# Immediate effects of intonational prominence in a visual search task

*Kiwako Ito & Shari R. Speer*

Department of Linguistics
Ohio State University, USA
{ito; speer}@ling.ohio-state.edu

## Abstract

Previous observation of spontaneous speech has shown consistent use of pitch accent by speakers to mark the contrastive status of words. To investigate how listeners process accentual prominence marking a contrast, eye-movements were monitored while participants listened to spoken directions and searched for ornaments to decorate holiday trees. Eye movement latencies to the target ornament cells were shorter when the intonation felicitously marked contrast on the color (e.g. First, hang the green drum → Next, hang the ORANGE drum.) than when it did not (→ orange DRUM). Felicitous pitch accent placement also induced earlier fixations to the target compared to trials that simply lacked the emphatic accent (→ orange drum). In addition, the infelicitous use of the accent on the color modifier (e.g. green drum → ORANGE ball) led to incorrect initial fixations to the preceding ornament cell (e.g. drum) *before* the noun itself was processed. These results demonstrate the immediate processing of accentual information on a modifier that leads to a strong expectation about the upcoming discourse entity.

## 1. Introduction

Understanding the role of intonation in discourse comprehension requires mapping the correspondence between the physical aspects of spontaneously produced intonation and online responses of listeners to perceived accentual patterns. Previous investigation of spontaneous productions during a tree-decoration instruction task showed that adjectives were more frequently accented than the following head nouns, whether they appear as novel entities or as already-introduced entities in the discourse [1]. Also, speakers were more likely to use the emphatic accent L+H* [2, 3], when a word conveyed a contrast than when it did not (e.g., "… orange candy. Next is a GREEN candy." Here, the word with L+H* is designated by CAPITAL letters.)

To investigate how prosodic information is processed during speech comprehension, listeners' responses to various intonation patterns must be monitored online. Eye-movement monitoring serves as a highly informative measurement of the converser's immediate responses to auditory input. A previous study with a simple visual search task demonstrated that listeners started fixating the target object (e.g., beaker) after 200 ms from the onset of the word [4]. In the presence of a lexical cohort member (e.g., beetle), fixation proportions to the target started diverging from those of the cohort at about 200ms after the point where unique phonetic information identified the target. Another study investigated the effect of pitch accent during an object-moving task with phonetic cohorts serving as target/competitor pairs (e.g., candy & candle) [5]. Participants heard paired instructions such as "Put the candy below the triangle, …Now put the CANDLE/candle above the square/ABOVE THE SQUARE." Upon hearing the second instruction, participants fixated the unmentioned competitor when the target object was marked with prominent accent (CAN…). In contrast, when the target was not produced with accentual prominence (can…), fixations were directed to the previously-mentioned object. These results suggest that accentual prominence immediately evokes reference to an alternative entity in the discourse representation.

In the present study, two eye-tracking experiments were conducted to test how the emphatic accent L+H* on a modifier adjective (e.g., BLUE ball) affects visual search during a tree-decoration task. If L+H* on a modifier (e.g., BLUE) is immediately processed as expressing contrastive information, listeners should be able to build an expectation that the modified object will be one of a set that has been recently mentioned, and this immediate prosodic processing may lead to anticipatory eye-movements to the target object (e.g., ball) even *before* participants hear the noun. The data demonstrate a robust effect of L+H* leading to such anticipatory eye fixations.

## 2. Experiment 1

### 2.1. Design and Materials

Participants were told to follow audio instructions to decorate holiday trees. Each participant had four trees to decorate, and each tree had a total of 24 ornaments on it. For each tree, a set of ornaments were sorted by object type and displayed on a grid. Eight color adjectives (blue, red, green, orange, gold, silver, brown, grey) were used to paint eight objects (ball, drum, angel, bell, stocking, onion, candy, egg). Three additional colors (purple, white, yellow) were used for four fillers (snowmen, lightbulb, tree, star) and 'dummy' ornaments that remained unmentioned on the grid. Within each tree, there were four sequences that repeated the object noun, serving as a context where the color adjective should convey contrastive information (e.g., green drum → orange drum). There were also four sequences that repeated the color adjective, serving as a context where the object noun should convey contrastive information (e.g., brown ball → brown angel). Across the four trees, each adjective and each noun was used twice to make these contrastive sequences.

The intonational patterns of the instructions were determined based on the accentual patterns observed in the previous spontaneous production study [1]. For the above two types of contrastive sequencers, one half of the critical trials were presented with felicitous intonation, while the other half were presented with infelicitous intonation. When the contrast was on color adjective with a repeated object noun, the target noun phrase was felicitously produced with L+H* on the adjective followed by no accent on the noun (e.g., green drum → ORANGE drum.) In infelicitous renditions, the noun

phrase was produced with H* on the adjective followed by L+H* on the repeated noun (e.g., green drum → orange DRUM). When the contrast was on the object noun with a repeated color adjective, the target noun phrase was felicitously produced with [H* L+H*] (e.g., brown ball → brown ANGEL). In infelicitous trials, the noun phrase was produced with [L+H* no-accent] pattern (e.g., brown ball → BROWN angel).

The audio instructions were recorded using SoundEdit16, Version II (22.05KHz, 16 bits) with a trained female phonetician. Two ToBI transcribers annotated instruction utterances to ensure the accentual patterns of the instructions using Praat 4.2.17 [6]. Example ToBI transcriptions of [L+H* no-accent] and [H* L+H*] are given in Figure 1.

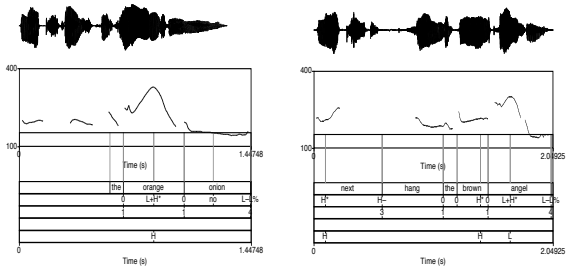


Figure 1: *Example ToBI annotations*.

Table 1 summarizes the mean word duration and mean F0 peaks of the target noun phrases of the four critical conditions in Experiment 1.

Table 1: *Mean duration and F0 of target phrases in Experiment 1.*

| Condition | Adjdur (ms) | AdjF0 (Hz) | Ndur (ms) | NF0 (Hz) |
|---|---|---|---|---|
| Felicitous Adj Cont | 347 | 304 | 401 | --- |
| Infelicitous Adj Cont | 307 | 207 | 573 | 291 |
| Felicitous N Cont | 328 | 205 | 498 | 298 |
| Infelicitous N Cont | 358 | 295 | 425 | --- |

### 2.2. Participants

36 undergraduate students at Ohio State University, who are native American English speakers participated as a part of their linguistics course requirement.

### 2.3. Eye-tracking procedure

Participants sat in front of a drafting table with the top tilted at 35 degrees to support the ornament display board. They wore light headgear with a small eye-camera and a magnet that functioned to correct measured eye positions for head movement. Participants were told to follow instructions from a loud speaker to choose a specified ornament from the board and place it on the tree located to their left. The x and y coordinates of eye-fixations on the board were recorded at 60 Hz using ASL e5000 data-collection system. The target area of interest for each trial was pre-specified to coincide with the relevant ornament cell on the board. Participants had to say "O.K." every time they finished placing one ornament, so that the experimenter, who sat on the other side of the drafting table, could play the next instruction. The experimenter also monitored the participant's performance via a ceiling-mounted camera. An example image from the ceiling camera is presented in Figure 2. Eye position is represented by cross-hairs.

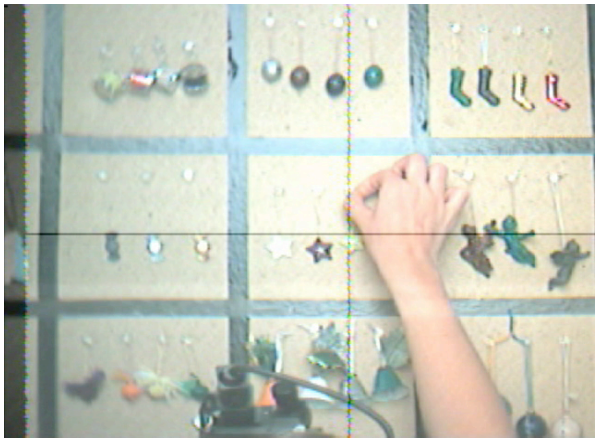

Figure 2*: Example image of the ornament display taken from the ceiling-mounted camera.*

### 2.4. Results

Each of the 36 participants had 8 trials in each of the four critical conditions. The fixation proportion was calculated for each time point by dividing the total number of fixations to the target by the number of possible fixations (288=8 x 36). A comparison of the two conditions where the contrast was on the color adjective showed a robust advantage for felicitous L+H* (e.g. First, hang the green drum. → Next, hang the ORANGE drum.) compared to infelicitous L+H* (e.g. → Next, hang the orange DRUM). Figure 3 shows that fixation proportions for the felicitous accentual pattern increased earlier than those for the infelicitous pattern. The two lines start diverging at around 200 ms after the onset of the target noun (e.g., drum). In the condition where L+H* felicitously marked contrast on the color adjective, the fixation proportion reached its high plateau before 600 ms into the noun, whereas

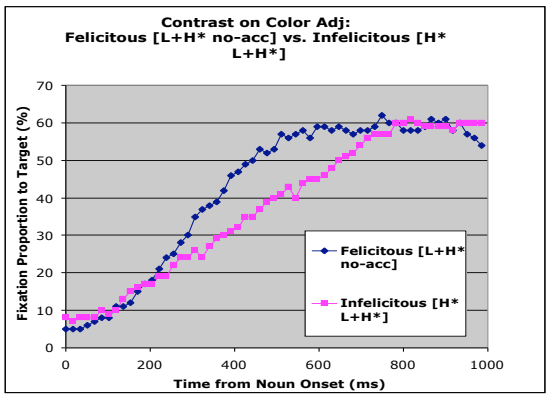


Figure 3*:Contrast on Adjective: Felicitous [L+H* no accent] vs. Infelicitous [H* L+H*].*

it did not reach the equivalent level until almost 800ms into the noun in the trials where L+H* was infelicitously used. Planned comparison following a one-way repeated measures ANOVA showed a significant difference in mean fixation proportion for the 300-600 ms window ($F(1,7) = 27.55$, $p < .001$). Interestingly, when the color was repeated and thus the object noun conveyed contrast, no facilitation in the timing of fixations was found for the felicitous intonation [H* L+H*] (e.g., First, hang the brown ball. → Next, hang the brown ANGEL) as compared to the infelicitous intonation [L+H* no-accent] (→ Next, hang the BROWN angel). As shown in Figure 4, fixation proportions increased relatively slowly in both conditions, and the difference between fixation proportions in the two conditions did not reach statistical significance ($F(1,7) = 3.07$, $p = .081$).

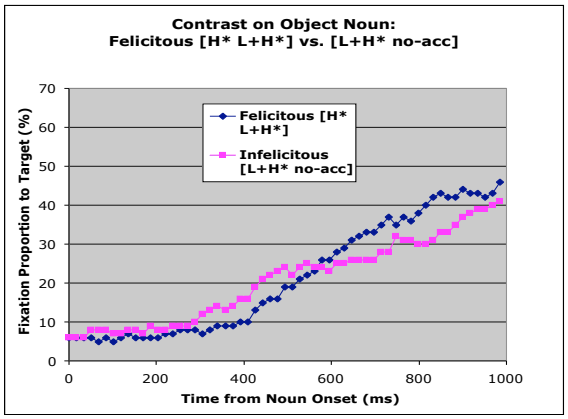


Figure 4: *Contrast on Noun: Felicitous [H* L+H*] vs. Infelicitous [L+H* no-accent].*

The null effect of intonation for Contrast-on-Noun conditions was probably due to the setup of the object manipulation. Since the ornaments were sorted by object type on the grid, the accentual pattern of the repeated color adjective did not serve as informative cue to narrow down the possible candidates for the upcoming noun. This difference between the Contrast-on-Adjective and Contrast-on-Noun trials suggests that listeners can 'tune' their sensitivity to use the prosodic signals that are relevant to the discourse and the task they carry out.

## 3. Experiment 2

In Experiment 1, the infelicitous intonation for Contrast-on-Adjective trials had L+H* on the noun. Therefore, the difference in the timing of increase in fixation proportion may have emerged due to the *delay* in fixations triggered by the infelicitous use of L+H*. To eliminate this possibility, Experiment 2 included trials that simply lacked infelicitous L+H* on nouns. In addition, we tested whether infelicitous L+H* on adjectives could lead to a strong enough anticipation to induce early fixations to incorrect targets. The results confirmed a robust anticipatory effect of L+H* on adjectives.

### 3.1. Design and Materials

Same sets of ornaments as Experiment 1 were used in Experiment 2. Each participant decorated four trees with 26 ornaments on each. In Experiment 2, eye-fixation patterns were compared between felicitous intonation [L+H* no-accent] (e.g., hang the green drum → Next, hang the ORANGE drum) and infelicitous intonation [H* !H*] (e.g., → Next, hang the orange drum). Fixation patterns were also monitored for trials with infelicitous L+H* on the adjectives where the context did not evoke immediate contrast (e.g., Hang the green drum → Next, hang the ORANGE ball). All instructions were newly recorded with the same speaker. The mean duration of words and mean F0 peaks of the target noun phrases in the three critical conditions are presented in Table 2.

Table 2: *Mean duration and F0 of target phrases in Experiment2.*

| Condition | Adj dur (ms) | Adj F0 (Hz) | N dur (ms) | N F0 (Hz) |
|---|---|---|---|---|
| Felicitous Adj Cont | 302 | 354 | 473 | --- |
| Infelicitous Adj Cont | 344 | 225 | 484 | 191 |
| Infelicitous no Cont | 338 | 360 | 482 | --- |

### 3.2. Participants & Eye-tracking procedure

36 undergraduate students at Ohio State University, who were native American English speakers participated as a part of their linguistics course requirement. None of them had participated in Experiment 1. The eye-tracking procedure in Experiment 2 was identical to Experiment 1.

### 3.3. Results

The advantage of felicitous L+H* on the adjective was confirmed in comparison with trials without L+H*. As shown in Figure 5, fixation proportion increased earlier and reached its high plateau of 60% at about 600 ms into the noun for the trials with felicitous intonation, whereas it increased much slower and reached its plateau at about 800 ms in trials where Contrast-on-Adjective was not marked with L+H*. Planned comparison following a one-way repeated measures ANOVA for data in the 300-600 ms window showed significantly more fixations to the target with felicitous L+H* than with H* ($F(1,7) = 38.37$, $p < .001$).

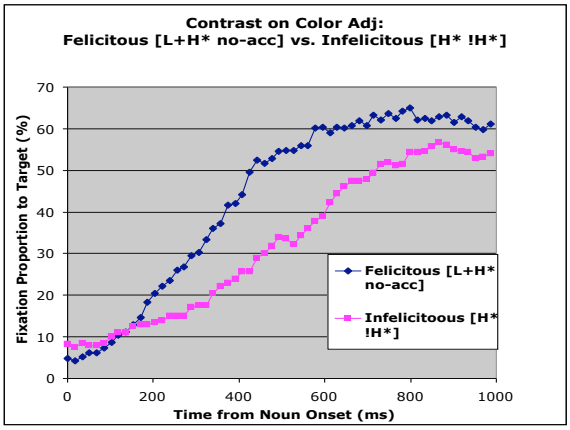


Figure 5: *Contrast on Adjective: Felicitous [L+H* no-accent] vs. Infelicitous [H*!H*].*

The eye-fixation patterns of the trials with infelicitous L+H* on the adjective revealed a robust anticipatory process. Figure 6 compares the fixation proportions to target cell with the proportions of re-fixation of the ornament cell mentioned in

the immediately preceding trial (e.g., 'drum' in the sequence 'green drum → ORANGE ball'). In the presence of infelicitous L+H* on the color adjective (e.g. ORANGE), the re-fixations of the preceding target cell (e.g., drum) started increasing shortly after the onset of the noun. A one-way repeated measures analysis of variance comparing fixation proportions for the 500 ms window beginning at the onset of the noun showed significantly more fixations to the incorrect ornament cell than to the target cell ($F(1,35) = 29.36$, $p < .001$). Note that this early fixation increase took place *before* 200 ms, the time when eye-movements based on the phonetic information of the noun can be executed [4]. That is, participants planned and executed the eye-movements to the false targets (re-fixating the cell mentioned in the immediately preceding trials) solely based on the accentual information of the modifier adjective. Due to these early false fixations, the fixations to the real targets were remarkably delayed.

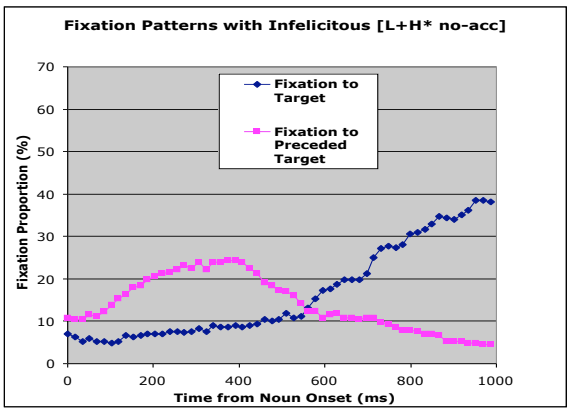


Figure 6: *Infelicitous L+H* on no-Contrast: Target vs. Preceded target*

## 4. Discussion

The results presented here indicate that felicitous intonational prominence can facilitate fixations to target objects, both as compared to infelicitous intonational promience due to mis-placed emphatic L+H* and as compared to infelicitous use of a non-emphatic pitch accent, H*. In particular, when a prominent pitch accent was assigned to a word that conveyed contrastive discourse information, looks to the target occurred sooner and more often than in comparison conditions. These findings confirm and extend previous results showing the immediate processing of emphatic accent on nouns in discourse context [5].

In addition, results show that the presence of an emphatic accent on adjectives in contrastive discourse context produces a robust anticipatory effect. Fixations to target cells in felicitous L+H* conditions were earlier and more frequent. In addition, infelicitous use of the L+H* accent on the adjective preceding a non-repeated noun caused re-fixation to the target cell of the immediately preceding noun. These effects show that listeners can use emphatic accent information in combination with the informational status of a word to predictively determine an upcoming discourse entity, and that such use can be very early in processing, even before the name of the entity (here, the noun information) is encountered.

## 5. Conclusions

The current studies clearly demonstrate the merit of eye-movement monitoring technique for investigating the effect of pitch accent placement on discourse comprehension during visual search. Future research should further examine the scope of intonationally-triggered anticipatory effects like those shown here, specifying the critical factors that may constrain the sensitive "tuning' to intonational cues.

## 6. References


[1] Ito, K & Speer, S. R. (to appear). Using interactive tasks to elicit natural dialogue. In P.Augurzky & D. Lenertova (eds), *Methods in Empirical Prosody Research*, Mouton de Gruyter.

[2] Beckman, M. E., & Ayers, G. M. 1997. *Guidelines for ToBI labelling,* [manuscript]: Ohio State University.

[3] Pierrehumbert, J., & Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 342-365). Cambridge: MIT Press.

[4] Allopenna, P.D., Magnuson, J.S., and Tanenhaus, M. K.1998. Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language,38,*419-439.

[5] 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.

[6] Boersma, P. W., & Weenick, D. (1992-2005). Praat doing phonetics by computer (Version 4.2.17)