Friday, December 10th, 2022

To: Editorial Board of *Cortex*

Dear Drs. Guediche and Caffarra,

# We are re-submitting our manuscript, CORTEX-D-21-00884 “Most experiments on exposure effects in speech perception do not distinguish between underlying mechanisms: A computational review”, authored by Xin Xie, Florian Jaeger, and Chigusa Kurumada for consideration for the special issue *Mapping sound to meaning under challenging conditions: converging findings and open questions across methods*. The manuscript is original, not previously published, and not under concurrent consideration elsewhere.

We are grateful for the highly constructive (and encouraging) reviews we received from you and the reviewers. As you summarized in your letter, both reviewers pointed to a need to (1) clarify the goals and scope of this manuscript, (2) clarify the take-home points, in particular whether there exist conditions for which any of the proposed mechanisms can be ruled out, and (3) shorten the manuscript to make it more accessible, potentially by moving parts of the framework into the SI. Finally, R2 pointed out that (4) our presentation of neuro-imaging work was lacking and not well integrated with the rest of the manuscript. We summarize the revisions we made to address these points here, and then respond to the remaining comments point by point. Given the substantial revisions, we have not tracked changes. Only Sections 2-4 have remained relatively unchanged.

1. **Clarifying goals, scope, and contributions.** This was particularly helpful feedback. We have completely revised the introduction:

* After the first introductory paragraph, the introduction now provides a high-level overview of (a) our long-term goals to understand the *mechanisms* underlying adaptive speech perception and (b) the three more immediate goals of the present study: (i) to introduce a new analytical framework—which we now call *ASP* for *Adaptive Speech Perception*—that can support our long-term goals, (ii) to demonstrate the use of this framework through two simulation-based case studies, and (iii) to provide initial guidance on what factors determine whether an experiment can decide between competing hypotheses about adaptive speech perception.
* Following this overview, a new subsection describes the “State of the field(s)”. This section reviews the field and states our contributions. This includes our classification of dozens of competing hypotheses into three qualitatively different types of hypotheses about adaptive speech perception (normalization, representational changes, and changes in decision-making)—something that we failed to highlight at all in the previous version.
* Importantly, the previous version of the manuscript did not clearly state what we consider an important contribution (reflected in feedback from R1). We are now clear that previous work left open whether its signature results distinguish between competing hypotheses, resulting in a *possible* empirical indeterminacy. We believe that our case studies are the first to *show* that the signature results of two influential lines of research can be accounted for by any of the hypotheses. That is, we show that existing data from experiments on perceptual recalibration and accent adaptation are indeed empirically indeterminate. This was *not* previously known.
* We have also substantially revised the text introducing each hypothesis to be clearer *why it matters which of these hypotheses is supported*.

1. **Clarify take-home points.** Guided by their comments, we have completely restructured it to clarify our take home points. After briefly summarizing the findings of our two case studies, the revised discussion now *begins* the discussion with a summary of recommendation for future work. These recommendations necessarily remain general:

* We are now clear right from the start of this section that we are not yet in a position to point to trivial design choices that are guaranteed to distinguish between all three mechanisms *without the need of quantitative model comparison* (something that the reviewers asked). So far, we have been able to identify only one *potential* ‘shortcut’ that does not require quantitative model comparisons, and might be able to reject one of the hypotheses (normalization) as sufficient explanation for a specific adaptive behavior. We now mention this approach, as part of our recommendations (p. XXX).
* Even under our most optimistic estimate, this still leaves at least two mechanisms to be distinguished between by means of quantitative model comparison. Additionally, model comparison will likely be required to distinguish between more specific alternatives *within* each of the three hypotheses—for example, to test whether category expansion or category shift explain a given adaptive behavior.
* To facilitate such model comparisons, we make the same general recommendations as in the previous manuscript. We have, however, revised them to be clearer about the overall take-home point: in essence, our recommendations boil down to the *when*sand *where*sof exposure and test: i.e., after how much exposure to test (*when*), whether to test repeatedly with intervening additional exposure (*when*), andthe location of exposure and test stimuli in the acoustic space (*where*). Computational frameworks like ASP that spell out the linking hypotheses necessary to map acoustic properties to expected categorization responses will likely be a critical component in guiding these stimulus and design decisions.
* Finally, the revised general discussion is now more specific *how ASP or similar frameworks can* *help w*ith such model comparisons, and *why this is not trivial* (Figure XXX on p. XXX)—i.e., why we don’t already provide more specific design recommendations in this paper.

Following these recommendations, we discuss how our findings relate to other efforts to determine the mechanisms underlying adaptive speech perception, incl. neuroimaging research that is explicitly framed in that way but also behavioral findings that—while not framed in these terms—might shed light on the question of whether, e.g., normalization is sufficient to explain adaptive speech perception.

1. **Shorten the manuscript and make it more accessible.** Both reviewers mentioned that the manuscript was long and challenging to read/review. The manuscript bridges research from three theorical perspectives that have largely proceeded in separate and draws on two lines of experimental research, while combining behavioral, neuroimaging, and computational findings. While we have not been able to drastically reduce the length of the manuscript, we have aimed to implement reviewers’ helpful suggestions wherever possible. The main text of the document has been shortened from 67 to XXX double-spaced pages. This was achieved primarily by:

* Restructuring the introduction. For example, we had originally introduced the experimental paradigms for the two case studies in the introduction. In the revised manuscript, we instead introduce each paradigm in the sections where it becomes relevant (3 and 4).
* Simplifying the change model for decision-making, which also simplified and shortened its presentation (Section 2). Moving non-critical technical details into footnotes or the appendix (this mostly affected Section 2, with smaller changes in Sections 3 and 4).
* Removing various asides throughout the manuscript.
* As a result of its scope, almost XXX% of the manuscript length are references. If recommended by the reviewers, we could further cut background information on the different lines of research, which would also cut the length of the bibliography.

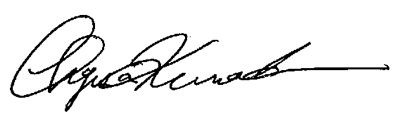
For reasons that are now stated in the introduction, we have deliberately kept the somewhat tutorial-like style of Section 2. We feel that there is a substantial need to bridge the gap between computational and experimental research, and we hope that the relatively verbose exposition of our framework can serve that purpose. Both in our conversations with experimenters, and in our own experience reading computational papers, we often feel the frustration that comes with a lack of shared backgrounds. We have also tried to further improve our figures and animations to that end. However, if the manuscript is still too long or inaccessible, we could move further details of the framework into an appendix OR collect them in a methods article (if that is suitable).

1. **Better integration of neuroimaging research.** R1 suggested we either omit or improve our presentation of neuroimaging research. We have aimed for the latter. This has primarily affected the introduction (p. XXX) and the general discussion (p. XXX-XXX). Additionally, we have aimed to integrate relevant neuroimaging research throughout the paper where relevant (e.g., at the start of Sections 3 and 4). In particular, the introduction now also clarifies that:

*“Compared to the behavioral research …, it is more common in neuroimaging work to directly contrast hypotheses about different mechanisms … However, in contrast to behavioral work, neuroimaging research tends to not distinguish between hypotheses (A) and (B), grouping both hypotheses together as functionally distinct from higher-level, decision-related mechanisms further downstream (C).*” [footnote 3]

Next, we respond to the remaining comments point by point.

Sincerely,





Xin Xie T. Florian Jaeger and Chigusa Kurumada

**Suggested reviewers:**

Eleanore Chodroff (eleanor.chodroff@york.ac.uk)

James McQueen (j.mcqueen@donders.ru.nl)  
Jessamyn Schertz (jessamyn.schertz@utronto.ca)   
Matthew Goldrick (matt-goldrick@northwestern.edu)  
Naomi Feldman (nhf@umd.edu)  
Rachel Theodore (rachel.theodore@uconn.edu)  
Sahil Luthra (sahilluthra@cmu.edu)

**Responses to the remaining comments of Reviewer 1**

I would imagine that there is a good case to argue that adaptation could occur at all three levels simultaneously, at least to some degree.

We completely agree. In fact, we consider ASP’s ability to model any *combination* of mechanisms a *major* strength. As we have shown in our cases studies, even the empirical coverage of individual mechanisms is more powerful than previously assumed, and exhibits dynamics that go beyond what can easily grasped by intuitions. This is even more true when the effects of combinations of these mechanisms are considered. We have revised the introduction (p. xXX) and general discussion (p. XXX) to state this clearly.

-when introducing the lapse rate parameter on p. 21, I was not initially sure of why this parameter would receive such prominent treatment in the paper, although the case was nicely made later on in the paper.  Given the importance of lapses was not discussed in detail earlier in the paper, it could be helpful to foreshadow the importance of this parameter earlier on.

Thank you. We now anticipate the importance of this parameter when we first introduce it (Section 2.1.3). We have also revised the section where we demonstrate its importance (2.2.3) to be clearer how the introduction of attentional lapses means that changes in decision-making can explain adaptive changes in speech perception that are not traditionally considered a consequence of decision-making.

-on p.36: Examining Figure 14A, my impression was that different stimuli were used in the /d/-shifted vs. /t/-shifted panels of the figure.  Would the tightest control not contain the same base stimuli shifted in either direction?

We first answer whether this would be a good idea, and then whether that is what is done in perceptual recalibration (PR) experiments.

Unfortunately, it is unclear what “same” stimuli would even mean. E.g., shifting the /d/ VOTs 10msecs up is not the same as shifting the /t/ stimuli 10msecs down (because /d/ and /t/ typically differ in their variance). One could aim for equivalent shifts in the subject probability of being identified as the targeted category, which would require detailed norming of many stimuli and likely entail different amounts of shifts for /d/ and /t/. Comparable approaches exists—typically under a different name and using somewhat different paradigms (e.g., unsupervised distributional learning paradigms or dimension-based statistical learning paradigms, e.g., Clayards et al., 2008; Idemaru & Holt, 2011). Our manuscript originally contained another case study on these alternative paradigms, confirming that the indeterminacy we describe generally extends to those paradigms (but, as the reviewers pointed out, the manuscript is long as-is).

Regardless of the *possibility* of changing PR experiments to employ an approach more akin to what the reviewer suggests, it is not what has been done so far. PR experiments do *not* typically parametrically manipulate the acoustic-phonetic properties of stimuli. The most common approach to the generation of exposure stimuli is to (i) record typical /d/ and /t/ versions of each stimulus (e.g., *lemonade* and *lemonate*), (ii) blend these two stimuli together under various amplitude weightings (from 100% *lemonade-* 0% *lemonate* to 0% *lemonade*-100% *lemonate*), (iii) select based on experimenters’ intuition or a small norming study the most ambiguous blend for each stimulus and call it the “shifted”, “ambiguous”, or “atypical” stimulus version (with 100% *lemonade* remaining the “typical” or “unshifted” stimulus). There are rare exceptions to this (for a review and critique of this approach, see Theodore, 2021).

In short, PR experiments do neither carefully select the tokens *within* each category, nor is there any form of counter-balancing *across* the two categories. As we state on p. XXX, it is extremely rare that the acoustic properties are even measured. Our computational simulations capture the qualitative approach taken in PR experiments (in a separate project, we *have* measured the acoustic properties for our and dozens of other PR experiments to confirms this).

p. 37: The focus on the simulations is on the beginning of the test phase; however, should the model not also be able to account for performance throughout the test phase?  If not, why not?  Is this reflective of some additional parameter not included in the model (e.g., a reluctance to keep changing beyond a certain point?), particularly in the face of repeated stimuli?

This is another interesting point. In essence, this is a point that has received very little attention in previous experimental work. In Liu & Jaeger (2018), we showed that repeated testing reduced the effects of exposure. This and subsequent studies by us and other researchers suggest that at least 2-3 factors contribute to this:

(1) continued unsupervised adaptation over the unlabeled input with non-bi-modally distributed acoustic properties. Test stimuli tend to span some continuum, with each location along that continuum being repeated equally often. Even when some locations are repeated more often, it tends to be those in the center of the continuum leading to a uni-modal distribution. Either way, the distribution of test stimuli violate listeners’ expectations based on lifelong input and deviate from the exposure distributions.

(2) Meta-expectations, including expectations specific to the task structure of experiments: e.g., the expectation that a 2AFC task with two possible answer displayed on the screen likely means that each option will occur about equally often.

(3) Dis-engagement due to the repetition of highly similar sounding stimuli (i.e., increased lapse rates).

As we now clarify, none of these factors are modeled in our study (all can be added to ASP, and some are already implemented). Instead, we tend to use short test session with many participants to overcome this problem experimentally. Since this point is not critical for the purpose of this article, we have moved it into a larger footnote on p. XXX at the start of the result section.

p.55 the authors state "the highest accuracy is obtained for the fastest changes, and it matches that observed for changes in decision making."  Looking at the data, I am not sure that the match is especially strong, but I may be misinterpreting the data being referenced here or the level of "match" that the authors are referring to.  Perhaps this could be clarified?

The reviewer is correct. We have revised the presentation of this result to be clear that we mean qualitative similarities (the fact that L2-accented exposure conveys an overall benefit, compared to L1 exposure, and that this benefit is most pronounced for /d/--the category that differs in the L2-accent).   

I had several issues using the pdf document, including generating a printed copy.  I suggest the journal and the authors be mindful of this if this paper is moved to production.  I was on windows 10 using the current version of Adobe Acrobat when I encountered these issues.

Thank you for making us aware of this. We apologize for the inconvenience. We have noticed that some printers struggle when printing the PDF *double-sided*. We suspect that this is due to the size of the manuscript, which is in turn due to the use of animations.

We have a back-up strategy (alternative figures) in case the animations will not be accepted by *Cortex*. For now, we have also made available in OSF a PDF for printing that we hope will avoid the problems (LINK)? We have also change the default state of the figures to show the *end* state of the animations, i.e., the state of maximal differences. The captions have been adjusted accordingly. This is more informative.

**Responses to the remaining comments of Reviewer 2**