



Open Library of Humanities

Standing out in context: Prominence in the production and perception of public speech

Suyeon Im*, Department of English Language and Literature, Soongsil University, South Korea, sim@ssu.ac.kr

Jennifer Cole, Department of Linguistics, Northwestern University, Evanston, IL, USA, jennifer.cole1@northwestern.edu

Stefan Baumann, IfL Phonetik, University of Cologne, Germany, stefan.baumann@uni-koeln.de

*Corresponding author.

This study investigates the relationship between signal-based factors and discourse context in the production and perception of prominence, extending prior work on prominence in American English by examining speech from complete TED Talks as examples of a public speech style. First, we conduct a production study. The TED Talk speech samples are analyzed for relationships between (1) phonological pitch accent, (2) acoustic prosodic features, and (3) information structure, modeled here in terms of referential and lexical givenness as well as focus related to semantic alternatives. Second, we conduct a perception study. Word-level prominence ratings obtained from untrained listeners are analyzed for the effects of these (1) phonological, (2) acoustic, and (3) information structural factors. The aim of this study is to evaluate the interaction of the three kinds of factors on prominence ratings with speech that is fully contextualized, allowing for information about the speaker, the situational context, and the discourse context to interact with and possibly mediate signal-based cues to prominence. Results from the production analysis show a weak probabilistic relation between pitch accent assignment and information structure for these TED Talk speakers, but only in the presence versus absence of pitch accent with differences between lexical and referential information status. No systematic relationship is found between a word's phonological pitch accent type and information structure category, though we do find some evidence of the expected gradient variation in the acoustic expression of prominence: Words that are more informative tend to have higher pitch and a wider pitch range. Despite the weak association between signal-based prominence and information structure in production, listeners perceive and rate prominence in a manner that is, to varying degrees, in accordance with the Accentual Prominence and Givenness hierarchies. In sum, the findings contribute new evidence of the mediating effects of contextual factors relating to the phonological status of pitch accents, information structure, and speech style on the interpretation of acoustic cues to prominence.



A word produced in a phrasal context is perceived as prominent to the extent that it stands out among neighboring words. Many factors are known to contribute to the perceptual prominence of a word, including those related to the speech signal (acoustic prominence), the phonological pitch accent category (accentual prominence), the position of the word in the prosodic phrase (structural prominence), and the semantic properties of the word, such as animacy or thematic role (semantic prominence) (Cole et al., 2019; Luchkina & Cole, 2017). These factors are all local to the word, or to the prosodic phrase or syntactic clause to which the word belongs. Other factors associated with prominence are non-local, in that they have to do with how a word relates to the prior discourse context, e.g., as conveying given or new information, or as invoking contextually salient focus alternatives (contrastive focus) (see Büring, 2007, for an overview). Beyond information status (givenness) and semantic alternatives, there are other properties of discourse context, e.g., the discourse roles of the speaker and addressee, their social identities, and their relationship to one another, yet there is relatively little work investigating effects of such discourse factors on listeners' perception of prominence. The effects of psycho-social and situational properties of discourse context on speech processing are admittedly challenging to investigate using standard experimental methods, but must be considered alongside lexical and grammatical factors, for the goal of achieving a comprehensive model of prominence perception. We take a step towards this goal in the present study, through an approach that is part case study and part experimental study. The case-study component uses corpus methods to describe the relationships among (1) accentual prominence, (2) acoustic prominence, and (3) information structure, in two speech samples that are authentic examples of an engaging public speaking style, each comprising a complete discourse. The experimental component uses a real-time prominence rating task to assess spontaneous prominence judgments in the same speech sample presented to listeners as a complete discourse.

To preview the methods described in greater detail below, this study investigates listeners' perception of prosodic prominence in relation to local, signal-driven factors and in relation to salient contextual factors, in two complete public speeches by speakers of American English (AE). Prosodic prominence is rated by AE listeners who have no special training or linguistic expertise, using Rapid Prosody Transcription (Cole & Shattuck-Hufnagel, 2016; see also Roy, Cole, & Mahrt, 2017; Cole, Mo, & Hasegawa-Johnson, 2010). As signal-driven factors, phonological features of intonation (pitch accents and phrase boundary tones) and acoustic cues to prominence are examined. Structural features related to the position of pitch accent and the type of prosodic boundary are also taken into account. As factors related to the discourse context, we examine information structure (IS), modeled as referential and lexical givenness as well as contrastive focus, using the RefLex scheme (Riester & Baumann, 2017). We investigate these factors, and their interaction, as perceptual cues to prominence in speech presented to the listener as a complete discourse, which allows other salient dimensions of the situational context to come

into play, while also providing a rich discourse context for determining IS, on which givenness and focus distinctions are based. The TED Talk speeches chosen for this study represent a public speech style intended to engage the listener, and in this regard, the present study differs from prior studies that examine prominence perception in de-contextualized conversational excerpts or scripted sentences produced in a laboratory setting. This study represents a step towards a much larger goal of understanding the influence of speech style on prominence perception.

Some key findings of this study are summarized below. In these TED Talk speeches, listeners' perception of a word's prosodic prominence was more likely with:

- An enhancement of the word's acoustic correlates of prominence,
- The word being new or alternative-eliciting in the discourse context,
- The presence of a high-toned pitch accent (H*, L + H* and L + H*) on the word, compared to low-toned and downstepped pitch accents (L* and !H*),
- The presence of a nuclear pitch accent on the word (in phrase-final position), compared to prenuclear pitch accents (in phrase-initial and phrase-medial positions),
- The presence of a rising edge tone (H-H% and L-H%), compared to low or falling (L- and L-L%) and mid-flat edge tones (H-, H-L%),
- The immediately preceding word being perceived as prominent.

Overall, this study presents new evidence of the mediating effect of context relating to phonological factors (pitch accent status and type), as well as discourse factors (IS) on the interpretation of acoustic cues to prominence.

Section 1 reviews key findings from prior studies on the relationships among prominence, pitch accents, and IS. The methods and results for the two components of this study follow in Sections 2 and 3. Section 2 introduces the TED Talk speech materials and presents the methods and results of the analysis of those production data, showing the relationship between the pitch accent status of a word, its acoustic prosodic properties, and its information status. Section 3 presents the methods and results of the perceptual prominence rating experiment, where we examine how the same factors influence prominence rating for these speech samples. The production and perception data that are analyzed in this study are schematized in **Figure 1**. Section 4 discusses the results, and we finish with concluding remarks in Section 5.

1. Prominence in prior work

1.1. Pitch accents in American English (AE)

As used here, prominence refers to the perceptual salience of a word in relation to surrounding words in an utterance, as grounded in its structural and acoustic properties (Cangemi &

Baumann, 2020). We are concerned with phrase-level prominence, distinguished from word-level prominence, following the early seminal work by Bolinger (1958). In Autosegmental-Metrical theory (Lieberman, 1975; Pierrehumbert, 1980; term coined by Ladd, 2008), prominence is associated with metrically strong positions in an utterance. In English, the nuclear prominence is the rightmost prominence in a prosodic phrase. On the other hand, prenuclear prominences, which appear to be optional, tend to occur in the initial or early position of a prosodic phrase (Shattuck-Hufnagel, 1995; Shattuck-Hufnagel, Ostendorf, & Ross, 1994), or in positions that produce a rhythmic alternation of prominence (Calhoun, 2010; Vogel, Bunnell, & Hoskins, 1995).

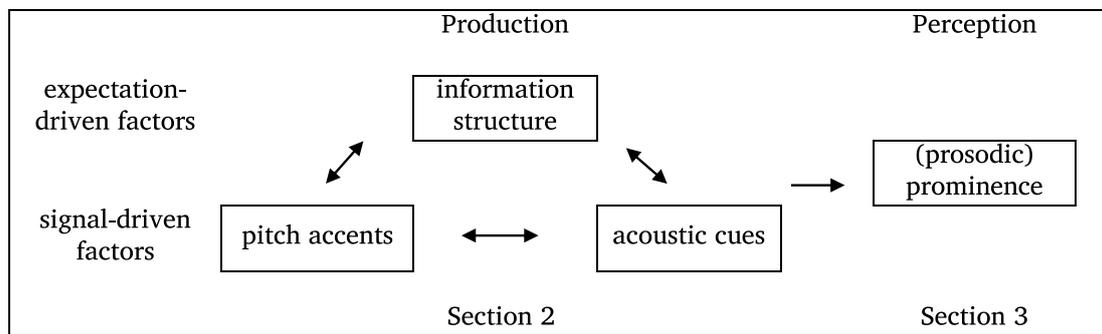


Figure 1: Schematic representation of the current study, relating IS (referential and lexical givenness, plus contrastive focus) as expectation-driven factors, and pitch accents and acoustic cues as signal-driven factors for prominence perceived by linguistically non-expert listeners of AE.

Words in metrically strong positions are eligible for the assignment of a tonal pitch accent, which specifies the pitch pattern realized on or around the accented syllable. There is general agreement that AE uses a set of tonally distinct pitch accents to mark prominences, and one characterization of this inventory has been reflected in the widely adopted Tones and Break Indices (ToBI) system for annotating intonation in Mainstream AE (MAE: Beckman & Ayers Elam, 1997; Silverman et al., 1992; Veilleux, Shattuck-Hufnagel, & Brugos, 2006), though the distinctive status of some pitch accent types has been disputed.¹ Recent experimental findings from AE show that accent type influences prominence perception (Bishop, Kuo, & Kim, 2020; Cole et al., 2019; Hualde et al., 2016; Turnbull, Royer, Ito, & Speer, 2017, among others). In Cole et al.

¹ For instance, although Pierrehumbert (1980) proposes three accent types, H*, L*+H, and L+H* that involve a rising pitch contour, and though some studies offer empirical support for distinctions among these accents (Arvaniti & Garding, 2007; Beckman & Pierrehumbert, 1986; Cole et al., 2019; Dainora, 2001; Dilley & Heffner, 2013), other authors cast doubt on the status of one or more accents in this set (Bartels & Kingston, 1994; Calhoun, 2006; Ladd & Schepman, 2003, among others). Another contested distinction is that between H* and its downstepped counterpart, !H*, proposed by Ladd (1980; see also Liberman & Pierrehumbert, 1984; Yoon & Cole, 2006), but considered as an allotonic variant of H* by Dainora (2001).

(2019), prominence ratings of words from non-expert listeners were compared with pitch accents from a ToBI annotation produced by trained annotators. Results showed that accented words were more likely to be perceived as prominent than unaccented words, and among accented words, those with a nuclear pitch accent (i.e., the final accent in the intermediate phrase—the level below the intonational phrase) showed a greater likelihood of perceived prominence than prenuclear accented words. Also, L + H*, the pitch accent described as marking contrastive focus (Pierrehumbert & Hirschberg, 1990), had the largest effect of boosting the likelihood of prominence perception compared to other pitch accent types. Similar results were reported for the comparison of prominence ratings and ToBI pitch accent labels in Bishop et al. (2020), with data from a public speech of former U.S. President Barack Obama. The results from that study showed the likelihood of perceived prominence increasing across distinctions in the accent status of a word: Unaccented < prenuclear accented < nuclear accented. Moreover, among the accented words, the likelihood of perceived prominence increased incrementally across accent categories: L* or !H* < H* < L + H*. In addition, the prominence ratings for H* depended on the position of the accent. Ratings for prenuclear H* were similar to those for prenuclear L* or !H*, while ratings for nuclear H* were similar to those of nuclear L + H*. Henceforth, we use the term “Accentual Prominence hierarchy” to refer to the ranking of pitch accent types based on the likelihood of perceived prominence.

1.2. Acoustic correlates

Acoustic correlates of accentual prominence are reported in numerous studies showing that accented words have longer duration, greater intensity, steeper spectral slope, and are hyper-articulated, compared to unaccented words (Beckman, 1986; Breen, Fedorenko, Wagner, & Gibson, 2010; Cole, Kim, Choi, & Hasegawa-Johnson, 2007; Kochanski, Grabe, Coleman, & Rosner, 2005; Sluijter & van Heuven, 1996; Turk & White, 1999, among others). F0 movement is also considered a correlate of accentual status (Ladd, 2008; Ladd & Morton, 1997; Ladd, Verhoeven, & Jacobs, 1994; Pierrehumbert, 1980), though the empirical evidence has been inconsistent (Calhoun, 2006; Cole et al., 2010; Cole et al., 2019; Kochanski et al., 2005). Kochanski et al. (2005) examined acoustic correlates of syllables marked as prominent by trained phoneticians in a spontaneous speech corpus of British and Irish English. Their findings showed that loudness (based on a normalized energy measure) and phone duration were important in distinguishing prominent from non-prominent syllables, while measures of F0, aperiodicity (as a measure of voicing), and spectral slope showed little or no such effect. Cole et al. (2019) examined acoustic correlates of perceptual prominence ratings for words in excerpts from the Buckeye corpus of AE conversational speech. The word phone rate (as a measure of local tempo), along with intensity and maximum F0 in the stressed syllable, were found to be significant factors predicting prominence ratings, with word phone rate showing the largest effect on perceived prominence.

Finally, Bishop et al.'s (2020) study of political speech investigated prominence ratings in relation to the acoustic properties and accent status of a word and found significant effects of maximum F0, duration, and intensity on the perceived prominence of accented words, with somewhat different patterns for accented words in prenuclear and nuclear positions. Summarizing findings from these and other studies, acoustic correlates of perceived prominence in AE have been noted in measures of duration, intensity (or related energy measures), and maximum F0, though with variation across studies in the significance and relative strength of effect among these measures. Below, we use the term “prosodic prominence” to refer jointly to accentual prominence and acoustic prominence.

1.3. Information structure and discourse meaning

Prosodic prominence in AE conveys referential meaning as it relates to discourse context, suggesting a role for discourse meaning in prominence perception. While a full account of intonational meaning awaits further developments in semantic and pragmatic meaning, we briefly review some of the key insights and empirical findings here. Evidence for the contribution of prominence to referential meaning is illustrated in the phenomenon of reference resolution for pronouns in AE. For example, in the sentence “Sue threw Laura a purple hat. She (...),” the antecedent of the word *she* is more likely to be *Laura*, if *Laura* carries prosodic prominence, as shown by Schafer, Camp, Rohde, and Grüter (2019). In their compositional theory of intonational meaning, Pierrehumbert and Hirschberg (1990) propose a one-to-one mapping between pitch accent type and meaning distinctions related to IS. In their analysis, a high-toned accent, H*, is used to mark words that contribute new information to the common ground, while a low-toned accent, L*, is used when a speaker attempts to make an item salient but does not wish to include the item in the predication. The bitonal rising accents, L*+H and L+H*, are said to invoke a scalar interpretation, with L*+H conveying speaker uncertainty, while L+H* marks corrective or contrastive focus. These and other IS distinctions are also manifested in acoustic correlates of pitch accents, with greater intensity, longer duration, and higher mean and maximum F0 for words with narrow or contrastive focus compared to words without focus that are discourse-new or -given (Breen et al., 2010).

More recently, evidence from German and AE challenge the notion of a one-to-one relationship between pitch accents and IS. For example, in a corpus analysis of read German, Baumann and Riester (2013; see also Mücke & Grice, 2014; Roessig, Mücke, & Grice, 2019) found that each attested accent type (H*, !H*, H+L*, L*) was associated with multiple levels of givenness, although with different patterns of distribution over IS categories: H* was more often used for referentially and lexically new items, while the falling accent, H+L*, was more often used for accessible words, and given words were more often unaccented. In an analysis of read aloud short narratives in AE, Chodroff and Cole (2018, 2019) similarly observe that each attested

accent type (H^* , L^* , $L^* + H$, $L + H^*$) occurs with focus and with multiple levels of givenness, both for prenuclear and nuclear accented words. Findings like these point to probabilistic rather than deterministic mapping between pitch accents and discourse meaning, including IS.

Another line of research argues that the binary given-new distinction is insufficient to capture the range of prosodically encoded IS distinctions, calling for a more complex set of distinctions among degrees or types of givenness (Baumann & Riester, 2012, 2013; Calhoun, Nissim, Steedman, & Brenier, 2005; Chafe, 1976, 1994; Dipper, Götze, & Skopeteas, 2007; Lambrecht, 1994; Prince, 1981; Riester & Baumann, 2017). Chafe (1976) proposes a three-way given-accessible-new distinction based on the active state of an idea in discourse context (see also Prince, 1981). Baumann and Riester (2012, 2013) further develop this model to include an additional distinction: *Given* words are those that are either explicitly mentioned or situationally salient in discourse context; an *accessible* (or *bridging*) word is one that is inferable from previously mentioned words (e.g., whole-part relationship); a *new* word can be either *brand new* (assumed to be unknown to the hearer) or *unused* (assumed to be known to the hearer but not yet activated in the hearer's consciousness). These IS labels can be ranked in order of informativity as *given* < *bridging* < *unused* < *new*, yielding a "givenness hierarchy" (using the same term but different categories as Gundel, Hedberg, & Zacharski, 1993). Baumann and Riester (2012, 2013; also Riester & Baumann, 2013, 2017) propose a two-level model, with givenness distinctions at the referential and lexical levels. The referential level deals with the information status (i.e., the assumed level of cognitive activation) of discourse referents or entities—such as people, states, or events—during communication, while the lexical level describes the information status of a lexical item or concept in a given context. Empirical evidence of distinctions in prominence between referential and lexical information status were shown in read and spontaneous speech in German (Baumann & Riester, 2013) based on the type and location of pitch accent for prenuclear or nuclear pitch accents. A greater degree of prominence was associated with lower levels of lexical givenness, and within each level of lexical givenness, a greater degree of prominence was associated with lower levels of referential givenness.

A final comment on the contribution of pitch accents to discourse meaning is that this function depends to some extent on the intonational context of the accent. There is a long tradition of research on intonational meaning in AE and British English that associates meaning with the nuclear accent in its combination with the following phrase accent and boundary tone, i.e., the *nuclear tune* (Bolinger, 1958; Ladd, 1980; O'Connor & Arnold, 1961). Nuclear tune meaning has been variously described in terms of speech act meaning (Lieberman & Sag, 1974; Sag & Lieberman, 1975), focus of attention (Jackendoff, 1972; Ladd, 1980), and the speaker's attitudes towards a hearer (Lieberman, 1975; O'Connor & Arnold, 1961), or towards the propositional content of an utterance (Ward & Hirschberg, 1985). An alternative approach derives tune meaning compositionally (Pierrehumbert & Hirschberg, 1990), from the independent meaning

contributions of the pitch accent, phrase accent, and boundary tone. Empirical evidence for nuclear tune meaning in AE is not plentiful but results so far suggest the potential for variability in tune-meaning mapping depending on the discourse context (Arvaniti & Garding, 2007; Reed, 2017; Ward & Hirschberg, 1985).

1.4. Speech style

Speech styles differ with respect to phonological and phonetic sound patterns. In clear speech, speakers enhance phonological and phonetic contrasts more than in conversational speech, promoting intelligibility and listeners' understanding by adjusting the signal to meet the demands of the context (Cho, Lee, & Kim, 2011; Cole et al., 2007; de Ruiter, 2015; Hirschberg, 1993; Smiljanić & Bradlow, 2005). Style-related effects can be observed for prosodic as well as segmental contrasts. At the phonetic level, prosodic enhancement associated with clear speech style may manifest in an increase in pitch range, longer segment, and pause duration compared with conversational speech (Smiljanić & Bradlow, 2005). Phonological effects of speech style related to prosody are also observed. For example, de Ruiter's (2015) study of German finds different usage patterns of pitch accents in relation to givenness, and different preferred boundary tones in read speech compared to spontaneous speech. In a similar vein, Hirschberg (1993) reports the frequent use of accent with *given* words in a corpus of broadcast radio speech in English. Style-based variation is also reported by Chodroff and Cole (2018, 2019) in their data from short read aloud narratives, where L + H* is more frequent in trials where participants were instructed to speak in a lively speaking speech style compared with a neutral, conversational style, regardless of the focus status of the word (very similar results were obtained in a study on German with an equivalent setup; Baumann, Mertens, & Kalbertodt, 2019). A related finding from Arvaniti and Garding (2007) links the use of L + H* to emphatic discourse contexts for speakers from Southern California (but not for speakers from Minnesota).

1.5. Multiple cues to prominence perception

The works cited above, among others, show that prominence in speech production arises from various sources related to the pitch accent assigned to a word, its acoustic implementation, and its information structure in relation to the discourse context (see also Watson, 2010). These sources contribute cues that may function independently and jointly to influence listeners' perception of prominence (Cole et al., 2010). Some of the cues reside in the speech signal (e.g., pitch accents, acoustic correlates of prominence), while other cues reflect a listener's expectations about prominence as it is associated with contextual features such as IS or word predictability. There is also growing evidence that all of these factors influence prominence perception (Aylett & Turk,

2004; Baumann & Winter, 2018; Bishop et al., 2020; Breen et al., 2010; Cole et al., 2019; Cole et al., 2010; Turnbull et al., 2017; Watson, Arnold, & Tanenhaus, 2008, among many others). Cole et al. (2010) found that listeners weigh expectation-driven factors to a greater degree than the acoustic measures. Bishop et al. (2020) found that pitch accent type had a significant direct effect on prominence ratings, while acoustic cues and extrinsic factors showed effects mediated by pitch accent, e.g., F0 effects were greater for accented words than for unaccented words. Turnbull et al. (2017) observed predicted effects of pitch accent type on prominence perception, but the pitch accent effect was stronger for words expressing a contrastive focus, and this added effect of IS was especially noted for materials presented in an engaging two-party narrative dialogue.

In the current study, we investigate perceived prominence in relation to IS as an expectation-based factor, and in relation to phonological and phonetic cues as signal-driven factors, in a sample of AE speech comprising two complete TED Talks. These factors were selected for the following reasons. First, we are interested in the influence of multi-level distinctions in IS on perceived prominence. For this, we adopt the two-level givenness hierarchy of Baumann and Riester (2013) modeling the referential and lexical givenness of a word, as well as its marking for contrastive focus that invokes salient semantic alternatives (Rooth, 1992). As noted above, lexical givenness is defined based on the presence or absence of an explicit prior mention of a lexical word, while referential givenness corresponds to the givenness status of the discourse referent of a word. Second, we examine the relationship between the type of pitch accent (i.e., its phonological category) and the degree of givenness of a word. We formulate the relationship between prominence and IS as a mapping between ranked elements in each domain, as shown in **Figure 2**. In the hypothesized mapping, the absence of accent (an unaccented word) is associated with referentially and lexically *given* information; L* and !H* (the accents with the lowest degree of prominence, following Bishop et al., 2020) are associated with the IS category *bridging*; H* with *unused*; and L* + H and L + H* (the highest degree of prominence) with *new* information. We wish to examine the relationship between these hierarchies in speech from the public speech style.

Accentual Prominence hierarchy	unaccented	<	L*, !H*	<	H*	<	L* + H, L + H*
	↓		↓		↓		↓
Givenness hierarchy	given	<	bridging	<	unused	<	new

Figure 2: Hypothesized relationship between the Accentual Prominence hierarchy and the Givenness hierarchy.

Third, we examine IS using speech samples that differ from those of most prior work in two ways. Where previous studies have used minimally contextualized utterances (e.g., question-answer sentences or brief excerpts from a corpus), we use a complete discourse (monologue) to provide listeners with richer context to ground distinctions in IS and other factors that may influence prominence perception. Our prominence rating study is designed as parallel to prior work on prominence rating for conversational speech (Cole et al., 2019; Cole et al., 2010) and political speech (Bishop et al., 2020), allowing a comparison between studies to assess potential differences in cue patterning in relation to speech style. Taken together, this study seeks to answer the following questions about prominence in the public speech style represented in two TED Talks:

Q1 [prominence production]: How are distinctions in IS associated with phonological distinctions (type of pitch accent) and with acoustic correlates of prominence?

Q2 [prominence perception]: What are the effects of IS, pitch accent, and acoustic prosodic measures on the perception of prominence by linguistically untrained listeners?

These questions are approached through confirmatory hypothesis testing, with a set of hypotheses drawn from work mentioned above on AE (Bishop et al., 2020; Cole et al., 2019) and German (Baumann & Riester, 2013). For prominence in speech production, our hypotheses are expressed in **Figure 2**, namely, that accents will be assigned corresponding to the IS status of a word, respecting the ranking relations at each level, and in addition, distinctions in accentual prominence will be reflected in their phonetic implementation, with the degree of acoustic enhancement varying with the accentual prominence.

From a perceptual perspective, the hypotheses can be formulated as follows:

The likelihood a word will be perceived as prominent increases with:

H1: Enhancement in the acoustic correlates of prominence.

H2: Its Givenness status, with given < new; and its status as alternative-eliciting (focused), with non-alternative-eliciting < alternative-eliciting.

H3: Its accent status and type on the Accentual Prominence hierarchy, from L* to L + H* (possibly mediated by its status as prenuclear or nuclear (H4), and by the following edge tones (H6)).²

² As our main goal is to examine the relationship between pitch accents and IS, we do not attempt to make an extensive examination of other factors, especially, the mapping between the tune (comprising the nuclear pitch accent, phrase accent and boundary tone) and its discourse meaning in the TED Talks, which we leave for future research.

H4: Its accent status and position in the phrase, with unaccented < non-initial prenuclear < phrase-initial prenuclear < nuclear.³

H5: A decreased likelihood that the preceding word is perceived as prominent, as an effect of prominence alternation or ‘rhythm.’⁴

In addition, there is an exploratory dimension to this study, with two additional questions for which neither prior theoretical work nor empirical work motivates specific hypotheses. One concerns the effect of different types of prosodic boundary on perceived prominence for words in pre-boundary (final) position, and the other concerns the expected prominence ranking among IS distinctions in the elaborated Givenness hierarchy adopted here. For prosodic boundaries, we propose a prominence-related ranking based on the layering of prosodic domains, with word boundaries at the lower end and intonational phrase boundaries at the higher end: Word boundary < {L-, H-} < {L-L%, H-H%, L-H%, H-L%}. This order of boundary-marking tones (also termed *edge tones*) is used in our statistical analysis. This treatment of edge tones is parallel to that of pitch accents, which we model not only in terms of a binary distinction (accented vs. non-accented), but also in terms of their tonal characterization. For prominence effects related to IS distinctions, we rely on the pitch accent type expected for each level on the Givenness hierarchy (**Figure 2**) to define the Givenness prominence ranking in direct relation to the Accentual Prominence ranking: Given < {accessible, bridging} < unused < new (see **Table 1** for definitions and examples). We formulate one more provisional hypothesis [H6], and a revised version of hypothesis 2, reflecting our best estimate of these effects.

The likelihood a word will be perceived as prominent increases with:

H6: The status of the immediately following boundary on the prominence-related ranking: word < phrase accent < boundary tone.

H2’: The word’s status on the expanded Givenness hierarchy, related to Accentual Prominence type, with given < {accessible, bridging} < unused < new.

³ Phrase-initial prenuclear words may gain a prominence boost from rhythmic factors, based on the observation that phrase-initial words frequently have a “rhythmic” accent (Shattuck-Hufnagel et al., 1994).

⁴ An RPT study on German (Baumann & Winter, 2018) found a strong preference for prominence judgments on words which did *not* follow other prominent words, supporting the prominence alternation hypothesis for West Germanic languages.

Level	Label	Description	Example
referential	r-given	coreferring entity present in discourse	A car was waiting in front of the hotel. I could see a woman in the car .
	r-bridging	accessible entity present in discourse	I put the key in the lock but the door wouldn't open.
	r-unused	unique (definite), new entity in discourse	President Barack Obama delivered a brilliant speech in Tucson .
	r-new	non-unique (indefinite), new entity in discourse	After the holidays, John arrived in a new car and Harry had also bought a new car .
lexical	l-given	active expression in discourse	A cat makes for a popular pet. Moreover, a cat is quite independent.
	l-accessible	semi-active expression in discourse	I tried to open the door but the lock was rusty.
	l-new	inactive expression in discourse	Smith was very optimistic. The polls showed a solid majority for the politician .
alternative-eliciting	alt	clearly identifiable alternative entity present in discourse	Did you call John ? No, I called Mary .

Table 1: IS labels adopted from the RefLex scheme.

2. Prominence analyzed in production data: Pitch accent, acoustics, and IS

2.1. Materials

The speech materials selected for this study consist of the complete speech content from two TED Talks.⁵ Portions of the transcripts for these TED Talks are shown in (1) and (2).

- (1) 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.

⁵ “Try something new for thirty days” (https://www.ted.com/talks/matt_cutts_try_something_new_for_30_days) and “Talk nerdy to me” (https://www.ted.com/talks/melissa_marshall_talk_nerdy_to_me)

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. (...)

- (2) Five years ago, I experienced a bit of what it must have been like to be Alice in Wonderland. Penn State asked me, a communications teacher, to teach a communications class for engineering students. And I was scared. Really scared. Scared of these students with their big brains and their big books and their big, unfamiliar words. But as these conversations unfolded, I experienced what Alice must have when she went down that rabbit hole and saw that door to a whole new world. That's just how I felt as I had those conversations with the students. I was amazed at the ideas that they that they had, and I wanted others to experience this wonderland as well. (...)

These TED Talks were delivered by two speakers of AE, one male and one female, in a clear and engaging manner. It is typical for TED Talks to be rehearsed and delivered with the aid of a teleprompter, and we assume that was the case for the TED Talks we selected. We chose these speech samples for the following reasons. First, the TED Talks are planned monologues, each covering a non-technical topic for a general audience. We assume that the speakers had to structure the information they planned to convey in a manner that would be coherent for the audience, who could not be expected to have shared background knowledge. This type of speech provides an ideal test case for our goal to examine prominence in relation to the semantic and pragmatic relationships between words in the presence of rich discourse context. Second, the speech samples are produced in a clear and engaging speech style for a large audience. Clear speech is characterized by phonological features and phonetic cues that are enhanced relative to those of conversational speech (Baumann & Riester, 2013; Cho et al., 2011; Cole et al., 2007; de Ruiter, 2015; Hirschberg, 1993; Smiljanić & Bradlow, 2005). We assume that this would be beneficial for participants in our prominence rating experiment, in making perceptual judgments of prominence. Combined, the two TED Talks consist of 1018 words ($t = 7' 14''$): 361 words from the male speaker ($t = 2' 56''$) and 657 words from the female speaker ($t = 4' 18''$).

2.2. Methods

2.2.1. Annotation of information structure

The speech materials were annotated for information structure by one of the authors (SB) using a simplified version of the RefLex scheme (Riester & Baumann, 2017). Three levels of IS—

referential, lexical, and contrastive (or “alternative-eliciting”, see Riester & Baumann, 2013)—were considered for each word in the speech samples. As discussed above, the referential (*r-*) level describes the information status of referring expressions, which (in most cases) are noun phrases that denote discourse referents. Words that are not in referring expressions, e.g., idioms and expletives, along with verbs and adverbs that are not part of a noun phrase, are not defined for referential givenness. We assign the label *NR* (*r-none*) to such words in this dataset. The lexical (*l-*) level applies to the domain of content words (nouns, adjectives, adverbs with lexical semantic content, and verbs), describing the degree to which the word is “activated” in the mind of the speaker/hearer, by prior mention of the same lexical word or a word that is semantically related, e.g., as a synonym, hyponym, hypernym or holonym. Outside of this set are pronouns, discourse markers and function words, which we designate with the label *NL* (*l-none*). In addition to the *r-* and *l-* levels, the alternative (*alt-*) level marks word pairs that occur in a parallel construction and which are thus potentially in contrastive focus, as well as verum focus constituents or words that are associated with a focus-sensitive particle. Words that do not have any of these focus types are labeled as *non-alt*. The descriptions and examples of *r-*, *l-*, and *alt-* labels are summarized in **Table 1**. The labels are presented in increasing order of informativity (i.e., decreasing order of givenness) for each level (e.g., *r-given* < *r-bridging* < *r-unused* < *r-new*). Words that are labeled as *NR*, *NL*, and *non-alt* are undefined for givenness at their respective level. The words in bold provide illustrative examples of each label (not from the TED Talks).

2.2.2. Annotation of pitch accents, phrase accents, and boundary tones

Pitch accents, phrase accents, and boundary tones were annotated for all words in the TED Talks ($n = 1018$) by one of the authors (JC) following ToBI annotation conventions (Veilleux et al., 2006). Among the accented words ($n = 542$; $n = 184$ from the male speaker and $n = 358$ from the female speaker), six accent types— L^* , $!H^*$, H^* , $H+!H^*$, L^*+H and $L+H^*$ —were observed. $H+!H^*$ ($n = 11$) was reassigned to $!H^*$ with the same starred tones, due to its low frequency in the materials. If the reassigned $!H^*$ appeared in the initial position of an (intonational or intermediate) prosodic phrase, it was further modified to H^* following the ToBI annotation conventions. The position of a pitch accent in the intermediate phrase was also labeled. Three accent positions were used: *Final* for the nuclear pitch accent ($n = 248$), and *initial* ($n = 172$) and *middle* ($n = 122$) for the prenuclear pitch accents. If there was one pitch accent in an intermediate phrase, this accent was labeled as *final*. If there were two pitch accents, the initial pitch accent was considered as *initial* and the final accent as *final*. If there were more than two pitch accents, all pitch accents other than the final and initial accents were labeled as *middle*. Lastly, for phrase accents and boundary tones ($n = 253$), six types— $L-$, $H-$, $H-L\%$, $L-L\%$, $L-H\%$, $H-H\%$ —were found in our data.

2.2.3. Acoustic measures

Four acoustic measures—max F0, F0 range, phone rate (as a measure of local tempo), and mean intensity—were selected as acoustic correlates of prominence, following prior work on AE. Specifically, Cole et al. (2019) showed that max F0, phone rate, and mean intensity are significant predictors of variation in prominence ratings in conversational speech from the Buckeye corpus (Pitt, Johnson, Hume, Kiesling, & Raymond, 2005). F0 range is also included in the present study. Max F0 and F0 range (in semitones), and mean intensity (in dB) were measured from the stressed syllable of each word in the two TED Talk samples. For the F0 measures, pitch halving or pitch doubling were manually checked and corrected using *Praat* (Boersma & Weenink, 2019). For phone rate, the number of phones in a word was divided by the entire duration of the word, with the phone counts calculated based on *the CMU Pronouncing Dictionary* (Weide, 2005).⁶ All acoustic measurements were obtained using *ProsodyPro* (Xu, 2013). The four acoustic measures were normalized in two steps. The first step was to measure the acoustic prominence of a word relative to the neighboring words. For this, the F0 and intensity measures of each word (the target) were normalized using the means and standard deviations of the adjacent words in a five-word window centered on the target word. The phone rate of each word was normalized for the speech rate of an entire utterance, using a *Praat* script written by Tim Mahrt, an author of the Cole et al. (2019) paper. The second step was to take into account the different units of the acoustic measures. The four acoustic measures of each word were centered and scaled using the means and standard deviations of the same measures from all words in the speech sample, separately for each speaker. Finally, based on visual inspection of the graphed data, the F0 and intensity measures were found to be positively correlated with perceived prominence while phone rate was negatively correlated. To have a consistent sign of the correlation coefficient among the acoustic measures, the inverse phone rate (positively correlated with prominence) was used in statistical analysis.

2.3. Results

2.3.1. Distribution of pitch accents and IS categories

Below, we present analyses of Pearson's chi-squared tests and multivariate multiple regressions, which provide information on whether and how the speakers convey IS distinctions through pitch accents and their phonetic implementation. We first investigate the relationship between pitch accents and IS. **Table 2** shows the frequency of words associated with specific pitch

⁶ Phone rate provides a measure of word duration that is normalized for the speech rate of the utterance using Pfitzinger's (1998) RateLR formula. Word phone rate values are z-scored values of phones/second and indicate how much the word is lengthened or shortened relative to the speech rate of the utterance (including function words) and taking into account the number of phones in the word. There are of course additional factors that contribute to word duration, including - importantly - segment identity, which we have not modeled.

accents and their IS labels. These data were submitted to a series of Pearson's chi-squared tests (Pearson, 1900) to test for differences in the distribution of accent and accent types among the IS categories. Fisher's exact test (Fisher, 1934) was employed instead for the analysis of IS levels with less than five tokens. As noted above, the labels *NR*, *NL*, and *non-alt* are assigned to words for which referential, lexical, and alternative-eliciting information is not defined, respectively.

Level	Label	Unacc	L*	!H*	H*	L* + H	L + H*	Acc	Total
r-level	NR	356	25	43	83	17	73	241	597
	r-given	98	12	12	24	11	39	98	196
	r-bridging	4	4	13	8	5	6	36	40
	r-unused	6	11	19	16	3	23	72	78
	r-new	12	13	16	26	14	26	95	107
	total	476	65	103	157	50	167	542	1018
l-level	NL	424	12	15	66	5	65	163	587
	l-given	11	10	14	14	8	12	58	69
	l-accessible	3	3	4	7	3	4	21	24
	l-new	38	40	70	70	34	86	300	338
		total	476	65	103	157	50	167	542
alt-level	non-alt	468	56	93	145	42	147	483	951
	alt	8	9	10	12	8	20	59	67
		total	476	65	103	157	50	167	542

Table 2: Frequency of words by pitch accent and IS labels. Unacc stands for unaccented words and Acc indicates accented words.

From **Table 2** we can already see substantial variation in the distribution of accented and unaccented words across the various IS categories, which means that if there is a systematic relationship between pitch accent and IS, it will be probabilistic, and not deterministic. This is further visualized in the heat map of **Figure 3**, which shows the distribution of accented words grouped by *r*-level (x-axis) and *l*-level (y-axis). The color coding of the heat map represents the proportion of accented words in the whole corpus ($n = 1018$), with darker blue expressing higher proportions. In each cell, the number before the slash shows the number of accented words in a category (e.g., $n = 69$ for *r-new/l-new*), and the number after the slash is the total number of words in each category (e.g., $n = 76$ for *r-new/l-new*). If there was a systematic relationship between pitch accent and IS, we would observe that the proportion of accented words would

gradually increase (i.e., the cell color in the heat map would gradually become darker) from the word with the lowest assumed level of ‘informativeness’ (i.e., *NR/NL* on the lower left) to that with the highest level of informativeness (i.e., *r-new/l-new* on the upper right cell), which turns out not to be the case.

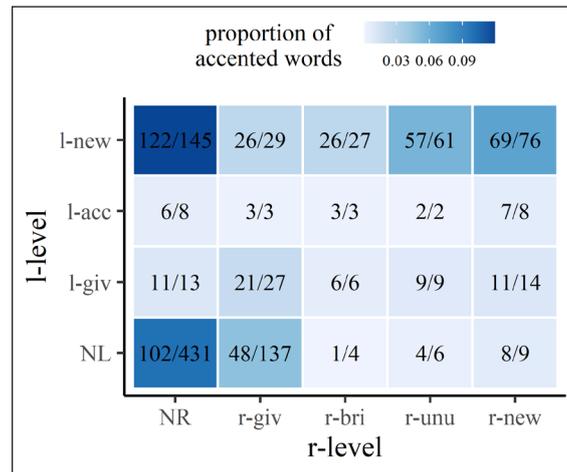


Figure 3: Distribution of accented words grouped by *r*-level (x-axis) and *l*-level categories (y-axis).

Along these lines, if pitch accent is used to convey distinctions in IS at any level, then words that convey IS distinctions are expected to be accented with greater frequency than words that do not convey IS distinctions. In terms of the IS labels in our data, this means that a pitch accent of any type is expected to be more frequent in words that are specified for any of the referential, lexical, or alternative IS labels, than in words that are specified as *NR*, *NL*, and *non-alt* (i.e., undefined for IS). To test this prediction, we first categorized all words into two types: *Accented* (includes all of the pitch accent types), and *unaccented*. We similarly categorized words according to their IS label as *none* (corresponding to *NR*, *NL*, and *non-alt*), and *any* (all other IS labels at the *r*-, *l*-, and *alt*-levels), and tested for differences in the distribution of these categories using Pearson’s chi-squared test. Three independent Pearson’s chi-squared tests were submitted for the *r*-, *l*-, and *alt*-levels.

In **Table 3**, the results show that the presence or absence of accent varies significantly based on IS status, for the *r*-, *l*-, and *alt*-levels. Looking at the frequency data in **Table 2**, we see that words that convey distinctions in referential or lexical givenness, and words under contrastive focus are more likely to be accented, whereas words that do not are less likely to be accented. This result shows that the TED Talk speakers utilize accent in relation to givenness and focus distinctions.

	Level	χ^2	<i>df</i>	<i>p</i>
none/any	r-level	94.85	1	< .001
	l-level	358.97	1	< .001
	alt-level	33.44	1	< .001

Table 3: Chi-squared tests assessing a relationship between accent status (*accented* versus *unaccented*) and IS status (*any* versus *none*). See text for description of these labels.

2.3.2. Distribution of pitch accents in relation to givenness

Another way accented and unaccented words may pattern differently in relation to IS relates to the *given* category. Looking again at **Table 2**, we see that the *r-given* and *l-given* categories include both accented and unaccented words, but we are interested to know if unaccented words are more frequent in the *given* categories compared to the other categories at each level, leaving aside the *none* levels (*NR* and *NL*). To test this, we conducted a second Pearson's chi-squared test on IS labels reassigned to two categories: *Given* versus *non-given*. The *given* category corresponds to *r-given* for the referential level and *l-given* for the lexical level. The *alt*-level is not included in this analysis, since alternatives (focus) indexes a different dimension of information structure than the givenness dimension. The *non-given* category includes *r-bridging*, *r-unused*, and *r-new* for the referential level, and *l-new* for the lexical level. This analysis also used the recategorized pitch accent labels, *unaccented* versus *accented*. Two Pearson's chi-squared tests were run, for the *r*-level and *l*-level.

In **Table 4**, the results show that the distribution of accented and unaccented words is significantly different between the *given* versus *non-given* categories defined at the referential level, but not at the lexical level. In other words, the TED Talk speakers preferentially treat co-referring expressions (*r-given*) as unaccented, and they are relatively more likely to assign pitch accents to words in referential expressions when the referent is either new, or accessible from the discourse context. However, the speakers do not show a similar preference to treat lexically given words as unaccented. The results indicate that these speakers consider referential and lexical givenness differently for purposes of accent assignment, supporting a model of IS that distinguishes the referential and lexical levels, as proposed by the RefLex scheme.

	Level	χ^2	<i>df</i>	<i>p</i>
given/non-given	r-level	81.2	1	< .001
	l-level	.77	1	<i>n.s.</i>

Table 4: Chi-squared tests for accentedness of words delivering given information.

2.3.3. Distribution of pitch accent types at each IS level

Our final test of differences in the distribution of accented and unaccented words looks for differences in the type of pitch accent associated with distinct IS categories. For this, we ran a series of Fisher's exact tests. Words that are not defined for givenness (*NR*, *NL*) or focus (*non-alt*) are excluded in this analysis. The numerical values in **Table 2** are plotted in **Figure 4** for the referential, lexical, and alternative levels, respectively (omitting the *NR*, *NL*, and *non-alt* labels). The two panels in each figure plot the same data, grouped by accent type (left panel), or by IS category (right panel) on the y-axis. For both panels, the x-axis shows the count number of occurrences.

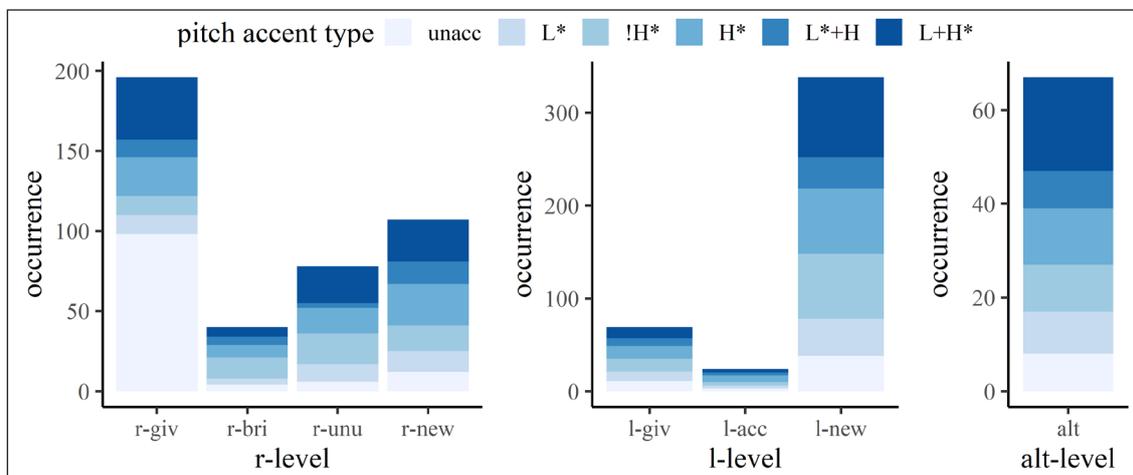


Figure 4: Distribution of pitch accent type and *r*-level (left), *l*-level (middle), or *alt*-level (right) categories for words.

Figure 4 confirms the finding already apparent in **Table 2**, namely that the hypothesized deterministic mapping between the Prominence and Givenness hierarchies (as sketched in **Figure 2**) does not hold in the TED Talk speech samples. Rather, all accent types are observed in each IS category at the referential, lexical, and alternative levels. To test for differences in the distribution of accent types across the referential IS categories, Fisher's exact test was conducted for the referential and lexical IS levels, retaining the full set of accent labels and IS labels (excluding *NR* and *NL*). We do not test the association between pitch accents and the *alt*-level using Fisher's tests because the *alt*-level has only one category (*alt*).

At the referential level ($n = 421$, excluding 597 *NR* words), Fisher's exact test shows a marginally significant association between accent types and *r*-labels (two-tailed $p = .05$) that weakly conforms to the predictions from the simple mapping shown in **Figure 2**. For instance,

Figure 4 (left) shows !H* as the most frequent accent type for *r-bridging*, and L+H* as most frequent for *r-new*. These non-significant numeric trends in our data are in line with those of Baumann and Riester's (2013) study on German, which finds accent types to be probabilistically, not exclusively, associated with IS. A couple of findings from this analysis deserve special mention: As shown in **Figure 4** (left), L+H* is the most frequent accent type for *r-unused* as well as *r-new*. This finding is parallel to previous findings from a radio news corpus (Hirschberg, 1993) showing a probabilistic association between L+H* and proper nouns (labeled as *r-unused* in this study). Second, as shown in the left panel of **Figure 4**, L* frequently occurs with *r-unused* and *r-new*. In these TED Talks, we notice the frequent occurrence of rising declaratives in non-question contexts, which impressionistically signals discourse continuation. We will see below (**Figure 8**) that many instances of L* occur in such contexts, as nuclear accents marking new referents, followed by high edge tones (H-, H-H%) that seem impressionistically to signal discourse continuation.

At the lexical level ($n = 431$, excluding 587 *NL* words), Fisher's exact test again shows no significant association between accent types and *l*-labels (two-tailed $p = .88$), but with numeric differences that conform to predictions, e.g., L+H* is the most frequent accent type for *l-new* (**Figure 4**, middle).⁷ At the *alt*-level, as shown in the right panel of **Figure 4**, speakers use all accent types in association with focus, though favoring L+H*. A surprising finding is the frequent occurrence of unaccented words under focus. These are mostly function words (e.g., *was*, *have*, *can*) marked for contrast with other function words.

2.3.4. Distribution of pitch accent types at each IS level in relation to accent position

Figure 4 reveals a probabilistic relationship between pitch accent types and IS at each level of IS analysis. One may wonder if such relationships vary according to the position of the pitch accent in the prosodic phrase. For instance, it is possible that nuclear pitch accents are more strongly related to IS than are prenuclear pitch accents, which are sometimes described as serving a 'rhythmic' function, or just as 'ornamental' (Calhoun, 2010; Chodroff & Cole, 2018; Vogel et al., 1995). **Figures 5–7** show the relationship between the distribution of pitch accent types and IS categories at the referential, lexical, and alternatives levels, now paneled by pitch accent

⁷ There is an unexpected finding in the distribution of accents for the *l-given* category, where we predicted mostly (or only) unaccented words. Instead, *l-given* occurs most frequently with accents that are fairly high on the Prominence hierarchy: H*, !H* and L+H*. Further inspection shows that many instances of *l-given* are words that the speakers repeat often in the TED Talks (e.g., *thirty*, *day*, *month*), where the use of prominent accents seems to signal topical emphasis. The finding of no significant difference in the distribution of accented words in the *l-given* category compared with the combined *l-accessible* and *l-new* categories (Table 4) is likely due to the frequency of these emphasized *l-given* words.

position: *Initial* and *middle* for prenuclear pitch accents, and *final* for the nuclear pitch accents, as described in 2.2.2.⁸

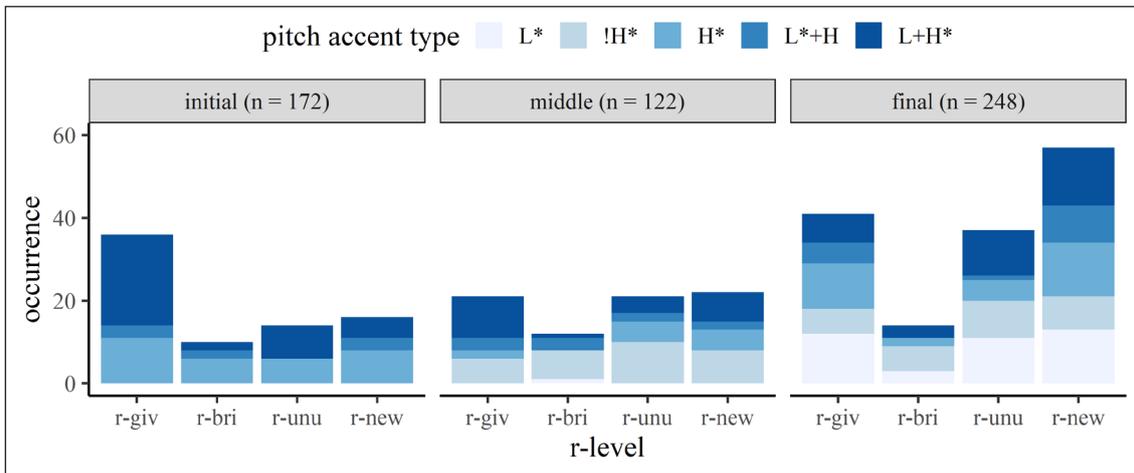


Figure 5: Distribution of pitch accent type and *r*-level categories for words, paneled by accent position in the utterance.

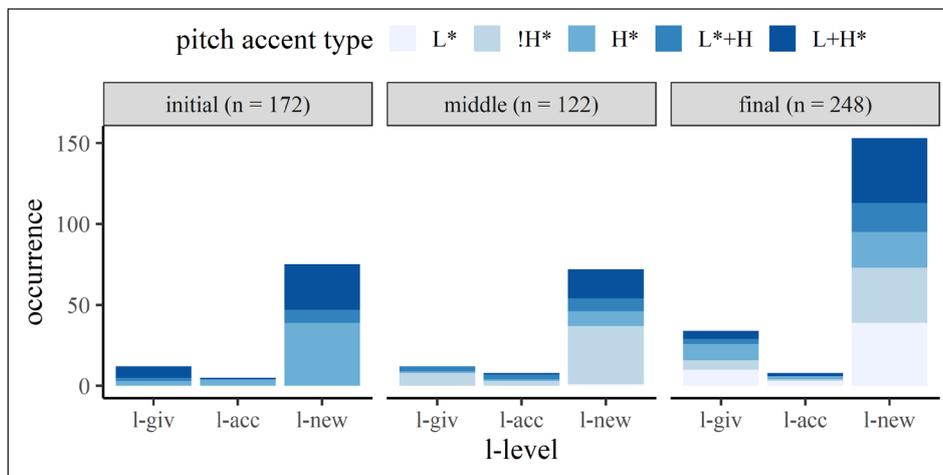


Figure 6: Distribution of pitch accent type and *l*-level categories for words, paneled by accent position in the utterance.

⁸ The number of words for each accent position is indicated in parentheses. In our speech materials, there are many prosodic phrases that have only a single pitch accent. Consequently, there are more final pitch accents than there are initial and middle pitch accents.

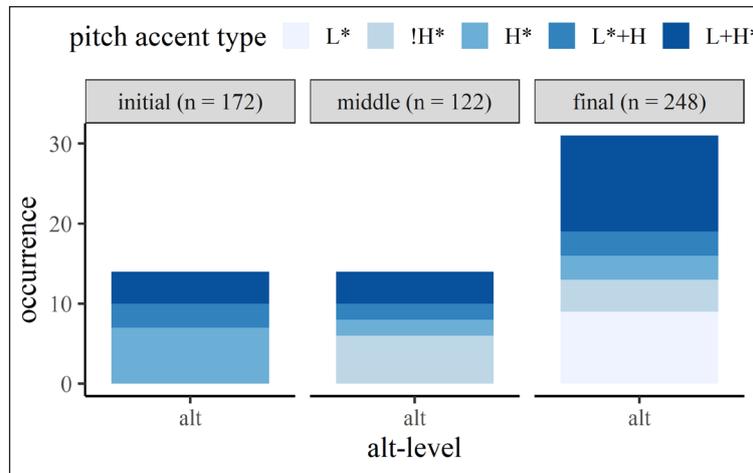


Figure 7: Distribution of pitch accent type and the single *alt*-level category for words, for comparison with Figures 5 and 6. Graph is paneled by accent position in the utterance.

In Figures 5–7, as observed in Figure 4, all or nearly all accent types occur in each IS category, both for accents in nuclear (*final*) position, and for prenuclear pitch accents (*initial*, *middle*), indicating at best a probabilistic association between accent type and IS category regardless of accent position. We do, however, find that the specific associations between pitch accents and IS predicted from the simple hypothesized model in Figure 2 are slightly more frequent with nuclear pitch accents than prenuclear ones in these data. For instance, Figure 5 shows that among nuclear (*final*) pitch accents, L* is the most frequent accent type for *r-given*, !H* for *r-bridging*, and L + H* for *r-new*. Similarly, Figure 6 shows that the most frequent nuclear (*final*) accent types for *l-given* are L* and H*, while L + H* is most frequent for *l-new*. At the alternative level (Figure 7), L + H* is the most frequent accent for nuclear words with the *alt* label, though not for prenuclear *alt* words.

These weak asymmetries observed in the distribution of accent types across IS categories in nuclear position are not apparent in the distribution of prenuclear (*initial* and *middle*) accents. At the *r*-level, as shown in Figure 5, among *initial* pitch accents, H* is the most frequent accent type for *r-bridging* and *r-new*, but L + H* is most frequent for both *r-given* and *r-unused*. L* never occurs in the initial position, even for *r-given* words. In *middle* position, !H* is the most frequent accent type for all *r*-labels, except *r-given*, where L + H* is unexpectedly the most frequent accent type. At the *l*-level, shown in Figure 6, among *middle* accents, !H* is the most frequent accent type for all *l*-labels, while among *initial* accents L + H* is again unexpectedly the most frequent accent type for *l-given*, while H* is the most frequent accent for *l-accessible* and *l-new*. Finally, turning to the *alt*-level, Figure 7 shows that for prenuclear pitch accents, H* is the most frequent accent type associated with the *alt* category in initial position, while !H* is most frequent in non-initial

(middle) position. Taken together, these results indicate a weak, probabilistic relationship between accent type and IS for words with nuclear (*final*) accent, while words with prenuclear accents exhibit somewhat fewer accent types, which however do not appear to depend on IS category.

2.3.5. Distribution of nuclear pitch accent types at each IS level in relation to edge tones

The final step in the distributional analysis of accents in relation to IS categories takes into consideration effects from the combination of nuclear pitch accent and edge tones—the nuclear tune. As noted above, **Figure 4** (left) shows L* occurs frequently with *r-unused* and *r-new* words, which we suggested may reflect a pattern of phrase-final rising pitch, in sequences like L* H- or L* H-H%. **Figures 8–10** show the relationship between nuclear pitch accent type and IS categories at the referential, lexical, and alternative levels, paneled for the edge tones following the nuclear accent. These figures confirm that at all IS levels, L* is the most frequent accent type preceding the high edge tones H- and H-H%. This context accounts for 58% of all L* tokens in our speech materials. **Figure 8** shows that the rising tunes (L* H- and L* H-H*) are most frequently associated with referentially new or accessible (*r-unused*) information, while **Figure 9** shows falling tunes (!H* L- and !H* L-L%) are most frequent with lexically new information. At the alternative level, **Figure 10** shows no clearly preferred pairing of the L+H* pitch accent and edge tones. This analysis reveals different pairings of nuclear tunes with IS categories related to givenness, though not alternatives, which provides some evidence for a model of IS that distinguishes referential and lexical givenness.

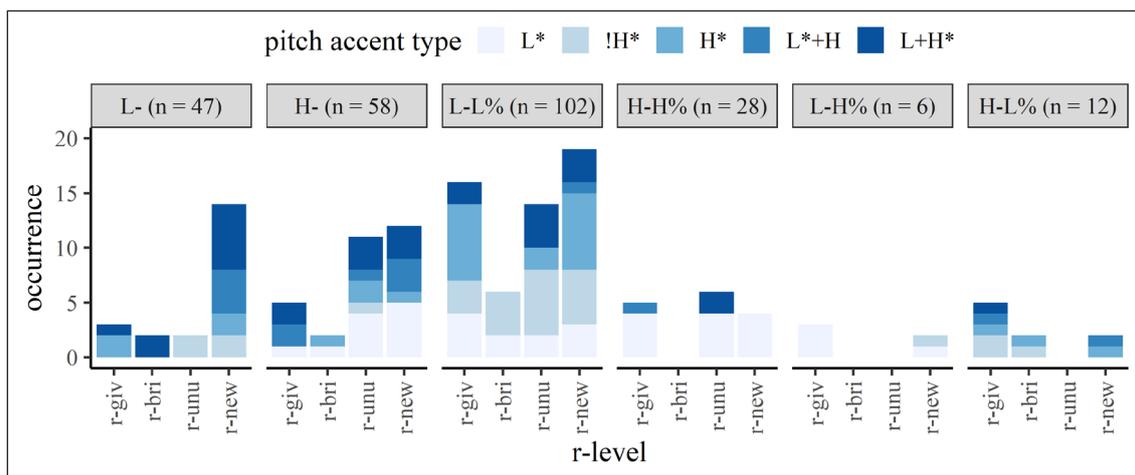


Figure 8: Number of words (y-axis) associated with *r*-labels and nuclear pitch accents, paneled by edge tones.

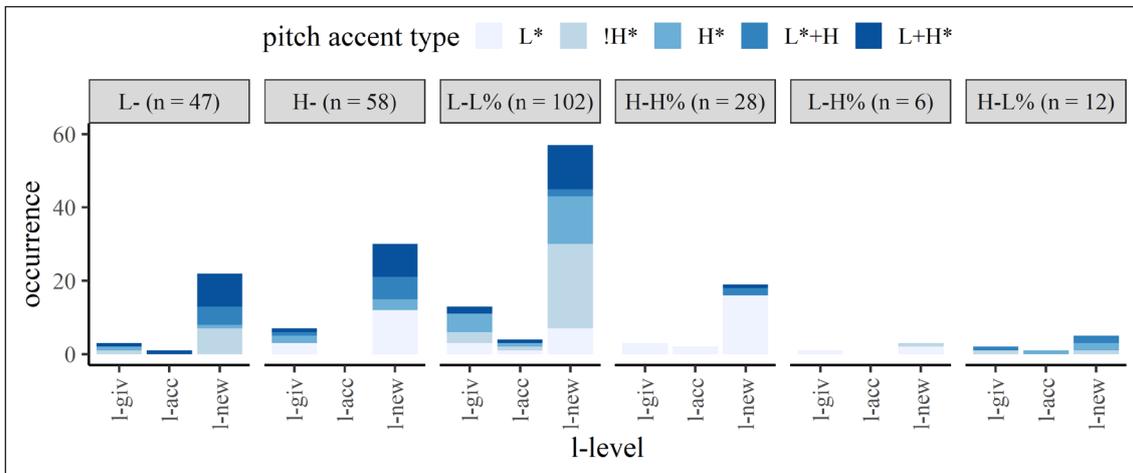


Figure 9: Number of words (y-axis) associated with *l*-labels and nuclear pitch accents paneled by following edge tones.

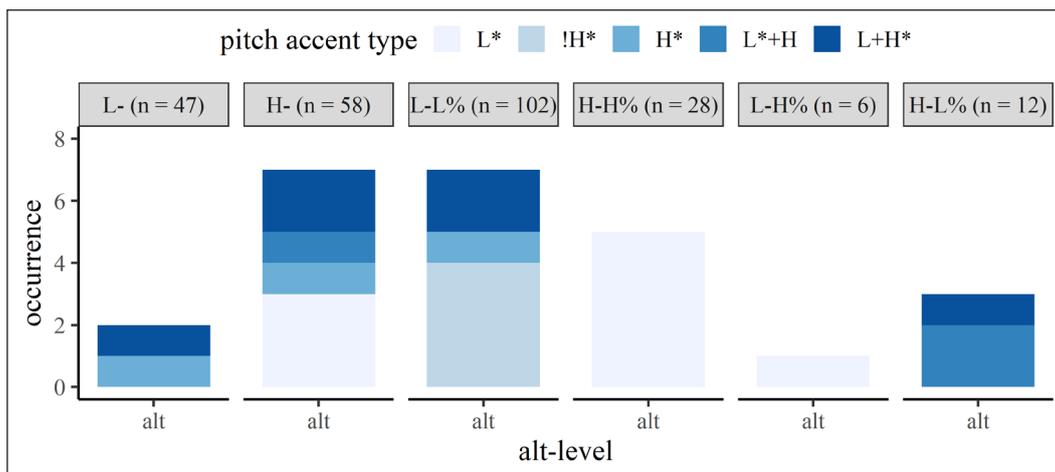


Figure 10: Occurrences of words (y-axis) for *alt*-labels and nuclear pitch accents in relation to edge tones.

Overall, our analysis of the distributional relationship between accent type and IS category confirms that the speakers in these TED Talks use pitch accents systematically in relation to IS, but primarily in terms of the presence versus absence of accent. Accent, especially, the nuclear pitch accent, is preferentially used with words that convey IS distinctions, and among those, referentially or lexically given words are preferentially unaccented. Looking at tonally specified pitch accent types, we observe non-significant numeric preferences for the hypothesized,

conventional mappings between accent types and IS (as sketched in **Figure 2**), but only for nuclear accents. There is no apparent relationship between prenuclear accent type and IS categories.

2.3.6. Acoustic correlates of prominence

2.3.6.1. Models for the entire dataset

We now move on to examine how IS and pitch accents are reflected in the acoustic implementation of prominence distinctions in our speech samples. For this, we carried out a multivariate multiple regression to model variation in acoustic cues (max F0, F0 range, phone rate, and mean intensity) as the dependent variables in relation to the TED Talk speaker (male, female), the IS category of the word (*r*-, *l*-, *alt*-labels), its pitch accent (unaccented word, L*, !H*, H*, L* + H, L + H*), edge tone (word boundary, L-, H-, L-L%, H-H%, L-H%, H-L%), and interaction between IS category and pitch accent as independent variables. Adopting the hypothesized prominence ranking over IS categories and pitch accents, these factors were modeled using successive difference contrast coding, where each level within a factor was compared to the mean of the previous level. The ordering of levels for IS and pitch accents followed the Givenness and Prominence hierarchies in **Figure 2**. The factors of TED Talk speakers and edge tones were modeled using dummy coding, where each level was compared to the reference level, the male speaker for the TED Talk speaker, and the (phrase-medial) word boundary for the edge tones. The multivariate multiple regression model was run in *R* (R Core Team, 2019) using the *lm* function⁹ and was visualized using the *visreg* package (Breheny & Burchett, 2017). As this model was run with all the words in the dataset as input, we refer to this as the *All-words* model. The model output includes results from univariate multiple regression for each dependent variable separately; these are combined and submitted to Type II MANOVA (using the *Anova* function in *R*), which tests for each independent variable whether it affects all the dependent variables, jointly. We are interested in both results: A significant effect from the univariate analysis identifies an effect on an individual acoustic dimension of prominence (e.g., max F0), while a significant effect from MANOVA indicates an independent variable that has a general, joint effect on all the measured acoustic dimensions of prominence. We report only significant effects from the univariate models in **Table 5**, and the full table of MANOVA results in **Table 6**. Among

⁹ The specified R function for the *All-words* multivariate multiple regression is as follows: `lm(cbind(max F0, F0 range, phone rate, mean intensity) ~ TED Talk speaker + r-level + l-level + alt-level + pitch accent + edge tone + r-level: pitch accent + l-level: pitch accent + alt-level: pitch accent)`

the significant effects from the univariate models in **Table 5**, we visualize a couple of key significant effects in **Figure 11**.

	IS	Pitch accent	Edge tone	IS x Pitch accent
max F0	l-access < l-new	L* < unacc L* < H* L* + H < L + H*	L-L% < word	NL: L* + H < l-given: L + H*
F0 range	r-new < r-unused	H* < L* + H	word < L- word < L-L% word < H-H% word < L-H%	r-bridg: L* < r-giv: unacc r-giv: L* < r-bridg: !H* r-bridg: unacc < r-unus: L* r-unus: !H* < r-bridg: L*
phone rate	—	—	word < L-L%	l-access: L* < l-giv: unacc l-new: L* < l-giv: unacc
mean intensity	—	—	L-L% < word	l-giv: L* < NL: unacc NL: L* < l-giv: !H*

Table 5: Summary of results from the *All-words* univariate models for each dependent variable in the multivariate multiple regression modeling acoustic cues as a function of speaker, IS, pitch accent, edge tone and interaction between IS and pitch accent, over all words in the TED Talk samples. Significant results are indicated for each acoustic variable (rows) and all predictor variables (columns). See text for details. Bolded effects are in the direction predicted by the hypothesized prominence rankings; effects in red are opposite to those predictions. There were no significant effects of speaker on any acoustic variable.

Variable	Pillai's value	F	Hypothesis df	Error df	p
speaker	.00	.22	4	954	<i>n.s.</i>
r-level	.02	1.17	16	3828	<i>n.s.</i>
l-level	.03	2.34	12	2868	< .01
alt-level	.00	.83	4	954	<i>n.s.</i>
pitch accent	.28	14.26	20	3828	< .001
edge tone	.16	6.57	24	3828	< .001
r-level: pitch accent	.09	1.08	80	3828	<i>n.s.</i>
l-level: pitch accent	.08	1.37	60	3828	< .05
alt-level: pitch accent	.02	.93	20	3828	<i>n.s.</i>

Table 6: Results from the *All-words* MANOVA modeling acoustic cues as a function of speaker, IS, pitch accent, edge tone, and interaction between IS and pitch accent over all words in the TED Talk samples. Bold face indicates statistically significant effects.

Significant results from the univariate *All-words* models are summarized in **Table 5** for each acoustic variable (rows) and all predictor variables (columns). Significant effects ($p < .05$) are listed in the cells, indicating the level and direction of the effect (e.g., in the max F0 row, *l-acc < l-new* indicates a significant effect of IS, where lexically accessible words are estimated to have a lower max F0 than lexically new words, when other effects are held at their average). Note that phone rate effects are reported as inverse of the rate measure, so that a higher rate corresponds to shorter durations, and a lower rate to longer durations. Effects that are in the direction predicted by the hypothesized prominence rankings are bolded while effects that are in opposite direction to those predictions are in red. Complete model output, including effect sizes, is included in the Online Supplement, **Table 1**. Relative to our hypotheses, there are two important observations from **Table 5**:

- There are relatively few main effects of IS or pitch accent; they are observed only with F0 measures, and only some are in the predicted direction. There are many significant interactions of IS and pitch accent, but again, only some are in the predicted direction. Notice that many of the effects that are in the opposite direction involve the L* accent and its interactions with IS. The unexpected acoustic prominence of these L* words may relate to their frequent occurrence in phrase-final rising nuclear tunes (L* H- and L* H-H%, see **Figure 8**), an intonation style resembling “uptalk.” These rising tunes are frequently observed with unused and new IS (**Figures 8, 9**); hence, the acoustic prominence of these L* words may relate more to their IS status and/or edge tone condition than to their accent status.
- The most common effect of edge tone is the distinction between the word boundary condition (with no marked tone feature) and tonally marked boundaries, especially for F0 range. With two exceptions (max F0 and mean intensity), all of the edge tone effects are in the predicted direction. The two exceptions concern the relationship between the word boundary and L-L% conditions. A phrase-final word marked for L-L% has lower max F0 and lower intensity than a non-final word, which is not surprising. In this case, the low F0 and intensity are good cues for the phonological low tones at this location. This finding suggests that it is not sufficient to formulate blanket hypotheses about prominence rankings among prosodic boundary levels that do not also take into account the tonal specification of those boundaries (which were not tested as interaction terms in our statistical model).

We observe in **Table 5** that only F0 measures show main effects of IS or pitch accent, which is further visualized in **Figure 11**. The figure shows the predicted values of two F0 measures (max F0 and F0 range) in relation to IS (x-axis) and pitch accent type (y-axis) based on the results from the univariate models. For each cell, the number indicates the total number of words in a category. In the left figure, the predicted max F0 is indeed likely to be higher (the shade of cell

is darker) for *l-new* than *l-accessible* (the significant difference indicated in **Table 5**). In the right figure, the predicted F0 range tends to be higher for *r-unused* than *r-new* (cf. **Table 5**).

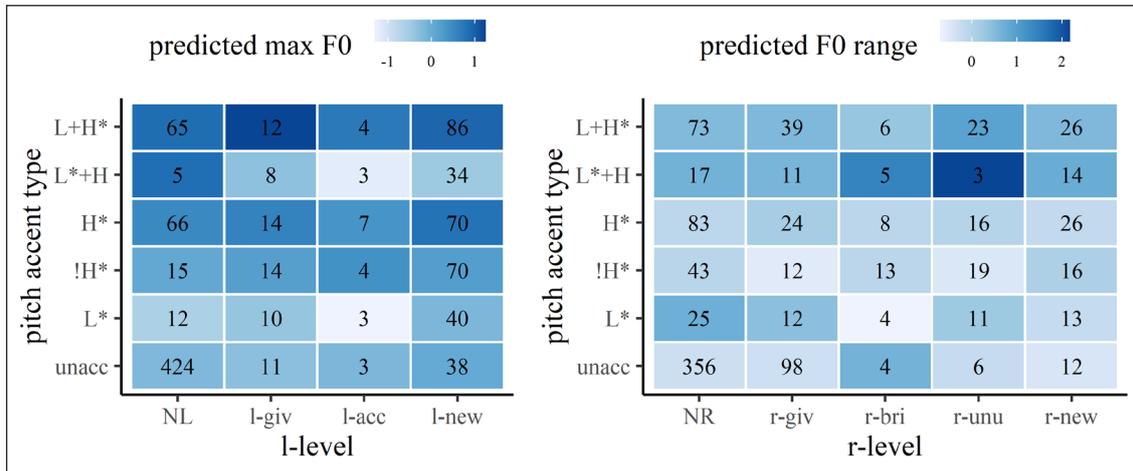


Figure 11: Predicted F0 measures for *All-words* grouped by pitch accent type and *l-level* (left) and by pitch accent type and *r-level* (right).

In **Table 6**, the MANOVA test on the combined output of the multivariate multiple regression confirms main effects of lexical givenness (*l-level*), pitch accent, and edge tone on the joint set of acoustic variables. There are also effects from the interaction between *l-level* and pitch accent. These effects from the multivariate model were also significant in the univariate models for at least some of the individual dependent variables (**Table 5**). The multivariate model returns no effects of referential givenness or focus (*alt*), and no interaction effects between *r-level* and pitch accent, or focus (*alt*) and pitch accent. The absence of effects of *r-level* and *alt-level* suggests that any general acoustic effects related to referential givenness and focus are already captured in the main effects from the phonological factors of pitch accent and edge tones. In contrast, lexical givenness (via prior mention of a lexeme or related word) has a direct effect on the joint acoustic correlates of prominence, above and beyond the variation due to pitch accent and edge tones, with additional effects in the interaction of *l-level* and pitch accent. The absence of an effect of speaker indicates that the male and female TED Talk speakers have similar patterns of variation for these acoustic parameters in relation to the independent variables.

2.3.6.2. Models for accented words

We further explored the effect of phonological factors on the acoustic correlates of prominence in accented words only to examine the effect of accent position. The *Accented-words* model excludes unaccented words and otherwise differs from the *All-words* model only in the addition of accent

position (middle, initial, final) as an independent factor, using successive difference coding.¹⁰ The *Accented-words* model has around half the number of observations ($n = 476$) as the *All-words* model ($n = 1018$, including unaccented words).

Significant results from the *Accented-words* univariate models are summarized in **Table 7**, following the format of **Table 5**. Though there are few main effects of IS and pitch accent on these acoustic variables, the observed effects are similar to what we found for the model of all words (**Table 5**), with the difference being that all but one of the effects on accented words are in the predicted direction (given that the edge tone effects were expected after all, see previous section). There is only one effect of pitch accent position, in greater max F0 with phrase-initial accents compared to medial, which may be attributed to compression of the pitch range across the extent of the prosodic phrase. Edge tone effects are reduced in this dataset, preserving only the effects on max F0 and mean intensity, but not F0 range or phone rate, of the full dataset in **Table 5**. Finally, this smaller dataset has only a few interaction effects, and all in the predicted direction (unlike **Table 5**). One new effect appears in the accented word data—the female TED Talk speaker produces accented words with lower (normalized) mean intensity than the male TED Talk speaker.

Figure 12 is a visualization of the two F0 measures (max F0 and F0 range) in relation to IS (x-axis) and pitch accent type (y-axis) based on the results from the univariate models. The patterns are similar with those in **Figure 11** above from the *All-words* model. In the left figure, the predicted max F0 is likely to be higher for *l-new* than *l-accessible* (as indicated in **Table 7**). In the right figure, the predicted F0 range tends to be higher for *r-unused* than *r-new* (as shown in **Table 7** as well).

Table 8 shows the *Accented-words* MANOVA results for joint effects on the acoustic variables, which are mostly the same for accented words as in the *All-words* model (**Table 6**). There are significant effects of lexical givenness (*l-level*), pitch accent, and edge tone. We see a main effect of the male vs. female TED Talk speaker, which was observed only for mean intensity in the univariate models. There are no direct effects of referential givenness, focus, accent position, nor are there any significant interactions between IS and pitch accent.

It would be interesting to test for similar effects on acoustic variation among only the nuclear accented words, excluding phrase-medial words (i.e., eliminating the word-boundary level of the edge tone condition from the model of all accented words in **Table 7**). Unfortunately, given the distribution of the 134 nuclear accented words in our dataset, there is insufficient data to test for effects across even a minimal number of pitch accent and edge tone levels.

¹⁰ The specified R function for the *Accented-words* multivariate multiple regression is as follows: `lm(cbind(max F0, F0 range, phone rate, mean intensity) ~ TED Talk speaker + r-level + l-level + alt-level + pitch accent + accent position + edge tone + r-level: pitch accent + l-level: pitch accent + alt-level: pitch accent)`

	Speaker	IS	Pitch accent	Accent position	Edge tone	IS x Pitch accent
max F0	—	l-access < l-new	L* < !H* L* + H < L + H*	mid < init	L-L% < word	NL:L* + H < l-giv: L + H*
F0 range	—	r-new < r-unus l-given < NL	H* < L* + H	—	—	r-giv:L* < r-bridg: !H*
phone rate	—	—	—	—	—	—
mean intensity	F < M	—	—	—	L-L% < word	NL: L* < l-giv: !H*

Table 7: Univariate results for each dependent variable in the *Accented-words* multivariate multiple regression modeling acoustic cues as a function of speaker (male, female), IS, pitch accent, accent position, edge tone, and interaction between IS and pitch accent, from accented words only. Formatting as in Table 5.

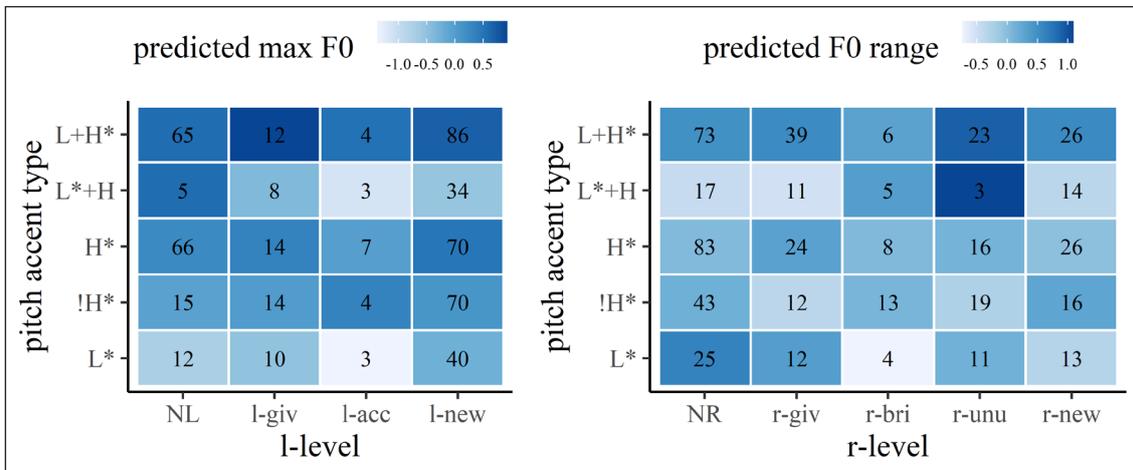


Figure 12: Predicted F0 measures for *Accented-words* grouped by pitch accent type and *l*-level (left), and by pitch accent type and *r*-level (right).

Variable	Pillai's value	<i>F</i>	Hypothesis <i>df</i>	Error <i>df</i>	<i>p</i>
speaker	.02	2.64	4	485	< .05
r-level	.03	.83	16	1952	<i>n.s.</i>
l-level	.05	2.24	12	1461	< .01
alt-level	.01	.84	4	485	<i>n.s.</i>
pitch accent	.25	8.09	16	1952	< .001
accent position	.02	1.42	8	972	<i>n.s.</i>
edge tone	.13	2.68	24	1952	< .001
r-level: pitch accent	.12	.93	64	1952	<i>n.s.</i>
l-level: pitch accent	.10	1.08	48	1952	<i>n.s.</i>
alt-level: pitch accent	.03	.87	16	1952	<i>n.s.</i>

Table 8: Results from the *Accented-words* MANOVA modeling acoustic cues as a function of speaker, IS, pitch accent, edge tone, accent position, and interaction between IS and pitch accent from accented words only.

2.3.7. Summary of production study results

The above distributional analyses of words based on their specification for pitch accent, edge tones and IS, and the regression models showing effects of phonological and IS factors on acoustic prosodic measures, complete our study of variation in prominence production in these TED Talks. To summarize the main findings, we observe (1) a significant, though probabilistic,

relationship in the distribution of pitch accents across levels in each dimension of IS, but only in the distinction of the presence versus absence of accent, and not in accent type; (2) a non-significant numerical relationship between accentual prominence and informational prominence for nuclear accented words that is not observed for prenuclear accented words; (3) a wide variety of pitch accents and edge tones in all givenness conditions and with focus; (4) sporadic evidence of variation in acoustic measures of prominence (especially F0) as a function of pitch accent and IS, specific combinations of the two, and edge tone, with most but not all of the observed effects in the direction predicted by the hypothesized prominence rankings—enhanced acoustic prominence for words with greater prominence with respect to accent status, edge tones, and IS.

3. Prominence analyzed in perceptual ratings: Pitch accent, acoustics, and IS

3.1. Materials

The TED Talk speech materials analyzed in Section 2 were submitted to listeners in a perception experiment designed as a prominence rating task. For this task, the TED Talk speech materials were broken down into smaller excerpts of 30–39 seconds each and presented to listeners in succession, following their order of presentation in the respective TED Talks. An orthographic transcription of the speech was also obtained from the TED Talks web archive and presented to participants after removal of all capitalization and punctuation. This was done following Cole et al. (2010), to avoid the use of these text features as potential cues to prominence due to their frequent occurrence at syntactic clause boundaries, which frequently map onto prosodic phrase boundaries, and the corresponding frequent alignment of prominence with prosodic phrase boundaries.

3.2. Participants and task

Seventy native speakers of AE (30 males and 40 females, mean age 29) participated in the perception experiment. All the participants self-reported no known deficits in hearing or vision. Listeners were instructed to mark words for prominence in real time, while listening to the speech excerpts, using the method of Rapid Prosody Transcription (Cole et al., 2010; Cole & Shattuck-Hufnagel, 2016). 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.” We refer to this task as prominence *rating*. The participants had no prior experience with prosody, and during the experiment, no training or feedback was offered. Using Rapid Prosody Transcription, Cole et al. (2019) found differences in prominence ratings when prominence was described in terms of discourse meaning versus acoustic properties. Given that, and considering that our instructions referred to the auditory properties of speech,

we focused our analysis on phonological and phonetic criteria for prominence, rather than on semantic or pragmatic factors. Participants were divided into two groups of 35, each listening to and rating prominence for only one of the TED Talks. They were able to listen to each speech excerpt twice, in entirety and without control of the audio playback, while viewing a transcript of the same excerpt on a computer screen. The experiment was conducted online using the custom online interface called Language Markup and Experimental Design Software (Mahrt, 2013). Participants performed the task remotely, in a quiet location of their choosing, using their own computer and headphones. The experiment took less than 30 minutes. All the participants received monetary compensation.

3.3. Results

Section 2 examined how signal-based properties (pitch accents and acoustic cues) pattern in relation to information structure (IS) in the speech produced by the two TED Talk speakers. We now turn to examine how IS, pitch accents, and acoustic cues might influence prominence perception by linguistically untrained listeners. For this, two generalized linear mixed-effects models (GLMMs) were run in *R* (R Core Team, 2019) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), an *All-words* model examining prominence ratings over all the words in the dataset ($n = 35630$) and an *Accented-words* model examining prominence ratings for only the accented words ($n = 18970$).¹¹ As in the analysis of the production data in Section 2.3, the *Accented-words* model included the additional independent variable of accent position, and was otherwise identical to the *All-words* model. Both models showed significant effects similar in size and in the same direction for some but not all levels of referential and lexical givenness, pitch accent type, and acoustic variables. Both models also showed significant interaction effects between IS factors, pitch accent type, and acoustic variables, and although the models differed in which levels of each factor were involved in a significant interaction, the models were similar in the direction of effects, and in the overall pattern. In the interest of space, we present only

¹¹ We also ran a model using only nuclear accented words to test for a possible interaction between pitch accent type and the following edge tone on prominence ratings. Unfortunately, because of sparse or missing data for some of the combinations of pitch accent and edge tones, we were unable to obtain a clear analysis of this interaction in our dataset. The specified R functions for the two GLMMs are as follows: For the *All-words* model, `glmer(perceived prominence ~ TED Talk speaker + r-level + l-level + alt-level + pitch accent + edge tone + rhythm + PC1 + PC2 + PC3 + r-level: PC1: PC2: PC3 + l-level: PC1: PC2: PC3 + alt-level: PC1: PC2: PC3 + pitch accent: PC1: PC2: PC3 + r-level: pitch accent + l-level: pitch accent + alt-level: pitch accent + (1 | subject))`; For the *Accented-words* model, `glmer(perceived prominence ~ TED Talk speaker + r-level + l-level + alt-level + pitch accent + edge tone + rhythm + accent position + PC1 + PC2 + PC3 + r-level: PC1: PC2: PC3 + l-level: PC1: PC2: PC3 + alt-level: PC1: PC2: PC3 + pitch accent: PC1: PC2: PC3 + r-level: pitch accent + l-level: pitch accent + alt-level: pitch accent + pitch accent: accent position + (1 | subject))`.

results from the *Accented-words* model here, and we refer the reader to the Online Supplement for results from the *All-words* model.¹²

The dependent variable (in both models) was the binary response of perceived prominence, coded as 1 for words rated as prominent and 0 for those not marked as prominent. Fixed factors in the *Accented-words* model were TED Talk speaker, IS (*r*-, *L*-, and *alt*-levels), pitch accent type, edge tone type, rhythm, accent position, and acoustic cues, including interactions between IS and pitch accent type, pitch accent type and accent position, IS and acoustic variables, and pitch accent type and acoustic variables. Similar to the coding scheme for the *Accented-words* model in production (Section 2.3.6.2), we used the successive difference coding for all the fixed factors, following the hypotheses in Section 1 (H1-H6), except TED Talk speaker and edge tone type, which were modeled using dummy coding. To analyze the effects of acoustic cues on prominence ratings, we used principal component analysis (PCA) in *R* (R Core Team, 2019) to reduce dimensionality and avoid potential issues of collinearity among the four acoustic variables measured (F0 max, F0 range, phone rate, intensity).¹³ Three PCs explained 89% of the variance in the acoustic variables. With positive loadings for all four acoustic measures, PC1 is tracking a general pattern of acoustic enhancement, though dominated by the F0 variables (max and range), which had the highest factor loadings in this component. PC2 captures an inverse pattern of co-variation between tempo (inverse phone rate) and intensity as the dominant factors—decreased tempo (locally slower speech rate) patterns with decreased intensity. This acoustic pattern is characteristic of words in phrase-final position that undergo boundary-related lengthening unrelated to accentual prominence. The first two components capture 67% of the variation among the four acoustic variables. PC3 captures another 21%, and like PC2, the dominant factor loadings are for tempo (inverse phone rate) and intensity, which however co-vary in PC3, with negative loadings—increased tempo patterns with lower intensity, a characteristic of unaccented words, low on the Prominence hierarchy. These three PCs were submitted to the models as independent variables, in place of the four acoustic variables. Two additional factors, the TED Talk speaker and rhythm, were also included as fixed factors. The rhythm factor captures whether an individual listener's

¹² To briefly summarize the differences between these models, there was a significant main effect of pitch accent (for *L**) and one effect related to acoustic measures (tempo and intensity) in the *All-words* model that were not present in the *Accented-words* model, while the *Accented-words* model had a significant main effect of referential givenness (for *r-given*) not observed in the *All-words* model. In addition, the *All-words* model returns several more significant interaction effects between levels of referential givenness and pitch accent type, and interactions of acoustic measures with levels of IS and pitch accent type. Overall, the *Accented-words* model discussed here presents a more conservative pattern of effects on prominence ratings.

¹³ As a measure of local tempo, we coded phone rate using the inverse of the phone rate measure. An increase in the inverse phone rate measure corresponds to a local tempo slowing, resulting in increased segment and syllable duration. Using the inverse phone rate measures makes it easier to interpret the principal component analysis, since for all four acoustic variables, increased values contribute to enhanced acoustic prominence—higher F0 peaks, larger F0 range, longer duration, greater intensity.

prominence ratings are alternating (rhythmic) or clumpy over adjacent words. Words have the value of 1 for the rhythm factor if the preceding word is marked as prominent by the same listener, and otherwise have the value of 0.

Significant results for fixed factors and interactions from the GLMM for the *Accented-words* model are presented in **Table 9**.

Level	est.	SE	z	p
(intercept)	-2.02	9.55	-.21	<i>n.s.</i>
r-given	-.21	.07	-3.12	< .01
r-bridging	-.34	.12	-2.85	< .01
r-unused	.58	.13	4.42	< .001
l-given	.39	.09	4.15	< .001
l-new	.53	.11	4.90	< .001
alt	.49	.07	6.90	< .001
H*	.52	.13	3.98	< .001
L-	.87	.11	8.00	< .001
H-	.62	.10	5.95	< .001
L-L%	.99	.10	9.91	< .001
H-H%	1.23	.13	9.17	< .001
L-H%	1.15	.20	5.84	< .001
H-L%	.32	.16	2.05	< .05
prominent-preceding-word	1.03	.05	21.28	< .001
PC1	.23	.02	11.90	< .001
PC3	-.10	.03	-3.48	< .001
r-given: H*	-.63	.19	-3.36	< .001
r-unused: H*	.80	.27	2.93	< .01
r-new: H*	-.43	.20	-2.13	< .05
r-given: L* + H	.61	.21	2.86	< .01
r-bridging: L* + H	-1.29	.37	-3.47	< .001
r-unused: L* + H	-1.29	.45	-2.88	< .01
r-new: L* + H	2.40	.33	7.31	< .001
r-given: L + H*	-.74	.20	-3.65	< .001

(Contd.)

Level	est.	SE	z	p
r-bridging: L + H*	1.73	.35	4.90	< .001
r-new: L + H*	-1.27	.32	-3.99	< .001
l-accessible: !H*	1.40	.40	3.50	< .001
l-new: !H*	-1.08	.34	-3.16	< .01
l-given: H*	-.71	.24	-2.96	< .01
l-given: L* + H	1.31	.33	3.99	< .001
l-accessible: L + H*	-1.74	.44	-3.90	< .001
l-new: L + H*	1.34	.40	3.32	< .001
alt: !H*	-.57	.22	-2.56	< .05
alt: H*	.51	.20	2.59	< .01
H*: initial	.53	.15	3.43	< .001
L + H*: initial	-.46	.21	-2.22	< .05
L* + H: final	-.85	.17	-4.99	< .001
L + H*: final	.92	.17	5.49	< .001
r-bridging: PC1: PC2: PC3	-.68	.17	-3.97	< .001
r-unused: PC1: PC2: PC3	.48	.17	2.76	< .01
!H*: PC1: PC2: PC3	.28	.10	2.86	< .01
L* + H: PC1: PC2: PC3	.30	.12	2.46	< .05
L + H*: PC1: PC2: PC3	-.37	.12	-3.00	< .01

Table 9: Significant results from the *Accented-words* GLMM modeling prominence ratings of accented words as a function of TED Talk speakers, IS, pitch accent type, edge tone type, rhythm, accent position, acoustic cues (PC1, PC2, PC3), and interaction. PC1-3 as described in text.

Significant effects from this model are illustrated in **Figures 13–23** below. The main effects of IS, pitch accent type, pitch accent position, edge tone type, rhythm, and acoustic cues on prominence ratings are illustrated in **Figures 13–17** from the model of accented words only (**Table 9**). These figures plot the model estimated effect of an independent (predictor) variable (x-axis) on the probability of prominence rating (y-axis), which was converted from log-odds of the outcome (prominence rating of 1) in *R* using the *visreg* package (Breheny & Burchett, 2017). Note that **Figures 13–17** use color to redundantly code the dependent variable (probability of prominence rating) for ease of comparison with the heatmaps in **Figures 19–23**, which show the interaction of two predictor variables on the same dependent variable.

Figure 13 shows the model estimates for the effects of acoustic measures on prominence rating, with acoustic measures represented by the three PCA components discussed above, shown in z-scores on the x-axis. **Figure 13** illustrates the significant effects of PC1 (left panel) and PC3 (right panel). In the left panel, higher values of PC1 (i.e., overall acoustic enhancement, notably in higher max F0 and F0 range) are associated with a greater likelihood of prominence rating. In the middle panel, there is no significant effect from PC2; acoustic patterns of slower tempo and lower intensity (higher values of PC2), which as noted above are characteristic of phrase-final position, do not increase prominence ratings. The right panel of **Figure 13** shows a negative association between PC3 and prominence rating; acoustic patterns of lower intensity and faster tempo, which as noted above are a pattern of function words, are associated with a lower likelihood of prominence rating.

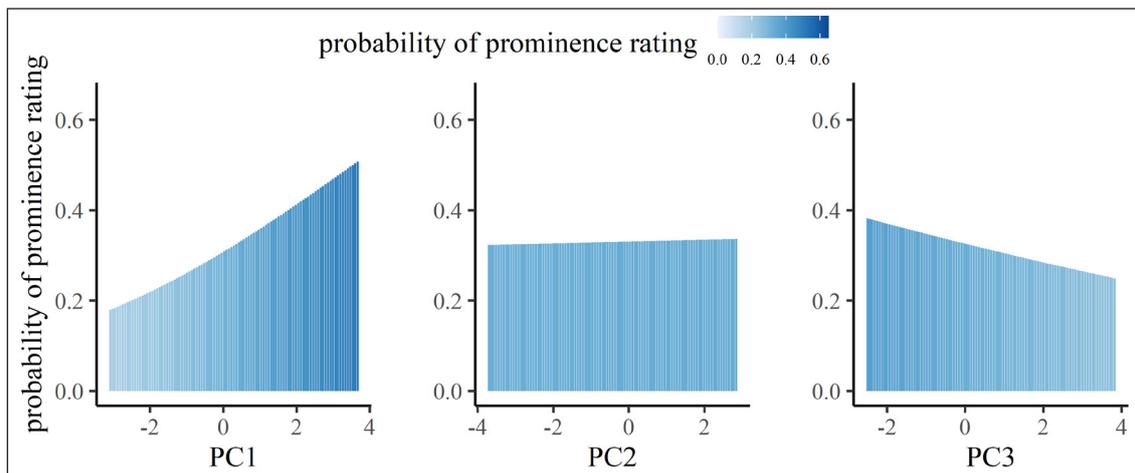


Figure 13: Predicted effects of acoustic prominence in PC1 (left), PC2 (center), and PC3 (right) on the probability of prominence ratings (y-axis and color coded, see text) for accented words (Table 9).

Figure 14 shows the predicted effects of the referential (left panel), lexical (center), and alternative (right) IS conditions on the probability of prominence ratings, with IS levels ordered left-to-right on the x-axis according to the Givenness hierarchy (**Figure 2**). These plots illustrate significant effects at each level in the likelihood of prominence ratings increasing with position on the Givenness hierarchy, from left to right, with two exceptions: Words undefined for referential givenness (*NR*) and lexically given words (*l-giv*) have higher than expected estimates for prominence rating. The lower estimate for lexically accessible words (*l-acc*) is also unexpected, but this effect was not significant in the model.

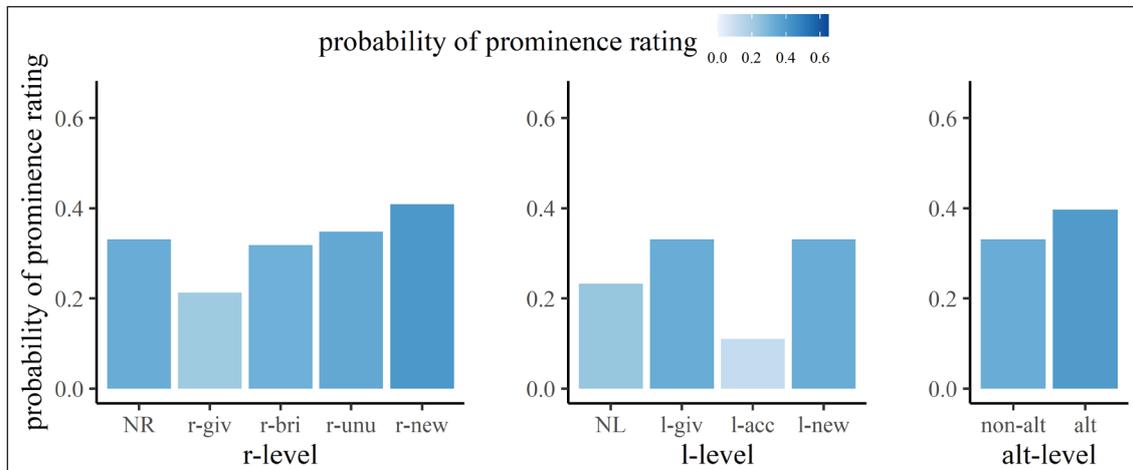


Figure 14: Predicted effects of referential givenness (left), lexical givenness (center), and contrastive alternatives (right) on the probability of prominence ratings (y-axis and color coded) for accented words (Table 9). The *r*-labels and *l*-labels are ordered left-to-right according to the Givenness hierarchy (Figure 2).

Model estimates for the effect of pitch accent type are shown in **Figure 15**, with pitch accent types ordered left-to-right on the x-axis according to the Prominence hierarchy (**Figure 2**). The statistical model reports a significant difference in the probability of prominence ratings at only one location on the Accentual Prominence hierarchy, with a higher prominence rating estimate for H^* compared to $!H^*$. In other words, based on prominence ratings, the pitch accents can be categorized into two groups: $\{L^*, !H^*\}$ and $\{H^*, L^* + H, L + H^*\}$, associated with, respectively, lower and higher likelihood of prominence rating. The model estimates plotted in **Figure 15** suggest that this grouping may result in part from the unexpectedly high prominence ratings for L^* , which we return to discuss below. **Figure 15** also shows a numeric trend with a higher likelihood of prominence rating for $L + H^*$, which, however, was not significant in the model.

Figure 16 shows the estimated effects of pitch accent position and boundary tone. Only nuclear pitch accents, i.e., those that occur in the final position of an intermediate or intonational phrase, are followed by edge tones. In **Figure 16** (right), non-final accents, i.e., those in initial or middle position in the phrase, are grouped together in the *word* level of the boundary tone type condition, while the nuclear accents in final position are grouped according to the type of edge tone that follows the pitch accent. There are some cases where the nuclear accents in final position are not immediately followed by edge tones, which are also grouped in the *word* level of the boundary tone type condition. The statistical model (**Table 9**) shows no significant differences in the effect of accent position on prominence ratings, when effects of all other variables are held constant, though a numeric trend is apparent in **Figure 16** (left) with greater effects for nuclear accented words (*final*) compared to accented words in initial and middle positions. Considering differences in the presence and type of boundary tone at the end of an accented word, the model

reports significant effects at each level, from left to right in **Figure 16** (right). Non-final accented words (in the *word* condition) have the lowest probability of prominence rating. In increasing order of effects are: Accented words followed by a H- phrase accent that end in a mid-level pitch (H-, H-L%); those with a L- phrase accent ending in a falling pitch (L-, L-L%); and those that end in a pitch contour that rises to a high final pitch (H-H%, L-H%). We return to discuss this finding, which does not follow the predicted ranking based on boundary type (word, intermediate phrase, intonational phrase), in Section 4.

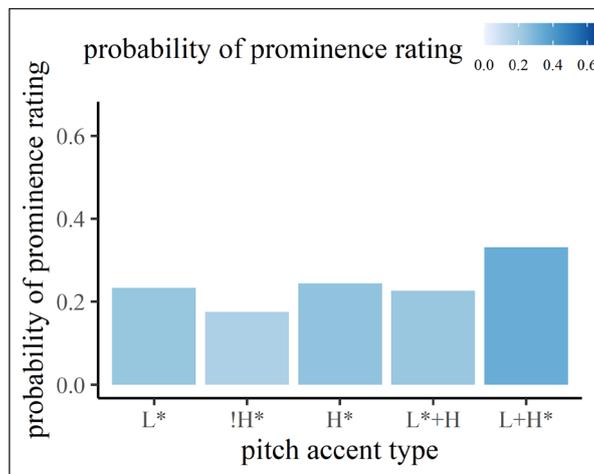


Figure 15: Predicted effects of pitch accent types (x-axis) on the probability of prominence ratings (y-axis and color coded) for accented words (Table 9). Accent types are ordered left-to-right according to the Accentual Prominence hierarchy (Figure 2).

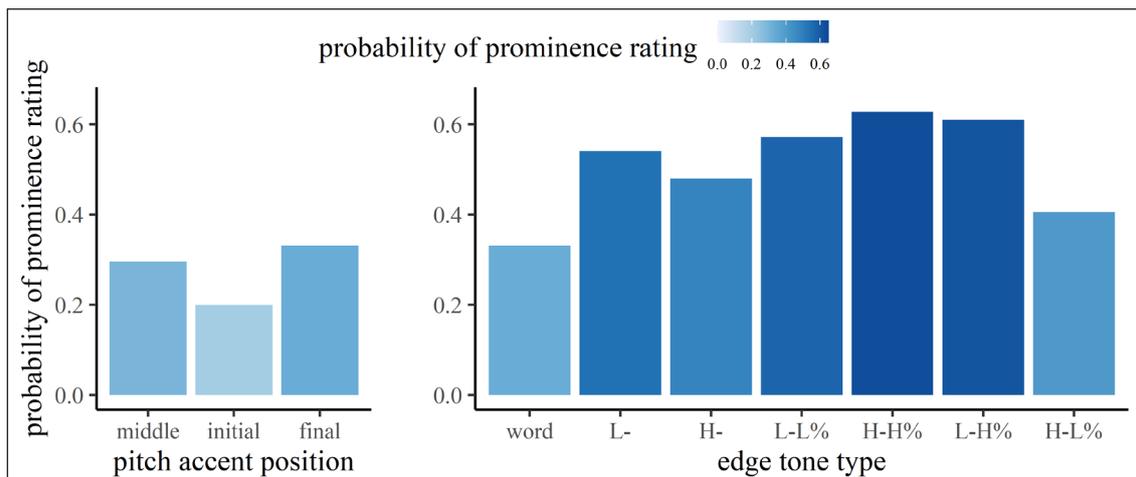


Figure 16: Predicted effects of pitch accent positions (left panel) and edge tones (right panel) on the probability of prominence ratings (y-axis and color coded) for accented words (Table 9).

The effect of perceived rhythm on prominence ratings is modeled based on whether or not the same listener rated the immediately preceding word as prominent. The model estimate for the rhythm factor is shown in **Figure 17** and shows an effect opposite to rhythm: A word is more likely to be rated as prominent if the preceding word was also rated as prominent by the same listener.

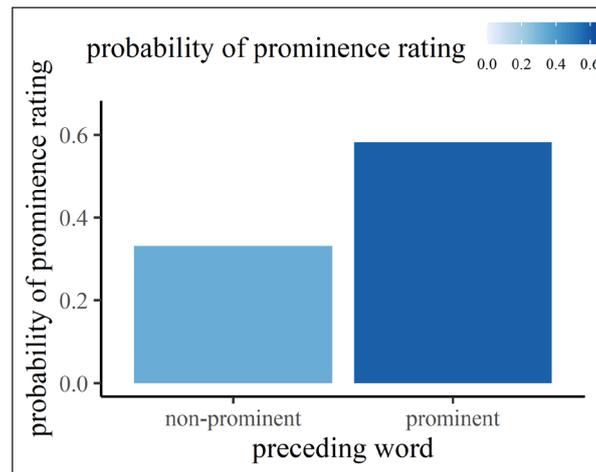


Figure 17: Predicted effects of the preceding word's prominence rating (x-axis) on the probability of the current word's prominence ratings (y-axis and color coded) for accented words (Table 9).

We now turn to the effects of interactions among IS, pitch accent type, pitch accent position, and acoustic cues on the likelihood of prominence rating. While the statistical model reports many significant interactions among these variables, for some combinations of these variables (L^* and $!H^*$) our data are sparse, which calls for cautious interpretation of the findings. First, we present the interaction between accent type and accent position. As illustrated above in **Figure 16** (left), the model shows no significant main effects of accent position on the likelihood of prominence rating, indicating that listeners weigh the nuclear and the prenuclear pitch accents similarly in their prominence ratings, holding all other factors constant. Yet as illustrated in **Figure 18**, the effect of accent type differs across the three accent positions for H^* , $L^* + H$, and $L + H^*$.¹⁴ Starting with H^* , we observe that H^* and $!H^*$ pattern alike in boosting the likelihood of prominence rating by only a small amount in middle and initial positions of the phrase, while in phrase-final position (as the nuclear accent), H^* has a larger boosting effect on prominence rating, similar to the bitonal accent $L^* + H$ (see also Bishop et

¹⁴ Note that there are no instances of L^* or $!H^*$ in phrase-initial position in our TED Talk data, although their predicted effects are estimated by the model.

al., 2020 for a similar finding). Next, while among all accents, $L+H^*$ is associated with the highest likelihood of prominence rating in the middle and final positions, a different pattern of effects is observed in initial position, where $L+H^*$ has a more modest effect on prominence rating and where L^*+H is the accent associated with the highest likelihood of prominence rating.

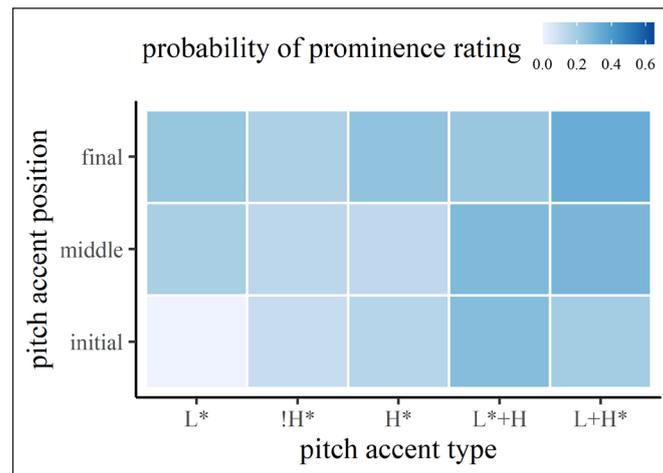


Figure 18: Predicted effects of pitch accent type (x-axis) on the probability of prominence ratings (color) by pitch accent position (y-axis) for accented words (Table 9).

We move on to the interaction between pitch accent type and acoustic measures. The statistical model finds a limited effect of accent type on the likelihood of prominence rating, and an effect of acoustic correlates of prominence in PC1 and PC3, but the model also finds significant interactions between these variables. In Section 2.3.6.2, we saw results from the production model (Table 5) showing that effects of pitch accent type on acoustic measures of prominence were observed only for F0. Accordingly, to illustrate the interaction between acoustic variables and perceptual prominence ratings, we show the interaction between pitch accent type and PC1, the F0-dominant component. Note, however, that the results from the statistical model are for the four-way interaction between pitch accent type and PCs 1–3. Figure 19 shows the estimated effects of PC1 (x-axis) conditioned on pitch accent type (y-axis), holding other factors constant. If the effect of F0 on the likelihood of prominence rating were similar across accent types, we would expect to find similar model fits (as revealed by the color shade) across the accent types. Instead, we observe differences in the effect of F0 across accent types. The color-coded model predictions show the main effect of PC1 on prominence ratings that we saw above in Figure 13, where an increase in PC1 (higher max F0, bigger F0 range) predicts an increased likelihood of prominence ratings. But the PC1 effect is not uniform across

pitch accent types, with a noticeably lesser effect (lighter shade) for !H*, and the greatest effect (darkest shade) with L+H*. In other words, the combination of increased F0 with the accents that are lowest (L*) and highest (L+H*) on the Prominence hierarchy has the greatest boosting effect on prominence ratings.

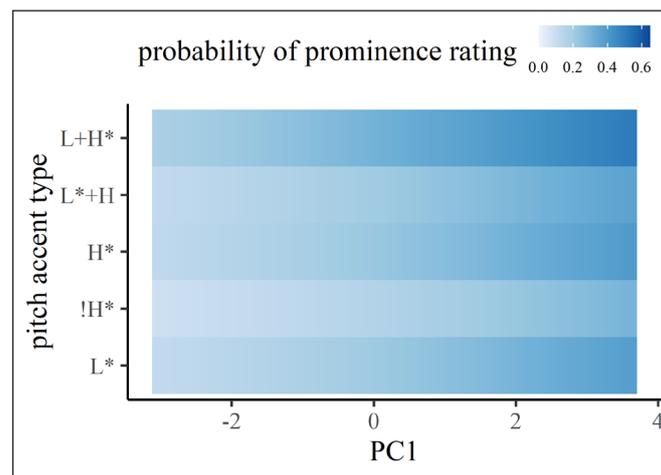


Figure 19: Predicted effects of PC1 (x-axis) on the probability of prominence ratings (color) in relation to pitch accent type (y-axis) for accented words (Table 9).

Next, we consider the interaction of IS (referential and lexical givenness, and contrastive alternatives) and acoustic measures. Recall that the production model (Table 5) shows effects of IS on F0 measures, but not on intensity or phone rate, so we focus here on interaction of IS with PC1 in the effects on prominence rating. Further, we focus on the interaction between PC1 and referential givenness in particular, since an interaction between acoustic measures (PC1:PC2:PC3) and IS was significant only for referential givenness in the statistical model (i.e., for *r-bridging* and *r-unused*, Table 9), while there were no significant interactions with lexical givenness or contrastive alternatives. Figure 20 shows estimated effects of PC1 (x-axis) on prominence ratings (shade) for each level of referential givenness (y-axis). Differences in the estimated effect (color shade) across the *r*-labels indicate that the effects of PC1 (F0 max and range) on prominence ratings vary according to the referential givenness of the word. The effects on F0 are greater than expected for non-referential words (*NR*), and they increase from *r-given* to *r-new*. In other words, enhanced F0 has a greater boosting effect on prominence ratings for non-referential *NR* words than for *r-given* words, holding all other factors constant, and this boosting effect gets successively stronger for words that are higher along the referential Givenness hierarchy.

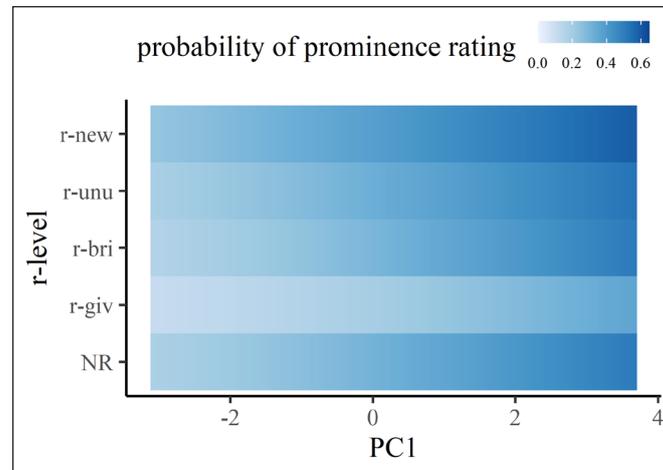


Figure 20: Predicted effects of PC1 (x-axis) on the probability of prominence ratings (color) in relation to referential givenness (y-axis) for accented words (Table 9).

Finally, we present the interaction between accent type and IS (referential, lexical, or alternative information) on prominence rating. Under the strictest interpretation of prominence relations (**Figure 2**), IS categories map one-to-one onto accent types, in which case we expect to find within each dimension of IS (referential, lexical, alternatives) a high estimate for the likelihood of prominence rating for only one accent type within that IS dimension. **Figures 21–23** show that this is not the case. Note that the *none* labels (*NR*, *NL*, *non-alt*) are not under examination here, because they are not expected to be associated with any accent type. For all three dimensions of IS, we observe a general effect of accentual prominence on the probability of prominence rating, with increasing model estimates from L^* to $L+H^*$, left to right in each panel of **Figures 21–23**. For referential givenness (**Figure 21**), the deviations from this trend involve the unexpectedly high prominence boost from H^* (*r-unused*) and inconsistent behavior of L^*+H as boosting (*r-given*, *r-new*) or damping (*r-bridging*, *r-unused*), which correspond to significant interactions in the statistical model output (**Table 9**). The same general trend of increasing estimates of prominence rating across accent types ranked by prominence is observed for lexical givenness in **Figure 22**, with deviations mainly due to the inconsistent behavior of L^* and $L+H^*$, which each show a strong boosting effect on prominence ratings with *l-given* and *l-new*, and a damping effect with *l-accessible*. For words marked for contrastive alternatives (*alt*, **Figure 23**), there is no general trend of increasing model estimates across accent types, as was observed for the referential and lexical dimensions of IS. Instead, the sole effect of accent type in this IS category is related to the unexpected damping effect of $!H^*$ on prominence rating. We return to these interactions in the Discussion section below.

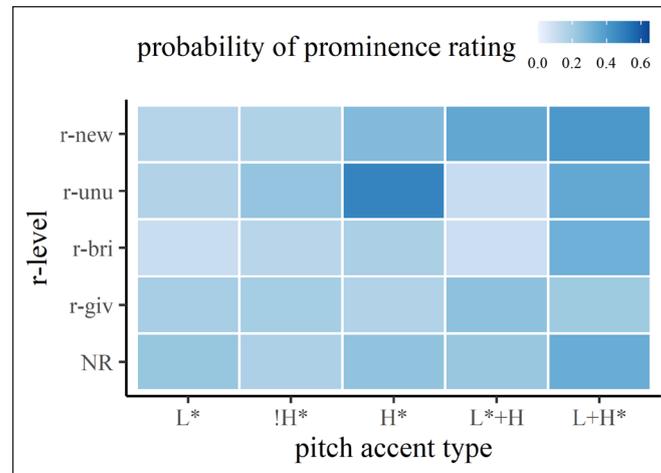


Figure 21: Predicted effects of accent types (x-axis) on the probability of prominence ratings (color) in relation to referential information. The *r*-labels are presented along the Givenness hierarchy in five panels. These results are based on the *Accented-words* model (Table 9).

To broadly summarize the findings presented in this section, we observe that the likelihood of a word being rated as prominent in our experimental task varies as predicted in relation to the word's acoustic prominence. Prominence rating also varies in relation to the prominence of a word on the Givenness hierarchy and its Accentual Prominence hierarchy, for at least some of the tested levels, with effects most often but not always in the predicted direction. In addition, the finding of significant interactions among predictor variables in the prominence rating models suggests that prominence ratings are made in consideration of the combined properties of a word, including its IS category and pitch accent type, pitch accent type and position, and pitch accent type and tonal specification of the immediately following prosodic boundary.

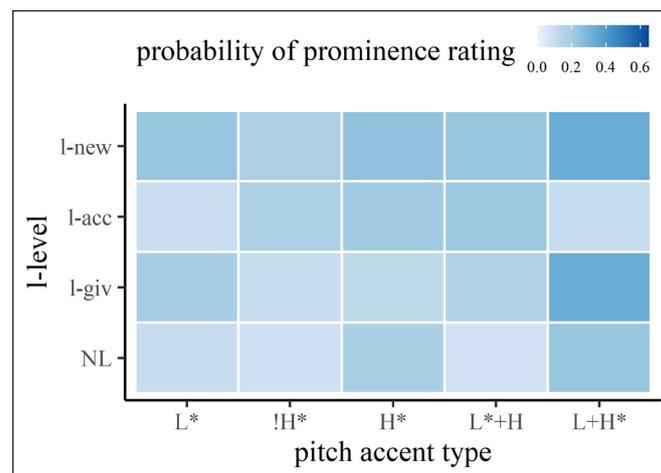


Figure 22: Predicted effects of accent types (x-axis) on the probability of prominence ratings (color) in relation to lexical information. The *l*-labels are presented along the Givenness hierarchy in four panels. These results are based on the *Accented-words* model (Table 9).

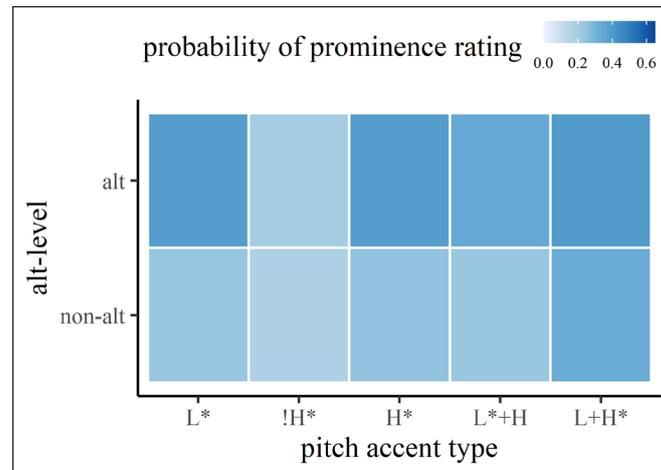


Figure 23: Predicted effects of accent types (x-axis) on the probability of prominence ratings (color) in relation to the status of a word as alternative-eliciting. The *alt*-labels are shown in two panels. These results are based on the *Accented-words* model (Table 9).

4. Discussion

In this study, we assessed the relationships between prominence related to information structure (IS: Referential and lexical givenness, alternative-eliciting) and accentual prominence, and the potential influence of both on the acoustic correlates of prosodic prominence in speech from two TED Talks, as examples of a public, performative speech style. We then explored how linguistically untrained listeners rate the prominence of words in these speech samples, in relation to the IS category, pitch accent status and type, and acoustic prominence of a word. We were particularly interested in how the IS factors, which specify how a word relates to the prior discourse context, might interact with signal-based factors of accentual and acoustic prominence in this task, where listeners were presented with the complete narrative of each TED Talk. Another goal of this work was to examine the relationship among IS, phonological, and acoustic cues in a public speech style, which has not been widely studied in prior work. In this section, we begin with a review of the key findings from the production study, followed by a discussion of key findings from the perception study as they relate to the hypotheses given in Section 1.

4.1. Prominence in production

In the introduction, we reviewed conflicting evidence from previous studies on the one-to-one mapping between pitch accents and IS categories in English and other languages. We set out to test the hypothesis of a direct association between the ranking of pitch accents on the Accentual Prominence hierarchy and the ranking of IS categories on the Givenness hierarchy. The evidence presented above does not support this hypothesis. There was no significant relationship between the pitch accent type and IS category, considering the IS categories of referential and lexical givenness and alternative-eliciting. Nonetheless, numeric differences in

the frequency of accent types by IS category were observed, and though these differences did not reach significance, we note in all cases the count asymmetries went in the predicted direction: Pitch accents ranked higher on the Accentual Prominence hierarchy were more often associated with IS categories ranked higher on the Givenness hierarchy. Furthermore, this tendency was observed with nuclear pitch accents, but not prenuclear pitch accents, which appear to be unrelated to IS, confirming the structural or “rhythmic” role of the prenuclear pitch accent (Calhoun, 2010; Chodroff & Cole, 2018; Vogel et al., 1995). In short, these TED Talk speakers showed a weak preference for using more prominent pitch accents on more informative nuclear accented words.

On a strict reading, our findings do not support even the weaker claim of a probabilistic relationship between pitch accent type and IS category, as put forth by Cangemi and Grice (2016) for Neapolitan Italian, and for German by Baumann and Riester (2013). Recall from Section 2.3.4 (**Figure 4**) that our data show only a statistically marginal association between accent type and levels of referential givenness, and no significant associations between accent type and lexical givenness, though numeric trends are in the predicted direction. What seems clear from our data is that these TED Talk speakers do not effectively employ different types of pitch accent for the purpose of conveying distinctions in meaning related to IS. They make use of the variety of pitch accent types available in English, but the criteria for choosing which pitch accent to assign to a particular word seems to have little to do with the IS category status of that word. Impressionistically, it appears that the speakers exploit variation in pitch accent type for other purposes, possibly to enhance the clarity and liveliness of speech.

Despite the lack of a significant relationship between pitch accent type and IS category, there is evidence that the speakers do not entirely ignore IS in the assignment of pitch accent. Pitch accents are statistically more likely to be assigned to words that specify information structure (words that receive a label at the referential, lexical, or alternative level) compared to words that are not defined for IS. There is further evidence for a relationship between pitch accent assignment and IS in observed differences between referential and lexical givenness in the mapping between IS categories and pitch accent. For instance, pitch accents are more likely to be assigned to referentially *new* words than referentially *given* words, as expected. But a different pattern obtains for words that are lexically *given*, which are equally likely to be accented as are lexically *new* words. Overall, we observe that IS may play a role in the assignment of pitch accent, but only in a coarse manner, in that pitch accents of any type are preferentially assigned to words that convey IS, and words that do not convey IS are comparatively less likely to bear pitch accent.

Looking beyond IS category distinctions, we observe other differences in pitch accent assignment in our data, which, however, we have not systematically analyzed here, such as the association of pitch accent with discourse markers, intensifiers, and certain other lexical categories, independent of IS. In particular, we observe many instances where these TED Talk speakers produce proper nouns, numerals, negations, and discourse markers with greater emphasis than other word classes. Our observations along these lines are reminiscent of findings from previous studies that describe certain part-of-speech categories as strong attractors for accent assignment (Hirschberg, 1993; Sityaev, 2000). Discourse markers have also been noted as being associated with prominence-lending phonological features and phonetic patterns (Calhoun & Schweitzer, 2012).

4.2. Prominence in perception

We hypothesized that listeners perceive a word's prominence in direct relation to its acoustic prominence, its ranking on the Prominence and Givenness hierarchies, and its properties related to prosodic structure. We review the key significant findings from the statistical models of prominence ratings as they relate to each of the six hypotheses introduced in Section 1.

Although this study draws on speech samples from only two speakers, one male and one female, we were interested in knowing whether listeners would rate prominence differently for the two speakers. Towards this end, Speaker was entered as a predictor variable in both statistical models of prominence ratings (*All-words* and *Accented-words*), each pooled over the speaker-specific datasets, but there was no significant effect of Speaker. This null effect indicates that listeners were using similar criteria for assigning prominence ratings for the two speakers, despite differences between them in the intensity measure of their speech (**Table 7**). In other words, listeners seem to calibrate their prominence ratings to the speaker-specific values of acoustic variables in their speech.

Enhancement of acoustic correlates of prominence (PC1: Increased F0 max and range, locally slowed tempo, increased intensity) increases the likelihood that a word will be rated as prominent, as expected. A complementary finding is that lower intensity and faster tempo (PC3), which is a characteristic of function words, appear to inhibit prominence ratings. In contrast, there is no significant impact on prominence ratings from the acoustic pattern that pairs slower tempo with lower intensity (PC2). This finding is expected inasmuch as PC2 captures an acoustic pattern associated with phrase-boundary effects unrelated to prominence. Together, these results (**Figure 13**) show that listeners are sensitive to variation in prominence conveyed through acoustic measures of pitch, tempo, and loudness.

The probability of prominence rating increases in relation to the information status of a word, more or less according to our hypothesis, grounded in the Givenness hierarchy sketched in **Figure 2** for the referential and lexical dimensions of IS (**Figure 14**). Prominence rating is also more likely for words that elicit alternatives compared to those that do not (**Figure 14**). The observed effects of lexical givenness in our data, though conforming to predictions, are surprising considering that pitch accent assignment did not vary across levels of lexical givenness. Lexically given words are just as likely to be accented as lexically new words; nonetheless, listeners are less likely to rate a word as prominent if it is lexically given.

Although the effects of IS on prominence rating are overall in the expected direction, there were some exceptions. First, the *NR* category has an unexpectedly high estimate for prominence rating, which may relate to the fact that the non-referential (*NR*) words consist mainly of function words, among which negations (e.g., *never, not, but*) and discourse markers (e.g., *also, instead, really*) tend to be rated as prominent by most listeners, more often than words that are referentially *given* or *bridging*.

Second, words in the *l-given* category (previously mentioned words) have a surprisingly high likelihood of being rated as prominent, nearly equal to that of *l-new* (previously unmentioned) words. The effect of lexical givenness on prominence rating is not what we expect from the relationship between lexical givenness and pitch accent in the production data, where we observed that *l-given* words are just as likely to be accented as previously unmentioned words, including *l-accessible* and *l-new* (**Table 4**), with no differences in the type of accent assigned to words across the three levels of lexical givenness (**Figure 4**, middle). Thus, considering just the accent status of words across level of lexical givenness in the production data, we expect similarity in prominence rating across all levels of lexical givenness (*l-given, l-accessible, l-new*). Instead, we find prominence ratings are much less likely for *l-accessible* words than for *l-given* and *l-new* words, which have similarly high prominence estimates. The high probability of prominence rating for *l-given* words may be explained by the frequent occurrence in this class of words related to the themes of the two TED Talks, such as the words in the expression “*thirty days*” in Matt Cutt’s TED Talk. We suggest that the prominence rating of these words reflects their topical salience, over-riding the expected effect of lexical givenness. In comparison, the *l-accessible* words, lacking status as topically salient, have a low probability of prominence rating, which may more directly reflect their lower status relative to *l-new* words on the IS hierarchy.

There was limited evidence for significant effects of accent type on the probability of prominence rating, with lower prominence estimates for the two accent types lowest on the Prominence hierarchy (**Figure 2**), L^* and $!H^*$, compared to the three other accents, H^* , $L^* + H$,

and L + H*. We offer two comments here. First, the estimate for prominence with the L* accent is still relatively high, but this is not entirely surprising in light of similar findings reported for AE conversational speech (Cole et al., 2019) and political speech (Bishop et al., 2020). In the present study, the production analysis shows that L* occurs most frequently as the nuclear pitch accent (**Figures 5–7**), in combination with high edge tones, H- or H-H% (**Figures 8–9**), and associated with lexically new information (**Figures 5, 6**). In other words, L* is used in the phrase-final rising tune (L*H- or L*H-H%), a frequent pattern in these TED Talks, possibly serving to signal meaning dependencies with following material. Second, the lack of any further difference on prominence rating among the set {H*, L* + H, L + H*} is best understood in relation to their interaction with accent position, as discussed below, where opposing effects in different positions in the prosodic phrase may cancel out the expected differences between them as a main effect.

Contrary to the hypothesis that nuclear accented words (those in phrase-final position) are more likely to be rated as prominent than prenuclear accented words (in phrase-initial and -medial positions), there was no significant effect of position on prominence rating (**Figure 16**, left), though we note a numeric trend in the expected direction. The lack of a significant effect is surprising considering that the production analysis found a relationship between pitch accent and IS status for nuclear words, but not for prenuclear accented words, similar to findings from other production studies of AE (Calhoun, 2010; Chodroff & Cole, 2018; Vogel et al., 1995). We speculate that the extensive and stylistic use of pitch accents across levels in all IS dimensions in these samples of TED Talk speech might have led listeners to discount the phrase position and its relationship with IS in assessing prominence.

There was no evidence that prominence ratings were rhythmic in their pattern over successive words (**Figure 17**); the likelihood of prominence rating was not lower when the immediately preceding word was rated as prominent (by the same listener). Instead, the opposite pattern was observed, where listeners were more likely to rate a word as prominent if they also rated the preceding word as prominent. This is an unexpected result that suggests prominence perception is ‘clumpy’, and not rhythmic. Rhythmic factors do not seem to play a role in pitch accent assignment in the production data either, as most content words are accented. The clumpiness of perceived prominence may instead reflect clustering at the level of IS, in clusters of words with similar information value, e.g., sequences of previously mentioned words, or sequences of words introducing new referents.

Phrase-final accented words as a group were more likely to be rated as prominent compared to non-final accented words. Furthermore, when accented words are grouped according to the tonal specification of the following boundary, additional significant differences in prominence rating

emerge (**Figure 16**, right). In this more fine-grained analysis, non-final words (those in phrase-initial or -medial position) had the lowest likelihood of prominence rating, with prominence ratings estimates increasing across words based on the following ranking of edge tones: {H-, H-L%} < {L-, L-L%} < {H-H%, L-H%}. This ranking does not align with Hypothesis 6, which predicted differences in prominence rating based only on boundary type: Word < {H-, L-} < {L-L%, L-H%, H-H%, H-L%}. Instead, the observed ranking suggests that, for accented words, it is the direction of the final pitch movement (mid-flat, falling, rising) that matters most for prominence rating, and not the boundary level (phrase accent vs. boundary tone), or the specific tonal specification.

Next, we turn to discuss several notable interaction effects related to our hypotheses. While there was only a limited effect of accent type on prominence rating [H3], distinguishing low-prominence from high-prominence accents, and no significant main effect of accent position [H4], additional distinctions emerge in the interaction of these two factors (**Figure 18**). Two pitch accents in particular exhibit different effects on prominence rating in different phrase positions. In final position, where it is most frequent, H* boosts the likelihood of prominence rating, suggesting that listeners do in fact register the perceptual salience of H* as a nuclear accent, much more than when it is used in prenuclear positions. A similar positional asymmetry is observed with L+H*, which is expected to have the greatest perceptual prominence (see **Figure 2**). A boosting effect of L+H* is indeed observed in phrase-medial and -final positions, but not in phrase-initial position, where L+H* is by far the most frequent accent type used with referentially given words (**Figure 5**). This combination of initial position and referential givenness is dominant among instances of L+H*, which may explain the null effect of this accent on prominence rating, when considered independent of other factors.

The effects of acoustic cues on prominence rating were investigated with variables derived from principal component analysis over four acoustic variables. Two of these acoustic components, related (primarily) to enhanced F0 (PC1) and diminished tempo/intensity (PC3), had significant effects on prominence rating. But looking at the interaction of these acoustic effects with pitch accent type reveals additional distinctions. We illustrated those interactions for one acoustic component (PC1) as a function of accent type in **Figure 19**, where we observed that the effects of enhanced F0 (higher F0 max, expanded F0 range) on prominence rating are greatest for the accents that are lowest (L*) and highest (L+H*) on the Accentual Prominence hierarchy (**Figure 2**). At a more general level, the significant four-way interaction between acoustic components (PCs 1–3) and accent type indicates that the same acoustic properties are weighed differently as cues to prominence depending on accent type. In other words, accent type

appears to mediate the effect of acoustic cues to prominence, as previously reported by Bishop et al. (2020).

The effect of acoustic cues on prominence rating also varies as a function of the IS category of a word. We illustrated this interaction for PC1, the F0-dominant acoustic component (**Figure 20**), where we observed a stronger effect of enhanced F0 on prominence rating for words in the lowest ranked category of referential givenness, *NR* (words defined as non-referential) and words in the highest ranked category, *r-new*. Here also it appears that listeners weigh the contribution of acoustic cues to prominence in relation to the referential givenness status of a word.

We found fairly robust effects of IS on prominence rating for all three dimensions (referential, lexical, alternatives), and a limited effect of accent type. In looking at the interaction of these variables, we were particularly interested in testing the prediction that distinctions in IS level would correspond with distinctions in accentual prominence, as shown in **Figure 2**. There was at best weak evidence for this hypothesis in the production data, as already noted, but there is also limited evidence that the effect of accent type on prominence rating depends on the IS category of the word. For example, based on its position on the Accentual Prominence hierarchy, H* is predicted to have a moderate effect on boosting prominence rating—less than the bitonal accents. Yet, while that pattern is observed for *r-given*, *r-bridging*, and *r-new* words (**Figure 21**), H* has an unexpectedly strong boosting effect on prominence rating for *r-unused* words. This is particularly surprising given that H* is not high on the Accentual Prominence hierarchy, nor is *r-unused* high on the Givenness hierarchy (**Figure 2**). Another unexpected finding is the lack of a boosting effect of the L+H* accent on alternative-eliciting words (**Figure 23**). L+H* is considered the preferred accent for encoding contrastive focus, yet perhaps due to the high frequency of L+H* in these TED Talks, listeners seem to disregard L+H*, or any other accent type, as a cue to prominence for alternatives-eliciting words (recall that the *alt* label is assigned based on syntactic properties). We do not have an explanation for all of the particular interactions of accent type and IS observed in our data, but we expect that understanding prominence at this level of detail will require close examination of a wide range of factors related to the discourse and situational context.

4.3. Summary of key findings

We have discussed the complex relation between phonological pitch accent types, acoustic prosodic features, and information structure in the production (Section 4.1) and perception (Section 4.2) of prominence in a public speech style. Key findings are summarized in **Table 10**, where the term “prominence perception” is used to abbreviate “the likelihood a listener rates a word as prominent” (see Methods description in Section 3.2).

	Production	Perception
Acoustic cues	Acoustic measures of prominence (especially F0) varied as a function of pitch accent and IS, specific combinations of the two, and edge tone, with most but not all of the observed effects in the direction predicted by the Accentual Prominence hierarchy (Tables 5, 7).	<i>Main effect:</i> Enhancement of acoustic correlates of prominence (PC1: Increased F0 max and range, locally slowed tempo, increased intensity) increased prominence perception (Figure 13). <i>Interaction:</i> Acoustic effects on prominence perception varied with pitch accent types (Figure 19) and IS categories (Figure 20).
IS	Pitch accents were statistically more likely to occur with words that do specify information structure than those that do not (Table 3); same for words that are referentially <i>new</i> or <i>accessible</i> compared to referentially <i>given</i> ones (Table 4). Pitch accents were equally likely to occur with lexically <i>given</i> words versus lexically <i>new</i> or <i>accessible</i> words (Table 4).	<i>Main effect:</i> The probability of prominence increased in relation to the information status of a word, more or less according to the Givenness hierarchy for referential and lexical givenness. Prominence perception is also more likely for words that elicit alternatives compared to those that do not (Figure 14). <i>Interaction:</i> Depending on IS categories, acoustic effects on prominence perception varied (Figure 20). Limited evidence was found that depending on IS categories, the effect of accent type on prominence perception systematically varies (Figures 21–23).
Pitch accent	Limited evidence of a deterministic relation between pitch accent type and IS category was found, though there were non-significant numeric differences in the frequency of accent types by IS category in the predicted direction of the Accentual Prominence and Givenness hierarchies (Figure 4).	<i>Main effect:</i> Limited evidence was found for significant effects of accent type on prominence perception, distinguishing low(er)-prominence from high(er)-prominence accents: $\{L^*, !H^*\} < \{H^*, L^* + H, L + H^*\}$ (Figure 15). <i>Interaction:</i> Depending on pitch accent type, acoustic effects on prominence perception varied (Figure 19). Limited evidence was found that the effect of accent type on prominence perception depends on the IS category of the word (Figures 21–23).
Accent position	A non-significant numerical relation between accentual and informational prominence was found for accented words in nuclear position but not in prenuclear position (Figures 5, 6).	<i>Main effect:</i> No significant effect of accent position on prominence perception was found, though we note a numeric trend in the expected direction: prenuclear accent $<$ nuclear accent (Figure 16, left). <i>Interaction:</i> Effects of some accent types differed across accent positions (Figure 18).
Rhythm	(<i>not tested in production</i>)	<i>Main effect:</i> The probability a current word is perceived as prominent increased if the preceding word was perceived as prominent by the same listener (Figure 17).

(Contd.)

	Production	Perception
Edge tone	A wide variety of mappings between pitch accents and edge tones was found in all information structural conditions (Figure 8–10).	Main effect: The likelihood of prominence perception increased across words based on the following ranking of edge tones: non-final words (phrase-initial or -medial) < {H-, H-L%} < {L-, L-L%} < {H-H%, L-H%} (Figure 16, right).

Table 10: Summary of key findings.

4.4. Comparison with prominence perception in conversational speech—the effect of speech style

Our prominence rating data for TED Talk speech can be compared with prominence rating data for other types of speech data, specifically, from conversational speech (Cole et al., 2019; Cole et al., 2010; Hualde et al., 2016; Im, Cole, & Baumann, 2018; Turnbull et al., 2017) and political speech (Bishop et al., 2020). First, in our data, F0 measures play a significant role in prominence rating, which aligns with findings from Bishop et al. (2020). This is counter to findings from conversational speech where durational measures (stressed vowel duration or word phone rate) are found to play a bigger role in predicting perceived prominence (Cole et al., 2019; Cole et al., 2010). We find that acoustic correlates of prominence in this public speech are different from what has been reported from studies investigating laboratory and conversational speech. In the present TED Talk speech, F0 is the most robust acoustic correlate of pitch accent type and IS category, while phone rate (as a durational cue) and intensity are not strong correlates of any accent distinctions. The TED Talk speakers speak slowly and loudly throughout their narrative and apparently have little opportunity to further modulate tempo or intensity to convey distinctions in accentual prominence or givenness. A practical implication from this finding is that a successful model of prominence in speech production must consider acoustic cues in relation to the phonological accent type, and in relation to contextual factors, including IS and speech style.

Second, we find that the influence of accent type on prominence rating differs depending on the position of the accented word in the prosodic phrase, which also aligns with findings from Bishop et al. (2020). In failing to make a strong association between distinctions in accent type and distinctions in IS category, while at the same time using a high rate of accenting with accents of different types, these TED Talk speakers produce a pattern of phrasal intonation characterized by frequent pitch modulation that do not convey specific IS distinctions. To our ears, this produces a lively speech style that engages our attention, and which sounds very different from a more relaxed style of conversational speech. Notably, the TED Talk speakers favor L + H*. The

greater frequency of L+H* in the TED Talk speech aligns with the finding from Chodroff and Cole (2018, 2019) showing that AE speakers produce far more L+H* accents when they are asked to read aloud stories in a lively speaking style compared to a neutral, relaxed style. Similar findings across these two studies suggest that the TED Talk speakers may be producing frequent L+H* pitch accents to express engagement and liveliness, in an effort to attract and maintain listeners' attention. The distribution of accents in phrase-initial position is of special interest. It shows that despite the greater frequency of initial L+H* in the TED Talk samples (Figures 5, 6), phrase-initial words with the L+H* accent have a relatively low probability of being rated as prominent (Figure 18; see also Im et al., 2018, comparing prominence ratings of one TED Talk with the conversational speech of the Buckeye corpus). In other words, listeners are apparently not responding to the prevalence of words with greater accentual and acoustic prominence in the TED Talk samples by rating more words as prominent (which would boost the likelihood of prominence ratings). Instead, our findings suggest that listeners calibrate their prominence ratings to the speech style of the speakers, effectively discounting the more frequent L+H* accents and higher max F0 of the TED Talks. This calibration is only possible if listeners are able to take contextual information into account when rating prominence, considering factors like speech style as it relates to the situation (e.g., a TED Talk) or to the habits of the individual speaker.

Lastly, lexical repetition (captured by *l*-labels in the present study) has a significant effect on the prominence ratings in these TED Talks, consistent with prominence ratings for conversational speech from Cole et al. (2010). The repetition of a lexical expression, but not a referential expression, was correlated with a higher likelihood of prominence rating. In addition to lexical givenness, our data show that alternatives-eliciting words have a higher likelihood of being rated as prominent, aligning with conversational speech (Turnbull et al., 2017).

4.5. Limitations of the present study

This paper reports on two case studies, both explored with respect to production and perception of prominence in performative speech. The data analyzed here reflect the particular patterns of association between pitch accents, information structure (givenness and contrastive focus), and acoustic correlates of prominence produced by two speakers in their respective TED Talks. While the statistical analysis found only one significant difference between these two speakers, in intensity as an acoustic correlate of prominence, it is possible that other speakers may present different patterns of association over the parameters in question. Such differences may arise due to dialectal differences, in relation to the topics under discussion, or even the communication setting. We think these are fruitful avenues for future research on prominence in AE speech.

5. Conclusion

In the current study, we investigated prominence in a public speech style of AE, using a case study approach with speech data from two speakers. Speech materials, as production data, were analyzed for relationships between phonological pitch accents and acoustic prosodic cues as signal-based correlates of prominence, and between those signal-based factors and information structure distinctions related to prominence in three meaning domains: referential, lexical and alternatives-eliciting. While we find no systematic relationship between pitch accent type and information structure categories in these samples of public speech, the speakers do make distinctions in the phonetic implementation of pitch accent in relation to the IS category of the word, especially in pitch. Moreover, effects of information structure are observed in the likelihood that a word is associated with pitch accent (of any type), and those effects are different for lexical versus referential information status. Despite the weak relationship between accent types and IS in production, listeners perceive and rate prominence in a manner that is, to varying degrees, in accordance with the Accentual Prominence and Givenness hierarchies. Prominence rating is more likely: for words that are more informative (new or alternatives-eliciting) compared to words that express given information; for words with high-toned pitch accents (H^* , $L+H^*$ and $L+H^*$) compared to low-toned and downstepped pitch accents (L^* and $!H^*$); and for acoustically enhanced words compared with words with less acoustic enhancement. Moreover, listeners appear to differentially weigh acoustic cues and pitch accent status as cues for prominence depending on the type of accent, on the position of accent, on the type of the following edge tone, and on the IS category of the accented word. Altogether, the findings from this study contribute new evidence of the mediating effects of phonological context (pitch accent status and type) as well as discourse- and situation-related contextual factors (IS, speech style) on the interpretation of acoustic cues to prominence.

Supplementary materials

The complete output from statistical models discussed here can be found in the online repository, <https://osf.io/hakpw/>.

Acknowledgements

This study was supported by NSF BCS 12-51343 to Jennifer Cole.

Competing Interests

The authors have no competing interests to declare.

References

- Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. Hualde (Eds.), *Papers in laboratory phonology 9* (pp. 547–576). Berlin & New York: Mouton de Gruyter.
- 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. DOI: <https://doi.org/10.1177/00238309040470010201>
- Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. *The Journal of the Acoustical Society of America, 95*, 2973. DOI: <https://doi.org/10.1121/1.408967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Baumann, S., Mertens, J., & Kalbertodt, J. (2019). Informativeness and speaking style affect the realization of nuclear and prenuclear accents in German. *Proceedings of the 19th International Congress on Phonetic Sciences* (pp. 1580–1584).
- 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. DOI: <https://doi.org/10.1515/9783110261790.119>
- Baumann, S., & Riester, A. (2013). Coreference, lexical givenness and prosody in German. *Lingua, 136*, 16–37. DOI: <https://doi.org/10.1016/j.lingua.2013.07.012>
- Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics, 70*, 20–38. DOI: <https://doi.org/10.1016/j.wocn.2018.05.004>
- Beckman, M., & Ayers Elam, G. (1997). *Guidelines for ToBI labeling (Version 3)*. The Ohio State University. Unpublished ms.
- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255–309. DOI: <https://doi.org/10.1017/S095267570000066X>

- Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Dordrecht, Netherlands: Fortis. DOI: <https://doi.org/10.1515/9783110874020>
- Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, 82, 100977. DOI: <https://doi.org/10.1016/j.wocn.2020.100977>
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program]. Retrieved from <http://www.praat.org/>
- Bolinger, D. (1958). A theory of pitch accent in English. *Word*, 14(2–3), 109–149. DOI: <https://doi.org/10.1080/00437956.1958.11659660>
- Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25(7–9), 1044–1098. DOI: <https://doi.org/10.1080/01690965.2010.504378>
- Breheny, P., & Burchett, W. (2017). Visualization of regression models using visreg. *The R Journal*, 9, 56–71. DOI: <https://doi.org/10.32614/RJ-2017-046>
- Büring, D. (2007). Intonation, semantics, and information structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (pp. 445–473). Oxford, UK: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199247455.013.0015>
- Calhoun, S. (2006). *Information structure and the prosodic structure of English: A probabilistic relationship* (Doctoral dissertation, University of Edinburgh, Edinburgh, UK).
- Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86(1), 1–42. DOI: <https://doi.org/10.1353/lan.0.0197>
- Calhoun, S., Nissim, M., Steedman, M., & Brenier, J. (2005). A framework for annotating information structure in discourse. *Proceedings of Frontiers in corpus annotation II: Pie in the sky, ACL2005 conference workshop*. DOI: <https://doi.org/10.3115/1608829.1608836>
- Calhoun, S., & Schweitzer, A. (2012). Can intonation contours be lexicalised? Implications for discourse meanings. *Prosody and Meaning*, 271–328. DOI: <https://doi.org/10.1515/9783110261790>
- Cangemi, F., & Baumann, S. (2020). Integrating phonetics and phonology in the study of linguistic prominence. *Journal of Phonetics*, 81, 100993. DOI: <https://doi.org/10.1016/j.wocn.2020.100993>
- Cangemi, F., & Grice, M. (2016). The importance of a distributional approach to categoriality in Autosegmental-Metrical accounts of intonation. *Laboratory Phonology*, 7(1), 9. DOI: <https://doi.org/10.5334/labphon.28>
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (Ed.), *Subject and topic* (pp. 25–55). New York, NY: Academic Press.
- Chafe, W. (1994). *Discourse, Consciousness, and Time*. Chicago/London: University of Chicago Press.
- Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, 39(3), 344–361. DOI: <https://doi.org/10.1016/j.wocn.2011.02.005>

- Chodroff, E., & Cole, J. (2018). Information structure, affect, and prenuclear prominence in American English. *Proceedings of Interspeech 2018* (pp. 1848–1852). DOI: <https://doi.org/10.21437/Interspeech.2018-1529>
- Chodroff, E., & Cole, J. (2019). The phonological and phonetic encoding of information structure in American English nuclear accents. *Proceedings of the 19th International Congress on Phonetic Sciences* (pp. 1570–1574).
- Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & de Souza, R. N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, 75, 113–147. DOI: <https://doi.org/10.1016/j.wocn.2019.05.002>
- 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. DOI: <https://doi.org/10.1016/j.wocn.2006.03.004>
- 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. DOI: <https://doi.org/10.1515/labphon.2010.022>
- Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology*, 7(1), 8. DOI: <https://doi.org/10.5334/labphon.29>
- Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. (Doctoral dissertation, The University of Chicago, Chicago, IL).
- de Ruiter, L. E. (2015). Information status marking in spontaneous vs. read speech in storytelling tasks—Evidence from intonation analysis using GToBI. *Journal of Phonetics*, 48, 29–44. DOI: <https://doi.org/10.1016/j.wocn.2014.10.008>
- Dilley, L. C., & Heffner, C. C. (2013). The role of f0 alignment in distinguishing intonation categories: evidence from American English. *Journal of Speech Sciences*, 3(1), 3–67. DOI: <https://doi.org/10.20396/joss.v3i1.15039>
- Dipper, S., Götze, M., & Skopeteas, S. (2007). Information structure in cross-linguistic corpora: Annotation guidelines for phonology, morphology, syntax. In *Interdisciplinary Studies on Information Structure: Working Papers of the SFB632*. Potsdam, Germany: University of Potsdam.
- Fisher, R. A. (1934). *Statistical Methods for Research Workers* (5th Ed.). Edinburgh, UK: Oliver & Boyd.
- Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307. DOI: <https://doi.org/10.2307/416535>
- Hirschberg, J. (1993). Pitch accent in context predicting intonational prominence from text. *Artificial Intelligence*, 63(1–2), 305–340. DOI: [https://doi.org/10.1016/0004-3702\(93\)90020-C](https://doi.org/10.1016/0004-3702(93)90020-C)
- Hualde, J., 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 J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of speech prosody 2016* (pp. 459–463). DOI: <https://doi.org/10.21437/SpeechProsody.2016-94>

- Im, S., Cole, J., & Baumann, S. (2018). Probabilistic relationship between pitch accents and information status in public speech. *Proceedings of Speech Prosody 9* (pp. 508–511). DOI: <https://doi.org/10.21437/SpeechProsody.2018-103>
- Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*, 1038. DOI: <https://doi.org/10.1121/1.1923349>
- Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington, IN: Indiana University Press.
- Ladd, D. R. (2008). *Intonational Phonology* (2nd Ed.; 1st Ed. 1996). Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511808814>
- Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*(3), 313–342. DOI: <https://doi.org/10.1006/jpho.1997.0046>
- Ladd, D. R., & Schepman, A. (2003). “Sagging transitions” between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*, 81–112. DOI: [https://doi.org/10.1016/S0095-4470\(02\)00073-6](https://doi.org/10.1016/S0095-4470(02)00073-6)
- Ladd, D. R., Verhoeven, J., & Jacobs, K. (1994). Influence of adjacent pitch accents on each other’s perceived prominence: Two contradictory effects. *Journal of Phonetics*, *22*(1), 87–99. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30268-2](https://doi.org/10.1016/S0095-4470(19)30268-2)
- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511620607>
- Lieberman, M., & Sag, I. (1974). Prosodic form and discourse function. *Papers from the 10th Regional Meeting of the Chicago Linguistics Society* (pp. 416–427).
- Lieberman, M. Y. (1975). *The intonational system of English* (Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA).
- Lieberman, M. Y., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157–234). Cambridge: MIT Press.
- Luchkina, T., & Cole, J. (2017). Structural and referent-based effects on prosodic expression in Russian. *Phonetica*, *73*, 279–313. DOI: <https://doi.org/10.1159/000449104>
- Mahrt, T. (2013). Language markup and experimental design software [Computer software]. Retrieved from <http://www.timmahrt.com/lmeds.html>
- Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, *44*, 47–61. DOI: <https://doi.org/10.1016/j.wocn.2014.02.003>
- O’Connor, J. D., & Arnold, G. F. (1961). *Intonation of Colloquial English*. London, UK: Longmans.

- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157–175. DOI: <https://doi.org/10.1080/14786440009463897>
- Pfitzinger, H. R. (1998). Local speech rate as a combination of syllable and phone rate. *Proceedings of the 5th International Conference on Spoken Language Processing* (pp. 1087–1090). DOI: <https://doi.org/10.21437/ICSLP.1998-545>
- Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA).
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- 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. DOI: <https://doi.org/10.1016/j.specom.2004.09.001>
- Prince, Ellen F. (1981). Towards a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–256). New York, NY: Academic Press.
- R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Reed, P. E. (2017). The influence of regional identity on Appalachian intonation. *The Journal of the Acoustical Society of America*, 142, 2678. DOI: <https://doi.org/10.1121/1.5014763>
- Riester, A., & Baumann, S. (2013). Focus triggers and focus types from a corpus perspective. *Dialogue and Discourse*, 4(2), 215–248. DOI: <https://doi.org/10.5087/dad.2013.210>
- Riester, A., & Baumann, S. (2017). The RefLex scheme – annotation guidelines. In *SinSpeC: Working Papers of the SFB 732 “Incremental Specification in Context.”* Stuttgart, Germany: University of Stuttgart. DOI: <https://doi.org/10.18419/opus-9011>
- Roessig, S., Mücke, D., & Grice, M. (2019). The dynamics of intonation: Categorical and continuous variation in an attractor-based model. *PLoS ONE*, 14(5), e0216859. DOI: <https://doi.org/10.1371/journal.pone.0216859>
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1), 75–116. DOI: <https://doi.org/10.1007/BF02342617>
- Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology*, 8(1), 22. DOI: <https://doi.org/10.5334/labphon.108>
- Sag, I., & Liberman, M. (1975). The intonational disambiguation of indirect speech acts. *Papers from the 11th Regional Meeting of the Chicago Linguistics Society* (pp. 487–497).
- Schafer, A. J., Camp, A., Rohde, H., & Grüter, T. (2019). Contrastive prosody and the subsequent mention of alternatives during discourse processing. In K. Carlson, C. Clifton Jr., & J. Fodor, (Eds.), *Grammatical approaches to language processing* (pp. 29–44). New York, NY: Springer. DOI: https://doi.org/10.1007/978-3-030-01563-3_3

- Shattuck-Hufnagel, S. (1995). The importance of phonological transcription in empirical approaches to “stress shift” versus “early accent”: Comments on Grabe and Warren, and Vogel, Bunnell, and Hoskins. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV* (pp. 128–140). Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511554315.010>
- Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*, 22(4), 357–388. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30291-8](https://doi.org/10.1016/S0095-4470(19)30291-8)
- Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., ... & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *The 2nd International Conference on Spoken Language Processing* (Vol. 2, pp. 867–870). DOI: <https://doi.org/10.21437/ICSLP.1992-260>
- Sityaev, D. (2000). The relationship between accentuation and information status of discourse referents: A corpus-based study. *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. *The Journal of the Acoustical Society of America*, 100(4), 2471–2485. DOI: <https://doi.org/10.1121/1.417955>
- Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, 118(3), 1677–1688. DOI: <https://doi.org/10.1121/1.2000788>
- Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2), 171–206. DOI: <https://doi.org/10.1006/jpho.1999.0093>
- 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. DOI: <https://doi.org/10.1080/23273798.2017.1279341>
- 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>
- Vogel, I., Bunnell, T., & Hoskins, S. (1995). The phonology and phonetics of the rhythm rule. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV* (pp. 111–127). Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511554315.009>
- Ward, G. & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 61, 747–776. DOI: <https://doi.org/10.2307/414489>
- Watson, D. G. (2010) The many roads to prominence: Understanding emphasis in conversation. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 163–183). Amsterdam, Netherlands: Elsevier. DOI: [https://doi.org/10.1016/S0079-7421\(10\)52004-8](https://doi.org/10.1016/S0079-7421(10)52004-8)
- 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. DOI: <https://doi.org/10.1016/j.cognition.2007.06.009>
- Weide, R. (2005). The Carnegie Mellon Pronouncing Dictionary [cmudict. .6]. Carnegie Mellon University. Retrieved from <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

Xu, Y. (2013). ProsodyPro – A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (pp. 7–10).

Yoon, T., & Cole, J. (2006). Downstepped pitch accent in American English is categorical and predictable. *The 10th Conference on Laboratory Phonology*.

