001	p < .05
gender	1	341.59	5.38	.02	.027	p < .05

Table J2.3 Conditional and marginal R2

marginal R2	conditional R2
.69	.79

Table J2.4 Predicted probabilities of nuc_H_reg (reg tier)

mode	predicted	conf.low	conf.high	std.error
MDC	< 0.01	< 0.01	0.01	0.95
MWH	< 0.01	< 0.01	< 0.01	1.67
MYN	0.1	0.03	0.28	0.62
MDQ	0.5	0.24	0.76	0.59

Table J2.5 Predicted probabilities of nuc_H_reg (reg tier)

gender	**predicted**	**conf.low**	**conf.high**	**std.error**
F	< 0.01	< 0.01	0.01	0.95
M	< 0.01	< 0.01	0.06	0.95

Table J2.6 predicted probability of nuc_H_reg

group	**estimate**	**predicted**	**conf.low**	**conf.high**	**std.error**
F	MDC	< 0.01	< 0.01	0.01	0.95
F	MWH	< 0.01	< 0.01	< 0.01	1.67
F	MYN	0.1	0.03	0.28	0.62
F	MDQ	0.5	0.24	0.76	0.59
M	MDC	< 0.01	< 0.01	0.06	0.95
M	MWH	< 0.01	< 0.01	0.04	1.66
M	MYN	0.39	0.15	0.69	0.64
M	MDQ	0.85	0.6	0.95	0.66

Table J2.7 b1: nuc_H_reg ~ mode + gender + (1 | speaker) + (1 | prompt)

intercept	**slope**	**estimate**	**conf.low**	**conf.high**	**std.error**	**z.value**	**p.value**
modeMDC	modeMWH	0.18	0.01	4.1	0.29	-1.07	.283
modeMDC	modeMYN	69.60	15.70	309.1	52.94	5.58	< .001
modeMDC	modeMDQ	594.40	127.20	2777.2	467.55	8.12	< .001
modeMWH	modeMYN	68.90	15.50	306.4	52.47	5.56	< .001
modeMWH	modeMDQ	588.60	125.80	2753.6	463.37	8.1	< .001
modeMYN	modeMDQ	9.14	4.84	17.2	2.96	6.83	< .001
intercept	genderM	5.43	1.24	23.8	4.09	2.25	.025

J3. BGLMM Model of likelihood of L%

Table J3.1 Printout of Model

```
Cov prior  : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : prompt ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Prior dev  : -4.5085

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: `L%` ~ mode + gender + (1 | speaker) + (1 | prompt)
   Data: m_corpus_boundaries
Control:
glmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

     AIC      BIC   logLik deviance df.resid
   424.9    456.1   -205.4    410.9      632

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7394 -0.3245 -0.1098 -0.0274  4.1437

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 9.403    3.066
 prompt  (Intercept) 2.148    1.466
Number of obs: 639, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.5112     1.5701  -2.236   0.0253 *
modeMWH      -0.4147     0.4035  -1.028   0.3041
modeMYN       0.6048     0.3725   1.623   0.1045
modeMDQ       2.3111     0.3887   5.946 2.75e-09 ***
genderM      -0.4412     1.9455  -0.227   0.8206
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.113
modeMYN -0.135  0.491
modeMDQ -0.172  0.458  0.541
genderM -0.547  0.000  0.000  0.013
```

Table J3.2 Drop 1 ChiSq test of model: L% ~ mode + gender + (1 | speaker) + (1 | prompt)

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
mode	3	481.66	62.77	< .001	< .001	p < .05
gender	1	422.93	0.04	.838	.914	

Table J3.3 Conditional and marginal R2

marginal R2	conditional R2
.07	.79

Table J3.4 Predicted probabilities of L% (reg tier)

mode	predicted	conf.low	conf.high	std.error
MDC	0.03	< 0.01	0.39	1.57
MWH	0.02	< 0.01	0.3	1.58
MYN	0.05	< 0.01	0.54	1.56
MDQ	0.23	0.01	0.86	1.55

Table J3.5 Predicted probabilities of L% (reg tier)

gender	predicted	conf.low	conf.high	std.error
F	0.03	< 0.01	0.39	1.57
M	0.02	< 0.01	0.35	1.71

Table J3.6 predicted probability of L%

group	estimate	predicted	conf.low	conf.high	std.error
F	MDC	0.03	< 0.01	0.39	1.57
F	MWH	0.02	< 0.01	0.3	1.58
F	MYN	0.05	< 0.01	0.54	1.56
F	MDQ	0.23	0.01	0.86	1.55
M	MDC	0.02	< 0.01	0.35	1.71
M	MWH	0.01	< 0.01	0.27	1.71
M	MYN	0.03	< 0.01	0.5	1.7
M	MDQ	0.16	< 0.01	0.84	1.69

Table J3.7 b1: L% ~ mode + gender + (1 | speaker) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	z.value	p.value
modeMDC	modeMWH	0.66	0.30	1.46	0.27	-1.03	.304
modeMDC	modeMYN	1.83	0.88	3.80	0.68	1.62	.104
modeMDC	modeMDQ	10.10	4.71	21.60	3.92	5.95	< .001
modeMWH	modeMYN	2.77	1.28	5.98	1.09	2.6	.009
modeMWH	modeMDQ	15.30	6.80	34.30	6.3	6.6	< .001
modeMYN	modeMDQ	5.51	2.69	11.30	2.01	4.68	< .001
intercept	genderM	0.64	0.01	29.10	1.25	-0.23	.821

J4. Likelihood of H%, non-register-tier analysis, mode-only model

Table J4.1 summary of: H% ~ mode + (1 | speaker) + (1 | prompt)

term	estimate	conf.low	conf.high	std.error	z.value	p.value
(Intercept)	0.00	0.00	0.01	0	-5.17	< .001
modeMWH	0.26	0.01	7.43	0.44	-0.79	.43
modeMYN	6.82	0.96	48.30	6.81	1.92	.054
modeMDQ	63.20	9.84	405.80	59.96	4.37	< .001

Table J4.2 ANOVA of model: H% ~ mode + (1 | speaker) + (1 | prompt)

term	npar	AIC	BIC	logLik	deviance	chi2	df	p.value	p.adj (BH)
fin_phon_model _H_or_not	6	165.64	192.4	-76.82	153.64	65.3	3	< .001	< .001

J5. Likelihood of high register, register-tier analysis, mode-only model

Table J5.1 summary of: nuc_H_reg ~ mode + (1 | speaker) + (1 | prompt)

term	estimate	conf.low	conf.high	std.error	z.value	p.value
(Intercept)	0.00	0.00	0.02	0	-6.04	< .001
modeMWH	0.18	0.01	4.12	0.29	-1.07	.285
modeMYN	69.60	15.60	309.90	53.03	5.57	< .001
modeMDQ	591.20	126.10	2771.80	466.06	8.1	< .001

Table J5.2 ANOVA of model: nuc_H_reg ~ mode + (1 | speaker) + (1 | prompt)

term	npar	AIC	BIC	logLik	deviance	chi2	df	p.	p.adj (BH)
h_reg_model	6	341.6	368.34	-164.8	329.6	318.6	3	< .001	< .001

J6. Likelihood of nuclear-PA-only IP

Table J6.1 Printout of Model

```
Cov prior  : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : prompt ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Fixef prior: normal(sd = c(10, 2.5, ...), corr = c(0 ...), common.scale = FALSE)
Prior dev  : 21.9236

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: nuc_PA_only ~ mode + gender + (1 | speaker) + (1 | prompt)
   Data: m_corpus
Control:
glmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

     AIC      BIC   logLik deviance df.resid
   468.6    499.8   -227.3    454.6      632

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.7675 -0.4036 -0.0991 -0.0094  8.3518

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 2.665    1.632
 prompt  (Intercept) 1.974    1.405
Number of obs: 639, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1314     1.0763  -1.051  0.29316
modeMWH      -3.7780     0.8654  -4.366 1.27e-05 ***
modeMYN       0.9085     0.3042   2.986  0.00282 **
modeMDQ       0.9282     0.3065   3.028  0.00246 **
genderM      -2.1253     1.0567  -2.011  0.04430 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.051
modeMYN -0.154  0.174
modeMDQ -0.155  0.173  0.551
genderM -0.404  0.010 -0.022 -0.016
```

Table J6.2 Drop 1 ChiSq test of model: nuc_PA_only ~ mode + gender + (1 | speaker) + (1 | prompt)

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
mode	3	576.7	114.1	< .001	< .001	p < .05
gender	1	472.06	5.47	.019	.027	p < .05

Table J6.3 Conditional and marginal R2

marginal R2	conditional R2
.39	.75

Table J6.4 Predicted probabilities of nuc_PA_only (reg tier)

mode	predicted	conf.low	conf.high	std.error
MDC	0.244	0.038	0.727	1.076
MWH	0.007	< 0.001	0.094	1.346
MYN	0.445	0.089	0.867	1.072
MDQ	0.449	0.091	0.87	1.072

Table J6.5 Predicted probabilities of nuc_PA_only (reg tier)

gender	predicted	conf.low	conf.high	std.error
F	0.244	0.038	0.727	1.076
M	0.037	0.004	0.274	1.164

Table J6.6 predicted probability of nuc_PA_only

group	estimate	predicted	conf.low	conf.high	std.error
F	MDC	0.244	0.038	0.727	1.076
F	MWH	0.007	< 0.001	0.094	1.346
F	MYN	0.445	0.089	0.867	1.072
F	MDQ	0.449	0.091	0.87	1.072
M	MDC	0.037	0.004	0.274	1.164
M	MWH	< 0.001	< 0.001	0.014	1.424
M	MYN	0.087	0.01	0.479	1.155
M	MDQ	0.089	0.01	0.484	1.156

Table J6.7 b1: nuc_PA_only ~ mode + gender + (1 | speaker) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	z.value	p.value
modeMDC	modeMWH	0.02	0.00	0.12	0.02	-4.37	< .001
modeMDC	modeMYN	2.48	1.37	4.50	0.75	2.99	.003
modeMDC	modeMDQ	2.53	1.39	4.61	0.78	3.03	.002
modeMWH	modeMYN	56.80	15.20	212.00	38.17	6.01	< .001
modeMWH	modeMDQ	57.90	15.50	216.50	38.95	6.03	< .001
modeMYN	modeMDQ	1.06	0.60	1.86	0.31	0.19	.847
intercept	genderM	0.12	0.01	0.95	0.13	-2.01	.044

J7. Probability of Either nuclear-PA-only IP or High Register (Register-Tier Analysis)

Table J7.1 Printout of Model

```
Cov prior  : speaker ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
           : prompt ~ wishart(df = 3.5, scale = Inf, posterior.scale = cov,
common.scale = TRUE)
Prior dev  : -1.515

Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [bglmerMod]
 Family: binomial  ( logit )
Formula: at_least_1_strat ~ mode + gender + (1 | speaker) + (1 | prompt)
   Data: m_corpus
Control:
glmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

     AIC      BIC   logLik deviance df.resid
   487.7    519.0   -236.9    473.7      632

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.0349 -0.3533 -0.0470  0.3933 15.8488

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 1.655    1.287
 prompt  (Intercept) 1.659    1.288
Number of obs: 639, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.68626    0.94468  -1.785   0.0743 .
modeMWH     -4.24418    1.08942  -3.896 9.79e-05 ***
modeMYN      2.06629    0.29979   6.893 5.48e-12 ***
modeMDQ      3.57313    0.35405  10.092  < 2e-16 ***
genderM     -0.06275    0.81600  -0.077   0.9387
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.035
modeMYN -0.191  0.113
modeMDQ -0.178  0.085  0.575
genderM -0.391 -0.010  0.005  0.009
```

Table J7.2 Drop 1 ChiSq test of model: at_least_1_strat ~ mode + gender + (1 | speaker) + (1 | prompt)

factor	npar	AIC	LRT	Pr(Chi)	p.adj (BH)	signif.
mode	3	821.22	339.47	< .001	< .001	p < .05
gender	1	485.76	0.01	.932	.932	

Table J7.3 Conditional and marginal R2

marginal R2	conditional R2
.57	.78

Table J7.4 Predicted probabilities of at_least_1_strat (reg tier)

mode	predicted	conf.low	conf.high	std.error
MDC	0.156	0.028	0.541	0.945
MWH	0.003	< 0.001	0.041	1.417
MYN	0.594	0.19	0.901	0.935
MDQ	0.868	0.507	0.977	0.948

Table J7.5 Predicted probabilities of at_least_1_strat (reg tier)

gender	predicted	conf.low	conf.high	std.error
F	0.156	0.028	0.541	0.945
M	0.148	0.025	0.541	0.977

Table J7.6 b1: at_least_1_strat ~ mode + gender + (1 | speaker) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	z.value	p.value
modeMDC	modeMWH	0.01	0.00	0.12	0.02	-3.9	< .001
modeMDC	modeMYN	7.89	4.39	14.20	2.37	6.89	< .001
modeMDC	modeMDQ	35.60	17.80	71.30	12.61	10.09	< .001
modeMWH	modeMYN	550.30	64.10	4722.90	603.57	5.75	< .001
modeMWH	modeMDQ	2483.20	278.40	22149.10	2772	7	< .001
modeMYN	modeMDQ	4.51	2.48	8.21	1.38	4.94	< .001
intercept	genderM	0.94	0.19	4.65	0.77	-0.08	.939

Appendix K. LMEMs of Nuclear Phonetic Parameters in M-Corpus

K1. Mode-only model: `l_t`

Table K1.1 Model Printout

```
Formula:
l_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: l_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)
   Data: m_corpus %>% filter(abs(scale(resid(l_t_mode_only_mdl))) <= 3.5)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 5435.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.85851 -0.65843 -0.00912  0.63640  3.12130

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept) 211.220  14.533
 fin_phon (Intercept)  48.700   6.979
 prompt   (Intercept)   3.363   1.834
 Residual             320.056  17.890
Number of obs: 629, groups:  speaker, 11; fin_phon, 3; prompt, 3

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  90.4416     7.5326   9.9604  12.007 3.02e-07 ***
modeMWH       0.4117     2.0049 610.9682   0.205 0.837381
modeMYN      -2.2791     2.0187 611.7813  -1.129 0.259340
modeMDQ     -20.9825     2.1448 598.8245  -9.783  < 2e-16 ***
genderM     -53.5779     8.9184   8.9697  -6.008 0.000203 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.130
modeMYN -0.146  0.500
modeMDQ -0.182  0.467  0.492
genderM -0.537 -0.004  0.000  0.004
```

Table K1.2 ANOVA of model: l_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	41965	13988	3	607.5	43.71	< .001	< .001	p < .05
gender	11551	11551	1	8.97	36.09	< .001	< .001	p < .05

Table K1.3 Conditional and marginal R2

marginal R2	conditional R2
.57	.76

Table K1.4 Predicted values of l_t (mode only model)

mode	predicted	conf.low	conf.high	std.error
MDC	90	76	105	8
MWH	91	76	106	8
MYN	88	73	103	8
MDQ	69	55	84	7

Table K1.5 Predicted values of l_t (mode only model)

gender	predicted	conf.low	conf.high	std.error
F	90	76	105	8
M	37	21	53	8

Table K1.6 b1: l_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	0.4	-3.5	4.3	2	0.21	610.97	.837
modeMDC	modeMYN	-2.3	-6.2	1.7	2.02	-1.13	611.78	.259
modeMDC	modeMDQ	-21.0	-25.2	-16.8	2.14	-9.78	598.82	< .001
modeMWH	modeMYN	-2.7	-6.6	1.3	2.01	-1.34	612.04	.182
modeMWH	modeMDQ	-21.4	-25.6	-17.2	2.14	-9.97	597.81	< .001
modeMYN	modeMDQ	-18.7	-22.8	-14.6	2.1	-8.9	609.62	< .001
intercept	genderM	-53.6	-73.8	-33.4	8.92	-6.01	8.97	< .001

K2. Mode-only model: l_f0

Table K2.1 Model Printout

```
Formula:
l_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: l_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)
   Data: m_corpus %>% filter(abs(scale(resid(l_f0_mode_only_mdl))) <=      3)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2251.2

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.85679 -0.61593 -0.02573  0.59624  3.12747

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept) 0.340    0.5831
 fin_phon (Intercept) 1.938    1.3922
 Residual             2.000    1.4143
Number of obs: 627, groups:  speaker, 11; fin_phon, 3

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  -2.2654     0.8544   2.4149  -2.651   0.0967 .
modeMWH       0.1218     0.1582 611.1553   0.770   0.4417
modeMYN       1.6236     0.1589 611.3276  10.219   <2e-16 ***
modeMDQ       2.4891     0.1713 615.7998  14.535   <2e-16 ***
genderM      -0.0676     0.3711   8.9975  -0.182   0.8595
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.091
modeMYN -0.103  0.500
modeMDQ -0.128  0.460  0.488
genderM -0.197 -0.006 -0.001  0.005
```

Table K2.2 ANOVA of model: l_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	592.57	197.52	3	613.35	98.75	< .001	< .001	p < .05
gender	0.07	0.07	1	9	0.03	.859	.859	

Table K2.3 Conditional and marginal R2

marginal R2	conditional R2
.2	.63

Table K2.4 Predicted values of l_f0 (mode only model)

mode	predicted	conf.low	conf.high	std.error
MDC	-2.3	-3.9	-0.6	0.9
MWH	-2.1	-3.8	-0.5	0.9
MYN	-0.6	-2.3	1	0.9
MDQ	0.2	-1.4	1.9	0.8

Table K2.5 Predicted values of l_f0 (mode only model)

gender	predicted	conf.low	conf.high	std.error
F	-2.3	-3.9	-0.6	0.9
M	-2.3	-4	-0.6	0.9

Table K2.6 b1: l_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	0.1	-0.2	0.4	0.16	0.77	611.16	.442
modeMDC	modeMYN	1.6	1.3	1.9	0.16	10.22	611.33	< .001
modeMDC	modeMDQ	2.5	2.2	2.8	0.17	14.53	615.8	< .001
modeMWH	modeMYN	1.5	1.2	1.8	0.16	9.47	611.7	< .001
modeMWH	modeMDQ	2.4	2.0	2.7	0.17	13.8	616.57	< .001
modeMYN	modeMDQ	0.9	0.5	1.2	0.17	5.17	614.44	< .001
intercept	genderM	-0.1	-0.9	0.8	0.37	-0.18	9	.859

K3. Mode-only model: `h_t`

Table K3.1 Model Printout

```
Formula:
h_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)
   Data: m_corpus %>% filter(abs(scale(resid(h_t_mode_only_mdl))) <= 3.5)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 5957.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4843 -0.5948  0.0135  0.6804  3.3832

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept)  679.4   26.06
 fin_phon (Intercept)  814.0   28.53
 prompt   (Intercept) 1651.3   40.64
 Residual              697.6   26.41
Number of obs: 631, groups:  speaker, 11; fin_phon, 3; prompt, 3

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) 294.2086    30.7325   4.6619   9.573 0.000307 ***
modeMWH      -0.3732     2.9553 613.0833  -0.126 0.899551
modeMYN      -2.4635     2.9720 613.1963  -0.829 0.407489
modeMDQ     -16.7871     3.1789 615.2306  -5.281 1.79e-07 ***
genderM     -59.8829    15.9259   8.9878  -3.760 0.004495 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.047
modeMYN -0.053  0.499
modeMDQ -0.067  0.463  0.490
genderM -0.235 -0.003  0.000  0.003
```

Table K3.2 ANOVA of model: h_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	25182	8394	3	614.09	12.03	< .001	< .001	p < .05
gender	9863	9863	1	8.99	14.14	.004	.006	p < .05

Table K3.3 Conditional and marginal R2

marginal R2	conditional R2
.19	.85

Table K3.4 Predicted values of h_t (mode only model)

mode	predicted	conf.low	conf.high	std.error
MDC	294	234	354	31
MWH	294	234	354	31
MYN	292	232	352	31
MDQ	277	217	338	31

Table K3.5 Predicted values of h_t (mode only model)

gender	predicted	conf.low	conf.high	std.error
F	294	234	354	31
M	234	173	295	31

Table K3.6 b1: h_t ~ mode + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	-0.4	-6.2	5.4	2.96	-0.13	613.08	.9
modeMDC	modeMYN	-2.5	-8.3	3.4	2.97	-0.83	613.2	.407
modeMDC	modeMDQ	-16.8	-23.0	-10.5	3.18	-5.28	615.23	< .001
modeMWH	modeMYN	-2.1	-7.9	3.7	2.97	-0.7	613.29	.482
modeMWH	modeMDQ	-16.4	-22.7	-10.2	3.18	-5.15	615.43	< .001
modeMYN	modeMDQ	-14.3	-20.4	-8.2	3.11	-4.6	614.63	< .001
intercept	genderM	-59.9	-95.9	-23.8	15.93	-3.76	8.99	.004

K4. Mode-only model: h_f0

Table K4.1 Model Printout

```
Formula:
h_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)
   Data: m_corpus
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2699.8

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.09388 -0.58125  0.00071  0.62306  3.11189

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept) 1.855    1.362
 fin_phon (Intercept) 3.460    1.860
 Residual             3.923    1.981
Number of obs: 632, groups:  speaker, 11; fin_phon, 3

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept)   3.2424     1.2309   3.1953   2.634   0.0731 .
modeMWH       0.4050     0.2216 616.0547   1.828   0.0681 .
modeMYN       1.7187     0.2225 616.2127   7.724 4.57e-14 ***
modeMDQ       4.5107     0.2382 619.0647  18.938  < 2e-16 ***
genderM       0.6758     0.8399   9.0127   0.805   0.4418
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
        (Intr) modMWH modMYN modMDQ
modeMWH -0.088
modeMYN -0.100  0.499
modeMDQ -0.126  0.463  0.491
genderM -0.310 -0.004  0.000  0.005
```

Table K4.2 ANOVA of model: h_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	1652	550.82	3	617.49	140.4	< .001	< .001	p < .05
gender	2.54	2.54	1	9.01	0.65	.442	.485	

Table K4.3 Conditional and marginal R2

marginal R2	conditional R2
.25	.68

Table K4.4 Predicted values of h_f0 (mode only model)

mode	predicted	conf.low	conf.high	std.error
MDC	3.2	0.8	5.7	1.2
MWH	3.6	1.2	6.1	1.2
MYN	5	2.6	7.4	1.2
MDQ	7.8	5.4	10.2	1.2

Table K4.5 Predicted values of h_f0 (mode only model)

gender	predicted	conf.low	conf.high	std.error
F	3.2	0.8	5.7	1.2
M	3.9	1.5	6.4	1.3

Table K4.6 b1: h_f0 ~ mode + gender + (1 | speaker) + (1 | fin_phon)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	0.4	0.0	0.8	0.22	1.83	616.05	.068
modeMDC	modeMYN	1.7	1.3	2.2	0.22	7.72	616.21	< .001
modeMDC	modeMDQ	4.5	4.0	5.0	0.24	18.94	619.06	< .001
modeMWH	modeMYN	1.3	0.9	1.8	0.22	5.91	616.38	< .001
modeMWH	modeMDQ	4.1	3.6	4.6	0.24	17.21	619.41	< .001
modeMYN	modeMDQ	2.8	2.3	3.2	0.23	11.99	618.27	< .001
intercept	genderM	0.7	-1.2	2.6	0.84	0.8	9.01	.442

K5. Mode-plus-phonology model: l_t

Table K5.1 Model Printout

```
Formula:
l_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | prompt) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: l_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | prompt) +
    (1 | fin_phon)
   Data: m_corpus %>% filter(abs(scale(resid(l_t_mode_phon_mdl))) <= 3.25)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 5407.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1953 -0.6353  0.0100  0.6158  3.0377

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept) 203.588  14.268
 prompt   (Intercept)   4.017   2.004
 fin_phon (Intercept)  45.037   6.711
 Residual             314.526  17.735
Number of obs: 629, groups:  speaker, 11; prompt, 3; fin_phon, 3

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)     89.92937    7.37635  10.20665  12.192 2.07e-07 ***
modeMWH          0.35594    1.98763 608.03182   0.179 0.857937
modeMYN          0.08794    2.10780 609.06516   0.042 0.966734
modeMDQ        -15.66954    2.58873 608.29016  -6.053 2.49e-09 ***
acc_phon^[L*]H  -3.48555    8.16881 613.63132  -0.427 0.669755
acc_phonL*^[H]  -7.08523    3.86835 613.06507  -1.832 0.067498 .
acc_phon^[L*H]  -9.27957    2.54784 612.04755  -3.642 0.000293 ***
genderM        -52.09089    8.77432   9.02508  -5.937 0.000216 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) modMWH modMYN modMDQ a_^[L*] a_L*^[ a_^[L*H
modeMWH     -0.132
modeMYN     -0.146  0.472
modeMDQ     -0.161  0.380  0.556
acc_p^[L*]H -0.001  0.006 -0.124 -0.141
acc_pL*^[H]  0.002  0.002 -0.146 -0.382  0.066
acc_p^[L*H]  0.021  0.008 -0.301 -0.515  0.188   0.290
genderM     -0.539 -0.004  0.016  0.026 -0.032   0.004 -0.053
```

ANOVA of model: l_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | prompt) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	17393	5798	3	610	18.43	< .001	< .001	p < .05
acc_phon	4403	1468	3	613.17	4.67	.003	.004	p < .05
gender	11085	11085	1	9.03	35.24	< .001	< .001	p < .05

Table K5.2 Conditional and marginal R2

marginal R2	conditional R2
.58	.77

Table K5.3 Predicted values of l_t (mode-and-phonology model)

mode	predicted	conf.low	conf.high	std.error
MDC	90	75	104	7
MWH	90	76	105	7
MYN	90	76	104	7
MDQ	74	60	89	7

Table K5.4 Predicted values of l_t (mode-and-phonology model)

acc_phon	predicted	conf.low	conf.high	std.error
L*H	90	75	104	7
^[L*]H	86	65	108	11
L*^[H]	83	67	99	8
^[L*H]	81	65	96	8

Table K5.5 Predicted values of l_t (mode-and-phonology model)

gender	predicted	conf.low	conf.high	std.error
F	90	75	104	7
M	38	22	53	8

K6. Mode-plus-phonology model: l_f0

Table K6.1 Model Printout

```
Formula:
l_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: l_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)
   Data: m_corpus %>% filter(abs(scale(resid(l_f0_mode_phon_mdl))) <=      3)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2085.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.0301 -0.6172 -0.0333  0.5920  3.9661

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept) 0.498    0.7057
 fin_phon (Intercept) 1.190    1.0907
 Residual             1.539    1.2407
Number of obs: 625, groups:  speaker, 11; fin_phon, 3

Fixed effects:
               Estimate Std. Error       df t value Pr(>|t|)
(Intercept)     -2.3348     0.7080   2.9344  -3.298   0.0473 *
modeMWH          0.1095     0.1390 606.1004   0.788   0.4312
modeMYN          1.0703     0.1471 606.9555   7.277 1.06e-12 ***
modeMDQ          1.3682     0.1836 610.5712   7.452 3.17e-13 ***
acc_phon^[L*]H   2.8679     0.5697 613.6987   5.034 6.31e-07 ***
acc_phonL*^[H]   0.3470     0.2769 612.0962   1.253   0.2107
acc_phon^[L*H]   2.2452     0.1790 611.6621  12.544  < 2e-16 ***
genderM         -0.4799     0.4405   9.1074  -1.089   0.3039
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) modMWH modMYN modMDQ a_^[L*] a_L*^[ a_^[L*H
modeMWH     -0.095
modeMYN     -0.106  0.470
modeMDQ     -0.113  0.372  0.556
acc_p^[L*]H -0.001  0.006 -0.121 -0.142
acc_pL*^[H] -0.004  0.002 -0.151 -0.394  0.063
acc_p^[L*H]  0.010  0.008 -0.307 -0.524  0.190   0.307
genderM     -0.282 -0.006  0.022  0.035 -0.044   0.007 -0.071
```

Table K6.2 ANOVA of model: l_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	127.8	42.6	3	608.1	27.67	< .001	< .001	p < .05
acc_phon	264.95	88.32	3	612.23	57.37	< .001	< .001	p < .05
gender	1.83	1.83	1	9.11	1.19	.304	.356	

Table K6.3 Conditional and marginal R2

marginal R2	conditional R2
.33	.68

Table K6.4 Predicted values of l_f0 (mode-and-phonology model)

mode	predicted	conf.low	conf.high	std.error
MDC	-2.3	-3.7	-0.9	0.7
MWH	-2.2	-3.6	-0.8	0.7
MYN	-1.3	-2.7	0.1	0.7
MDQ	-1	-2.4	0.4	0.7

Table K6.5 Predicted values of l_f0 (mode-and-phonology model)

acc_phon	predicted	conf.low	conf.high	std.error
L*H	-2.3	-3.7	-0.9	0.7
^[L*]H	0.5	-1.2	2.3	0.9
L*^[H]	-2	-3.5	-0.5	0.8
^[L*H]	< 0.1	-1.5	1.3	0.7

Table K6.6 Predicted values of l_f0 (mode-and-phonology model)

gender	predicted	conf.low	conf.high	std.error
F	-2.3	-3.7	-0.9	0.7
M	-2.8	-4.2	-1.4	0.7

Table K6.7 b1: l_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	0.1	-0.2	0.4	0.14	0.79	606.1	.431
modeMDC	modeMYN	1.1	0.8	1.4	0.15	7.28	606.96	< .001
modeMDC	modeMDQ	1.4	1.0	1.7	0.18	7.45	610.57	< .001
modeMWH	modeMYN	1.0	0.7	1.2	0.15	6.52	607.28	< .001
modeMWH	modeMDQ	1.3	0.9	1.6	0.18	6.82	610.99	< .001
modeMYN	modeMDQ	0.3	0.0	0.6	0.16	1.87	609.17	.062
acc_phonL*H	acc_phon^[L*]H	2.9	1.7	4.0	0.57	5.03	613.7	< .001
acc_phonL*H	acc_phonL*^[H]	0.3	-0.2	0.9	0.28	1.25	612.1	.211
acc_phonL*H	acc_phon^[L*H]	2.2	1.9	2.6	0.18	12.54	611.66	< .001
acc_phon^[L*]H	acc_phonL*^[H]	-2.5	-3.7	-1.3	0.62	-4.08	613.83	< .001
acc_phon^[L*]H	acc_phon^[L*H]	-0.6	-1.7	0.5	0.56	-1.1	613.16	.27
acc_phonL*^[H]	acc_phon^[L*H]	1.9	1.3	2.4	0.28	6.78	611.16	< .001
intercept	genderM	-0.5	-1.5	0.5	0.44	-1.09	9.11	.304

K7. Mode-plus-phonology model: h_t

Table K7.1 Model Printout

```
Formula:
h_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon) +
    (1 | prompt)
   Data: m_corpus %>% filter(abs(scale(resid(h_t_mode_phon_mdl))) <= 3.5)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 5935.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4398 -0.5743 -0.0116  0.6965  3.3907

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept)  678.7   26.05
 fin_phon (Intercept)  700.6   26.47
 prompt   (Intercept) 1693.7   41.15
 Residual              684.1   26.16
Number of obs: 632, groups:  speaker, 11; fin_phon, 3; prompt, 3

Fixed effects:
               Estimate Std. Error       df t value Pr(>|t|)
(Intercept)    294.0512    30.3408   4.4719   9.692 0.000361 ***
modeMWH         -0.4437     2.9267 611.0976  -0.152 0.879560
modeMYN         -1.0557     3.0993 611.4414  -0.341 0.733497
modeMDQ        -13.5875     3.8229 613.3851  -3.554 0.000408 ***
acc_phon^[L*]H -71.7856    12.0905 614.6459  -5.937 4.85e-09 ***
acc_phonL*^[H]  -2.5198     5.7084 613.1265  -0.441 0.659064
acc_phon^[L*H]  -3.5558     3.7364 613.1198  -0.952 0.341641
genderM        -58.0674    15.9342   9.0193  -3.644 0.005346 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) modMWH modMYN modMDQ a_^[L*] a_L*^[ a_^[L*H
modeMWH     -0.047
modeMYN     -0.053  0.471
modeMDQ     -0.059  0.378  0.557
acc_p^[L*]H  0.000  0.005 -0.122 -0.143
acc_pL*^[H]  0.001  0.002 -0.147 -0.379  0.067
acc_p^[L*H]  0.007  0.007 -0.304 -0.511  0.187   0.287
genderM     -0.238 -0.003  0.013  0.022 -0.026   0.003 -0.043
```

Table K7.2 ANOVA of model: h_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)
mode	11604	3868	3	612.12	5.65	< .001	.001
acc_phon	24142	8047	3	613.56	11.76	< .001	< .001
gender	9085	9085	1	9.02	13.28	.005	.007

Table K7.3 Conditional and marginal R2

marginal R2	conditional R2
.21	.86

Table K7.4 Predicted values of h_t (mode-and-phonology model)

mode	predicted	conf.low	conf.high	std.error
MDC	294	235	354	30
MWH	294	234	353	30
MYN	293	234	352	30
MDQ	280	221	340	30

Table K7.5 Predicted values of h_t (mode-and-phonology model)

acc_phon	predicted	conf.low	conf.high	std.error
L*H	294	235	354	30
^[L*]H	222	158	286	33
L*^[H]	292	231	352	31
^[L*H]	290	231	350	31

Table K7.6 Predicted values of h_t (mode-and-phonology model)

gender	predicted	conf.low	conf.high	std.error
F	294	235	354	30
M	236	176	296	31

Table K7.7 b1: h_t ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	-0.4	-6.2	5.3	2.93	-0.15	611.1	.88
modeMDC	modeMYN	-1.1	-7.1	5.0	3.1	-0.34	611.44	.733
modeMDC	modeMDQ	-13.6	-21.1	-6.1	3.82	-3.55	613.39	< .001
modeMWH	modeMYN	-0.6	-6.7	5.5	3.1	-0.2	611.56	.844
modeMWH	modeMDQ	-13.1	-20.7	-5.6	3.84	-3.42	613.57	< .001
modeMYN	modeMDQ	-12.5	-19.1	-6.0	3.32	-3.78	612.78	< .001
acc_phonL*H	acc_phon^[L*]H	-71.8	-95.5	-48.0	12.09	-5.94	614.65	< .001
acc_phonL*H	acc_phonL*^[H]	-2.5	-13.7	8.7	5.71	-0.44	613.13	.659
acc_phonL*H	acc_phon^[L*H]	-3.6	-10.9	3.8	3.74	-0.95	613.12	.342
acc_phon^[L*]H	acc_phonL*^[H]	69.3	43.7	94.8	13.02	5.32	614.61	< .001
acc_phon^[L*]H	acc_phon^[L*H]	68.2	44.7	91.7	11.97	5.7	614.28	< .001
acc_phonL*^[H]	acc_phon^[L*H]	-1.0	-12.5	10.5	5.86	-0.18	613.04	.86
intercept	genderM	-58.1	-94.1	-22.0	15.93	-3.64	9.02	.005

K8. Mode-plus-phonology model: l_f0

Table K8.1 Model Printout

```
Formula:
h_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)
   Data: m_corpus
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2549.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6284 -0.6663 -0.0212  0.6513  3.2294

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept) 1.506    1.227
 fin_phon (Intercept) 2.894    1.701
 Residual             3.100    1.761
Number of obs: 632, groups:  speaker, 11; fin_phon, 3

Fixed effects:
               Estimate Std. Error       df t value Pr(>|t|)
(Intercept)      3.4168     1.1215   3.1566   3.047    0.052 .
modeMWH          0.4186     0.1970 613.0484   2.125    0.034 *
modeMYN          0.9502     0.2085 613.6502   4.556 6.28e-06 ***
modeMDQ          2.6830     0.2571 616.5119  10.435  < 2e-16 ***
acc_phon^[L*]H  -0.1984     0.8093 618.9536  -0.245    0.806
acc_phonL*^[H]   3.2700     0.3829 616.7286   8.541  < 2e-16 ***
acc_phon^[L*H]   2.8945     0.2509 616.8778  11.535  < 2e-16 ***
genderM          0.2900     0.7584   9.0762   0.382    0.711
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) modMWH modMYN modMDQ a_^[L*] a_L*^[ a_^[L*H
modeMWH     -0.086
modeMYN     -0.096  0.471
modeMDQ     -0.107  0.378  0.557
acc_p^[L*]H -0.001  0.006 -0.120 -0.141
acc_pL*^[H]  0.001  0.002 -0.146 -0.378  0.061
acc_p^[L*H]  0.014  0.007 -0.305 -0.512  0.191   0.289
genderM     -0.306 -0.004  0.018  0.031 -0.037   0.005 -0.061
```

Table K8.2 ANOVA of model: h_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	352.01	117.34	3	614.61	37.85	< .001	< .001	p < .05
acc_phon	524.02	174.67	3	617.38	56.34	< .001	< .001	p < .05
gender	0.45	0.45	1	9.08	0.15	.711	.733	

Table K8.3 Conditional and marginal R2

marginal R2	conditional R2
.35	.73

Table K8.4 Predicted values of h_f0 (mode-and-phonology model)

mode	predicted	conf.low	conf.high	std.error
MDC	3.4	1.2	5.6	1.1
MWH	3.8	1.6	6	1.1
MYN	4.4	2.2	6.6	1.1
MDQ	6.1	3.9	8.3	1.1

Table K8.5 Predicted values of h_f0 (mode-and-phonology model)

acc_phon	predicted	conf.low	conf.high	std.error
L*H	3.4	1.2	5.6	1.1
^[L*]H	3.2	0.5	5.9	1.4
L*^[H]	6.7	4.4	9	1.2
^[L*H]	6.3	4.1	8.6	1.2

Table K8.6 Predicted values of h_f0 (mode-and-phonology model)

gender	predicted	conf.low	conf.high	std.error
F	3.4	1.2	5.6	1.1
M	3.7	1.5	6	1.1

Table K8.7 b1: h_f0 ~ mode + acc_phon + gender + (1 | speaker) + (1 | fin_phon)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	0.4	0.0	0.8	0.2	2.12	613.05	.034
modeMDC	modeMYN	0.9	0.5	1.4	0.21	4.56	613.65	< .001
modeMDC	modeMDQ	2.7	2.2	3.2	0.26	10.43	616.51	< .001
modeMWH	modeMYN	0.5	0.1	0.9	0.21	2.55	613.87	.011
modeMWH	modeMDQ	2.3	1.8	2.8	0.26	8.77	616.82	< .001
modeMYN	modeMDQ	1.7	1.3	2.2	0.22	7.76	615.48	< .001
acc_phonL*H	acc_phon^[L*]H	-0.2	-1.8	1.4	0.81	-0.25	618.95	.806
acc_phonL*H	acc_phonL*^[H]	3.3	2.5	4.0	0.38	8.54	616.73	< .001
acc_phonL*H	acc_phon^[L*H]	2.9	2.4	3.4	0.25	11.53	616.88	< .001
acc_phon^[L*]H	acc_phonL*^[H]	3.5	1.8	5.2	0.87	3.97	618.85	< .001
acc_phon^[L*]H	acc_phon^[L*H]	3.1	1.5	4.7	0.8	3.87	618.38	< .001
acc_phonL*^[H]	acc_phon^[L*H]	-0.4	-1.1	0.4	0.39	-0.96	616.57	.339
intercept	genderM	0.3	-1.4	2.0	0.76	0.38	9.08	.711

Appendix L. LMEMs of Global Phonetic Parameters in M-Corpus

L1. Mode-only model: mean f_0

Table L1.1 Model Printout

```
Formula:
utt_mean_f0 ~ mode + gender + (1 + mode | speaker) + (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: utt_mean_f0 ~ mode + gender + (1 + mode | speaker) + (1 | prompt)
   Data: m_corpus %>% filter(abs(scale(resid(utt_mean_f0_mode_mdl))) <=
    2.9)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 1973.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6495 -0.6228 -0.0442  0.6076  3.4917

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 speaker  (Intercept) 0.146217 0.38238
          modeMWH     0.899279 0.94830  -0.42
          modeMYN     0.522309 0.72271  -0.82  0.64
          modeMDQ     1.728819 1.31485   0.08 -0.46 -0.58
 prompt   (Intercept) 0.003706 0.06088
 Residual             1.178717 1.08569
Number of obs: 632, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -0.5235     0.1590 10.5796  -3.292 0.007542 **
modeMWH       0.4738     0.3104 10.0649   1.526 0.157742
modeMYN       0.7610     0.2490 10.0581   3.056 0.012050 *
modeMDQ       2.1973     0.4160 10.0327   5.282 0.000353 ***
genderM       0.1256     0.1306  9.0656   0.962 0.361089
```

Table L1.2 ANOVA of model: utt_mean_f0 ~ mode + gender + (1 + mode | speaker) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	72.72	24.24	3	10.04	20.56	< .001	< .001	p < .05
gender	1.09	1.09	1	9.07	0.92	.361	.409	

Table L1.3 Conditional and marginal R2

marginal R2	conditional R2
.25	.54

Table L1.4 Predicted values of utt_mean_f0 (mode only model)

mode	predicted	conf.low	conf.high	std.error
MDC	-0.5	-0.8	-0.2	0.2
MWH	< 0.1	-0.6	0.5	0.3
MYN	0.2	-0.1	0.6	0.2
MDQ	1.7	0.8	2.5	0.4

Table L1.5 Predicted values of utt_mean_f0 (mode only model)

gender	predicted	conf.low	conf.high	std.error
F	-0.5	-0.8	-0.2	0.2
M	-0.4	-0.7	< 0.1	0.2

Table L1.6 b1: utt_mean_f0 ~ mode + gender + (1 + mode | speaker) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	0.5	-0.2	1.2	0.31	1.53	10.06	.158
modeMDC	modeMYN	0.8	0.2	1.3	0.25	3.06	10.06	.012
modeMDC	modeMDQ	2.2	1.3	3.1	0.42	5.28	10.03	< .001
modeMWH	modeMYN	0.3	-0.3	0.9	0.25	1.14	9.98	.282
modeMWH	modeMDQ	1.7	0.4	3.1	0.6	2.87	10.02	.017
modeMYN	modeMDQ	1.4	0.2	2.7	0.57	2.53	10.05	.03
intercept	genderM	0.1	-0.2	0.4	0.13	0.96	9.07	.361

L2. Mode-only model: slope f_0(t)

Table L2.1 Model Printout

```
Formula:
utt_slope ~ mode + gender + (1 + mode | speaker) + (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: utt_slope ~ mode + gender + (1 + mode | speaker) + (1 | prompt)
   Data: m_corpus %>% filter(abs(scale(resid(utt_slope_mode_mdl))) <=      2.5)
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2744.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8638 -0.5558  0.0373  0.5528  2.8622

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 speaker  (Intercept)  5.6582  2.3787
          modeMWH     16.0881  4.0110   -0.55
          modeMYN      7.1224  2.6688   -0.81  0.07
          modeMDQ     14.0526  3.7487   -0.46 -0.11  0.71
 prompt   (Intercept)  0.1671  0.4087
 Residual              3.9782  1.9945
Number of obs: 622, groups:  speaker, 11; prompt, 3

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -1.0847     0.8026 11.8526  -1.352 0.201735
modeMWH      -2.4239     1.2297  9.9951  -1.971 0.077008 .
modeMYN       3.5498     0.8350  9.9852   4.251 0.001692 **
modeMDQ       6.6087     1.1547 10.0827   5.723 0.000186 ***
genderM       1.4541     0.4890  9.0729   2.974 0.015475 *
```

Table L2.2 ANOVA of model: utt_slope ~ mode + gender + (1 + mode | speaker) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	140.58	46.86	3	9.95	11.78	.001	.002	p < .05
gender	35.18	35.18	1	9.07	8.84	.015	.019	p < .05

Table L2.3 Conditional and marginal R2

marginal R2	conditional R2
.5	.83

Table L2.4 Predicted values of utt_slope (mode only model)

mode	predicted	conf.low	conf.high	std.error
MDC	-1.1	-2.7	0.5	0.8
MWH	-3.5	-5.6	-1.4	1.1
MYN	2.5	1.3	3.6	0.6
MDQ	5.5	3.4	7.7	1.1

Table L2.5 Predicted values of utt_slope (mode only model)

gender	predicted	conf.low	conf.high	std.error
F	-1.1	-2.7	0.5	0.8
M	0.4	-1.2	2	0.8

Table L2.6 b1: utt_slope ~ mode + gender + (1 + mode | speaker) + (1 | prompt)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
modeMDC	modeMWH	-2.4	-5.2	0.3	1.23	-1.97	10	.077
modeMDC	modeMYN	3.5	1.7	5.4	0.83	4.25	9.99	.002
modeMDC	modeMDQ	6.6	4.0	9.2	1.15	5.72	10.08	< .001
modeMWH	modeMYN	6.0	2.8	9.1	1.42	4.19	10	.002
modeMWH	modeMDQ	9.0	5.1	13.0	1.76	5.13	10.03	< .001
modeMYN	modeMDQ	3.1	1.2	4.9	0.83	3.69	9.89	.004
intercept	genderM	1.5	0.3	2.6	0.49	2.97	9.07	.015

L3. Mode-plus-phonology model: mean f_0

Table L3.1 Model Printout

```
Formula:
utt_mean_f0 ~ mode + h_start + acc_phon + fin_phon + gender + (1 + mode |
speaker) + (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: utt_mean_f0 ~ mode + h_start + acc_phon + fin_phon + gender +
    (1 + mode | speaker) + (1 | prompt)
   Data: m_corpus
Control:
lmerControl(optimizer = "optimx", calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: 2035.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4087 -0.6025 -0.0326  0.6131  3.6202

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 speaker  (Intercept) 0.090196 0.30033
          modeMWH     0.825785 0.90873  -0.14
          modeMYN     0.305546 0.55276  -0.66  0.67
          modeMDQ     1.565800 1.25132   0.09 -0.55 -0.71
 prompt   (Intercept) 0.006215 0.07884
 Residual             1.271941 1.12780
Number of obs: 639, groups:  speaker, 11; prompt, 3

Fixed effects:
                   Estimate Std. Error        df t value Pr(>|t|)
(Intercept)        -0.38442    0.17917  20.45107  -2.146 0.044077 *
modeMWH             0.40351    0.30680  10.71275   1.315 0.215885
modeMYN             0.58589    0.21574  10.96919   2.716 0.020127 *
modeMDQ             1.84224    0.41386  11.50021   4.451 0.000877 ***
h_startno H start  -0.13878    0.12920 270.29705  -1.074 0.283719
acc_phon^[L*H]      0.79535    0.17859 184.83268   4.453 1.46e-05 ***
acc_phonL*^[H]      0.07327    0.25463 418.27622   0.288 0.773683
acc_phon>H*        -0.95567    0.53978 421.25668  -1.770 0.077370 .
acc_phonH*         -1.21410    0.84003 526.46818  -1.445 0.148968
acc_phon^[L*]H     -0.95138    0.51145 245.11536  -1.860 0.064059 .
fin_phonL%          0.45603    0.14958 220.41952   3.049 0.002579 **
genderM            -0.06861    0.16330  10.04770  -0.420 0.683240
```

Table L3.2 ANOVA of model: utt_mean_f0 ~ mode + h_start + acc_phon + fin_phon + gender + (1 + mode | speaker) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)
mode	53.84	17.94	3	11.8	14.11	< .001	< .001
h_start	1.47	1.47	1	270.3	1.15	.284	.345
acc_phon	44.3	8.86	5	350.17	6.97	< .001	< .001
fin_phon	11.82	11.82	1	220.42	9.3	.003	.004
gender	0.22	0.22	1	10.05	0.18	.683	.726

Table L3.3 Conditional and marginal R2

marginal R2	conditional R2
.3	.55

Table L3.4 Predicted values of utt_mean_f0 (mode + phonology model)

mode	predicted	conf.low	conf.high	std.error
MDC	-0.4	-0.7	< 0.1	0.2
MWH	< 0.1	-0.6	0.6	0.3
MYN	0.2	-0.2	0.6	0.2
MDQ	1.5	0.6	2.3	0.4

Table L3.5 Predicted values of utt_mean_f0 (mode + phonology model)

h_start	predicted	conf.low	conf.high	std.error
H start	-0.4	-0.7	< 0.1	0.2
no H start	-0.5	-0.8	-0.2	0.2

Table L3.6 Predicted values of utt_mean_f0 (mode + phonology model)

acc_phon	predicted	conf.low	conf.high	std.error
L*H	-0.4	-0.7	< 0.1	0.2
^[L*H]	0.4	-0.1	0.9	0.3
L*^[H]	-0.3	-0.9	0.3	0.3
^H*	-1.3	-2.5	-0.2	0.6
H*	-1.6	-3.3	< 0.1	0.9
^[L*]H	-1.3	-2.4	-0.2	0.6

Table L3.7 Predicted values of utt_mean_f0 (mode + phonology model)

fin_phon	predicted	conf.low	conf.high	std.error
%	-0.4	-0.7	< 0.1	0.2
L%	< 0.1	-0.4	0.5	0.2

Table L3.8 Predicted values of utt_mean_f0 (mode + phonology model)

gender	predicted	conf.low	conf.high	std.error
F	-0.4	-0.7	< 0.1	0.2
M	-0.5	-0.8	-0.1	0.2

Table L3.9 b1: utt_mean_f0 ~ mode + h_start + acc_phon + fin_phon + gender + (1 + mode | speaker) + (1 | prompt)

intercept	**slope**	**Est.**	**conf. low**	**conf. high**	**std. error**	**t. value**	**df**	**p. value**
modeMDC	modeMWH	0.4	-0.3	1.1	0.31	1.32	10.71	.216
modeMDC	modeMYN	0.6	0.1	1.1	0.22	2.72	10.97	.02
modeMDC	modeMDQ	1.8	0.9	2.7	0.41	4.45	11.5	< .001
modeMWH	modeMYN	0.2	-0.4	0.7	0.26	0.7	13.23	.495
modeMWH	modeMDQ	1.4	0.1	2.8	0.61	2.37	11.14	.037
modeMYN	modeMDQ	1.3	0.1	2.4	0.53	2.38	10.24	.038
acc_phonL*H	acc_phon^[L*H]	0.8	0.4	1.1	0.18	4.45	184.83	< .001
acc_phonL*H	acc_phonL*^[H]	0.1	-0.4	0.6	0.25	0.29	418.28	.774
acc_phonL*H	acc_phon>H*	-1.0	-2.0	0.1	0.54	-1.77	421.26	.077
acc_phonL*H	acc_phonH*	-1.2	-2.9	0.4	0.84	-1.45	526.47	.149
acc_phonL*H	acc_phon^[L*]H	-1.0	-2.0	0.1	0.51	-1.86	245.12	.064
acc_phon^[L*H]	acc_phonL*^[H]	-0.7	-1.3	-0.2	0.27	-2.64	383.07	.009
acc_phon^[L*H]	acc_phon>H*	-1.8	-2.8	-0.7	0.55	-3.21	530.93	.001
acc_phon^[L*H]	acc_phonH*	-2.0	-3.7	-0.3	0.85	-2.36	523.84	.019
acc_phon^[L*H]	acc_phon^[L*]H	-1.7	-2.7	-0.8	0.51	-3.46	402.93	< .001
acc_phonL*^[H]	acc_phon>H*	-1.0	-2.2	0.1	0.59	-1.74	452.76	.082
acc_phonL*^[H]	acc_phonH*	-1.3	-3.0	0.4	0.88	-1.46	550.07	.144
acc_phonL*^[H]	acc_phon^[L*]H	-1.0	-2.1	0.1	0.56	-1.84	291.98	.067
acc_phon>H*	acc_phonH*	-0.3	-2.1	1.6	0.96	-0.27	583.59	.787
acc_phon>H*	acc_phon^[L*]H	0.0	-1.4	1.4	0.69	0.01	595.84	.995
acc_phonH*	acc_phon^[L*]H	0.3	-1.6	2.2	0.96	0.27	434.15	.785
intercept	h_startno H start	-0.1	-0.4	0.1	0.13	-1.07	270.3	.284
intercept	fin_phonL%	0.5	0.2	0.8	0.15	3.05	220.44	.003
intercept	genderM	-0.1	-0.4	0.3	0.16	-0.42	10.05	.683

L4. Mode-plus-phonology model: f_0(t)

Table L4.1 Model Printout

```
Formula:
utt_slope ~ mode + h_start + acc_phon + fin_phon + gender + (1 + mode | speaker)
+ (1 | prompt)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: utt_slope ~ mode + h_start + acc_phon + fin_phon + gender + (1 +
    mode | speaker) + (1 | prompt)
   Data: m_corpus %>% filter(abs(scale(resid(utt_slope_full_mdl))) <=      3)
Control:
lmerControl(optCtrl = list(maxit = 1e+09, maxfun = 1e+09, xtol_abs = 1e-09,
    ftol_abs = 1e-09))

REML criterion at convergence: 2627.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.4522 -0.5892 -0.0002  0.5646  3.1413

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 speaker  (Intercept)  5.1184  2.2624
          modeMWH     16.7184  4.0888   -0.44
          modeMYN      5.4323  2.3307   -0.92  0.18
          modeMDQ      9.3523  3.0581   -0.47 -0.09  0.71
 prompt   (Intercept)  0.2996  0.5474
 Residual              3.2407  1.8002
Number of obs: 627, groups:  speaker, 11; prompt, 3

Fixed effects:
                  Estimate Std. Error       df t value Pr(>|t|)
(Intercept)        -1.8878     0.8035  12.0173  -2.350 0.036713 *
modeMWH            -2.1157     1.2529  10.1143  -1.689 0.121823
modeMYN             2.9481     0.7372  10.1221   3.999 0.002463 **
modeMDQ             5.2143     0.9643  10.6883   5.407 0.000237 ***
h_startno H start   0.9716     0.2223 416.1406   4.370 1.57e-05 ***
acc_phon^[L*H]      1.8065     0.3099 220.7479   5.830 1.95e-08 ***
acc_phonL*^[H]      3.2764     0.4221 578.9512   7.763 3.78e-14 ***
acc_phon>H*        -4.2319     0.8974 518.7216  -4.716 3.10e-06 ***
acc_phonH*         -2.5721     1.3850 582.2952  -1.857 0.063805 .
acc_phon^[L*]H     -7.9614     0.8832 424.7222  -9.014  < 2e-16 ***
fin_phonL%         -1.0806     0.2645 408.6294  -4.085 5.30e-05 ***
genderM             2.3869     0.4193   9.9159   5.693 0.000207 ***
```

Table L4.2 ANOVA of model: utt_slope ~ mode + h_start + acc_phon + fin_phon + gender + (1 + mode | speaker) + (1 | prompt)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
mode	102.33	34.11	3	10.08	10.53	.002	.003	p < .05
h_start	61.89	61.89	1	416.14	19.1	< .001	< .001	p < .05
acc_phon	692.06	138.41	5	460.07	42.71	< .001	< .001	p < .05
fin_phon	54.09	54.09	1	408.63	16.69	< .001	< .001	p < .05
gender	105.03	105.03	1	9.92	32.41	< .001	< .001	p < .05

Table L4.3 Conditional and marginal R2

marginal R2	conditional R2
.57	.87

Table L4.4 Predicted values of utt_slope (mode + phonology model)

mode	predicted	conf.low	conf.high	std.error
MDC	-1.89	-3.46	-0.31	0.8
MWH	-4	-6.33	-1.68	1.19
MYN	1.06	0.04	2.08	0.52
MDQ	3.33	1.42	5.23	0.97

Table L4.5 Predicted values of utt_slope (mode + phonology model)

h_start	predicted	conf.low	conf.high	std.error
H start	-1.89	-3.46	-0.31	0.8
no H start	-0.92	-2.47	0.64	0.79

Table L4.6 Predicted values of utt_slope (mode + phonology model)

acc_phon	predicted	conf.low	conf.high	std.error
L*H	-1.89	-3.46	-0.31	0.8
^[L*H]	-0.08	-1.8	1.64	0.88
L*^[H]	1.39	-0.39	3.17	0.91
^H*	-6.12	-8.51	-3.73	1.22
H*	-4.46	-7.58	-1.34	1.59
^[L*]H	-9.85	-12.24	-7.46	1.22

Table L4.7 Predicted values of utt_slope (mode + phonology model)

fin_phon	predicted	conf.low	conf.high	std.error
%	-1.89	-3.46	-0.31	0.8
L%	-2.97	-4.6	-1.34	0.83

Table L4.8 Predicted values of utt_slope (mode + phonology model)

gender	predicted	conf.low	conf.high	std.error
F	-1.89	-3.46	-0.31	0.8
M	0.5	-1.09	2.08	0.81

Table L4.9 b1: utt_slope ~ mode + h_start + acc_phon + fin_phon + gender + (1 + mode | speaker) + (1 | prompt)

intercept	slope	estimate	Conf. low	conf. high	std. error	t.value	df	p.value
modeMDC	modeMWH	-2.12	-4.90	0.67	1.25	-1.69	10.11	.122
modeMDC	modeMYN	2.95	1.31	4.59	0.74	4	10.12	.002
modeMDC	modeMDQ	5.21	3.08	7.34	0.96	5.41	10.69	< .001
modeMWH	modeMYN	5.06	2.10	8.02	1.33	3.8	10.24	.003
modeMWH	modeMDQ	7.33	3.71	10.90	1.63	4.49	10.39	.001
modeMYN	modeMDQ	2.27	0.72	3.81	0.69	3.28	10	.008
acc_phonL*H	acc_phon^[L*H]	1.81	1.20	2.42	0.31	5.83	220.75	< .001
acc_phonL*H	acc_phonL*^[H]	3.28	2.45	4.11	0.42	7.76	578.95	< .001
acc_phonL*H	acc_phon>H*	-4.23	-6.00	-2.47	0.9	-4.72	518.72	< .001
acc_phonL*H	acc_phonH*	-2.57	-5.29	0.15	1.39	-1.86	582.3	.064
acc_phonL*H	acc_phon^[L*]H	-7.96	-9.70	-6.22	0.88	-9.01	424.72	< .001
acc_phon^[L*H]	acc_phonL*^[H]	1.47	0.57	2.37	0.46	3.2	537.71	.001
acc_phon^[L*H]	acc_phon>H*	-6.04	-7.80	-4.28	0.9	-6.73	576.53	< .001
acc_phon^[L*H]	acc_phonH*	-4.38	-7.15	-1.61	1.41	-3.11	585.14	.002
acc_phon^[L*H]	acc_phon^[L*]H	-9.77	-11.40	-8.09	0.85	-11.43	548.32	< .001
acc_phonL*^[H]	acc_phon>H*	-7.51	-9.43	-5.58	0.98	-7.66	557.45	< .001
acc_phonL*^[H]	acc_phonH*	-5.85	-8.70	-3.00	1.45	-4.03	584.1	< .001
acc_phonL*^[H]	acc_phon^[L*]H	-11.20	-13.10	-9.36	0.96	-11.76	500.93	< .001
acc_phon>H*	acc_phonH*	1.66	-1.39	4.71	1.55	1.07	580.14	.286
acc_phon>H*	acc_phon^[L*]H	-3.73	-5.94	-1.52	1.12	-3.32	578.34	< .001
acc_phonH*	acc_phon^[L*]H	-5.39	-8.55	-2.23	1.61	-3.35	577.54	< .001
intercept	h_startno H start	0.97	0.54	1.41	0.22	4.37	416.22	< .001
intercept	fin_phonL%	-1.08	-1.60	-0.56	0.26	-4.08	408.69	< .001
intercept	genderM	2.39	1.45	3.32	0.42	5.69	9.92	< .001

Appendix M. LMEMs of Peak Alignment as a Proportion of the Foot or Voicing

M1. PN Peak timing proportional to foot

Table M1.1 Model Printout

```
Formula:
`h_t as ratio of foot` ~ foot_syls + (1 | speaker) + (1 | gender) + (1 |
ana_syls) + (1 | pn_str_syl) + (1 | wrd_end_syl)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: pn_h_t_re_ft_dur_mdl
   Data: pn_h_data_lh %>% filter(abs(scale(resid(pn_h_t_re_ft_dur_mdl))) <=
    2.5)
Control:
lmerControl(optimizer = optimizer, calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: -1077

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9940 -0.6212 -0.1071  0.4719  3.0756

Random effects:
 Groups      Name        Variance  Std.Dev.
 speaker     (Intercept) 0.0008662 0.02943
 pn_str_syl  (Intercept) 0.0093642 0.09677
 ana_syls    (Intercept) 0.0038695 0.06220
 wrd_end_syl (Intercept) 0.0038242 0.06184
 gender      (Intercept) 0.0085106 0.09225
 Residual                0.0049165 0.07012
Number of obs: 472, groups:
speaker, 11; pn_str_syl, 8; ana_syls, 4; wrd_end_syl, 3; gender, 2

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)   0.83873    0.09041   3.25231   9.277   0.0019 **
foot_syls2   -0.03377    0.02087 441.01904  -1.618   0.1063
foot_syls3   -0.14539    0.02007 444.16038  -7.246 1.91e-12 ***
foot_syls4   -0.26288    0.02460 430.27360 -10.688  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
           (Intr) ft_sy2 ft_sy3
foot_syls2 -0.149
foot_syls3 -0.152  0.595
foot_syls4 -0.142  0.487  0.807
```

Table M1.2 ANOVA of model: h_t as ratio of foot ~ foot_syls + (1 | speaker) + (1 | gender) + (1 | ana_syls) + (1 | pn_str_syl) + (1 | wrd_end_syl)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	0.65	0.22	3	435.05	43.92	< .001	< .001	p < .05

Table M1.3 Omega2 (partial) for h_t as ratio of foot.

Parameter	ω2p	95% CI
foot_syls	.23	[.16, .29]

Table M1.4 Conditional and marginal R2

marginal R2	conditional R2
.22	.88

Table M1.5 Predicted values of h_t as ratio of foot

foot_syls	predicted	conf.low	conf.high	std.error
1	0.84	0.66	1.02	0.09
2	0.8	0.63	0.98	0.09
3	0.69	0.52	0.87	0.09
4	0.58	0.4	0.75	0.09

Table M1.6 b1: h_t as ratio of foot ~ foot_syls + (1 | speaker) + (1 | gender) + (1 | ana_syls) + (1 | pn_str_syl) + (1 | wrd_end_syl)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	-0.03	-0.07	0.01	0.02	-1.62	441.02	.106
foot_syls1	foot_syls3	-0.14	-0.18	-0.11	0.02	-7.25	444.16	< .001
foot_syls1	foot_syls4	-0.26	-0.31	-0.22	0.02	-10.69	430.27	< .001
foot_syls2	foot_syls3	-0.11	-0.15	-0.07	0.02	-6.06	421.86	< .001
foot_syls2	foot_syls4	-0.23	-0.28	-0.18	0.02	-9.86	405.55	< .001
foot_syls3	foot_syls4	-0.12	-0.15	-0.09	0.01	-8.1	452.18	< .001

M2. Nuclear `H_t` as proportion of foot

Table M2.1 Model Printout

```
Formula:
`h_t as ratio of foot` ~ foot_syls + (1 | speaker) + (1 | gender) + (1 |
pre_syls) + (1 | nuc_new_word) + (1 | nuc_str_syl) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: h_t_re_ft_dur_equation
   Data: nuc_data
Control:
lmerControl(optimizer = optimizer, calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: -2094.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.8713 -0.6488 -0.0029  0.6316  3.3229

Random effects:
 Groups       Name        Variance  Std.Dev.
 speaker      (Intercept) 2.676e-03 0.051727
 pre_syls     (Intercept) 8.928e-04 0.029880
 nuc_str_syl  (Intercept) 1.830e-03 0.042781
 fin_phon     (Intercept) 1.962e-03 0.044294
 nuc_new_word (Intercept) 1.341e-05 0.003662
 gender       (Intercept) 9.273e-04 0.030451
 Residual                 3.648e-03 0.060399
Number of obs: 788, groups:
speaker, 11; pre_syls, 4; nuc_str_syl, 3; fin_phon, 2; nuc_new_word, 2; gender, 2

Fixed effects:
              Estimate Std. Error         df t value Pr(>|t|)
(Intercept)   0.841044   0.051299   3.868275  16.395 0.000102 ***
foot_syls2   -0.140159   0.010685 369.158998 -13.118  < 2e-16 ***
foot_syls3   -0.120633   0.008744 768.070838 -13.797  < 2e-16 ***
foot_syls4    0.008050   0.011848   6.948712   0.679 0.518838
```

Table M2.2 Table A1.1 ANOVA of model: h_t as ratio of foot ~ foot_syls + (1 | speaker) + (1 | gender) + (1 | pre_syls) + (1 | nuc_new_word) + (1 | nuc_str_syl) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	1.28	0.43	3	19.06	117.38	< .001	< .001	p < .05

Table M2.3 Omega2 (partial) for h_t as ratio of foot.

Parameter	ω2p	95% CI
foot_syls	.94	[.87, .96]

Table M2.4 Conditional and marginal R2

marginal R2	conditional R2
.21	.76

Table M2.5 Predicted values of h_t as ratio of foot

foot_syls	predicted	conf.low	conf.high	std.error
1	0.84	0.74	0.94	0.05
2	0.7	0.6	0.8	0.05
3	0.72	0.62	0.82	0.05
4	0.85	0.75	0.95	0.05

Table M2.6 b1: h_t as ratio of foot ~ foot_syls + (1 | speaker) + (1 | gender) + (1 | pre_syls) + (1 | nuc_new_word) + (1 | nuc_str_syl) + (1 | fin_phon)

intercept	slope	estimate	conf.low	conf.high	std.error	t	df	p
foot_syls1	foot_syls2	-0.14	-0.16	-0.12	0.01	-13.12	369.16	< .001
foot_syls1	foot_syls3	-0.12	-0.14	-0.10	0.01	-13.8	768.07	< .001
foot_syls1	foot_syls4	0.01	-0.02	0.04	0.01	0.68	6.95	.519
foot_syls2	foot_syls3	0.02	0.00	0.04	0.01	1.72	368.82	.086
foot_syls2	foot_syls4	0.15	0.12	0.18	0.01	12.23	9.14	< .001
foot_syls3	foot_syls4	0.13	0.10	0.16	0.01	10.34	8.47	< .001

M3. Nuclear `h_t` as ratio of voicing

Table M3.1 Model Printout

```
Formula:
`h_t as ratio of voicing` ~ foot_syls + (1 | speaker) + (1 | gender) + (1 |
pre_syls) + (1 | nuc_new_word) + (1 | nuc_str_syl) + (1 | fin_phon)

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: nuc_h_t_re_v_dur_equation
   Data: nuc_nuc_h_t_re_v_dur.trimmed
Control:
lmerControl(optimizer = optimizer, calc.derivs = F, optCtrl = list(method =
"nlminb",
    starttests = F, kkt = F))

REML criterion at convergence: -2053.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3050 -0.6513 -0.0300  0.6474  3.0147

Random effects:
 Groups       Name        Variance  Std.Dev.
 speaker      (Intercept) 1.945e-03 0.044105
 pre_syls     (Intercept) 6.732e-05 0.008205
 nuc_str_syl  (Intercept) 1.817e-03 0.042631
 fin_phon     (Intercept) 6.676e-03 0.081706
 nuc_new_word (Intercept) 2.851e-04 0.016885
 gender       (Intercept) 2.471e-03 0.049705
 Residual                 3.765e-03 0.061357
Number of obs: 780, groups:
speaker, 11; pre_syls, 4; nuc_str_syl, 3; fin_phon, 2; nuc_new_word, 2; gender, 2

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  7.548e-01  7.509e-02  2.323e+00  10.051  0.00577 **
foot_syls2   2.398e-02  9.097e-03  1.649e+01   2.636  0.01766 *
foot_syls3  -7.502e-03  8.866e-03  7.624e+02  -0.846  0.39771
foot_syls4  -4.333e-04  1.349e-02  1.487e+02  -0.032  0.97442
```

Table M3.2 ANOVA of model: h_t as ratio of voicing ~ foot_syls + (1 | speaker) + (1 | gender) + (1 | pre_syls) + (1 | nuc_new_word) + (1 | nuc_str_syl) + (1 | fin_phon)

term	sumsq	meansq	NumDF	DenDF	F value	p.value	p.adj (BH)	signif.
foot_syls	0.05	0.02	3	15.37	4.45	.019	.034	p < .05

Table M3.3 Omega2 (partial) for h_t as ratio of voicing.

Parameter	ω2p	95% CI
foot_syls	.35	[0, .6]

Table M3.4 Conditional and marginal R2

marginal R2	conditional R2
.01	.78

Table M3.5 Predicted values of h_t as ratio of voicing

foot_syls	predicted	conf.low	conf.high	std.error
1	0.75	0.61	0.9	0.08
2	0.78	0.63	0.93	0.07
3	0.75	0.6	0.89	0.08
4	0.75	0.61	0.9	0.08

Table M3.6 b1: h_t as ratio of voicing ~ foot_syls + (1 | speaker) + (1 | gender) + (1 | pre_syls) + (1 | nuc_new_word) + (1 | nuc_str_syl) + (1 | fin_phon)

intercept	slope	estimate	conf.low	conf.high	std.error	t.value	df	p.value
foot_syls1	foot_syls2	0.02	0.00	0.04	0.01	2.64	16.49	.018
foot_syls1	foot_syls3	-0.01	-0.03	0.01	0.01	-0.85	762.4	.398
foot_syls1	foot_syls4	0.00	-0.03	0.03	0.01	-0.03	148.66	.974
foot_syls2	foot_syls3	-0.03	-0.05	-0.01	0.01	-3.26	15.94	.005
foot_syls2	foot_syls4	-0.02	-0.05	0.00	0.01	-2.02	19.42	.057
foot_syls3	foot_syls4	0.01	-0.02	0.04	0.01	0.51	126.41	.613

Appendix N. Speech Prosody 2020 Paper on K-Max (Rodgers 2020)

This paper is also available at: https://doi.org/10.21437/SpeechProsody.2020-46.

10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

K-Max: a tool for estimating, analysing, and evaluating tonal targets

Antoin Eoin Rodgers[1]

[1] Phonetics and Speech Laboratory, Trinity College, Dublin
rodgeran@tcd.ie

Abstract

This paper presents a novel approach to the identification of tonal targets within the Autosegmental Metrical (AM) framework using the second time derivative of the f0 contour. The approach is implemented through an interactive Praat script called K-Max, which allows users to annotate salient turning points on a text grid as well as correct tracking errors and remove microprosodic events on the pitch contour. The script also generates a resynthesized model of the pitch contour based on the annotation of turning points, which is not typical in the AM approach. The theoretical rationale for the overall approach is presented, followed by a description of its implementation. The paper then discusses the success of the technique in identifying tonal targets in relation to user intuitions and acceptability judgments regarding the resynthesized f_0 contours. Finally, it provides examples of its potential application regarding issues such as downstep and f_0 plateaux, arguing that the over-specification of tonal targets required in the resynthesis component can facilitate analysis of the relationship between underlying phonological structures and their realisation in the f_0 contour.

Index Terms: AM phonology, intonation, Praat, pitch, fundamental frequency (f_0), tonal targets, parameter extraction, contour modelling.

1. Introduction

1.1. AM and identification of tonal targets

In the Autosegmental Metrical approach (AM) to intonation [1], the fundamental frequency (f_0) contour is viewed as being, in part, the result of the phonetic implementation of a string of underlying low and high tonal primitives (L and H) which are independent of but associated with landmarks in the segmental string and metrical structure. These tonal primitives can form pitch accents, which can be monotonal (H*, L*) or bi-tonal (L*+H, H*+L, etc.), and occasionally tri-tonal. The starred tone indicates the tone associated with a stressed syllable in the metrical structure. Tones can also manifest as edge tones, such as boundary tones and phrase accents. While the tonal sequence is considered highly abstract [2], at the same time, however, much research has been conducted to ascertain how phonological tones are linked to or aligned with elements in the metrical structure and segmental string [3]. This involves precise measurement of the timing of L and H targets, typically in pitch accents, in relation to landmarks such as syllable onsets or the onset of vowels in stressed syllables. In other words, manifestations of these highly abstract primitives can still be identified and measured empirically. As such, tonal targets are identified in the f_0 contour in terms of turning points. These can be either f_0 maxima and minima, or elbows in the contour, the latter referring to points where there is a distinct shift in f_0 trajectory.

As implied in the use of the terms High and Low, f_0 maxima and minima are the archetypal turning points, and elbows their less ideal manifestations. f_0 maxima and minima are also easier to identify than that of elbows. Within the AM approach, one common means of estimating elbows without simply eyeballing the curve is via a technique which involves line fitting two lines inside a designated portion of the contour [4]–[6]. The intersection of the best fit lines is taken to indicate the timing of the elbow. An alternative means of measuring turning points includes using the extrema of the second time derivative of a smoothed f_0 contour, or $f_0''(t)$ [7].

1.2. K-Max: aims and guiding principles

K-Max [8] is a tool developed using Praat [9] scripts. It was developed based on the AM assumption that intonation can be described as a sequence of H and L targets. However, rather than viewing f_0 extrema as prototypical turning points, it takes the view that f_0 extrema just happen to be the most salient form of turning point. For this reason, as will be discussed in section 2.1 below, turning points are estimated using $f_0''(t)$.

K-Max is designed primarily to facilitate the identification and analysis of tonal targets in pitch accents. However, it is also designed for the analysis of phenomena which are potentially more problematic within the AM approach, most notably, f_0 plateaux and valleys. The difficulty with plateaux and valleys is that they have duration, and as such are not readily identifiable as tonal targets. AM approaches have tried to account for the apparent duration of tonal targets, such as via tonal spreading, in which one tonal target—typically a trailing tone—is seen to extend beyond the initial target time [10]. Different techniques have been used to quantify plateaux, such as [11], in which the plateau edge is measured heuristically in terms of a percentage fall from f_0 peak. Since K-Max provides a method of identifying and quantify turning points in a principled manner, not only for pitch accents, but also for plateaux and valleys wherever they appear salient, it can also help provide empirical data in the analysis of features such as tone spreading.

A further aim in developing K-Max was to permit analysis by resynthesis using turning points identified as salient by the analyst. In part, this stems from the view that empirical analysis of tonal targets should contribute towards the effective and efficient modelling of the pitch contour for (re)synthesis, a view not necessarily widely held within AM [2]. More importantly, this goal is based on the view that it is important to demonstrate that a contour can be modelled effectively using theoretical principles of the AM approach to help demonstrate its validity. It is believed that in modelling the contour more comprehensively, elements of phonetic implementation which might otherwise be over-looked will have to be accounted for.

 10.21437/SpeechProsody.2020-46

2. Quantifying tonal targets and slopes

2.1. Turning points and the second derivative of f_0

If we consider f_0 or the rate of vibration of the vocal folds in terms of (angular) velocity [12], its first derivative is the rate of change of velocity—i.e., acceleration—while the second derivative, $f_0''(t)$, is the rate of change of acceleration, also described as jerk. The intuition that jerk is salient can be understood by analogy: in term of motion, jerk is experienced as the sensation one gets in a car as it begins accelerates, while peak jerk occurs when a car breaks suddenly.

$F_0''(t)$ extrema align temporally with points of maximum curvature (concavity or convexity) in the f_0 contour. It should be noted that while $f_0''(t)$ is not a direct measure of curvature, for convenience, these time points of the extrema will be referred to as K_{max}. In the f_0 contour, these points are evident as turning points at f_0 extrema and elbows. The nature of this relationship between $f_0''(t)$ and turning points is readily apparent if we consider the linear stylised contour shown in Figure 1. $f_0''(t)$, shown in red, spikes at turning points in the f_0 contour (black), regardless of whether they are elbows or f_0 extrema. Negative spikes coincide with the most convex points—often f_0 maxima—and positive spikes with the most concave points—typically f_0 minima. Thus, both f_0 elbows and extrema are manifestations of the same phenomenon, and the polarity of $f_0''(t)$ indicates whether the turning point is more H-like (negative) or L-like (positive) at that time point. It is worth noting that in real-world cases f_0 extrema often occur near rather than at K_{max}, so f_0 maxima and minima may even sometimes be viewed as symptoms of or epiphenomena around K_{max}.

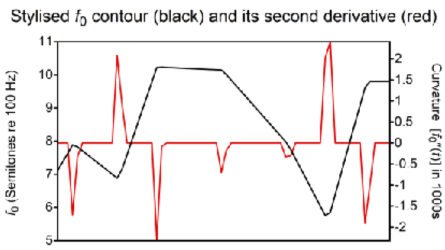


Figure 1: *Stylised f_0 contour and its second derivative.*

2.2. Tonal targets, physiological constraints, and planning

In order to distinguish tonal targets as described in the PENTA model from those of the AM approach, Xu describes articulatory targets in the PENTA model as *covert* as opposed to the *overt* tonal targets of the AM approach [13]. Overt targets, on this view, are measured in terms of literal f_0 maxima and minima as realised in the contour. Covert targets, on the other hand, are underlying articulatory targets which are not realised literally in the f_0 contour. This is because physiological constraints on the larynx limit the speed at which pitch trajectories can change [14], resulting in phenomena such as tonal undershoot which occurs as speakers must adjust their f_0 trajectory over time [15].

The argument that tonal targets should not be viewed as coequal with surface manifestations of f_0 is compelling. However, it does not seem incompatible with the AM approach. While it has indeed been that case that AM analyses of tonal targets have measured tonal alignment and scaling in terms of literal targets, such as in [16]–[18] *inter alia*, this may sometimes be more a case of methodological convenience than anything else. In fact, within AM literature, the f_0 contour itself is viewed as the result of the phonetic implementation of a phonological surface representation [19]. As such, it is reasonable to argue that adjustments made to f_0 during phonetic implementation as a result of physiological constraints are responsible for differences between phonological surface representation and its realisation in the f_0 contour. Thus, within the AM approach, one might well expect a mismatch between *literal* targets in the acoustic signal and the *ideal* targets of the phonological surface representation.

2.3. Trajectories and Inflexion points

One potential strategy then is to estimate the ideal trajectory of the f_0 contour towards the ideal tone minus the effects of physiological constraints. Bearing in mind that $f_0''(t)$ minima and maxima indicate points of maximum convexity and concavity respectively in the f_0 contour, the points of zero curvature between these two points—mathematical inflexion points identifiable using the roots of the second derivative—represent the times at which the f_0 contour appears to be under least pressure to change trajectory. Thus, inflexion points can be taken to represent moments where the contour is least affected by physiological constraints and is most 'on course' towards the *ideal* target. Consequently, the linear slope (or tangent) at the inflexion point can be viewed as an *ideal* slope towards an *ideal* target.

This is exemplified in Figure 2, which shows f_0 and $f_0''(t)$ contours of a model curve. The blue lines indicate the tangents at inflexion points between times maximum curvature. The intersection of these two lines can be viewed as the unrealised *ideal* target which would be achieved if physiological constraints and segmental pressure did not cause the speaker to make adjustments to the f_0 trajectory. Using these tangents and their intersections, an *idealised* linear interpolation of the contour can be estimated.

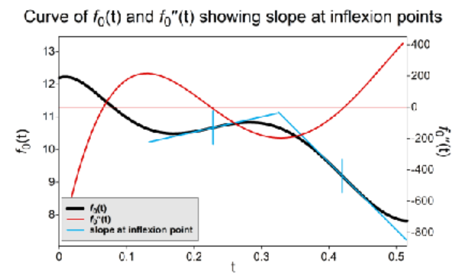


Figure 2: *Example of f₀ contour (black line), its second derivative (red line), and linear slopes projected from inflexion points (light blue line).*

3. Implementation

This section outlines the implementation of K-Max, in terms of how it manages f_0 contour correction, estimates turning points, permits user intervention, and generates an idealised and smoothed pitch contour (see Figure 3). It should be noted that all f_0 processing is carried out using semitones re 100 Hz.

A corrected f_0 contour is generated which is then smoothed and used in the rest of the analysis, in an approach similar to

226

that used by the IPO in generating stylized contours [20]. The corrected contour is produced via the procedure **@fixPitch**. This exploits the 'To Manipulation' function of Praat and provides the option for the user to correct the original contour by removing segmental effects not associated with intonation—such as those caused during voiced fricatives or at the onset of voicing after voiceless stops—and to correct pitch tracking errors such as pitch halving and pitch doubling. To facilitate the estimation of $f_0''(t)$, the corrected pitch contour is interpolated and smoothed using Praat's in-built functions.

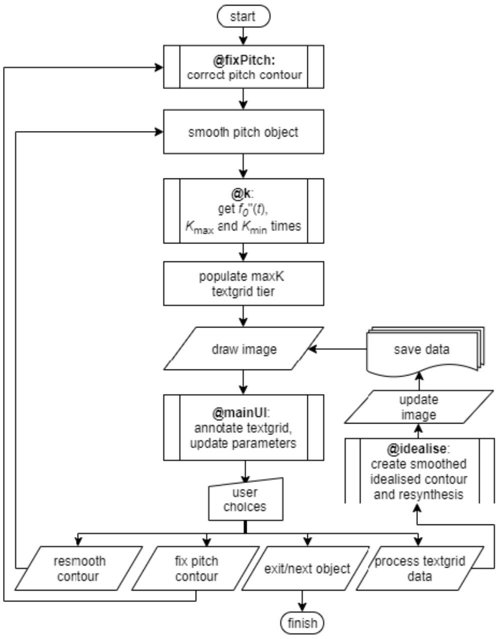


Figure 3*: Flowchart of the analysis, annotation, and processing for a single utterance.*

After this, the procedure **@k** is called, which estimates times of K_{max} and K_{min} using $f_0''(t)$. This includes near-roots, which are points where $f_0''(t)$ approaches but does not reach zero. Even after smoothing, the script will likely identify more K_{max} time points than there are salient tonal events. Conversely, as a result of over-smoothing, **@k** may occasionally identify too few turning points. User intervention during the procedure **@mainUI** helps deal with such problems.

To resolve the problem of multiple K_{max} time points, the user is prompted to annotate a tonal tier using only turning points which appear salient (see Figure 4). This includes the obligatory annotation of boundaries, but otherwise, the user is free to use any annotation convention they see fit here. If there are too few K_{max} time points, the user can adjust the Praat smoothing parameter in the user interface and then re-smooth the f_0 contour for processing using the new smoothing parameter.

During the user intervention stage, the picture window always displays the pitch contour and a single target tier. f_0 is indicated on the y-axis while shape size and colour intensity are used to indicate Cepstral Peak Prominence (see Figure 5). This helps distinguish more periodic components of the contour from less periodic ones (such as during voiced frication) and thus identify those parts of the contour which may be more relevant to intonation [21], [22]. The corrected contour can also be displayed in the picture window, so if there appear to be errors in the corrected contour, the user can re-run **@fixPitch**.

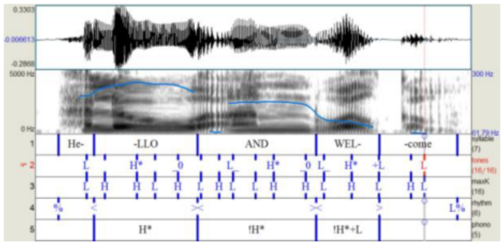


Figure 4: *Text grid showing estimated turning points ('MaxK' tier) and those marked salient ('tones' tiers).*

Once the salient turning points have been identified and annotated, the idealised contour is estimated by the procedure **@idealise**, which uses the principles set out in section 2.3 above to estimate *ideal* slopes and targets. This procedure also deals with several problems which may arise during estimation. First, there will be instances with more than one inflexion point between two turning points. Therefore, **@idealise** identifies the first and last roots of f_0'' between two extrema. It then performs a linear regression between these points on the f_0 contour to calculate the *ideal* f_0 slope between the two turning points. Conversely, there may be no inflexion points between turning points, such as occurs in plateaux (see Figure 1), in which case near-roots are used. If there is neither a root nor near-root, which can occur at boundaries, slope is calculated using the two edge-most frames. These slopes are used to identify *ideal* targets. If the procedure can still not adequately generate *ideal* targets, the user is warned to make adjustments.

Using *ideal* targets and slopes, an *idealised* f_0 contour is generated. Since this contour does not account for physiological constraints, it tends to sound like a much more exaggerated version of the original. At present, to simulate the effects of physiological constraints, a triangular moving point average (MPA) smoothing function is passed across the idealised contour. The width of the MPA size (in frames) can be changed manually in the main UI window in order to ensure that a reasonable approximation of the original contour is achieved.

Once the smoothed idealised f_0 contour has been generated, the utterance is resynthesized and the picture window is updated. Using visual and auditory judgments, the user can decide if the location of the *ideal* targets and the resynthesis are acceptable. Finally, the updated text grid is stored along with tables containing f_0 and time data to facilitate statistical analysis. This includes text grid annotation, time, and f_0 of turning points as well as the data regarding the ideal contour, i.e., ideal time points, f_0 values and slopes.

3.1. Testing and Effectiveness

A test set of 83 utterances was analysed and processed using K-Max by the author. There were 7-8 utterances from 11 speakers of northern Irish English (5M, 6F). The pitch accents of the utterances had previously been analysed by the author and another trained phonetician using IViE annotation conventions [23], but the original analyses were not consulted during this process. Since the quality of the resynthesis is dependent on the number of turning points specified by the user, the author attempted to mark only turning points which appeared salient as edge tones, pitch accents, and the edges of plateaux / valleys.

As K-Max involves a semi-iterative process which allows the user to test and retest output visually and auditorily and to adjust smoothing parameters on the fly, the results tended to be very satisfactory. That is, it was always possible to identify salient turning points from the 'maxK' text grid tier (see Figure 4) which agreed with the author's intuitions. Furthermore, the final resynthesized output tended to sound practically indistinguishable the original. In only two cases was it necessary to include a turning point which was not readily identifiable as belonging to the prescribed categories. In each case, these were in nuclear accents occurring in feet with four syllables. This may be a fault of the underlying approach or may reflect the weakness of using of linear interpolation between *ideal* targets.

So far only a small impressionistic test of the idealised and smoothed resynthesis has been conducted. 11 utterances, one each from each speaker, were selected randomly. Each original and resynthesized version was played to four colleagues at the Speech and Phonetics Laboratory. Two listeners judged all acceptable, while two judged one each to have boundaries which were slightly different from the original (for example, see Figure 6). However, in these cases, they felt that this did not affect the overall interpretation of the contour.

4. Applications

As intended, the data provided by K-Max can be used in standard AM analysis of temporal alignment and scaling of tonal targets. However, further applications of K-Max stem from the fact that, to create a reasonable resynthesized pitch contour, it requires over-specification of turning points when compared with typical AM analysis. This consequence was intended by design and its benefit will be exemplified with two short illustrations.

Pierrehumbert's original analysis of downstep [10] argued that it was the obligatory result of a sequence of H and L tonal targets; however, Ladd critiqued this [24], observing that in her data, L targets were sometimes added *ad hoc* in order to justify such an analysis. Ladd felt the problem boiled down to the difficulty in "reconciling goals of phonetic specification and linguistic generalisation" [24, p. 725]. If one considers the contour in Figure 5, there is a clear sequence of down-stepped H* pitch, which can be represented in AM phonology as H* !H* !H*L L% (following IViE annotation conventions, [25]). Yet in order to be able to resynthesize the contour adequately, it is necessary to annotate turning points at the edges of the down-stepped plateaux. The left edge of each down-stepped plateau has been annotated as L_ to show the turning point is associated with the upcoming H* and is not the tail of the previous pitch accent. In essence, this surface L_H* sequence can be viewed as the phonetic implementation of an underlying !H* pitch accent without compromising the phonology. In fact, it provides a means by which one might be able to identify more precisely the mechanisms through which the underlying phonology is implemented, or, in other words, reconcile phonetic specification with linguistic generalisation.

As a second example, K-max can be used to quantify and analyse plateau boundaries. In Figure 5 and Figure 6, the right edges of the plateaux have been annotated with _0. Again, the underlying phonology is still evident (L*H L*H % in the latter case), but the inclusion of _0 turning points is needed for resynthesis. In each utterance, _0 occurs at or near the right edge of the stress-containing word. If nothing else, this hints at a need for a more detailed analysis of the alignment of plateau edges and their potential role in signalling lexical boundaries.

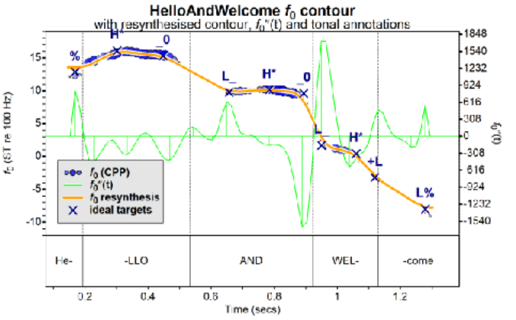


Figure 5: *Example of output from the picture window showing pitch contours and tonal targets.*

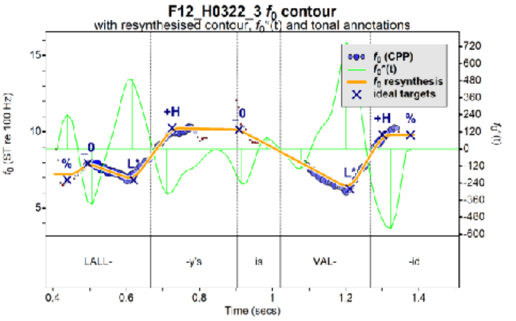


Figure 6: *Original and smoothed idealized contour of an utterance from a northern Irish English speaker*

In short, the inclusion of a resynthesis component forces the user to identify a minimal number of turning points which are required for contour realisation. This, in turn, encourages the user to consider the role of such turning points and their relationship to the underlying phonological forms.

5. Conclusions

This paper has outlined the rationale behind an AM-based intonation analysis tool, K-Max. It identifies turning points using $f_0''(t)$ and resynthesizes the contour using *ideal* tonal targets and slopes along with a simulation of physiological constraints on f_0 variation. It is argued that the inclusion of a synthesis component, while not typical of an AM approach, helps draw attention to and facilitate the analysis of intonationally significant components of the f_0 contour which are under-analysed in the AM approach.

Refinements will be made to the physiological constraints function and further objective and subjective tests of the resynthesis will also be conducted. Moreover, K-Max will be used to analyse larger corpora, which will provide more quantitative data regarding its effectiveness, and hopefully demonstrate its usefulness.

6. References

[1] D. R. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 2008.

[2] A. Arvaniti, "Autosegmental-Metrical Model of Intonational Phonology," in *Prosodic Theory and Practice*, S. Shattuck-Hufnagel and J. Barnes, Eds. Cambridge MA: MIT Press, to appear.

[3] P. Prieto, "Tonal Alignment," in *The Blackwell Companion to Phonology*, M. von Oostendorp, C. J. Ewen, E. Hume, and K. Rice, Eds. John Wiley & Sons, 2011, pp. 1–19.

[4] M. D'Imperio, "The Role of Perception in Defining Tonal Targets and Their Alignment," Ohio State University, 2000.

[5] M. E. Beckman and J. B. Pierrehumbert, "Intonation structure in Japanese and English," *Phonology*, vol. 3, no. 01. pp. 255–309, 1986.

[6] P. Welby, "The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation," Ohio State University, 2003.

[7] B. Ahn, N. Veilleux, and S. Shattuck-hufnagel, "Annotating Prosody With Polar : Conventions for a Decompositional Annotation System," in *Proceedings of ICPhS 2019*, 2019, pp. 1302–1306.

[8] A. E. Rodgers, "K-Max," 2019. [Online]. Available: https://github.com/AERodgers/Praat-K-Max.

[9] D. Boersma, Paul & Weenink, "Praat: doing phonetics by computer (v. 6.1.03)." 2019.

[10] J. B. Pierrehumbert, "The Phonology and Phonetics of English Intonation," MIT, Cambridge MA, 1980.

[11] R. A. Knight and F. Nolan, "The effect of pitch span on intonational plateaux," *J. Int. Phon. Assoc.*, vol. 36, no. 2004, pp. 21–38, 2006.

[12] W. J. Hardcastle, *Physiology of Speech Production*. London: Academic Press, 1976.

[13] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *J. Phon.*, vol. 33, no. 2, pp. 159–197, 2005.

[14] Y. Xu and X. Sun, "Maximum speed of pitch change and how it may relate to speech.," *J. Acoust. Soc. Am.*, vol. 111, no. 3, pp. 1399–1413, 2002.

[15] Y. Xu, "Speech melody as articulatorily implemented communicative functions," *Speech Commun.*, vol. 46, no. 3–4, pp. 220–251, 2005.

[16] D. R. Ladd, I. Mennen, and A. Schepman, "Phonological conditioning of peak alignment in rising pitch accents in Dutch.," *J. Acoust. Soc. Am.*, vol. 107, no. 5, pp. 2685–2696, 2000.

[17] J. N. Sullivan, "Variability of F0 Valleys: The Case of Belfast English," in *CamLing 2007: Proceedings of the Fifth University of Cambridge Postgraduate Conference in Language Research*, no. L, N. Hilton, R. Arscott, K. Barden, A. Krishna, S. Shah, and M. Zellers, Eds. Cambridge: Cambridge Institute of Language Research, 2007, pp. 245–252.

[18] P. Prieto, J. van Santen, and J. Hirschberg, "Tonal alignment patterns in Spanish," *J. Phon.*, vol. 23, no. 4, pp. 429–451, 1995.

[19] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[20] J. 't Hart, R. Collier, and A. Cohen, *A perceptual study of intonation*. Cambridge: Cambridge University Press, 1990.

[21] A. Albert, F. Cangemi, and M. Grice, "Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration," in *Proceedings of the 9th International Conference on Speech Prosody*, 2018, pp. 804–808.

[22] F. Cangemi, A. Albert, and M. Grice, "Modelling intonation: Beyond segments and tonal targets," in *ICPhS 2019*, 2019, pp. 572–576.

[23] A. E. Rodgers, "The effects of anacrusis and foot size on prenuclear pitch accents in northern Irish English (Derry City)," in *ICPhS 2019*, 2019, pp. 1307–1311.

[24] D. R. Ladd, "Phonological Features of Intonational Peaks," *Language (Baltim).*, vol. 59, no. 4, pp. 721–759, 1983.

[25] E. Grabe, "The IViE Labelling Guide (Version 3)," 2001. [Online]. Available: http://www.phon.ox.ac.uk/files/apps/IViE/guide.html.