Response to Reviews:

**(Previous) Point 8.**

**Thank you, I understand that each reponse was coded between 0 and 1 (in 0.25 steps), and the shift was only made to associate negative scores with "incorrect" and positive scores with "correct". It will not make any difference statistically, but an alternative process, potentially more intuitive, would be to code "definately incorrect" as -1 and "definately correct" with 1, and then use steps of 0.5 (instead of 0.25).**

*Response: Yes, indeed, this would have been an alternative approach that could have been more intuitive. But since that does not make a statistical difference, we keep the present one because that its consistent with the previous paper.*

**What will make a difference statistically, however, is that a score of 0.5 (before subtracting 0.5) represents either "exactly bewteen definately incorrect and definately correct" or "don't know". So "don't know" and "neutral" are scored identically. I therefore would ask the authors to do one of the following:**

**1. Motivate why neutral opinions to questions are identical to "don't know". This is a theoretical assumption.**

**2. Otherwise, remove all "don't know" from the analyses (alternatively introduce imputations) or there will be a bias with scores leaning to the mean and with smaller standard deviations affecting the statistical tests.**

Response: In the PROMS, participants had five response options in the “definitely same”, “maybe same”, “don’t know”, “maybe different”, and “definitely different”. There is no “neutral” response option. But, if one participant had pressed always the “definitely” response options, but would be correct in 50% of trials only, they would have the same score as someone who always pressed “don’t know”, which is exactly 0 in both cases. From a theoretical point of view, these two participants differ in their response style, but not in their sensitivity to musically expressed information. In terms of sensitivity, both are at chance level, but while one is more cautious (by pressing “don’t know”), the other one is more confident and thus sometimes confidently wrong. In signal detection theory, this is often referred to as the “criterion”. Our measure of interest was about sensitivity, and not about differences in the response style, so we consider our approach valid.

Concerning your second suggestion, we included some additional analyses on OSF (R-files “01c\_PROMS\_data\_analysis\_amateurs\_partI.R”). Importantly, we checked whether the number of “don’t know” responses differs between groups. There are no differences in any of the PROMS subtests between singers and instrumentalists, and thus we conclude that they had a similar response style as well.