



慶應義塾
Keio University

5322 Endo, Fujisawa
Kanagawa 252-0882, Japan
(+81) 80-6551-4063

July 1st, 2023

RE: <https://doi.org/10.31234/osf.io/jr9x7>, *Globally, songs and instrumental melodies are slower, higher, and use more stable pitches than speech*

Dear Dr. Chambers,

We appreciate you managing this review process and providing the final feedback on our submission. We are grateful for the chance to use additional constructive reviewer feedback to refine our manuscript. Note that in order to address the issues raised by reviewers, we have made some changes in the Stage 1 materials. While we are aware that these materials should ideally have been frozen, we have taken the following actions. None of these make any substantive change in the content: 1) added clarification in the "2. Methods" section, and 2) modified the terms to make them more intuitive, which also appear in Stage 1 Supplementary Materials (Section S6). Incidentally, accompanied by clarifying terms, we have also slightly updated Stage 2 Supplementary Materials (Section S8). We have also realized that one particular data point in the map figure created during Stage 1 (Fig. S1) was mistakenly placed, so we have corrected it. These changes also do not make any substantive changes in the content. All changes are tracked for transparency.

We feel that the review process finally made our manuscript deliverable to readers. We hope you will find the revised manuscript acceptable for formal Stage 2 Recommendation.

Sincerely,

Yuto Ozaki and Patrick E. Savage
(on behalf of the authors)

Editor summary (Chris Chambers)

I have now received two evaluations of your Stage 2 submission by the reviewers were assessed the proposal at Stage 1. As you can see, both are very positive about the completed work, with one reviewer happy with the submission in its current state and the other raising some minor issues concerning terminology and clarification of the analyses. Please address these points in a revised manuscript and response, and we should be able to award a positive Stage 2 recommendation without further review.

Once again, we appreciate you managing the review process. We believe we have incorporated all suggestions (see below for details).

Reviewer #1 (Bob Slevc):

I was excited about the first stage of this registered report, and have enjoyed seeing how the data turned out. For the stage 2 review process, I guess the most important question is whether the work was carried out as proposed in the stage 1 manuscript. This certainly seems to be the case! Changes from stage 1 are clearly described (and seem reasonable), the manuscript includes plenty of detail, and the recordings and analyses scripts are not only made openly available but are appropriately documented. I don't think it's really necessary for me to comment on the results and interpretation, but I will just note as an aside that my stage 1 concerns about the specific SESOI chosen appear to have been unfounded -- effects are larger than I would have expected! That said, I do appreciate the inclusion of manipulated examples and discussion of how the SESOI was chosen (e.g., in section S7).

Overall, I enjoyed this paper and appreciate all the work that went into this project. I expect this will prove to be a really useful resource for many of us in the field(s)!

We sincerely appreciate you taking the time to refine our manuscript at both Stage 1 and Stage 2.

Reviewer #2 (Nai Ding):

In general, I think the paper is almost ready to publish.

We sincerely appreciate you taking the time to refine our manuscript at both Stage 1 and Stage 2.

A number of issues, however, still need to be addressed and most of these issues are terminology issues. For example, it should be explicitly stated that the rhythm measures general refer to the rhythm of breath instead of the rhythm of sound (if I understood it correctly).

1. For the inter-onset interval, please specify the unit (i.e., the onset of what?). If it's the onset of a breath, I wonder why it can reflect the speed of speech or music. Suppose I take a breath after every syllable in one condition and take a break after a sentence in a second condition. I may breathe more frequently in condition 1, but the speech rate, e.g., measured by the number of syllables per second, may still be higher in condition two.

Similarly, in Fig. 8, what is called an onset and break annotation? Does it mean the duration within a breath?

We thank the reviewer for bringing up this point. Actually, most rhythm measures do not refer to breaths but rather to syllable/note onsets. While we have already provided details in the supplementary section, we have added the following explanation to the Methods section to clarify the points raised by the reviewer.

<2. Methods section>

“The unit of IOI is seconds and IOI rate is the reciprocal of IOI. Onset represents the perceptual center (P-center) of an acoustic unit (e.g., syllables, mora, note), which represents the subjective moment when the sound is perceived to begin. The P-center can be interpreted to reflect the onset of linguistic units (e.g., syllable, mora) and musical units (e.g., note), with the segmentation of acoustic units determined by the person who made the recording. This measure includes the interval between a break and the onset immediately preceding the break. Breaks were defined as relatively long pauses between sounds. For vocal recordings, that would typically constitute when the participant would inhale.”

I don't know how the Fourier transform can be used to length normalization and interpolation.

Interpolation using the Fourier transform is a technique established in the field of signal processing (Fraser, 1989; Schafer & Rabiner, 1973). In brief, changing the length of a signal and performing necessary interpolation can be achieved by applying zero padding in the frequency domain and reducing sampling rate in the time domain. We have added the reference and described the implementation we used (i.e., interpft function of Matlab).

<Caption of Figure 8>

“The extracted f0 contours were normalized to the length of 512 samples using interpolation by Fourier transform and resampling (Fraser, 1989; Schafer & Rabiner, 1973). The implementation by the MATLAB function interpft is used.”

phrase length -> breath duration

interval regularity -> this one is particularly confusing since interval also refers to the breath signal in IOI

loudness -> intensity or just short-term energy

We appreciate the reviewer’s suggestions for improving the clarity. We have updated the terms as follows. Note that we chose not to change “phrase length” to “breath duration” as the latter might imply the length of time taken to breath BETWEEN phrases - instead we have added a parenthetical clarification as below:

phrase length -> phrase length (duration between two breaths/breaks)

interval regularity -> pitch interval regularity

loudness -> intensity

2. Abstract: "relative to speech, songs consistently use" Consider to replace consistently with generally. If I understood the results correctly, there is high variability and the result is not consistent within every participant (e.g., Fig. 7).

We agree with the reviewer's concern. To make this more concise, we have simply removed "consistently": "songs consistently use -> songs use", and "speech consistently used -> speech used".

<Abstract>

"Of our six pre-registered predictions, five were strongly supported: relative to speech, songs use 1) higher pitch, 2) slower temporal rate, and 3) more stable pitches, while both songs and speech used similar 4) pitch interval size, and 5) timbral brightness."