Manuscript #378, Second Resubmission

Effects of Auditory Stimuli During Submaximal Exercise on Cerebral Oxygenation

PCI Registered Reports

Recommender's Comments:

Your revised submission has now been re-evaluated by the reviewers from the previous round. As you can see, most points have been addressed and we are now within reach of Stage 1 in-principle acceptance. There are however a few remaining issues to resolve, including further specification of methodological details and clarification (and likely some further revision) to the statistical sampling plan. We look forward to receiving your response and revised manuscript in due course.

Response: We want to express our gratitude for your invitation to submit a revised draft of our Stage 1 manuscript. The input that we received from you and the two expert reviewers has been incredibly valuable and duly thought provoking. We have addressed point-by-point each of the reviewers' comments (see below).

Reviewers' Comments:

Reviewer 1 (Anonymous)

In their revised manuscript, authors meticulously addressed each comment and concerns point-by-point. Overall, their answers are both clear and precise enough and improved the manuscript reading and full understanding of the protocol.

Some last comments are proposed for further discussion.

Response: We would like to express our sincere gratitude for the time you dedicated to reading our contribution, and for providing us with your erudite remarks.

I suggest authors to indicate when needed the type of exercise they manipulate, rather to state "exercise protocol". As such, constant workload exercise at 5% above VT1 is preferred, and has to be written when appropriate.

Note that with reference to the physiological events of VT1, the true and exact measurements come from gas exchange measurement, even they can be identified indirectly by means of HRV. For sure, HRV could be a reliable, non-invasive, and low-cost method of assessing VT1 (and VT2).

Response: We have amended the text in accord with your helpful suggestion (see e.g., p. 4, l. 69–70; p. 6, l. 146–147). We appreciate your acknowledgment of the possibility to indirectly identify the physiological event of VT1 through HRV measurement.

VO2max criteria/exhaustion assesment: references included in Supplementary File 2 did not assess VO2max according to usual criteria (see Taylor et al., 1955) encountered in exercise physiology.

Response: Kindly note that the studies referenced in Supplementary File 2 were selected to align with our focus on a distinct physiological construct; one that differs from the evaluation of maximal oxygen uptake. It is worth mentioning that the concept of maximal oxygen uptake was not addressed in Supplementary File 2. In our assessment, we adopt an incremental test as proposed by Karapetian et al. (2008) to identify participants' first ventilatory threshold using heart rate variability. It is important to recognize that multiple physical tests exist for this specific purpose; however, the protocol we have chosen has undergone rigorous testing, validation and replication across experimental studies (e.g., Barreto-Silva et al., 2018; Bigliassi et al., 2017). Our goal with this test is to ensure that participants move beyond their first ventilatory threshold, which will allow us to standardise our experimental trials based on this physiological index.

It is likely that end point of exercise (voluntary exhaustion) will differ according to the profile of the subjects. Do you expect to provide verbal encouragement, ask RPE score?

Response: Participants will be pre-screened to ensure that they are recreationally active (see p. 7, l. 168–171) and our final sample will be homogeneous in this regard. Nonetheless, we do agree with you that the point of voluntary exhaustion is likely to vary in accord with each individual's physiological capacity. This is the reason for which we proposed to remove outliers in terms of exercise duration prior to subsequent analyses (now located on p. 14–15, l. 361–365) and to compute our dependent variable as a percentage of the 5%-above-VT1 phase rather than as an absolute time (now located on p. 14, l. 348–351).

Verbal encouragement will not be provided to participants, given that this could serve as a confounding factor and have a differential effect across participants (see e.g., Midgley et al., 2018). We will take psychophysical measures in the form of RPE (CR-10 scale; now located on p. 10, l. 241–246) at 2.5-min intervals starting with 2.5 min after VT1 has been reached (see Figure 2, p. 29; “Experimental Protocol for the Proposed Study”). This will serve as another means by which to check whether our desired intensities are reached.

Authors replied that VT1 is the exercise intensity (i.e., power output, cycling ergometer) to maintain (start point of exercise protocol) and that will soon drift to RCP and beyond, towards volitional exhaustion. Here, there is an issue! We know in exercise physiology that VO2 steady state at or slightly above 5%) VT1 can be maintained for a while; the slow drift you reported occurs at or above VT2 (see the extensive research work of Jones, Poole and co-authors on the VO2 kinetics during constant work rate exercise. Maintaining power output at VT1 is belonging to the moderate exercise domain, without important drift in any physiological variables (respiratory, cardiac, muscle..) except after more than 20-25 minutes, especially in Lab settings without cooling effect. In this case, a drift in HR (maybe HRV) can occur due to different underlying origins (e.g., heat-induced hypervolemia, passive-heat stress increases heart rate), such as blood flow redistribution among tissue that can affect cerebral oxygenation. It means that authors should report respiratory rate and heart rate time course as suggested, and verified the changes over time in these two variables with respect to the cerebral oxygenation response. Note that minute ventilation and cardiac output are more robust and complete (volume x rate changes) into the cardiorespiratory monitoring.

Response: Both respiratory rate and heart rate will be recorded (now located on p. 10–11, l. 250–261) to provide the means by which to report time-course changes and relate them to cerebral responses, as you kindly suggest (see p. 15, l. 369–370). Thank you for your remark regarding the cardiorespiratory monitoring. Note that with a subtle form of manipulation such

as auditory stimuli, the use of online gas analysis can obfuscate the influence of music, given the “attentional demands and potential anxiety-inducing nature of the apparatus” that is required (see Terry et al., 2020, p. 103).

Competing hypothesis not to rule out.
With exhaustion, whatever the exercise protocol (incremental, constant work, rate), reduced CO2 levels may occur and result in vasoconstriction and reduced cerebral blood flow, and so in NIRS-parameters (O2Hb).

Response: The proposed protocol is predicated on a repeated-measures design and will provide the means to examine the specific effect of music (vs. audiobook and control) at a similar level of exercise intensity. Variations in CO_2 levels that are not due to the experimental manipulation should be similar across conditions; we will be able to verify that this is indeed the case by analysing the non-cortical haemodynamic responses monitored by use of the photoplethysmograph sensor (see p. 11, l. 262–268; p. 15, l. 369–370).

Reviewer 2 (David Mehler)

Thank you for addressing all my concerns in detail, they are sufficiently addressed now. Only the sampling plan requires in my view further clarification and possibly a revision.

Response: We thank you for the time devoted to evaluating our contribution. Your insightful feedback and highly thoughtful suggestions have significantly improved the quality and clarity of our work.

4) The sampling plan describes a smallest telescope approach to establish a SESOI, which I think is a val approach given the risk for bias and the challenge in establishing a mechanistic SESOI. Yet, sample size estimates seem based on effect size estimates in previous literature.

Could the authors please clarify? Further, given their design, I am wondering whether (nested) Bayesian hypothesis testing may be more sensitive and robust, as it provides flexible stopping options?

Response authors: To justify our sample size, we decided to rely on statistical power, namely the probability of detecting an effect (i.e., not accepting the null hypothesis) provided that this effect exists. The sample size computation was not performed using the SESOI because the latter corresponds to the effect size that an earlier similar study would have had 33% power to detect. However, for a sample size justification, we want the power to be fixed at 80%, which corresponds to a β level of .20. We made this choice in a way that the Type I error is four times less likely to occur than the Type II error (see Cohen, 1988). The expected effect size was estimated from previous similar studies in terms of variables of interest and experimental design, as recommended by Lakens (2022).

Thank you for the explanation. It is, however, not in line with what is stated in manuscript, which includes a power calculation for 90% (which is the PCI RR requirement). Also the explanation of the sampling plan in the manuscript is not very clear "The small telescopes approach was used to determine the smallest effect size of interest (SESOI; i.e., the difference that is considered too small to be meaningful; Simonsohn, 2015). Accordingly, the SESOI was set to the effect size that an earlier study would have had 33% power to detect (Lakens et

al., 2018).". Please explain the telescope approach and why the SESOI was set to effect size that an ealier study would have had 33% power to detect it. What is the value for this effect size?

Response: Thank you for pointing this out – this is indeed an oversight on our part in the responses to reviewers' comments document. As stated within the manuscript (now located on p. 8, l. 179–190), the power is fixed at 90%, which corresponds to a β level of .10.

The small telescopes approach states that the smallest effect size of interest (SESOI) should be set to the effect size that an earlier, similar study would have had 33% power to detect, meaning that "the odds are at least 2:1 against obtaining a statistically significant effect" (Simonsohn, 2015, p. 562). For example, the earlier study we used to compute the SESOI using the small telescopes approach for H_1 and H_2 across the dorsolateral prefrontal cortex is that of Oh et al. (2018), with a four-cell, between-subjects design and 20 observations per cell. Thus, an effect size of 0.38 would give this study 33% power (for $\alpha = .02$ and a one-tailed test), which corresponds with our SESOI. Note that there was an error in the manuscript (but not in Table 1, p. 25–26): the results of Oh et al. (2018) were used as parameters to compute the SESOI for H_1–H_2, but *only* across the dorsolateral prefrontal cortex. For H_1–H_2 across the medial prefrontal cortex and H_3, the results of Ozawa et al. (2019) were used. For H_4, the results of Guérin et al. (2021) were used (in line with the articles used to derive the effect sizes for the power analysis). We have made all necessary amendments to the text (see p. 8, l. 194–199).

Table 1 lists the mapping of hypotheses to the sampling plan. Some assumed effect sizes are unbeliably large (>1.3) and it not clear where these values stem from. As a general point, it is also very questionable to base the sampling plan on previous effect size estimates from literature that was not preregistered. Simulations and meta-analyses suggest that on average non-preregistered literature (that may be subject to all forms of biases) yields effect size that 2-3 times larger compared to preregistered work (https://journals.physiology.org/doi/full/10.1152/jn.00765.2017; https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00813/full; https://www.nature.com/articles/s41562-019-0787-z). Please clarify this point and consider revising the sampling plan accordingly.

Response: The references used to estimate the effect sizes are cited within the manuscript (now located on p. 8, l. 182–185). Kindly note that the effect size used to compute the required sample size for H_4 does originate from a registered report (Guérin et al., 2021). Nonetheless, we fully agree with you that in the (non-preregistered) literature, effect sizes usually gravitate towards higher values. The large effect sizes ($d > 1.3$) you referred to did indeed result in a small required sample size ($N = 9$). It is however important to note that our final sample will be $N = 36$ (driven by the number of participants needed to test H_4, for which a registered report effect size was used). Accordingly, this sample size will allow us to detect effect sizes that are superior or equal to $d = 0.57$ ($\alpha = .02$; $1\text{-}\beta = .90$; paired-samples t test), which is still inferior to an effect size that is two times smaller than the assumed ones (i.e., 0.69 and 0.685 for H_1–H_2 and H_3, respectively). Thus, we are confident that our final sample of participants will allow us to test our research hypotheses, even if the effect sizes we used for specific hypotheses are inflated due to being derived from non-preregistered articles.

References

Barreto-Silva, V., Bigliassi, M., Chierotti, P., & Altimari, L. R. (2018). Psychophysiological effects of audiovisual stimuli during cycle exercise. *European Journal of Sport Science, 18*(4), 560–568. https://doi.org/10.1080/17461391.2018.1439534

Bigliassi, M., Karageorghis, C. I., Wright, M. J., Orgs, G., & Nowicky, A. V. (2017). Effects of auditory stimuli on electrical activity in the brain during cycle ergometry. *Physiology & Behavior, 177,* 135–147. https://doi.org/10.1016/j.physbeh.2017.04.023

Guérin, S. M. R., Vincent, M. A., Karageorghis, C. I., & Delevoye-Turrell, Y. N. (2021). Effects of motor tempo on frontal brain activity: An *f*NIRS study. *NeuroImage, 230,* Article 117597. https://doi.org/10.1016/j.neuroimage.2020.117597

Oh, S., Song, M., & Kim, J. (2018). Validating attentive locomotion training using interactive treadmill: An *f*NIRS study. *Journal of Neuroengineering and Rehabilitation, 15,* Article 122. https://doi.org/10.1186/s12984-018-0472-x

Ozawa, S., Kanayama, N., & Hiraki, K. (2019). Emotion-related cerebral blood flow changes in the ventral medial prefrontal cortex: An NIRS study. *Brain and Cognition, 134,* 21–28. https://doi.org/10.1016/j.bandc.2019.05.001

Midgley, A. W., Marchant, D. C., & Levy, A. R. (2018). A call to action towards an evidence-based approach to using verbal encouragement during maximal exercise testing. *Clinical Physiology and Functional Imaging, 38*(4), 547–553. https://doi.org/10.1111/cpf.12454

Terry, P. C., Karageorghis, C. I., Curran, M. L., Martin, O. V., & Parsons-Smith, R. L. (2020). Effects of music in exercise and sport: A meta-analytic review. *Psychological Bulletin, 146*(2), 91–117. https://doi.org/10.1037/bul0000216

Simonsohn, U., 2015. Small telescopes: Detectability and the evaluation of replication results. *Psychological Science, 26*(5), 559–569. https://doi.org/10.1177/0956797614567341