

Dear Miss Mitra,

BLC-21-0133 entitled "Mixed language processing increases cross-language phonetic transfer in Bengali-English bilinguals" which you submitted to *Bilingualism: Language and Cognition*, has been reviewed. The comments of the reviewers are included at the end of this letter.

As you will see in the attached comments, although the reviewers point out that your paper has the potential to make a significant contribution to the field, they have raised a number of major concerns. Specifically, the reviewers ask you to address the following aspects of the study that they find problematic:

1. The main concern is the small sample size in the study. If it is underpowered, this may be responsible for the null results in the comparisons of the outcomes of two tasks and of the unilingual and bilingual trials. This possibility needs to be ruled out before any meaningful conclusions can be justified. Accordingly, you need to provide a power analysis demonstrating that the study is not underpowered. Alternatively, you may consider collecting additional data to increase your sample size.
2. The proficiency and the levels of exposure of the participants are not specified. This is problematic for any generalizations to be drawn beyond your sample.
3. The baseline data come from a different study with all female participants. It is unclear how it was determined that the two samples can be directly compared (beyond the background data). Could the differences in gender have biased the results?
4. The Discussion section is short and does not provide a deep analysis and interpretation of the findings. It should start with a short summary of the findings.
5. Some of the claims about prior research are unfounded (see the references that the reviewers provide for the literature on the vowel shift).
6. Consider the suggestion of reviewer 3 to run a single model for the five measurement times rather than separate models.

In light of these concerns and recommendations from the reviewers, I'm sorry that we will not be able to accept your paper for publication in *Bilingualism: Language and Cognition* at this stage. However, I encourage you to consider the reviewers' comments carefully and revise and resubmit your manuscript. Please bear in mind the following two points when you revise the paper: First, inviting resubmission does not entail that the next version, or any subsequent version, will be accepted for publication. Moreover, the clarifications that result from the revision may reveal new issues, hitherto unnoticed, that preclude publication. Second, please ensure you submit a cover letter detailing the reviewers' points and how you have dealt with each one.

To revise your manuscript, log into <https://mc.manuscriptcentral.com/blc> and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a major revision. You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Please do not directly mark the changes inside the revised manuscript. Instead, provide a detailed cover letter that lists the original reviewers' concerns and how you have addressed these concerns. This cover letter should be uploaded to either the Author's Cover Letter section, or Author's Response section, to be viewable by the editor and reviewers for your revised manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewers.

Because we are trying to facilitate timely publication of manuscripts submitted to *Bilingualism: Language and Cognition*, your revised manuscript should be uploaded as soon as possible. If you plan to resubmit but anticipate a significant delay in doing so (e.g., in more than 3 months), please let me know.

Thank you for considering *Bilingualism: Language and Cognition* as the outlet of your research.

Sincerely,

Dr. Kira Gor

Associate Editor, *Bilingualism: Language and Cognition*

kiragor@umd.edu

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author

I will focus my comments here on a few things I believe should be improved. Let me say first that I am quite impressed by the quality of the writing and the insight demonstrated. This is a very fine project and it has many positive things going on for it. There are a number of issues the authors would do well to address, though.

1. The sample size is just too small. The study recruited only 10 participants. This is not much less than comparable studies, I know. The standards in our field, however, are quickly changing. This study **sets out to compare the sizes of the effects of two experimental treatments or conditions**. Effect sizes are likely to be very small in the population, if they exist at all. This is designed as a confirmatory study, rather than an exploratory study. It is one thing to explore an issue for the first or second time, and another to confirm a specific effect others believe they have found while exploring the possibility of its existence. A confirmatory experiment needs a power analyses, or at the very least the sample size must be explicitly justified. I **suspect the sample is too small to detect any differences between the two treatments** (picture naming vs. code-switching), and the **differences between the unilingual and the bilingual trials is apparently found to be very small**. The likelihood of detecting a significant difference increases with a more powerful experimental design; and if no differences are detected one may propose that none exist in the population. And, for significant differences, a more accurate estimate can be found in more powerful designs--those with larger samples. I'd like to encourage the authors to consider the possibility to return to the field and to continue to collect data, if at all feasible. Otherwise, at the very least, acknowledge the limitations openly and justify your sample size. Changes are this won't be enough, but only the editorial team can decide whether they continue to accept low-power studies or not.

2. We need to know a lot more about the participants. Participant information takes a single paragraph. Readers are not necessarily familiar with this particular bilingual population. What's more, there are no indicators of bilingualism or English or Bengali proficiency. The authors would do well to collect information from the participants utilizing one of the available bilingualism surveys, such as the BLP or the LEAP-Q. Also, linguistic proficiency should be assessed somehow; perhaps a vocabulary size measure or a grammar knowledge measure. Or at least a self-assessed proficiency questionnaire. This is of course optional, particularly in native bilingual populations. In this case, though, the reader needs to learn a lot more about the society and the bilingual conditions of both the population and the specific sample: language use patterns, ages and contexts of acquisition, literacy, origin, ages, education levels and languages of education, etc. Finally, consider the possibility of reporting the inclusion and exclusion criteria openly and explicitly.

3. The Discussion section is brief and shallow. I am impressed by the quality of the writing in the Introduction, but the Discussion and Conclusions are no match for it. The Discussion needs to be much longer, much deeper, it needs to engage with theory and contextual findings explicitly and in depth, it needs to interpret the specific findings and explain their implications. I suggest that the "limitations" paragraph be moved out of the last position in the article. It could perhaps be placed in the Discussion section, but not in the Conclusions section. The Conclusions section needs to focus on the positive findings: what have we learned here? What can we say after having conducted this study? Etc. Finally, regarding the Discussion section, there should be a separate subsection entitled "Summary of Findings" (of course, whether there is such a separate section or not is a personal matter, but my point is that there need to be a few paragraphs in which the authors summarize the findings (perhaps by providing standardized effect sizes) in plain English while refraining from interpreting or explaining the data. There needs to be a subsection in which the data are summarized so that the reader can remember the findings before engaging in a discussion of their implications.

Except for my suggestion to keep collecting data to obtain a larger sample (as well as collecting more information from the participants, in the form of some bilingualism survey), I believe my other suggestions are relatively minor. I must say that I was quite impressed while reading the manuscript until (1) I saw the Participants section (small sample size and sparse information about the linguistic histories and experiences of the participants), and (2) I read the Discussion (brief, shallow, avoiding a deep discussion of the implications of the study and lacking a summary of findings devoid of interpretation before moving on to their interpretation). I sincerely hope the authors are willing to consider some of my comments. I believe this is, overall, a fine study.

Reviewer: 3

Comments to the Author

The paper reports two tasks – cued-picture naming and code switching – investigating cross-language phonetic transfer in bilingual (Bengali-English) speakers. The study focused on the F1 and F2 of two vowels of the L2 (Indian English) language. Both vowels were subject to changes, with a shift toward the L1 norms. The same changes occurred independently of the task. The study thus reports first evidence of transfer effects on L2 vowels.

I've found the study interesting, and the findings having a good potential for significantly contributing to models of bilingual language control. There are some relatively minor issues (mostly methodological) that I think deserve attention and need to be addressed.

1. As baseline data, the authors use Dutta's (2021) dataset, which is composed of production data from 5 female speakers. In the present study, however, participants are 5 females and 5 males. Because of the effects of gender on vowel realization, do the authors think that this difference between the two groups (Dutta vs the present one) might have contributed to the effects reported in the present study?

2. Participants.

a) How was the sample size calculated?

b) Participants should be described better, especially for what concerns their L2 experience. Currently, there is no information about their proficiency level, as well as their degree of exposure and use of their two languages.

3. Materials, how many of the 20 words contained each of the two target vowels? Please specify this information in the text.

4. Procedures.

a) If I correctly understand, each word was repeated 4 times in each task (i.e., the same word was produced for a total of 8 times by each participant). If this is the case, can the authors exclude any repetition effect? And why did they opt for a so massive repetition, instead of, e.g., increasing the number of items?

b) Cued-picture naming: why was the picture presented for a so short time (50 ms)? What words were used as language cue and as target?

5. Analyses.

a) To analyze the changes at different times during the vowel production, I would suggest to run a single model in which the time of measurement (5%, 15%, 25%, 35%, 45%) is entered as a numeric predictor (from 1 to 5). This will reduce the number of statistical models to run and, more importantly, will allow to directly study the time course of the changes.

b) p. 22 ls 21-23. The authors should specify which interactions they are referring to and report the results of the statistical tests.

c) first analysis (p. 22, ls 24 and following). Given its relevance for the hypotheses, why not including the interaction term for the random slope since the beginning of the analysis?

d) Tables reporting statistical results, please, specify in the legend what the numbers are (I assume the beta values and the standard errors (in parenthesis), but this should be spelt out).

e) I think the readers may appreciate few lines commenting the baseline measure (i.e., Dutta's (2021) data). Moreover, the authors should clarify whether and how these baseline data were compared to the results of their study.

6. I have found the discussion a bit unbalanced, as approximately half of it is dedicated to a secondary question (i.e., the impact of the paradigm on the transfer effects). I suggest the authors to re-balance the space, expanding the discussion of their main research question – they might further develop the theoretical implications of their findings, also offering some explanation of the results they consider unexpected.

Minor

- p. 3 ls 39-43, To improve readiness, I would suggest the authors to consistently use only one term throughout the manuscript.

- in some places, the authors use capital letters to highlight some concepts, but I don't think they are really needed. Simple lower-case letters may be fine as well.

- Introduction, when discussing previous studies, the authors might clarify what acoustic measure each study has investigated (VOT, ...)

- Acoustic analyses, third line, the authors report a total of 2333 target to be analyzed. How many of them were there for each target vowel?

- p. 22, first line, please change “mediated” with “affected” or “moderated” or a similar term – mediated refers to mediation analysis, which is not the analysis run by the authors.

This study set out “to ascertain if L1 influence on L2 increases during mixed-language use, relative to a participant’s multilingual baseline production of L2.” The major claim (page 24) is that “Both vowels in this study showed systematic shifts in the mixed-language condition compared to baseline unilingual productions.”

I am concerned about whether the results actually support this claim. As the ms. makes very clear, the data on baseline production of these vowels in IE was taken from a different study of five female speakers. The participants in this study were ten male speakers. While we are told that the speakers in this study “were selected on the basis of responses to a Language Background Questionnaire to ensure comparable LSRW...skills in L1 and L2 across participants), we are not given any information on the criteria used in this selection, and no information beyond gender is provided for the baseline group. The ms. seems to assume that the 5 female speakers in the baseline study can be taken as representative of all L1 IE speakers, but we have no evidence for that.

A second claim (p. 25) is that “As expected, we find a greater degree of shift in [ʌ], possibly owing to its position in a less dense part of the IE vowel space.” This is a very interesting claim but it would be much more convincing if we were provided explicit information on the IE vowel space, in the form of a chart of all the vowels in this language. We also need to know whether acoustic differences in addition to F1 and F2 (such as duration) play a role in distinguishing vowels.

Finally, at a number of points the ms. makes reference to the lack of studies of cross-language influence on vowels (as opposed to VOT). I’m not sure this is a fair characterization. While studies of the effect of language switching on vowels may be lacking, there are certainly a number of studies investigating phonetic drift in vowels, for example:

Grace E. Oh, Susan Guion-Anderson, Katsura Aoyama, James E. Flege, Reiko Akahane-Yamada, Tsuneo Yamada. 2011. A one-year longitudinal study of English and Japanese vowel

production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics* 39: 156-167.

Natalia Kartushina and Clara D. Martin. 2019. Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics* 77: 100920.

Benjamin Lang and Lisa Davidson. 2019. Effects of Exposure and Vowel Space Distribution on Phonetic Drift: Evidence from American English Learners of French. *Language and Speech* 62: 30-60.

Specific comments

page 6: I found the sentence in (ii) confusing.

page 6: 'monopthongs' should be 'monophthongs'

page 9: First sentence: "bilinguals speakers'..."

page 9: "The research discussed above concerns situations where the bilingual speaker is operating in any one of their languages." It seems odd to refer to studies investigating the first five weeks of learning a language as involving bilinguals.

page 13: It wasn't clear to me how the discussion of the Bengali laryngeal contrast was relevant here. from a different

page 14: "individual sounds pairs"