UNIVERSITY of ROCHESTER

DEPARTMENT OF
BRAIN AND COGNITIVE SCIENCES

To: Editorial Board of *JASA*

Dear Editors,

We are submitting our manuscript "Comparing accounts of formant normalization against US English listeners' vowel perception", jointly authored by Anna Persson, Santiago Barreda, and Florian Jaeger, for consideration in *JASA*. This is the first submission of our manuscript. It presents original work.

The manuscript presents an **evaluation of 20 influential accounts of formant normalization against L1 listeners' perception of natural and synthesized US English monophthongs**. It is now uncontroversial *that* subcortical auditory normalization mechanisms contribute to robust speech perception—helping listeners to overcome substantial inter-talker variability, specifically variability caused by differences in vocal tract size and shape. However, what computations these mechanisms involve has remained an open question. Some accounts hold that simple auditory transformations are sufficient; other accounts allow more complex operations but still limit normalization to information available *in the moment* (intrinsic normalization); and yet other accounts assume that listeners track and store talker-specific formant information across time (extrinsic normalization). Even within each of these broad classes of accounts, there are substantial differences in the amount of information listeners are assumed to track and apply to normalization.

We present two 8-way alternative forced-choice experiments that assess listeners perception of L1-US English vowel production—either natural productions from a single talker (Exp 1a) or resynthesized variants of these vowels (Exp 1b). We then use a recently developed, general-purpose computational framework for adaptive speech perception (Xie et al., 2023-*Cortex*) to compare how well different normalization accounts predict listeners' responses in these two experiments. We find that extrinsic accounts—i.e., **accounts that assume learning and storage of talker-specific formant information—best predict listeners' responses.** Of these accounts, however, the **computationally *least complex* accounts explain human behavior best.** These results inform theories of adaptive speech perception by narrowing down what types of information human speech perception tracks.

Beyond these immediate results, we hope to contribute by putting an emphasis on *reproducibility*. All materials, stimuli, experiment code, and analyses are documented and shared via OSF. The manuscript itself is written in R Markdown, making all of our analyses reproducible with the press of a button in freely available software. Other researchers can substitute their own data sources and rerun our code; they can add additional normalization accounts, or revisit and revise any of the decisions we made during data analysis.

Sincerely,

Anna Persson, Santiago Barreda, and Florian Jaeger

Meliora Hall
University of Rochester
Rochester NY 14627-0268

Phone: (585) 276 3611
Fax: (585) 442 9216