Are you paying attention? Investigating the extent of automatic perceptual recalibration

Pronunciation is variable: even talkers with similar language backgrounds will produce the same phonological categories differently, for example the amount of spectral energy they use to differentiate /s/ from /ʃ/. Listeners' perception will often flexibly accommodate this variation. Our ability to recalibrate the perceptual boundaries between phonemes may appear surprisingly automatic: research on lexically guided perceptual recalibration (PR) suggests that recalibration is not inhibited by distractions [1], lack of intention [2], or exposure to multiple talkers [3, 4]. Here we explore the extent of this automaticity. Previous work has concluded that PR is automatic provided lexical access takes place [5]. We test this prediction by exposing listeners to simultaneous speech from two distinct talkers, while asking listeners to attend to only one of the talkers—an extreme version of the cocktail party problem.

Experiment (n=60 participants). Following the structure of a typical PR experiment, we first exposed listeners to speech from unfamiliar talkers (*Figure 1*) through a series of two alternative forced-choice (2AFC) word recognition tasks, then measured the listeners' perception via 2AFC discrimination tasks during a subsequent test phase (*Figure 2)*. However, unlike typical PR experiments, each exposure trial consisted of isolated word recordings from *two different simulated talkers* (labelled as male and female in the experiment) *played at the same time* (one talker presented in the left ear and one talker in the right ear; talker-to-ear-assignment was balanced across trials). Participants were asked to always attend to the same talker (e.g., always the female talker). In the exposure phase, we manipulated whether the talker produced ʃ-biased (producing /s/ as /ʃ/-like) or S-biased (producing /ʃ/ as /s/-like) words between participants. Previous studies found PR when the two audial streams partially overlap (unattended talker's speech started with speech onset asynchrony [SOA] of 200ms after the attended talker's speech [5]; our SOA is 0ms). In the test phase, we assessed listeners' PR via *asi-ashi* phonetic continua produced by both talkers. Each test trial presented a recording from *either* the female or male talker. Each trial block contained a full six-step *asi-ashi* continuum sampled once without replacement in random order. The talker in the recording alternated every 2 blocks, for a total of 12 blocks (6 blocks per talker).

Results. Mixed-effects logistic regression of participants' *ashi*-responses during test found no significant PR—neither for the attended, nor for the unattended talker (*p*s > 0.05, *Figure 2)*. This stands in contrast to previous work which found PR for dual talker recordings with only 200ms SOA [5]. Critically, participants' lexical decision accuracy for the attended talker during exposure was above chance in our experiment (≥80% accuracy). While this performance is lower than in PR experiments with single-talker recordings (~85-95% accuracy), it suggests that lexical access took place on most or all exposure trials, without leading to PR. Notably, previous work has found PR with single-talker recordings with as few as half of the exposure tokens we used [6]. Additionally, our experiment had 50% more participants than previous work that reliably found the effect at longer SOAs.

Conclusion. Previous work has concluded that PR is automatic provided lexical access takes place. The present findings are at odds with that: attentional resources might, after all, mediate PR even when lexical access is successful—contrary to [5].

References. **[1]** Zhang, X., & Samuel, A. G. (2014). *JEP: HPP* **[2]** McAuliffe, M., & Babel, M. (2016). *JASA*. **[3]** Cummings, S. N., & Theodore, R. M. (2022). *Attention, perception & psychophysics*. **[4]** Kraljic, T., & Samuel, A. G. (2007). *JML*. **[5]** Samuel, A. G. (2016). C*ognitive Psych.* **[6]** Cummings, S. N., & Theodore, R. M. (2023).

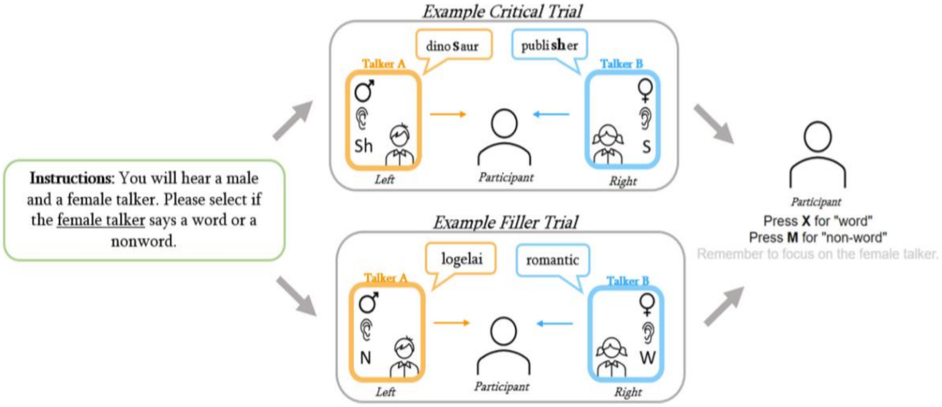

Figure 1**: Procedure during exposure phases**. Participants listened to recordings that played single word utterances from two talkers. All exposure trials consisted of a stereo recording of a female and a male talker. Participants were asked to either always attend to the male (♂) or always to the female (♀) talker, counterbalanced across participants. Critical trials (top) consisted of one ʃ-word and one S-word, while filler trials (bottom) consisted of a word and a non-word. Participants responded whether the *attended* talker produced a word or a nonword.

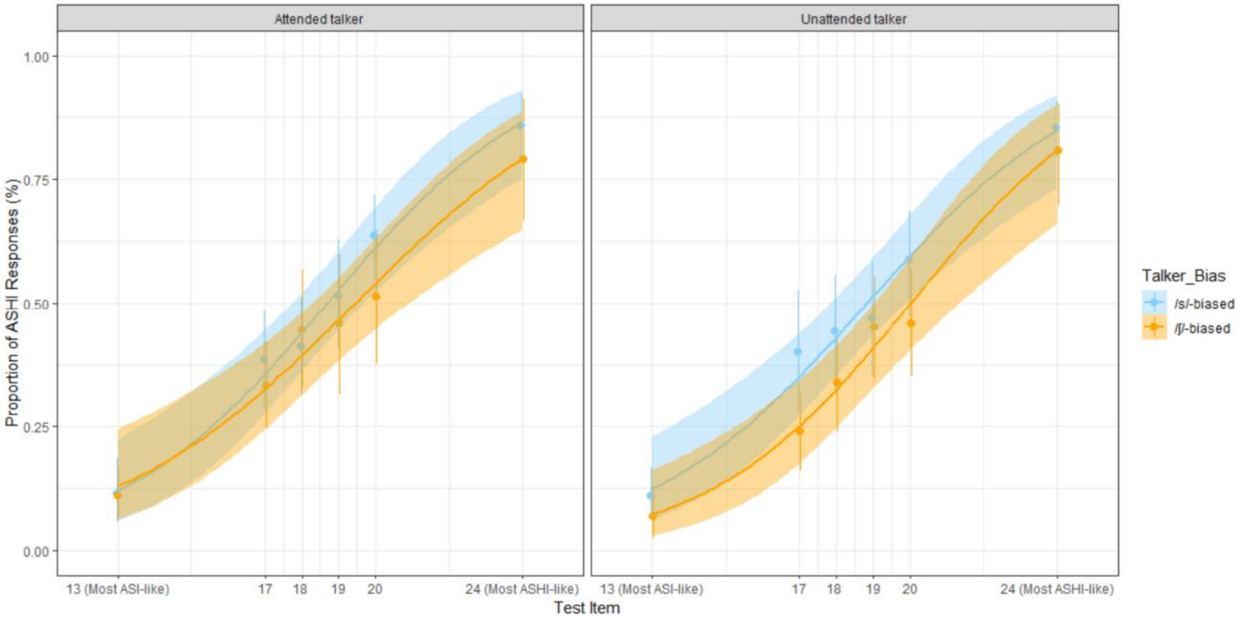

Figure 2: **Results.** The average proportion of *ashi* responses during test for the attended talker and unattended talker, depending on whether the talker produced /ʃ/-biased words during exposure or /s/-biased words. Point ranges show means and bootstrapped 95% CIs of by-participant means. Lines show best fit of a mixed-effect logistic regression with 95% confidence intervals.