Prominence perception is dependent on phonology, semantics, and awareness of discourse

Rory Turnbull , Adam J. Royer, Kiwako Ito & Shari R. Speer

Routledge
Taylor & Francis Group

Check for updates

REGULAR ARTICLE

Prominence perception is dependent on phonology, semantics, and awareness of discourse

Rory Turnbull [a,b], Adam J. Royer[a,c], Kiwako Ito[a] and Shari R. Speer [a]

aDepartment of Linguistics, Ohio State University, Columbus, OH, USA; bLaboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Ecole Normale Supérieure, PSL Research University, Paris, France; cDepartment of Linguistics, University of California Los Angeles, Los Angeles, CA, USA

ABSTRACT
The perception of prosodic prominence is thought to be influenced by multiple competing factors. Three experiments tested the effects of phonological salience, discourse context and listener's knowledge about the discourse on prosodic prominence judgements, using short adjective–noun phrases extracted from a corpus of spontaneous speech. These phrases had either a prominent L + H* 0 or a less prominent H* !H* pitch accent contour. The phrases were presented in a discourse context which either supported or did not support a contrastive interpretation of the adjective. Effects of the contrastive context to increase the perception of prominence only emerged for the phrases with the phonologically prominent L + H* 0 pitch accent sequence. Additionally, the magnitude of the contrast effect was correlated with the listener's awareness of the discourse context, suggesting an ample interplay between linguistic context, pragmatic context, and phonology in prominence perception.

Introduction

It is a truism that, for a given utterance, listeners perceive some words as more prosodically prominent than other words. This variability is reflected in the acoustics of utterance prosody, where fluctuations in fundamental frequency (f0), duration, and intensity – among other factors – are used to mark phonological events, such as phrasing, and to signal information-structural components, such as focus. While the acoustics of prominence production are somewhat well-studied (see Beckman & Venditti, 2010, for review), the perception of prosodic prominence is not well-understood. In particular, the extent to which acoustic cues interact with non-acoustic contextual cues in prominence perception is unclear. The present study aims to address this gap in understanding, with the ultimate agenda of advancing our knowledge of the speech perception–interpretation interface.

This lack of clarity is due, in part, to terminology: "prominence" can mean many things to many people (see Wagner et al., 2015, for review). For the present study, we restrict our definition of prominence to include only prosodic prominence – other kinds of prominence are beyond the scope of this paper – and we define prominence as a purely perceptual phenomenon. Prominence thus defined is *relational* – a word or phrase cannot be prominent a priori, but must be perceived as prominent *relative to* another word or phrase. This definition also entails that prominence cannot be directly observed in the acoustic signal, although acoustics remain useful as a means of indirect inference. In this way, prominence is a quale, similar to other qualia such as pitch or loudness. A purely acoustical or phonological theory of prominence perception would predict that prominence judgements can be inferred from the acoustics or phonology[1] of an utterance alone. An alternative view is that factors at other levels, such as pragmatic context, which includes both discourse context itself and the listener's knowledge about the discourse background, have a role to play in prominence perception.

In this study, we used a prominence-rating task to examine the extent to which prominence perception is affected by phonological factors and by contextual factors. The stimuli were short phrases with different pitch accent contours, thus varying their phonological features. Three contextual factors were also manipulated. The first factor, a manipulation of local context, was whether the sequence of the paired phrases in each trial did or did not repeat the noun: a repetition of the head noun (e.g. blue drum, brown drum) may lead to a contrastive reading of the second phrase. The second factor was the insertion of a brief utterance spoken by another voice (such as "Okay, next?") between the two

phrases. The third factor manipulated the instruction provided to the participants prior to the task, specifically regarding background information (or lack thereof) that may affect their interpretation of the illocutionary function of utterances. The insertion of another voice and the instruction on discourse background turned the sequential noun phrases (e.g. blue drum, brown drum) into meaningful turns in a discourse (e.g. "Blue drum". "Okay, next?" "Brown drum"). To preface our results, we find evidence that prominence perception is influenced by phonological content (pitch accent combination), discourse context (contrastive or non-contrastive readings), and the listeners' knowledge about the discourse background (whether they interpret communicative intent behind the phrases): words that are phonologically salient are perceived as being more prominent than words that are phonologically less salient regardless of the context; words that are in a contrastive word sequence are perceived as being more prominent than those in a non-contrastive sequence, independently of pitch accent; additionally, the magnitude of this effect is enhanced when participants believe that they are listening to a real dialogue rather than a monologue.

The contribution of pitch accenting to prominence perception

Within the autosegmental-metrical framework of American English intonational phonology, local tonal prominences in speech are due to phonological entities called *pitch accents* (Ladd, 2008; Pierrehumbert, 1980). The widely used Tones and Break Indices (ToBI; Beckman & Ayers, 1997) prosodic annotation system for American English distinguishes several pitch accents. In the present study, we compared two pitch accent sequences on two-word adjective–noun phrases: [H* !H*] and [L + H* 0] (where "0" represents an unaccented word). The sequence [H* !H*] has been described as a "citation form" (Ladd, 2008: 98) and is commonly regarded as a relatively neutral prosodic contour, serving simply to introduce information to the discourse without complex inferences. The exclamation mark on the !H* accent denotes "downstep", whereby this accent is realised with a slightly lower f0 peak than the preceding H*. A lower f0 peak does not necessarily lead to a perception of diminished prominence (Gussenhoven, Repp, Rietveld, Rump, & Terken, 1997; Pierrehumbert, 1979; Terken, 1991, 1994), although the second pitch accent of the sequence [H* !H*] is theoretically less prominent than the second pitch accent of the sequence [H* H*].

Before continuing, it should be noted that the distinction between L + H* and H* has been a source of continued debate in the literature. Principally, the dispute centres around whether the distinction is one of

phonological type (e.g. Arvaniti & Garding, 2007), or a distinction of phonetic prominence alone (e.g. Calhoun, 2012; Ladd & Schepman, 2003). These issues are largely orthogonal to those of the present study. It is clear that, regardless of whether the distinction is truly phonological, a [L + H* 0] sequence should give perceptual prominence to the adjective, relative to a [H* !H*] sequence.

Pierrehumbert and Hirschberg (1990, 296) described the L + H* accent as being used "to convey that the accented item – and not some alternative related item – should be mutually believed", and noted that it commonly denotes "a correction or contrast". This wording of contrast is incidentally very similar to Rooth's (1992) conception of focus-semantic value in his framework of alternative semantics. Perhaps in part due to this semantic prominence which is often associated with L + H*, or due to particular phonetic features of this accent (discussed below), this accent is commonly regarded as being highly prominent.

There is substantial experimental evidence that the L + H* accent, in opposition to the relatively neutral H* accent, evokes a contrastive interpretation (Bock & Mazzella, 1983; Ito & Speer, 2008, 2011; Ito, Bibyk, Wagner, & Speer, 2014; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Weber, Braun, & Crocker, 2006; cf. Ladd & Schepman, 2003). In an eyetracking study, Ito and Speer (2008) demonstrated that upon hearing a L + H* pitch accented adjective, listeners anticipated upcoming material consistent with a contrastive reading. That is, in a sequence like "hang the blue egg, now hang the GREEN [L + H*] … ", listeners would look toward the set of eggs on the visual display before hearing the actual second noun. Ito and Speer (2008) interpreted these results as the L + H* evoking a contrast between the accented entity (in this case, the colour *green*) and the most pertinent alternative (in this case, the previously mentioned colour *blue*). This contrast-evoking effect of L + H* is concordant with Pierrehumbert and Hirschberg's (1990, 296) characterisation of L + H* as denoting a "correction or contrast". Ito and Speer's (2008) stimuli consisted of tightly controlled recordings made by a trained phonetician, so-called lab speech. A reasonable question is whether these stimuli were "prosodically hyperarticulated" and somehow unnatural – in other words, whether a naturally produced L + H* is prominent in the same way as a L + H* derived from lab speech. Pitch accents are phonological entities, and as such they have variable acoustic phonetic correlates. Generally, L + H* has been asserted to have a higher and later pitch peak than H* (Arvaniti & Garding, 2007; Burdin et al., 2015; Turnbull, Burdin, Clopper, & Tonhauser, 2015), while unaccented words (labelled with 0) have been reported to have reduced vowels (Bolinger, 1981; Lindblom, 1963) and lower pitch peaks than

accented words, such as those with !H* (Burdin & Clopper, 2015). The extent to which these patterns hold for spontaneous speech is currently unknown, especially in the realm of perception.

The present study used spontaneous speech, derived from a naturalistic task, as stimuli in a prominence-rating task. The use of natural spontaneous speech allowed us to sidestep potential issues of "prosodic hyperarticulation" which typically occurs with lab speech, and allowed us to examine natural variation in the pronunciation of pitch accents such as H* and L + H*.

The contribution of linguistic context to prominence perception

Bock and Mazzella (1983) reported the results of two comprehension experiments where target sentences were preceded by context sentences which varied in prosodic appropriateness. For example, the sentence *DORIS fixed the radio*, where capitals denote a L + H* accent, was comprehended faster when it was preceded by the contextually appropriate sentence *ARNOLD didn't fix the radio* than when the preceding sentence had a contextually inappropriate accent (L + H* on *fix*) or when there was no strong accent in the sentence at all. The contextually appropriate context sentence evoked anticipation for a contrastive entity in the target sentence, leading to the facilitation in comprehension. This early investigation into the role of linguistic context suggests that prosodic perception is not completely signal-driven.

More recently, Cole, Mo, and Hasegawa-Johnson (2010) and Bishop (2012) have examined the balance between signal and contextual factors in the perception of prosodic prominence. Cole, Mo, and Hasegawa-Johnson (2010) presented naïve listeners with short monologues excerpted from the Buckeye Speech Corpus (Pitt et al., 2007), and asked the listeners to mark, on a transcript, which words they perceived as prominent. Their results showed considerable influence of the acoustic signal on prominence markings, but also that expectation-based factors, such as word frequency, influenced prominence independently of the signal. While the experimenters did not have direct control over the linguistic context of the stimuli, their findings suggest that participants were sensitive to differences in acoustic prominence that occur in spontaneous speech.

Bishop's (2012) study involved a prominence-rating task where listeners were presented recordings of short dialogues. In contrast with Cole, Mo, and Hasegawa-Johnson (2010), these recordings were lab speech, not spontaneous speech. However, this design allowed the preceding linguistic context to be tightly controlled.

Each dialogue consisted of a question–answer pair. The answer was a sentence such as "I bought a motorcycle", while the question entailed a contrast on the whole sentence ("What happened yesterday?"), the verb phrase ("What did you do yesterday"), or the object ("What did you buy yesterday"). Bishop (2012) found that the word "motorcycle" was perceived as more prominent in the object contrast condition than in the others, suggesting that the linguistic context (in this case, the information structure that leads to contrast) influenced listeners' perceptions. In all three conditions, the same recording of the answer was presented, thus ruling out any potential acoustic or phonological influence on prominence perception. These results strongly suggest that discourse context plays an important role in prominence perception.

The use of simple dialogues in Bishop's (2012) study demonstrated that preceding context utterances alone can lead to differences in the perception of prominence in the target utterance. However, it is not clear how much pragmatic context is required to establish focus interpretation. Do listeners anticipate particular discourse entities to be under focus even in a considerably impoverished discourse context, such as one which simply presents phrases in sequence with no semblance of a dialogue? Additionally, Bishop (2012) did not compare the effects of context to effects of phonological prominence, prompting the related line of inquiry into how listeners balance the acoustic and non-acoustic cues when making their judgements: whether these effects are additive, if one can "override" the other, and what listeners do in cases where the cues conflict. The present study used a combination of acoustic and linguistic contextual cues in a prominence-rating task. Over three experiments, we varied the discourse context to investigate the extent to which listeners require pragmatic support for contrastive interpretations of phrases, and how these interpretations influence the perception of prosodic prominence.

Measuring prominence perception

The present study adopted a modified version of the Rapid Prosodic Transcription (RPT) paradigm (Cole, Mo, & Baek, 2010, Cole, Mo, and Hasegawa-Johnson, 2010) for obtaining prominence judgements from listeners (for related paradigms, see also Baumann & Röhr, 2015; Buhman et al., 2002; Kohler, 2008; Portele, 1998; Streefkerk, Pols, & ten Bosch, 1997; Swerts, 1997). In this task, naïve, untrained listeners hear a stimulus phrase and are given minimal instructions to highlight the words that they perceive to be prominent on a written transcript of the speech. This method has been used to

acquire prominence ratings for samples of spontaneous speech in American English (Cole, Mo, & Baek, 2010, Cole, Mo, & Hasegawa-Johnson, 2010; Hualde et al., 2016; Kimball & Cole, 2014), French (Hualde et al., 2016; Roux, Bertrand, Ghio, & Astésano, 2016; Smith, 2013), Hindi (Jyoti, Cole, Hasegawa-Johnson, & Puri, 2014; Luchkina, Puri, Jyoti, & Cole, 2015), Korean (You, 2012), Spanish (Hualde et al., 2016), and Russian (Luchkina et al., 2015; Luchkina & Cole, 2014), and samples of read speech in German (Baumann, 2014).

The majority of research using the RPT paradigm, cited above, has used speech extracts with no explicit manipulation of linguistic or pragmatic context. However, with some small modifications, the task can investigate the effects of context at multiple levels with short phrases. The RPT paradigm is a quick and effective method for obtaining prominence judgements from naïve listeners, who may adopt different marking strategies when asked to attend to different cues (Cole, Mahrt, & Hualde, 2014; Smith, 2013).

Hypotheses and predictions

In this study, we propose three hypotheses about the relationship between acoustics, context, and prominence: (1) the phonology-primary hypothesis, (2) the context-primary hypothesis, and (3) the balanced-cue hypothesis. In the first, phonological (or acoustic) salience is primary and is a prerequisite for prominence. Utterances without phonological salience, regardless of context, cannot be perceived as prominent. For salient utterances, however, both discourse context and listener's knowledge about the discourse can provide secondary cues and thus enhance or diminish the perception of prominence accordingly, as demonstrated by Bock and Mazzella (1983) and Bishop (2012). This hypothesis therefore predicts main effects of phonological salience, but that significant effects of contextual factors should only be observed in interaction with the phonological salience. Under the second hypothesis, prompted in Lieberman (1965), context is primary, and phonological salience has little role in influencing the prominence of words. Here, main effects of contrastive context and of listener's discourse knowledge are predicted, and the effect of phonological salience is expected only when the contexts elicit the perception of prominence. The third hypothesis is a balanced one, where phonological salience and context (both discourse context and listener's knowledge) are equally weighted in terms of their influence on prominence. This hypothesis predicts that each factor should have an independent main effect, and an additive contribution to the perception of prominence when combined.

Experiment 1: monologue

Participants

A total of 43 monolingual speakers of American English participated in the experiment (27 female; aged 18–40, $M = 21.9$). Participants were recruited from among the Ohio State University community, and were compensated by either $10 or partial course credit.

Materials

Materials were drawn from a ToBI-annotated corpus of spontaneous speech (Ito & Speer, 2006). This speech corpus consists of spontaneous dialogues where a naïve participant directed a confederate in the decoration of a Christmas tree. The participants were presented with slides on a computer screen that indicated an ornament and the location to place it on the tree one by one, and gave instructions to the confederate accordingly. The ordering of the ornaments requested, and their position relative to each other, was manipulated by the experimenters to alter the information status of the referring expressions used to denote the ornaments. Although the labels for the ornaments, such as "green egg", were suggested in text on the screen, the participants were free to speak as they pleased, and thus the speech was truly a spontaneous interactive dialogue. (Indeed, one participant routinely invented new names for the ornaments, such as *periwinkle tumour* and *disco ball*, establishing "conceptual pacts" (Brennan & Clark, 1996) with the confederate over the course of their dialogue.)

For the present experiment, simple adjective–noun phrases, such as "green egg", with the pitch accent sequence [H* !H*] or [L + H* 0] (where 0 represents an unaccented word) were selected from one female talker in the corpus. This particular talker was chosen because of her relatively common use of these contours; other talkers represented in the corpus did not have enough distinct phrases using these contours to allow for a balanced design. A total of 22 adjective–noun phrases were selected. Ten of these phrases had a leading article *a* or *an*, while 12 others were bare. A total of 48 trials were created by pairing these utterances in sequences of two phrases each. The content of the phrases and their pitch accent tunes is listed in Appendix A. Figure 1 shows two representative samples of stimuli with pitch accent sequence [H* !H* and [L + H* 0].
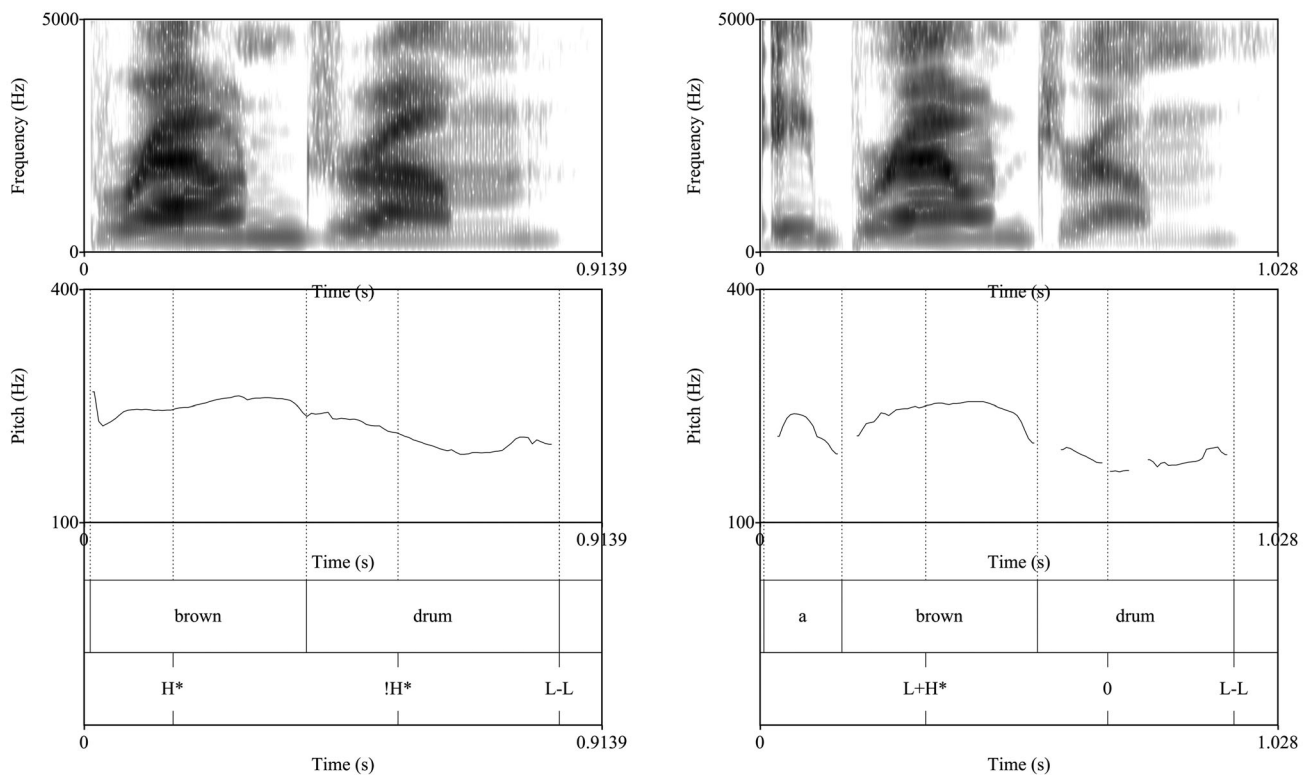
Figure 1. Spectrograms and f0 traces of two stimulus phrases used in the current experiments. Left: "brown drum" with [H* !H*]; right: "a brown drum" with [L + H* 0].

Each trial consisted of two phrases, "context" and "target", which were separated by 500 ms of silence. Each context-target sequence either elicited or did not elicit a contrastive interpretation of the adjective in the target phrase. A contrastive sequence repeated the noun (e.g. *blue egg, green egg*), which supported the notion that the referent *green egg* is being denoted in opposition to contextually determined alternatives (such as the *blue egg*, see Rooth, 1992). A non-contrastive sequence involved no repetition (e.g. *blue ball, green egg*). It is important to bear in mind that the term "contrastive" is used here to define the word sequence type, and not the pitch accent contour, which was manipulated orthogonally. If phrase sequence influences listeners' perception of prosodic prominence, the word *green* in the sequence *blue egg, green egg* may be rated as more prominent than the same word in the

sequence *blue ball, green egg*. Within each type of sequence (contrastive and non-contrastive), both the context and the target phrases appeared with either [H* !H*] or [L + H* 0] contour, yielding the design with eight conditions (Table 1). There were 6 trials in each condition, for a total of 48 trials in the experiment.

As can be seen, the current design investigates contrast on the adjective, but not on the noun. This choice was motivated simply by the availability of contour types in the original corpus of spontaneous speech. [L + H* 0] and [H* !H*] were the two most common contour types on two-word noun phrases, and they lend themselves naturally to an investigation of adjective contrast. Other contours which would have allowed an investigation of noun contrast (such as [H* L + H*]) were not common enough to make possible such an investigation. These constraints constitute one drawback of using stimuli drawn from a corpus of naturalistic, spontaneous speech.

Stimulus acoustics

To investigate the weight of psychoacoustic factors in the perception of prominence, each stimulus token underwent detailed acoustic analysis. On each of the adjective and the noun, a total of seven measures were

Table 1. Summary of design.

	Context phrase	Target phrase
Non-contrastive	blue ball [H* !H*]	green egg [H* !H*]
	BLUE ball [L + H* 0]	green egg [H* !H*]
	blue ball [H* !H*]	GREEN egg [L + H* 0]
	BLUE ball [L + H* 0]	GREEN egg [L + H* 0]
Contrastive	blue egg [H* !H*]	green egg [H* !H*]
	BLUE egg [L + H* 0]	green egg [H* !H*]
	blue egg [H* !H*]	GREEN egg [L + H* 0]
	BLUE egg [L + H* 0]	GREEN egg [L + H* 0]

taken: (1) word duration, (2) vowel duration, (3) peak f0, (4) mean f0 during the vowel, (5) peak f0 alignment relative to vowel onset, and (6 and 7) two measures of spectral tilt. The spectral tilt measures, following Cole, Mo, and Baek (2010), were the difference between the mean intensity of two different spectral bands, either 2 kHz in bandwidth (i.e. 0–2 kHz minus 2–4 kHz) or 4 kHz in bandwidth (i.e. 0–4 kHz minus 4–8 kHz). Spectral tilt is known to be related to hyperarticulation in general and prosodic prominence in particular (Cole, Mo, & Baek, 2010; Sluijter & Van Heuven, 1996; Tamburini, 2005).

In addition to the above 14 measures, two f0 slope measures were taken to capture the gross pitch shape from the adjective to the noun: the slope from the adjective peak f0 to the noun peak f0; and the slope from the adjective vowel mean f0 to the noun vowel mean f0 for the interval between the adjective's vowel midpoint to the noun's vowel midpoint. Both of these slopes were expressed as a change in Hz per second. Finally, pitch excursion – the f0 change in Hz from trough to peak – was measured for the adjectives. (Excursion is not reported for nouns as no substantial f0 movement was observed on unaccented words or words with !H* accents.)

Before discussing the results of this analysis in detail, it is worthwhile to bear in mind that it is quite normal for phonological categories to have overlapping phonetic cues. The difference between the vowels in *beat* and *bit* are clearly different to the native listener, but the acoustic properties of F1, F2, and duration cannot uniquely bisect the acoustic space of tokens of these sounds. This overlap is especially noticeable in spontaneous, naturally produced speech. Nevertheless, with enough tokens, it is possible to use statistical methods to establish that the means of these distributions differ, and therefore conclude that the acoustic dimensions of F1, F2, and duration are somewhat relevant to distinguishing *beat* from *bit*.

The stimuli in the current experiment are, for the purposes of establishing a phonetic distinction, relatively few – only 14 [H* !H*] tokens and 8 [L + H* 0] tokens. A power analysis shows that for a sample of this size, a t-test has power of .574 to find a significant effect of Cohen's $d = 1$, assuming an alpha of .05. Note that Cohen deemed an effect size of $d = 0.8$ to be "large". These facts mean that the observation of significant differences between these two categories is unlikely, regardless of whether there is a true difference between the means. Further, it is well-known from phonetic studies of segmental phenomena, such as the vowels that distinguish *beat* and *bit*, that variability and overlap between distinct categories is extremely common.

Despite these limitations, significant differences were observed between the [H* !H*] and the [L + H* 0] stimuli for two of these 17 acoustic variables: noun mean vowel f0 ($t(12.476) = 2.388$, $p = .034$) and vowel-to-vowel mean f0 slope ($t(13.545) = 2.234$, $p = .043$). Noun mean vowel f0 was higher for [H* !H*] stimuli (208 Hz) than for [L + H* 0] stimuli (182.6 Hz), and vowel-to-vowel mean f0 slope was shallower for [H* !H*] stimuli (−164.2 Hz/s) than for [L + H* 0] stimuli (−252.3 Hz/s). The 15 other acoustic variables were not significantly different between the two contour types (all $p > .05$). Means and standard errors of each measure are shown in Appendix B.

These results could be argued to suggest that the difference between these contours is realised solely on the noun. That is, the distinction is cued by the noun f0 height, meaning that the relevant property which distinguishes these contours is the distinction between !H* and 0, while the adjectives do not differ in pitch accent (contrary to the transcription). To test this possibility, we constructed a support vector machine to predict contour type using only the acoustic features of the adjective.[2] The model achieved an accuracy significantly above chance, at 86.3% (19 out of 22). While none of the adjective acoustic measures can alone separate the categories, which is hardly surprising given the small sample size and the inherent variability in spontaneous speech, the measures can accurately predict category when used together. Although the precise interplay of features within a support vector machine can be difficult to interpret, it is possible to assess the importance of particular features to the classification by examining the magnitude of the weight for that feature, and by comparison to the accuracy of a machine constructed without that feature. By these metrics, the acoustic measures of peak f0, word duration, and both spectral tilt measures were all important to the accurate prediction of pitch accent category for the adjective.

Two alternative conclusions can be drawn from these analyses. The first is that the pitch contours represented in this stimulus set may not differ in a large number of acoustic variables, including those purported to be relevant for the H*/L + H* distinction, such as adjective f0 peak, mean f0, peak alignment, and duration. Therefore, any distinct behavioural response between the contours may not be due to mere psychoacoustic prominence of the words that are labelled with H* and L + H*, but rather the listeners' gestalt perception of the relative prominence within a continuous pitch contour. The second possibility is that the current acoustic cues used to identify the differences between the pitch accents are inadequate, and while they may be appropriate for some talkers, they do not yield consistent identification of pitch accents for this particular talker that provided

the stimuli. In either case, these acoustic analyses suggest that systematic variation in these stimuli is primarily in phonological pitch accent, rather than raw acoustic cues. Any results can therefore be inferred to be due to these phonological distinctions among the stimuli, and not mere psychoacoustic perceptual salience.

Procedure

The task instructions provided to the participants were as follows:

> In normal speech, speakers pronounce some word or words in a sentence with more prominence than others. The prominent words are in a sense highlighted for the listener, and stand out from other, non-prominent words. In this experiment, you will listen to phrases and be asked to identify which words are prominent.

These instructions were followed by opportunities to ask questions, and further instructions on the logistical details of how to respond.

After the instructions and consent procedure were completed, participants were seated in front of a computer screen in a sound-attenuated booth. On each trial, a pair of phrases was presented over headphones. The four words that composed the context-target phrase sequence appeared on the screen (e.g. blue ball green egg). Participants used a button box to mark which of the four words they perceived to be prominent. Pressing a particular button (e.g. number 1) would select a particular word on the screen (e.g. adjective in the context phrase); pressing the button again would deselect the word. There were no limits on the participants' selections: all four words could be selected, or none, or any possible combination of markings. Figure 2 shows an example screen display from a trial, with the word "brown" selected as prominent. The marking was self-paced and participants could take as much time as they wanted to ponder their selection. Once participants finished making their selection, they pressed the fifth,

"continue" button, which replayed the stimulus and gave the participants another chance to alter their markings. A second press of the "continue" button progressed the experiment to the next trial. Thus, each stimulus pair was presented exactly twice to each participant, and each participant had ample opportunity to respond.

Results

A mixed effects logistic regression model tested whether or not the adjective of the target phrase (e.g. the *green* of *blue egg, green egg*) was marked as prominent.[3] Fixed effects were the sequence type (contrastive or non-contrastive), the contour type of the context phrase ([H* !H*] or [L + H* 0]), and the contour type of the target phrase ([H* !H*] or [L + H* 0]). Since a number of participants reported during debriefing that they attended to the article *a* at the beginning of some phrases, the presence or absence of the article at the beginning of the first phrase and the second phrase were also added as two *ad hoc* fixed effects. Also included were two-way interactions between sequence type and context contour type, and between sequence type and target contour type. Random intercepts of item and participant were included, as were random slopes of target contour type and sequence type for participant. All effects were sum-contrast coded, with the contrastive sequence, [L + H* 0] contours, and phrases with articles coded with +0.5, and the non-contrastive sequence, [H* !H*] contours, and phrases without articles coded with −0.5. A total of 2064 prominence judgements (43 participants × 48 trials) were entered into the analysis.

The output of the mixed effects logistic regression model is reported in Table 2. Two predictor factors – target contour type and the presence of an article in the context phrase, showed a significant effect on the likelihood of prominence marking. First, the adjective in a [L + H* 0] target contour was more likely to be marked as prominent (80.8%) than the adjective in a [H* !H*] target contour (56.3%). This result confirms

green egg   brown drum

Figure 2. Example display screen from a trial in Experiment 1. Here, the word "brown" has been selected as prominent (highlighted in red), while the other words have not been selected.

Table 2. Output of logistic regression model for experiment 1.

	β	Std. Error	z	p
Intercept	1.029	0.405	2.544	**.011**
Sequence	0.215	0.124	1.727	.084
Context pitch accent contour	0.063	0.118	0.530	.596
Target pitch accent contour	1.621	0.273	5.933	**<.001**
Context article	−0.664	0.140	−4.738	**<.001**
Target article	0.138	0.222	0.622	.534
Sequence × context pitch accent contour	−0.281	0.221	−1.270	.204
Sequence × target pitch accent contour	0.553	0.248	2.233	**.026**

Significant *p*-values are in boldface.

that the phonological status of the contour predicted the prominence perception. Second, the presence of an article in the context phrase showed an unexpected effect on prominence marking. When preceded by a context phrase with an article, adjectives in target phrases were rated as less prominent (63.8%) than when the context phrase had no article (72.3%). That is, the word *green* was rated as more prominent in *a blue ball, green egg* than in *blue ball, green egg*.

In addition, a significant interaction between target contour and sequence type was observed. While sequence type alone did not influence prominence markings for the [H* !H*] contour (mean ratings of 55.8% and 56.8% for non-contrastive and contrastive focus respectively), adjectives in the [L + H* 0] contour in the contrastive sequence were perceived to be more prominent (82.4%) than adjectives in [L + H* 0] in the non-contrastive sequence (79.3%) (see Figure 3). This interpretation was confirmed by a *post hoc* analysis of the [L + H* 0] and [H* !H*] subsets separately, repeating the above logistic regression model (without a term for target pitch accent contour). In the model for the [L + H* 0] phrases, sequence type was observed to be significant ($\beta = 0.629$, $z = 2.618$, $p < .01$), but sequence type was not significant in the [H* !H*] model ($\beta = -0.054$, $z = -0.348$, $p = .728$).

Discussion

When two phrases were presented side-by-side with minimal discourse context, prominence markings of the adjective in the target phrase were primarily influenced by the pitch contour type of the target phrase. Those
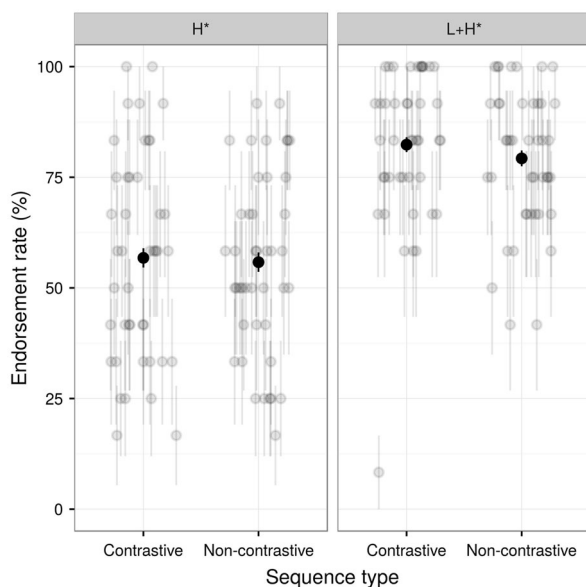
Figure 3. Prominence endorsement rates for target adjectives in Experiment 1, split by sequence type and pitch accent.

with the [L + H* 0] contour were much more likely to have the adjective marked as prominent than those with the [H* !H*] contour. Thus, naïve listeners were sensitive to the difference between the two contour types produced by a naïve speaker. This shows that the trained ToBI-labellers' phonological distinction between the two contours bear perceptual reality in ordinary listeners. However, this result does not necessarily imply that participants of the present experiment made a phonological distinction between L + H* and H* for the adjective. The higher prominence marking for the adjectives with L + H* than for those with H* may have resulted from the perception of *relative* prominence on the adjective in comparison to the following noun after the entire phrase was processed. That is, because our stimulus phrases are either [L + H* 0] or [H* !H*], it is not possible to separate effects due to the pitch accent type on the adjective versus the noun. The simplest interpretation is that the prominence of the adjective is mediated by the pitch accent type of the adjective, but this cannot be determined with absolute certainty. In any case, the results demonstrate a strong effect of contour type on perceived adjective prominence.

The present data showed no main effect of sequence, that is, the repetition of the noun alone did not lead to more frequent perception of prominence on the target phrase's adjective (blue egg, green egg). However, the significant interaction between sequence and target pitch accent type reveals that participants were sensitive to the information structure of the phrases, but only when the prosody supported the contrastive interpretation: adjectives in contrastive [L + H* 0] phrases were endorsed as prominent more often (82.4%) than those in non-contrastive [L + H* 0] phrases (79.3%) ($\beta = 1.621$, $z = 5.933$, $p < .001$). For [H* !H*] contours, sequence type had no effect on prominence marking (56.8% for contrastive vs. 55.8% for non-contrastive). There are two alternative interpretations for this asymmetry between [L + H* 0] and [H* !H*]: First, the result may support the phonology-primary hypothesis, which predicts that acoustic salience (that lead to phonological distinction) must be present for evoking prominence whereas the sequence of referential expressions has only an additive effect to enhance the perception of prominence. However, the present result contradicts previous findings that information structure influences prominence perception independently of acoustics (Bishop, 2012; Bock & Mazzella, 1983). An alternative account is that a discourse context such as contrastive sequence has an effect to enhance the perception of prominence independently of the presence or absence of acoustic salience, yet the current task failed to establish a discourse context within which a sequence that repeats the noun can be interpreted

contrastively. Since the stimuli were presented in an extremely impoverished discourse context, participants may have not processed each phrase pair as part of a coherent discourse. To further explore the effect of local context, experiment 2 inserted a brief response from the male confederate between the context and the target phrases to make the same phrase sequences sound like an extraction from a conversation between two speakers.

Experiment 2: plain dialogue

Participants

A total of 42 monolingual speakers of American English (28 female; aged 18–34, $M = 21.0$) participated in the experiment, none of whom had participated in experiment 1. Participants were recruited from among the Ohio State University community, and were compensated by either $10 or partial course credit.

Materials

Experiment 2 used the same stimuli from Experiment 1, but a short feedback phrase was inserted between each context and target phrase to create the percept of an ongoing dialogue, rather than a monologue. The short utterances such as "Alrighty, next" and "Okay, next" were produced by the male confederate (the tree decorator) in Ito and Speer (2006), and thus were part of the same corpus that provided the female talker's stimuli phrases. Six such phrases were inserted between the context and target phrases. The mean duration of the inserted phrase was 968 ms (min: 575 ms; max: 1589 ms). All other aspects of the stimuli were identical to those of Experiment 1.

Procedure

The procedure for Experiment 2 was identical to that of Experiment 1. No additional instruction was provided regarding the discourse background of the stimuli.

Results

A mixed effects logistic regression model tested whether or not the target adjective was marked as prominent. Model structure was identical to that of Experiment 1. A total of 2016 prominence judgements (42 participants × 48 trials) were entered into the analysis.

The output of the mixed effects logistic regression model is reported in Table 3. The overall pattern of results was very similar to Experiment 1, showing the main effects of the pitch contour type and the presence

Table 3. Output of logistic regression model for experiment 2.

	β	Std. Error	z	P
Intercept	0.751	0.499	1.507	0.132
Sequence	0.171	0.123	1.391	0.164
Context pitch accent contour	0.063	0.125	0.500	0.617
Target pitch accent contour	2.097	0.280	7.477	**<0.001**
Context article	−0.562	0.147	−3.828	**<0.001**
Target article	0.764	0.260	2.939	**0.003**
Sequence × context pitch accent contour	0.107	0.228	0.470	0.638
Sequence × target pitch accent contour	0.956	0.249	3.835	**<0.001**

Significant *p*-values are in boldface.

of an article for the context phrase: adjectives with L + H* were more likely to be marked (76.9%) than adjectives with H* (48.7%); and adjectives in target phrases preceded by context phrases with an article were less likely to be marked (59.5%) than those preceded by context phrases with no article (65.3%). In addition, the presence of an article in the target phrase showed an effect in Experiment 2: adjectives in target phrases with articles were perceived as more prominent (69.1%) than adjectives in target phrases without articles (53.9%). This effect is in the opposite direction to that of the article presence on the context phrase. Taken together, these effects suggest than an article enhances the prominence of its own phrase while diminishing the prominence of other phrases. Again, there was a significant interaction between sequence type and contour type in the target phrase suggesting that adjectives with L + H* were perceived as more prominent when appearing in contrastive sequences (79.6%) than in non-contrastive sequences (74.2%) (see Figure 4).
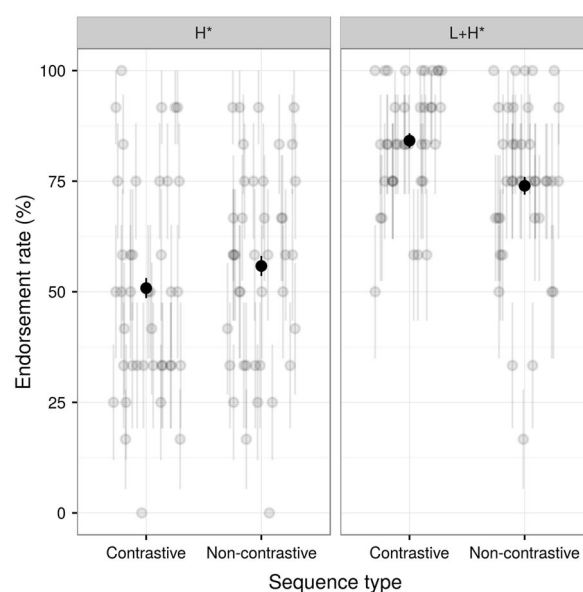
Figure 4. Prominence endorsement rates for target adjectives in Experiment 2, split by sequence type and pitch accent.

Discussion

The results of this experiment were qualitatively similar to those of Experiment 1. Again, adjectives in phrases with the [L + H* 0] contour were marked as more prominent (76.9%) than those in phrases with the [H* !H*] contour (48.7%). Also, adjectives in contrastive [L + H* 0] phrases were marked as more prominent than those in non-contrastive [L + H* 0] phrases, with no effect of sequence on [H* !H*] phrases. However, the magnitude of these effects appears to be greater in Experiment 2 than in Experiment 1. While in Experiment 1, the contrastive sequences led to a 3.1 percentage point increase for endorsement of [L + H* 0] phrases as compared to non-contrastive sequences, in Experiment 2 the sequence manipulation led to an increase of 5.4 percentage points. While this small increase in the magnitude of the effect may be due to differences between subject groups in the two experiments, it may also suggest that the insertion of a confederate's utterance between the two phrases may have led listeners to process the sequence as a coherent discourse more than in Experiment 1. As in Experiment 1, there was no effect of sequence on [H* !H*] phrases, that is, the effect of contrastive sequence was not shown unless the target phrase had phonetic prominence. Again, we are faced with the two alternative interpretations of the interaction: either that prosody serves as the primary cue to prominence whereas the discourse context alone can only add to the effect by evoking contrast (contra Bishop, 2012; Bock & Mazzella, 1983); or that the stimuli presentation of Experiments 1 and 2 simply did not allow the participants to interpret the sequence as naturalistic discourse.

In favour of the latter alternative are comments received in the post-experiment debriefing. While several participants reported that they believed the recordings to be excerpted from a true and real dialogue, some claimed that they found the juxtaposition of phrases implausible or unexpected. Further, several participants mentioned that they found the inserted feedback utterances to be "annoying" or "distracting". Quantitative comparison of the results of this experiment to those of Experiment 1 suggests that there was overall less prominence marking in this experiment: target adjectives were marked as prominent 68.6% of the time in Experiment 1 but only 62.9% of the time in Experiment 2 ($\chi^2(1) = 14.755$, $p < .001$). Thus, it is possible that the inserted phrases were distracting and unexpected, and disrupted the participants' attention and the flow between the phrases, resulting in fewer overall markings. Since participants of Experiment 2 had no information as to what communicative purpose those sequential utterances had, assessing prominence of two phrases that were separated by another voice may have been rather difficult. Experiment 3 therefore was designed to provide the participants with an explicit explanation of the purpose and the context of discourse between the two speakers, *without a change in stimuli*. If listeners' awareness of the discourse background is tightly related to their processing of discourse structure, the contrastive sequence should lead to the largest effect in enhancing the percept of prominence when participants have the richest information about the corpus. Importantly, as predicted by the context-primary hypothesis, phonological prominence may not remain as the primary cue to prosodic prominence when listeners' knowledge about the task background is sufficient for processing the phrase sequences as a coherent discourse.

Experiment 3: narrative dialogue

Participants

A total of 40 monolingual speakers of American English (26 female; aged 17–35, $M = 20.7$) were recruited from among the Ohio State University community, and were compensated by either $10 or partial course credit. None of them had participated in Experiments 1 or 2.

Materials

Auditory stimuli were the identical to those used in Experiment 2.

Procedure

The procedure for this experiment was the same as that of Experiment 2, with the sole exception of the task instructions. In this experiment, participants were told that the voices they would hear are from a previous experiment where two people talked to each other to decorate a Christmas tree. They were presented with around 40 seconds of the spontaneous conversation between the instructor and the decorator, a diagram of the experimental setup, and a description of how the slides on the screen elicited the speech from the instructor. Participants were encouraged to imagine that they were physically present with both the instructor and the decorator, and to mark "the words that stood out the most" (see Appendix C for the specific instructions presented to participants in Experiment 3).[4] This manipulation in the instruction offered the

participants more background information on the recordings, allowing them to be aware of a discourse context to a much greater degree than in Experiments 1 or 2.

Results

A mixed effects logistic regression model with a structure identical to that of Experiments 1 and 2 tested how the same predictor factors affected the likelihood of prominence marking. A total of 1,920 prominence judgements (40 participants × 48 trials) were entered into the analysis.

The output of the model is reported in Table 4. The same effects observed in Experiments 1 and 2 were again observed in this experiment: significant simple effects of target pitch contour, context phrase article presence, and target phrase article presence, and a significant interaction between sequence and target pitch accent contour. That is, adjectives with L + H* were overall more likely to be marked (79.1%) than adjectives with H* (53.3%); adjectives in target phrases preceded by context phrases with an article were less likely to be marked (64.0%) than those preceded by phrases with no article (67.9%); adjectives in target phrases with an article were more likely to be marked (69.6%) than those in targets phrases with no article (61.5%); and adjectives with L + H* were perceived as more prominent (84.2%) in contrastive sequences than in non-contrastive sequences (74.0%) (see Figure 5). Importantly, a significant main effect of sequence type, which was not observed to be significant in Experiments 1 and 2 was observed in this experiment. That is, adjectives in the contrastive sequences were marked as more prominent (67.5%) than those in non-contrastive sequences (64.9%).

The use of sum contrasts in the model design means that each of these effects can be interpreted as equivalent to an ANOVA "main effect" (see Clopper, 2013, for discussion of contrast structure in mixed effects modelling). However, it appears from Figure 5 that this main effect of sequence type is driven solely by the [L + H*

Table 4. Output of logistic regression model for experiment 3.

	β	Std. error	z	P
Intercept	0.827	0.466	1.774	.076
Sequence	0.316	0.127	2.491	**.013**
Context pitch contour	−0.078	0.129	−0.603	.546
Target pitch contour	1.897	0.272	6.984	**<.001**
Context article	−0.455	0.151	−3.016	**.003**
Target article	0.532	0.251	2.120	**.034**
Sequence × context pitch accent contour	0.069	0.236	0.290	.772
Sequence × target pitch accent contour	1.376	0.262	5.246	**<.001**

Significant *p*-values are in boldface.
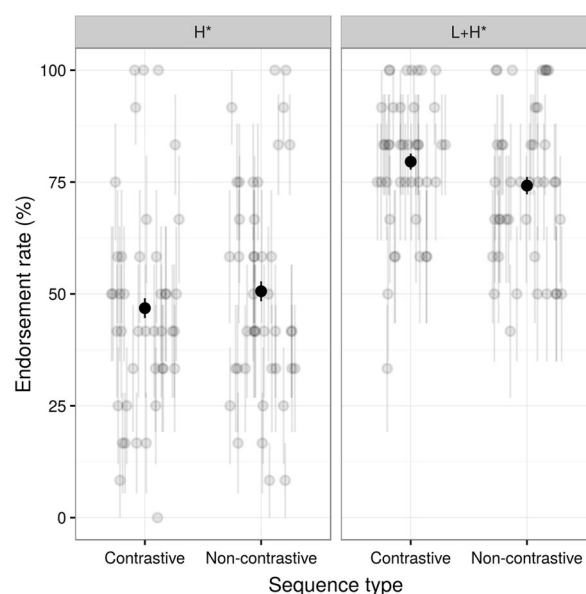
Figure 5. Prominence endorsement rates for target adjectives in Experiment 3, split by sequence type and pitch accent.

0] phrases. To test for this hypothesis, the analysis was repeated but the crucial variables were re-coded with treatment contrasts: target phrase pitch accent contour (with [H* !H*] as baseline), and sequence type (with non-contrastive sequence as baseline). Under this coding scheme, the sequence type coefficient is interpreted as the effect of the sequence type solely on the [H* !H*] phrases, with any effects of sequence type on [L + H* 0] trials that deviate being indicated in the interaction term between pitch accent and sequence type. Under this analysis, sequence type was not significant for the [H* !H*] trials ($p = .073$), but it was significant for the [L + H* 0] trials ($\beta = 1.081$, $z = 4.546$, $p < .001$).

Discussion

As observed in Experiments 1 and 2, [L + H* 0] phrases were more often marked as prominent than [H* !H*] phrases. Again, an interaction between sequence and pitch contour was observed, with adjectives in contrastive [L + H* 0] phrases being marked as more prominent than those in non-contrastive [L + H* 0] phrases. That is, the sequence manipulation had an effect on phrases with [L + H* 0], but not on those with [H* !H*].

These effects are all the more striking when considering that the stimuli herein were exactly the same as those in Experiment 2, and the only difference was that participants were made aware of the provenance of the recordings and asked to imagine themselves listening to an ongoing dialogue. Experiment 2 saw an increase of 5.4 percentage points for [L + H* 0] phrases in contrastive vs non-contrastive sequences (compared

to Experiment 1), while Experiment 3 saw a 10.2 percentage point increase (again compared to Experiment 1). These effects speak to the influence of participants' beliefs about the stimuli they are exposed to: when listeners process utterances as part of a coherent discourse where they believe the speaker to have communicative intent, the informational status of words assigned by a local context can impact the perception of prominence (see also Speer, Warren, & Schafer, 2011).

These results also suggest that even in the presence of sufficient discourse context, the influence of information structure on prominence perception is conditioned on the presence of acoustic prominence. When listeners process a discourse, a congruence between acoustic prominence and information structure seems to lead to the highest likelihood of prominence perception.

General discussion

In comparing the three experiments that varied in the presence or absence of the second voice of the confederate and in the presence or absence of explicit description of the discourse background, some general trends appear. Foremost among these is the effect of pitch contour. In all three experiments, the adjective in a [L + H* 0] phrase was endorsed as prominent more often than the adjective in a [H* !H*] phrase. As discussed in the stimulus analysis in Experiment 1, this result cannot be accounted for simply by isolated acoustic cues such as the adjective's f0, intensity, and duration, since the [L + H* 0] phrases and the [H* !H*] phrases did not differ significantly along these dimensions. Rather, we maintain that the contrast between these two contour types is a phonological distinction, cued by a complex of multiple phonetic correlates.

Another clear trend across the experiments is the lack of effect of pitch contour of the context phrase. The pitch accent contour of the preceding phrase never significantly influenced the prominence marking of the target phrase. This result suggests that, while prominence is relational, it is bounded in scope, presumably at the level of the intonational phrase. In this connection, it is worthwhile to note that participants were asked to mark the presence or absence of prominence on all four words – two each in the context and target phrases – in each trial. Their attention was therefore directed towards considering all four words as a single entity, but the pitch accent contour of the context phrase appears to never have influenced the perception of prominence of the target phrase.

The lack of pitch contour effects of the context phrase is surprising given Bock and Mazzella's (1983) finding that a prominent element in a context phrase can trigger expectation of a prominent element in the target phrase (cf. Foltz, 2010, showing complex interactions of prosodic prominence and edge tones in the processing of context vs. target adjectives in successive sentences). It is possible to carry out a *post hoc* analysis to determine if the *perceived prominence* of the context adjective (rather than the pitch contour) has an influence on the perceived prominence of the target adjective. This possibility was examined by examining rates of prominence marking on target adjectives with L + H*, split by whether or not the context adjective was marked as prominent. When the context was marked as prominent, the target adjectives were marked as prominent 77.0% of the time; when the context was not marked as prominent, the target adjectives were marked as prominent 82.2% of the time. This pattern provides further evidence suggesting that prominence is relational, not absolute, although in this case, it is not possible to determine whether the context adjective prominence affects the target adjective prominence or vice versa. This pattern also suggests that the prominent element in the context phrase did not prompt expectation of a (contrastive and) prominent element in the target phrase. This result is counter to that of Bock and Mazzella but can be explained as being due to differences in sentential content between the experiments. That is, Bock and Mazzella's stimuli were essentially corrective utterances ("Arnold didn't fix the radio, Doris fixed the radio"), whereas the current stimuli involve no such illocutionary force, suggesting that the expectation effect is driven by both sentential context and prosodic prominence.

Nevertheless, one aspect of the context phrase – the presence or absence of an article – did influence the perception of prominence in the target phrase. In Experiments 2 and 3, the presence of an article on the target phrase also affected the prominence rating. While the presence of an article in the context phrase decreased the prominence marking rate of the target adjective, an article in the target phrase increased it. These data suggest an effect of article presence on the construction of rhythmic structure – another factor that can influence the perception of prominence. While we cannot examine the effect of projected rhythmic structure and its interaction with acoustics or discourse context with the present data, these remain interesting topics for future research. These patterns further confirm the relative nature of prominence: the articles were unaccented and gave rise to the salience of the following material (i.e. adjective). In the context phrase, the heightened prominence of the context adjective due to the article may have made the target adjective relatively less prominent, leading to a decrease in prominence markings. Similarly, it could be that the article – either *a* or *an* – implied that

the noun is new to the discourse, and was therefore perceived to be more prominent (Calhoun, 2010; cf. Turnbull, in press). Table 5 shows percentage prominence marking rates for each word in each trial, split by the presence of an article at the beginning of the word's phrase. As can be seen, adjectives are generally more likely to be marked as prominent when preceded by an article.

In terms of information structure, although the contrastive sequence context alone did not reliably increase prominence ratings for target adjectives, it did produce a consistent numerical trend across experiments. In addition, a significant interaction between the pitch contour type and sequence was observed consistently in all three experiments: Contrastive [L + H* 0] was marked as more prominent than non-contrastive [L + H* 0]. Importantly, the size of this effect varied with the listener's awareness of discourse context. The difference between contrastive and non-contrastive [L + H* 0] was the smallest in the monologue condition (Experiment 1), slightly larger with inserted turns (Experiment 2) and the largest when a detailed scenario directed participants to interpret the stimuli sequence as part of a task-oriented dialogue (Experiment 3). Thus, listeners required stronger support of dialogue context in order to assign information status to the sequential phrases. When listeners process the incoming stream of speech as components of a coherent discourse, the confluence of phonological cues, such as pitch accent, with semantic cues, such as contrast, becomes important. Given the present results, we argue that the phonological and semantic cues must be "aligned" in some manner for prominence to be enhanced.

Overall patterns of pitch accent prominence

To determine whether the interaction between the target phrase's pitch contour type and the sequence was reliably enhanced across the experiments, the data from all three experiments were pooled into one large dataset. With the merged dataset, the robust effect of phonological (and phonetic) properties of the contours on the prominence perception was confirmed. Table 6 shows endorsement rates for adjectives and noun in all
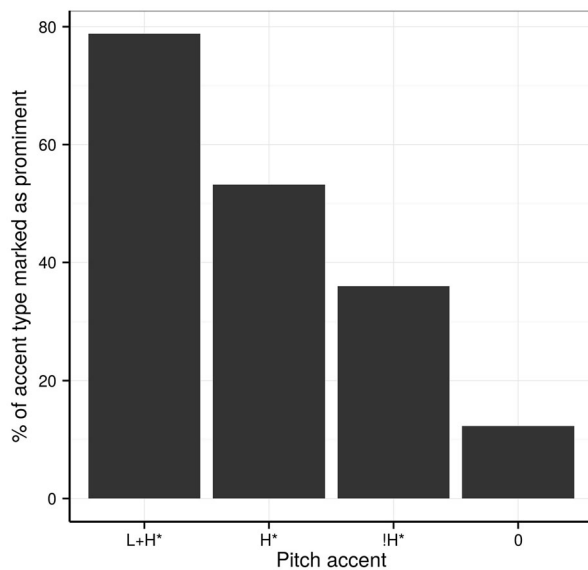
target phrases split by contrast and pitch accent contour. Figure 6 shows mean prominence markings for all pitch accents (L + H*, H*, !H*, and 0) on all words collapsed over all three experiments, from a total of 24,000 judgements in total ([(43 participants × 48 trials) + (42 participants × 48 trials) + (40 participants × 48 trials)] × 4 words). A clear hierarchy of pitch accent prominence emerges, such that the likelihood of prominence marking was highest to lowest in the order of L + H* > H* > !H* > 0. This pattern is consistent with the expected patterns of prominence in American English. Since all of the L + H*s and H*s were associated with phrase-initial adjectives, and all of the !H*s and unaccented words were phrase-final nouns, this hierarchy also suggests that the early adjectives were overall more likely to be marked as prominent than the later nouns. It is not clear whether this pattern is due to the part of speech or the linear sequence within a phrase. Further, since L + H* adjectives always preceded an unaccented noun (in [L + H* 0]), and the H* adjectives always preceded a noun with !H* (in [H !H*]), the hierarchy may also reflect the relative assessment of prominence within each phrase (e.g. the unaccented noun within [L + H* 0] contour sounded less prominent than the noun in [H* !H*]). Thus, the present data do not demonstrate absolute prominence distinctions among the pitch accent types for isolated words. Rather, they show consistent differences in the likelihood of prominence perception for words produced in the two contours. Importantly, the likelihood of prominence marking by our naïve participants corresponded to the different pitch accent types identified by the trained ToBI annotators, validating the practice of inferring relative prominence levels from ToBI annotations.

Between experiment analysis

To test the effect of listeners' beliefs about the discourse status of spoken stimuli, a mixed effects logistic regression model structure was constructed to include factors identical to the previous analyses, with one additional fixed effect of experiment (treatment coded: monologue (baseline), plain dialogue, or narrative dialogue). The model also included three kinds of two-way interactions between experiment and context contour

Table 5. Percentage endorsement rates for each word in each trial, split by the presence of an article at the beginning of the word's phrase.

		No article	Article
Context	Adjective	56.0	79.1
	Noun	25.7	22.2
Target	Adjective	59.1	70.7
	Noun	26.0	22.9

Table 6. Percentage endorsement rates (standard deviations in parentheses) for adjectives and noun in all target phrases in all three experiments, split by contrast and pitch accent contour.

	Adjective		Noun	
	H* !H*	**L + H*** 0	H* **!H***	L + H* **0**
Non-contrastive	54.1 (1.3)	75.9 (1.1)	34.5 (1.2)	12.2 (0.8)
Contrastive	51.5 (1.3)	82.0 (1.0)	39.0 (1.3)	10.9 (0.8)

Figure 6. Overall prominence endorsement rates for all words, split by pitch accent.

Table 7. Output of logistic regression of grand analysis of all three experiments.

	β	Std.Error	z	p
Intercept	0.990	0.437	2.265	**.024**
Sequence	0.201	0.117	1.717	.086
Context pitch accent contour	0.052	0.115	0.454	.650
Target pitch accent contour	1.806	0.245	7.376	**<.001**
Narrative dialogue experiment	−0.145	0.221	−0.657	.511
Plain dialogue experiment	−0.334	0.218	−1.532	.126
Context article	−0.536	0.084	−6.345	**<.001**
Target article	0.379	0.171	2.216	**.027**
Context pitch accent contour × narrative dialogue experiment	−0.099	0.161	−0.615	.539
Context pitch accent contour × plain dialogue experiment	0.001	0.158	0.006	.996
Target pitch accent contour × narrative dialogue experiment	0.032	0.303	0.106	.916
Target pitch accent contour × plain dialogue experiment	0.073	0.299	0.246	.806
Sequence × context pitch accent contour	−0.237	0.223	−1.060	.289
Sequence × target pitch accent contour	0.508	0.237	2.146	**.032**
Sequence × narrative dialogue experiment	0.132	0.166	0.797	.426
Sequence × plain dialogue experiment	−0.050	0.162	−0.308	.758
Sequence × context pitch accent contour × Narrative dialogue experiment	0.278	0.322	0.864	.388
Sequence × context pitch accent contour × Plain dialogue experiment	0.325	0.316	1.028	.304
Sequence × target pitch accent contour × Narrative dialogue experiment	0.899	0.332	2.712	**.007**
Sequence × target pitch accent contour × Plain dialogue experiment	0.394	0.323	1.220	.222

Significant *p*-values are in boldface.

type, experiment and target contour type, and experiment and sequence type, and two kinds of three-way interactions between experiment, sequence type, and context contour type, and between experiment, sequence type, and target contour type.

In addition to the effects of target pitch contour type and article and the interaction between the sequence and pitch contour type, the results of cross-experiment analysis (Table 7) showed a significant three-way interaction between sequence type, target pitch contour type, and experiment. The boost in prominence afforded to contrastive [L + H* 0] phrases (relative to non-contrastive [L + H* 0] phrases) was significantly larger in Experiment 3 (Narrative Dialogue), relative to Experiment 1 (Monologue). This prominence boost was numerically larger in Experiment 2 (Plain Dialogue) than in Experiment 1 (Monologue), but was not significant (β = 0.394, p = .222).[5] Thus, the insertion of the confederate's turns *and* the detailed instructions about the discourse background altered the participants' beliefs about the stimuli, and enhanced the effects of contrastive sequence and [L + H* 0] to hear prominence on the target adjective.

The role of awareness of dialogue

Across these three experiments, we have shown that a participant's awareness of discourse can turn mere concatenated phrases (Experiment 1) into meaningful discourse units. In all three experiments, participants judged an adjective with a [L + H*] pitch accent to be more prominent when it was in a contrastive context (*blue egg …*

green egg) than when it was in a non-contrastive context (*blue ball … green egg*). Crucially, the magnitude of this effect increased with the enrichment of the discourse context, or the extent to which the participants believed that they were listening to a genuine dialogue with true illocutionary goals. In the first experiment, participants listened to concatenated phrases spoken by a single talker with no prior expectations about what they would hear. In the second experiment, the voice of another talker intervened the phrases, providing feedback to the first talker and thus made each sequence a dialogue. Finally, the third experiment provided participants with much more background information about the recordings, and instructed them explicitly to imagine directly observing the conversation. Across these manipulations, the magnitude of the effect of contrast on prominence increased from 3.1 percentage points (Experiment 1), to 5.4 percentage points (Experiment 2), and to 10.2 percentage points (Experiment 3).

These differences arose in spite of the fact that the target stimuli were acoustically identical in all three experiments. Prominence perception is therefore sensitive to the phonological form of the stimulus itself, the semantic context of the stimulus, and the extent to which the listener infers communicative intent on the part of the talker.

Conclusion

We have reported on three experiments that investigated the perception of prominence, and how this perception is influenced by pitch accenting and discourse-driven contrast. Stimulus phrases, consisting of an adjective and a noun (e.g. *green egg*) either bore a [L + H* 0] contour or a [H* !H*] contour, and, given the preceding phrase, the adjective was either contrastive or non-contrastive. In all three experiments, adjectives in [L + H* 0] phrases were more often marked as prominent than adjectives in [H* !H*] phrases. Additionally, participants marked adjectives in [L + H* 0] phrases as more prominent when the adjective was contrastive than when the adjective was not contrastive. Crucially, the magnitude of this effect increased as the discourse context was enriched (the between experiment manipulation). That is, the extent to which the listener had reason to believe that they were listening to a genuine dialogue with real illocutionary goals, rather than a contextless monologue, mediated the listener's responses to the contrastive [L + H* 0] phrases: the size of the prominence enhancement for these phrases due to contrast increased monotonically from experiments 1–3. These results suggest that the mechanism of prominence processing is reliant on pragmatic (metalinguistic) contextual factors, such as listener's awareness of discourse, as well as more linguistic contextual factors, such as contrast in a discourse.

Taken together, these results are consistent with the phonology-primary hypothesis, where phonological factors play the primary role in the perception of prominence, with context providing a supporting role when phonological prominence has been established. The context-primary and balanced-cue hypotheses both predict that context alone ought to be able to trigger prominence in the absence of supporting phonological factors, but this prediction is not borne out in our data. Importantly, the present data revealed that the effect of linguistic context that elicits the processing of information structure (which in turn influences the perception of prominence) is constrained by listeners' awareness of the discourse context.

The precise relationship between prominence perception and acoustic salience is yet to be explored; such an investigation must take into account both within-word prominence (due to, for example, lexical stress or the salience of individual phonemes, Kohler, 2008) and perceived vocal effort (Eriksson & Traunmüller, 2002). Future work that uses stimuli from multiple talkers, thus introducing intra-speaker variation, is also warranted. The extent to which the local context cues a target phrase as more or less predictable may be able

to influence prominence perception independently of other factors, but this is not currently known (see Turnbull et al., 2015, for an investigation of the role of predictability in prosody production). It is clear that prominence perception is a function of multiple factors, including acoustic salience, phonology, semantics, and awareness of discourse context. Future work should consider how these factors interact with each other in more enriched discourse context.

Notes

1. By 'acoustics', we refer to physical properties of the signal, such as fundamental frequency and harmonic structure. By 'phonology', we refer to the utterance's abstract properties, such as pitch accents and phonemic content.
2. A support vector machine is a type of machine learning model which classifies multi-dimensional tokens into a set number of categories (in this case, 2). The model finds the optimal separating hyperplane to bisect the space into the cleanest separation of categories. Relatively accessible introductions can be found in Chapter 12 of Hastie, Tibshirani, and Friedman (2009) and Chapter 5 of Styler (2015).
3. Since our conception of prominence is inherently relational, it is also possible to carry out an analysis of the reported prominence of the adjective relative to the noun. That is, if the adjective is marked as prominent and the noun is not, the relative prominence score is 1; if the noun is marked as prominent and the adjective is not, the score is −1; and if both or neither of the words are marked as prominent, the score is 0. Such an analysis yields the same pattern of results as those reported here for all three experiments and the cross-experiment analysis conducted in the discussion section. At the request of a reviewer, these models are reported in Appendix D. However, since the distribution of this relative prominence score fails to meet the assumption of normality required for linear regression modelling, we do not discuss these analyses here.
4. A reviewer asked whether the phrasing of 'words that stood out the most' could have encouraged participants to focus on the single most prominent word, leading to different response strategies from the other experiments. This hypothesis predicts that there should be overall fewer markings in Experiment 3 relative to the other experiments, as participants are biased to mark one and only one word. However, this prediction is not borne out in the data: Target adjectives were marked 66.2% of the time in Experiment 3, versus 68.6% and 62.8% in Experiments 1 and 2 respectively. In Experiment 3, the mean number of words that were marked as prominent (from a minimum of 0 to a maximum of 4) was 1.78; the means for Experiments 1 and 2 were 1.86 and 1.77, respectively. It does not appear, then, that the instructions necessarily had any influence upon the participants' overall propensity to mark words as prominent or not.
5. To allow the statistical comparison of Experiments 2 and 3, the model was re-constructed with Experiment 2 as

the baseline for the fixed effect of experiment. This model revealed that the prominence boost for [L + H* 0] phrases in contrastive contexts was not significantly larger in Experiment 3 relative to Experiment 2 ($\beta =$ −0.047, $p = .885$).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Rory Turnbull ⓘ http://orcid.org/0000-0002-2574-9064
Shari R. Speer ⓘ http://orcid.org/0000-0002-8737-1503

References

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. *Laboratory Phonology*, 9, 547–575. New York: Mouton de Gruyter.

Baumann, S. (2014, May). *The importance of tonal cues for untrained listeners in judging prominence*. Proceedings of the 10th international seminar on speech production, Cologne, Germany (pp. 21–24).

Baumann, S., & Röhr, C. T. (2015, August). *The perceptual prominence of pitch accent types in German*. Proceedings of ICPhS XVIII, Glasgow, Scotland.

Beckman, M. E., & Ayers, E. G. (1997). *Guidelines for ToBI labeling, vers 3.0* [manuscript]. Columbus, OH: Ohio State University.

Beckman, M. E., & Venditti, J. J. (2010). Tone and intonation. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 603–652). Somerset, NJ: Wiley. doi:10.1002/9781444317251.ch16

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 239–270). Berlin: Mouton de Gruyter. doi:10.1515/9783110261790.239

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, 11, 64–76. doi:10.3758/BF03197663

Bolinger, D. (1981). *Two kinds of vowels, two kinds of rhythm*. Reproduced by Indiana University Linguistics Club, number 289. Bloomington, IN: Indiana Linguistics Club, Indiana University.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493. doi:10.1037/0278-7393.22.6.1482

Buhmann, J., Caspers, J., van Heuven, V. J., Hoekstra, H., Martens, J.-P., & Swerts, M. (2002, May). *Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus*. Proceedings of LREC, Las Palmas, Canary Islands, Spain.

Burdin, R. S., & Clopper, C. G. (2015, August). *Phonetic reduction, vowel duration, and prosodic structure*. Proceedings of ICPhS XVIII, Glasgow, Scotland.

Burdin, R. S., Phillips-Bourass, S., Turnbull, R., Yasavul, M., Clopper, C. G., & Tonhauser, J. (2015). Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua*, 165, 254–276. doi:10.1016/j.lingua.2014.10.001

Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, 25, 1099–1140. doi:10.1080/01690965.2010.491682

Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, 40, 329–349. doi:10.1016/j.wocn.2011.12.001

Clopper, C. G. (2013). Modeling multi-level factors using linear mixed effects. *Proceedings of Meetings on Acoustics*. 19, 060028. doi:10.1121/1.4799729

Cole, J., Mahrt, T., & Hualde, J. I. (2014). Listening for sound, listening for meaning: Task effects on prosodic transcription. *Proceedings of Speech Prosody* 7, 859–863.

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, 25, 1141–1177. doi:10.1080/01690960903525507

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1, 425–452. doi:10.1515/labphon.2010.022

Eriksson, A., & Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Perception & Psychophysics*, 64 (1), 131–139. doi:10.3758/BF03194562

Foltz, A. (2010). *How listeners resolve reference: Effects of pitch accent, edge tones, and lexical contrast* (PhD thesis). Ohio State University Columbus, OH.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, W. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *The Journal of the Acoustical Society of America*, 102, 3009–3022. doi:10.1121/1.420355

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer. doi:10.1007/978-0-387-84858-7

Hualde, J. I., Cole, J., Smith, C. L., Eager, C. D., Mahrt, T., & Napoleão de Souza, R. (2016). The perception of phrasal prominence in English, Spanish and French conversational speech. *Proceedings of Speech Prosody*, 8, 459–463. doi:10.21437/SpeechProsody.2016-94

Ito, K., Bibyk, S., Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, 41 (1), 84–110. doi:10.1017/S0305000912000554

Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in empirical prosody research* (pp. 229–257). Berlin: Mouton de Gruyter.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of*

Memory and Language, 58, 541–573. doi:10.1016/j.jml.2007.06.013

Ito, K., & Speer, S. R. (2011). Semantically-independent but contextually-dependent interpretation of contrastive accent. In P. Prieto, S. Frota, & G. Elordieta (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 69–92). New York, NY: Springer. doi:10.1007/978-94-007-0137-3_4

Jyoti, P., Cole, J., Hasegawa-Johnson, M., & Puri, V. (2014). An investigation of prosody in Hindi narrative speech. *Proceedings of Speech Prosody*, 7, 623–627.

Kimball, A., & Cole, J. (2014). Avoidance of stress class in perception of conversational American English. *Proceedings of Speech Prosody*, 7, 497–501.

Kohler, K. (2008). The perception of prominence patterns. *Phonetica*, 65, 257–269. doi:10.1159/000192795

Kurumada, C., Brown, M., Bibyk, S. A., Pontillo, D., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335–342. doi:10.1016/j.cognition.2014.05.017

Ladd, D. R. (2008). *Intonational phonology* (2nd Ed.). Cambridge, UK: Cambridge University Press. doi:10.1017/cbo9780511808814

Ladd, D. R., & Schepman, A. (2003). 'Sagging transitions' between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, 31, 81–112. doi:10.1016/S0095-4470(02)00073-6

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. *WORD*, 21(1), 40–54. doi10.1080/00437956.1965.11435417

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(11), 1773–1781. doi:10.1121/1.1918816

Luchkina, T., & Cole, J. (2014). Structural and prosodic correlates of prominence in free word order language discourse. *Proceedings of Speech Prosody*, 7, 1119–1123.

Luchkina, T., Puri, V., Jyoti, P., & Cole, J. S. (2015, August). *Prosodic and structural correlates of perceived prominence in Russian and hindi*. Proceedings of ICPhS XVIII, Glasgow, Scotland.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America*, 66, 363–369. doi:10.1121/1.383670

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (PhD thesis). Massachusetts Institute of Technology.

Pierrehumbert, J. B., & 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. 342–365). Cambridge, MA: MIT Press.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech* (2nd release) [www.buckeyecorpus.osu.edu]. Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Portele, T. (1998, November-December). *Perceived prominence and acoustic parameters in American English*. Proceedings of interspeech 1998, Sydney, Australia.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, 75–116. doi:10.1007/BF02342617

Roux, G., Bertrand, R., Ghio, A., & Astésano, C. (2016). Naïve listeners' perception of prominence and boundary in French spontaneous speech. *Proceedings of Speech Prosody*, 8, 912–916. doi:10.21437/SpeechProsody.2016-187

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:10.1121/1.417955

Smith, C. (2013). French listeners' perceptions of prominence and phrasing are differentially affected by instruction set. *Proceedings of Meetings on Acoustics*, 19, 060191. doi:10.1121/1.4799041

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology*, 2, 35–98. doi:10.1515/labphon.2011.002

Streefkerk, B. M., Pols, L. C. W., & ten Bosch, L. F. M. (1997). Prominence in read aloud sentences, as marked by listeners and classified automatically. *Proceedings of the Institute of Phonetic Sciences*, 21, 101–116.

Styler, W. (2015). *On the acoustical and perceptual features of vowel nasality* (PhD thesis). University of Colorado at Boulder.

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, 101, 514–521. doi:10.1121/1.418114

Tamburini, F. (2005). Automatic prominence identification and prosodic typology. *Proceedings of Interspeech*, 2005, 1813–1816.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *The Journal of the Acoustical Society of America*, 89, 1768–1776. doi10.1121/1.401019

Terken, J. (1994). Fundamental frequency and perceived prominence of accented syllables. II. Nonfinal accents. *The Journal of the Acoustical Society of America*, 95, 3662–3665. doi:10.1121/1.409936

Turnbull, R., Burdin, R. S., Clopper, C. G., & Tonhauser, J. (2015). Contextual predictability and the prosodic realisation of focus: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, 30(9), 1061–1076. doi:10.1080/23273798.2015.1071856

Turnbull, R. (in press). The role of predictability in intonational variability. *Language and Speech*. doi:10.1177/0023830916647079

Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., … Vainio, M. (2015, August). *Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence*. Proceedings of ICPhS XVIII, Glasgow, Scotland.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, 49(3), 367–3392. doi:10.1177/00238309060490030301

You, H.-J. (2012). Determining prominence and prosodic boundaries in Korean by non-expert rapid prosodic transcription. *Proceedings of Speech Prosody*, 6, PS2B–PS16.