Boundary tone processing during online comprehension

S. Bibyk* (Vanderbilt Univeristy), W. Heeren (Utrecht University), A. Albin (Kobe University), W. Rankinen (GVSU), A.P. Salverda, C. Gunlogson, & M.K. Tanenhaus (University of Rochester)

Intonational contours play a central role in real-time spoken language processing[1,2,3]. In the ToBI framework, a minimal contour includes a pitch accent, a phrasal accent, and a boundary tone. Boundary tones by themselves play a crucial role in signaling a speaker's intended meaning, for example indicating the difference between a statement (low tone) and a question (high tone)[4]. However, a high tone does not exclusively signal a "question" meaning (e.g. consider uptalk[5]), and thus it is not yet clear how boundary tone information is processed in conjunction with other prosodic, lexical, syntactic, and discourse cues to the speaker's intention. Studying the online processing of boundary tones is also challenging because they occur in conjunction with many other cues to the speaker's intention (e.g. subject-auxiliary inversion in yes-no questions) and their acoustic realization (e.g., changes in F0) may be distributed over long stretches of speech. In order to examine how information about boundary tones is identified, used, and combined with other prosodic and non-prosodic information in sentence processing, we developed a visual-world targeted language game[6]. This paradigm demonstrated that eye movements can be used to determine when and how listeners integrate rising and falling intonation with other cues to the speaker's intention.

In the current experiment we ask *when* during the processing of the acoustic signal listeners distinguish rising from falling intonation: is it at the pitch accent, the boundary tone, or after the full contour? Participants (n=24) played a go-fish-like card game against a computer. The goal of the game was to get rid of one's hand of "playing cards" (Figure 1, bottom left) by matching them to the card displayed in the center of the screen. Whenever the computer declared a match (e.g., "Got an armadillo."), the participant looked at their "blocking card" (Figure 1, bottom right), to try to block the match. If none of the computer's playing cards matched the center card, it asked the participant to make a match instead (e.g. "Got an armadillo?"), prompting the participant to look at their playing cards. Participants thus were required to fixate on different objects depending on whether the computer asked a question or made a statement. On critical trials, the computer used the elliptical "Got an X" construction, which can be interpreted as either a *question* or a *statement*, but was distinguished by the boundary (high for questions, low for statements). Four 4-syllable words with penultimate stress (*armad'illo, baller'ina, orig'ami*, and *ravi'oli*) were used as target words. As seen in Figure 2 (colors representing each of the target words), the pitch contours for the critical utterances had flat pitch over the first four syllables, with reliable cues distinguishing the two contours appearing on the final stressed syllable (e.g. the "dil" in armadillo). The stressed syllables contained identifiable *turning points* (an elbow in the pitch from the accent before the final rise, and a peak in the accent before the final fall), allowing us to specify the earliest reliable acoustic information that marked the transition between the nuclear accent and the boundary tone.

We established in previous work[6] that the "Got an X" construction is biased in favor of a question interpretation, and we thus focused our analyses on the *statements* where the boundary tone was predicted to override this initial question interpretation. The plot in Figure 3 shows that, after the initial bias to fixate the playing cards (the competitor, dotted red line), participants then shift to fixating the blocking card (the target, solid red line) approximately 50 ms *after* the turning point (confirmed by a breakpoint analysis selecting the model with the lowest BIC). Thus, we argue that listeners make inferences about the speaker's intention to ask a question or make a statement at least as early as the onset of the boundary tone, and possibly during the pitch accent. This research opens the door to many questions with regard to how intonational contours are processed on-line, allowing us to ask when listeners combine boundary tone information with expectations based on prior lexical, syntactic, and prosodic information.

*sarah.a.bibyk@vanderbilt.edu

[1] Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*, 292–314.
[2] Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *Journal of Memory and Language, 58*, 542–573.
[3] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & 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.
[4] 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. 271–311). Cambridge, MA: MIT Press.
[5] McLemore, C. A. (1991). *The pragmatic interpretation of English intonation: Sorority speech* (Doctoral dissertation). University of Texas Austin.
[6] Heeren, W. F. L., Bibyk, S. A., Gunlogson, C. and Tanenhaus, M. K. (2015). Asking or telling—Real-time processing of prosodically distinguished questions and statements. *Language and Speech, 58* (4), 474-501

Figure 1

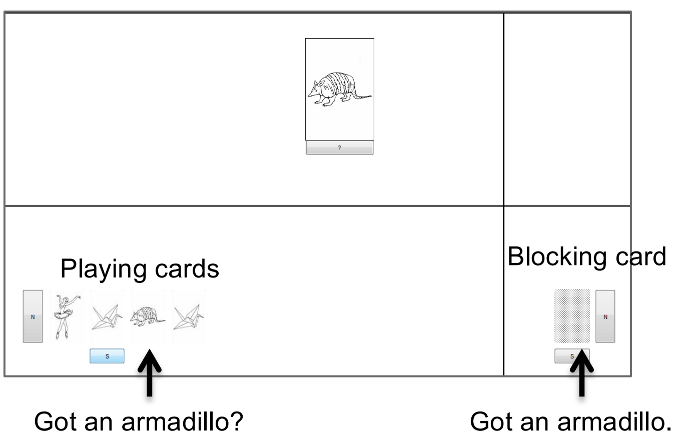

Figure 2

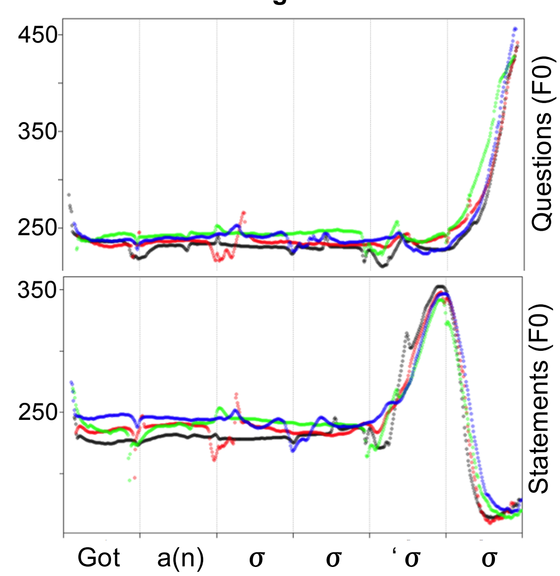

Figure 3