Laboratory Phonology
Journal of the Association for Laboratory Phonology

Ito, K et al 2017 Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 6, pp.1–29, DOI: https://doi.org/10.5334/labphon.86

JOURNAL ARTICLE

Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors

Kiwako Ito[1], Rory Turnbull[1,2] and Shari R. Speer[1]

[1] Ohio State University Dept of Linguistics, Columbus, Ohio, US

[2] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Ecole Normale Supérieure, PSL Research University, Paris, FR

Corresponding author: Kiwako Ito (ito.19@osu.edu)

A prominent pitch accent is known to trigger immediate contrastive interpretation of the accented referential expression. Previous experimental demonstrations of this effect, where [L+H* unaccented] contours led to an increase in earlier responses than [H* !H*] contours in contrastive context, may have benefited from the use of laboratory speech with stylized, homogenous pitch contours as well as data collected from a uniform participant group—college students. The present study tested visitors to a science museum, who better represent the general public, comparing lab and spontaneous speech to replicate the contrast-evoking effect of prominent pitch accent. Across two eye-tracking experiments where participants followed spoken instructions to decorate Christmas trees, spontaneous two-word [L+H* unaccented] contours led to faster eye-movements to contrastive ornament sets than [H* !H*] contours with no delay as compared to lab speech. The differences in the fixation functions were overall smaller than those in a previous study that used clear lab speech in richer contexts. Detailed acoustic analyses indicated that the lab speech tune types were distinguishable by any of several independent F0 measures on the adjective and by F0 slope. In contrast, no single phonetic measure on the spontaneous speech adjective distinguished between tune types, which were best classified according to independent noun-based measures. However, a non-linear combination of the adjective measures was shown to be equal to the noun measures in distinguishing between the [H* !H*] and [L+H* unaccented] tunes. The eye-movement data suggest that naïve listeners were comparably sensitive to both lab and spontaneous prosodic cues on the adjective and made anticipatory eye-movements accordingly.

Keywords: online prosodic processing; eye-tracking; spontaneous speech stimuli

1. Introduction

1.1. Invariance in lab-based studies: Stimuli and participants

Laboratory investigation of speech perception or comprehension is often bound to two types of invariance: One comes by choice and another as a practical constraint. The former is the set of stimuli researchers use: Under a traditional factorial design, researchers typically manipulate some aspect of the speech signal and make sure that their stimuli exhibit intended difference(s) across the levels of the manipulated factor. In practice, stimuli are carefully handpicked such that all items in an experimental condition conform to the intended sound pattern and differ from the items of a comparison condition in a consistent manner. Thus, any given item from a particular condition should sound equally different from any given item from the comparison condition. Since the effect of the manipulation should be attributed only to the manipulated factor (in principle), researchers pay close attention to the potential confounding factors and often try to demonstrate that their stimuli bear a statistically valid variation only along the targeted acoustic dimension and not along the others. In order to clear this non-trivial requirement, researchers often ask highly trained laboratory personnel to produce the stimuli (or synthesize/artificially modulate particular acoustic parameters of natural speech) for their experiments.

The second type of invariance results from conventionalized participant recruitment. Many experimental behavioral studies, including those that investigate speech comprehension, collect data from students of the institutes that the researchers are affiliated with because that is the most convenient way to recruit participants. Students can make a great participant group because they come from a narrow age range, are assumed to have similar educational and socio-economical background, have an easy access to the lab, do not need to be instructed on how to use a computer mouse, and do not generally exhibit much anxiety for participating in a study (as they get multiple opportunities to do so for course credit). These factors are often taken for granted while researchers assume that their rather homogeneous participant sample represents a much more general population such as 'native speakers of language X.'

These conventions for experiment designing and participant recruitment are so commonly practiced that they almost serve as the field's norms. However, in terms of scientific validity, such methodological routines come at a price. Researchers of speech processing who practice these conventions are faced with questions such as *Did we obtain the effect because of the extreme clarity and predictability of lab speech? Can we generalize our results to real world speech processing? Do most people really make use of the acoustic cues we manipulated in their daily conversations?* Addressing these questions is important especially for the investigation of prosody, an aspect of spoken language known to be highly variable within and across productions by individual speakers and comprehension by individual listeners (e.g., Speer et al, 2011; Speer & Foltz, 2015). The current study aims to make a small step toward the generalization of laboratory findings about the interpretation of prosodic prominence to the every-day communicative function of interpreting contrast. In particular, we try to achieve this goal by introducing two types of variability: One in stimuli for an eye-tracking experiment, and another in the participants, recruited at a local science museum.[1] With non-laboratory speech and non-institution-bound participants, we test whether spontaneously produced intonation contours lead to contrastive interpretation of speech in participants who may better represent the general public than a pool of college students from a particular university.

1.2. Contrastive prosody: categorical distinction despite variant phonetic cues

The phenomenon that the present study investigates is the processing of prosodic prominence that expresses contrast in American English. The present study follows the terminology of the autosegmental metrical theory of prosody (e.g., Bruce, 1977; Goldsmith, 1976, 1990; Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Ladd, 1996, 2008;

[1] Generally, experiments in a museum must be brief and cannot include detailed questionnaires and additional behavioral assessments. Thus, we cannot test hypotheses such as how participants' general cognitive function and verbal skills are related to their prosodic processing abilities (e.g., Stojanovik, 2010; Diehl & Paul, 2013) or how their socioeconomic status affects prosodic processing in the current study. It is left to future studies to explore these important factors that lead to variability in responses to prosodic cues in speech.

Gussenhoven, 2004) and the prosodic annotation system ToBI (Tones and Break Indices: Beckman & Ayers, 1997; Beckman et al., 2005), which distinguishes *pitch accents,* the local pitch excursions that lead to the perception of prominence, from *phrasal* or *boundary tones* that mark the edges of prosodic phrases. The inventories of pitch accents and boundary tones are known to be highly language specific (see Jun, 2005, 2014 for the intonational cross-linguistic typology). As for American English, pitch accents (as well as boundary tones) are assumed to be functionally distinct prosodic categories. Since Pierrehumbert and Hirschberg's description of pragmatic meaning of pitch accents (1990), it has been widely assumed that pitch accents are prosodic morphemes, which bear communicative function for expressing informational status of words or phrases within a mutual belief space. The important implication of this view is that the ultimate pragmatic interpretation of a given intonation pattern is computed by combining the function of all pitch accents and phrasal tones that participate in the contour. While Pierrehumbert and Hirschberg (and developers of ToBI annotation systems) have explicitly stated that any given pitch accent can convey multiple meanings and there is no one-to-one mapping between tune and meaning, the general assumption is that the intonational meaning of an utterance can be derived from basic pragmatic functions unique to local prosodic events, such as pitch accents.

While the field seems to have achieved a gross agreement as to what phonological elements (such as pitch accent and phrasal/boundary tones) participate in an intonation contour of an utterance (e.g., English pitch accents align to stressed syllables while edge tones may be realized across multiple segments that occupy the space between the accented syllable and the end of the phrase), researchers do not always agree on how listeners achieve categorical distinctions among particular sets of pitch accents (cf. Braun & Tagliapietra, 2010; Calhoun, 2012; Krahmer & Swerts, 2001; Watson et al., 2008). For example, Ladd and Morton (1997) showed that listeners could assign degree of emphasis in a gradient manner consistent with a continuum of F0 excursions that varied in the peak height between two rise-fall contours, which respectively represented a canonical H* accent and a L+H* accent. However, when listeners were asked to make a choice between 'everyday event' and 'unusual event' for each F0 contour, their responses showed a sharp switch from one interpretation to another in the middle of the same continuum. Thus, results from Ladd and Morton suggest that gradient levels of prominence among similar pitch contours are perceivable yet they may fall into two interpretational categories. According to Ladd and Morton, the categorical distinction between the two types of pitch accents is made not at the phonetic level, but at the level where an interpretation is assigned to the tune.

Results of recent research suggest that listeners are sensitive to the shape of F0 curvatures and their alignment to aspects of the stress-carrying syllable, and they may make use of these cues together with segmental makeup of the utterance for distinguishing pitch accents. A recent study by Barnes et al. (2015) showed that the shape of a rise toward the F0 peak and the shape of a fall thereafter interact with the peak alignment for the perception of accent categories. For example, when the F0 peak is aligned within the stressed syllable onset and nucleus, a domed rise or a scooped fall is perceived more often as H+!H* whereas a scooped rise or a domed fall is perceived more often as L+H*. Barnes et al. also report that the sonority of tone-hosting segments and F0 curvature of rise and fall affect the categorical distinction that has been annotated as an alignment difference (e.g., L+H* vs. L*+H). While the distributions of the domed and scooped rises and falls in spontaneous speech are yet to be catalogued, Barnes et al. clearly demonstrate that the processing of prosodic morphemes may not rely solely on the detection of point targets such as F0 minima and maxima, and propose Tonal Center of Gravity (TCoG) as a model of F0 contour perception.

1.3. Prominence perception of spontaneous speech: Turnbull et al. (2014, 2017)

While the findings of Ladd and Morton (1997) and Barnes et al. (2015) together suggest that phonetic variability in F0 height and contour shape can be perceived (or interpreted) categorically, another recent study by Turnbull et al. (2014, 2017) adds to the complexity of prominence perception by showing that both the listener's belief about the discourse status of stimuli and the F0 level of an adjacent word may matter for the assignment of prominence. Turnbull et al. adapted the prominence-marking task of Cole et al. (2010) where listeners hear auditory stimuli and mark the words that sound prominent on a transcript. The stimuli of Turnbull et al. were taken from the spontaneous speech production corpus of Ito and Speer (2006), where naïve participants were asked to give instructions on how to decorate Christmas trees according to the visual slides that specified the ornament and its location on the tree (see **Figure 1**). All recordings of target noun phrases (e.g., brown drum) had been ToBI-annotated by blind labelers in this corpus.

Using the transcribed spontaneous speech produced by a female undergraduate student, Turnbull et al. created four types of noun phrase pairs by crossing the two sequence types (contrastive and non-contrastive) with the two pitch contour types ([L + H* unaccented] and [H* !H*]). In the contrastive sequences, the noun was repeated across the two phrases, leading to contrast on the color adjective (e.g., blue drum, green drum). Non-contrastive sequences did not have any repetition (e.g., grey house, brown drum). The second 'adjective + noun' phrase had been annotated as either [L + H* unaccented] or [H* !H*]. Turnbull et al. found that the adjectives produced in the [L + H* unaccented] contours were more likely to be marked as prominent than the adjectives produced in the [H* !H*] contours, irrespective of the sequence type. Interestingly, the sequence type (i.e., whether the adjective was contrastive or not) did not affect the prominence rating, *unless* listeners were explicitly told that stimuli sequences with an intervening confederate's speech (e.g., "Blue drum." "O.K. Next?" "Green drum.") were extracted from a dialogue that actually took place between the two speakers. When the same Speaker A–Speaker B–Speaker A sequences were presented without the background information about the discourse context (i.e., the nature of the conversation from which the utterances were extracted), the repetition of the noun (e.g., drum in the above example) did not give rise to higher prominence rating on the second adjective (e.g., green). Turnbull et al.'s findings therefore suggest that the mere sequence of phrases would not evoke the notion of contrast in listeners when the communicative purposes of the utterances are unclear to them.

As for the effect of pitch contour type, Turnbull et al. confirmed that the distinction between [L + H* unaccented] and [H* !H*] that ToBI annotators had made had some perceptual bearing in naïve listeners' speech processing. To find out what acoustic features most reliably distinguished the two contours, Turnbull et al. entered a total of 14 acoustic

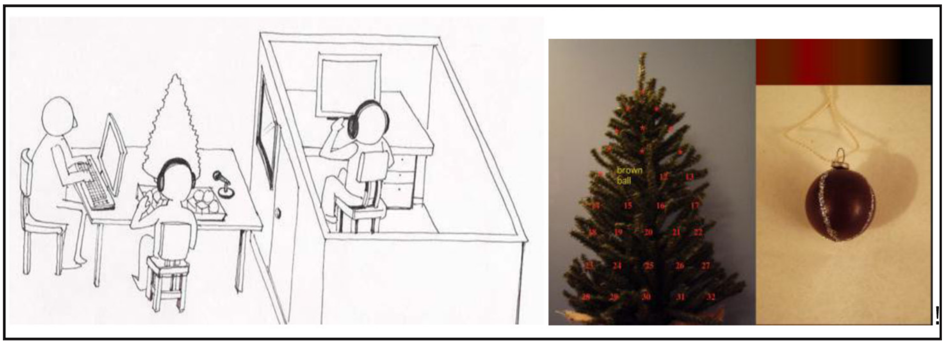

Figure 1: Spontaneous speech elicitation (Ito & Speer, 2006).

measures (each word's duration, each vowel's duration, each word's F0 peak, the mean F0 of the vowel, two measures of spectral tilt,[2] the F0 slope from the adjective's peak to the noun's peak, and the F0 slope from the adjective vowel mean to the noun vowel mean) into a classification tree analysis (Therneau & Atkinson, 1997). Surprisingly, it was neither the F0 values of the adjectives nor the F0 slopes, but was the *noun's* F0 peak that best differentiated the contour types. When the noun's F0 peak replaced the contour type as a continuous predictor factor in the logistic mixed effects model that tested the likelihood of prominence assignment, it showed a reliable effect (i.e., the lower the noun F0 peak was, the more likely the preceding adjective was marked as prominent).

Turnbull et al.'s findings are not completely in line with the past prominence perception studies such as Gussenhoven et al.'s (1997), which reports that the prominence rating of earlier word correlates with its F0 peak while it is not affected much by the contour-offset F0 value. Given the similarity in the task between Turnbull et al. and Gussenhoven et al. (where participants reported which word in each utterance sounded emphasized), the discrepancy in their findings may have resulted from the difference in the nature of stimuli (natural, spontaneous speech in Turnbull et al. vs. synthesized F0 contours in Gussenhoven et al.) and also from the structure of stimuli (two consecutive phrases in Turnbull et al. vs. isolated sentences in Gussenhoven et al.). Crucially, Turnbull et al. report that the absolute values of adjective's F0 peak were not consistently higher for [L+H* unaccented] than [H* !H*], indicating the possibility that the two contours were not readily distinguishable up to the noun's F0 peak. That is, in spontaneous speech, distinctive phonetic differences between the contours that could be annotated as [L+H* unaccented] and [H* !H*] may not be observed within the words that are labeled with different pitch accents (e.g., adjective with L+H* vs adjective with H*). In a nutshell, Turnbull et al.'s findings suggest that the perception of prominence of a phrase-initial word may be affected largely by the level of tonal and vowel reduction of the following word, whereas the contextual factor such as phrase sequence may trigger contrastive interpretation only when listeners process the sequence as a meaningful discourse rather than isolated word pairs.

1.4. Testing the online response to spontaneous prosody: An eye-tracking paradigm

The present study makes use of the spontaneous speech from Turnbull et al. (2014, 2017) to compare the responses to [L+H* unaccented] and [H* !H*] patterns between spontaneous and typical laboratory speech. The two types of speech were used as instructions for a Christmas tree decoration task, adapted from a previous eye-tracking study by Ito and Speer (2008). In Ito and Speer, which used typical lab speech, a large F0 excursion over the adjective and following pitch range compression over the noun led to faster looks to the target ornament set in a contrastive sequence such as "Hang a blue drum." "Now, to its right, hang a GREEN$_{\mathrm{L+H*}}$ drum$_{\mathrm{unacc}}$." This [L+H* unaccented] pattern induced incorrect fixations to the just-mentioned ornament set in non-contrastive sequences such as "Hang a blue drum." "Now hang a GREEN$_{\mathrm{L+H*}}$ onion$_{\mathrm{unacc}}$." (i.e., upon hearing GREEN, participants often looked at drums first). Because the looks to the preceding ornament set increased from the end of the prominent color adjective to the beginning of the noun, Ito and Speer argued that the prosodic prominence of the adjective is processed immediately and can lead to anticipatory eye-movements. The present study tests whether spontaneous [L+H* unaccented] sequences, which do *not* show an overt F0 excursion for L+H* like the stimuli in Ito and Speer (2008), can lead to similar anticipatory responses.

[2] The spectral tilt measures were the difference between the mean intensity of two different spectral bands, either 2kHz in bandwidth (i.e., 0–2kHz minus 2–4kHz) or 4kHz in bandwidth (i.e., 0–4kHz minus 4–8kHz).

While the procedure of the present study is overall very similar to that of Ito and Speer (2008) (i.e., participants sat in front of a real-world board with an array of ornament sets and decorated a tree following the pre-recorded instructions), the present study differs largely from Ito and Speer (2008) in the degree of interactivity. In Ito and Speer, the pre-recorded instructions included many conversational utterances for filler trials such as "So to its left, hang a red onion, … it looks like jewelry." According to the authors, such extra speech was included to blend the strictly controlled critical prosodic contours into a naturalistic interactive set of instructions. In the present study, we used only isolated bare noun phrases such as 'grey house' and 'brown drum,' without any surrounding context. This was due to the fact that many of the phrases used by Turnbull et al. were produced spontaneously as isolated utterances, and because we decided to test whether the anticipatory effect reported by Ito and Speer can be replicated even with minimally engaging speech input. If the anticipatory effect is confirmed with the spontaneous speech that does not have either the distinct F0 excursion like lab speech or conversational context phrases, such effect may serve as a stronger evidence for the listeners' sensitivity to subtle prosodic cues to prominence that can be interpreted contrastively. Alternatively, the mere sequential presentation of noun phrases may not replicate the anticipatory effect if, as Turnbull et al. suggest contrastive interpretation of a phrase is achieved only in a coherent communicative discourse context.

2. Experiment 1

2.1. Participants

A total of 160 participants (age range: 18–65, average 28;04) were recruited at a local science museum. These participants were recruited by four undergraduate assistants, who were trained to interact with the museum visitors as part of an outreach program of a language science laboratory. Participants volunteered to take part in the experiment and did not receive any incentives. A total of 79 participants were assigned to the list that presented lab speech as stimuli, while a total of 81 participants were assigned to the list with spontaneous speech stimuli such that the two groups' age ranges roughly match.

2.2. Materials and design

A subset of the stimuli from Turnbull et al. was combined to create a total of 24 critical sequences across four conditions (2 contour types x 2 sequence types) for each type of speech: Lab and spontaneous (See **Table 1** for the design). Due to the restricted availability of phrases from the spontaneous speech corpus (Ito & Speer, 2006; Turnbull et al., 2014, 2017), the words participating in the critical phrase sequences could not be matched between the [L+H* unaccented] and the [H* !H*] sets. However, the entire sequence of phrases experienced within an experiment was identical between the lab speech and spontaneous speech sessions (See Appendix A for the entire sequence of presentation): After the entire phrase sequence was constructed with the spontaneous utterances, a trained female phonetician, who often has served as a trained speaker for experiments in our laboratory, recorded the same sequence, with canonical [L+H* unaccented] and [H* !H*]. The phonetician's productions were prompted by text and ToBI annotations; she was not presented with examples from the spontaneous talker.

	Contrastive	**Non-Contrastive**
[L+H* unaccented]	blue drum, **GREEN drum**	clear house, **BROWN ball**
[H* !H*]	grey house, **orange house**	green bell, **navy onion**

Table 1: Four conditions (2 contour types for the **target phrase** x 2 sequence types). The phrase preceding the critical phrase always had a [H !H*] contour.

The acoustic analysis of the critical stimuli is summarized in **Table 2**. While most of the measures are self-explanatory, some require further explanation. F0 peak alignment was defined as the time of the F0 peak relative to the onset of the vowel. Spectral tilt refers to the difference in dB between the lower and upper halves of the (low-pass filtered) spectrum; for the 2 kHz bandwidth the measure is thus a comparison of the mean dB of the 0–2 kHz bin with that of the 2–4 kHz bin, and for the 4 kHz bandwidth the measure compares the 0–4 kHz bin to the 4–8 kHz bin. The peak-to-peak and vowel-to-vowel F0 slopes are taken from Turnbull et al. (2017), described above in section 1.3.

Following Turnbull et al. (2014, 2017), a classification tree analysis was carried out to determine which acoustic cues could reliably linearly discriminate between the two tune types, [H* !H*] and [L+H* unaccented]. Due to the small sample size (only 14 [H* !H*] tokens and 8 [L+H* unaccented] tokens for each talker), there was little chance of finding significant differences between mean values (A power analysis shows that for a sample of this size and an alpha value of .05, the power of finding an effect size of $d=1$ is only .574. For reference, note that Cohen regarded an effect size of $d=0.8$ to be 'large.'). We know from phonetic studies of segmental phonetic features (e.g., the beat~bit contrast) that overlap and variability is the norm. Surprisingly, then, the tune type distinction was perfectly linearly separable in the lab speech. The four variables—adjective peak F0, adjective vowel mean F0, vowel-to-vowel mean F0 slope, and peak-to-peak F0 slope—were individually and independently able to provide 100% classification accuracy.

While the spontaneous speech was more variable than the lab speech, above-chance prediction of category from acoustic features was still observed. The best classifier in this case was the F0 peak of the noun, which correctly classifies 19 out of the 22 tokens (86.3%). Other effective measures are the mean-to-mean F0 slope (17 correct, 77.3%) and the adjective 2 kHz bandwidth spectral balance (16 correct, 72.7%). As the best classifier is a noun-based measure, it could be argued that the brunt of the contrast is being carried on the noun, while the L+H* vs H* distinction, which is carried on the adjective, is worth little. However, this is not the case: A support vector machine[3] was constructed to predict category using acoustic features from the adjective alone (that is, adjective F0 peak height, adjective vowel mean F0 height, adjective word duration, adjective vowel duration, adjective excursion size, adjective spectral tilt [2 kHz bandwidth], and adjective spectral tilt [4 kHz bandwidth]). This support vector machine was tuned using 10-fold cross-validation to prevent overfitting. The resulting model correctly predicted the category of 19 (86.3%) out of the 22 tokens, performance equal to that of the noun peak F0 measure. Informal inspection of the model suggested that of the 7 input variables, the adjective peak F0 and both spectral tilt measures were particularly important to ensuring correct predictions. This finding suggests that while none of the acoustic variables that were measured provide a perfect linear separation, non-linear transformations of the variables are able to predict the contour categories at levels significantly greater than chance.

Taken together, then, despite the small sample size, despite the phonetic overlap common between phonological categories (especially prosodic categories), and despite the intrinsic variability inherent to spontaneous and natural speech, all the evidence suggests that these two categories are acoustically distinct, and that furthermore, the adjectives in particular are acoustically distinct in meaningful ways.

[3] Support vector machines are machine learning models which convert variables into high-dimensional vector spaces and find a best-fitting hyperplane(s) to divide the space into predicted categories.

	Lab Speech				Spontaneous Speech			
	Adjective		Noun		Adjective		Noun	
	[L+H* unacc]	[H* !H*]	[L+H* unacc]	[H* !H*]	[L+H* unacc]	[H* !H*]	[L+H* unacc]	[H* !H*]
Peak F0 (Hz)	320.94 (17.17)	212.82 (6.87)	172.57 (13.84)	198.15 (9.89)	267.8 (18.25)	275.15 (14.33)	203.98 (35.42)	233.62 (15.37)
Vowel mean F0 (Hz)	277.83 (10.86)	198.42 (8.75)	152.25 (8.1)	171.86 (7.14)	257.14 (15.85)	258.7 (14.27)	182.57 (25.54)	207.97 (21.05)
Excursion size (Hz)	90.30 (34.26)	25.81 (8.37)	– –	– –	37.25 (14.52)	38.42 (11.34)	– –	– –
Word Duration (ms)	403.43 (98.58)	358.58 (45.31)	476.81 (126.68)	508.08 (96.21)	357.99 (69.47)	325.32 (93.49)	464.59 (65.19)	488.14 (121.89)
Vowel Duration (ms)	228.64 (31.75)	211.28 (36.67)	193.36 (35.97)	211.7 (39.86)	152.79 (54.2)	148.33 (38.39)	148.54 (19.36)	157.62 (40.17)
F0 peak align (ms)	181.3 (46)	181.65 (88.84)	21.43 (20.91)	24.18 (17.72)	105.91 (50.65)	126.46 (76.95)	26.76 (79.21)	20.62 (70.08)
Word duration (proportion of total phrase duration)	0.46 (0.12)	0.42 (0.07)	0.54 (0.12)	0.58 (0.07)	0.43 (0.07)	0.4 (0.1)	0.57 (0.07)	0.6 (0.1)
Vowel duration (proportion of noun+adj vowel duration)	0.6 (0.17)	0.59 (0.09)	0.45 (0.19)	0.43 (0.13)	0.43 (0.14)	0.47 (0.12)	0.32 (0.06)	0.35 (0.17)
Spectral tilt, 2 kHz bandwidth (dB)	15.33 (3)	14.61 (4.8)	20.96 (6.09)	18.01 (5.52)	7.74 (2.36)	12.2 (7.47)	10.3 (6.06)	13.14 (6.1)
Spectral tilt, 4 kHz bandwidth (dB)	24.91 (2.44)	28.34 (2.75)	28.87 (5.29)	27.22 (4.15)	17.86 (7.67)	21.65 (4.96)	20.35 (7.18)	23.28 (6.11)

F0 slopes	Lab Speech		Spontaneous Speech	
	[L+H* deacc]	[H* !H*]	[L+H* deacc]	[H* !H*]
Peak-to-peak	–92.45 (75.33)	–92.45 (75.33)	–269.3 (131.99)	–202.59 (91.94)
Vowel-to-vowel	–360.05 (102.26)	–79.39 (31.18)	–252.27 (91.86)	–164.24 (83.5)

Table 2: Acoustic analysis of stimuli.
All the numbers in this table show the average value for the given measure. Values in parentheses are standard deviations.

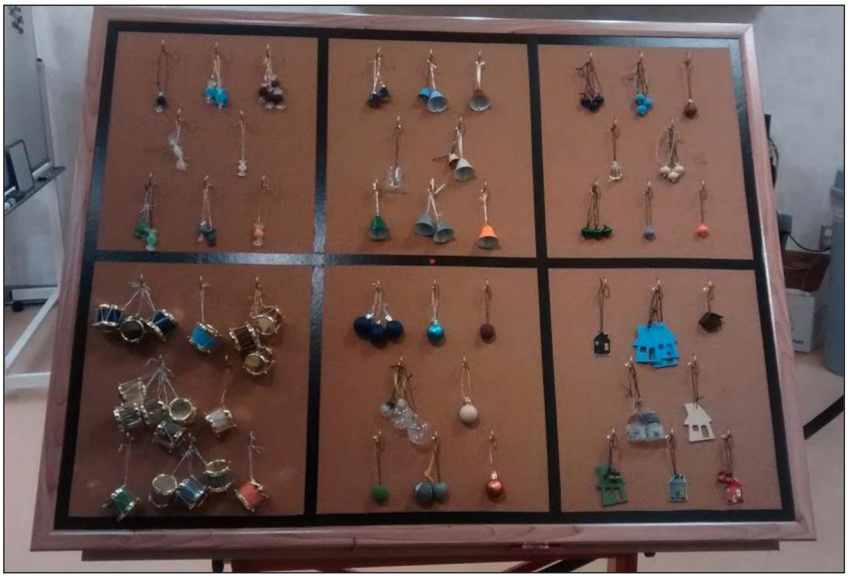

Figure 2: Experiment 1 and 2 ornament set.

2.3. Procedure

The experiment took place in a laboratory space located in the science museum (Center of Science and Industry: http://bln.osu.edu/LanguagePod.php). Each participant was seated in front of a 36 x 48 inch (approximately 91 x 122cm) corkboard on which a total of 93 ornaments were sorted into six 15.6 inch by 13.8 inch (approximately 40 x 35cm) cells by object type (candy, bell, onion, drum, ball and house: See **Figure 2**). The critical 24 sequences were embedded in a larger sequence with 13 filler utterances that together mentioned a total of 37 ornaments.

Participants were asked to wear a head-mounted ASL EyeVision XG 6.0.7.3 system. After a quick calibration, participants were told that they would hear names of ornaments one by one, and after hearing each name, they should locate and choose the corresponding ornament from the board and place it on the artificial tree that was next to the board. Participants were asked to face back to the board and look at a small red star on the center of the black border between the two central cells after placing each ornament on the tree to indicate they were ready for the next trial. After the experimenter confirmed that the participant understood the task, headphones (Bose Quiet Comfort 15 Acoustic Noise Cancelling Headphones) were placed over the participant's headgear and the experiment started. Each session was completed within 15 minutes (average duration of the task: 7min 45 sec). Participants' eye movements were recorded at 30Hz throughout the tree decoration task.

2.4. Results: Experiment 1

Data from twenty-eight participants were excluded from the analysis due to frequent track loss (15), calibration problems (4), other system failure (3), and other reasons such as a participant being a non-native speaker of American English, color blind, under age, etc. (4), leaving 66 participants in each speech type group (lab speech: Age 18–63, average 27;06, spontaneous speech: Age 18–65, average 29;01). The data in the following figures are aligned from the noun onset. The average duration of the noun in the two contour types are indicated with vertical lines, with a red line for [H* !H*] and a blue line for [L + H* unaccented].

The data from the two speech types showed the expected faster looks to the target ornament set for [L + H* unaccented] than for [H* !H*] contour type for the Contrastive sequences. **Figure 3** shows the proportion of fixations to the target ornament cell in the Contrastive sequences for the two groups. In both groups, the fixation proportions started increasing about 200ms after the noun onset, and reached ceiling level by around 1400–1600ms post noun onset. Importantly, the two groups showed faster looks to the target cell when the phrase with a repeated noun had the [L+H* unaccented] contour (e.g., blue drum, GREEN drum) than when it had the [H* !H*] contour (green drum).

For the statistical testing, participants' fixation proportion functions were submitted to growth curve analyses (GCA), which has been adapted to eye-tracking data analysis by Mirman and colleagues (Mirman et al., 2008; Mirman, 2014). Each participant's average fixation proportion across trials of a particular condition was calculated for each time point with 20ms intervals. The changes in the fixation proportions as a function of time (80 time bins for 0–1600ms from the onset of the noun) were submitted to a mixed effects model that included orthogonal polynomials of time bins as predictive factors together with sum-coded pitch accent type (PA: L+H* unaccented vs. H* !H*) and sum-coded Speech Type (lab vs. spontaneous). The models included participant random effects on polynomial terms and participant-by-PA random effects on the first two polynomial terms (higher order polynomials were excluded from the random effect calculation as

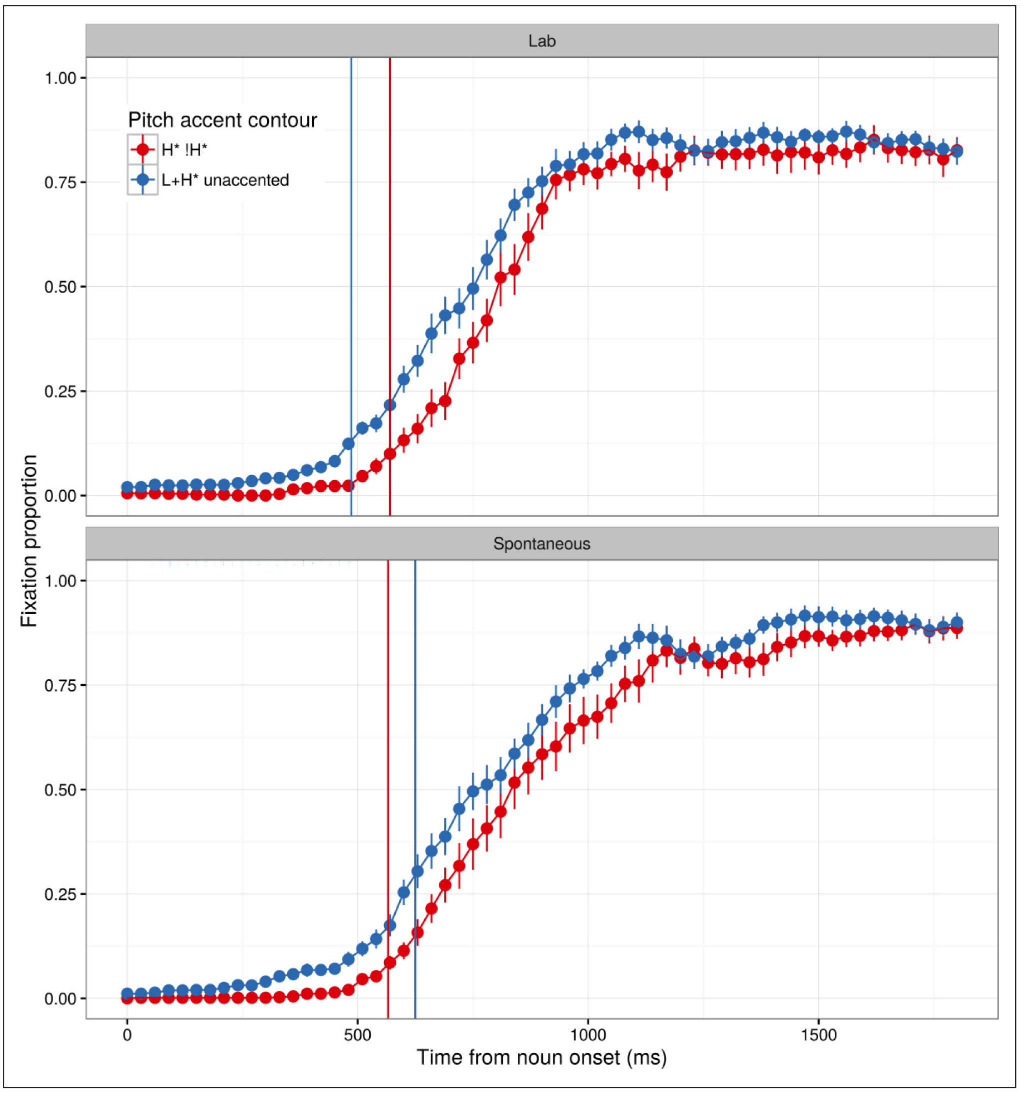

Figure 3: Experiment 1: Fixations to the target ornament cell in Contrastive sequences for Lab and Spontaneous speech.

they increase the computational cost and would capture less relevant variation in the tails (see Mirman, 2014, for details). Although the fixation proportions for the target grossly showed typical sigmoid functions in **Figure 3**, a model comparison revealed that a model that includes up to the quartic polynomial term better fit the data than a model with up to the cubic term. Thus, the full models used in the present study included up to the fourth order orthogonal polynomials (ot1–4) for both the target and the competitor analyses unless specified otherwise.

The output of a model for the target fixations in the Contrastive sequences is shown in **Table 3**. There was a main effect of pitch accent type (PA), which indicates the overall higher fixations to the target with [L+H* unaccented] than with [H* !H*]. The pitch accent type interacted with the second, third, and fourth order polynomials, showing that the difference in the prosodic contour led to changes in quadratic, cubic and quartic components of the functions. There was no main effect of Speech Type, i.e., the lab speech and spontaneous speech groups had overall similar amounts of fixations to the target. The lack of interaction between pitch accent type (PA) and Speech Type indicates that

Fixed effects:	**Estimate**	***SE***	***df***	***t* value**	**Pr(>|t|)**
(Intercept)	4.585e–01	9.729e–03	1.330e+02	47.127	< 2e–16 ***
ot1	2.473e+00	4.424e–02	1.370e+02	55.902	< 2e–16 ***
ot2	–1.873e–01	4.550e–02	1.360e+02	–4.117	6.62e–05 ***
ot3	–6.367e–01	2.936e–02	1.320e+02	–21.687	< 2e–16 ***
ot4	1.587e–01	3.148e–02	1.320e+02	5.042	1.49e–06 ***
PA	6.646e–02	1.043e–02	1.390e+02	6.374	2.52e–09 ***
SpeechType	7.309e–03	1.946e–02	1.330e+02	0.376	0.7078
ot1:PA	–4.918e–02	5.636e–02	1.650e+02	–0.873	0.3842
ot2:PA	–2.965e–01	4.829e–02	2.210e+02	–6.139	3.80e–09 ***
ot3:PA	8.666e–02	1.937e–02	1.315e+04	4.473	7.78e–06 ***
ot4:PA	1.486e–01	1.936e–02	1.314e+04	7.676	1.75e–14 ***
ot1:SpeechType	1.149e–01	8.848e–02	1.370e+02	1.298	0.1964
ot2:SpeechType	6.027e–02	9.099e–02	1.360e+02	0.662	0.5089
ot3:SpeechType	7.615e–03	5.872e–02	1.320e+02	0.130	0.8970
ot4:SpeechType	–6.707e–03	6.295e–02	1.320e+02	–0.107	0.9153
PA:SpeechType	–4.284e–03	2.085e–02	1.390e+02	–0.205	0.8375
ot1:PA:SpeechType	1.058e–01	1.127e–01	1.650e+02	0.939	0.3492
ot2:PA:SpeechType	1.115e–01	9.659e–02	2.210e+02	1.155	0.2495
ot3:PA:SpeechType	7.418e–03	3.875e–02	1.315e+04	0.191	0.8482
ot4:PA:SpeechType	–7.625e–02	3.872e–02	1.314e+04	–1.969	0.0489 *

Table 3: Experiment 1: Output of a mixed effect model for the looks to the target in Contrastive sequences.
Model Structure: meanFixation ~ (ot1 + ot2 + ot3 + ot4) * PA * SpeechType + (1 + ot1 + ot2 + ot3 + ot4 | File) + (1 + ot1 + ot2 | File:PA).
Number of obs: 14148, groups:File:PA, 264; File, 132.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

the facilitative effect of L + H* did not differ between the lab and the spontaneous speech groups. A significant three-way interaction between the fourth order polynomial, (ot4) pitch accent type (PA) and Speech Type indicates that the difference in the pitch contour affected the quartic component of the functions differently between the lab and spontaneous speech groups. We suspect that this effect comes from the difference in the shape of functions for [L + H* unaccented] condition for the later region of 1000–1600ms after the noun onset in **Figure 3**.

For the Non-Contrastive sequences (e.g., 'clear house,' 'brown ball'), the two groups showed the expected higher looks to the preceding ornament set (e.g., 'house') when the target phrase had the [L + H* unacc] contour ('BROWN ball') than when it had the [H* !H*] contour ('brown ball'), although the magnitude of this trend was very small. **Figure 4** shows that the looks to the competitor (i.e., the preceding ornament) started increasing during the noun in both groups. Unlike Ito and Speer (2008), however, the incorrect looks to the previous ornament set did not delay the looks to the correct target. Instead, the looks to the correct target reached a ceiling level relatively more slowly with [H* !H*] than with [L + H* unaccented] contours.

For the statistical analysis of data from Non-Contrastive sequences, both the fixation proportions for the target and those for the competitor were calculated for the 80 time

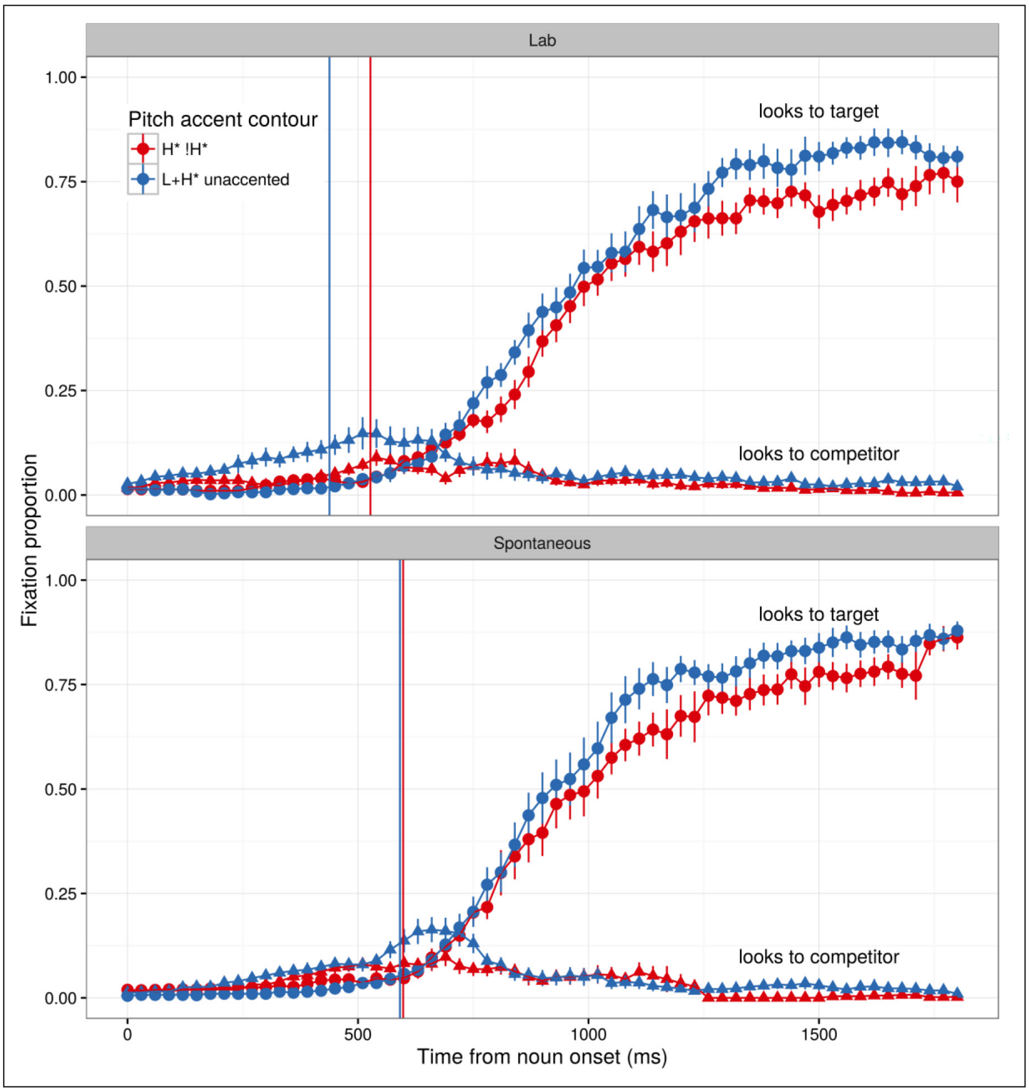

Figure 4: Experiment 1: Fixations to the target and competitor cells in Non-Contrastive sequences.

bins, and they were submitted to the mixed effects models that contained up to the fourth order polynomials as the predictor factors (see **Table 4** for the target, and **Table 5** for the competitor analyses).

The results of the model showed a main effect of PA and a main effect of Speech Type on the looks to the target in the Non-Contrastive sequences. These indicate that the [L + H* unaccented] contour led to more looks to the target than the [H* !H*] contour, and the spontaneous speech group showed overall more looks to the target than the lab speech group. The lack of interaction between PA and Speech Type suggests that the overall effect of PA did not differ between the two groups. The two-way interactions between the polynomial terms and PA suggest that the pitch contour difference affected linear, cubic, and quartic components of functions. The two-way interaction between ot1 and Speech Type suggests that the linear component of the function was steeper for the spontaneous speech group. A three-way interaction (with the negative estimate) between the third order polynomial term, PA, and Speech Type indicates that the effect of pitch contour type on the cubic component of the function was smaller for the spontaneous group.

Fixed effects:	**Estimate**	***SE***	***df***	***t* value**	**Pr(>|t|)**
(Intercept)	3.456e–01	9.258e–03	1.340e + 02	37.329	< 2e–16 ***
ot1	2.177e + 00	4.953e–02	1.360e + 02	43.960	< 2e–16 ***
ot2	2.733e–01	3.917e–02	1.380e + 02	6.976	1.16e–10 ***
ot3	–5.063e–01	3.181e–02	1.320e + 02	–15.917	< 2e–16 ***
ot4	–9.713e–02	2.418e–02	1.320e + 02	–4.018	9.81e–05 ***
PA	3.254e–02	1.008e–02	1.590e + 02	3.227	0.00152 **
SpeechType	4.243e–02	1.852e–02	1.340e + 02	2.291	0.02350 *
ot1:PA	2.657e–01	6.503e–02	1.560e + 02	4.086	6.99e–05 ***
ot2:PA	–2.260e–02	4.453e–02	2.220e + 02	–0.507	0.61232
ot3:PA	–8.753e–02	1.906e–02	1.315e + 04	–4.593	4.42e–06 ***
ot4:PA	5.520e–02	1.905e–02	1.314e + 04	2.897	0.00377 **
ot1:SpeechType	1.994e–01	9.906e–02	1.360e + 02	2.013	0.04607 *
ot2:SpeechType	–3.121e–02	7.834e–02	1.380e + 02	–0.398	0.69097
ot3:SpeechType	–5.376e–02	6.362e–02	1.320e + 02	–0.845	0.39959
ot4:SpeechType	3.141e–02	4.835e–02	1.320e + 02	0.650	0.51714
PA:SpeechType	1.922e–02	2.017e–02	1.590e + 02	0.953	0.34203
ot1:PA:SpeechType	1.624e–02	1.301e–01	1.560e + 02	0.125	0.90077
ot2:PA:SpeechType	–1.492e–01	8.905e–02	2.220e + 02	–1.676	0.09520
ot3:PA:SpeechType	–7.586e–02	3.812e–02	1.315e + 04	–1.990	0.04659 *
ot4:PA:SpeechType	3.532e–02	3.810e–02	1.314e + 04	0.927	0.35403

Table 4: Experiment 1: Output of a mixed effect model for the looks to the target in Non-Contrastive sequences.

Model Structure: meanFixation ~ (ot1 + ot2 + ot3 + ot4) * PA * SpeechType + (1 + ot1 + ot2 + ot3 + ot4 | File) + (1 + ot1 + ot2 | File:PA).

Number of obs: 14135, groups: File:PA, 264; File, 132.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’1.

Fixed effects:	Estimate	*SE*	*df*	*t* value	Pr(>|t|)
(Intercept)	5.367e–02	3.896e–03	1.320e + 02	13.777	< 2e–16 ***
ot1	–7.036e–02	1.996e–02	1.530e + 02	–3.525	0.000559 ***
ot2	–1.663e–01	1.909e–02	1.460e + 02	–8.714	5.77e–15 ***
ot3	8.773e–02	1.634e–02	1.320e + 02	5.368	3.47e–07 ***
ot4	4.766e–02	1.392e–02	1.320e + 02	3.424	0.000823 ***
PA	2.719e–02	5.405e–03	1.320e + 02	5.031	1.56e–06 ***
SpeechType	–2.377e–03	7.791e–03	1.320e + 02	–0.305	0.760734
ot1:PA	–2.984e–02	2.949e–02	2.440e + 02	–1.012	0.312737
ot2:PA	–5.663e–02	2.797e–02	2.010e + 02	–2.025	0.044234 *
ot3:PA	6.347e–02	1.089e–02	1.312e + 04	5.829	5.69e–09 ***
ot4:PA	–2.426e–02	1.089e–02	1.311e + 04	–2.229	0.025854 *
ot1:SpeechType	–1.969e–02	3.992e–02	1.530e + 02	–0.493	0.622603
ot2:SpeechType	–6.957e–02	3.817e–02	1.460e + 02	–1.823	0.070403
ot3:SpeechType	2.205e–02	3.268e–02	1.320e + 02	0.675	0.501040
ot4:SpeechType	3.814e–02	2.784e–02	1.320e + 02	1.370	0.172970
PA:SpeechType	8.996e–03	1.081e–02	1.320e + 02	0.832	0.406834
ot1:PA:SpeechType	7.998e–02	5.899e–02	2.440e + 02	1.356	0.176417
ot2:PA:SpeechType	–6.906e–02	5.594e–02	2.010e + 02	–1.235	0.218427
ot3:PA:SpeechType	–4.815e–02	2.178e–02	1.312e + 04	–2.211	0.027045 *
ot4:PA:SpeechType	3.229e–02	2.177e–02	1.311e + 04	1.483	0.138103

Table 5: Experiment 1: Output of a mixed effect model for the looks to the competitor in Non-Contrastive sequences.
Model Structure: meanFixation ~ (ot1 + ot2 + ot3 + ot4) * PA * SpeechType + (1 + ot1 + ot2 + ot3 + ot4 | File) + (1 + ot1 + ot2 | File:PA).
Number of obs: 14135, groups: File:PA, 264; File, 132.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1.

The results of a mixed effects model for the competitor showed a main effect of PA, i.e., more looks to the previously mentioned ornament cell with the [L + H* unaccented] than with [H* !H*] contours. There was no main effect of Speech Type, nor Speech Type interaction with PA, which indicates that the two groups did not differ in the overall amount of looks to the competitor and the effect of PA on the looks to the competitor. The two-way interactions between the polynomial terms and PA suggest that the pitch contour affected the quadratic, cubic, and quartic components of the functions. The three-way interaction between ot3, PA, and Speech Type indicates that the effect of pitch contour for the cubic component was smaller for the spontaneous speech group.

Although the [L + H* unaccented] contour led to the expected facilitative effect on the looks to the target in the Contrastive sequences and the misleading effect on the looks to the competitor in the Non-Contrastive sequences, the magnitudes of these effects were surprisingly small, given the similarity in the task between the present study and Ito and Speer (2008) and the large number of participants in the present experiment. Upon careful inspection of the data, the researchers noticed that many participants remained looking at the central red star until the end of each auditory noun phrase. Thus, we decided to explore the data by excluding fixations on locations other than the six ornament cells (e.g., fixations on the center or off the board) for the calculation of fixation proportions. This subset analysis revealed much larger differences in fixation proportions across contour types for both Contrastive and Non-Contrastive sequences. For Contrastive sequences (**Figure 5**), the looks to the target ornament set started increasing at about 200ms into the noun with [L+H* unaccented] contours, while the increase was delayed and slower for with [H* !H*] contours in both groups. For the Non-Contrastive sequences (**Figure 6**), the incorrect looks to the competitor increased visibly during the critical noun only with [L+H* unaccented] contours in both groups. Thus, when participants moved their eyes as soon as they heard the speech input, the differences in the pitch contour facilitated the detection of target in Contrastive sequences and lead to anticipatory fixations in Non-Contrastive sequences, regardless of the speech type.

We conducted GCA on these subsets of data as well. The outcome of a model confirmed a main effect of PA, but it also showed a main effect of Speech Type on the looks to the target in the Contrastive sequences (**Table 6**). Thus, while participants generally looked more to the target with [L+H* unaccented] contours, those who heard the

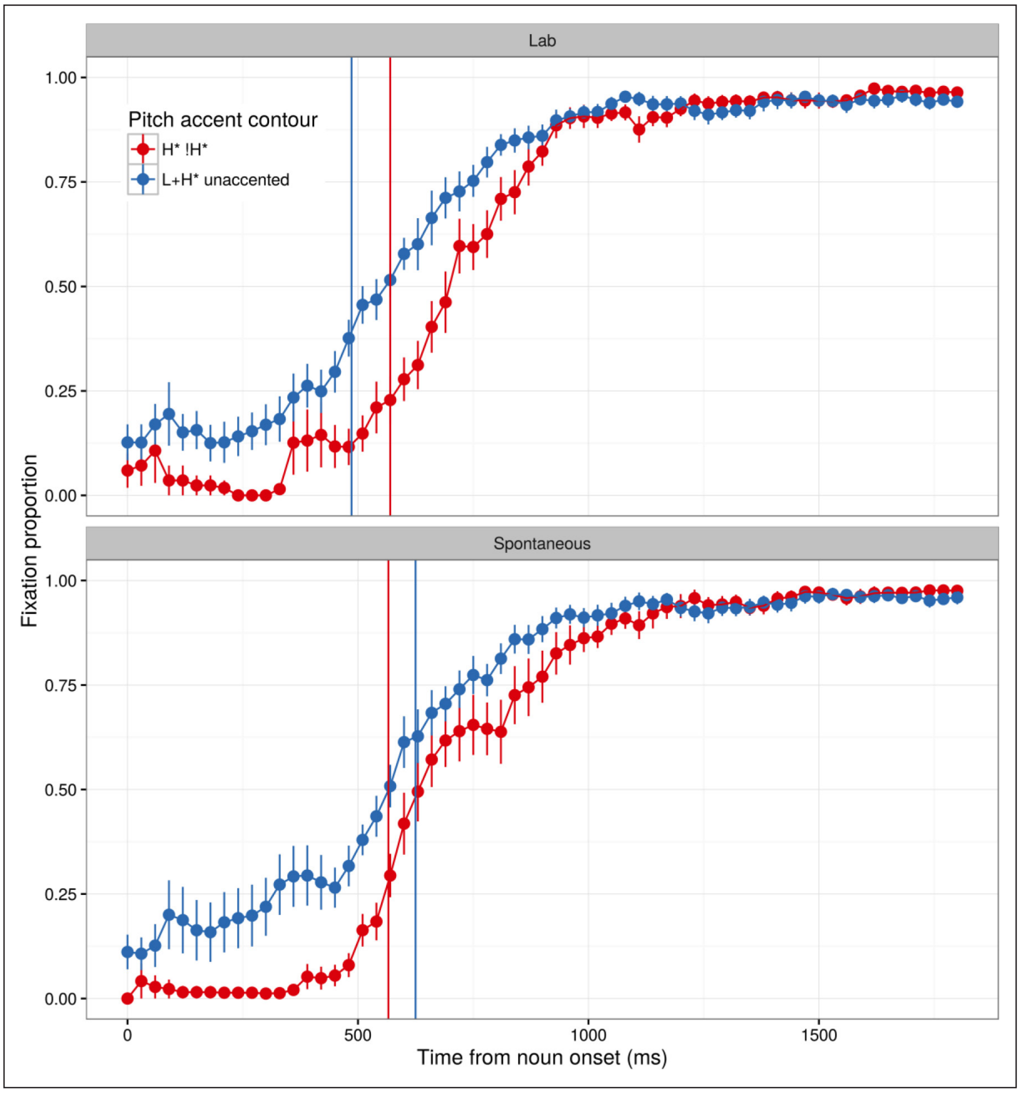

Figure 5: Experiment 1: Fixations to the target cell in Contrastive sequences after removing off-AOI fixations.

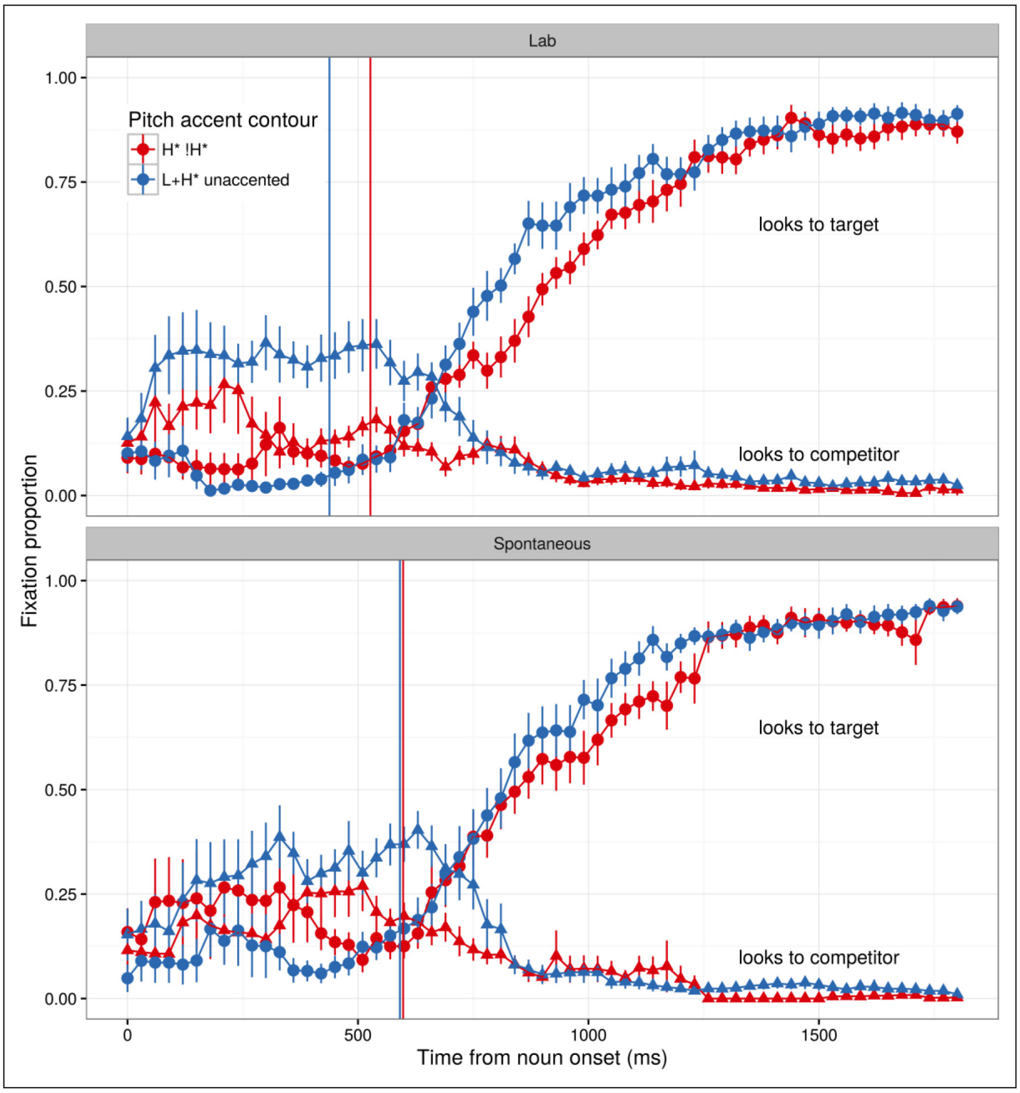

Figure 6: Experiment 1: Fixations to the target and competitor cells in Non-Contrastive sequences after removing off-AOI fixations.

spontaneous speech looked more to the target than those who heard lab speech. In fact, **Figure 5** shows that the increase in the looks to the target was steeper for the spontaneous speech group than for the lab speech group especially within the critical noun. While both PA and Speech Type showed some two-way and three-way interactions with the polynomial terms, the lack of interaction between PA and Speech Type suggests that the overall facilitative effect of [L+H* unaccented] contour on the looks to the target did not differ between the two groups. The three-way interactions suggest that the pitch contour affected different components of fixation functions differently across the two groups.

For the Non-Contrastive sequences, the outcome of subset analysis showed a main effect of PA and a main effect of Speech Type on the looks to the target (**Table 7a**). Again, participants in the spontaneous speech group showed a faster increase in looks to the target than those in the lab speech group. The analysis of looks to the competitor showed a main effect of PA, a main effect of Speech Type, and no interaction between the two (**Table 7b**). Again, [L+H* unaccented] led to more looks to the previously mentioned ornaments, and those who heard spontaneous speech made more looks to

Fixed effects:	**Estimate**	***SE***	***df***	***t* value**	**Pr(>|t|)**
(Intercept)	5.210e–01	1.002e–02	1.310e + 02	52.009	< 2e–16 ***
ot1	2.472e + 00	5.345e–02	1.330e + 02	46.241	< 2e–16 ***
ot2	–3.503e–01	5.066e–02	1.420e + 02	–6.914	1.48e–10 ***
ot3	–6.099e–01	3.368e–02	1.120e + 02	–18.108	< 2e–16 ***
ot4	2.632e–01	3.229e–02	1.290e + 02	8.151	2.71e–13 ***
PA	8.100e–02	1.353e–02	1.310e + 02	5.986	1.95e–08 ***
SpeechType	1.323e–01	2.004e–02	1.310e + 02	6.604	9.15e–10 ***
ot1:PA	–3.170e–01	7.983e–02	1.610e + 02	–3.971	0.000108 ***
ot2:PA	–3.049e–01	6.721e–02	2.040e + 02	–4.537	9.72e–06 ***
ot3:PA	2.169e–01	2.529e–02	7.213e + 03	8.576	< 2e–16 ***
ot4:PA	4.185e–02	2.314e–02	1.183e + 04	1.808	0.070562
ot1:SpeechType	1.118e–01	1.069e–01	1.330e + 02	1.045	0.297773
ot2:SpeechType	–2.657e–01	1.013e–01	1.420e + 02	–2.623	0.009673 **
ot3:SpeechType	6.107e–02	6.737e–02	1.120e + 02	0.907	0.366609
ot4:SpeechType	2.022e–01	6.458e–02	1.290e + 02	3.130	0.002158 **
PA:SpeechType	2.468e–02	2.707e–02	1.310e + 02	0.912	0.363446
ot1:PA:SpeechType	–4.303e–01	1.597e–01	1.610e + 02	–2.695	0.007788 **
ot2:PA:SpeechType	9.506e–02	1.344e–01	2.040e + 02	0.707	0.480250
ot3:PA:SpeechType	2.685e–01	5.059e–02	7.213e + 03	5.307	1.15e–07 ***
ot4:PA:SpeechType	–2.898e–01	4.628e–02	1.183e + 04	–6.262	3.94e–10 ***

Table 6: Experiment 1: Summary of mixed effect models for the looks to the target (vs. other AOIs) in Contrastive sequences.
Model Structure: meanFixation ~ (ot1 + ot2 + ot3 + ot4) * PA * SpeechType + (1 + ot1 + ot2 + ot3 + ot4 | File) + (1 + ot1 + ot2 | File:PA).
Number of obs: 12623, groups: File: PA, 264; File, 132.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

the contrastive competitors. Importantly, there was no three-way interaction between PA, Speech Type, and the quadradic term, i.e., the overall dome-shaped function (rise and fall) for the competitor was not affected differently by pitch contour across the two speech type groups.

In the above subset analysis, the number of trials was reduced to less than a half of those included in the earlier analysis. A closer observation of the video data informed us that many of the participants who had little contribution to the subset analysis kept looking at the center of the board until they heard the entire phrase. We suspected that our instruction to look at the red star was the cause of this common behavior, and thus decided to recode the data to examine the fixations to the narrow rectangular window that surrounded the red star. **Figure 7** shows the fixation proportions for the center rectangle (in red) for Contrastive and Non-Contrastive sequences. These figures confirm that participants of both groups were looking mostly at the center star while listening to the critical phrases, regardless of the sequence type and the contour type.

Fixed effects:	**Estimate**	***SE***	***df***	***t* value**	**Pr(>|t|)**
(Intercept)	3.964e–01	1.060e–02	1.290e + 02	37.392	< 2e–16 ***
ot1	2.152e + 00	6.549e–02	1.270e + 02	32.865	< 2e–16 ***
ot2	3.150e–01	4.586e–02	1.480e + 02	6.870	1.68e–10 ***
ot3	–5.578e–01	3.631e–02	1.210e + 02	–15.360	< 2e–16 ***
ot4	–3.278e–02	3.071e–02	1.260e + 02	–1.067	0.287829
PA	3.962e–02	1.615e–02	1.270e + 02	2.454	0.015497 *
SpeechType	1.441e–01	2.120e–02	1.290e + 02	6.796	3.61e–10 ***
ot1:PA	9.924e–02	1.074e–01	1.280e + 02	0.924	0.357087
ot2:PA	1.026e–02	7.017e–02	1.590e + 02	0.146	0.883917
ot3:PA	–1.387e–01	2.556e–02	6.478e + 03	–5.429	5.87e–08 ***
ot4:PA	1.287e–01	2.327e–02	1.190e + 04	5.529	3.28e–08 ***
ot1:SpeechType	1.490e–01	1.310e–01	1.270e + 02	1.138	0.257471
ot2:SpeechType	5.137e–02	9.171e–02	1.480e + 02	0.560	0.576270
ot3:SpeechType	–1.575e–01	7.263e–02	1.210e + 02	–2.169	0.032067 *
ot4:SpeechType	1.600e–01	6.142e–02	1.260e + 02	2.605	0.010295 *
PA:SpeechType	3.340e–02	3.230e–02	1.270e + 02	1.034	0.303070
ot1:PA:SpeechType	–3.158e–01	2.147e–01	1.280e + 02	–1.471	0.143868
ot2:PA:SpeechType	–8.160e–02	1.404e–01	1.590e + 02	–0.581	0.561797
ot3:PA:SpeechType	–1.768e–01	5.111e–02	6.478e + 03	–3.459	0.000546 ***
ot4:PA:SpeechType	1.823e–01	4.654e–02	1.190e + 04	3.918	8.97e–05 ***

Table 7a: Experiment 1: Outcome of a mixed effect model for the looks to the target (vs. other AOIs) in Non-Contrastive sequences.
Model Structure: meanFixation ~ (ot1 + ot2 + ot3 + ot4) * PA * SpeechType + (1 + ot1 + ot2 + ot3 + ot4 | File) + (1 + ot1 + ot2 | File:PA).
Number of obs: 12803, groups: File:PA, 264; File, 132.
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

2.5. Discussion: Experiment 1

Experiment 1 confirmed the predicted eye-movement patterns that suggest the contrastive interpretation of the [L+H* unaccented] contours in both lab speech and spontaneous speech: In the Contrastive sequences, the looks to the target ornament cell increased faster with the [L+H* unaccented] contours than with the [H* !H*] contours, while in the Non-Contrastive sequences, [L+H* unaccented] contours led to more incorrect looks to the previously mentioned ornament set than the [H* !H*] contours. Unfortunately, the instruction to look at the star on the ornament board, which was included for checking the calibration accuracy between trials, prevented the immediate eye-movement responses to speech input in many participants. The analysis of the subset data (excluding the fixations on the center) suggested that when participants did not keep fixating the star, they made immediate eye-movements upon hearing [L+H* unaccented] contours in a similar manner to the previous findings by Ito and Speer (2008). Importantly, the overall timing of fixations tended to be faster for the spontaneous speech group than for the lab speech group. It was particularly interesting that the [L+H* unaccented] contours of spontaneous speech, which did not have a large F0 excursion during the adjective, led to an immediate increase in looks to the previously mentioned ornament set during the noun at timing very similar to that for the

Fixed effects:	Estimate	*SE*	*df*	*t* value	Pr(>|t|)
(Intercept)	9.974e–02	8.676e–03	1.250e + 02	11.496	< 2e–16 ***
ot1	–3.562e–01	6.025e–02	1.320e + 02	–5.912	2.73e–08 ***
ot2	–1.789e–01	3.641e–02	1.150e + 02	–4.913	2.99e–06 ***
ot3	2.512e–01	3.662e–02	9.900e + 01	6.861	5.95e–10 ***
ot4	–5.568e–02	2.695e–02	1.180e + 02	–2.066	0.041038 *
PA	5.722e–02	1.432e–02	1.330e + 02	3.995	0.000107 ***
SpeechType	8.979e–02	1.735e–02	1.250e + 02	5.175	8.79e–07 ***
ot1:PA	–2.088e–01	1.085e–01	1.400e + 02	–1.924	0.056341
ot2:PA	–6.348e–02	5.860e–02	1.690e + 02	–1.083	0.280224
ot3:PA	1.793e–01	1.869e–02	1.032e + 04	9.596	< 2e–16 ***
ot4:PA	–9.944e–02	1.649e–02	1.216e + 04	–6.030	1.68e–09 ***
ot1:SpeechType	–5.909e–01	1.205e–01	1.320e + 02	–4.903	2.72e–06 ***
ot2:SpeechType	–9.427e–02	7.282e–02	1.150e + 02	–1.295	0.198075
ot3:SpeechType	3.494e–01	7.324e–02	9.900e + 01	4.770	6.32e–06 ***
ot4:SpeechType	–1.684e–01	5.390e–02	1.180e + 02	–3.123	0.002250 **
PA:SpeechType	6.897e–02	2.865e–02	1.330e + 02	2.408	0.017426 *
ot1:PA:SpeechType	–2.789e–01	2.170e–01	1.400e + 02	–1.285	0.200951
ot2:PA:SpeechType	–8.369e–02	1.172e–01	1.690e + 02	–0.714	0.476117
ot3:PA:SpeechType	1.829e–01	3.738e–02	1.032e + 04	4.892	1.01e–06 ***
ot4:PA:SpeechType	–1.185e–01	3.298e–02	1.216e + 04	–3.593	0.000329 ***

Table 7b: Experiment 1: Outcome of a mixed effect model for the looks to the competitor (vs. other AOIs) in Non-Contrastive sequences.
Number of obs: 12803, groups: File:PA, 264; File, 132.
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1.

lab speech (**Figure 6**). This suggests that listeners ‘garden-pathed’ on the basis of adjective information—they did not wait until they processed the F0 information of the noun, which best distinguished the contour types according to the classification tree analysis, to interpret the contour as signaling contrast. In order to confirm the equivalent timing of responses to the [L + H* unaccented] contours across lab speech and spontaneous speech, Experiment 2 was conducted without the instruction to look at the central star between trials.

3. Experiment II

3.1. Participants

A total of 77 participants (age range: 16–73, average 29) were recruited at a local science museum. Participants volunteered to take part in the experiment and did not receive any incentives. A total of 36 participants were assigned to the lab speech presentation list, while a total of 41 participants were assigned to the spontaneous speech presentation list such that the two groups matched for the age range.

3.2. Materials and design

The auditory and visual stimuli and the design were identical to those of Experiment 1, *except* that the red star was removed from the center of the ornament board.

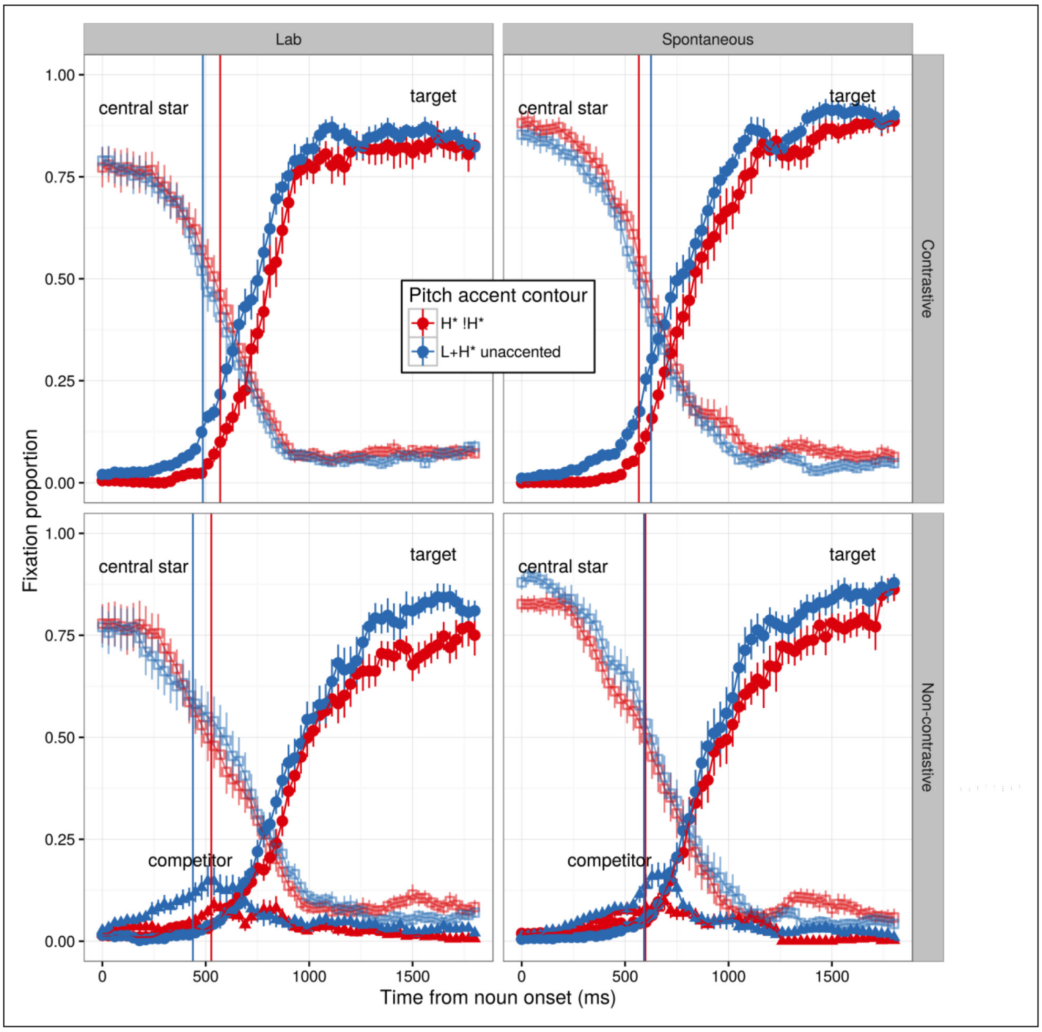

Figure 7: Experiment 1: Fixations to the center red star (shown with empty circles) across conditions.

3.3. Procedure

The experimental procedure was identical to that of Experiment 1, *except* that participants were simply told to return to facing the board after hanging each ornament on the tree. No instruction was given as to where they should look before each trial.

3.4. Results: Experiment 2

Data from 14 participants were excluded from the analyses due to frequent track loss (6), system failure (4), color-blindness (1), being a non-native speaker of American English (1), and being underage (2). Data from 63 participants (31 Lab: age 18–58, average 28;04;32 Spontaneous: age 19–73, average 30;10) were submitted to the following analyses.

The overall fixation patterns of participants in Experiment 2 resembled the results after the removal of off-AOI fixations (to the center and off the board) in Experiment 1. In Contrastive sequences, [L+H* unaccented] contours led to a faster increase in the looks to the target ornament cell in both lab and spontaneous speech groups (**Figure 8**: Denominators included only the fixations on the 6 AOIs). However, the magnitude of this facilitative effect was rather small, especially for the spontaneous speech group.

Since both groups reached their ceiling levels of looks by 1200ms after the noun onset, we submitted the fixation proportions for 0–1200ms post noun onset to GCA. The results showed a marginal main effect of PA, which interacted with ot2 (marginal), ot3, and

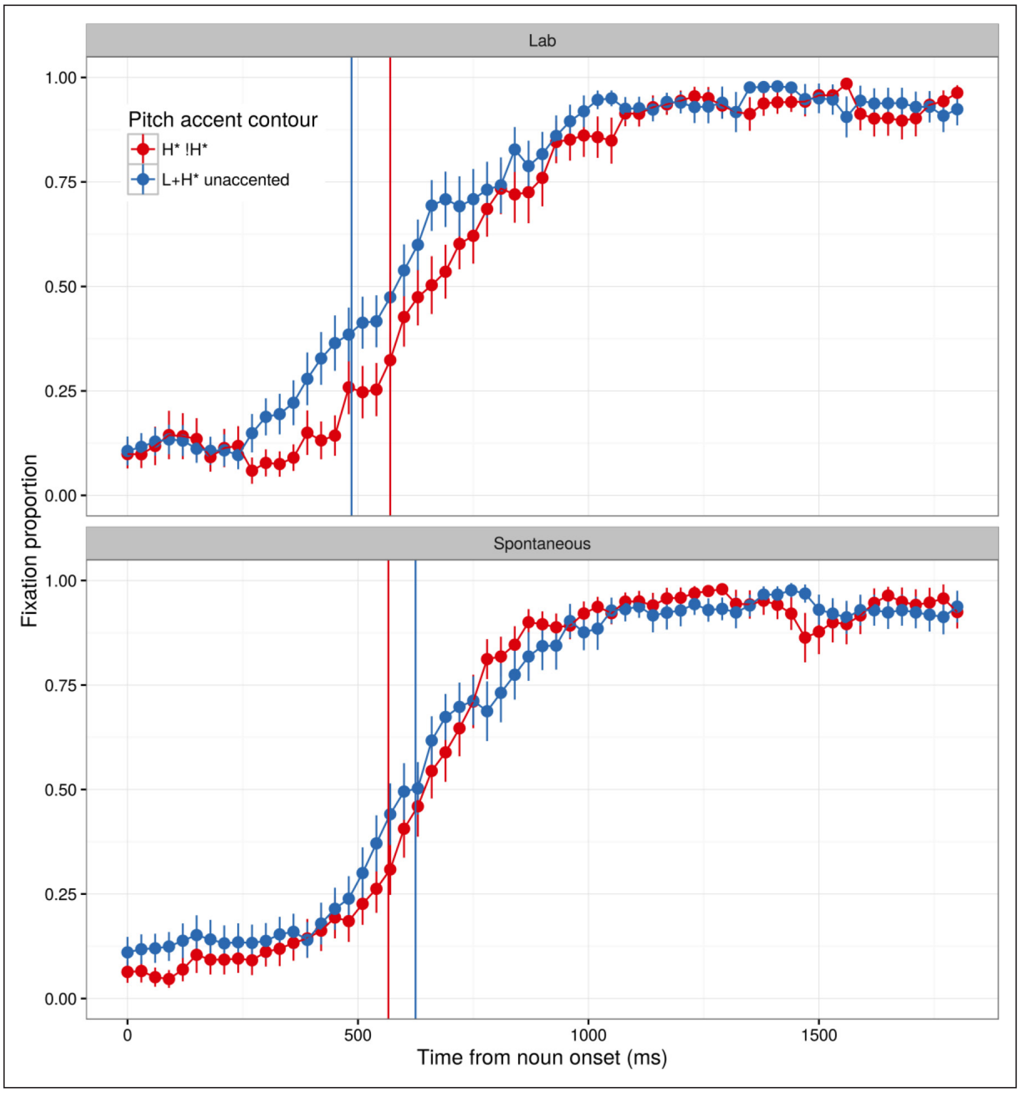

Figure 8: Experiment 2: Fixations to the target cell in Contrastive sequences.

ot4 (**Table 8**). Thus, like in Experiment 1, [L+H* unaccented] contours led to overall more looks to the target in the Contrastive sequences than [H* !H*] contours, yet the difference in the pitch contour did not seem to affect the steepness of the linear function.

In the Non-Contrastive sequences, the [L+H* unaccented] contours led to a quick increase in the looks to the contrastive competitor cells (i.e., previously mentioned ornament set) in both groups (**Figure 9**). In the lab speech group, the looks to the competitor kept increasing until 200–300ms past the noun offset. In both groups, the looks to the competitor remained at a relatively higher level for the [L+H* unaccented] than for the [H* !H*] trials for an extended period of time, and decreased more slowly for the [L+H* unaccented] than for the [H* !H*] trials.

The outcome of the mixed effects model (0–1400ms post noun onset) for the competitor in the Non-Contrastive sequences confirmed the main effect of PA and no effect of Speech Type, i.e., the [L+H* unaccented] led to overall higher looks to the previously mentioned ornament cell than [H* !H*] and there was no overall difference in the looks to the competitor between the groups (**Table 9**). The interaction between PA and ot2 indicates that the [L+H* unaccented] affected the quadratic (i.e., the dome-shape) function. The three-way interaction between ot3, PA, and Speech Type suggests that the effect of pitch contour on the cubic component was larger for the spontaneous speech group.

Fixed effects:	**Estimate**	***SE***	***df***	***t* value**	**Pr(>|t|)**
(Intercept)	4.748e–01	1.520e–02	6.200e + 01	31.243	< 2e–16 ***
ot1	2.009e + 00	6.169e–02	7.300e + 01	32.563	< 2e–16 ***
ot2	2.014e–01	7.160e–02	6.400e + 01	2.813	0.00651 **
ot3	–4.191e–01	4.935e–02	6.200e + 01	–8.492	5.85e–12 ***
ot4	–5.430e–02	3.832e–02	6.300e + 01	–1.417	0.16151
PA	4.304e–02	2.400e–02	6.300e + 01	1.793	0.07782
SpeechType	–6.648e–03	3.040e–02	6.200e + 01	–0.219	0.82761
ot1:PA	–9.362e–02	9.221e–02	1.090e + 02	–1.015	0.31220
ot2:PA	–1.607e–01	8.584e–02	8.700e + 01	–1.872	0.06451
ot3:PA	6.639e–02	2.957e–02	4.323e + 03	2.246	0.02478 *
ot4:PA	1.291e–01	2.942e–02	4.271e + 03	4.388	1.17e–05 ***
ot1:SpeechType	1.206e–01	1.234e–01	7.300e + 01	0.977	0.33166
ot2:SpeechType	4.267e–02	1.432e–01	6.400e + 01	0.298	0.76669
ot3:SpeechType	–7.812e–02	9.870e–02	6.200e + 01	–0.791	0.43171
ot4:SpeechType	–1.041e–01	7.665e–02	6.300e + 01	–1.358	0.17940
PA:SpeechType	–5.246e–02	4.801e–02	6.300e + 01	–1.093	0.27875
ot1:PA:SpeechType	–2.331e–01	1.844e–01	1.090e + 02	–1.264	0.20897
ot2:PA:SpeechType	2.252e–01	1.717e–01	8.700e + 01	1.312	0.19302
ot3:PA:SpeechType	–2.019e–02	5.913e–02	4.323e + 03	–0.341	0.73278
ot4:PA:SpeechType	–6.362e–02	5.883e–02	4.271e + 03	–1.081	0.27960

Table 8: Experiment 2: Outcome of a mixed effect model for the looks to the target (vs. other AOIs) in Contrastive sequences.
Number of obs: 4704, groups: File:PA, 125; File, 63.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

To further test whether the level of initial increase of the looks to the competitor differed between the lab and spontaneous speech groups, fixation proportions up to 600ms after the noun onset were submitted to another mixed effects model. The outcome confirmed the main effect of PA (*Est.* = 1.197e–01, *SE* = 2.940e–02, *t* = 4.073, *p* < .001), no effect of Speech Type, and no interaction between PA, Speech Type, and polynomial terms. Thus, [L + H* unaccented] increased the overall looks to the competitor as compared to [H*!H*] equally for both groups, and changes over time did not differ across the two groups.

3.5. Discussion: Experiment 2

Without the instruction to look at the central star, Experiment 2 confirmed that the [L + H* unaccented] contour leads to incorrect fixations to the contrastive competitor in Non-Contrastive sequences, even with many fewer participants (about half of those included in Experiment 1). In both groups, fixations to the previously mentioned ornament set (e.g., houses in 'clear house,' 'BROWN ball') started increasing during the noun (e.g., ball) despite the conflicting segmental information, and those fixations to the contrastive competitors were sustained relatively higher than in the [H* !H*] trials until a few hundred milliseconds after the noun offset. While the degree of increase in the fixations to the competitor appeared larger for the lab speech group (**Figure 9**), neither the main

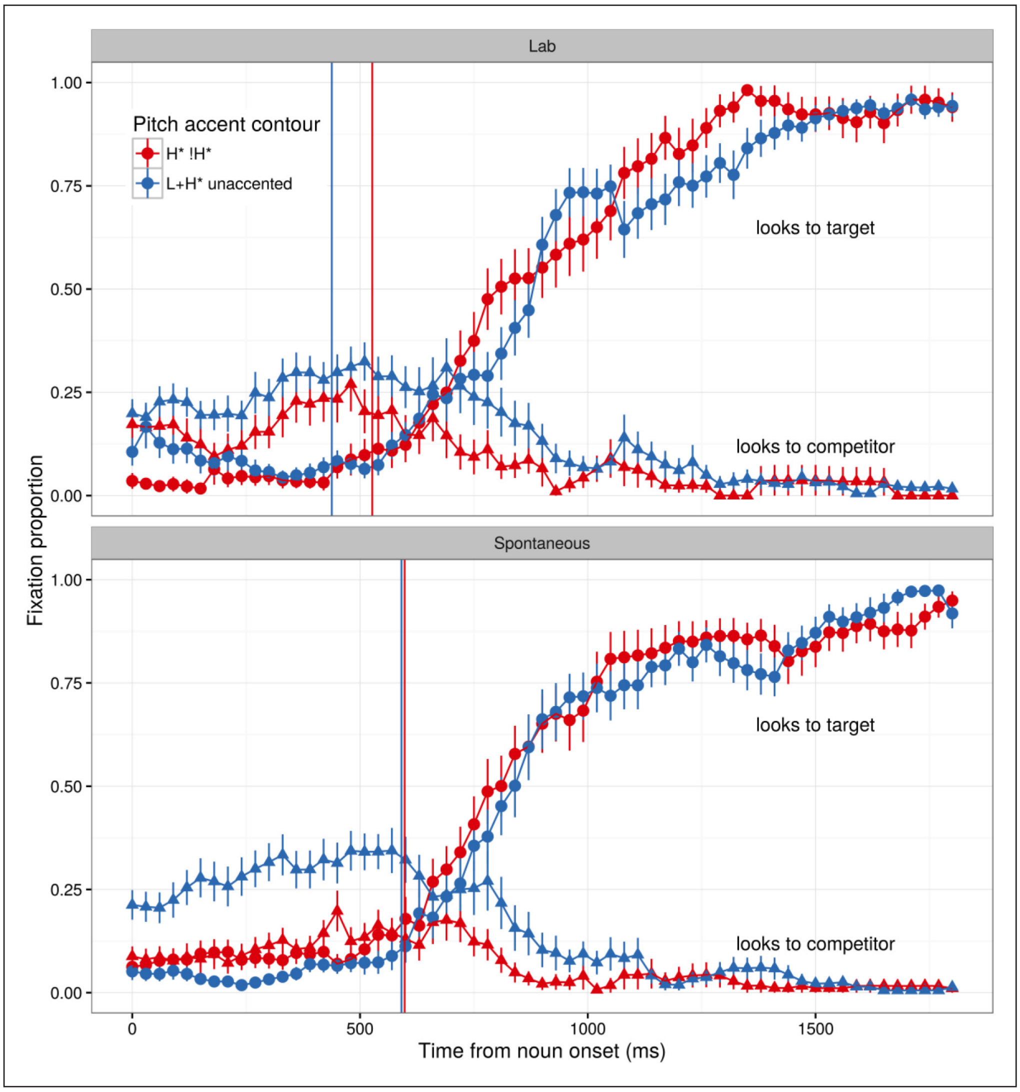

Figure 9: Experiment 2: Fixations to the target and competitor cells in Non-Contrastive sequences.

effect of Speech Type nor the interaction between PA and Speech Type was confirmed. In addition, the relatively longer looks to the previous target set seemed to delay the looks to the correct target set at a similar timing across the two groups. The results of Experiment 2 therefore confirmed no critical differences in the timing of fixations between the two speech type groups.

4. General Discussion

The present study tested whether spontaneously produced pitch contours, which have been ToBI-annotated as [L+H* unaccented], and rated as relatively more prominent by naïve listeners (Turnbull et al., 2014, 2017), yet do not have an extreme pitch rise and fall for the adjective labeled with L+H*, can be interpreted contrastively by a broad range of adult participants recruited at a local science museum. The timings of responses to pitch contours were compared between the laboratory speech group and the spontaneous speech group. Unlike the study by Ito and Speer (2008) that originally used the Christmas tree decoration task, stimuli in the present study were limited to noun phrases that consisted of a color adjective and an ornament noun, and included no conversational context. According to Turnbull et al. (2014, 2017), listeners may not process a sequence of mere isolated phrases such as “green drum, brown drum” contrastively unless they believe that the sequence is a part of a communicative discourse. If a naturalistic discourse environ-

Fixed effects:	Estimate	*SE*	*df*	*t* value	Pr(>|t|)
(Intercept)	1.613e–01	1.430e–02	6.300e + 01	11.281	< 2e–16 ***
ot1	–2.799e–01	4.831e–02	8.300e + 01	–5.794	1.19e–07 ***
ot2	–2.593e–01	5.004e–02	6.300e + 01	–5.182	2.44e–06 ***
ot3	8.114e–02	3.215e–02	6.000e + 01	2.524	0.0143 *
ot4	1.336e–01	2.795e–02	6.100e + 01	4.780	1.15e–05 ***
PA	9.969e–02	2.223e–02	6.300e + 01	4.484	3.17e–05 ***
SpeechType	–1.589e–02	2.860e–02	6.300e + 01	–0.556	0.5804
ot1:PA	–1.522e–01	8.985e–02	9.400e + 01	–1.694	0.0936
ot2:PA	–1.465e–01	6.284e–02	7.500e + 01	–2.332	0.0224 *
ot3:PA	3.275e–02	2.695e–02	4.353e + 03	1.215	0.2244
ot4:PA	1.186e–03	2.669e–02	4.308e + 03	0.044	0.9646
ot1:SpeechType	–1.016e–02	9.662e–02	8.300e + 01	–0.105	0.9165
ot2:SpeechType	–3.509e–02	1.001e–01	6.300e + 01	–0.351	0.7271
ot3:SpeechType	2.369e–02	6.429e–02	6.000e + 01	0.368	0.7139
ot4:SpeechType	–1.768e–02	5.591e–02	6.100e + 01	–0.316	0.7529
PA:SpeechType	4.942e–02	4.446e–02	6.300e + 01	1.112	0.2706
ot1:PA:SpeechType	–2.483e–01	1.797e–01	9.400e + 01	–1.382	0.1704
ot2:PA:SpeechType	–9.813e–03	1.257e–01	7.500e + 01	–0.078	0.9380
ot3:PA:SpeechType	1.066e–01	5.390e–02	4.353e + 03	1.977	0.0481 *
ot4:PA:SpeechType	–6.218e–02	5.338e–02	4.308e + 03	–1.165	0.2441

Table 9: Experiment 2: Outcome of a mixed effect model for the looks to the competitor (vs. other AOIs) in Non-Contrastive sequences.
Number of obs: 4714, groups: File:PA, 126; File, 63.
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

ment were an absolute precondition for comprehending the semantics of pitch contours, then the minimally engaging sequential presentation of pre-recorded phrases would not lead to contrastive interpretation of [L + H* unaccented], whether it is laboratory speech or spontaneous speech.

The overall patterns of the present results, however, suggest that adult listeners recruited outside a typical college community processed the [L + H* unaccented] contours contrastively, regardless of speech type. Thus, the present tree decoration task, while eliciting no linguistic responses from participants, may have provided a sufficient visual environment and task structure to allow participants to interpret the sequence of bare [color + ornament] noun phrases as a component in a coherent discourse. More importantly, the present results support the validity and generalizability of past work that has demonstrated the effect of pitch prominence to evoke contrastive interpretation of a referential expression produced in lab speech in a more narrowly selected group of participants (e.g., Dahan et al., 2002; Ito & Speer, 2008, 2011; Ito et al., 2012; Ito et al., 2014; Kurumada et al., 2014; Weber et al., 2006).

Because the present stimuli were presented without a carrier phrase such as "Hang the …," which would have provided the pitch range cues that facilitates the assessment of the prominence of the following words, we were surprised to see the steady increase in the looks to the competitor for trials where [L+H* unaccented] tokens appeared in Non-Contrastive sequences, especially with the spontaneous speech. While the increase did not appear as large as in the lab speech group (**Figure 9**), the looks to the contrastive competitor with the spontaneous speech started rising gradually from the noun onset and kept increasing while incompatible segmental information was unfolding. This result indicates that listeners interpreted the pitch contour of [L+H* unaccented] trials contrastively based on information available early in the adjective-noun phrase, and executed anticipatory eye-movements before fully integrating information from later in the phrase. Such a finding was not independently predicted by individual factors in our acoustic analysis of the stimuli where the phonetic measures from the adjective such as word/vowel duration, F0 height, and spectral tilts did not reliably distinguish the two contour types for spontaneous speech. However, a non-linear combination of acoustic feature differences between the adjectives in the two tune types was successful in distinguishing them. The present data suggest that listeners were sensitive to these cues to contrast in the earlier part of the [L+H* unaccented] spontaneous contours.

Since our acoustic analysis did not include the F0 curvature that may affect the categorical perception of pitch contours, we re-examined our stimuli to see whether the early part of [L+H* unaccented] pitch contour exhibited a consistent form. The F0 contours of our stimuli from the spontaneous speech did not show either a scoopy rise or a domed fall as a consistent cue to L+H* (Appendix B). In contrast, our lab speech stimuli exhibited a more uniform, clear scoopy rise and fall followed by a pitch range compression for the [L+H* unaccented] items, and a consistent 'hat pattern' with a downstep for the [H* !H*] items. With the F0 contour patterns alone, it is difficult to determine what cues in the spontaneous speech stimuli led to the immediate contrastive interpretation of the [L+H* unaccented] items. Because the spontaneous speech stimuli were so much less uniform than the laboratory speech stimuli, it is possible that listeners responded to a common cue present in all lab speech trials, but to a more varied constellation of cues present in subsets of the spontaneous speech trials. While the visual inspection of F0 curvatures may not discern contour types, future studies may benefit from computational models of perceptual correlates that incorporate multiple aspects of signals that interact with the F0 curvature, such as TCoG (Barnes et al., 2015).

While the present study replicated both the facilitative effect of the [L+H* unaccented] for detecting the target in Contrastive sequences and the 'garden-path' effect to increase incorrect eye-movements in Non-Contrastive sequences, the magnitudes of the facilitative effect were rather small across experiments. We suspect that the smaller facilitative effects in the present study resulted from the overall slower responses to speech input in the experiments, which left little room to show the prosody-driven differences in the eye-movements for the repeated, thus easier-to-detect ornament sets. Despite the similarities in the phrase structure, the overall timing of fixation increase in the Contrastive sequences was slower in the present study as compared to the results of Ito and Speer (2008): In the present study, the fixations to the target set started increasing at about 200–300ms after the onset of the noun with both speech types, whereas the looks to the target in an equivalent condition in Ito and Speer started rising sharply at the noun onset.[4] A comparison of stimuli across the two studies revealed that the average duration of adjectives was shortest in Ito and Speer (323ms), longest in the present lab speech (380ms), and in between for the present spontaneous speech (341ms), whereas the duration of the noun did not differ as much across the studies (present lab speech: 498ms, present spontaneous

[4] The general difference in the fixation timing between the present study and Ito and Speer (2008) was also observed in Non-Contrastive sequences.

speech, 477ms, and Ito & Speer lab speech: 478 ms). As compared to the lag in the timing of fixations (200–300ms), the durational differences in the speech stimuli were rather small. Therefore, it is unlikely that the differences in the fixation timing across the studies were due to the differences in the duration of stimuli. It is also counter-intuitive that relatively longer adjectives led to slower fixation increases in the present study. If listeners relied on the prosodic cues from the adjective to program saccades, longer word duration could have benefited anticipatory eye-movements as it would provide listeners with extra time to process the signal.

We suspect that the overall delay in fixations in the present study resulted from multiple factors. First, although the speech stimuli were presented over noise-cancelling headphones, the science museum environment has much more noise and many more distractions than the quiet university laboratory. While participants were seated in front of the ornament board, their family members and friends could be waiting for the session to be over in the nearby visible waiting area or could be interacting with the other researchers in the same large lab space. The museum visitors outside the glass wall were often gathering behind the participant to see what was happening in the lab (and they sometimes tapped on the glass wall to greet the experimenters!). Thus, the experimental environment of the present study may have required extra attention to the task and stimuli, which may have been difficult to maintain throughout the session. Second, the slower responses may have reflected the greater difficulty in perceiving the relative prominence of phrase-initial adjectives, which were presented without any preceding carrier phrase. The pre-recorded speech stimuli in Ito and Speer included abundant acoustic context that may have made the calibration of the speaker's pitch range and the perception of relative prominence easier even with shorter word duration. While the present data demonstrated that listeners appropriately processed the semantics of pitch contours of isolated phrases, spoken context preceding the critical phrase could have made their responses even faster. Third, the present participant group may have had a wider range of individual differences in response timing than the participant group of Ito and Speer, which may be closely related to the wider age range. To quickly examine whether older participants had overall slower fixation increase, participant's age was added to the mixed effects models for predicting the likelihood of fixation on the target in the Contrastive sequence in Experiment 2. The results showed no significant main effect of age, nor any interactions with contour type or speech type. Thus, among the individuals tested in Experiment 2, age was not a reliable predictor of the response speed or sensitivity to prosodic pattern and speech style. Future study is required to explore how other individual factors such as attention and verbal memory in a noisy environment are related to prosodic processing of speech.

Finally, the present study demonstrated the impact of the task instruction on participants' behavior. Although the experimenter did not tell participants to keep looking at the center red star while listening to the ornament names in Experiment 1, most participants habitually sustained their gaze on the star across the trials. It is possible that participants misunderstood the instruction, but they also may have kept looking at the star while trying hard to concentrate on the speech input. If this was a general strategy that participants adopted for processing spoken instruction in a noisy or distractive environment, it may have also contributed to the equally slow fixation timing even without the star in Experiment 2.

To conclude, we would like to emphasize that naïve speakers' natural speech without dramatic F0 excursions can lead to the contrastive interpretation of prosodic prominence in naïve listeners. The exact acoustic cues that triggered the contrastive interpretation of spontaneous speech is still to be explored, yet the present results strongly suggest that the mechanism of prominence processing may be flexible and listeners can tune to a different set of acoustic cues (i.e., allophonic prosodic cues) in different voices for interpreting the same intention. Future studies therefore should make use of a wider range of spontaneous speech from multiple speakers to test whether a set of particular acoustic properties emerges as the invariant primary cues to contrast or whether a combination of cues to contrast is highly variable across speakers. If the latter is the case, we will be given a further complex research problem as to how listeners adapt to the speaker-specific prosodic cues to contrast in various environments.

Competing Interests

The authors have no competing interests to declare.

References

Barnes, J., Veilleux, N., Brugos, A. and Shattuck-Hufnagel, S. 2015. Interpreting patterns of variability in the realization of English intonation contours. Poster presented at Experimental and Theoretical Advances in Prosody 3, University of Illinois, Urbana-Champaign.

Beckman, M. E. and Ayers, E. G. 1997. Guidelines for ToBI labeling, vers 3.0 [manuscript]: Ohio State University.

Beckman, M. E., Hirschberg, J. and Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A., (ed.), *Prosodic typology; the phonology of intonation and phrasing*, 9–54. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0002

Beckman, M. E. and Pierrehumbert, J. B. 1986. Intonational structure in Japanese and English, *Phonology Yearbook*, Vol. 3, 255–309. DOI: https://doi.org/10.1017/S095267570000066X

Braun, B. and Tagliapietra, L. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes*, 25: 1024–1043. DOI: https://doi.org/10.1080/01690960903036836

Bruce, G. 1977. *Swedish word accents in sentence perspective*. Geleerup: Lund.

Calhoun, S. 2012. The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, 40(2): 329–349. DOI: https://doi.org/10.1016/j.wocn.2011.12.001

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

Dahan, D., Tanenhaus, M. K. and Chambers, C. G. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47, 292–314. DOI: https://doi.org/10.1016/S0749-596X(02)00001-3

Goldsmith, J. A. 1976. Autosegmental phonology (Doctoral dissertation). MIT, Cambridge, MA.

Goldsmith, J. A. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell.

Gussenhoven, C. 2004. The Phonology of tone and intonation. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511616983

Gussenhoven, C., Repp, B. H., Rietvelt, A. C. M., Rump, H. H. and Terken, J. 1997. The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America* 102, 3009–3021. DOI: https://doi.org/10.1121/1.420355

Ito, K., Bibyk, S., Wagner, L. and Speer, S. R. 2014. Interpretation of contrastive pitch accent in 6- to 11-year-old English speaking children (and adults). *Journal of Child Language, 41*(1), 84–110. DOI: https://doi.org/10.1017/S0305000912000554

Ito, K., Jincho, N., Minai, U., Yamane, N. and Mazuka, R. 2012. Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds. *Journal of Memory and Language*, *66*(1), 265–284. DOI: https://doi.org/10.1016/j.jml.2011.09.002

Ito, K. and Speer, S. R. 2006. Using interactive tasks to elicit natural dialogue. In: Sudhoff, S., Lenertova, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N. and Schlieβer, J. eds., *Methods in empirical prosody research*, 229–257. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110914641.229

Ito, K. and Speer, S. R. 2008. Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language,* 58*,* 541–573. DOI: https://doi.org/10.1016/j.jml.2007.06.013

Ito, K. and Speer, S. R. 2011. Semantically-independent but contextually-dependent interpretation of contrastive accent. In: Prieto, P., Frota, S. and Elordieta, G., (eds.) *Prosodic categories: production, perception and comprehension,* 69–92. Springer. DOI: https://doi.org/10.1007/978-94-007-0137-3_4

Jun, S.-A. 2005. *Introduction. Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Jun, S.-A. 2014. *Prosodic typology: By prominence type, word prosody, and macro-rhythm. Prosodic Typology 2*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199567300.001.0001

Krahmer, E. and Swerts, M. 2001. On the alleged existence of contrastive accents. *Speech Communication,* 34(4): 391–405. DOI: https://doi.org/10.1016/S0167-6393(00)00058-3

Kurumada, C., Brown, M., Bibyk, S. A., Pontillo, D. and 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: https://doi.org/10.1016/j.cognition.2014.05.017

Ladd, R. D. 1996. *Intonational phonology.* Cambridge: Cambridge University Press.

Ladd, R. D. 2008. *Intonational phonology*. Cambridge: Cambridge University Press, 2nd edition. DOI: https://doi.org/10.1017/CBO9780511808814

Ladd, R. D. and Morton, R. 1997. The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, 25, 313–342. DOI: https://doi.org/10.1006/jpho.1997.0046

Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R*. Chapman and Hall / CRC.

Mirman, D., Dixon, J. A. and Magnuson, J. S. 2008. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475–494. DOI: https://doi.org/10.1016/j.jml.2007.11.006

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

Pierrehumbert, J. B. and Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P. R., Morgan, J. and Pollack, M. E., (eds.), *Intentions in communication*, 342–65. Cambridge, MA: MIT Press.

Speer, S. R. and Foltz, A. 2015. The implicit prosody of corrective contrast primes appropriately intonated probes (for some readers). In: Frazier, L. and Gibson, T., (eds.): *Explicit and implicit prosody in sentence comprehension: Studies in honor of Janet Fodor.* DOI: https://doi.org/10.1007/978-3-319-12961-7_14

Speer, S. R., Warren, P. and Schafer, A. J. 2011. Situationally independent prosodic phrasing. *Laboratory Phonology* 2(1): 35–98. DOI: https://doi.org/10.1515/labphon.2011.002

Therneau, T. M. and Atkinson, E. J. 1997. An introduction to recursive partitioning using the RPART routines. Technical Report 61, Section of Biostatistics, Mayo Clinic, Rochester, MN.

Turnbull, R., Royer, A., Ito, K. and Speer, S. R.. 2014. Prominence perception in and out of context. *Speech Prosody* 7. Dublin, May.

Turnbull, R., Royer, A., Ito, K. and Speer, S. R. 2017. Prominence perception is dependent on phonology, semantics, and beliefs about discourse. *Language, Cognition, and Neuroscience*. DOI: https://doi.org/10.1080/23273798.2017.1279341

Watson, D., Tanenhaus, M. K. and Gunlogson, C. 2008. Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science 32*: 1232–1244. DOI: https://doi.org/10.1080/03640210802138755

Weber, A., Braun, B. and Crocker, M. W. 2006. Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, 49(3), 367–92. DOI: https://doi.org/10.1177/00238309060490030301

How to cite this article: Ito, K, Turnbull, R and Speer, S R 2017 Allophonic tunes of contrast: Lab and spontaneous speech lead to equivalent fixation responses in museum visitors. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 6, pp. 1–29, DOI: https://doi.org/10.5334/labphon.86

Published: 04 April 2017

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS