One of the longest-standing puzzles in human speech perception is how listeners manage to cope with the often extreme differences in how individual talkers use acoustic cues to realize their linguistic intentions. A number of solutions have been proposed, including the recent proposal that listeners quickly *adapt* to unfamiliar talkers by *learning* the distributions of acoustic cues that they produce (their "accent") (Kleinschmidt & Jaeger, 2015).

This can be formalized as a kind of statistical inference, where listeners try to infer which of all possible accents best explains a talker's speech. Prior experience is helpful because it narrows down the range of possibilities that a listener needs to consider (in Bayesian jargon, it provides an *informative prior* on accents). We test a critical prediction of this view: when an unfamiliar talker's accent falls *outside* the range of typical variation across talkers, listeners should adapt only partially. Specifically, listeners' phonetic classifications should reflect a compromise between listeners' prior expectations and the actual accent they hear. We also, in doing so, demonstrate a novel technique for measuring listeners' subjective prior expectations about an unfamiliar talker's accent.

We use a /b/-/p/ VOT distributional learning paradigm (Clayards, Tanenhaus, Aslin, & Jacobs, 2008), where listeners hear a bimodal distribution over an acoustic cue (VOT), with a cluster at a low value implicitly corresponding to /b/ and another at a high value corresponding to /p/. By varying the location of these clusters, we create "accents" that are more or less like those produced by a typical American English talker (as measured by, e.g., Kronrod, Coppess, & Feldman, 2012).

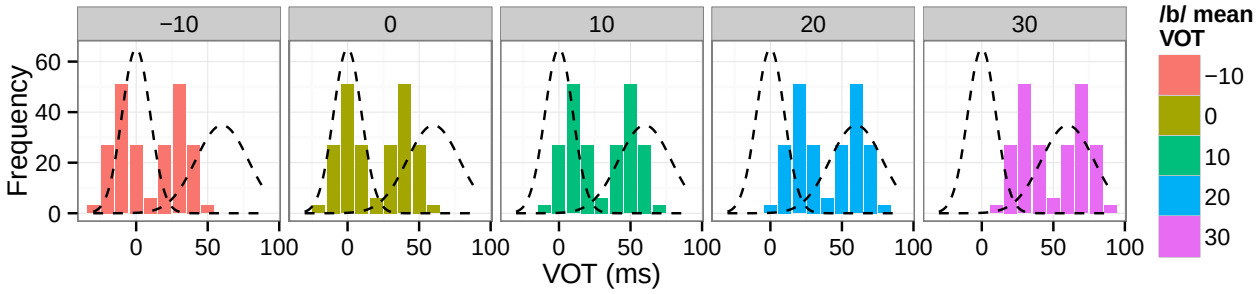

Figure 1: Listeners heard one of these five synthetic "accents", which differed only in the location of (implied) /b/ and /p/ clusters of VOTs (colored histograms) relative to a typical English talkers' VOT distributions (dashed black lines).

We measure how well listeners *learn* these accents by comparing their classification functions to the ideal boundaries implied by the distributions (as in Clayards et al., 2008). As predicted, when the VOT clusters were unusually high or low, listeners actual category boundaries reflected a compromise between the boundary of a typical talker and the boundary implied by the input distributions they heard.

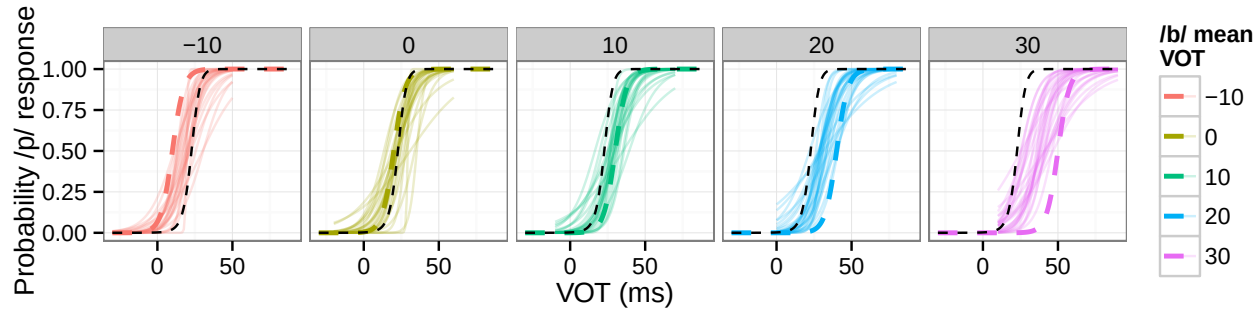

Figure 2: After exposure, listeners' /b/-/p/ classifications (thin lines) reflected a compromise between the typical (dashed black) and experimental (dashed colored).

Second, we used a Bayesian belief-updating model to work backwards from the patterns of adaptation to different accents in order to infer what listeners' starting beliefs were, and how confident they were in those

beliefs. The inferred prior expectations matched the range seen across typical American English talkers, including the counterintuitive fact that listeners were *more* uncertain about the /b/ mean VOT than the /p/, corresponding to the fact that there's high variance in the VOT of /b/ *across* talkers due to some talkers prevoicing.

The ability to measure listeners' prior expectations potentially provides an important and heretofore missing tool in the toolbox of sociophonetics: it directly links the measurable variability in *production* of linguistic variables with listeners' subjective expectations about those variables, in both cases potentially conditional on *social* variables. We demonstrate a proof-of-concept here using standard American English as a reference, but the same procedure can be applied to more specific variables like gender, region, class, etc., by providing information to the listener about *who* the talker is (information which listeners do in fact use to guide speech perception, Hay & Drager, 2010; Niedzielski, 1999).

References

Clayards, M. A., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. a. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–9. doi:10.1016/j.cognition.2008.04.004

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865–892. doi:10.1515/ling.2010.027

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. doi:10.1037/a0038695

Kronrod, Y., Coppess, E., & Feldman, N. H. (2012). A Unified Model of Categorical Effects in Consonant and Vowel Perception. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 629–634). Austin, TX: Cognitive Science Society.

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. doi:10.1177/0261927X99018001005