Dear Dr. Guérin,

My sincere apologies for the delay in providing feedback on your RR. Securing reviewers proved challenging, especially towards the year's end. However, once suitable individuals were found, their evaluations were detailed and promptly completed. I believe these reviews will greatly assist you in revising this Stage 1 plan.

Although both reviewers provided substantial recommendations for enhancement, their overall assessment of the proposed studies is positive. Reviewer #1 highlights several important concerns, including acknowledging the concept of active sensing, considering recent evidence on metre perception, clarifying the state of knowledge on brain oscillatory interactions, distinguishing between related theories, reconsidering movement and brain activity as the gold standard for measuring beat perception, refining participant exclusion criteria, addressing potential confounds related to metronome use, clarifying technical details, and considering Bayesian ANOVAs to address potential issues with sample size and interpretation of null findings.

Dr. Anne Keitel, the second reviewer, suggests adding a control condition without movement to rule out alternative interpretations, replacing pure tone stimuli with real drum sequences for better ecological validity, providing clearer information on participant samples and stimuli in abstracts and introductions, clarifying inclusion criteria regarding cultural exposure, considering self-report questionnaires for music exposure, and justifying the chosen alpha value of .02.

My own reading was mostly concentrated on the RR and statistical analyses because I am not a specialist in the topical area. The primary issues that I would like you to consider are outlined below:

1. Most hypotheses are proposed to be tested by modelling the experiment as either pairwise *t*-tests or Mixed-model ANOVA. However, this approach, which only models fixed effects, has some limitations. I believe fitting mixed models (Bolker, 2015) that include random effects, such as participants, may be preferable, as it would allow for the generalisation of results to a wider population of participants (Barr et al., 2013; DeBruine & Barr, 2021). Linear mixed models could be fitted with random intercepts per participant, or even random intercepts and random slopes between sessions for each participant. To implement this in R, you could use packages such as `brms` (Bürkner, 2017) if you decide to follow Reviewer #1's suggestion of using a Bayesian approach. Alternatively, packages such as `lme4` (for general or generalised models; Bates et al., 2015) or `lmerTest` (only general —normal/Gaussian— linear models; Kuznetsova et al., 2017) could be used if you decide to maintain a frequentist approach. Regardless of the chosen approach, whether frequentist or Bayesian, *post-hoc* comparisons can be tested using packages like `emmeans` (Lenth, 2024).

   For example, for the Movement Condition × Session models, you could use one of these options:

```
Unset
# For a frequentist approach
## Load package
library(lmerTest)
## Fit model
model <- lmer(variable ~ Movement_Condition*Session +
             (1 + Session | Participant),
             data = XXX)
## See model results in ANOVA-type table
anova(model)
## Plot the estimated marginal means
ggplot(data.frame(emmeans::emmeans(model,
                       pairwise~Session|Movement_Condition),
       aes(x = Movement_Condition, emmean, color = Session)) +
```

```
  geom_point(position = position_dodge(width = 0.5)) +
  geom_errorbar(aes(ymin = emmean-SE, ymax = emmean+SE),
                width = 0.4,
                position = position_dodge(width = 0.5))

# For a Bayesian approach (without specifying priors)
## Load package
library(brms)
## Fit model
model <- brm(variable ~ Movement_Condition*Session +
            (1 + Session | Participant),
            data = XXX,
            family = gaussian())
## See model results
summary(model)
## Plot the estimated marginal means
plot(marginal_effects(model))

# For post-hoc comparisons between sessions (regardless of whether
# you used lme4, lmertest or brms to fit the models)
## Load package
library(emmeans)
## Contrast between sessions
emmeans(model, pairwise~Session)
## Contrast between sessions by movement condition
emmeans(model, pairwise~Session | Movement_Condition)
### Note: Please keep in mind that, by defect, emmeans uses Tukey to
### adjust p-values. If you prefer to use Bonferroni correction,
### simply add the parameter 'adjust = "bonferroni"'.
```

2. On page 27 (line 740) you state that "Normality will be checked using the Shapiro–Wilk test" and that "if violated, the data will be normalised using a transformation that will be contingent on data distribution curves". I want to point out that when fitting general models, such as an ANOVA or a linear mixed model, the assumption of normality pertains to the residuals of the model, not necessarily the dependent variable itself. For this reason, I believe this approach may not be the most appropriate.

   To test the normality of the residuals (and other assumptions of a model fit) in R, packages such as `performance` (Lüdecke et al., 2021), have very convenient functions, including `check_model()`.

Unset
```
# Load package
library(performance)
# Check normality of residuals
check_normality(model)
# Visual check of model assumptions
check_model(model)
```

If the assumption of normality of residuals is violated, there are several options. First, you could use a function such as `check_distribution(model)` (also from the package `performance`), to get some idea of alternative distributions to then fit a generalised linear model. Alternatively, and perhaps preferably, you can use bootstrap techniques. This is useful because (1) bootstrapping the models could really help in dealing with issues such a (potential) non-normal residual distribution (see e.g., J. Fox, 2016 Chapter 21), but also because (2) having CIs would help in assessing effects (even in the absence of *p* values).

Minor points:

A. As mentioned by Dr. Keitel, I wonder what the rationale was behind setting an alpha of 0.02. To clarify, I am not against this decision, but I believe it should be explained in the text.

B. In a few places, you mention that some computations were performed using RStudio (e.g., lines 440 and 724). However, you should cite R instead, as it is the software that performs all computations. RStudio is a very useful IDE that aids in interacting with R, but while all your code can be run in R (regardless of whether you are using RStudio or a different IDE or interface, or even just directly from a command line), your code cannot be executed in RStudio without R.

C. On page 18 (line 474), you cite Figure 2, Panel A. However, unlike Fig. 1, Fig. 2 does not include panel labels, so perhaps these are to be added if you want to refer to a specific panel.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bolker, B. M. (2015). Linear and generalized linear mixed models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological Statistics: Contemporary theory and application* (pp. 309–333). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software*, *100*, 1–54. https://doi.org/10.18637/jss.v100.i05

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920965119. https://doi.org/10.1177/2515245920965119

Fox, J. (2016). Bootstrapping Regression Models. In *Applied Regression Analysis and Generalized Linear Models* (3rd ed., pp. 587–606). Sage. https://us.sagepub.com/sites/default/files/upm-binaries/68018_Fox_Chapter_21.pdf

Kurz, A. S. (2019). *Statistical Rethinking with brms, ggplot2, and the tidyverse* (version 1.0.1). https://bookdown.org/ajkurz/Statistical_Rethinking_recoded/

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.10.0) [Computer software]. https://CRAN.R-project.org/package=emmeans

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139