LISTENERS' BELIEFS INFLUENCE PROSODIC ADAPTATION: ANTICIPATORY USE OF CONTRASTIVE ACCENT DURING VISUAL SEARCH

Chie Nakamura[1,2], Jesse A. Harris[1], and Sun-Ah Jun[1]

[1]University of California, Los Angeles, [2]Japan Society for the Promotion of Science
chienak@mit.edu

ABSTRACT

This study tested whether listeners adapt to speaker-specific prosody in anticipatory processing, and, if so, whether adaptation is modulated by belief about the speaker's intention. In three visual-world eye-tracking experiments, we compared how listeners responded to proper uses of contrastive accent on an adjective-noun pair (e.g., *First, find the red cat. Next, find the PURPLE*$_{\text{L+H*}}$ *cat*) in Experiment 1 to deviant or improper uses of contrastive accent (e.g., *First, find the red cat. Next, find the PURPLE*$_{L+H*}$ *pig*) in Experiments 2 and 3. Experiment 1 confirmed that proper uses of contrastive accent lead anticipatory looks to the target object. Experiment 2 showed no anticipatory effects of contrastive accent when the adjective accent was deviant. Experiment 3, where participants were informed before the experiment that the speaker was not trustworthy, showed participants learned to anticipate the upcoming referent (*pig*) with improper contrastive accent over the course of the experiment.

Keywords: Prosody, contrastive accent, adaptation. anticipatory processing, eye-tracking

1. INTRODUCTION

There are a growing number of studies reporting the anticipatory use of prosody, in which listeners process prosodic cues rapidly enough to predict a speaker-intended referent, even before encountering disambiguating lexical material [1,2]. These studies support the view that language users draw on general statistical knowledge of prosodic functions in conjunction with other information, such as contextual information and world knowledge, to generate predictions about upcoming linguistic material. In addition, recent work in sentence processing has addressed the language processing change in response to language input, known as *linguistic adaptation.* The adaptation effect has been primarily observed at lower levels of processing, such as phonemic categorization and word recognition [3,4]. However, recent research has found that adaptation is also present at higher levels of representations, such as sentence level syntactic processing [5,6]. The findings suggest that language users update their expectations for future input according to their experience within the experiment. Similarly, studies investigating *prosodic adaptation* indicate that listeners down-weight prosodic cues when the information they convey is not reliable [7,8]. Listeners also appear to adjust the degree to which they use boundary tone information in proportion to how informative the cue is in the experiment [9].

An important remaining research question concerns whether and how listeners make use of deviant or improper prosodic marking, and whether the outcome of prosodic adaptation is modulated by listeners' beliefs about the speaker's intention. On the one hand, listeners might simply put less weight on prosody in making structural judgement with improper prosodic marking. On the other, listeners might learn speaker-specific prosodic patterns to make anticipatory judgments about the intended meaning. In order to address these possibilities, the current study manipulated the contrastive accent marking (L+H*) and the listener's belief about the speaker's intention in three experiments.

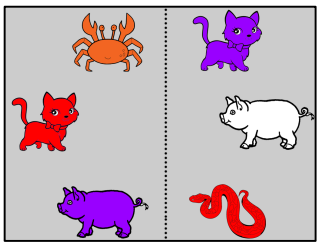
In a series of visual-world eye-tracking experiments [10], we compared how listeners responded to proper and improper uses of contrastive accent on an adjective-noun pair. Experiment 1 tested whether proper use of contrastive accent (1; *red cat* followed by *PURPLE cat* with Fig. 1) leads to anticipatory looks to a target object. Experiment 2 tested how listeners respond to the "improper" use of contrastive accent (2; *red cat* followed by *PURPLE pig* with Fig. 1). Using exactly the same materials as in Experiment 2, participants in Experiment 3 were informed before the experiment that the speaker was not trustworthy. This manipulation allowed us to explore whether listeners' awareness of the speaker's intention influences how much they adapt to the speaker-specific deviant use of contrastive accent.

Proper contrastive accent (Experiment 1)
(1) *First, find the red cat.*
Next, find the PURPLE$_{\text{L+H*}}$ *cat.*

Improper contrastive accent (Experiments 2 and 3)

(2) *First, find the red cat.*
Next, find the PURPLE $_{L+H*}$ *pig.*

Figure 1: Visual array presented with (1) and (2)



2. EXPERIMENTS

2.1. Experiment 1

In Experiment 1, we tested whether proper use of contrastive accent (L+H*) on the adjective (1b) leads to anticipatory looks to a target object (e.g., a purple cat) as compared to the control condition that had new information (H*) accent (1a).

(1) *First, find the red cat.*
a. *Next, find the purple*$_{H*}$ *cat.*
b. *Next, find the PURPLE*$_{L+H*}$ *cat.*

We predicted that listeners use contrastive accent to predict an upcoming referent during visual search. We expected to observe anticipatory eye movements to the target object prior to the onset of the target noun when the sentence had contrastive accent on a contrastive adjective (1b) compared to the control condition with new information (H*) accent (1a).

2.1.1. Participants

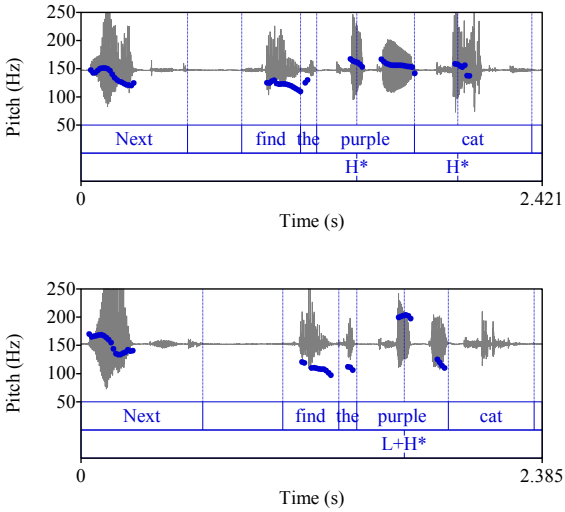
Thirty-six native speakers of English with unimpaired vision and hearing participated in the experiment for course credit.

2.1.2. Stimuli

Thirty-six experimental items were created. Each item consisted of a sound file of a sentence and a corresponding visual scene. The auditory stimuli were recorded by a male native speaker of English trained in English ToBI [11]. Figure 2 shows the F0 contours of the sentence (1) in each condition.

The visual scenes were prepared using clip art images. The position of the objects was counter-balanced across the items. Two experimental lists were created following the Latin square design including 36 target items and 54 filler items. The 90 items in each list were presented in a pseudo-random order.

Figure 2: Waveform, pitch track, and accent type for (1a, top) and (1b, bottom).



2.1.3 Procedure

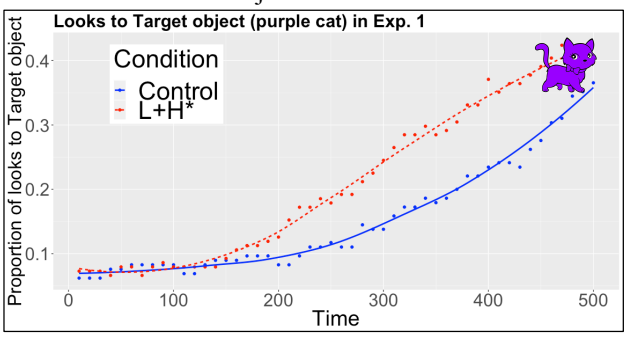
Participants were told to listen to the sentences carefully while attending to the picture on the computer monitor. As soon as the sentence finished, participants responded by pressing the left or right bumper on a gamepad that corresponded to the position of the second-mentioned object. In each trial, the sentence was presented 3000ms after the picture onset of the computer screen. Participants' eye-movements around the screen were recorded with the EyeLink 1000 Plus (SR Research) at a sampling rate of 500 Hz. Each experimental session lasted approximately 30 minutes.

2.1.4. Data analysis and results

We summed the gazes to each entity in the scene and calculated the logit of looks to each entity out of looks to all the objects in the scene [12], including the background. Statistical analyses for the duration of anticipatory time window (from the onset of the color adjective until the minimum onset of the target noun) were conducted using linear mixed-effect regression models [13]. We included Prosody (with L+H* or H*) and Trial Order (1 to 90) as fixed effects, as well as their interaction. Participants and items were included as random factors.

Figure 3 shows the proportion of looks to the target object from the onset of the color adjective. There was a main effect of Prosody (β=2.37, SE=1.05, t=2.25, $p<0.05$); more looks to the target object were observed with contrastive accent than without it. Crucially, the effect was observed before the onset of any disambiguating information. As predicted, participants used contrastive accent to program an eye-movement in anticipation of the upcoming word that was most likely to contrast with the previous word in the visual array, replicating the results of previous research [1].

Figure 3: Proportion of looks to the target object from the onset of the color adjective to 500ms.



2.2. Experiment 2

Previous studies testing the adaptation effect in prosodic processing have shown that listeners track how informative or reliable prosodic cues are, and adjust the extent to which they use prosody in structural analysis [7-9]. For example, Nakamura, Harris, and Jun [9] showed that listeners placed less weight on prosodic information in structural decision while processing target items when filler items in the experiment had a boundary tone that did not align with a syntactic unit. However, it is not yet clear whether listeners can learn the speaker-specific way of deviant but consistent prosodic marking patterns to make anticipatory judgements. In order to address this question, Experiment 2 tested how listeners responded to the improper use of contrastive accent (2b) compared to a control condition with new information (H*) accent (2a). The same visual scenes as in Experiment 1 were used in Experiment 2.

(2) *First, find the red cat.*
   a. *Next, find the purple*$_{H*}$ *pig.*
   b. *Next, find the PURPLE*$_{L+H*}$ *pig.*

If listeners adjust the degree to which they use prosodic cues based on how informative the prosody is in the experiment, listeners should put less weight on contrastive accent information in anticipatory processing in this experiment. Thus, we expected that participants in Experiment 2 would make less anticipatory eye-movements compared to the participants in Experiment 1.

2.2.1. Participants, Stimuli, and Procedure

Thirty-five native speakers of English participated in the experiment. As in Experiment 1, two experimental lists were created including 36 target items and 54 filler items, recorded by the same speaker as in Experiment 1. The procedure in Experiment 2 was identical to that in Experiment 1.

2.2.2. Data analysis and results

Analyses on fixations to each entity were conducted as in Experiment 1. The results for the looks to the target object (e.g., purple pig) and those to the competitor object (e.g., purple cat) showed no effect of Prosody or Trial Order. In order to compare the difference between Experiment 1 and 2, we conducted a combined analysis between the two experiments and compared the looks made to the object that is predicted by the proper use of contrastive accent (e.g., purple cat). There was a main effect of Experiment (β=6.38, SE=3.05, t=2.09, p<0.05), indicating that upon hearing the color adjective, participants in Experiment 2 looked less often to the object that contrasted with the previously mentioned object compared to participants in Experiment 1. However, no evidence for learning of the improper use of contrastive accent was observed.

2.3. Experiment 3

Previous studies found that listeners quickly recruit speaker-based knowledge in reference resolution [14]. Grodner & Sedivy [15] have further shown that when listeners had reason to believe the speaker in the experiment to be an unreliable communicative partner, they did not use a restrictive adjective modifier as an informative cue in determining the referent in contrastive context. In Experiment 3, we examined the possibility that listeners' beliefs about the speaker help them learn the speaker's specific use of deviant prosodic marking. Using exactly the same materials as in Experiment 2, participants in Experiment 3 were told before the experiment that they were going to play a game with "a devious opponent who will try to trick them". This manipulation allowed us to explore whether the adaptation process would be different depending on whether or not listeners are informed about the speaker's intention, and whether listeners use the speaker's deviant but consistent way of mapping between prosody and meaning in anticipatory processing. Thirty-three participants participated in Experiment 3.

2.3.1. Data analysis and results

Analyses on fixations to each entity were performed in the same manner for the same anticipatory time window as in Experiments 1 and 2. Figure 4 shows the proportion of looks to the target object from the onset of the color adjective. The results for the looks to the target object (e.g., purple pig) revealed an interaction between Prosody and Trial Order (β=1.45, SE=7.38, t=1.97, p<0.05). Participants learned how to predict the correct upcoming referent with improper use of contrastive accent, as they heard more tokens in the experiment (Figure 5).

Figure 4: Proportion of looks to the target object from the onset of color adjective to 500ms.

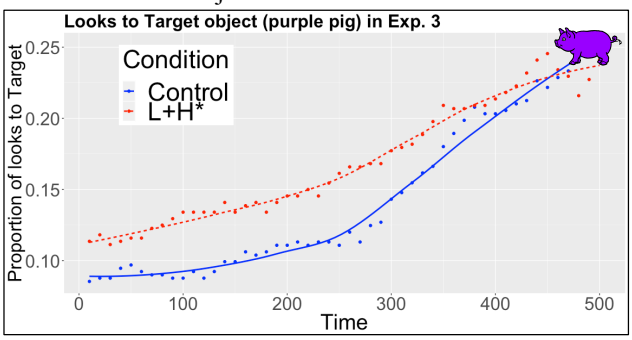
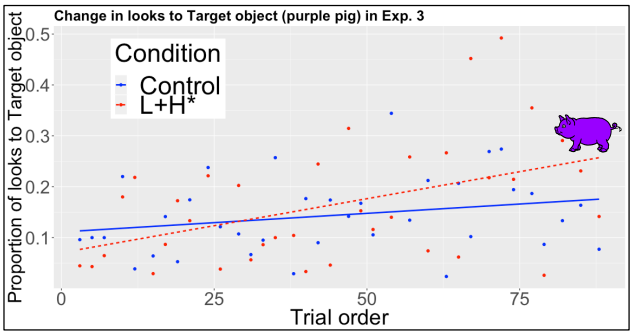


Figure 5: Change in looks to the target object from trial 1 to 90.



In order to compare the difference in looks to the target object (purple pig) between Experiment 2 and 3, we conducted a combined analysis between the two experiments. The analysis revealed a marginal three-way interaction between Prosody, Trial Order, and Experiment (β=1.43, *SE*=7.45, *t*=1.925, *p*=0.054), suggesting that learning effect of improper use of prosody was significant only in Experiment 3. The results further indicate that when listeners were aware that the speaker was uncooperative, they learned the speaker's deviant use of contrastive accent prosody and ascribed a meaning to the speaker-specific prosody.

3. DISCUSSION

The current study examined whether listeners adapt to speaker-specific use of prosody in anticipatory processing. We also tested whether the rate and outcome of adaptation is modulated by the listener's belief about the speaker's intention. The results of Experiment 1 showed that participants used contrastive accent to anticipate the upcoming word that was most likely to contrast with the previous word, given the visual array. The results of Experiment 2 showed that when the use of contrastive accent was unconventional or deviant, participants did not make anticipatory eye-movements to the target object nor to the competitor object. This suggests that when a reliable association between prosody and structure is broken, listeners resist making predictions based on prosody. The results of Experiment 3 showed that when listeners were aware that the speaker was uncooperative, they quickly identified the unconventional use of contrastive accent and learned how to predict the upcoming referent with improper contrastive accent over the course of the experiment.

The results of the current study provided evidence that listeners immediately used contrastive accent information to restrict the possibilities for subsequent reference in the visual context, supporting the view that the human language comprehension system integrates multiple sources of information in forming a representation of the sentence in a highly incremental and efficient manner. The results further show that listeners' beliefs about the speaker modulated the degree of adaptation to the speaker's specific way of unconventional prosodic marking.

The results provide important implications for the rate at which listeners adapt to their interlocutors [6,16]. Comprehenders appear to rapidly adjust their expectations for upcoming information as they receive more information about the speech and goals of the speaker, as well as the propensity to use particular structures [see also 17,18,19]. The current study finds that when prosody is used in a non-standard way, listeners need more than input alone in order to adapt to the speaker-specific deviant use of prosody in anticipatory processing.

It is known from previous studies on prosodic adaptation that the comprehension system reduces the informational value of prosodic cue when it is unreliable or inconsistent. The current study provided evidence that listeners adjust to the speaker's specific mapping between prosody and meaning, allowing them to anticipate an upcoming referent on the basis of improper, but consistent, contrastive accent. In all, our results support the view that online interpretation of prosodic information is modulated by how informative prosodic cues are, and further showed that the assessment of the speaker's communicative intentions influences the extent to which listeners adjust to non-conventional uses of such cues.

4. ACKNOWLEDGEMENTS

We thank Canaan Breiss for help recording the stimuli for the experiment, and the undergraduate research assistants at the UCLA Language Processing Lab for administering the experiment: Katie Barnett, Janice Chen, Anthony Chadwick, Durgesh Rajandiran, Lalitha Balachandran, Arjun Gananathan, Delaney Warren, Livia Witteveen, Pranav Singh, Chenchen Wang, Joonhwa Kim, and Rebecca Wu. We received valuable feedback on this project from the UCLA Psycholinguistics Seminar. This work was supported by JSPS KAKENHI Grant-in-Aid for Young Scientists (A), grant number 15H05381.

5. REFERENCES

[1] Ito, K. Speer, S. R. 2008. Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58, 541–573.

[2] Snedeker, J., Trueswell, J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48, 103–130.

[3] Norris, D., McQueen, J. M., Cutler, A. 2015. Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience* 31, 4–18.

[4] Kraljic, T., Samuel, A. G. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review* 13, 262–268.

[5] Kaschak, M. P., Glenberg, A. 2004. This construction needs learned. *Journal of Experimental Psychology: General* 133, 450–467.

[6] Fine, A. B., Jaeger, T. F., Farmer, T., Qian, T. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE, 8*.

[7] Tanenhaus M., Kurumada C., Brown M. 2015. Prosody and Intention Recognition. In: Frazier L., Gibson E. (eds) *Explicit and Implicit Prosody in Sentence Processing*, Springer, 99-118.

[8] Roettger, T., Franke, M. 2018. Dynamic speech adaptation to unreliable cues during intonational processing. To appear in *Proceedings of 40th Annual Conference of the Cognitive Science Society.*

[9] Nakamura, C., Harris, J.A., Jun, S.-A. 2018. Cue reliability affects anticipatory use of prosody in processing globally ambiguous sentences. CUNY presentation. UC Davis.

[10] Altmann, G.T.M., Kamide, Y. 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264.

[11] Beckman, M., Hirschberg, J., 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*. Oxford: Oxford University Press, 9-54.

[12] Barr, D. J. 2008. Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language* 59, 457–474.

[13] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390–412.

[14] Hanna, J.E., Tanenhaus, M.K., Trueswell, J.C. 2003. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language* 49, 43–61.

[15] Grodner, D., Sedivy, J. 2011. The effect of speaker-specific information on pragmratic inferences. In: Pearlmutter, N., Gibson, E. (eds). *The Processing and Acquisition of Reference.* Cambridge, MA: MIT Press, 239–272.

[16] Kamide, Y. 2012. Learning individual talkers' structural preferences. *Cognition* 124, 66–71.

[17] Branigan, H.P., Pickering, M.J., Cleland, A.A. 1999. Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review* 6, 635–640.

[18] Chang, F., Dell, G.S., Bock, K. 2006 Becoming syntactic. *Psychological Review* 113, 234–272.

[19] Kaschak, M.P., Loney, R.A., Borreggine, K.L. 2006. Recent experience affects the strength of structural priming. *Cognition* 99, B73–B82.