**Supervised and unsupervised learning of linguistic cue distributions** Dave F. Kleinschmidt & T. Florian Jaeger (University of Rochester) [dkleinschmidt@bcs.rochester.edu](mailto:dkleinschmidt@bcs.rochester.edu) *Keywords go here*

During language acquisition, people learn the probabilistic mapping between observable cues and underlying linguistic structures. Recent work has shown that this learning continues into adulthood, with people rapidly adapting their linguistic expectations at many different levels in response to experience (with, e.g., syntactic structures, phonetic categories, pragmatic interpretations, etc.). While there are many similarities between the sort of learning that occurs during acquisition and adaptation in adults, there is one major difference: acquisition is *unsupervised*, while adaptation is *supervised*, at least as it's typically studied. During adaptation, people have many sources of information that *label* each observed cue value, such as lexical context or audio-visual input (for phonetic adaptation) or sentential contexts that are not globally syntactically ambiguous (for syntactic expectation adaptation). People do take advantage of such top-down labels during *processing*, integrating them with bottom-up cues to change the inferences they make (cf. the Ganong and Mcgurk effects, **citation for mike stuff**). However, it is not known whether people additionally take advantage of them during adaptation, or whether adaptation is purely a bottom-up learning process.

To investigate the extent to which people take advantage of labels during linguistic adaptation, we used a phonetic adaptation task that could be easily adapted to be either supervised or unsupervised. Subjects heard spoken words, all members of /b/-/p/ minimal pairs (beach/peach, bees/peas, and beak/peak) synthesized with VOTs ranging from -20ms to 90ms in 10ms steps. On each trial, two pictures portraying one /b/ word and one /p/ word were shown, and subjects were instructed to click on the picture that matched the word they heard. There were two types of trials. On *unlabeled* trials, both response options were members of the same minimal pair. On *labeled* trials, one response picture matched the minimal pair of the stimulus, and the other did not. For instance, for a stimulus from the beach-peach continuum, an unlabeled trial would have pictures of a beach and a peach, while a labeled trial might have a picture of a beach and pees. Each subject was assigned to either the *unsupervised* condition, where all trials were unlabeled, or the *supervised* condition, where half were labeled and half unlabeled. The particular VOT values that each subject heard were determined by one of four bimodal distributions. These distributions were distinguished only by the mean VOT values of their high and low VOT clusters (corresponding to /p/ and /b/). These were: 0/40ms, 10/50ms, 20/60ms, and 30/70ms, with implied b/p category boundaries of 20ms, 30ms, 40ms, and 50ms, respectively. The lowest distribution is similar to natural distributions, while the highest distribution is rather unusual.

Listeners learned in all conditions, showing higher category boundaries (measured only on unlabeled trials) in higher-VOT conditions. Learning was nearly perfect in the 20ms and 30ms boundary conditions, but in the 40ms and 50ms boundary conditions listeners' category boundaries were substantially lower than expected based on the distributions they heard. In the supervised condition, listeners consistently used the label information provided on label trials to guide their responses (98% consistent with label). However, learning was neither more complete nor faster in the supervised than in the unsupervised condition, and in fact by one measure (category boundary slope) listeners in the supervised condition performed *worse* (shallower boundaries) than in the unsupervised condition.

The most straightforward interpretation of these results is that the label information we provided to subjects is not available to or used for the learning processes that result in adaptation, and that these processes are primarily bottom-up. This is surprising given that labels are highly informative about the statistical structure of the environment, and substantially reduce the difficulty of the learning problem. However, we hesitate to conclude this for two reasons. First, we provide labeling information through the context, or task structure, rather than as part of the linguistic signal per se. In other studies of phonetic adaptation, labels are typically provided by the linguistic signal itself, either lexical or audio-visual. Second, the cue distributions subjects heard may have been either too far outside the normal range to be adapted to at all (very high mean VOT values), or adapted to so quickly that no effect of supervision would have been detectable.