Review by Chris Chambers, 09 Jun 2023 09:51

I'm satisfied with the authors' response to my points.

Review by Markus Ploner, 14 Jun 2023 11:03

The authors have convincingly addressed my comments. I wish them good luck with their study and I am looking forward to seeing the results.

Review by Zoltan Dienes, 18 Jun 2023 17:55

The authors have spelt out more concretely - and usefully - a power calculation for the interaction. But my issue has not been addressed.

We apologize for not sufficiently addressing your concerns regarding the power calculation, to which we gave serious consideration. Please see the answers to the questions below.

1) Text has been added about how small the N has been in previous studies to detect the effect; but this does not tell us in itself how large the N must be to infer no effect should the effect be non-significant.

As suggested by Reviewer 1, we added a post hoc Bayesian interference analysis to the analysis plan in which we will compare H_0 (the model used in the main analysis, including the interaction term) to a model that only includes the main and random effects, but no interactions (H_1). The ratio between these two models expressed in a Bayes Factor (BF_{10}) will be used to assess the validity of the previous rejection of H_0 using the frequentist approach. The interpretation of the BF_{10} will be based on the interpretation table proposed by Lee and Wagenmakers (2013). This will allow us to confirm (or disprove) with higher confidence whether we rightfully rejected the null hypothesis. These adaptations are mentioned in the manuscript in the analysis table (pp. 19-20) as well as in the main text on p.17.

2) Power needs to be demonstrated for each test in the study design template, by a calculation for each test in itself, using the effect size for that specific test that one does not wish to miss out on.

As specified in our previous response, estimating effect sizes for such exploratory investigations as ours is rather difficult. Only very little literature on the modulation of ongoing oscillations assessed using a frequency-tagging of ongoing oscillations (FT-OO) approach is available. Crucially, the investigations that use this technique in combination with a cognitive modulation are non-existent. At this point, there is no literature available that would enable us to adequately estimate the effect of a cognitive task on the modulation of ongoing oscillations assessed using FT-OO and it is therefore extremely difficult to estimate effect sizes for such an investigation. Additionally, we cannot just assume that the effect of a cognitive task on the modulation of ongoing oscillations will be the same as - for example – the ones observed in studies investigating pain-induced event-related potentials. Therefore, while we did our best to estimate reasonable effects of the cognitive task on the modulation of ongoing oscillations, these are just “our best guesses” of possible effects.

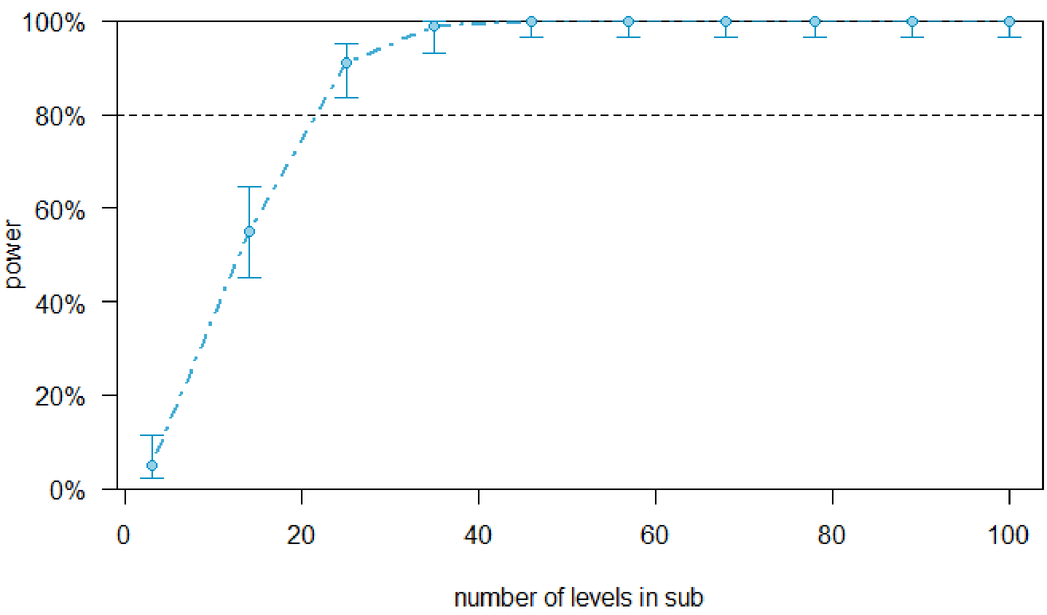
Currently, the only reported effect sizes of investigations using frequency-tagging during painful sustained periodic stimuli are reported in Mulders et al. (2020). More specifically, the effect size is only reported for the ANOVA conducted on the phase-locked response (η^2_p = 0.060 for the temperature*surface interaction) and not on the modulations of ongoing oscillations itself. We are therefore not able to calculate the sample size for each frequency band analysis based on effect sizes, since we simply don’t have reliable estimates. This is the reason why we resorted to the simulation approach described in the previous response.

Additionally, the data for the peak at the frequency of stimulation was only reported for the phase-locked response and not for the modulation of ongoing oscillations in the different frequency bands. Instead, the “area under the curve” (AUC) was reported for the modulation of ongoing oscillations. This measure does not reflect our outcome measure (peak amplitude at the frequency of stimulation) and can therefore not be used for the sample size calculation. The AUC of the alpha frequency band was mistakenly used for the data simulation in the previous submission, and we are grateful for the opportunity to correct our calculations.

The data for conditions HH and LL were thus simulated based on the peak amplitudes reported at the frequency of stimulation in the phase-locked response (HH: 0.59 ± 0.33 µV, LL: 0.41 ± 0.31 µV). We then tried to estimate the possible effect of the cognitive task based on the percentage change in ratings between the conditions observed in the pilot data, which was 18% between HM and LM. To reach HM and LM, the mean amplitude between HH and LL was calculated, and varied +9% for HM and +9% for LM (based on our assumption that the expectation of a higher rating would lead to an increase in amplitude at the frequency of stimulation). The same procedure was used to estimate the standard deviation of HM and LM. These values for mean and std. deviation for each condition were implemented to simulate a dataset including 2 levels of condition (matched, unmatched) and 2 levels of temperature (high, low). This simulated dataset was then used to create a Linear Mixed Model (LMM) using the same formula as proposed for the main investigation (see image below for details of the fitted model).

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: amplitude ~ temp * cue + (1 | sub)
   Data: sim_pl

REML criterion at convergence: 36.3

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.26060 -0.62681 -0.05643  0.73496  2.71843

Random effects:
 Groups   Name        Variance Std.Dev.
 sub      (Intercept) 0.0000   0.0000
 Residual             0.1066   0.3265
Number of obs: 48, groups:  sub, 12

Fixed effects:
                     Estimate Std. Error       df t value Pr(>|t|)
(Intercept)           0.80902    0.09424 44.00000   8.585 5.93e-11 ***
tempunmatched        -0.22772    0.13327 44.00000  -1.709 0.094559 .
cuelow               -0.48275    0.13327 44.00000  -3.622 0.000752 ***
tempunmatched:cuelow  0.44372    0.18848 44.00000   2.354 0.023089 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

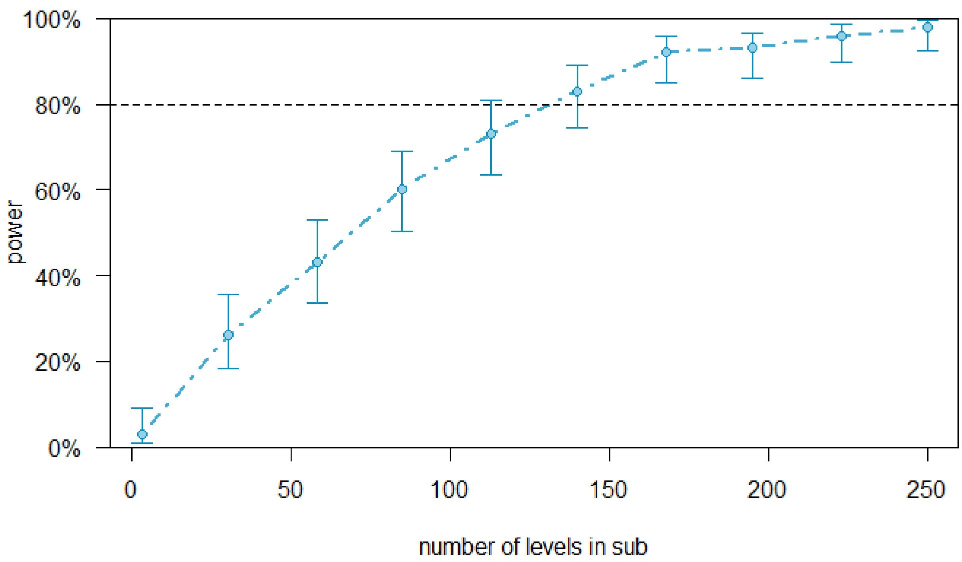
The LMM was then implemented in a sample size calculation using a target power of .9 and alpha of .02. According to this estimation, a minimum of 25 participants would have to be recruited to reach the targeted power (see figure below).



3) The power that is calculated is calculated for the size of interaction expected, not for the size that one does not want to miss out on.

We agree with the Reviewer that it would be important to calculate not just the expected effect but the effect size that one does not want to miss out on. We applied this calculation to our simulated data set, based on the suggestions of the Reviewer in the previous response. We thus simulated the LMM based on the data set of Mulders et al. (2020) (see answer above) and calculated the 80% confidence interval of the estimates of that LMM. We then replaced the estimates in the LMM with the lower bound estimates determined by the confidence interval. This changed the estimates for intercept (0.6855215) and slopes for temperature (-0.3991420), cue (-0.6554178) and their interaction (0.1804856).

Applying this updated model to the sample size calculation resulted in a suggestion to recruit over 150 subjects (illustrated in the figure below) to detect the smallest possible effect one does not want to miss out on. Due to the time-consuming nature of EEG experiments and limited financial resources, recruiting such a sample for one experiment would not be feasible for us to carry out. As a compromise, we would like to suggest raising the sample size from the minimum of 25 participants calculated above to 40 participants, to at least get a little bit closer to the sample size that would be able to detect the smallest effect one would not want to miss out on.



The sample size section of the manuscript (pp.6-8) has been adapted to reflect the changes described in this response.

References

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press. https://books.google.be/books?id=50tkAgAAQBAJ

Mulders, D., de Bodt, C., Lejeune, N., Courtin, A., Liberati, G., Verleysen, M., & Mouraux, A. (2020). Dynamics of the perception and EEG signals triggered by tonic warm and cool stimulation. *PLOS ONE*, *15*(4), e0231698. https://doi.org/10.1371/journal.pone.0231698