**Supplementary Material**

To examine the test-retest reliability of the sound discrimination and reproduction tests used in the current study, we decided to conduct a follow-up research project. We recruited and asked a total of 30 L1 and L2 English users to take the same tests twice for two consecutive days (Time1, Time 2). Our assumption was that their Time 1 and Time 2 scores should tap into the same constructs of one’s auditory processing ability, and thus demonstrate significant associations. As in the current study, the sound discrimination and memory tasks comprised the following subcomponents:

**Table 1**

*Constructs of Explicit Auditory Processing and Its Measures*

|  |  |
| --- | --- |
| Type of processing | Measures |
| Sound discrimination | Duration discrimination |
|  | Amplitude rise time discrimination |
|  | Pitch discrimination |
|  | Formant discrimination |
| Music memory | Rhythm reproduction |
|  | Melody reproduction |

A total of 30 L1 and L2 English users (13 males, 17 females) included native speakers of Chinese (*n* = 8), Japanese (*n* = 8), English (*n* = 5), Polish (*n* = 4), French (*n* = 2), and other Indo-European languages (*n* = 3). Their chronological age widely ranged from 22 to 49 (*M* = 29.9 years; *SD* = 6.9). At the time of the project, eight participants were in English-as-a-Foreign-Language contexts (Japan or China), and 27 participants were residing in the UK. According to the results of Pearson correlation analyses summarized in Table 2, their overall scores (averaging across all the subcomponent scores) demonstrated mid-to-strong associations for their discrimination performance (*r* = .701) and reproduction performance (*r* = .863).

**Table 2**

*Summary of Test-Retest Correlations*

|  |  |
| --- | --- |
|  | *Correlation analysis* |
| *r* | *p* |
| Overall sound discrimination  | .701\* | < .001 |
| Formant | .619\* | < .001 |
| Pitch | .562\* | .001 |
| Duration | .284 | .128 |
| Amplitude rise time | .798\* | < .001 |
| Overall music memory | .863\* | < .001 |
| Melody | .907\* | < .001 |
| Rhythm | .775\* | < .001 |

When we took a close look at the participants’ subcomponent scores, however, it is important to note that the participants’ duration discrimination scores were not significantly correlated between T1 and T2. This indicates several possible scenarios. First, the specific task format (duration discrimination) may not be sensitive enough to capture participants’ duration-specific and temporal processing abilities. Second, as operationalized in the current study, it would be more reliable, and, by extension, valid to use combined/overall scores to proxy the multifaceted nature of one’s auditory processing abilities. Finally, the sample size could have been too small (*N* = 30). The information here is adapted from the following research report.

Saito, K., Sun, H., & Tierney, A. (2020). Brief report: Test-retest reliability of explicit auditory processing measures. *bioRxiv.*