Nice to *hear* you again: Listeners maintain prosodic adaptation over 7 days

Amelia Ostrow, Chigusa Kurumada (University of Rochester)

One of the persistent puzzles for theories of intonational phonology and its processing is **the phonetic variability of speech sounds** [1-3]. Even a seemingly straightforward distinction (e.g., question vs. statement) exhibits substantial variability conditioned on physiological and socio-indexical features (e.g., talker's body size, gender, age, dialect) (**Fig.1**). Recent studies have demonstrated that listeners are nevertheless surprisingly good at deciphering intonational meaning, highlighting the role of the rapid **adaptation** to cross-talker variability [4,5]. However, this mechanism will be of little use if listeners cannot retain the result of the adaptation and have to start the process all over again for the same speaker. Here, we investigated whether, and if so for how long, listeners are able to maintain their adapted prosody-meaning mapping across encounter episodes.

Experiment: A male talker in his 20's recorded pairs of statement (e.g., *It's cooking.*) and question (*It's cooking?*) utterances in American English. We manipulated the F0 and syllable duration so intonation contours gradually shifted from the typical falling to the rising tone values (**Fig.2**). A resulting 11-step continuum was then normed by 120 native speakers of American English to determine the maximally ambiguous stimulus (i.e., for which listeners were least certain about the intended meaning being a statement or a question).

[Adaptation] Subjects (N=368) were randomly assigned to one of the 6 between-subject conditions (2 training conditions * 3 delay conditions). In Pre-Test, all subjects heard 44 instances of one item type (e.g., *it's cooking*) sampled uniformly from the Statement-Question continuum (twice per step), and provided 2AFC (Question vs. Statement) judgments. In Training, subjects continued to make 2AFC judgments in the same format and received feedback: Subjects in the **Question-biasing** condition heard prototypical Statements (i.e., Step 1) and the acoustically ambiguous item (Step 6) disambiguated as Questions. In contrast, those in the **Statement-biasing** condition heard the prototypical Questions (Step 11) and the acoustically ambiguous items (Step 6) identified as Statements **Fig.3A**). Finally, in Post-Test, all subjects repeated the same procedure as in the Pre-Test (44 trials). There was no lexical overlap between the Training and Tests. All subjects were returned to the 2nd session after a delay of 3, 5, or 7 days and repeated the same procedure.

Results from the 1st session revealed that the two groups of subjects in the Question-biasing vs. Statement-biasing conditions provided opposite interpretations for the previously ambiguous items after the first training (p<.001, **Fig.3B**); i.e., their prosodic categorization adapted to the prosody-meaning mapping in the expected directions. Critically, **the adaptation was maintained in all three delay conditions**: Pretest responses in the 2nd session were already tuned in the trained direction (**Fig.3C**). The adaptation thus withstood the large number of talkers and the countless number of intervening utterances that subjects must have encountered during these delay periods. 7 days is among the longest periods tested for the retention of speech adaptation [6].

Our novel results thus support that prosodic processing is not only flexible, but it is also impressively stable over time. Just 6 minutes of exposure powerfully recalibrated the prosody-meaning mapping, which lasted for over a week. The ability to store talker-specific adjustments can also help explain how cumulative exposure to a given talker (and talker group) can lead to longer-term processing benefits (e.g., familiar voices and accents are processed more easily and accurately) [7]. We will discuss possible memory mechanisms underlying the long-term maintenance of prosodic adaptation as well as their links to L2 prosody acquisition and learning.
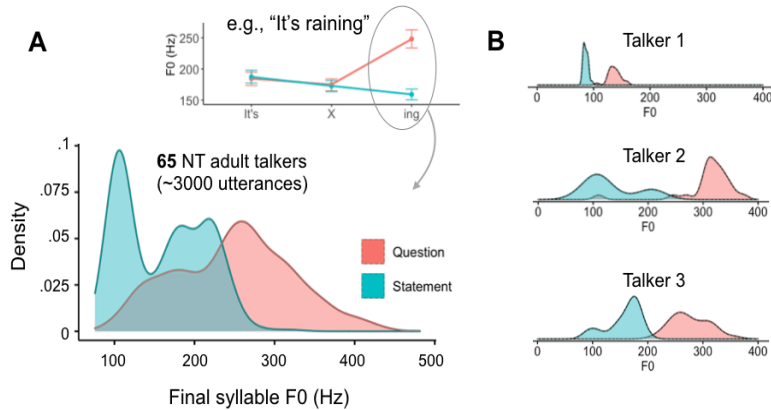
Fig. 1. A. Substantial category overlap in phonetic cue distributions from 65 native talkers (~3000 tokens), based on data from Xie et al., (2021) [5]. **B**. While better separation of question vs. statement observed within each talker, no simple baseline corrections can fully resolve the across-talker variability. This motivates the importance of adaptation in comprehension.
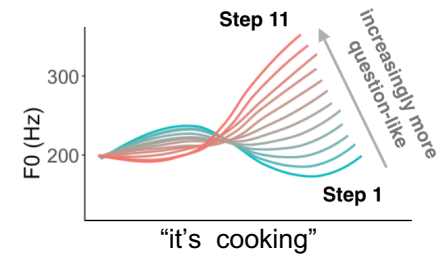
Fig. 2. Fundamental frequency (F0) and duration of an 11-step continuum of the stimuli. Step 1 and Step 11 were produced by a male native speaker of American English.

Fig. 3. Procedure and results of the experiment. **A.** Subjects completed the pre-test, training, post-test procedure twice with a delay of 3, 5, or 7 days in between. **B.** In each day, subjects in the question-biasing (pink) and the statement-biasing (turquoise) conditions showed significant adaptation. Here we show an example from Session 1 of the 7-day delay condition. The results from the 3-day and 5-day delay conditions were very similar. Error bars indicate 95% confidence intervals. **C.** Adaptation was maintained for up to 7 days.

[1] Ladd, D. R. (2008) *Intonational phonology*, CUP. [2] Arvaniti. A & Garding, G. (2007). Dialectal variation in the rising accents of American English, *Lab phone*. [3] Clopper, C. & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *J of Phonetics*. [4] Xie, Buxó-Lugo & Kurumada, (2021). Encoding and decoding of meaning in intonational speech prosody. *Cognition*. [5] Roettger, T. & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. *Cog. Sci.* [6] Witteman, M.J. et al., (2014). Automaticity and stability of adaptation to forein-accented speaker. *Language and Speech*. [7] Magnuson, J. et al., (2021). Talker familiarity and the accommodation of talker variability. *Attention, Perception, and Psychophysics*.