# PROSODIC PROMINENCE IN ENGLISH

BY

SUYEON IM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Jennifer Cole, Chair, Northwestern University
Professor José Ignacio Hualde
Associate Professor Chilin Shih
Associate Professor Ryan Shosted
Dr. Joseph Roy, American Association of University Professors

## ABSTRACT

In English, certain words are perceptually more salient than other neighboring words. The perceptual salience is signaled by acoustic cues. Prominent words are higher, longer, or louder than nonprominent words in English. Perceptual prominence is associated with meaning of a word in discourse context. Prominent words are usually new or contrastive information, while nonprominent words are given or noncontrastive information. This dissertation addresses English prominence in two separate studies. The first study investigates the prosodic prominence in relation to pitch accents, acoustic cues, and discourse meaning of a word in a public speech. The second study examines the cognitive representation of prosodic contour in a corpus of imitation.

Linguists claim that the information status of a word determines the types of pitch accents in English. Prior research informs us about prominence (1) in relation to the binary given-new distinction of lexical givenness, and (2) in minimally contextualized utterances such as question-answer prompts or excerpts from a corpus. The assignment of prominence, however, can vary in relation to referential meaning as well as lexical meaning of a word in natural, more contextualized speech. This study examines the prosodic prominence as a function of pitch accents, acoustic cues, and information status in a complete public speech. Information status is considered in relation to referential, lexical givenness and alternative-based contrastive focus. The results show that accent type is probabilistically associated with information status in this speech. The accent assignment differs between referentially vs. lexically given words. Despite the weak relationship between information status and pitch accents in the speech of the speaker, non-expert listeners perceive prominence as expected: they are more likely to perceive

prominence on words carrying new or contrastive information or words with high or bitonal pitch accents. Surprisingly, the listeners perceive acoustic cues differently depending on the information status or accent types of a word. Based on these results, the first study suggests that (1) the relationship between information status and accent type is not deterministic in English, (2) lexical givenness differs from referential givenness in production and perception of prominence, and (3) perceived prominence is influenced by information status, pitch accents, acoustic cues, and their interaction.

The second study examines how an intonational contour is represented in the mental lexicon of English speakers. Some linguists find that speakers are able to reproduce the phonetic details of intonational features, while in other research speakers are better at reproducing intonational features than imitating phonetic details of an utterance. This study investigates the domain of intonational encoding by comparing several prosodic domains in imitated utterances. I hypothesize that the domain which best captures the similarity of intonational contour between the model speaker and imitators is the target of imitation, and that imitation can be considered as the domain of intonational encoding in cognitive representation. The results show that the f0 distance between the model speaker and imitators is best explained over an intermediate phrase. Based on these results, the second study proposes that speakers encode a time-varying f0 contour over a prosodic phrase in their mental lexicon and supports the exemplar encoding of intonational contour.

## ACKNOWLEDGEMENTS

I was able to complete my PhD program thanks to advices and support from many people around me. Among those, I would like to express my deepest gratitude to my academic advisor, Jennifer Cole; my committee members, José I. Hualde, Joseph Roy, Chilin Shih, and Ryan Shosted; the head of department, James Yoon; the director of graduate studies, Tania Ionin; and my former academic advisor at Seoul National University, Hyunkee Ahn. I would like to also thank my mom, dad, and Wonjoon for endless love and support.

## TABLE OF CONTENTS

# 1. INTRODUCTION

When we listen to an utterance, some words have greater perceptual salience than other neighboring words. Linguists claim that prominent words are associated with "newsworthy" information, such as new information or contrastive focus, while non-prominent words are not. However, another line of research on information status proposes a more complex distinction than the binary lexical given-new distinction of information status. Some research suggests a three-way given-accessible-new distinction or, in other works, a referential vs. lexical distinction. More complex distinctions in information status call for a reexamination of information status's relationship with pitch accents in English. Moreover, prior research has investigated the relationship between information status and pitch accents in minimalized discourse contexts such as question-and-answer prompts and excerpts from corpora, which might not fully capture the information status of a word built on prior context. The first study of this dissertation addresses both issues by examining the production and perception of prosodic prominence in relation to three-way given-accessible-new distinctions of referential givenness, lexical givenness, and contrastive focus in a complete public speech.

The second study investigates the representation of phrasal pitch patterns in our mental lexicon. Linguistic analyses of phrasal intonation posit phonological representations consisting of a few tonal targets such as high, low, and bitonal pitch accents, that define f0 targets, with interpolated f0 transitions between the tonal targets. Pitch accents as the underlying targets of perceived or produced intonational contours are hypothesized as the units that are encoded in our mental lexicon. Another line of linguistic research presents evidence of exemplar encoding of

heard speech, including speaker-indexical and other nonlinguistic information. The exemplar model predicts that intonational contours are specified in full phonetic detail in our cognitive representation. We may ask, then, about the empirical evidence for the encoding of intonational detail. Is all perceived acoustic detail equally likely to be encoded? Or are some aspects of phonetic detail—for example, cues to contrastive features—more likely to be encoded than others? The second study of this dissertation addresses these questions by comparing models of phrasal intonation in terms of different prosodic domains, modeled by time-series analyses in a corpus of imitated speech.

This chapter reviews the literature for both studies. For the first study, sections 1.1-1.5 outline prosodic prominence, the relationship between prominence and information status, the model of prominence perception, the RefLex Scheme, and public speech style. For the second study, sections 1.6-1.8 review the encoding of pitch contours, evidence of abstract vs. exemplar encoding, and prosodic domains for f0 used in prior research.

## 1.1. Prosodic Prominence

In Autosegmental-Metrical theory (AM; Liberman 1975; Pierrehumbert, 1980), prosody is characterized in terms of prominence and boundaries. Boundaries indicate the edge of a prosodic constituent such as a word or a phrase. Prominence is assigned to the head of a prosodic constituent at a designated level. In this framework, a phonological representation is built up hierarchically and prominence is assigned to an element in relation to the surrounding elements. Words that have phrase-level prominence are the eligible landing position of pitch accents. Pitch accents are discrete units of phonological form characterized by changes in pitch. They are specified by High (H*) or Low (L*) tones, or bitonal combinations. Eight types of pitch accents

are observed in Mainstream American English: H*, L*, !H*, L+H*, L*+H, L+!H*, L*+!H, H+!H* (Veilleux, Shattuck-Hufnagel, & Brugos, 2006).

Accent types can be ranked in relation to prominence. Hualde et al. (2016) compare the accent types labeled by trained annotators using a ToBI annotation convention (Veilleux et al., 2006) with the prominence scores obtained from non-expert listeners, for a sample of spontaneous speech. The results show that the nuclear pitch accents (the rightmost pitch accent in a prosodic phrase) are more likely to be perceived as salient than prenuclear pitch accents (the pitch accents preceding the nuclear pitch accent in the same prosodic phrase), which in turn are more likely to be perceived as salient than unaccented words. Also, bitonal pitch accents are more likely to be perceived as salient than monotonal accents. The L+H* pitch accent is the most perceptually prominent accent, as the accent type is associated with narrow or contrastive focus. H+!H*, L*, H*, !H* are lower in the ranking of perceptual prominence, in decreasing order.

Based on Hualde et al.'s (2016) study, the accent types can be ranked in terms of perceived prominence as in Table 1.1.

Table 1.1.

*Pitch Accent Hierarchy.*

| Least Prominent ← | | | | → Most Prominent |
|---|---|---|---|---|
| L* | !H* | H* | H+!H* | L+H* |

In Table 1.1, the least perceptually prominent accent type L* is located on the left of the prominence continuum and the most perceptually prominent accent type L+H* is presented on

the right.[1] I call this rank of accent types in relation to perceived prominence the "pitch accent hierarchy" henceforth.

Prominence is expressed through several phonetic properties. In English, prominent words are associated with longer duration, greater intensity, steeper spectral slope, and hyper-articulation. F0 is also often included as a correlate, most notably in the analysis of pitch accent as a phonological feature encoding prominence, but experimental evidence for the relationship between prominence and f0 is not clear (Beckmann, 1986; Breen, Fedorenko, Wagner, & Gibson, 2010; Cole, Kim, Choi, & Hasegawa-Johnson, 2007; Cole, Mo, & Hasegawa-Johnson, 2010; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Heldner, 2003; Kochanski, Grabe, Coleman, & Rosner, 2005; Ladd, 2008; Silipo & Greenberg, 1999; 2000; Sluijter & van Heuven 1996; Turk & White, 1999). Breen et al. (2010) examine more than 20 acoustic measures associated with different categories of information structure in English. They find that greater intensity, longer duration, and mean and maximum f0 are reliable correlates of focus. Kochanski et al. (2005) examine five acoustic correlates of prominent syllables in a spontaneous speech corpus in British and Irish English. The results show that intensity and duration are stronger cues than f0 to predict prominent syllables in the corpus. Watson, Arnold & Tanenhaus (2008) investigate the acoustic correlates of prominent words in relation to attributes of the importance and predictability of a word relative to the discourse goal, in task-oriented speech. They find that

[1] The positioning of L* as the least prominent pitch accent is in line with analyses that claim L* as the pitch accent used for words that are explicitly given in, or highly accessible from, the discourse context. On the other hand, L* is the accent that is typically used in polar (yes/no) questions, marking the start of the phrase-final pitch rise. In this context, L* may have greater perceptual prominence. In this regard it is notable that Hualde et al.'s (2016) prominence rating study finds that words with the L* pitch accent are more frequently rated as prominent than words with a H* pitch accent. I leave this as an open question here.

words that deliver important information are produced with greater intensity while words that are unpredictable are produced with higher f0 and longer duration. Watson (2010) emphasizes that prominence arises from multiple source such as information status, predictability, and importance. The acoustic correlates of prominence can differ depending on the source of prominence. In addition, they can vary depending on contextual factors such as speakers and speech style. For a review of prosody, see Cole (2015) and Wagner and Watson (2010).

### 1.2. Prominence and Information Structure

Pitch accents encode discourse meaning. Halliday (1970) relates prominence (he uses the term "tonic") to information structure in British English. According to his observation, information is delivered over a unit called a "tone group." The tone group is a phonological domain defined by the speaker, which often coincides with the syntactic clause. If any word in a tone group is marked by the tonic, this tone group is new information. If not, this tone group is considered as given information. Bolinger (1958) shows that different intonational forms (pitch accents) are associated with different meanings. In this sense, pitch accents are morphemes as they are associated with certain meanings. This is different from segments, as any segment (e.g., /p/) does not have intrinsic meaning. Pierrehumbert and Hirschberg (1990) argue for a one-to-one mapping between pitch accents and the pragmatic meaning of a word. The item made salient by H* conveys new information. The L* accent is used when a speaker attempts to render an item salient but does not wish to include the item in his predication. The L* accent commonly occurs in canonical yes/no questions (e.g., "Do PRUNES have FEET?" where accented words are indicated in capital letters). Speakers ask the hearer to confirm or reject the predication. The bitonal pitch accents (L+H*, L*+H, H+L*, H*+L) invoke scaled interpretation in hearers'

beliefs. Among the bitonal pitch accents, the L+H* accent is used to convey that the accented items, not the alternative items, should be believed by hearers. The L+H* accent is known as a corrective or contrastive accent.

Empirical research (Ito & Speer, 2008; Terken & Nooteboom, 1987) shows supporting evidence of the link between pitch accents and information status. Terken and Nooteboom (1987) hypothesize that accented words are verified faster regardless of information status, as accenting can draw listeners' attention to acoustic properties of a word and facilitate processing of the word. The results show that new items are verified faster if the items are accented. Surprisingly, given items are verified faster only if the items are unaccented. They suggest that the absence of accent guides listeners to search the referent of a word in prior context, instead of focusing on the acoustic properties of a word. Ito and Speer (2008) show that the felicitous assignment of L+H* helps listeners' visual identification of a target. Listeners are quicker to find a target item on a computer screen in a felicitous condition where L+H* is used to make an item contrastive with the preceding item (e.g., "First, hang the green drum"; "Next, hang the BLUE drum") than in an infelicitous condition where L+H* is not used to make an item contrastive with another (e.g., "First, hang the green drum"; "Next, hang the blue DRUM"). This suggests that felicitous assignment of contrastive focus helps listeners to process discourse meaning of a word.

Another line of research challenges the assumption of the binary given-new distinction of information status in the research above and proposes more complex distinctions to capture relative degrees of information status (Baumann & Grice, 2006; Baumann & Riester, 2013; Calhoun, 2010; Chafe, 1976; Clark, 1975; Prince, 1981; Vieira & Poesio, 2000). Prince (1981)

claims that information status is gradient, and the dichotomous given-new distinction is not sufficient to capture the relative nature of information status. She proposes a three-way evoked-inferable-new distinction, and further develops subcategories of information status. Evoked items are the items already mentioned or situationally salient in discourse context. Inferable items are inferable from the previous items in the discourse context. New items can be either brand new (assumed to be unknown to hearer) or unused (assumed to be known to hearer but not to be in hearer's consciousness). Chafe (1976) also categorizes information status into three given-accessible-new groups based on activation cost. The activation cost is the speakers' cognitive load to activate an idea from its prior inactive state. The accessible category refers to words that are semi-active or accessible from prior discourse context. Baumann and Grice (2006) find empirical evidence of accessible information in German. The accessible category shows two patterns: the whole-part relation (e.g., book-page) or the scenario condition, where the referent is predictable, are conveyed by H+L*, H*, and unaccentedness in decreasing order. Part-whole, synonymy, and hypernym-hyponym are delivered by unaccentededness, H+L*, and H* in decreasing order. Based on these results, the authors claim that the H+L* accent conveys accessible information in German. In addition to the three-way given-accessible-new distinction, Baumann and Riester (2013) propose further distinctions between referential and lexical givenness. Referential givenness denotes the coreferential status of a word with an antecedent in discourse context. Lexical givenness refers to the repetition of the same word or a similar word in prior context. Baumann and Riester (2013) show that unaccentuation can arise from either coreference as in (1), or lexical repetition as in (2).

(1) A: Did you see Dr. Cremer to get your root canal?
B: Don't remind me. I'd like to STRANGLE the butcher.

(2) On my way home, a dog barked at me. It made me think of ANNA'S dog.

In sentence (1), "the butcher" (underlined) is unaccented as it corefers with "Dr. Cremer" in the preceding sentence. In sentence (2), "dog" is unaccented because it is lexically repeated, even though it does not corefer with the previous mention of "dog" in the context. Baumann and Riester (2013) present the RefLex scheme, developed to annotate the referential and lexical information status of a word at separate levels. Their analyses of German read and spontaneous speech show that accent patterns are aligned with lexical information status (the lexically given, accessible, and new labels) and that, within each lexical information status, the accent patterns are aligned with referential information status (the referentially given, accessible, and new labels). Although all the accent types (H*, !H*, H+L*, L*) are found across all the referential and lexical labels, the most frequent pitch accent is H* for the referentially and lexically new label, H+L* for the referentially and lexically accessible label, and no pitch accents for the referentially and lexically given label.

The research on the complex distinctions in information status prompts us to consider two following points: First, the information status of a word is relative to that of other words in discourse context, and information status categories can be ranked along a continuum of cognitive status. Gundel, Hedberg, and Zacharski (1993) propose the givenness hierarchy, which presents referring expressions in order based on assumed cognitive status. This hierarchy is adopted by Baumann and Riester (2012) and modified in their RefLex scheme. Table 1.2 is a simplified version of Baumann and Riester's givenness hierarchy.

Table 1.2.

*Givenness Hierarchy (modified from Baumann & Riester, 2012).*

| Activated | Uniquely Identifiable (but not activated) | Referential (but not uniquely identifiable) |
|---|---|---|
| Given | | |
| Bridging | | |
| | Unused | |
| | | New |

In Table 1.2, "given" information is the most cognitively activated information while "new" information is the least cognitively activated information. The "bridging" and "unused" categories are situated between given and new information. The givenness hierarchy (the information status hierarchy, henceforth) is crucial for understanding the relationship between information status and prominence, as the information status hierarchy can be directly compared to the pitch accent hierarchy in Table 1.1.

Table 1.3 shows the hypothetical relationship between the pitch accent hierarchy and the information status hierarchy.

Table 1.3.

*Hypothetical Relationship between Pitch Accents and Information Status.*

| Pitch Accent Hierarchy | L* | !H* | H* | L+H* |
|---|---|---|---|---|
| | ↕ | ↕ | ↕ | ↕ |
| Information Status Hierarchy | Given | Bridging | Unused | New |

The least perceptually salient pitch accent L* is linked to the least cognitively activated information status (given). The most perceptually salient pitch accent L+H* is matched to the most cognitively activated information status (new). Likewise, the other two pitch accents !H*, H* are linked to the other two information statuses (bridging, unused). The direct comparison of

the two hierarchies allows us to examine the one-to-one correspondence between gradient information status and pitch accents. If pitch accents are "morphemes," they must be exclusively associated with a certain information status despite the more complex distinctions of information status.

Second, there is emerging evidence that challenges the one-to-one mapping between prominence and information structure (Baumann & Grice, 2006; Baumann & Riester, 2013; Calhoun, 2010; Cangemi & Grice, 2016; Dahan, Tanenhaus, & Chambers, 2002; Féry & Samek-Lodovici, 2004; Hirschberg, 1993; Krahmer & Swerts, 2001; Riester & Piontek, 2015; Terken & Hirschberg, 1994). Baumann and Grice (2006) show that the H+L* accent conveys accessible information in German but that this pitch accent is not the only accent marking accessible information. Words designated as accessible information are accented by H* or may even be unaccented, although these two cases are less common than H+L*. Similar results are also found in Baumann and Riester (2013). All the pitch accents are found across all the referential and lexical information status labels although certain pitch accents are more frequent than others for certain information status categories. The H* accent is the most frequent pitch accent associated with the referentially and lexically new label, the H+L* accent is most frequent with the accessible label, and words with the given label are most frequently unaccented. In addition to this observation in German, Cangemi and Grice (2016) find the alignment of f0 peaks varies in interrogative sentences in Neapolitan Italian. In this variety of Italian, the f0 peak occurs later in interrogatives than declaratives. However, the results show that the f0 peak occurs even earlier in interrogatives than declaratives. The authors suggest that intonational tunes for interrogative sentences vary in Neapolitan Italian, refuting the one-to-one relation between form and meaning.

The evidence that challenges the one-to-one relation between pitch accents and information status leads us to raise the question of how pitch accents and information status influence listeners' perception of prominence. Pitch accents and acoustic cues are signal-driven factors, while information status is a meaning-driven factor. If there is no strong relationship between signal-driven factors and meaning-driven factors, how do these factors influence the perception of prominence? Do signal-driven and meaning-driven factors influence perceived prominence independently or do they interact? In the next section, I review the model of prominence perception.

### 1.3. Perception Model of Prosodic Prominence

Perception of prominence is a comprehensive process incorporating two different types of processes. The expectation-driven process is the one where listeners expect to hear prominence based on their prior linguistic knowledge (syntactic, semantic, pragmatic, or lexical knowledge). The signal-driven process is the one where listeners perceive prominence based on their online processing of phonological features or phonetic cues delivered by a speaker. If listeners recognize that the words carrying given information tend to be unaccented (built-up knowledge), they expect to hear no accent on such words (expectation-driven process) whether the words are acoustically salient or not (signal-driven process). In this sense, these two types of processes are independent. However, as listeners incorporate signal information rapidly when they hear a word, these two processes may interact with one another in the comprehensive process of perceiving prominence.

There are empirical studies that investigate the effects of expectation-driven factors (the predictability of a word, repetition of a word, lexical frequency) or signal-driven factors

(acoustic cues, pitch accents) on perceived prominence (Aylett & Turk, 2004; Bard & Aylett, 1999; Breen et al., 2010; Cole et al., 2010; De Ruiter, 2015; Greenberg, 1999; Turnbull, 2017; Turnbull, Royer, Ito & Speer, 2017; Watson et al., 2008). Cole et al. (2010) examine how information factors (word frequency, lexical repetition) and acoustic measures influence the perception of prominence judged by non-linguistic expert listeners. The results show that information factors (expectation factors) and acoustic cues (signal-driven factors) independently contribute to perceived prominence. Words judged by listeners as prominent are longer in vowel duration than nonprominent words. Low-frequency words are often perceived as prominent and associated with spectral emphasis in the high-frequency region, which suggests the speaker's increased vocal effort producing such words. This leads to the conclusion that listeners perceive prominence (1) if they hear unexpected words (less frequent or less repeated words) in discourse contexts or (2) if they hear enhanced acoustic cues on a word. As the perception of prominence involves speakers' signal information and listeners' built-up expectations, the authors suggest that prominence is speaker-based (signal-driven) and listener-based (expectation-driven). Based on these results, they propose a model of perceived prominence in relation to expectation-driven and signal-driven factors shown in Figure 1.1.

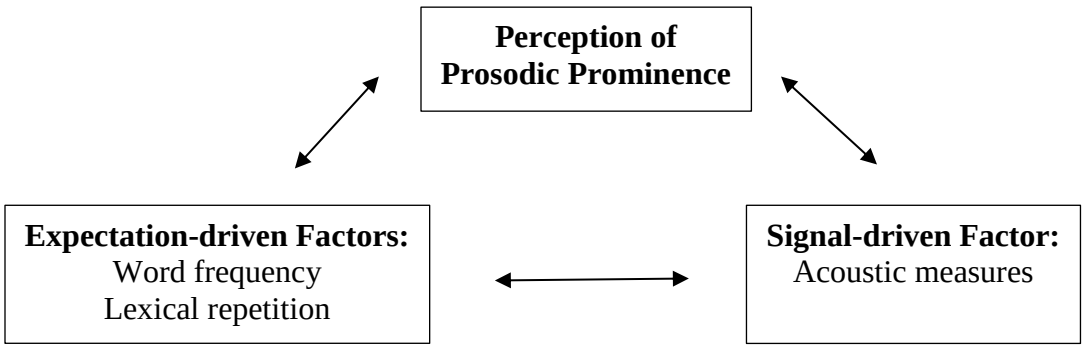


*Figure 1.1.* Perception model as a function of expectation-driven and signal-driven processes (adopted from Cole et al., 2010).

The first study of this dissertation adopts the perception model from Cole et al. (2010) and expands it in three regards: First, as expectation-driven factors, referential givenness, lexical givenness, and alternative-based contrastive focus (Rooth 1992) are examined using a simplified version of the RefLex scheme (Riester & Baumann, 2017). Most prior research above investigates the relationship between perceived prominence and lexical givenness only (the given-new distinction of a word, repetition of a word) and does not inform us how prominence interacts with other layers of information status (e.g., referentially given-accessible-new information status). To my knowledge, this is the first study which examines perceived prominence in relation to referential givenness, lexical givenness, and alternative-based contrastive focus together in English.

Second, as signal-driven factors, pitch accents and phonetic cues are considered. Most prior research above examines either pitch accents or phonetic cues and very few studies include both factors. Pitch accents and acoustic cues are highly correlated to each other, but they are not identical. Pitch accents are relative in nature. High or low tones are determined in relation to the pitch of neighboring tones. Acoustic cues can be measured in absolute value independent from neighboring sounds. By considering pitch accents and acoustic cues as signal-driven factors, I expect to find how these two factors are related to prominence and information status.

Third, expectation-driven and signal-driven factors are examined in a complete speech from TED talks. Most prior research above uses minimally contextualized utterances (e.g., question-and-answer prompts or experts from a corpus) and thus might not fully capture prominence and information status in more contextualized utterances. I come back to this point later in section 1.5.

In the next section, I review a simplified version of the RefLex scheme that I use to annotate information structure in the first study of this dissertation. This version is different from the original RefLex scheme (Riester & Baumann, 2017) in that it (1) uses basic labels such as given, bridging, unused, and new, and 2) includes alternative-based contrastive focus as a separate level in addition to the two original levels, one for the referential level and another for the lexical level. It is essential to be familiar with the basic labels of the scheme for the first study of this dissertation. For the complete version of this annotation scheme, see Riester and Baumann (2017).

### 1.4. The RefLex Scheme

There are several schemes for annotating the complex layers of information status (Calhoun, Nissim, Steedman, & Brenier, 2005; Dipper, Götze, & Skopeteas, 2007, Riester & Baumann, 2017). Calhoun et al. (2005) propose a framework for annotating information structure based on their analyses of Switchboard corpus. They consider a three-way old-mediated-new distinction of information status. In addition to this, they also distinguish rheme (background information) and theme (the information which links the utterance to the preceding context), and include *contrast* (contrastive focus) in their analyses of information structure. Similarly, Dipper et al. (2007) provide a guideline for annotating information structure in three layers based on their analyses of a multilingual corpus of various speech styles. The first layer considers given-accessible-new information status for referential expressions. The second and the third layers label topic and focus, respectively. Baumann and Riester (2017) present a slightly different annotation scheme of information status. Their scheme is also based on the three-way given-accessible-new distinction, but it differs from the previous schemes in that referential and

lexical givenness are considered at two separate levels, which is how this scheme is called the RefLex scheme. Consider the following example (Baumann & Riester, 2013):

(3) Smith was very optimistic. The polls showed a solid majority for the politician.

*referentially given*

*lexically new*

The underlined nominal expression "the politician" corefers with the preceding noun "Smith," so "the politician" is referentially given information. In comparison, "the politician" is lexically new information because it is a lexically different expression from "Smith." Empirical evidence shows that accent assignment is different between referential and lexical givenness in German (Baumann & Riester, 2013) and Russian (Luchkina, 2016). The RefLex scheme allows us to differentiate referential vs. lexical givenness in systematic manner in examining accent assignment.

Most prior research considers prominence in relation to lexical givenness only, for instance, in word frequency (Bell et al., 2003; Bybee, 2003; Cole et al., 2010; Wright, 2003), frequency of lexical repetition in given discourse context (Cole et al., 2010; Wright, 2003), or predictability of a word from surrounding words or preceding discourse context (Aylett & Turk, 2004; Bell et al., 2003; Ito & Speer, 2008; Watson et al., 2008). Although prior research yields insights on the relationship between prominence and lexical givenness, it does not inform us how prominence is related with other layers of information status. By adopting the RefLex scheme, the first study of this dissertation expands our understanding of the complex relationship between prominence and information structure in terms of referential givenness, lexical givenness, and contrastive focus in English.

The RefLex scheme has referential and lexical levels which consist of several labels in the given-new continuum. This section presents labels and examples of a simplified version of the RefLex scheme to help readers be familiar with the scheme. For the complete version of the RefLex scheme, see Baumann and Riester (2017).

### 1.4.1. The referential level

The referential level is to annotate the coreferential status of expressions. The referential labels are annotated on noun phrases. The basic referential labels are based on three-way given-accessible-new distinction with subdivided labels as shown in Table 1.4. Following the information status hierarchy, the most given label is presented on the top of the table and the newest label on the bottom. The examples are taken from Baumann and Riester (2013).

Table 1.4.
*Referential (r-) Labels of a Simplified Version of the RefLex Scheme.*

| Label | Description |
|---|---|
| R-given | Item coreferring with antecedent in discourse |
| R-generic | Generic item |
| R-bridging | Item accessible from prior item in discourse |
| R-unused | Item generally known |
| R-cataphor | Item whose referent is introduced later in discourse |
| R-new | New item not indefinable nor accessible from prior discourse |

The r-given label is to annotate an anaphor which is coreferential with an antecedent as in (4). The nominal expression "the car" is r-given as it is coreferential with the preceding noun "a car."

(4) A car was waiting in front of the hotel. I could see a woman in the car.

The r-generic label is assigned to generic terms as "a cat" in (5).

(5) A cat makes for a popular pet. Moreover, a cat is quite independent.

The r-bridging label is to annotate the item activated or accessible from the previous item. In (6), “the lock” is r-bridging as it is accessible from the conceptually relevant item “the door.”

(6) I tried to open the door but the lock was rusty.

The r-unused label is assigned to the entity whose referent is generally identifiable, such as “President Barack Obama” and “Tucson” in (7).

(7) President Barack Obama delivered a brilliant speech in Tucson.

The r-cataphor label is to annotate an item whose referent is identified in the upcoming discourse context. In (8), “she” is r-cataphor, as it corefers with “Coaster Semenya,” which comes later in discourse.

(8) Nine days after she won the women’s 800m world championship in Berlin, Coaster Semenya returned home to the plains of Limpopo.

Finally, the r-new label is assigned to a new referent, which is not identifiable nor accessible from previous items in discourse. Both noun phrases “a new car” in (9) are r-new, as they do not corefer with one another. They are also not accessible from previous items.

(9) After the holidays, John arrived in a new car and Harry had also bought a new car.

### 1.4.2. The lexical level

The lexical level is to annotate the lexically identifiable or activated status of expressions. Lexical labels are usually annotated on content words. The basic lexical labels are the three-way given-accessible-new distinction as shown in Table 1.5. The given label is at the top of the table and the new label is at the bottom.

Table 1.5.

*Lexical (l-) Labels of a Simplified Version of the RefLex Scheme.*

| Label | Description |
|---|---|
| L-given | Item repeated, synonymous, or semantically superordinate to prior item |
| L-accessible | Item semantically subordinate to prior item |
| L-new | New item not identifiable nor accessible from prior item in discourse |

The l-given label is to annotate a repeated item, synonym, or item that is semantically superordinate to its antecedent. In (10), "car" is l-given as it is repeated from a previous item.

(10) A car was waiting in front of the hotel. I could see a woman in the car.

The l-accessible label is assigned to an anaphor that is subordinate to its antecedent. In (11), "the viola" is labeled as l-accessible as it is a type of "stringed instrument," thus it can be lexically activated from its antecedent, "stringed instruments."

(11) Bach wrote many pieces for stringed instruments. He must have loved the viola.

Finally, the l-new label is to annotate a new lexical expression which is semantically unrelated to a previous expression. In (12), "politician" is a new lexical expression, as it is not identifiable nor accessible from previous items.

(12) Smith was very optimistic. The polls showed a solid majority for the politician.

### 1.4.3. The alternative level

The alternative level is not included in the current version of the RefLex scheme (Riester & Baumann, 2017) but it is added to mark alternative-based contrastive focus (Rooth, 1992) in the first study of this dissertation. The alternative label is annotated on noun phrases.

In (13), "Mary" is in alternation with "John." "Mary" has been called instead of "John." "Mary" and "John" are annotated as "alt" at the alternative level.

(13) Did you call John? No, I called Mary.

Following a version of the RefLex scheme used in the first study of this dissertation, a word can have up to three labels, one from each of the referential, lexical, and alternative levels. In the example replicated below, "Mary" is annotated with three information status labels, r-new (referentially-new), l-new (lexically-new), and alt (alternative).

(14) Did you call John? No, I called Mary.

*r-new*
*l-new*
*alt*

## 1.5. Public Speech Style

The information status of a word is built on prior context and it is crucial to examine its relationship with prominence in a complete discourse context. Prior research is limited in that it investigates prominence in minimally contextualized utterances (e.g., question-and-answer prompts, picture description tasks) or excerpts from conversational speech (Birch & Clifton, 1995; Breen et al., 2010; Cole et al., 2010; Ito & Speer, 2008; Terken & Nooteboom, 1987). These studies yield insights on the relationship between information status and pitch accents in controlled and refined discourse contexts, but they might not fully capture the richer discourse context that occurs in intact samples of extended and natural discourse. There might be discrepancies between the use of pitch accents in controlled, decontextualized speech and the use of pitch accents in natural, contextualized speech, with further potential for variation across different speech styles.

Accent pattern is influenced by different speech styles (or speech modes). Prior research finds the evidence of different accenting pattern in read and spontaneous speech (Baumann & Riester, 2013; Blaauw, 1994; De Ruiter, 2015; Hirschberg, 1993; Luchkina & Cole, 2016;

Silverman, Blaauw, Spitz, & Pitrelli, 1992; Sityaev, 2000, Swerts, Strangert, & Heldner, 1996). Baumann and Riester (2013) show that the early-peak pitch accents (H+!H*, H+L*) occur more frequently in read speech than spontaneous speech in German. The spontaneous utterances are usually shorter and end with continuation rise (L* H%). Similarly, De Ruiter (2015) finds that low boundary tones are predominant in read speech while high boundary tones are frequent in spontaneous speech in German. The early-peak accents (H+!H*, H+L*) occur frequently in read speech. The L*+H accent is frequent in spontaneous speech while it is almost absent in read speech. More strikingly, most given referents are unaccented in read speech while only one third of given referents are unaccented in spontaneous speech. The author suggests that the speakers in spontaneous speech have a more increased cognitive load than the speakers in read speech and, as a consequence, they are less likely to use intonation to reflect information status only. Different from the findings in German, Sityaev (2000) finds that given information is often accented in read speech in a corpus of English. Personal pronouns and proper nouns, which are reintroduced to discourse context, are mostly accented. In addition, other function words such as deictic demonstratives (e.g., *this*, *that*) and numerals also tend to be accented in this corpus. The author proposes that rhythm and contrast interact with information status, which results in the unexpected accent assignment of given information and function words. Also, Hirschberg (1993) shows that less than the half of given items are unaccented only in a corpus of broadcast radio speech in English. Proper nouns tend to be accented, although they have been introduced to the text. The author suggests that speakers in this corpus attempt to refocus recently mentioned persons (proper nouns) when they have mentioned other persons more recently. The speakers in this broadcast radio speech tend to assign accents to proper nouns regardless of their information status.

A public speech from a TED talk has characteristics of both read and spontaneous speech. It is a read speech, as the speaker usually writes a script and practices it beforehand. At the same time, it is a spontaneous speech, as most TED Talks speakers do not read from the script or teleprompter, and they look at the audience while they deliver their speech. In this sense, a TED talk is different from typical read speech such as broadcast speech or inaugural speech, where speakers look at the script or teleprompter and read it out. From this reasoning, we expect to find mixed accent patterns between read and spontaneous speech in a TED talk. The first study of this dissertation uses an intact TED talk delivered by a male speaker and examines the production and perception of prominence in it as a function of discourse meaning as well as the phonological and phonetic properties of its utterances. The speaker of this speech talks about his new experiences over 30 days and encourages members of the audience to plan their own 30-day challenges. The speaker mentions a series of events related to persons and places in an engaging and lively fashion. Although speech style is not the main focus of this study, the accent pattern of this public speech is compared to that of conversational speech from the Buckeye Corpus in the first study.

The next sections, 1.6-1.9, review the literature for the second study of this dissertation, which investigates the cognitive representation of intonational contour. I first outline the tension between the abstractionist and exemplar models on prosodic encoding, then move on to the prosodic domains that are used as the units of prosody modeling in prior research.

## 1.6. Abstract Encoding of Prosody

This section reviews the phonological representation of prosody for the second study. What is the cognitive representation of prosody that underlies the dynamic properties of its

acoustic correlates in speech? In the present study, I focus on f0 as an acoustic correlate of prosodic prominence and my analysis does not extend to other acoustic parameters, such as duration and intensity, which prior work shows are also part of the acoustic encoding of prominence in English.

AM theory (Liberman 1975; Pierrehumbert, 1980) proposes that the intonational contour of an utterance is composed of a sparse specification of intonational features (pitch accents and boundary tones) and interpolated f0 between the tonal targets of those features. According to AM theory, the intonational contour of the noun phrase in Figure 1.2 consists of three tonal targets, two pitch accents (H*), and one boundary tone (H-).

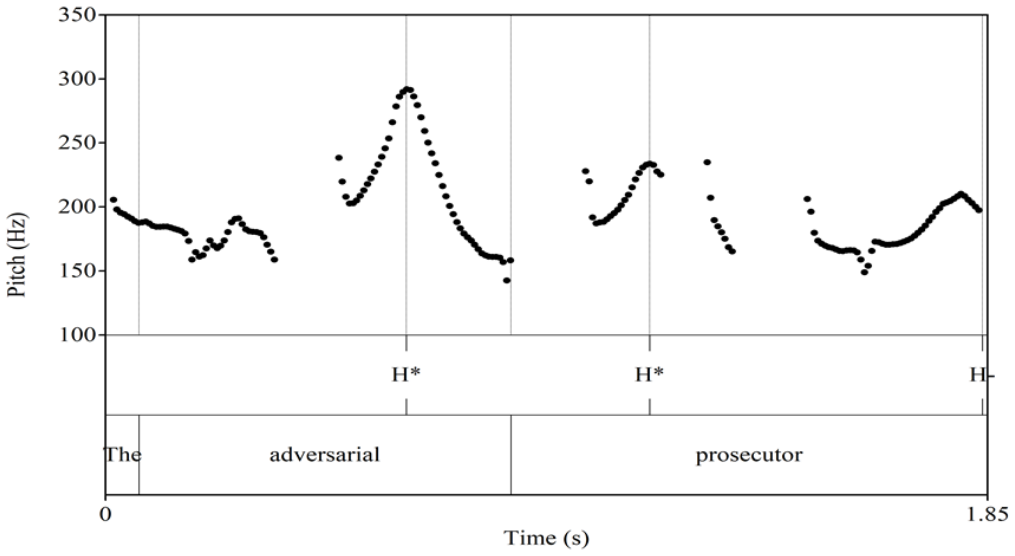


*Figure 1.2.* F0 contour over a noun phrase.

AM theory does not explicitly address the cognitive representation of the intonational contour, but the theory can be taken as a hypothesis that those intonational features are the information stored in the speaker's cognitive representation.

Prior research investigates the cognitive representation of prosody using an imitation paradigm. If listeners are able to reproduce intonational features in heard utterances, this

suggests that they must have perceived those intonational features and specified them in the memory representation of the utterance. There is no way to explain why listeners produce intonational features of heard utterances with such accuracy if we do not assume that the intonational features are encoded in memory representations. Recent evidence suggests that imitation of phonetic detail is incomplete, with a bias toward imitation that prioritizes primary cues to contrastive features over acoustic detail that is variable across utterances or speakers (Cole & Shattuck-Hufnagel, 2011, 2017; Michelas & Nguyen, 2011). In Cole and Shattuck-Hufnagel's (2011) study, listeners are asked to imitate spontaneously spoken utterances that they hear. The results show that listeners reproduce the phonological structure of the heard utterances more accurately than their phonetic details. Listeners are more accurate at imitating the location of pitch accents and boundary tones but less accurate at reproducing the duration of pauses and the occurrence of irregular pitch pulses as the acoustic correlates of intonational features in American English. Michelas and Nguyen (2011) examine whether listeners are able to reproduce initial high tone over an accentual phrase in French. Listeners are first asked to repeat stimulus noun phrases that they hear (repetition task); then they are asked to imitate the same stimulus in the similar way of the model speaker (imitation task). The results show that there are no significant differences between repetition vs. imitation tasks. Listeners are accurate at reproducing the initial high tone in both tasks. This provides evidence of the specification of the initial high tone in the cognitive representation of intonation among French listeners, despite its low frequency in everyday speech.

## 1.7. Exemplar Encoding of Prosody

Another line of research proposes that phonetic details over phonologically specified and unspecified regions of utterances are equally encoded in the mental lexicon. Exemplar theory (Goldinger, 1998) claims that all the phonetic detail of a heard utterance that a listener perceives is encoded and stored in their cognitive representation, including even nonlinguistic information such as the speaker's voice and background noise (Goldinger, 1998; Pufahl & Samuel, 2014). The perceived input creates a trace in phonetic space in mental representation and is recalled at the perception of similar linguistic inputs. Goldinger (1998) asks listeners to shadow and identify the words which vary in voice of the model speakers, lexical frequency, and number of repetitions. The results show that listeners' performance is better with the words produced with heard voice, low-frequency words, and more-repeated words. This suggests that the contextual information of words is stored in the mental lexicon and influences later speech production.

Studies of imitation show that people are capable of reproducing phonetic detail related to intonational contour in recently heard speech (Babel & Bulatov, 2011; Bosshardt, Sappok, Knipschild, & Hölscher, 1997; D'Imperio, Cavone, & Petrone, 2014; German, 2012; Gregory, Dagan, & Webster, 1997; Gregory, Webster, & Huang, 1993; Levitan et al., 2012). D'Imperio et al. (2014) examine whether Bari Italian speakers are able to imitate an unfamiliar pitch accent from Neapolitan Italian. Bari Italian has the L+H* accent (early peak) only while Neapolitan Italian has the L+H* and L*+H accents (late peak). The L+H* accent differs from the L*+H accent in the location of peak and, as a result, the L+H* accent has an early-rise shaped f0 contour while the L*+H accent has a late-rise f0 shape. The results show that Bari Italian speakers can imitate the Neapolitan Italian accent L*+H (late peak) by shifting the location of f0 peak. This suggest that speakers are able to reproduce an unfamiliar pitch contour, which does

not carry meaning in their regional dialect, by retaining the phonetic details of the heard pitch contour. German (2012) asks whether American English speakers can imitate unfamiliar intonational tune from Glasgow English. American English speakers typically produce a falling contour (H* L-L%) in a declarative sentence while Glasgow English speakers produce a rise-fall contour (L*+H H-L%). The results show that American English speakers can reproduce the rise-fall contour by shifting the f0 peak after they hear, and even produce the contour in new sentences. This suggests that speakers can rapidly learn and generalize a specific intonational pattern in an unfamiliar dialect.

The question of whether speech encoding is comprehensive over all perceived acoustic detail is interesting for intonation because the phonological specifications of intonational features are sparse. Relatively few tones define targets for an f0 contour that extends over an entire utterance. Does a listener encode all the details of an f0 contour that spans a prosodic phrase? Or does encoding privilege intervals of f0 that correspond to the targets of pitch accents and boundary tones while disregarding intervals of f0 interpolation between the tonal targets? If someone produces an intonational phrase with multiple accented and unaccented words, what are the domains in which the f0 contour is encoded? For this, AM theory and exemplar theory would predict two different domains of f0 encoding. Figure 1.3 shows two hypothetical domains of f0 encoding over a noun phrase.

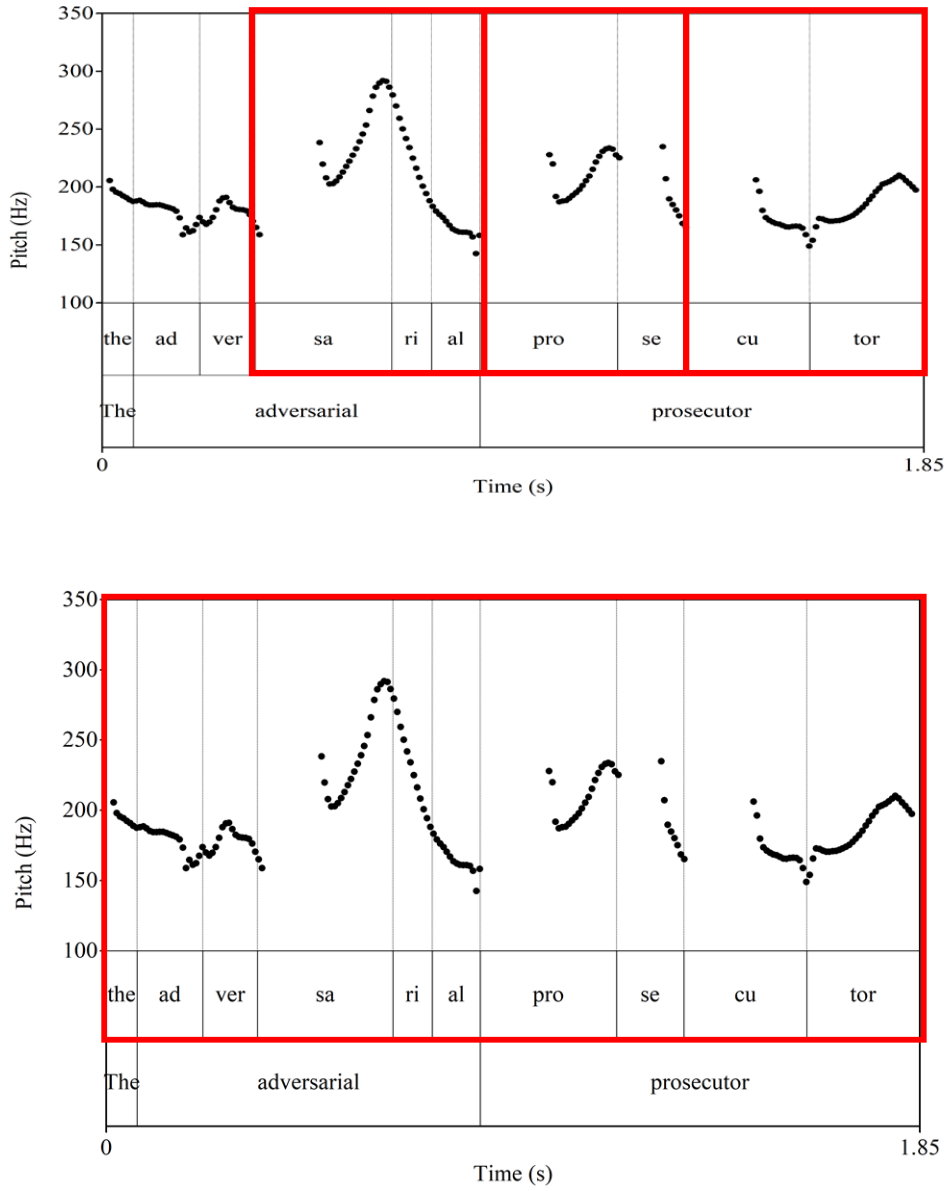


*Figure 1.3.* Hypothetical domains of f0 encoding. The domain is the intonational feature in the upper panel and the intermediate phrase in the lower panel.

AM theory predict that speakers will encode the regions of f0 contour carrying intonational features only while neglecting the other regions of f0 contour which do not specify intonational features. In the upper panel of Figure 1.3, the domain of f0 encoding is the intonational features as shown in red square. The domains cover the regions of intonational features ("-sarial

prosecutor") leaving out some preceding syllables in phonologically unspecified regions ("the adver-"). The f0 contour in each domain is simple with a couple of convex or concave as the f0 contour stretches over a few syllables. In comparison, exemplar theory predict that speaker will encode the details of f0 contour over the entire noun phrase. In the lower panel, the domain of f0 encoding is the intermediate phrase (in red square) covering the entire noun phrase ("the adversarial prosecutor"). In this phrase, the f0 contour is complex with multiple convex or concave because it the entire contour over a noun phrase. AM theory and exemplar theory can differ in their predictions, the domains of f0 encoding, and the f0 shape in the predicted domains.

F0 contour can be parsed into several different domains besides the intonational feature and the intermediate phrase. The next section reviews several prosodic domains that are used in different areas of prosodic research.

### 1.8. Domains of Prosodic Contour

F0 contours are modeled in work that examines imitation, entrainment (or convergence, assimilation), or intonation modeling and is measured with different domains of intonation, such as the prosodic phrase (Levitan & Hischberg, 2011; Gravano, Beňuš, Levitan, & Hirschberg, 2015), the accentable word (Reichel, 2011; Reichel & Cole, 2016), the intonational feature (Arvaniti & Ladd, 2009; Cole & Shattuck-Hufnagel, 2011; D'Imperio et al., 2014; German, 2012; Michelas & Nguyen, 2011), the stressed syllable (Kochanski et al., 2005), and the syllable (Andruski & Costello, 2004; Shih & Lu, 2015; Xu & Liu, 2006; Xu & Wang, 2001). Levitan and Hirschberg (2011) investigate entrainment between a pair of participants in a task-oriented speech. They measure mean and max intensity, mean and max pitch, voice quality, and speaking rate over participants' turns and task sessions. The participants' *turn* is defined as the unit of

speech separated from one another by at least 50 milliseconds. The acoustic measures are normalized by gender of participants and analyzed using a paired *t*-test or Pearson correlation coefficients based on the differences in absolute values of acoustic measures between participants. Reichel and Cole (2016) examine entrainment between cooperative vs. competitive conditions in a task-oriented speech. They examine intonational contours in the domain of the accentable word (Reichel, 2011). The accentable word is the domain that covers a content word and the preceding function words (e.g., "his categorical/ stance/ on protecting/ endangered/ animals/," where the slashes indicate the boundary location of each accentable word). The f0 values are transformed from Hz to semitones and submitted to a series of analyses: (1) The f0 values are modeled by the third-order polynomial regression, (2) the third-order polynomial coefficients are clustered into several classes (contour classes), and (3) the contour classes are quantified by standard string-based similarity metrics. Kochanski et al. (2005) investigate acoustic correlates of prominent syllables in read and spontaneous speech. They examine several acoustic measures including f0 over the stressed syllables, adopting a 452-millisecond fixed window centered on the stressed syllables. The f0 values are normalized and modeled using orthogonal polynomials. The polynomial coefficients are classified using a Bayesian classifier.

Prior research yields insights on f0 encoding over designated prosodic domains but it is limited in three regards: First, prior research uses prosodic domains based on theoretical assumption or analytical convenience, but these domains might not be the actual domains of prosodic encoding in cognitive representations of speech. To my knowledge, none of the research compares analyses of f0 in more than one prosodic domain to find the optimal domain for representing f0 contours. What is lacking to date is research that compares several prosodic domains and proposes the ideal domain of representing f0 contour, which can be further used as

the unit of f0 modeling in research on prosodic processing, modeling, imitation, or entrainment. Second, most prior research uses summarized or point measures of the f0 contour (max f0, mean f0). Although these measurements may potentially be sufficient to distinguish major tonal categories (e.g., H*, L*), more-detailed measurements are needed to test the hypothesis from exemplar theory, namely, that within-category phonetic details are also encoded. Third, a few prior studies use third-order polynomial coefficients to represent the time-varying prosodic contour (accent type), but this allows us to model the intonational tune with no more than one peak and one valley. There can be more complex intonational tune especially over large prosodic domains (e.g., intermediate phrase, intonational phrase) and needs to be modeled with higher-order polynomial coefficients.

The f0 contour is a nonlinear, time-series datum, as an f0 contour consists of f0 points at each time step. Two adjacent f0 points over a contour can be correlated (or autocorrelated) to one another as they are produced as a continuous event by a speaker. Figure 1.4 shows the hypothetical autocorrelation between two adjacent f0 values in time order, f0 at time point *t* and the other f0 point at the preceding time point *t*-1.

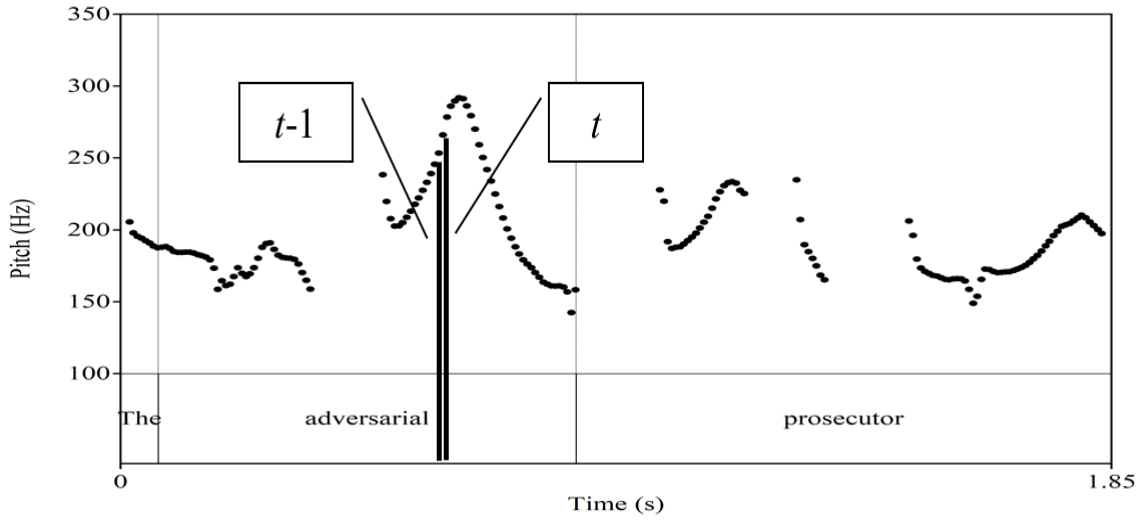


*Figure 1.4.* Hypothetical autocorrelation between adjacent f0 values.

Time-series analyses account for discrete time-point data with a possible internal structure (autocorrelation) that should be accounted for. The Generalized Additive Mixed Model (GAMM; Wood, 2017) is a time-series analysis and is a generalized linear model with a sum of smooth functions of covariates. The smooth functions are optimized by GAMM, which allows us to model complex f0 contour over a large prosodic domain. The maximum order of smooth functions can be pre-determined to prevent overfitted models.

## 1.9. Current Study

In this dissertation, I address two research questions in two studies: (1) How is the perception of prominence related to expectation-driven and signal-driven factors? (2) What is the cognitive representation of prosodic contour in English?

In the first study, in chapter 2, I investigate prominence in relation to expectation-driven factors (referential, lexical, alternative information status of a word) and signal-driven factors (phonological features, phonetic cues) in a complete TED talk. Two linguistic experts annotate

361 words of the speech in terms of referential, lexical, alternative information status using a simplified version of the RefLex scheme. Pitch accents are labeled by two other linguistic experts using a ToBI annotation convention. Phonetic cues (max f0, mean phone duration, mean word intensity) are extracted from the speech. I hypothesize that (1) pitch accents and acoustic cues (signal-driven factors) encode the information status of a word (expectation-driven factor) in the speech of the speaker and (2) expectation-driven and signal-driven factors independently contribute to the perception of prominence. Results inform us about the production and perception of prominence in relation to complex layers of information status, which may interact with pitch accents and acoustic cues in the public speech style outlined in chapter 2.

In the second study, in chapter 3, I examine the domain of f0 encoding in a previously collected corpus of imitated speech by comparing several prosodic domains proposed in prosodic theory and prior research. In the corpus, 33 speakers are asked to imitate the utterances produced by a model speaker of American English. The difference in f0 between the imitated and the model utterances are analyzed over six prosodic domains using GAMMs. The six prosodic domains are the intermediate phrase, accentable word, pitch accent, foot, stressed syllable, and syllable. GAMMs are used to capture the time-varying properties of the f0 contour, in which adjacent f0 values can be autocorrelated to one another. I hypothesize that the domain that best captures the imitation represents the actual domain of prosodic encoding. Results inform us about the domain of f0 encoding in the mental lexicon outlined in chapter 3.

# 2. PROSODIC PROMINENCE IN PUBLIC SPEECH

This study investigates prominence in relation to expectation-driven and signal-driven factors in an intact public speech from a TED talk as shown in Figure 2.1.

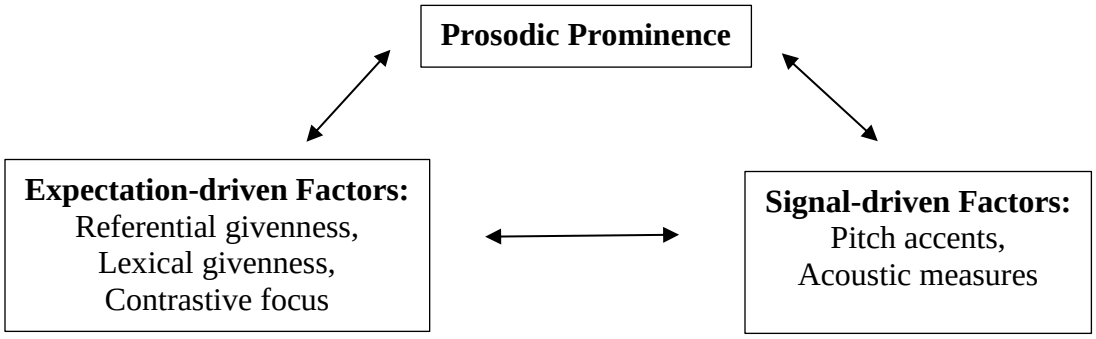


*Figure 2.1.* Design of the present study (modified from Cole et al., 2010).

For expectation-driven factors, the IS of a word is examined in relation to its referential meaning, lexical meaning, and alternative-based contrastive focus (Rooth 1992) using a simplified version of the RefLex scheme (Riester & Baumann, 2017). For signal-driven factors, the pitch accents labeled by a ToBI annotation convention (Veilleux et al., 2006) and three acoustic cues (max f0, mean phone duration, mean word intensity) are examined. Perceived prominence is measured by American English speakers who are not linguistic experts using the Rapid Prosody Transcription method (RPT; Cole et al., 2010).

The present study first examines the relationship among pitch accents, acoustic cues, and IS in the production of a speaker in chapter 2.1. It is essential to examine the production before the perception because the perception of prosodic prominence can be influenced by the way of speaker encoding IS using phonological features and phonetic cues. The IS, pitch accents, and

acoustic cues analyzed in the speech of the speaker are later used to model the perception of prominence rated by non-linguistic-expert listeners in chapter 2.2.

## 2.1. Prominence Produced in the Speech

### 2.1.1. Method

#### 2.1.1.1. Materials

The speech material, called “Try Something New for Thirty Days,” was obtained from TED Talks (https://www.ted.com/talks/matt_cutts_try_something_new_for_30_days) as shown in (15). It consists of 361 words, delivered by a male speaker of American English in a clear and engaging manner (*t* = 2’25”).

> (15) A few years ago, I felt like I was stuck in a rut, so I decided to follow in the footsteps of the great American philosopher, Morgan Spurlock, and try something new for 30 days. The idea is actually pretty simple. Think about something you've always wanted to add to your life and try it for the next 30 days. It turns out, 30 days is just about the right amount of time to add a new habit or subtract a habit—like watching the news—from your life. There's a few things I learned while doing these 30-day challenges. The first was, instead of the months flying by, forgotten, the time was much more memorable. This was part of a challenge I did to take a picture every day for a month. And I remember exactly where I was and what I was doing that day. I also noticed that as I started to do more and harder 30-day challenges, my self-confidence grew. I went from desk-dwelling computer nerd to the kind of guy who bikes to work—for fun. Even last year, I ended up hiking up Mt. Kilimanjaro, the highest mountain in Africa. I would never have been that adventurous before I started my 30-day challenges. I also figured out that if you really want something badly enough, you can do anything for 30 days. Have you ever wanted to write a novel? Every November, tens of thousands of people try to write their own 50,000-word novel from scratch in 30 days. It turns out, all you have to do is write 1,667 words a day for a month. So I did. By the way, the secret is not to go to sleep until you've written your words for the day. You might be sleep-deprived, but you'll finish your novel. Now is my book the next great American novel? No. I wrote it in a month. It's awful. But for the rest of my life, if I meet John Hodgman at a TED party, I don't have to say, "I'm a computer scientist." No, no, if I want to, I can say, "I'm a novelist."

2.1.1.2. Annotation of information status

The speech material was annotated for IS by two trained annotators using a simplified version of the RefLex scheme (Riester & Baumann, 2017). Three levels of IS—referential, lexical, and alternative—were considered as shown in Table 2.1.

Table 2.1.

*IS Annotation Labels Adapted from the RefLex Scheme (the words in bold correspond to the examples of each label).*

| Level | Label | Description | Example |
|---|---|---|---|
| R-level | Given | Coreferring item present in discourse | A car was waiting in front of the hotel. I could see a woman in **the car**. |
| | Bridging | Accessible item present in discourse | I tried to open the door but **the lock** was rusty. |
| | Unused | Globally unique new item in discourse | **President Barack Obama** delivered a brilliant speech in **Tucson**. |
| | New | Non-unique new item in discourse | After the holidays, John arrived in **a new car** and Harry had also bought **a new car**. |
| L-level | Given | Active expression in discourse | A car was waiting in front of the hotel. I could see a woman in the **car**. |
| | New | Inactive expression in discourse | Smith was very optimistic. The polls showed a solid majority for the **politician**. |
| Alt-level | Alt | Alternative expression in discourse | Did you call **John**? No, I called **Mary**. |

The referential (r-) level annotates the coreferential status of a word with the preceding words in discourse context. Five referential labels are used on individual nouns or noun phrases. The lexical (l-) level marks the lexically identifiable or activated status of a word. Two labels are tagged on each individual content word. Finally, the alternative (alt-) level was added to annotate alternative-based contrastive focus. One label is used on individual nouns or noun phrases. In Table 2.1, the labels are presented based on the IS hierarchy for each level (e.g., r-given < r-bridging < r-unused < r-new). There were a couple of additional labels observed in this speech,

but these labels obtained fewer than ten tokens (r-cataphor = 1, l-accessible = 7). They were merged with other labels (r-bridging for r-cataphor, l-given for l-accessible).

#### 2.1.1.3. Annotation of pitch accents

Pitch accents were annotated by two trained annotators following a ToBI annotation convention (Veilleux et al., 2006). The ToBI annotators were different from those who performed the annotation of IS. For some additional accent types, few tokens were found (H+!H* = 1, L*+H = 3). These items were reassigned to other accent types with the same starred tones (!H* for H+!H*) or with similar contour shapes (L+H* for L*+H).

#### 2.1.1.4. Acoustic measures

Acoustic measures of prominence were obtained using ProsodyPro (Xu, 2013). Max f0 (Hz) was manually inspected to check for pitch halving or doubling. Mean phone duration was obtained by dividing the entire duration of each word by the number of phones of the word. Mean word intensity was adopted from ProsodyPro without any modification. The acoustic cues were centered and scaled using the scale function in R (R Core Team, 2018).

#### 2.1.1.5. Analyses

Six models were used to examine prominence in relation to IS, pitch accents, and acoustic cues in the speech of the speaker as shown in Table 2.2.

Table 2.2.

*Summary of LMER Models and Distributional Analyses.*

| Model | Type | IV | | DV |
|---|---|---|---|---|
| | | Fixed Effects | Random Effects | |
| 1 | Mixed model | r-level + l-level + alt-level + accent + r-level:accent + l-level:accent + alt-level:accent | (1\|word) | F0 |
| 2 | | | | Duration |
| 3 | | | | Intensity |
| 4 | Chi-square test | Words with IS labels vs. without IS labels | | Accented vs. unaccented words |
| 5 | | Given words vs. non-given words (except the words without IS labels) | | |
| 6 | Fisher's exact test | Words with IS labels (except the words without IS labels) | | Accented words |

Three linear mixed-effects models (LMER; Models 1-3) were run, one for each acoustic cue as the DV (max f0, mean phone duration, mean word intensity). Acoustic measures were modeled as a function of IS (in red), accent type (in blue), and their interaction (in green) as fixed effects, and as a function of word as random effects. Three models had the same parameters but different DVs only. The following model was run for Model 1 using the lme4 package (Bates, Mächler, Bolker, & Walkers, 2015) in R: F0 ~ r-level + l-level + alt-level + accent + r-level:accent + l-level:accent + alt-level:accent + (1|word). Models 2 and 3 had duration and intensity as DVs, respectively.

The three linear mixed-effects models inform us about the relationship between (1) acoustic correlates and IS, and (2) acoustic correlates and pitch accents. In order to examine the relation between IS and pitch accents, three further analyses (Models 4-6) were run using Pearson's chi-square test with Yates's continuity correction and Fisher's exact test based on 2,000 replicates, based on the word frequency associated with those two categorical factors (see Table 2.6 below). Model 4 tests whether the words that are not eligible to carry IS following the

RefLex scheme are unaccented while the words delivering IS are accented. Models 5 and 6 examine the words delivering IS only. Model 5 tests whether words carrying given information are unaccented while the words carrying non-given information (accessible, new information) are accented. Finally, Model 6 tests whether different accent types are associated with different IS labels.

2.1.1.6. Predictions

Three predictions were made based on the pitch accent hierarchy and the IS hierarchy to examine the relationship among IS, pitch accents, and acoustic cues in the speech of the speaker.

First, the words with pitch accents ranked higher on the pitch accent hierarchy (L* < !H* < H* < L+H*) are produced by the speaker with more enhanced acoustic cues (i.e., higher f0, longer duration, higher intensity).

Second, the words with IS labels ranked higher on the IS hierarchy (given < bridging < unused < new) are produced by the speaker with more enhanced acoustic cues.

Third, the words with the higher-ranked pitch accents are associated with the words with higher-ranked IS labels.

### 2.1.2. Results

In this section, I first present the results from the LMER models (Models 1-3) that examine each acoustic cue in relation to IS, pitch accents, and their interactions. Then, I move on to the results from Pearson's chi-square test and Fisher's exact test (Models 4-6) that examine the relationship between IS and pitch accents.

#### 2.1.2.1. LMER models

The overall results from three LMER models are that acoustic cues are moderately correlated with pitch accents and IS in the speech of the speaker. Variation in mean phone duration is significantly associated with most IS labels and pitch accents. Variation in max f0 is associated with some pitch accents only. Mean word intensity is surprisingly associated with only very few IS labels.

Table 2.3 shows the results from Model 1 that examines max f0 in relation to IS and pitch accents.

Table 2.3.

*LMER Results for Modeling Max F0 as a Function of IS and Pitch Accents.*

| | est. | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | -.27 | .11 | 32.20 | -2.50 | <.05 |
| R-level | | | | | |
| r-given | .06 | .21 | 72.20 | .26 | .80 |
| r-bridging | -.52 | .42 | 320.70 | -1.24 | .22 |
| r-unused | -.41 | .55 | 311.20 | -.75 | .45 |
| r-new | .05 | .31 | 281.90 | .17 | .86 |
| L-level | | | | | |
| l-given | -.42 | .34 | 296.50 | -1.24 | .22 |
| l-new | -.16 | .25 | 300.40 | -.64 | .52 |
| Alt-level | | | | | |
| alt | .22 | .26 | 269.90 | .83 | .41 |
| Pitch accent | | | | | |
| L* | -.19 | .93 | 286.60 | -.21 | .84 |
| !H* | .06 | .40 | 316.20 | .15 | .88 |
| H* | .63 | .24 | 178.10 | 2.65 | <.01 |
| L+H* | 1.07 | .27 | 216.40 | 3.90 | <.01 |
| R-level:Pitch accent | | | | | |
| r-given:L* | -.42 | .88 | 320.60 | -.48 | .63 |
| r-bridging:L* | -.58 | 1.25 | 318.60 | -.46 | .64 |
| r-unused:L* | -.05 | .87 | 305.80 | -.05 | .96 |
| r-new:L* | -.28 | .68 | 296.40 | -.42 | .68 |
| r-given:!H* | .72 | .74 | 314.30 | .98 | .33 |
| r-bridging:!H* | .00 | .65 | 320.10 | .00 | 1.00 |
| r-unused:!H* | .02 | .71 | 309.60 | .03 | .98 |
| r-new:!H* | -.23 | .60 | 310.30 | -.38 | .71 |
| r-given:H* | -.46 | .39 | 312.80 | -1.16 | .25 |
| r-bridging:H* | -.39 | .63 | 317.50 | -.62 | .54 |
| r-unused:H* | .17 | .65 | 308.20 | .26 | .79 |
| r-new:H* | -.36 | .49 | 276.70 | -.74 | .46 |
| r-given:L+H* | .05 | .40 | 275.20 | .11 | .91 |
| r-bridging:L+H* | .05 | .67 | 313.20 | .08 | .94 |
| r-unused:L+H* | .07 | .67 | 299.90 | .10 | .92 |
| r-new:L+H* | -.84 | .44 | 305.20 | -1.89 | .06 |
| L-level:Pitch accent | | | | | |
| l-given:L* | .55 | 1.24 | 286.90 | .44 | .66 |
| l-new:L* | .81 | 1.08 | 294.10 | .76 | .45 |
| l-given:!H* | .63 | .62 | 319.00 | 1.01 | .31 |
| l-new:!H* | .50 | .48 | 319.20 | 1.03 | .30 |
| l-given:H* | .37 | .48 | 318.00 | .76 | .45 |
| l-new:H* | .24 | .37 | 307.20 | .66 | .51 |
| l-given:L+H* | .88 | .58 | 319.70 | 1.51 | .13 |
| l-new:L+H* | .44 | .38 | 309.20 | 1.18 | .24 |
| Alt-level:Pitch accent | | | | | |
| alt:L* | .03 | .78 | 307.90 | .04 | .97 |
| alt:!H* | -.14 | .75 | 316.50 | -.19 | .85 |
| alt:H* | -.44 | .49 | 318.50 | -.88 | .38 |
| alt:L+H* | .13 | .40 | 321.00 | .31 | .76 |

In Table 2.3, max f0 is significantly associated with a couple of pitch accents (H*, L+H*; in red), holding all other variables constant. Since pitch accents are characterized by changes in pitch, it is not surprising to find the meaningful relationship between max f0 and pitch accents. However, it is surprising to find that none of the IS categories is significantly associated with max f0. Among the pitch accents, the L+H* accent shows higher estimates than the H* accent, suggesting that L+H*, which is more highly ranked than H* on the pitch accent hierarchy, is associated with higher max f0.

Table 2.4 presents the results from Model 2 that examine the mean phone duration as a function of IS and pitch accents.

Table 2.4.

*LMER Results for Modeling Mean Phone Duration as a Function of IS and Pitch Accents.*

| | est. | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | -.5 | .13 | 159.80 | -4.09 | <.01 |
| R-level | | | | | |
| r-given | .06 | .22 | 296.30 | .28 | .78 |
| r-bridging | -.08 | .36 | 277.70 | -0.22 | .83 |
| r-unused | 1.15 | .50 | 319.40 | 2.31 | <.05 |
| r-new | -.14 | .29 | 316.90 | -.49 | .63 |
| L-level | | | | | |
| l-given | .88 | .31 | 320.90 | 2.85 | <.01 |
| l-new | .58 | .23 | 320.60 | 2.54 | <.05 |
| Alt-level | | | | | |
| alt | .60 | .21 | 216.20 | 2.82 | <.01 |
| Pitch accent | | | | | |
| L* | .48 | .91 | 224.90 | .53 | .60 |
| !H* | .74 | .35 | 292.60 | 2.11 | <.05 |
| H* | 1.16 | .23 | 285.50 | 4.95 | <.01 |
| L+H* | 1.11 | .27 | 265.30 | 4.09 | <.01 |
| R-level:Pitch accent | | | | | |
| r-given:L* | .12 | .78 | 320.60 | .16 | .88 |
| r-bridging:L* | .66 | 1.08 | 299.10 | .62 | .54 |
| r-unused:L* | -1.08 | .82 | 288.70 | -1.32 | .19 |
| r-new:L* | .01 | .64 | 284.90 | .02 | .98 |
| r-given:!H* | -.41 | .66 | 319.10 | -.62 | .54 |
| r-bridging:!H* | -.26 | .57 | 304.40 | -.45 | .65 |
| r-unused:!H* | -.37 | .65 | 316.40 | -.56 | .57 |
| r-new:!H* | -.03 | .55 | 318.90 | -.06 | .95 |
| r-given:H* | -.14 | .35 | 311.20 | -.41 | .68 |
| r-bridging:H* | -.46 | .56 | 319.00 | -.82 | .41 |
| r-unused:H* | -1.70 | .60 | 313.90 | -2.81 | <.01 |
| r-new:H* | -.20 | .46 | 313.30 | -.44 | .66 |
| r-given:L+H* | -.23 | .37 | 321.00 | -.61 | .54 |
| r-bridging:L+H* | .21 | .61 | 316.90 | .35 | .72 |
| r-unused:L+H* | -.74 | .62 | 310.50 | -1.19 | .23 |
| r-new:L+H* | .42 | .41 | 318.70 | 1.03 | .31 |
| L-level:Pitch accent | | | | | |
| l-given:L* | -.99 | 1.19 | 268.20 | -.83 | .40 |
| l-new:L* | -.05 | 1.05 | 233.30 | -.04 | .97 |
| l-given:!H* | -1.27 | .54 | 298.50 | -2.36 | <.05 |
| l-new:!H* | -.92 | .43 | 314.60 | -2.15 | <.05 |
| l-given:H* | -.55 | .42 | 288.80 | -1.32 | .19 |
| l-new:H* | -.84 | .34 | 318.40 | -2.48 | <.05 |
| l-given:L+H* | -.85 | .50 | 288.90 | -1.70 | .09 |
| l-new:L+H* | -.82 | .35 | 318.60 | -2.36 | <.05 |
| Alt-level:Pitch accent | | | | | |
| alt:L* | -.50 | .74 | 260.70 | -.68 | .50 |
| alt:!H* | -1.27 | .68 | 320.20 | -1.86 | .06 |
| alt:H* | -1.05 | .45 | 307.10 | -2.32 | <.05 |
| alt:L+H* | -.86 | .36 | 321.00 | -2.37 | <.05 |

In Table 2.4, variation in mean phone duration is significantly associated with many IS labels and pitch accents. For r-level, r-unused, which annotates proper nouns new to discourse context, is the only significant factor to predict variation in mean phone duration. For l-level, both l-given and l-new are significant predictors. Surprisingly, l-given shows higher estimates than l-new. In this speech sample, the speaker talks about his own experiences for 30 days and repeats certain expressions such as *I* and *thirty days*, labeled as l-given, with emphasis. For this reason, the l-given words seem to be acoustically enhanced, especially with longer duration, by the speaker. For alt-level, the alt label that annotates alternative, contrastive expressions is a significant factor. Finally, most pitch accents except L* are significant factors. Although H* shows a slightly higher estimate than L+H*, the overall increase of estimates among pitch accents are in line with the pitch accent hierarchy.

Table 2.5 shows the results from Model 3 that examines the mean word intensity in relation to IS and pitch accents.

Table 2.5.

*LMER Results for Modeling Mean Word Intensity as a Function of IS and Pitch Accents.*

| | est. | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept | -.16 | .13 | 97.30 | -1.26 | .21 |
| R-level | | | | | |
| r-given | .02 | .23 | 200.10 | .08 | .93 |
| r-bridging | -.96 | .44 | 316.10 | -2.16 | <.05 |
| r-unused | -.38 | .58 | 316.90 | -.67 | .51 |
| r-new | .21 | .33 | 306.40 | .65 | .52 |
| L-level | | | | | |
| l-given | -.20 | .36 | 314.70 | -.57 | .57 |
| l-new | .19 | .27 | 314.30 | .72 | .47 |
| Alt-level | | | | | |
| alt | -.06 | .27 | 277.30 | -.24 | .81 |
| Pitch accent | | | | | |
| L* | .30 | .99 | 288.90 | .30 | .76 |
| !H* | .18 | .42 | 321.00 | .43 | .67 |
| H* | .47 | .26 | 253.20 | 1.83 | .07 |
| L+H* | .47 | .30 | 265.10 | 1.57 | .12 |
| R-level:Pitch accent | | | | | |
| r-given:L* | .65 | .92 | 321.00 | .70 | .48 |
| r-bridging:L* | -1.73 | 1.30 | 316.80 | -1.33 | .19 |
| r-unused:L* | .89 | .92 | 308.80 | .96 | .34 |
| r-new:L* | -.11 | .72 | 304.60 | -.15 | .88 |
| r-given:!H* | 1.16 | .78 | 320.20 | 1.50 | .14 |
| r-bridging:!H* | 1.28 | .68 | 319.20 | 1.88 | .06 |
| r-unused:!H* | .62 | .75 | 314.90 | .84 | .40 |
| r-new:!H* | -.24 | .64 | 315.70 | -.37 | .71 |
| r-given:H* | -.11 | .41 | 320.90 | -.27 | .79 |
| r-bridging:H* | .23 | .66 | 320.50 | .35 | .72 |
| r-unused:H* | .17 | .69 | 314.20 | .25 | .80 |
| r-new:H* | -.35 | .52 | 301.70 | -.68 | .50 |
| r-given:L+H* | .14 | .43 | 310.80 | .34 | .74 |
| r-bridging:L+H* | .41 | .70 | 317.50 | .59 | .56 |
| r-unused:L+H* | -.06 | .71 | 309.80 | -.09 | .93 |
| r-new:L+H* | -.29 | .47 | 314.30 | -.61 | .54 |
| L-level:Pitch accent | | | | | |
| l-given:L* | -.71 | 1.32 | 297.50 | -.53 | .59 |
| l-new:L* | -.88 | 1.14 | 294.20 | -.77 | .44 |
| l-given:!H* | .19 | .65 | 318.30 | .29 | .77 |
| l-new:!H* | -.26 | .51 | 320.90 | -.51 | .61 |
| l-given:H* | .27 | .50 | 316.80 | .54 | .59 |
| l-new:H* | -.22 | .39 | 314.90 | -.56 | .57 |
| l-given:L+H* | .05 | .61 | 317.80 | .09 | .93 |
| l-new:L+H* | -.11 | .40 | 315.20 | -.28 | .78 |
| Alt-level:Pitch accent | | | | | |
| alt:L* | .21 | .83 | 306.90 | .25 | .80 |
| alt:!H* | -.04 | .79 | 318.30 | -.04 | .96 |
| alt:H* | -.09 | .52 | 318.20 | -.18 | .86 |
| alt:L+H* | .23 | .42 | 321.00 | .55 | .58 |

In Table 2.5, variation in mean word intensity is associated with only one IS label. R-bridging, used to annotate expressions activated from prior discourse context, shows a negative low estimate. In this speech sample, the speaker uses high intensity in general to address a large audience. The negative low estimate of r-bridging suggests that the speaker substantially decreases the volume of his voice when he talks about inferable expressions in rephrasing or giving examples.

Overall, mean phone duration is the most strongly correlated with both IS and pitch accents. Max f0 is correlated with pitch accents only. Mean word intensity is a weak correlate of IS and pitch accent in this speech sample. Considering the public speech style of the sample analyzed here, the mean word intensity might not be available to encode other linguistic information such as IS and pitch accents. Pitch accents are marked with longer mean phone duration and higher max f0, consistent with the pitch accent hierarchy. IS is not always marked with longer mean phone duration in line with the IS hierarchy. The next section presents the results from Models 4-6 that examine the relationship between IS and pitch accents.

#### 2.1.2.2. Distributional analyses

In this section, I first present the descriptive statistics of pitch accents in relation to IS labels and move on to the results from Models 4-6 using Pearson's chi-square tests and Fisher's exact tests.

Table 2.6 shows the frequency in this speech sample of words associated with pitch accents and IS labels. *Non-Referential* (NR) refers to the words that are not eligible to obtain any

r-labels following the RefLex scheme. Non-Lexical (NL) and Non-Alt are used for the words not eligible to obtain any l-labels and alt-labels, respectively.

Table 2.6.

*Distribution of Pitch Accents by IS Labels.*

| Level | Label | Unaccented | L* | !H* | H* | L+H* |
|---|---|---|---|---|---|---|
| R-level | NR | 124 | 5 | 15 | 31 | 26 |
| | R-given | 34 | 2 | 2 | 11 | 12 |
| | R-bridging | 5 | 1 | 6 | 5 | 4 |
| | R-unused | 3 | 5 | 6 | 11 | 8 |
| | R-new | 11 | 6 | 4 | 9 | 15 |
| L-level | NL | 152 | 1 | 7 | 27 | 20 |
| | L-given | 9 | 5 | 8 | 11 | 5 |
| | L-new | 16 | 13 | 18 | 29 | 40 |
| Alt-level | Non-Alt | 164 | 17 | 31 | 61 | 53 |
| | Alt | 13 | 2 | 2 | 6 | 12 |

In Table 2.6, it is surprising to observe that all accent types are used for all IS labels. The words labeled as NR, NL, and Non-Alt are mostly unaccented but sometimes accented by the speaker. The words labeled as given (r-given, l-given) show a different distribution for r- and l-levels. The words labeled as r-given are mostly unaccented while those labeled as l-given are mostly accented. In order to examine the pattern of accent assignment in relation to IS labels in more detail, the labels in Table 2.6 are recategorized into new labels and submitted to the Pearson's chi-square tests and Fisher's exact tests.

Model 4 tests the difference in the presence/absence of accent between the labels that are not eligible to carry IS and the labels that deliver IS. IS labels were recategorized using the new labels called "none" and "any." None includes NR for the r-level, NL for the l-level, and Non-Alt for the alt-level. Any contains all other labels except the none label (r-given, r-bridging, r-

unused, and r-new for the r-level; l-given and l-new for the l-level; alt for the alt-level). Pitch accent labels were recategorized using the new labels called "unaccented" and "accented." Unaccented includes the unaccented label. Accented contains all pitch accent labels (L*, !H*, H*, L+H*). Table 2.7 shows the results from Pearson's chi-square tests based on word frequency associated with those new labels.

Table 2.7.

*Chi-square Values for Labels as Carrying IS or not (None/Any) in Relation to Accent Assignment.*

| | | Unaccented/Accented | | |
|---|---|---|---|---|
| | | $\chi^2$ | *df* | *p* |
| None/Any | R-level | 27.96 | 1 | <.01 |
| | L-level | 113.32 | 1 | <.01 |
| | Alt-level | 1.70 | 1 | .19 |

In Table 2.7, the none label has a significantly different accent distribution from the any label at the r- and l-levels. This suggests that the words not eligible to carry IS (none) tend to be unaccented while the words carrying IS (any) tend to be accented at both r- and l-levels.

Model 5 tests the difference in the presence/absence of accent between the words carrying given information vs. the words carrying non-given information (accessible and new information). The IS labels were recategorized into two new labels called "given" and "non-given". Given includes r-given for the r-level, and l-given for the l-level. The alt-level is not examined, as there is not alt-given label. Non-given contains all other labels except given and none labels (r-bridging, r-unused, and r-new for the r-level; l-new for the l-level). The pitch accent labels were categorized as "unaccented" vs. "accented." Table 2.8 shows the results from Pearson's chi-square tests.

Table 2.8.

*Chi-square Values for Labels whether they are Given or not (Given/Non-given) in Relation to Accent Assignment.*

| | | Unaccented/Accented | | |
|---|---|---|---|---|
| | | $\chi^2$ | *df* | *p* |
| Given/Non-given | R-level | 21.14 | 1 | <.01 |
| | L-level | 1.40 | 1 | .24 |
| | Alt-level | *NA* | *NA* | *NA* |

In Table 2.8, given has significantly different distribution from non-given at the r-level only. This suggests that the words with r-given tend be unaccented while the words with l-given tend to be accented.

Model 6 tests the mapping between accent types and IS labels. The words with the none label were excluded. The words associated with different accent types and IS labels were submitted to Fisher's exact tests. There were no significant results for both r- ($p$ = .29) and l-levels ($p$ = .10). However, further qualitative analyses show the expected trend between accent types and IS labels. Figure 2.2 shows the word frequency of pitch accents in relation to r-labels. In the left panel, the accent types are arranged by r-labels and, in the right panel, the r-labels by accent types based on the same data.

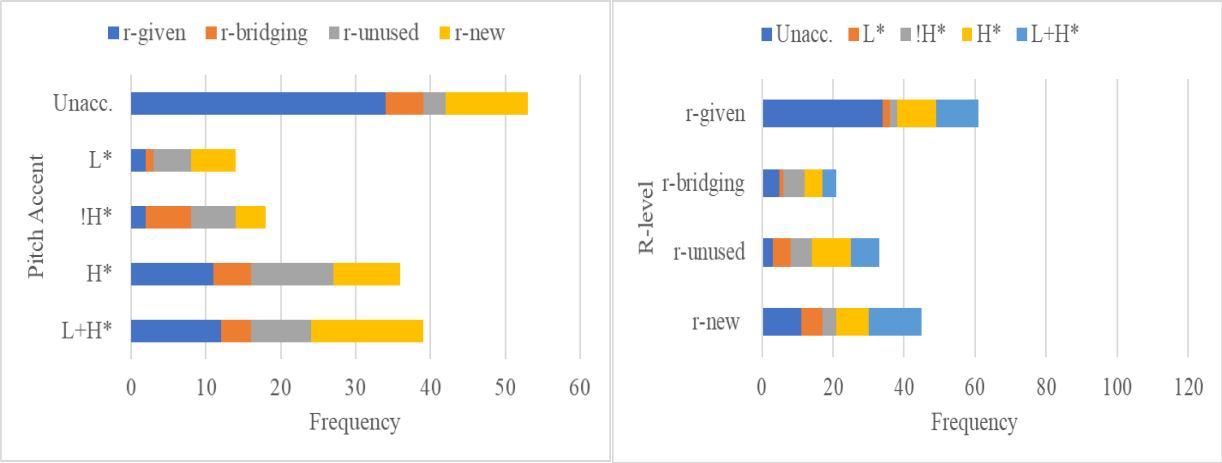


*Figure 2.2.* Distribution of pitch accents and r-labels excluding NR.

In Figure 2.2, !H* is the most frequent accent for r-bridging, H* for r-unused, and L+H* for r-new, consistent with the prediction that a pitch accent that is ranked higher on the pitch accent hierarchy is associated with an IS label that is also ranked higher on the IS hierarchy. Surprisingly, L* frequently occurs with r-unused and r-new. In this speech sample, the speaker tends to raise his pitch at the phrase- and utterance-final positions (L* H-, L* H-H%), which comprise 63 percent of the entire usage of L* by the speaker. L* is assigned to the rightmost word in an utterance due to the speaker's rising pattern and the rightmost word carries either r-unused or r-new. For this reason, L* seems to be frequently associated with r-unused and r-new. Besides r-unused and r-new, L* is next most frequently associated with r-given.

Figure 2.3 shows the accent types in relation to l-levels.

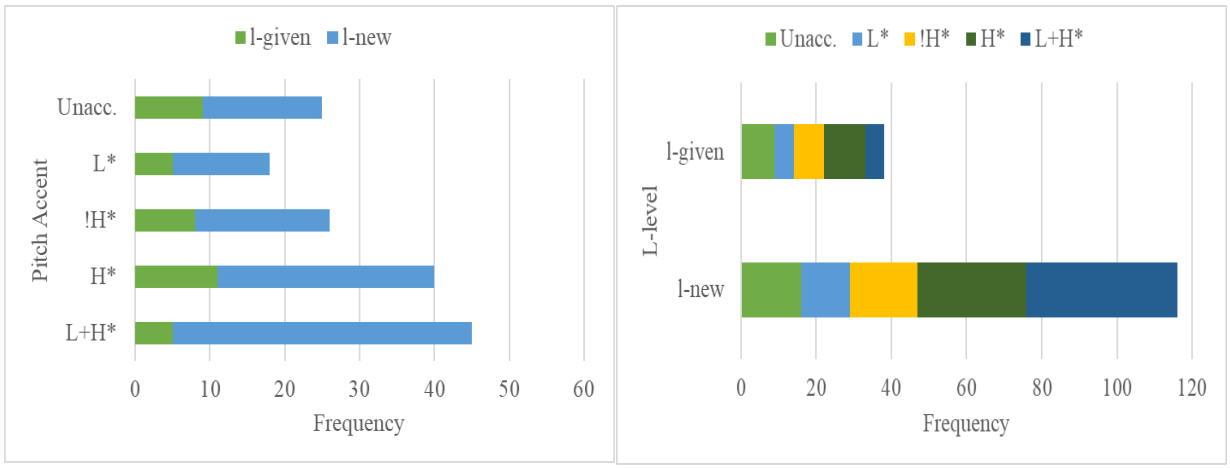


*Figure 2.3.* Distribution of pitch accents and l-labels excluding NL.

In Figure 2.3, L+H* occurs the most frequently for l-new. Surprisingly, H* is the most frequent accent for l-given. In this speech sample, the speaker is found to assign accents to lexically given words for emphasis (e.g., *I, thirty days*) and could have chosen the most neutral H* accent (Ladd, 2008).

Finally, Figure 2.4 shows the distribution of pitch accents and alt label.

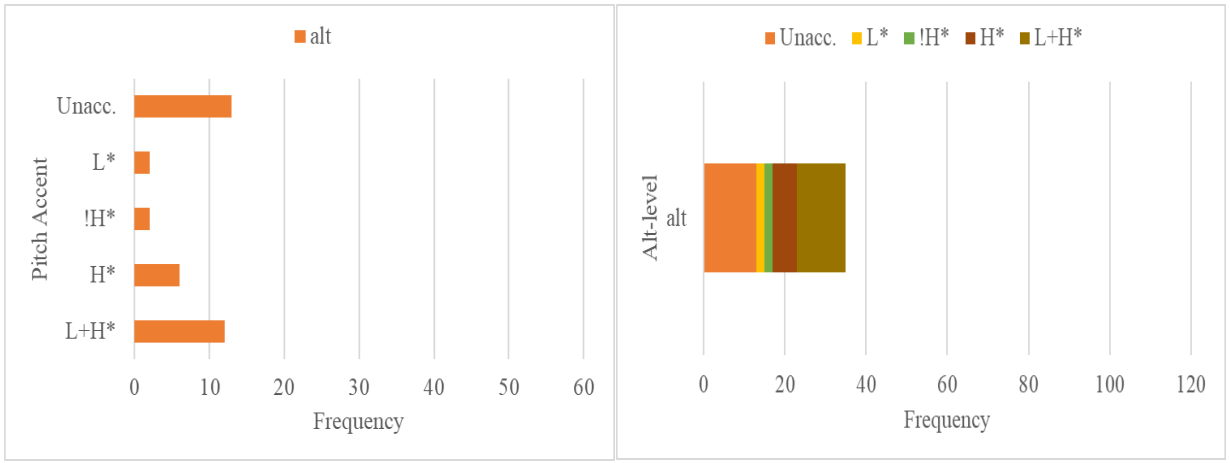


*Figure 2.4.* Distribution of pitch accents and alt label excluding Non-Alt.

In Figure 2.4, the L+H* accent, which has been described as marking contrastive focus, occurs the most frequently with alternative expressions. Unaccentedness is also frequently observed

with alternative expressions but this surprising finding can be understood as an artifact of the annotation scheme. The annotation of the alt label is applied to individual nouns or noun phrases. Within the same noun phrase, content words are accented the most frequently with L+H*, while function words are unaccented. The unaccentedness of alterative expressions in Figure 2.4 is driven by the unaccented function words.

Overall, pitch accents are found to be probabilistically related to IS labels. All accent types are observed for IS labels, but the most frequent pitch accent is associated with a certain IS label as claimed by the prior research. Words that are not eligible to carry IS are unaccented. Words delivering referentially given information in discourse context are also unaccented. In comparison, words carrying lexically given information are surprisingly accented in this speech sample. !H* and H* are likely to deliver bridging and unused information in discourse context, respectively. L+H* tends to be used for new information and contrastive expressions.

## 2.2. Prominence Perceived by Listeners

In the previous section, I presented the results that investigate the relationship among expectation-driven (IS) and signal-driven factors (pitch accents, acoustic measures) in the speech of the speaker. Mean phone duration and max f0 are associated with most IS and pitch accents while mean word intensity is surprisingly not correlated with most IS and pitch accents. Pitch accents are probabilistically associated with IS labels although they respect the mapping previously claimed, associating the given label with low-prominence pitch accents. Referential IS is different from lexical IS in that referentially given expressions are likely to be unaccented while lexically given expressions tend to be accented.

In the next section, I present the analyses of prominence perceived by untrained listeners in relation to IS, pitch accents, and acoustic measures as produced by the speaker. The same speech material was used for the perception experiment. IS, pitch accents, and acoustic measures that were obtained from the speech in the previous section were submitted to model their effects on perceived prominence.

### 2.2.1. Method

#### 2.2.1.1. Perception experiment

Thirty-five American English speakers from University of Illinois participated in a prominence rating experiment. They marked prominence while listening to the speech excerpt on the online interface LMEDS (Language Markup and Experimental Design Software; Mahrt, 2013). They were instructed to select "words that stand out in the speech stream by virtue of being louder, longer, more extreme in pitch, or more crisply articulated than other words in the same utterance." The speech sample was broken into four small excerpts of 30-39 seconds each, presented in natural order. Participants listened to each speech excerpt twice while viewing a transcript of the same excerpt on the computer screen. The transcript was presented without punctuation or capitalization. Figure 2.5 shows the screen capture of the experiment on the online interface.

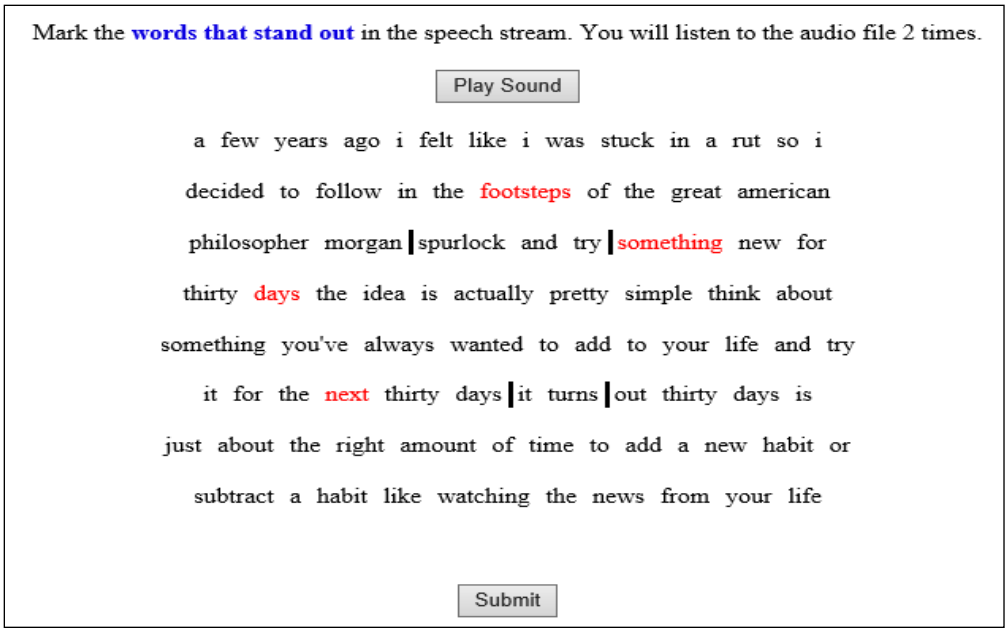


*Figure 2.5.* The screen capture of the prominence rating experiment on the online interface. Listeners marked prosodic prominence by clicking words that were perceived as prominent (the words shown in red). They were also asked to mark prosodic boundaries in the same experiment (shown as black vertical bar between words) but the boundaries were not analyzed in the current study.

2.2.1.2. Analyses

The prominence rated by linguistics nonexpert listeners was converted into binary coding, 1 for the words marked as prominent and 0 for the words marked as nonprominent, and submitted to three generalized linear mixed-effects models (GLMER; Models 7-9). In order to examine the effects of expectation-driven (r-, l-, alt-levels) and signal-driven factors (pitch accents, max f0, mean phone duration, mean word intensity) on perceived prominence, it is ideal to include all the factors and their interactions in one model, but due to a convergence issue, this was not possible. I was specifically interested in examining how the interaction between expectation-driven and signal-driven factors influence the perception of prominence and had to run three models that include the interactions in relation to each acoustic cue as shown in Table 2.9.

Table 2.9.

*Summary of GLMER Models.*

| Model | Type | IV | | DV |
|---|---|---|---|---|
| | | Fixed Effects | Random Effects | |
| 7 | Mixed model | r-level + l-level + alt-level + f0 + duration + intensity + accent + f0:r-level + f0:l-level + f0:alt-level + f0:accent | (1\|subject) | Perceived prominence |
| 8 | | r-level + l-level + alt-level + f0 + duration + intensity + accent + duration:r-level + duration:l-level + duration:alt-level + duration:accent | | |
| 9 | | r-level + l-level + alt-level + f0 + intensity + accent + intensity:r-level + intensity:l-level + intensity:alt-level + intensity:accent | | |

The binary rating of perceived prominence for each word, as rated by each individual annotator, was modeled in relation to IS, accent types, acoustic cues as fixed effects, and subjects as random effects. Three models had the same DV but slightly different interaction terms. Model 7 included interactions with max f0 (in red), Model 8 with mean phone duration (in blue), and Model 9 with mean word intensity (in green). The following model was run for Model 7 using the lme4 package in R: perceived prominence ~ r-level + l-level + alt-level + f0 + duration + intensity + accent + *f0:r-level* + *f0:l-level* + *f0:alt-level* + *f0:accent* + (1|subject). The italicized interaction terms were substituted with *duration:r-level* + *duration:l-level* + *duration:alt-level* + *duration:accent* for Model 8, and *intensity:r-level* + *intensity:l-level* + *intensity:alt-level* + *intensity:accent* for Model 9.

2.2.1.3. Predictions

Three predictions were made for the effects of expectation-driven and signal-driven factors on the perception of prominence.

First, the words with pitch accents ranked higher on the pitch accent hierarchy are more likely to be perceived as prominent.

Second, the words with more enhanced acoustic cues are more likely to be perceived as prominent.

Third, the words with IS labels ranked higher on the IS hierarchy are more likely to be perceived as prominent.

### 2.2.2. Results

The overall results from three GLMERs show that perceived prominence is significantly associated with most information statuses, pitch accents, acoustic cues, and their interactions. I first present the summaries from three GLMER models; then, I discuss the results based on the figures that are obtained from the same models presented later in this section.

Table 2.10-2.12 show the results from three GLMER models (Models 7-9) that examine perceived prominence in relation to IS, pitch accents, acoustic measure, and interactions. Table 2.10 shows the results modeling the interactions with max f0, Table 2.11 with mean phone duration, and Table 2.12 with mean word intensity.

Table 2.10.

*GLMER Results for Modeling Prominence Rating as a Function of IS, Pitch Accents, Acoustic Cues, and the Interactions with Max F0.*

| | est. | *SE* | *z* | *p* |
|---|---|---|---|---|
| (Intercept) | -3.09 | .13 | -23.57 | <.01 |
| R-level | | | | |
| r-given | -1.00 | .11 | -9.17 | <.01 |
| r-bridging | -.64 | .14 | -4.53 | <.01 |
| r-unused | .26 | .09 | 2.86 | <.01 |
| r-new | -.39 | .09 | -4.15 | <.01 |
| L-level | | | | |
| l-given | -.28 | .12 | -2.30 | <.05 |
| l-new | .74 | .08 | 9.64 | <.01 |
| Alt-level | | | | |
| alt | .62 | .09 | 7.11 | <.01 |
| Acoustic cue | | | | |
| f0 | .73 | .06 | 12.29 | <.01 |
| duration | .77 | .03 | 24.84 | <.01 |
| intensity | -.04 | .04 | -1.11 | .27 |
| Pitch Accent | | | | |
| L* | 1.69 | .13 | 13.05 | <.01 |
| !H* | .61 | .12 | 4.92 | <.01 |
| H* | 1.61 | .09 | 17.64 | <.01 |
| L+H* | 2.00 | .10 | 19.54 | <.01 |
| R-level:F0 | | | | |
| r-given:f0 | .30 | .08 | 3.76 | <.01 |
| r-bridging:f0 | -.48 | .17 | -2.81 | <.01 |
| r-unused:f0 | .13 | .09 | 1.40 | .16 |
| r-new:f0 | .44 | .11 | 3.98 | <.01 |
| L-level:F0 | | | | |
| l-given:f0 | .38 | .10 | 3.96 | <.01 |
| l-new:f0 | 0 | .07 | -.03 | .98 |
| Alt-level:F0 | | | | |
| alt:f0 | -.23 | .07 | -3.09 | <.01 |
| Pitch Accent:F0 | | | | |
| f0:L* | -.41 | .19 | -2.22 | <.05 |
| f0:!H* | -1.38 | .16 | -8.80 | <.01 |
| f0:H* | -.81 | .09 | -9.55 | <.01 |
| f0:L+H* | -.68 | .08 | -8.49 | <.01 |

Table 2.11.

*GLMER Results for Modeling Prominence Rating as a Function of IS, Pitch Accents, Acoustic Cues and the Interactions with Mean Phone Duration.*

| | est. | *SE* | *z* | *p* |
|---|---|---|---|---|
| (Intercept) | -3.08 | .13 | -23.51 | <.01 |
| R-level | | | | |
| r-given | -.47 | .10 | -4.56 | <.01 |
| r-bridging | -.58 | .17 | -3.45 | <.01 |
| r-unused | .62 | .09 | 6.80 | <.01 |
| r-new | -.04 | .09 | -.44 | .66 |
| L-level | | | | |
| l-given | -.17 | .13 | -1.34 | .18 |
| l-new | .58 | .08 | 7.26 | <.01 |
| Alt-level | | | | |
| alt | .60 | .09 | 6.95 | <.01 |
| Acoustic Cue | | | | |
| f0 | .21 | .03 | 6.43 | <.01 |
| duration | .91 | .07 | 12.51 | <.01 |
| intensity | .04 | .04 | 1.10 | .27 |
| Pitch Accent | | | | |
| L* | 1.54 | .15 | 10.62 | <.01 |
| !H* | .64 | .13 | 4.88 | <.01 |
| H* | 1.47 | .10 | 14.68 | <.01 |
| L+H* | 1.81 | .10 | 17.41 | <.01 |
| R-level:Duration | | | | |
| r-given:duration | -.55 | .11 | -5.04 | <.01 |
| r-bridging:duration | .12 | .22 | .54 | .59 |
| r-unused:duration | -.99 | .11 | -9.06 | <.01 |
| r-new:duration | -.42 | .10 | -4.24 | <.01 |
| L-level:Duration | | | | |
| l-given:duration | -.25 | .11 | -2.22 | <.05 |
| l-new:duration | .25 | .08 | 3.17 | <.01 |
| Alt-level:Duration | | | | |
| alt:duration | -.35 | .10 | -3.38 | <.01 |
| Pitch Accent:Duration | | | | |
| duration:L* | -.04 | .19 | -.23 | .82 |
| duration:!H* | .26 | .13 | 1.90 | .06 |
| duration:H* | -.01 | .09 | -.09 | .93 |
| duration:L+H* | .12 | .10 | 1.19 | .23 |

Table 2.12.

*GLMER Results for Modeling Prominence Rating as a Function of IS, Pitch Accents, Acoustic Cues and the Interactions with Mean Word Intensity.*

| | est. | *SE* | *z* | *p* |
|---|---|---|---|---|
| (Intercept) | -3.16 | .12 | -26.31 | <.01 |
| R-level | | | | |
| r-given | -.78 | .10 | -7.63 | <.01 |
| r-bridging | -.96 | .16 | -6.04 | <.01 |
| r-unused | .20 | .08 | 2.40 | <.05 |
| r-new | -.42 | .09 | -4.84 | <.01 |
| L-level | | | | |
| l-given | .06 | .11 | .53 | .60 |
| l-new | .73 | .07 | 10.29 | <.01 |
| Alt-level | | | | |
| alt | .46 | .08 | 5.57 | <.01 |
| Acoustic Cue | | | | |
| f0 | .34 | .03 | 10.73 | <.01 |
| intensity | .27 | .06 | 4.14 | <.01 |
| Pitch Accent | | | | |
| L* | 1.99 | .13 | 15.14 | <.01 |
| !H* | .86 | .12 | 7.24 | <.01 |
| H* | 1.83 | .09 | 20.66 | <.01 |
| L+H* | 2.33 | .09 | 25.32 | <.01 |
| R-level:Intensity | | | | |
| r-given:intensity | .17 | .09 | 1.87 | .06 |
| r-bridging:intensity | -.31 | .14 | -2.14 | <.05 |
| r-unused:intensity | .49 | .14 | 3.47 | <.01 |
| r-new:intensity | 1.52 | .13 | 11.70 | <.01 |
| L-level:Intensity | | | | |
| l-given:intensity | -.95 | .13 | -7.47 | <.01 |
| l-new:intensity | -.69 | .08 | -8.30 | <.01 |
| Alt-level:Intensity | | | | |
| alt:intensity | -.38 | .12 | -3.06 | <.01 |
| Pitch Accent:Intensity | | | | |
| intensity:L* | -.28 | .18 | -1.55 | .12 |
| intensity:!H* | .21 | .15 | 1.41 | .16 |
| intensity:H* | .19 | .09 | 2.25 | <.05 |
| intensity:L+H* | -.20 | .09 | -2.29 | <.05 |

Three GLMER models do not return the identical results for the main effects but they show the compatible results. Across all the three models, the perceived prominence is found to be significantly associated with most IS, pitch accents, acoustic cues. As the three models show similar results, I discuss the patterns in figures based on Model 7 (interaction with max f0) only for simplicity. Three GLMER models included different interactions and showed slightly different results for the interactions. I present the figures that describe the interactions in the three models and discuss the results in more detail.

Figure 2.6 shows the predicted probability in prominence rating from Model 7 in relation to pitch accents.

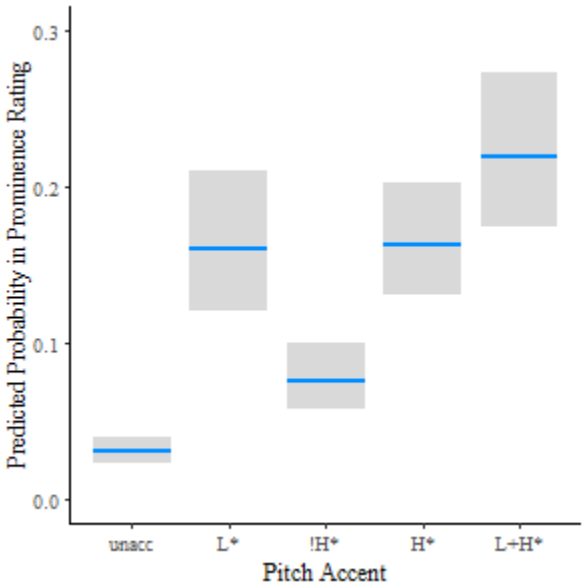


*Figure 2.6.* Relationship between pitch accents and predicted probability in prominence rating. *Unacc* stands for unaccented words.

There is increasing trend from unaccented (leftmost) to L+H* (rightmost), except L*, on the *x*-axis in relation to the predicted probability in prominence rating on the *y*-axis. This suggests that the words with higher-ranked pitch accents are more likely to be perceived as prominent. As

observed in the speech of the speaker, L* does not strictly follow the pitch accent hierarchy in this speech sample. From the speaker's words with rising pitch contour at the end of phrase or utterance (annotated as L* H- or L* H-H%), listeners could have perceived the words with L* as prominent in relation to the following high boundary tones. Or, they could have considered L* prominent because L* happens to be in the nuclear accent position, i.e., the structurally strong position which is often adjacent to the prosodic phrase boundary. These may have contributed to the increased prominence rating of L*.

Figure 2.7 shows the predicted probability of prominence rating from Model 7 as a function of acoustic cues.

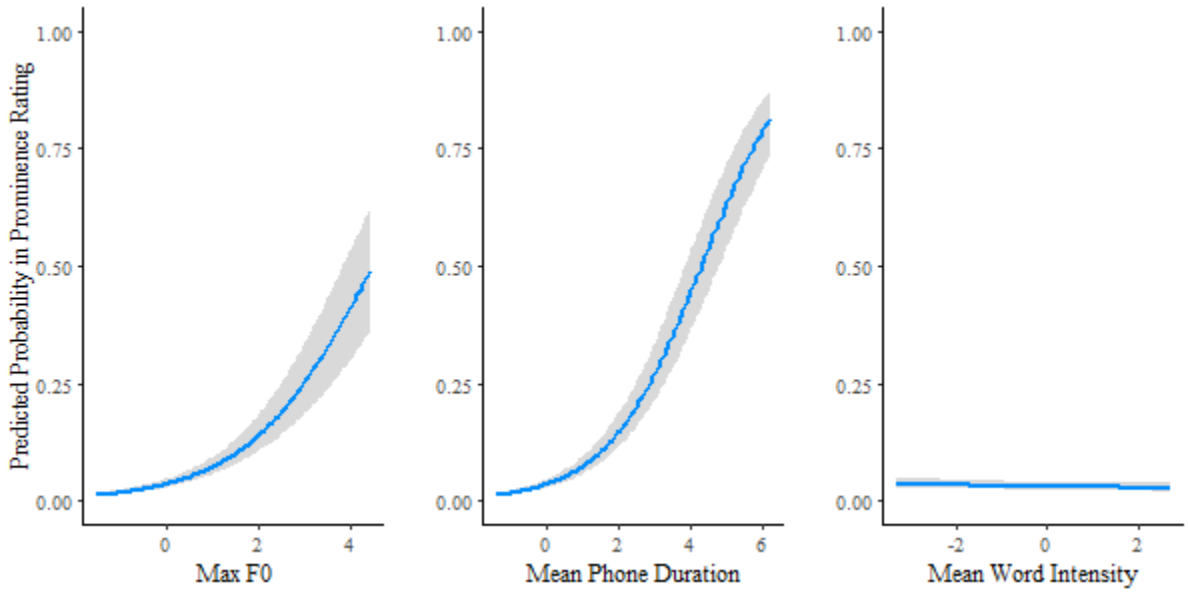


*Figure 2.7.* Relationship between acoustic cues and predicted probability in prominence rating.

In Figure 2.7, max f0 (left panel) and mean phone duration (middle panel) show an increasing trend from left to right on the *x*-axis in predicting probability of prominence rating on the *y*-axis. Put differently, the words with higher f0 and longer duration are more likely to be perceived as prominent than the words with lower f0 and shorter duration. Surprisingly, mean word intensity

does not contribute to the listeners' prominence rating. This parallels the findings from the speech of the speaker. The speaker uses high intensity throughout his narrative to address a large audience and so has less opportunity to use intensity to encode other information such as IS and pitch accents. As intensity is not a strong cue of the other information in the speech of the speaker, it seems to be less weighted by the listeners while judging prominence.

Figure 2.8 shows the predicted probability in prominence rating from Model 7 in relation to IS labels.

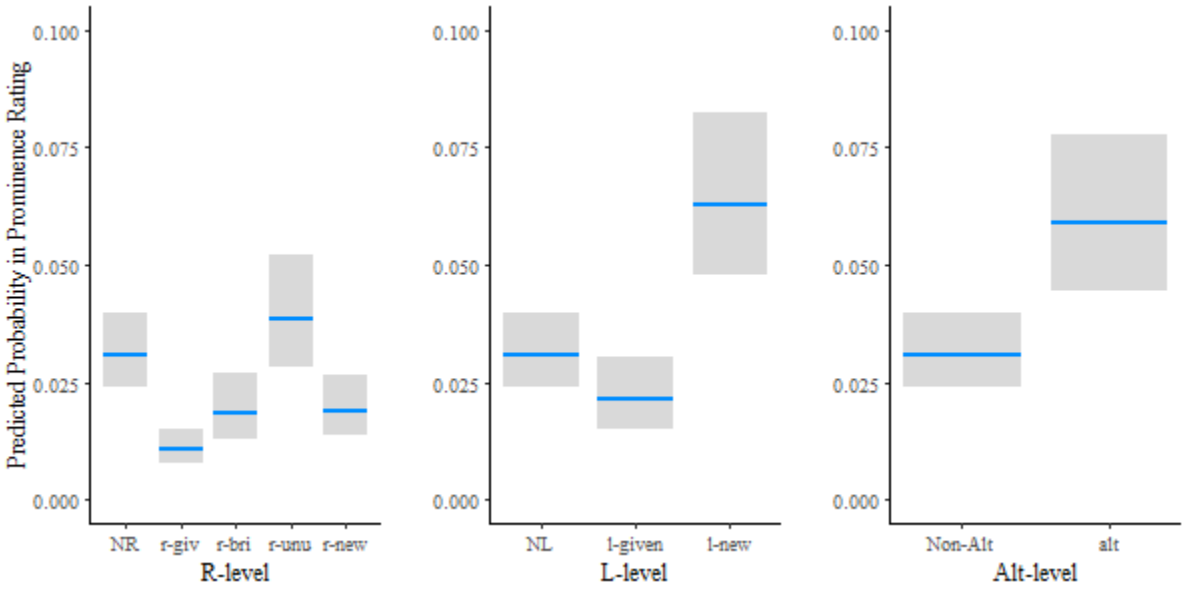


*Figure 2.8.* Relationship between IS and predicted probability in prominence rating. NR (left panel), NL (middle panel), and Non-Alt (right panel) stand for the words that are not eligible to obtain IS labels at a given level following the RefLex scheme.

Setting aside the words with the none label (NR, NL) and r-new, IS labels show the increasing trend from left to right on the *x*-axis in relation to the predicted probability of prominence rating on the *y*-axis. The words with higher-ranked IS labels are more likely to be perceived as prominent. There are some labels that need further consideration. First, the NR and NL labels show surprisingly higher estimates than the given or bridging labels in predicting probability in

prominence rating. The words with NR and NL are mostly function words or occasionally content words that do not refer to an entity. Among those words, the negations (e.g. *never, not, but*) and discourse markers (e.g., *also, instead, really*) seem to be perceived as prominent by most listeners. Second, the words with r-new are perceived as less prominent than the words with r-unused. The words with r-unused are especially marked with longer duration in the speech of the speaker and are expected to be perceived as prominent by most listeners. Listeners judge the words as prominent in relation to other surrounding words, and the referentially new words may happen to be next to other words such as proper nouns, discourse markers, that are mostly perceived as prominent. Finally, the words with l-given show surprisingly low estimates. In the speech of the speaker, the words with l-given are found to be mostly accented, especially with H*. However, the same words are not perceived to be as prominent as expected, suggesting listeners do not rely only on pitch accents while judging prominence.

So far, we have examined the results on the perceived prominence as a function of main effects. We now move on to analyzing the results in relation to interactions. Figure 2.9 shows the predicted probability in prominence rating from Model 7 as a function of the interactions between IS and max f0.

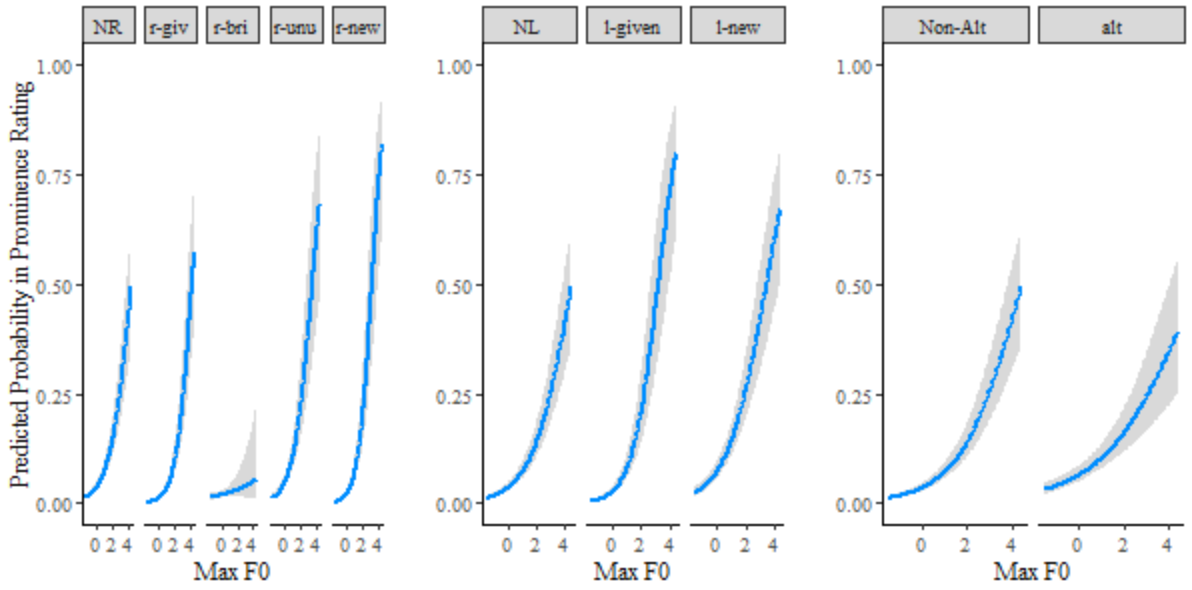


*Figure 2.9.* Predicted probability in prominence rating in relation to the interactions between IS and max f0. For r-level (left panel), *r-giv* stands for r-given, *r-bri* for r-bridging, and *r-unu* for r-unused.

In Figure 2.9, the effects of max f0 (in blue contour) are different from one another across different IS labels. This suggests that the same acoustic cue influences the perceived prominence differently depending on IS. For r-level (left panel), the slopes become steeper from NR to r-new, except r-bridging. Put differently, if the same amount of max f0 increases, the effects of max f0 on prominence rating is stronger for the label that is ranked higher on the IS hierarchy. For l-level (middle panel), the slopes are surprisingly steeper for l-given than l-new. There are the opposite patterns between the r- and l-levels. Max f0 shows a higher estimate on the prominence rating for r-new than that for r-given while it shows lower estimates for l-new than l-given. This suggests that max f0 influences perceived prominence differently between (1) given vs. new labels, and (2) referential vs. lexical levels. For alt-level (right panel), the effects of max f0 are surprisingly more gradual for alt than Non-Alt. Alternative expressions are usually marked by L+H* and are expected to have stronger effects of acoustic cues than non-alternative

expressions. The speaker seems to perform contrastive focus with acoustic diminishment. He exaggerates his speech and uses low pitch as a way of drawing listeners' attention.

Figure 2.10 shows the predicted probability in prominence rating from Model 8 in relation to the interactions between IS and mean phone duration.

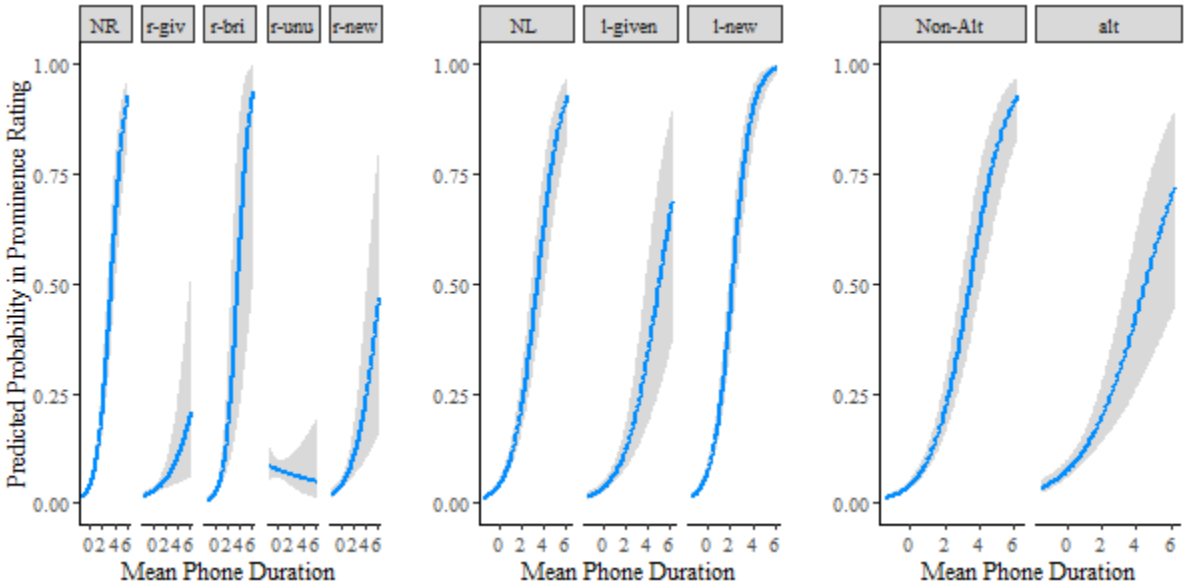


*Figure 2.10.* Predicted probability in prominence rating in relation to the interactions between IS and mean phone duration.

Similar to the findings from max f0, the effects of mean phone duration vary across IS labels. Also, the effects are more gradual for alternative expressions than non-alternative expressions. Different from the findings from max f0, the effects of mean phone duration are similar between given and new at both r- and l-level. The effects are stronger for new than given labels.

Figure 2.11 shows the predicted probability in prominence rating from Model 9 in relation to the interactions between IS and mean word intensity.

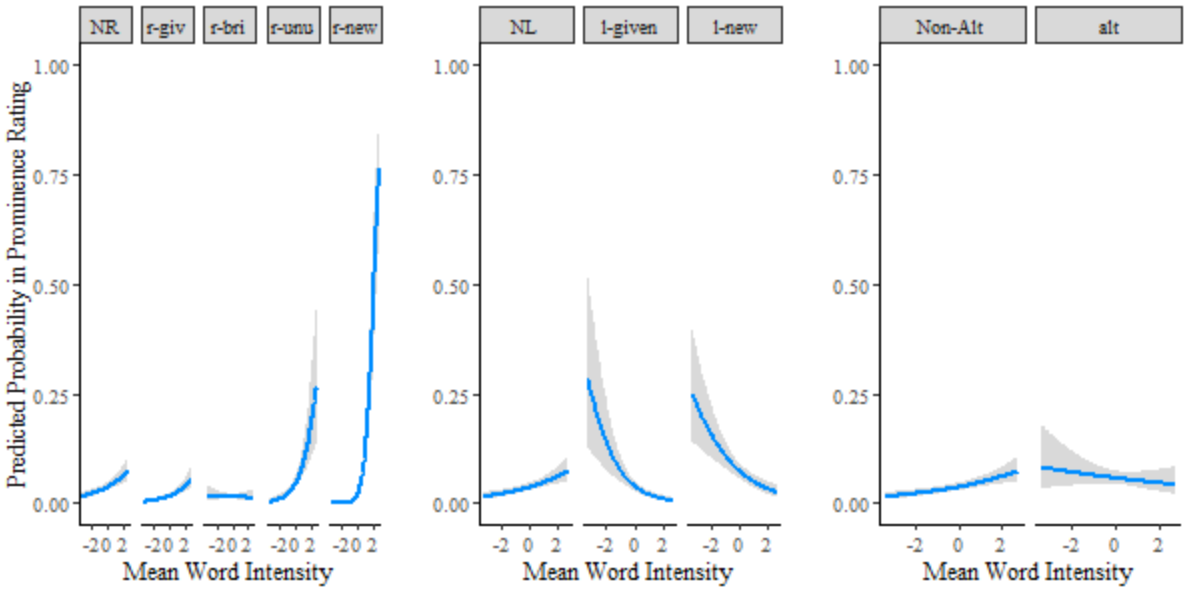


*Figure 2.11.* Predicted probability in prominence rating in relation to the interactions between IS and mean word intensity.

In Figure 2.11, the effects of mean word intensity are weak across most IS labels, except r-unused and r-new. This is not surprising because the results on the main effects (Figure 2.7) reveal that listeners less rely on intensity compared to other acoustic cues while judging prominence. Surprisingly, the effects of mean word intensity show a negative trend for alternative expressions. Put differently, alternative expressions associated with lower intensity tend to be perceived more prominent by listeners. In this speech sample, the speaker speaks loudly throughout his speech and softens his voice to attract listeners' attention.

Finally, Figure 2.12 shows the predicted probability prominence rating from Models 7-9 in relation to the interactions between acoustic cues and pitch accents.

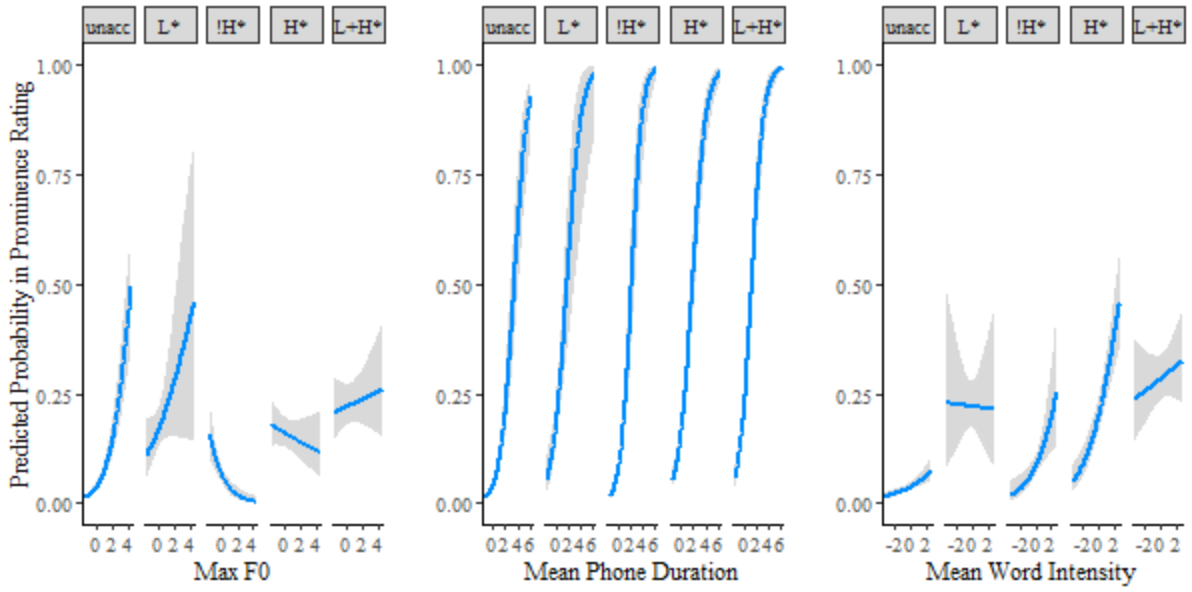


*Figure 2.12.* Predicted probability in prominence rating in relation to the interactions between pitch accents and acoustic cues. *Unacc* stands for unaccented words.

In Figure 2.12, the effects of max f0 (left panel) and mean word intensity (right panel) are different across accent types. This suggests that the same max f0 and mean word intensity influence perceived prominence differently depending on accent types. In comparison, mean phone duration surprisingly shows the similar effects across pitch accents.

Overall, listeners' perceived prominence is influenced by IS, pitch accents, acoustic cues, and their interaction. Listeners are more likely to perceive words as prominent if the words are associated with (1) new information on the IS hierarchy, (2) a higher-ranked pitch accent on the pitch accent hierarchy, and (3) enhanced max f0 and mean phone duration. The mean word intensity is not a significant correlate of perceived prominence with the speech style of the speaker. Listeners are also found to rely on acoustic cues differently depending on IS and pitch accents. They are more likely to mark words as prominent if they hear the same acoustic cues used for new information or higher-ranked pitch accents. Max f0 is found to be further weighted by listeners between referential vs. lexical meaning.

## 2.3. Discussion

This study investigates the question, "how is the perception of prosodic prominence influenced by expectation-driven (IS) and signal-driven factors (pitch accents, acoustic cues)?" In order to establish the relationship between expectation-driven and signal-driven factors, IS, pitch accents, and acoustic cues were analyzed in the speech of the speaker before the effects of those factors on perceived prominence were examined in this public speech.

### 2.3.1. How do speakers produce prosodic prominence in public speech style?

The IS of a word is an important factor that influences the speaker's use of prominence. Referential givenness is differentiated from lexical givenness in this speech sample in English in support of the RefLex scheme (Riester & Baumann, 2017). Referentially given information is likely to be unaccented while lexically given information tends to be accented. New information and contrastive expressions are phonologically marked, especially with H* and L+H*, and are produced with enhanced acoustic cues such as longer duration and higher intensity. In comparison, given information and accessible information are phonologically marked with !H* and L* and produced with relatively diminished acoustic cues.

There is scarce evidence in this sample of speech supporting the one-to-one mapping between IS and accent types. Pitch accents are probabilistically assigned to IS in line with previous findings from speech in Neapolitan Italian (Cangemi & Grice, 2016) and German (Baumann & Riester, 2013). All accent types occur across different IS labels, although certain accent types are indeed more frequently found with certain IS. Also, some words produced are prosodically salient regardless of there IS. Proper nouns, numerals, negations, and discourse

markers are usually produced with greater emphasis by the speaker. Prosodic prominence is also related with part of speech (Hirschberg, 1993; Sityaev, 2000) and discourse markers (Calhoun & Schweitzer, 2012).

Among acoustic cues, mean phone duration is the only reliable acoustic correlate with both IS and pitch accents. Max f0 is correlated with some pitch accents. Mean word intensity is not a strong correlate in this public speech. Duration and intensity are regarded as important correlates with prominent syllables in other speech styles (Kochanski et al., 2005) but in this sample of public speech style, the speaker speaks loudly throughout his narrative to address a large audience, thus cannot use intensity to encode other information such as IS and pitch accents. He rather softens his voice to produce contrastive focus and attract the attention of audience. This confirms that prominence is inherently relative. Speakers can achieve prominence by increasing acoustic cues in most speech styles. Speakers can also obtain similar effects by decreasing acoustic cues if they have to constantly speak loudly in a certain speech style. Therefore, if one attempts to model prominence using acoustic cues only, one should consider contextual factors such as speech style to determine which acoustic cues are significantly associated with prominence. For a more comprehensive review of prosody in context, see Cole (2015).

In this speech style, which is representative of a motivational and public speech style, acoustic cues and accenting patterns are found to be different from what has been reported from laboratory and conversational speech. Intensity is surprisingly not a strong predictor for the prosodic prominence. Pitch accents are assigned probabilistically to IS. To further explore the public speech style, comparisons are made between this speech sample (from one male) and

conversational speech from the Buckeye corpus (from eight males; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005). Figure 2.13 shows the occurrence of accent type (left panel), max f0 of a word (middle panel), and mean prominence rating of a word by non-expert listeners using RPT (right panel). ToBI annotation of the public speech and the speech from the Buckeye corpus was performed by the same labelers.

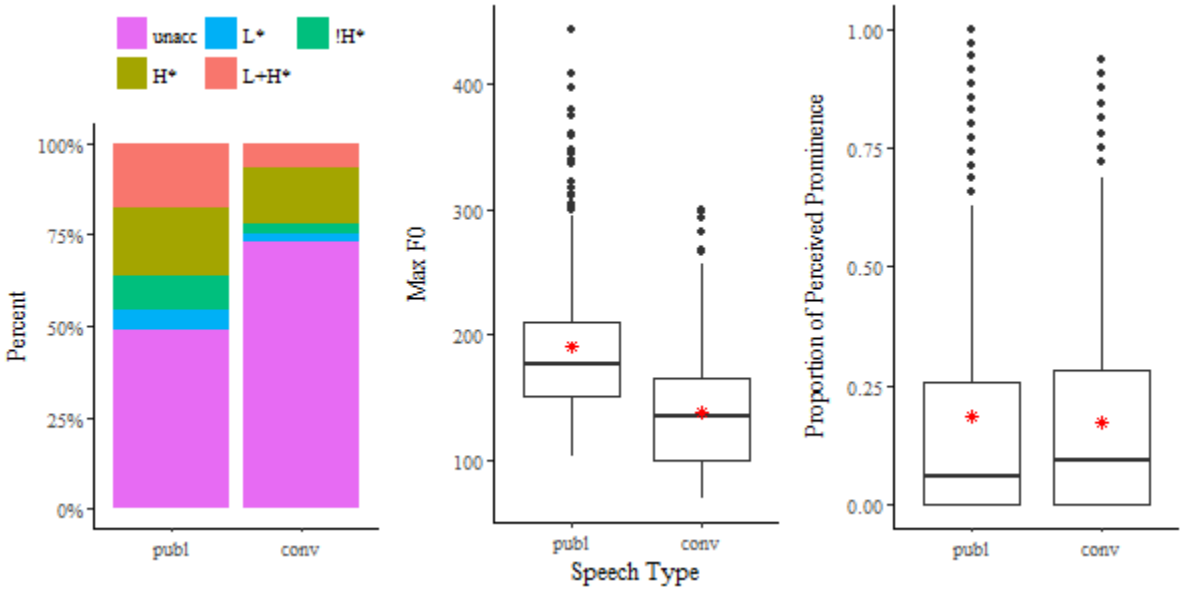


*Figure 2.13.* Comparison between the public speech (publ) and conversational speech (conv) from the Buckeye corpus.

In the left panel, words are accented more than half of the time in the public speech. L+H* is much more frequently used in the public speech than the conversational speech. In the middle panel, a higher max f0 is found in the public speech compared to the conversational speech. Surprisingly, in the right panel, the prominence rating of a word does not differ in the public and the conversational speech despite more occurrences of pitch accents and the higher max f0 used in the public speech. How do listeners consider various sources such as pitch accents, acoustic cues, IS, and speech style to judge prosodic prominence?

### 2.3.2. How do listeners perceive prosodic prominence in public speech style?

There are independent effects of IS, pitch accents, and acoustic cues on the perception of prominence in line with previous findings from a study of prominence perception in conversational speech in American English (Cole et al., 2010). Referential givenness is perceived differently from lexical expressions. Each level of discourse meaning contributes independently to prominence perception. New information or contrastive expressions are more likely to be perceived as prominent than given or accessible information. However, referentially new information is not always perceived as more prominent than referentially unique information in this speech sample.

As for signal-driven factors, pitch accents are perceived as prominent consistent with the pitch accent hierarchy (Hualde et al., 2016). L+H* is perceived as more salient than H*, which is in turn perceived as more prominent than !H*. The L* accent may be perceived as prominent if it is in nuclear position adjacent to a high boundary tone. Also, acoustic enhancement predicts perceived prominence. Intensity is not a strong cue for prominence in this sample of public speech, as the speaker speaks loudly throughout his talk and has less opportunity to use intensity to signal IS and accent distinction. Listeners seem to calibrate to this speech style and weigh intensity less as a cue for prominence.

In this study, there are parallel patterns between perception and production. The speaker cannot use intensity to encode most IS and pitch accents. Listeners also weigh intensity less to rate prominence. Words referring to places and names are marked with longer durations by the speaker, which seems to influence the overall enhanced prominence rating of this group of words by listeners. However, there is not always direct pairing between production and perception. The

words carrying lexically given information tend to be accented by the speaker, but the same words are not likely to be perceived as prominent by listeners. In Figure 2.13, the speaker's extensive use of pitch accents and max f0 seem to be considered as decorative and get filtered out by listeners for their prominence rating. Listeners must "normalize" the speaker's speech style, which needs further investigation in a future study.

Interaction between (1) acoustic cues and IS, and (2) acoustic cues and pitch accents also significantly contribute to the perception of prominence. Listeners weigh acoustic cues differently depending on givenness vs. newness as well as referential status vs. lexical status of a word. Also, max f0 and mean word intensity are perceived differently depending on types of pitch accents. This suggests that listeners perceive acoustic cues as mediated by IS and pitch accent type. Moreover, the acoustic cues within the same phonological feature are found to be perceived differently. Within the same phonological feature, more enhanced acoustic cues are more likely to be perceived as prominent by listeners.

This study informs us how prosodic prominence is produced and perceived in relation to IS, pitch accents and acoustic cues in public speech style. Prominence arises from multiple sources (Watson, 2010). Givenness, discourse meaning (i.e., referential or lexical meaning), and speech style (i.e. phonetic and accenting distinction) delivered by the speaker are confirmed to contribute to listeners' perception of prominence in this public speech style. Part of speech and discourse markers are observed as other sources of the perception of prominence. This study calls for the consideration of these sources, especially referential vs. lexical meanings and speech style in the analysis of prosodic prominence.

# 3. EXEMPLAR ENCODING OF INTONATION IN IMITATED SPEECH

This study turns to analyses of imitated speech to investigate the domain of f0 encoding, examining the similarity of an imitated utterance to a stimulus produced by different speakers in terms of differences in the shapes of f0 contours. Six prosodic domains (intermediate phrase, accentable word, pitch accent, foot, stressed syllable, and syllable) are examined from the imitated speech using Generalized Additive Mixed Model (GAMM; Wood, 2017), which allows us to model the time-varying patterns of f0 contour over the prosodic domains. The model comparison across the six domains is made based on goodness-of-fit (deviance explained value) evaluated by GAMMs. The hypothesis is that the domain in which imitated and stimulus f0 contours are the most similar corresponds to the target of cognitive encoding of f0. This study does not model f0 contours of individual utterances, but using GAMMs to model the similarity between two utterances. The goal of this study is not to measure how accurate any individual imitation is to its corresponding stimulus, but rather to determine how similarity of f0 contours should be evaluated. The present study addresses the question: what is the nature of the representation of sentence intonation that is the target of imitation in the mind of the imitator?

## 3.1. Method

### 3.1.1. Materials

This study examines f0 contour in the Illinois Imitation Corpus (Cole, Hualde, Eager, & Mahrt, 2015). Thirty-three American English speakers from University of Illinois (10 males, 23 females) participated in the experiment and imitated aural sentence stimuli produced by a female

American English speaker. The stimuli were 12 meaningful declarative sentences, 7-13 words in length, as shown in Table 3.1. Each stimulus sentence started with complex subject noun phrase, e.g., *the adversarial prosecutor,* shown in square bracket.

Table 3.1.

*Sentence Stimuli in Illinois Imitation Corpus.*

| Number | Stimuli | Word Count |
|---|---|---|
| 1 | [The realistic story] included a few untrue elements about George Clooney's hometown. | 12 |
| 2 | [The systematic tutors] always give clear instructions that even a beginner could follow. | 13 |
| 3 | [The Unitarian journalist] tried to be impartial in political disputes. | 10 |
| 4 | [The adversarial prosecutor] was not successful in making friends at the office. | 12 |
| 5 | [The professorial fashion] was never even noticed by most of the students. | 12 |
| 6 | [The inspirational speech] bored Alice out of her mind. | 9 |
| 7 | [The regulation of child labor] did not please everyone. | 9 |
| 8 | [The editorial column] reflected mainstream political views. | 7 |
| 9 | [His categorical stance] on protecting endangered animals admits no counter-arguments. | 10 |
| 10 | [The supplementary details] were unnecessary and made for a boring read. | 11 |
| 11 | [The automatic potato peeler] was too expensive for Johnny to buy. | 11 |
| 12 | [The disappointing performance] was depressing for Sue and the whole group. | 11 |

The complex subject noun phrase was produced by the stimulus speaker in one of three prosodic patterns as shown in Table 3.2. Accented syllables are noted with capital letters.

Table 3.2.

*Accent Patterns of Subject Noun Phrases in Stimuli.*

| Accent Pattern | Description | Example |
|---|---|---|
| Primary | Accent on the primary stress syllable | The adverSArial PROsecutor |
| Early high | Accent on the secondary stress syllable | The ADversarial PROsecutor |
| Unaccented | No accent on either primary or secondary syllables | The adversarial PROsecutor |

The first pattern is the "primary" accent pattern with a pitch accent H* on the primary stress syllable on *adversarial*. The second pattern has the "early high" accent pattern with a pitch accent H* on the secondary stress syllable of *adversarial*, which is distinct from the more typical accent pattern (adverSArial) that locates the pitch accent on the primary stress syllable. The third pattern is the "unaccented" pattern with no pitch accent produced for the first content word *adversarial*. Across all three patterns, the H* pitch accent was produced for the primary stress on the second content word (PROsecutor). There was a prosodic phrase break at the end of the subject noun phrase. Each prosodic pattern was produced on four sentences, and the assignment of prosodic pattern to sentence item was counterbalanced across three participant groups.

Participants had to reproduce the entire sentence after hearing it, but only the complex subject noun phrases of the stimuli were analyzed in this study. They were instructed to repeat what they heard in the manner the stimulus speaker said it. The instructions did not explicitly mention prosody, intonation or sentence melody. Participants advanced through the trials at a self-selected pace, reproducing each stimulus immediately after three successive aural presentations of the stimulus. There was no orthographic presentation of the stimulus during the experiment. Participants produced five imitations with incorrect words or long pause, and these items were excluded in the analyses of this study.

### 3.1.2. Prosodic domains

This study examines the similarity of imitated f0 contours and their corresponding stimulus in six analyses that differ in the scope of the prosodic domain in which f0 is modeled. The six domains are listed in Table 3.3 in decreasing order of phonological scope and specificity: intermediate phrase (ip), accentable word, pitch accent, foot, stressed syllable, and syllable.

Table 3.3.

*Six Prosodic Domains in Decreasing Order of Phonological Scope and Specificity.*

| Number | Domain | Description |
|---|---|---|
| 1 | Intermediate phrase | Entire subject noun phrase |
| 2 | Accentable word | A content word and the preceding function words |
| 3 | Pitch accent | A pitch-accented syllable and the following unaccented syllables |
| 4 | Foot | A stressed syllable and the following unstressed syllables |
| 5 | Stressed syllable | A stressed syllable with a 0.5-scond fixed window centered on the stressed syllable |
| 6 | Syllable | Syllable |

The subject noun phrases in the imitated utterances were manually segmented into these six domains using Praat (Boersma & Weenink, 2018) as shown in Figure 3.1. The label *NA* is assigned to intervals that are outside the domains of analyses. The label *OV* indicates overlaps between domains.

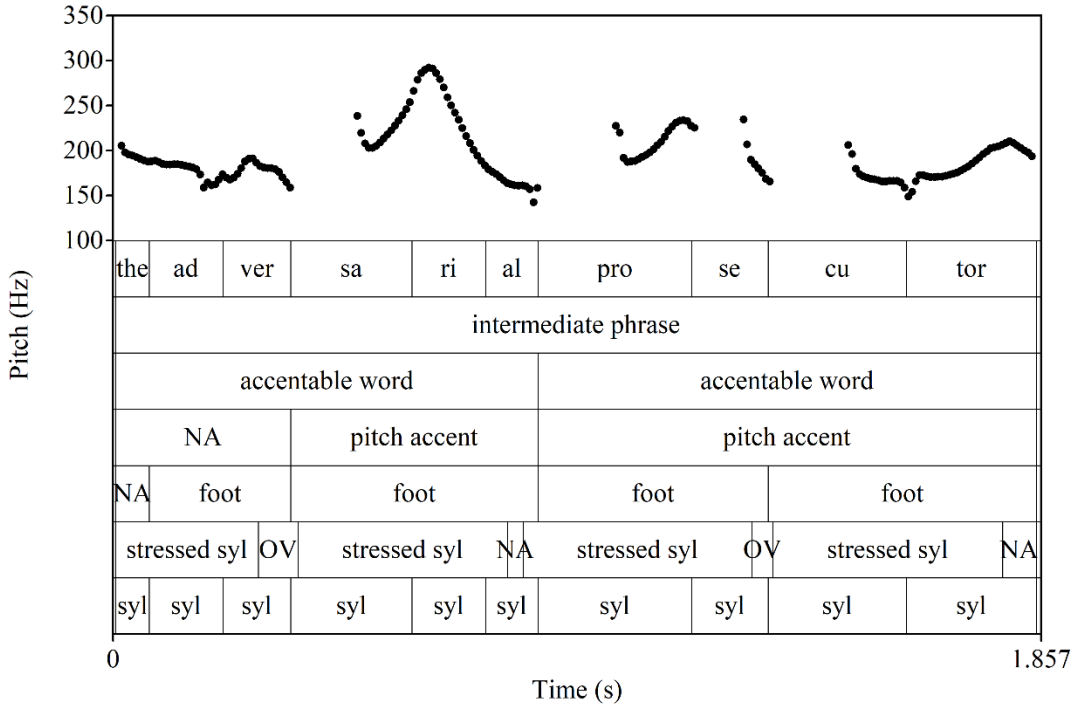


*Figure 3.1.* Six domains in decreasing order of phonological scope and specificity. The largest domain is the intermediate phrase (top) and the smallest domain is the syllable (syl; bottom). *NA* indicates the regions outside the domains of analyses; *OV* stands for the overlapping regions between domains.

The intermediate phrase domain covers the entire subject noun phrase. The accentable word domain contains a content word and the preceding function words if any exist. The pitch accent domain starts from the pitch-accented syllable up to the next-rightmost pitch-accented syllable. The domain consists of an accented syllable and any following unaccented syllables. The rightmost domain of subject noun phrases contains a boundary tone. The foot domain starts from primary or secondary stress syllables up to the next-rightmost stressed syllable. The stressed syllable domain contains primary or secondary stress syllables analyzed with a 0.5-second fixed window centered on the stressed syllable. Successive stressed syllable domains sometime overlap one another. The syllable domain covers each syllable.

The six domains show different coverage of the f0 contour over a subject noun phrase in Figure 3.1. Three domains cover the entire subject noun phrase while the other three domains do not. The intermediate phrase domain covers the entire subject noun phrase. Also, the accentable word domain and the syllable domain cover the entire subject noun phrase with smaller window sizes than that of the intermediate phrase domain. In comparison, the pitch accent domain, the foot domain, and the stressed syllable domain leave out some portion of the subject noun phrase. The temporal extent of the excluded intervals varies depending on the lexical content of the sentence (see Table 3.1). The pitch accent domain leaves out the syllables preceding the first pitch-accented syllable. The foot domain does not cover unstressed syllables preceding the first stressed syllable. The stressed syllable domain leaves out some unstressed syllables.

The six domains capture different complexity and granularity of the f0 contour. Figure 3.2 shows the f0 contour represented with 30 f0 points in each domain. The f0 contour of the leftmost interval for each domain in Figure 3.1 is shown in decreasing order in Figure 3.2.

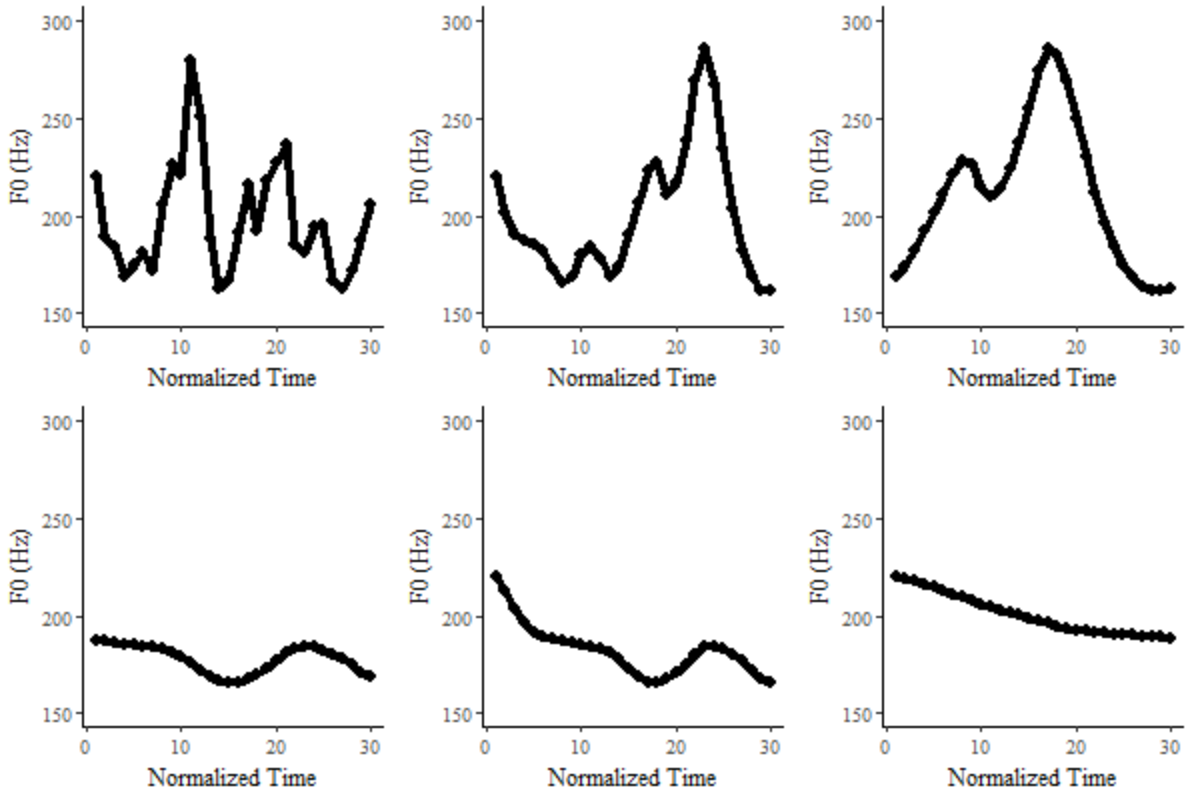


*Figure 3.2.* F0 contours represented with 30 f0 points for each domain in decreasing order. The intermediate phrase domain is presented at the top left and the syllable domain is located at the bottom right. The f0 contours of the leftmost item for each domain in Figure 3.1 are shown here. Each domain captures different complexity and granularity of the f0 contour.

The f0 contour for the intermediate phrase domain (top left panel) is the most complex, as it is the contour over the entire subject noun phrase. As this complex contour is represented with 30 f0 points only, the intermediate phrase domain is considered to be the most coarse-grained representation of the f0 contour among the six analyses presented here. In contrast, the f0 contour for the syllable domain (bottom right panel) is the simplest, as it is the contour over only one syllable. As this simple contour is represented with 30 f0 points, the syllable domain is considered to be the most fine-grained representation of the f0 contour among those presented here. The f0 contours for the accentable word domain (top middle panel) and the pitch accent domain (top right panel) are relatively complex, as they show the contours over large portions of

the entire subject noun phrase. The representations of the f0 contours are relatively coarse-grained, because the relatively complex contours are represented with 30 f0 points only. The f0 contours for the foot domain (bottom left panel) and the stressed syllable domain (bottom middle panel) are relatively simple, as they show the contours over small portions of subject noun phrase. The representations of the f0 contours are relatively fine-gained because the relatively simple contours are represented with 30 f0 points.

The coverage, complexity, and granularity of the f0 contour in six domains are summarized in Table 3.4.

Table 3.4.

*Coverage, Complexity, and Granularity of F0 Contour for Each Domain*

| | Intermediate Phrase | Accentable Word | Pitch Accent | Foot | Stressed Syllable | Syllable |
|---|---|---|---|---|---|---|
| Coverage | 100% | 100% | 70% | 80% | 60% | 100% |
| Complexity | High | Relatively high | Relatively high | Relatively low | Relatively low | Low |
| Granularity | Coarse | Relatively coarse | Relatively coarse | Relatively fine | Relatively fine | Fine |

### 3.1.3. Analyses

F0 values were processed in three steps using ProsodyPro (Xu, 2013). First, linear interpolation[2] was applied to fill in missing f0 values from voiceless regions in order to obtain continuous f0 contours over the entire subject noun phrase. Next, triangular smoothing was used to control micro-perturbation and smooth the entire f0 contour. Finally, time-normalization was

[2] The linear interpolation is a conventional method to minimally influence the entire shape of f0 contour. The interpolation method was identical for the stimulus and imitated contours, and the interpolated portion of the f0 was not considered to influence the analysis of similarity between the stimulus and imitated contours.

performed to obtain the same number of f0 samples regardless of the actual temporal duration of the modeling domain. For this, 30 f0 points were obtained at equal distance in each domain. Thirty was determined to be the optimal number of f0 samples for this study, based on the temporal extent of the largest domain (ip). Care was taken not to obtain too few from the largest domain (ip) or too many observations from the smallest domain (syllable). Finally, f0 points were converted from the Hertz scale to the ERB-rate scale to approximate the frequency selectivity of the auditory system (Hermes & Van Gestel, 1991).

The similarity of two f0 contours were examined using GAMMs. GAMM is the generalized linear model[3] with a sum of smooth functions of covariates, which allows us to model non-linear time-series data. An f0 contour is a non-linear time-series datum, as an f0 contour consists of f0 points at each time step. An f0 contour is modeled with the smoothing function optimized by GAMM.

In the GAMM models presented here, the dependent variable (DV) is the distance (on the ERB scale) between the stimulus and an imitated f0 contour. Figure 3.3 shows a hypothetical perfect imitation (left panel) and an actual imperfect imitation from our imitation corpus (right panel).

[3] For linear models, it is assumed that errors are independently and identically distributed (i.i.d.). F0 contour is the set of f0 points produced as a continuous event by a speaker and the f0 point at *t* may be autocorrelated with the f0 point at *t*+1. I performed AR(1) to model the autocorrelation in the data, but due to the nested structure of the data (i.e., subjects producing items which consist of normalized time steps) I did not find significant effects of AR(1) on the model. The Quantile Regression Model is a non-parametric regression model based on the estimation of either the median or quantile of DV and does not assume i.i.d. I ran Quantile Regression Models using qgam (Fasiolo, Goude, Nedellec, & Wood, 2017) in R and found similar results as the GAMM results, which are reported in Appendix A. I appreciate valuable suggestions from Harald Baayen.

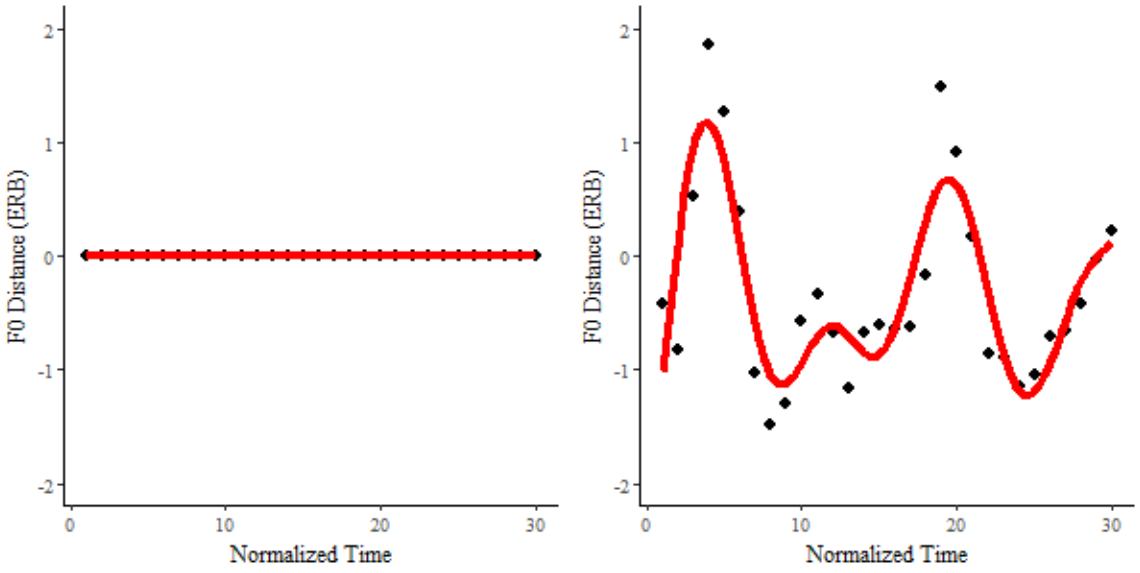


*Figure 3.3.* Perfect vs. imperfect imitation. The left panel shows the hypothetical perfect imitation, where there is no distance between the stimulus and imitated f0 contours, modeled by GAMM with a flat line. The right panel shows the real, imperfect imitation, where there is distance between the stimulus and imitated f0 contours, modeled by GAMM with a wiggly line.

If the stimulus is perfectly imitated, the stimulus and imitated f0 contours would be the same; thus, the distance between the two contours would be zero. This would be modeled by GAMM with a flat line. If the stimulus is not perfectly imitated, the stimulus and imitated f0 contours would not be the same; thus, the distance between the two contours would not be zero. This would be modeled by GAMM with a wiggly line. Figure 3.4 shows the f0 contours produced by the stimulus (top panel) and two different speakers from the corpus (middle and bottom panels).

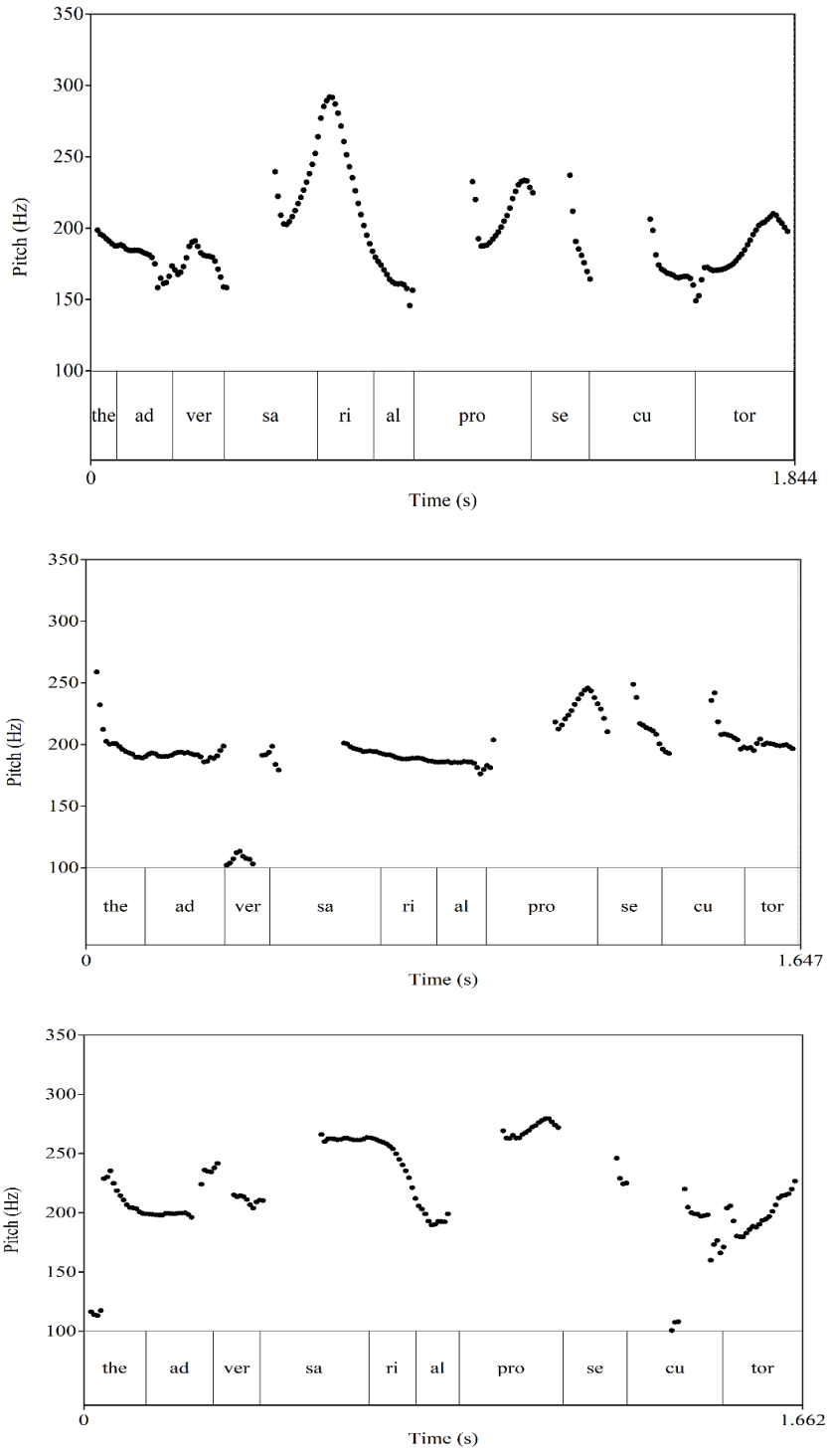


*Figure 3.4.* F0 contours produced by the stimulus (top panel) and two imitators (middle, bottom panels).

The stimulus f0 contour contains two pitch accents—that is, the first pitch accent at the syllable *sa* and the second pitch accent at the syllable *pro*. The first imitated f0 contour (middle panel) shows the omission of the first pitch accent. The second imitated f0 contour (bottom panel) shows an undershooting of the first pitch accent at the syllable *sa* and an overshooting of the second pitch accent at the syllable *pro*.

An imperfect imitation may result from several possible cases. If imitators undershoot the pitch accent of the stimulus or omit the pitch accent produced by the stimulus, these cases would result in positive values in the distance between the stimulus and imitated f0 contours, given that the distance is obtained by subtracting the imitated f0 from stimulus f0. If imitators overshoot the pitch accent of the stimulus or insert a pitch accent that is not produced by the stimulus, these cases would result in negative values in the distance between the stimulus and imitated f0 contours, given that the distance is obtained by subtracting the imitated f0 from stimulus f0.

In the GAMM models, two categorical factors and one smooth factor were submitted as fixed effects. The two categorical factors are the accent patterns produced by the stimulus and the gender of the participant. Effects of categorical factors would be seen as a shift of DV up or down on the *y*-axis. In the model of this study, this would be an overall increase or decrease in the f0 distance, relative to the sample mean. The accent factor is included because the sentence stimulus was produced by the model speaker with three different pitch accent patterns, and speakers may be more accurate when imitating one particular pitch accent pattern compared to another. The gender factor is included because of the intrinsic pitch differences between male and female speakers. Also, one factor is entered as a smooth term in the model. The smooth term in the model of this study is an interaction between normalized time, accent, and gender, to allow

for different wiggly patterns of f0 distance across time points, depending on the accent and gender of the participants. The smooth term allows us to model a variable effect on the DV for different values of the predictor, which can be visualized as a wiggly line which represents the non-linear function relating the predictor to the DV. In the model of this study, each normalized time point predicts the f0 distance, but f0 distance is allowed to vary in a non-linear pattern across the series of 30 time steps. Two factors were submitted as random effects, the intercept for subject and item. The following model was run using the mgcv package (Wood, 2017) in R (R Core Team, 2018): distance ~ accent + gender + s(normalizedtime, by = interaction(accent, gender, *k* = 10) + s(subject, bs = "re") + s(item, bs = "re").

Six GAMM models were run: one model for each domain. The same parameters were submitted with a different DV, that is the distance measured between the imitated and the stimulus f0 over 30 normalized-time points in each of the six domains. The six GAMM models were compared for goodness-of-fit using the deviance explained value, that is the measure of the proportion of variance that the model accounts for. The best model is the one with the highest deviance explained value, since the number of parameters is the same across all six models, and that model represents the domain in which the stimulus best predicts the imitated f0, allowing for systematic, non-linear divergences across the interval, and by phonological accent pattern and gender of participants.

### 3.1.4. Predictions

Two predictions are proposed, one supporting the abstractionist model and one supporting the exemplar model:

The first prediction supports the abstractionist model. The domains covering the regions of intonational features (i.e., the pitch accent domain, the foot domain, the stressed syllable domain) will show higher deviance explained values than domains covering the entire regions (i.e., the intermediate phrase domain, the accentable word domain, the syllable domain). If encoding privileges f0 over phonologically specified regions, the f0 contour over these regions would be more accurate than imitation over regions that lack a phonologically specified tone, and thus lack an explicit pitch target for imitation. Under the theory of sparse encoding of intonational features, a model that excludes regions that are not phonologically specified would yield lower variance (i.e., higher deviance explained value).

The second prediction is in support of the exemplar model. All six domains will show similarly high deviance explained values. If the exemplar model is true, the f0 contour over the entire region of the utterances is encoded, capturing all the perceived details of the f0 contour, uniformly across the utterance. The target of imitation is fine-grained under this hypothesis, and the similarity of the imitation to the stimulus should be accurate to the same degree whether measured in a small interval such as the syllable, or a longer interval such as the prosodic phrase. No matter the size of the domain in which f0 is measured, there should be low variance (i.e., high deviance explained value) across all domains, and little or no difference in deviance explained values between the domains capturing the entire region (i.e., the intermediate phrase domain, the accentable word domain, the syllable domain) and the domains capturing the phonologically specified regions (i.e., the pitch accent domain, the foot domain, the stressed syllable domain).

## 3.2. Results

The results show that all the six domains show high deviance explained values in support of the exemplar encoding. The deviance explained values for the six domains range from 75.7% to 79.2%. The intermediate phrase domain shows the highest deviance explained values (79.2%) across the six domains. The model summary of the six domains is provided in Table 3.5.

Table 3.5.

*GAMM Summary Table for Six Domains.*

| | Intermediate Phrase | | | | Accentable Word | | | | Pitch Accent | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Deviance explained** | 79.2 | | | | 76.7 | | | | 77.1 | | | |
| ***R*²** | .79 | | | | .77 | | | | .77 | | | |
| ***n*** | 11730 | | | | 23460 | | | | 19530 | | | |
| *Parametric coefficients* | | | | | | | | | | | | |
| | est. | *SE* | *t* | *p* | est. | *SE* | *t* | *p* | est. | *SE* | *t* | *p* |
| **Intercept** | -.18 | .07 | -2.73 | .01 | -.15 | .07 | -2.13 | .03 | -.13 | .07 | -1.84 | .07 |
| **Accent** | | | | | | | | | | | | |
| primary | -.10 | .01 | -8.12 | <.01 | -.11 | .01 | -12.45 | <.01 | -.01 | .06 | -.08 | .93 |
| unaccented | -.25 | .01 | -20.51 | <.01 | -.27 | .01 | -30.23 | <.01 | .02 | .07 | .26 | .79 |
| **Gender** | | | | | | | | | | | | |
| male | 1.99 | .12 | 16.97 | <.01 | 1.99 | .11 | 17.27 | <.01 | 1.99 | .11 | 18.28 | <.01 |
| *Approximate significance of smooth terms* | | | | | | | | | | | | |
| | EDF | *Df* | *F* | *p* | EDF | *df* | *F* | *p* | EDF | *df* | *F* | *p* |
| **s(normalizedtime)** | | | | | | | | | | | | |
| earlyhigh:female | 8.87 | 9.00 | 178.70 | <.01 | 8.62 | 8.96 | 71.33 | <.01 | 7.90 | 8.68 | 78.90 | <.01 |
| primary:female | 8.84 | 8.99 | 83.16 | <.01 | 8.24 | 8.84 | 105.77 | <.01 | 7.35 | 8.34 | 66.81 | <.01 |
| unaccented:female | 7.52 | 8.45 | 121.84 | <.01 | 4.82 | 5.89 | 44.05 | <.01 | 5.35 | 6.48 | 49.44 | <.01 |
| earlyhigh:male | 8.75 | 8.98 | 83.48 | <.01 | 8.20 | 8.83 | 60.76 | <.01 | 7.57 | 8.49 | 75.71 | <.01 |
| primary:male | 8.75 | 8.98 | 46.61 | <.01 | 7.83 | 8.64 | 50.12 | <.01 | 6.36 | 7.51 | 45.30 | <.01 |
| unaccented:male | 7.29 | 8.29 | 45.40 | <.01 | 2.25 | 2.81 | 8.31 | <.01 | 5.46 | 6.60 | 15.58 | <.01 |
| **s(subject)** | 30.74 | 31.00 | 119.77 | <.01 | 30.86 | 31.00 | 218.87 | <.01 | 30.40 | 31.00 | 253.37 | .03 |
| s(**item)** | 10.28 | 11.00 | 16.47 | <.01 | 22.75 | 23.00 | 93.06 | <.01 | 54.19 | 57.00 | 60.89 | .24 |

| | Foot | | | | Stressed Syllable | | | | Syllable | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Deviance explained** | 75.7 | | | | 77.7 | | | | 77.5 | | | |
| ***R*²** | .76 | | | | .78 | | | | .78 | | | |
| ***N*** | 39120 | | | | 39120 | | | | 95850 | | | |
| *Parametric coefficients* | | | | | | | | | | | | |
| | est. | *SE* | *t* | *p* | est. | *SE* | *t* | *p* | est. | *SE* | *t* | *p* |
| **Intercept** | -.14 | .07 | -1.87 | .06 | -.17 | .07 | -2.42 | .02 | -.22 | .07 | -3.06 | .01 |
| **Accent** | | | | | | | | | | | | |
| primary | -.19 | .01 | -27.62 | <.01 | -.17 | .01 | -25.80 | <.01 | -.12 | .01 | -27.09 | <.01 |
| unaccented | -.36 | .01 | -51.14 | <.01 | -.35 | .01 | -52.25 | <.01 | -.31 | .01 | -71.66 | <.01 |
| **Gender** | | | | | | | | | | | | |
| male | 1.99 | .12 | 16.69 | <.01 | 2.02 | .12 | 16.74 | <.01 | 2.01 | .12 | 16.53 | <.01 |
| *Approximate significance of smooth terms* | | | | | | | | | | | | |
| | EDF | *df* | *F* | *p* | EDF | *df* | *F* | *p* | EDF | *df* | *F* | *p* |
| **s(normalizedtime)** | | | | | | | | | | | | |
| earlyhigh:female | 6.34 | 7.49 | 13.78 | <.01 | 8.08 | 8.77 | 42.58 | <.01 | 3.76 | 4.65 | 6.47 | <.01 |
| primary:female | 6.09 | 7.25 | 39.52 | <.01 | 7.83 | 8.64 | 92.45 | <.01 | 3.50 | 4.34 | 5.87 | <.01 |
| unaccented:female | 5.03 | 6.13 | 26.59 | <.01 | 6.06 | 7.22 | 24.77 | <.01 | 1.00 | 1.01 | 9.57 | .01 |
| earlyhigh:male | 5.63 | 6.78 | 11.27 | <.01 | 6.44 | 7.59 | 25.63 | <.01 | 2.36 | 2.94 | 2.14 | .09 |
| primary:male | 5.41 | 6.54 | 23.62 | <.01 | 7.11 | 8.16 | 57.11 | <.01 | 2.58 | 3.21 | 3.40 | .02 |
| unaccented:male | 5.53 | 6.67 | 9.01 | <.01 | 5.61 | 6.76 | 10.03 | <.01 | 2.17 | 2.71 | 2.65 | .08 |
| **s(subject)** | 30.92 | 31.00 | 373.44 | <.01 | 30.92 | 31.00 | 415.41 | <.01 | 30.97 | 31.00 | 1049.98 | <.01 |
| s(**item)** | 38.65 | 39.00 | 116.06 | <.01 | 38.47 | 39.00 | 80.76 | <.01 | 96.42 | 97.00 | 196.81 | <.01 |

### 3.3. Discussion

All the domains show a good model fit, including the smallest domain, the syllable, which captures the temporal dynamics of f0 in the finest detail. In fact, the syllable model fares nearly as well as models with larger domains and coarser-grained representations of f0. This finding lends support for a theory of the cognitive encoding of f0 that represents temporal dynamics in fine detail. In comparison with the pitch accent domain, the foot domain, and the stressed syllable domain, the intermediate phrase domain does not leave out any portion of the f0 contour. The improved model fit for the intermediate phrase domain suggests that all regions of the f0 contour, even phonologically unspecified regions, are modeled as the target for imitation. In comparison with the syllable domain, the intermediate phrase domain models the f0 contour holistically, as a continuous, time-varying pattern over the whole phrase. The syllable domain models each syllable independently and does not capture the continuity of f0 contours in adjacent syllables.

Examining the imitation results in more detail, I found that there is variation in imitations of stimuli with different pitch accent patterns. All the three accent patterns are not perfectly imitated, but further qualitative analyses show that speakers more accurately imitate the early high pattern than other patterns. Speakers sometimes even replace another pattern with the early high pattern. Goldinger (1998) proposes that speakers' linguistic knowledge (e.g., word frequency) influences their degrees of imitation. In the same study, he finds that speakers imitate low-frequency words more faithfully than high-frequency words. In the current study, the early high pattern is a less-frequent pitch accent pattern, which is usually observed in radio news speech (Cole et al., 2015), and is found to be more accurately imitated by most speakers. This is

in line with previous studies' findings that speakers' linguistic knowledge and experience influence later production of speech (Goldinger, 1998; Nye & Fowler, 2003).

There is also variation in imitation accuracy by gender. Figure 3.5 shows two panels, the left panel for the variation in imitation accuracy by gender and the right panel for the variation in imitation accuracy among individuals, from the GAMM model with the intermediate phrase domain.

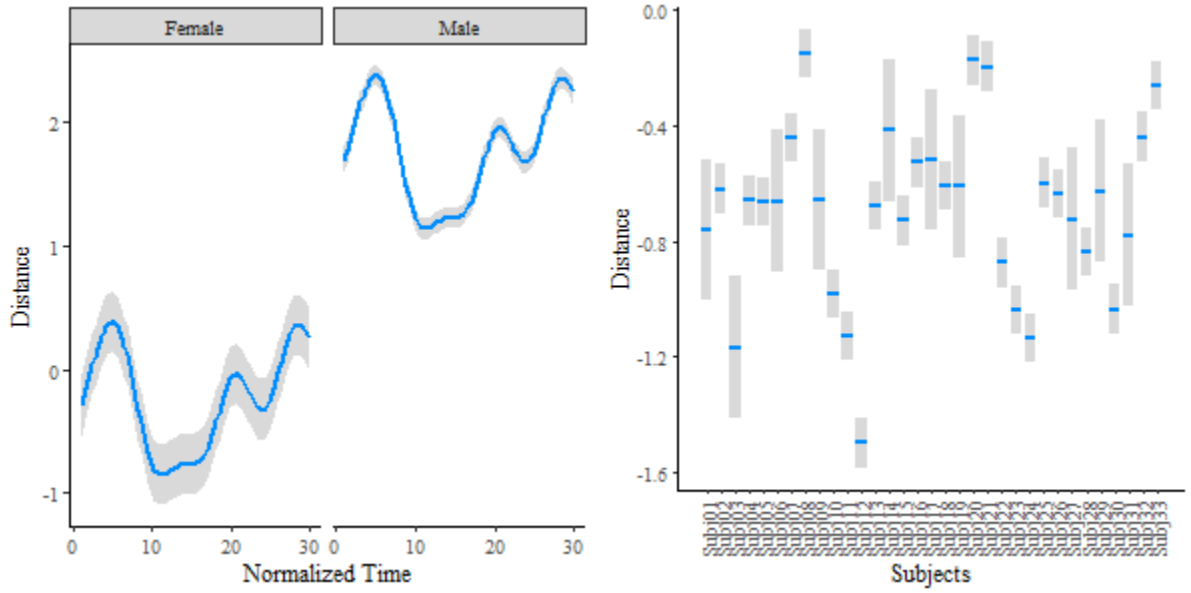


*Figure 3.5.* Variation in imitation accuracy by gender (left panel) and among individuals (right panel) from the GAMM model of imitation in the intermediate phrase domain.

In the left panel of Figure 3.5, the differences in the intercept is due to inherent pitch differences between males and females. The similarity in the shape suggests that males and females do not differ in imitating pitch accent patterns and make similar types of mistakes (e.g., undershooting, accent insertion). In the right panel, the individual speakers who exhibit a wider confidence interval are found to be all males. This suggests that males tend to produce spurious pitch accents inconsistently. Previous studies show conflicting evidence on the relationship between gender and imitation. Namy, Nygaard, and Sauerteig (2002) find that females converge to their model

speaker more than males in shadowing tasks. In comparison, Pardo (2006) finds that males converge more to their conversational partners than female speakers in task-oriented speech. The present study shows that males and females make similar types of mistakes in imitating speech, and males tend to converge less to the model speaker, and make mistakes more inconsistently than females.

Variation in imitation is also found across sentence stimuli (items). Figure 3.6 shows variation in imitation accuracy across items in the analysis of the intermediate phrase domain.

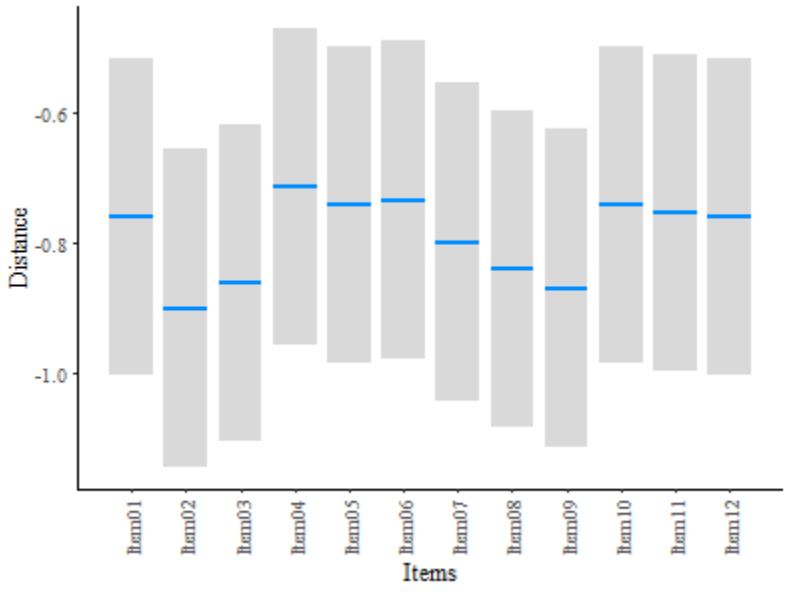


*Figure 3.6.* Variation in imitation accuracy across sentence stimuli (items) from the GAMM model of imitation in the intermediate phrase domain.

In Figure 3.6, each item shows a different intercept, and items 2, 3, 7, 8, and 9 show lower intercepts than the others. We might expect that imitation accuracy will be higher for lexically shorter sentences compared with longer sentences, but this was not the case. The items with lower intercepts are not always the longest utterances (items 1, 2, 4, and 5, with 12-13 words) nor are they always the utterances in which imitators make the most mistakes (items 5, 7, 9, and

10, with five mistakes across speakers). It is possible that the variation in imitation accuracy across items is related to word frequency or the semantic content of utterances.

There are a couple of factors which may influence exemplar encoding of the f0 contour. This study examines prosodic phrases produced with one or two H* pitch accents and the following L- boundary tone. Different types of pitch accents may draw more attention from listeners to their phonologically specified regions and inhibit the exemplar encoding of the f0 contour. For example, L+H* is more likely to be judged as prominent than H* by non-expert English listeners (Hualde et al., 2016). The effects of pitch accent type on the exemplar encoding of f0 are unknown and need to be examined in a future study.

Another factor that may influence the degree of exemplar encoding of the f0 contour is speech style or communication setting. This study examines the speech collected in a laboratory using an imitation paradigm. One may argue that speakers might have been more attentive to speech heard in a laboratory compared to speech heard in a live interaction with an interlocutor. A more attentive listener may perform a more detailed encoding of heard speech. However, prior studies show evidence of phonetic imitation of speech in everyday conversation as well as in task-oriented speech (Giles, Coupland, & Coupland, 1991; Levitan et al., 2012; Pardo, 2006). This suggests that speakers do pay attention to the phonetic details of speech under different communication settings, and are able to reproduce those details similarly. This is evidence of exemplar encoding.

The present study supports the exemplar model with evidence that speakers imitate f0 contours that extend over a prosodic phrase at a level of phonetic detail that is relatively consistent across the phrase. Previous studies have shown evidence that intonation is imitated in

its phonetic detail through evidence of summary or point measures of f0 (e.g., f0 peak and mean f0; D'Imperio et al., 2014; German, 2012). The present study examines f0 in a larger phonological domain, which comprises phonologically specified and unspecified regions, using measurements that capture the detailed f0 contour (i.e., f0 shape). The findings from this study point to the prosodic phrase as the domain that best captures the cognitive encoding of the target values for an f0 contour. The phrase is also the ideal domain for modeling convergence of f0 contours between speakers.

# 4. CONCLUSION

In this dissertation, I addressed two questions on prosodic prominence in English: (1) How is the prosodic prominence related to information status, pitch accents, and acoustic cues? (2) What is the ideal domain of f0 encoding?

The first study investigates the production and perception of prosodic prominence as a function of expectation-driven (IS) and signal-driven factors (pitch accents, acoustic cues) in an intact public speech in American English. The speaker is found to encode IS and pitch accents relying on f0 and duration, not intensity. He makes distinctions in accent assignments between the referential and lexical meaning of a word, but he associates accent types freely with IS, favoring L+H* in all IS categories. Listeners are found to perceive prosodic prominence in relation to IS, pitch accents, and acoustic cues, as well as their interaction. Acoustic cues are perceived differently depending on the givenness/newness and the referential/lexical statuses of a word. These findings contribute to previous studies showing the probabilistic relationship between accent assignment and discourse meaning, and the effects of interaction between expectation-driven and signal-driven factors on perception of prominence. This study calls for the consideration of a different speech style, and referential and lexical differentiation in discourse meaning in the research of prosodic prominence.

The second study investigates the phonological interval that defines the domain of cognitive encoding of intonational phonetic detail using imitated speech in American English. This study examines the similarity of f0 contours between imitated sentences and their stimuli

over domains of varying sizes and prosodic statuses, from the syllable to the prosodic phrase. Results show evidence for the cognitive encoding of the phonetically detailed f0 contour over an entire prosodic phrase (ip). The findings do not support a model of encoding that excludes phonologically unspecified regions. This study contributes to previous research showing speakers' adaptations to fine phonetic detail and calls for an extension of exemplar models to include phonetically detailed representations of f0 patterns.

## REFERENCES


Andruski, J. E., & Costello, J. (2004). Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association, 34*(2), 125–140.

Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology, 26*(1), 43–74.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231–248.

Bard, E. G., & Aylett, M. P. (1999, August). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In *Proceedings of the XIVth international congress of phonetic sciences*, San Francisco, pp. 1753–1756.

Bates, D., Mächler, M., Bolker, B., & Walkers, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics, 38*(10), 1636–1657.

Baumann, S., & Riester, A. (2012). Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (pp. 119–162). Berlin, New York: Mouton De Gruyter.

Baumann, S., & Riester, A. (2013). Coreference, lexical givenness and prosody in German. *Lingua, 136,* 16–37.

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*(2), 1001–1024.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*(4), 365–391.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication, 14*(4), 359–375.

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word, 14*(2–3), 109–149.

Boersma, P. & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.39, Retrieved from http://www.praat.org/

Bosshardt, H. G., Sappok, C., Knipschild, M., & Hölscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. *Journal of Psycholinguistic Research*, *26*(4), 425–448.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098.

Bybee, J. (2003). *Phonology and language use* (Vol. 94). Cambridge: Cambridge University Press.

Calhoun, S. (2010). The Centrality of Metrical Structure in Signaling Information Structure: A Probabilistic Perspective, *Language*, *86*(1), 1–42.

Calhoun, S., Nissim, M., Steedman, M., & Brenier, J. (2005). A framework for annotating information structure in discourse. In *Proceedings of the Workshop on Frontiers in Corpus Annotations 2: Pie in the Sky*, 45–52, Ann Arbor, MI.

Cangemi, F., & Grice, M. (2016). The importance of a distributional approach to categoriality in Autosegmental-Metrical accounts of intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *7*(1), 1–20.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, C. (Ed.), *Subject and Topic* (25–55). New York, NY: Academic Press.

Clark, H. H. (1975). Bridging. In *Proceedings of the 1975 workshop on Theoretical issues in natural language processing*, 169–174, Association for Computational Linguistics.

Cole, J. (2015). Prosody in Context: A Review. *Language, Cognition and Neuroscience, 30*(1–2), 1–31.

Cole, J., Hualde, J. I., Eager, C., & Mahrt, T. (2015). On the prominence of accent in stress reversal. In *Proceedings of the International Congress of Phonetic Sciences*. Glasgow, UK.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics, 35*(2), 180–209.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1*(2), 425–452.

Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: what do listeners imitate? In *12th Annual Conference of the International Speech Communication Association*. Florence, Italy.

Cole, J. & Shattuck-Hufnagel, S. (2017). Quantifying phonetic variation: Landmark labeling of imitated utterances. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking Reduction* Berlin: De Gruyter Mouton.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*(2), 292–314.

De Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks–Evidence from intonation analysis using GToBI. *Journal of Phonetics, 48,* 29–44.

D'Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology, 5*, 1–10.

Dipper, S., Götze, M., & Skopeteas, S. (2007). Information structure in cross-linguistic corpora: Annotation guidelines for phonology, morphology, syntax, semantics and information structure. *Interdisciplinary Studies on Information Structure* (Vol. 7), Potsdam, Germany: University of Potsdam.

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., & Lotts, D. W. (1986). Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *Language and speech, 29*(3), 233–251.

Fasiolo M., Goude Y., Nedellec R. & Wood S. N. (2017). Fast calibrated additive quantile regression. R package version. Retrieved from https://arxiv.org/abs/1707.03307

Féry, C., & Samek-Lodovici, V. (2006). Focus projection and prosodic prominence in nested foci. *Language, 82*(1), 131–150.

German, J. S. (2012). Dialect adaptation and two dimensions of tune. In Q. Ma, H. Ding & D. Hirst (Eds.), *Proceedings of the 6th International Conference on Speech Prosody* (pp. 430–433). Shanghai: Tongji University Press.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* Cambridge, UK: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Gravano, A., Beňuš, Š., Levitan, R., & Hirschberg, J. (2015). Backward mimicry and forward influence in prosodic contour choice in Standard American English. In *Sixteenth Annual Conference of the International Speech Communication Association.*

Greenberg, S. (1999). Speaking in shorthand–A syllable-centric perspective for understanding pronunciation variation. *Speech Communication, 29*(2–4), 159–176.

Gregory, S. W., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior, 21*(1), 23–43.

Gregory, S. W., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language & Communication, 13*(3), 195–217.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274–307.

Halliday, M. A. K. (1970). *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics, 31*(1), 39–62.

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.

Hirschberg, J. (1993). Pitch accent in context predicting intonational prominence from text. *Artificial Intelligence, 63*(1), 305–340.

Hualde, J. I., Cole, J., Smith, C. L., Eager, C. D., Mahrt, T., & de Souza, R. N. (2016). The perception of phrasal prominence in English, Spanish and French conversational speech. In *Proceedings of the International Conference on Speech Prosody* (Vol. 2016, pp. 459–463).

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*(2), 541–573.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*(2), 1038–1054.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication, 34*(4), 391–405.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Levitan, R., Gravano, A., Willson, L., Benus, S., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies* (pp. 11–19). Association for Computational Linguistics.

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Twelfth Annual Conference of the International Speech Communication Association*.

Liberman, M. Y. (1975). *The intonational system of English* (Doctoral dissertation). Retrieved from Massachusetts Institute of Technology.

Luchkina, T. V. (2016). *Prosodic and structural variability in free word order language discourse* (Doctoral dissertation). Retrieved from University of Illinois at Urbana-Champaign.

Luchkina, T., & Cole, J. S. (2016). Structural and referent-based effects on prosodic expression in Russian. *Phonetica, 73*(3–4), 279–313.

Mahrt, T. (2013). Language Markup and Experimental Design software (LMEDS). Retrieved from http://prosody.beckman.illinois.edu/lmeds.html.

Michelas, A., & Nguyen, N. (2011). Uncovering the effect of imitation on tonal patterns of French Accentual Phrases. In *Proceedings of Twelfth Annual Conference of the International Speech Communication Association*.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*(4), 422–432.

Nye, P. W., & Fowler, C. A. (2003). Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics, 31*(1), 63–79.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382–2393.

Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation* (Doctoral dissertation). Retrieved from Massachusetts Institute of Technology.

Pierrehumbert, J., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication,* 271–311.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication, 45*(1), 89–95.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223–256). New York: Academic Press.

Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology*, *70,* 1–30.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Reichel, U. (2011). The CoPaSul intonation model. *Elektronische Sprachverarbeitung*, 341–348.

Reichel, U. & Cole, J. (2016). Entrainment analysis of categorical intonation representations. In *Proceedings of Phonetik & Phonologie*, Munich, Germany. Retrieved from http://www.phonetik.uni-muenchen.de/~reichelu/publications/reichelColePuP.pdf.

Riester, A., & Baumann, S. (2017). The RefLex Scheme – Annotation Guidelines. University of Stuttgart. Retrieved from http://elib.uni-stuttgart.de/bitstream/11682/9028/1/RefLex-SinSpec14.pdf

Riester, A., & Piontek, J. (2015). Anarchy in the NP. When new nouns get deaccented and given nouns don't. *Lingua*, *165*, 230–253.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116.

Shih, C., & Lu, H. Y. D. (2015). Effects of talker-to-listener distance on tone. *Journal of Phonetics*, *51*, 6–35.

Silipo, R., & Greenberg, S. (1999). Automatic transcription of prosodic stress for spontaneous English discourse. In *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS99)* (Vol. 3, pp. 2351–2354).

Silipo, R., & Greenberg, S. (2000). Prosodic stress revisited: Reassessing the role of fundamental frequency. In *Proceedings of NIST Speech Transcription Workshop*.

Silverman, K. E., Blaauw, E., Spitz, J., & Pitrelli, J. F. (1992, February). Towards using prosody in speech recognition/understanding systems: Differences between read and spontaneous speech. In *Proceedings of the workshop on Speech and Natural Language* (pp. 435–440). Association for Computational Linguistics.

Sityaev, D. (2000). The relationship between accentuation and information status of discourse referents: A corpus-based study. In *UCL Working Papers in Linguistics 12*, 285–304.

Sluijter, A. M., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*(4), 2471–2485.

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. In *Proceedings of Fourth International Conference on Speech and Language Processing (ICSLP 96)* (Vol. 3, pp. 1501–1504).

Terken, J. & Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position. *Language and Speech, 37*(2), 125–145.

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*(3–4), 145–163.

Turnbull, R. (2017). The role of predictability in intonational variability. *Language and Speech, 60*(1), 123–153.

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience, 32*(8), 1017–1033.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics, 27*(2), 171–206.

Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). *6.911 Transcribing Prosodic Structure of Spoken Utterances with ToBI [PowerPoint slides]*. Retrieved from https://ocw.mit.edu.

Vieira, R., & Poesio, M. (2000). Processing definite descriptions in corpora.*Corpus-based and Computational Approaches to Discourse Anaphora*, 189–212.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7–9), 905–945.

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *The psychology of learning and motivation* (1st ed., Vol. 52, pp. 163–183). Elsevier Inc.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*(3), 1548–1557.

Wood, S. N. (2017) *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC.

Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. Local, R. Odgen, & R. Temple (Eds.) *Phonetic interpretation: Papers in laboratory phonology* (6th ed.) (pp. 26–50). Cambridge: Cambridge University Press.

Xu, Y. (2013). ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7–10).

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, *18*(1), 125–159.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337.

## APPENDIX A: QUANTILE REGRESSION MODELS

Further analyses were performed to check the effects of autocorrelation on the results obtained from GAMMs. The same data and models were submitted to Quantile Regression Models using qgam (QGAM, henceforth). The results were found to be similar between GAMMs and QGAMs as shown in Table A.1. All the domains show high deviance explained values. The intermediate phrase domain shows the highest deviance explained values for both analyses.

Table A.1.

*GAMM and QGAM Summary Table for Six Domains.*

| Domain | Deviance Explained Value (%) | |
|---|---|---|
| | GAMM | QGAM |
| Intermediate Phrase | 79.2 | 63.4 |
| Accentable Word | 76.7 | 61.1 |
| Pitch Accent | 77.1 | 61.3 |
| Foot | 75.7 | 60.6 |
| Stressed Syllable | 77.7 | 62.3 |
| Syllable | 77.5 | 63.0 |