No specialized mechanisms needed? Standard perceptual compensation might explain reduced perceptual learning

Shawn N. Cummings (UC Irvine) & T. Florian Jaeger (University of Rochester)

Some fifty years ago, Newell forcefully argued that the psychological sciences lack integrative theories that predict findings across paradigms; instead focusing on the identification of new phenomena, each attributed to new specialized mechanisms [1,2]. Recent work on speech perception has similarly identified a need for more integrative approaches [3,4]. We explore whether a model-driven approach can provide a parsimonious explanation for a phenomenon that has received considerable attention (>500 citations), and has been attributed to specialized learning or memory mechanisms [5-7]: the reduction of perceptual adaptation to unexpected pronunciations when the talker has a pen in the mouth.

Background. In perceptual learning experiments, listeners are exposed to lexically- or visually- labeled shifted pronunciation of a particular category (e.g., /s/ sounding /ʃ/-like in words like *dino?aur, epi?ode,* etc.). Subsequently, listeners accept more tokens—along e.g., an /ɑsi/-/ɑʃi/ continuum—as instances of the shifted category. This change is qualitatively and quantitatively well-predicted by distributional learning models [10,11]. It is robust to distractor tasks, suggesting a high degree of automaticity [8,9]. It is, however, strikingly reduced and sometimes even blocked when the talker has a pen in mouth during the shifted pronunciations [5-7]. Previous work has attributed this reduction to specialized mechanisms like context-specific storage in memory [5,6] and/or inferences about the causes for the observed pronunciations [7]. Critically, neither of these explanations has been shown to be *necessary* to explain reduced perceptual learning.

Model-driven approach. We thus ask whether *known* properties of speech perception suffice to explain reduced perceptual learning. Specifically, we integrated effects of (i) perceptual noise, (ii) perceptual compensation for audiovisual articulatory context, and (iii) lexical context (Ganong & word frequency effects) into a distributional learning model known to predict perceptual learning in the absence of a pen [10]. All of (i-iii) are known to operate during *perception* (see Figures 1 and 2). The question is whether their combined effects are sufficient to explain reduced *adaptation*, depending on the presence of a pen in the talker's mouth (which affects compensation).

To address this question, we reanalyzed a recently published experiment on the effect of the pen [7]. We used published estimates to *fix* both the (i) perceptual noise for all spectral cues (σ^2_{noise}= 878 Mel [12]) and (ii) perceptual compensation [13]. The latter provides a quantitative estimate of the resulting percept for each audio-visual stimulus, based on the acoustics of the speech stimulus, the visually evident effects of the targeted category (/s/ or /ʃ/), and the location of the pen (hand, mouth). Similarly, the two learning rate parameters of the ideal adaptor—κ & ν, determining how much the model adjusts category means and variances, respectively, based on the input during exposure—were *fixed* to those used in previous work [11,14]. This left one parameter not fully fixed from previous work: the influence of lexical context on categorization (π, expressed in log-odds). As recent work [15] reports a range of lexical context effects for fricatives ($.46 \leq \pi \leq 3.3$), we simulated distributional learning while varying π from 0 to 4 (n = 500 per π). Figures 1 and 2 demonstrate the consequences of (i-iii) on perceptual-learning relevant parameters.

Conclusion. Figure 3 qualitatively demonstrates that the blocking/reduction effect observed in previous work follows from known effects on speech perception (i-iii) even without positing new memory or learning mechanisms. Specifically, compensation (ii) reduced p(/ʃ/) across all groups, and lower values of π (iii) resulted in less difference between biasing groups. These effects interact in non-trivial ways, so that their joint effects—even while fixed based on previous work—only become evident once they were integrated into an existing model of adaptive speech perception.

References. **[1]** Newell, A. (1973). **[2]** Guest, O. & Martin, A. (2021). *Perspectives Psych. Sci.* **[3]** Bent, T. & Baese-Berk, M. (2021). *Handbook of Speech Percep.***[4]** Xie, X., Kurumada, C., & Jaeger, T.F. (2023) *Cortex*. **[5]** Kraljic, T., Brennan, S., & Samuel, A. (2008).**[6]** Kraljic, T. & Samuel, A. (2011). **[7]** Liu, L. & Jaeger, T.F. (2018) *Cognition.* **[8]** Samuel, A. (2016). *Cog. Psych*. **[9]** Zhang, X. & Samuel, A. (2014). *JEP:HPP*. **[10]** Kleinschmidt, D. & Jaeger, T.F. (2015). *Psych. Review.* **[11]** Cummings, S.N. & Theodore, R. (2023). *Cognition*. **[12]** Kronrod, Y., Coppess, E., & Feldman, N. (2016). *PBR.* **[13]** Cummings, S., Karboga, G., Yang, M., & Jaeger, T.F. (2025). *JEP:LMC.* **[14]** Cummings, S.N., et al. (2023) *JASA*. **[15]** Kingston, J. et al. (2016). *JEP:HPP*.

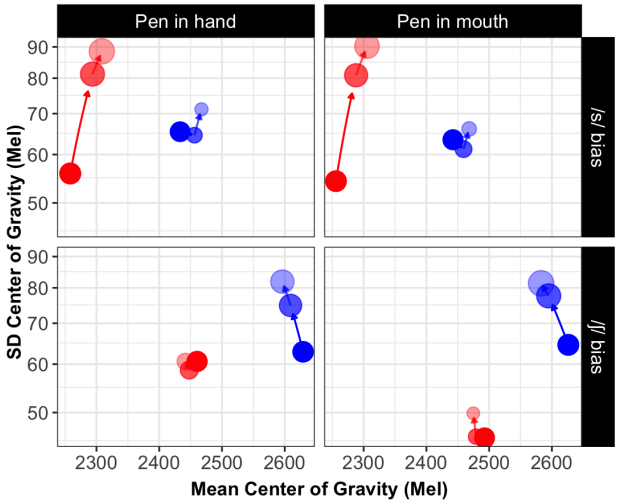

Figure 1 (left). Summary statistics of input across conditions of [7], demonstrating the joint consequences of perceptual compensation and lexical context. Arrows point to decreasing strength of lexical biases (0 = no impact of the lexicon on categorization of exposure input). Size indicates the expected accumulated evidence for a particular category (/s/ and /ʃ/ observations within a single condition and strength of lexical bias will always sum to 40, where 20 tokens were intended by the experimenters to be perceived as /s/ and 20 as /ʃ/). The character of this input, which diverges across conditions and lexical bias strengths, is expected to affect the outcome of distributional learning.

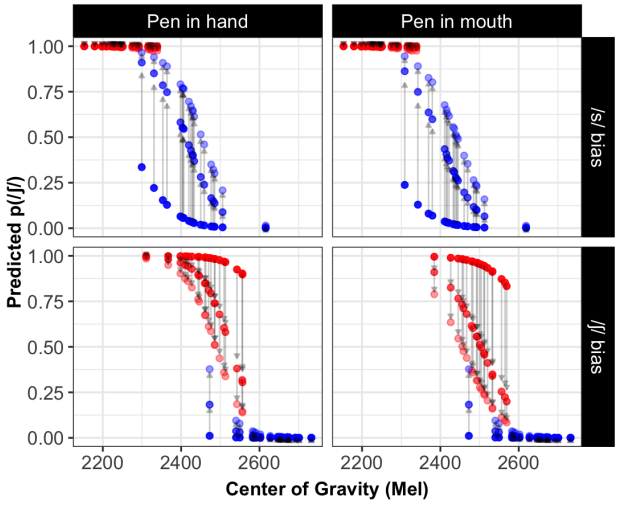

Figure 2 (left and below). Consequences of lexical bias and perceptual compensation on predicted perception of individual exposure tokens across conditions of [7]. Arrows point to decreasing strength of lexical biases. Perceptual compensation is apparent in that the same tokens are perceived as having a higher center of gravity when the pen is in the mouth [13]. This has ramifications for category endorsement, which is jointly determined by (compensated) acoustics and lexical bias.

Figure 3 (below). Comparison of predictions for behavior based on perceptual compensation. Left panels: predictions after distributional learning to the statistics summarized in Fig. 1, facetted by strength of lexical bias. Low-opacity shapes show individual simulations (n = 500), while lines connect averages. Right panel: grand means over subject means of human listener responses in the four conditions of [7]. At π=0, perceptual learning is almost completely blocked when the pen is in the mouth in exposure. Qualitatively, $(.5 \leq \pi \leq 1)$, appears to best approximate human behavior.