**1A. The scientific validity of the research question(s)**

This study aims to investigate the relationship between music, emotion and visual imagery across three different countries. This research question is interesting and bring together different fields. However, the scientific motivation behind the research questions is not always made entirely clear. For example, the sentence “By distinguishing between universality and culture-specificity of these associations, we aim to understand whether relationships between music, emotion, and visual imagery are cross-culturally universal or culturally specific.” is circular. Another example is the question regarding ‘visual density’; the authors state that investigating this “will add to our understanding of audio-visual associations…” but do not give any motivation for why this is the case. Why is this feature important to investigate and how exactly does it add to previous research on different visual features? Also, please explicitly define ‘visual density’ and ‘solo and group excerpts’ in the introduction.

More broadly, why is the relationship between music, culture, emotion and visual imagery important to characterize outside of the context of previous studies on this exact topic? Why is it important in general? Some of this is discussed in the fourth paragraph of the introduction where some more concrete findings regarding this relationship are mentioned. Tackling this earlier on might help in making the general motivation clearer.

**1B. The logic, rationale, and plausibility of the proposed hypotheses (where a submission proposes hypotheses)**

The motivation for the first hypothesis is clearly based on previous studies about the relationship between tempo and emotional arousal. The rationale for the second hypothesis is less clear to me, not least because the concept of ‘visual density’ has not been defined explicitly in the introduction. Furthermore, both hypotheses state that these relationships are universal but, as far as I understood, this is an empirical question and there is no clear evidence given to a priori expect universality or specificity between cultures.

**1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable)**

The proposed analyses are t-tests but this is a complicated design with more than 2 groups/measurements to compare. The authors added a foot-note for other potential analyses but have not provided any motivation for why these other analyses might be more suitable and which of them they will actually use. Some of the specific interpretations of the analysis outcomes are also incorrect. For example, “If our predicted effect is significant in each of the three cultural groups, we will conclude that tempo-arousal relationships are cross-culturally general.”. However, it is possible that the relationship is significant in each group but that there are still significant differences between the groups (e.g. the relationship is stronger in one culture than another), indicating cultural specificity. The next sentence regarding interpretation of a null-finding is also incorrect: the relationship might be significant in some but not other groups but if there is no significant difference between the groups you cannot claim cultural specificity. More generally, absence of evidence for an effect (i.e. a null-finding, or ‘statistically equivalent’) obtained with frequentist statistics does not provide evidence for the absence of an effect. I am also surprised by the small sample sizes proposed (N = 14 per group), especially since the pilot data already consisted of N = 9 per group and given that it is an online study. Taken together, I unfortunately have to conclude that the proposed analysis pipeline is not very sound.

**1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses**

The methodological details are mostly clear but for exact replication the musical excerpts and experimental code need to be shared. Furthermore, there is some unclarity and flexibility in the proposed statistical analysis (also see previous question).

**1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).**

The tempo of the excerpts is manipulated to be 20% slower or faster than the original. Is there a control that can ensure that these manipulations do not influence other aspects of the music, independent of tempo per se, such as the clarity of the lyrics, which might influence the results? Furthermore, the solo pieces were slower than the group pieces, so any difference between these two is likely to due to tempo rather than solo vs group. Relatedly, the pieces from the different countries also all have different tempos, which means that any difference here might also be due to tempo rather than country. In the absence of these controls, it is hard to draw clear conclusions about the effects of these variables.