Dear Dr. Chambers,

Thank you for your favorable reply and the opportunity to revise our manuscript. We answered the questions of the reviewers below and adjusted the manuscript accordingly. During the revision of the manuscript, we encountered a minor mistake in the results regarding the independent sample t-tests of the audiometric levels between our two groups. In the current version, we now report the correct statistics and further specified our procedure to compute the average hearing ability for our participants. This specification was not included by now and we believe that this further clarifies our matching approach.
We are very much looking forward to your further evaluation.

Sincerely and on behalf of the co-authors,
Lisa Reisinger

Reviewer 1:

This report is a great example of the scientific process. The hypotheses and aims were well laid out, based on prior work. Methods were well-planned in advance also, and deviations from the methods of the prior study are well-justified, and their importance clearly discussed. The aim of the study was to replicate previous novel and exciting findings with respect to possible tinnitus mechanisms, in the form of auditory predictive tendencies, and eliminate hearing loss as the explanation (which has turned out to be the basis of most previously reported 'tinnitus' studies that were not rigorously controlled). The study achieves its aims, in an independent group of participants, replicating the original findings. This is a big achievement. The discussion strikes a good balance of recognising the potential importance of the findings, whilst openly acknowledging the remaining uncertainties, and being clear between what its findings demonstrate, and what is still speculation.

I am supportive of publication, but would recommend some minor and moderate revisions first.

Thank you very much for your favorable evaluation and your constructive recommendations on how to further improve the manuscript.

Moderate:

Given that matching for hearing was the particular focus and novelty of this study, it would be useful to have more information about hearing matching, for instance:

- Were the groups only compared in terms of hearing for the average pure tone threshold across all frequencies (for each ear), or were significant differences in individual frequencies sought?

Thank you very much for pointing that out. Statistical comparisons between the groups were based on the average pure tone threshold for each ear. However, to guarantee the best possible

matching, participants in the tinnitus group were manually matched with controls by visually inspecting the individual hearing profiles. Therefore, we tried to acknowledge individual patterns in hearing ability and searched for the best match for each tinnitus patient. We included a statement in the corresponding paragraph in our manuscript to increase clarity.

Line 291-293:

"By visually inspecting individual patterns in the hearing profile, we were able to take into account individual outliers in specific frequencies and searched for a control participant with a similar hearing pattern."

- Can the authors please provide a figure showing the group mean and standard error of the hearing profiles for each ear?

Thank you for this suggestion. We agree with you that a figure is reasonable in this regard and included it in the according paragraph.

Page 12:

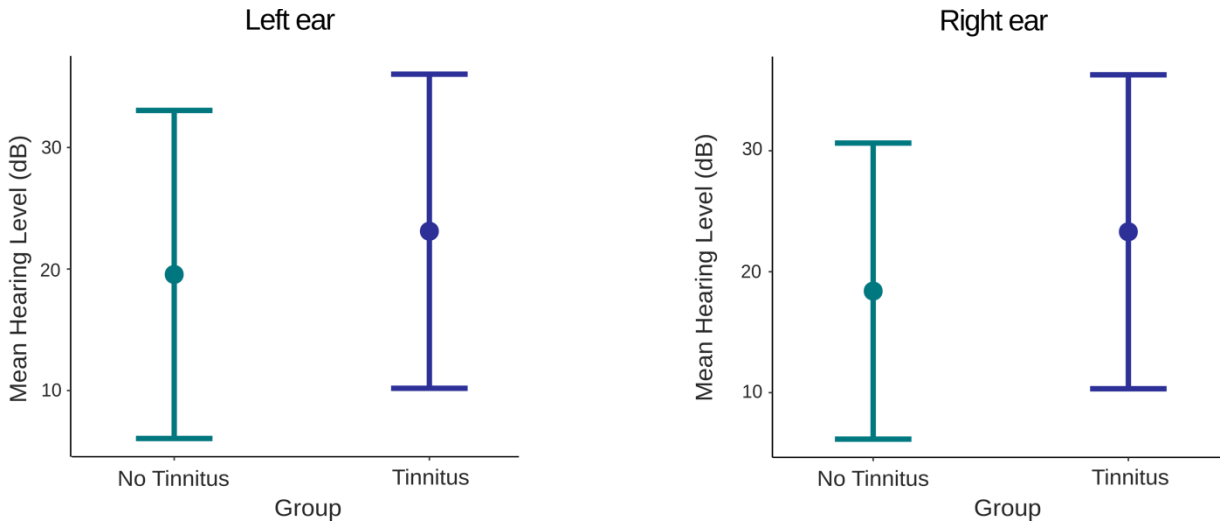

Fig. 1: *Mean hearing levels (dB) for the tinnitus and control group, displayed for each ear separately.*

What does it mean that controls have a negative anticipatory decoding accuracy difference between ordered and random? This is the one part that sits slightly uneasily with me. I do not doubt the findings, but rather I just wonder what it means that people are either 'un-representing', or 'contra-representing' upcoming stimuli based on learned regularities. As just one possibility of many, could this be because in the random condition it is more likely that the upcoming stimulus was presented 2-3 stimuli back (the probability of the immediately preceding stimuli clearly being exactly 25% in each condition, and therefore equivalent)? I am finding it hard to reconcile something that is less negatively predictive of the upcoming stimulus being an anticipatory prediction. Or, do the authors think that there are two opposing processes here? The authors do highlight this point and make quite a lot of discussion already about it, so perhaps they do not feel in a position to speculate further, but I would encourage them to ensure they have considered all possibilities as fully as they can.

We fully agree with you that this finding is quite difficult to discuss and interpret. Thank you for pointing out further possibilities, we highly appreciate your input. As you acknowledged, we already aimed to discuss our effect extensively and therefore we do not feel in a position to speculate a lot further about it. We encourage future work to widely examine this phenomenon with novel approaches to gain more insights and added a corresponding section.

Line 798-801:

"As indicated above, this interpretation remains speculative at this stage and requires further investigation. We acknowledge that there are other possibilities on how to interpret the below-chance decoding and therefore encourage future work to implement novel approaches to gain more insights into the exact underlying mechanisms driving this effect."

Minor:

In the discussion, I wonder about including the following considerations for future work

- Attention as another explanation for differences in tinnitus vs. control groups. (Though, MMN studies in tinnitus show diminished responses to sub-tinnitus frequency ranges in people with tinnitus in passive listening paradigms, which is hard to reconcile with increased predictive tendencies).

Thank you for your input. We agree with you that attention can be potentially influential and should not be neglected. We included a corresponding section in the discussion.

Line 856-863:

"Further, apart from the methodological specifications mentioned above, attention appears to be relevant in tinnitus as well, both in the generation and the formation of predictions (Durai et al., 2018; Roberts et al., 2013; Sedley et al., 2016). In the current study, we used a passive listening task including a movie to reduce attentional focus on the presented stimuli. Therefore, we can not draw conclusions whether differences in attention between the two groups had an influence on the effects. Future studies should include more manipulations of attention to investigate its relevance."

- Measuring hyperacusis scores in participants (but, these typically correlate with tinnitus distress, which already has not shown a significant correlation)?

Thank you for pointing that out. Even though hyperacusis is typically correlated with tinnitus distress, we agree with you that hyperacusis scores can provide further information about participant characteristics.

Line 864-870:

"Additionally, we rigorously controlled for hearing loss and ruled out any influence of tinnitus distress on the effects, however, we did not screen our participants for hyperacusis. This hypersensitivity to mild sounds is widely correlated with the sensation of tinnitus and underlying neural mechanisms are potentially intertwined with tinnitus processes (Schilling et al., 2023;

Yukhnovich et al., 2023; Zheng, 2020). Screening for hyperacusis in future work can therefore reveal more details on participant characteristics influencing predictive processing."

I think that future work using tones around the tinnitus frequency could be highly informative, as it would allow probing of specific differential anticipatory activations between tinnitus and non-tinnitus tones, and might highlight correlates of tinnitus frequency predictions themselves, to complement the existing findings relating to auditory predictive tendencies more generally.

We appreciate your feedback and agree with you on the importance of also including tinnitus-specific frequencies. We added this approach to our discussion.

Line 850-855:

"Next to implementing broader frequency spacing, future studies should also focus on the opposite direction, namely narrowing down the sound range around the individual tinnitus frequency. Recent MEG studies already focused on the tinnitus frequencies in their designs (Reisinger et al., 2023) and in terms of predictive coding, this approach can allow insights into the predictive processing of tinnitus vs non-tinnitus tones."

Minor

Abstract:

Clarify whether 80 is number per group or total number of participants (i.e. the latter)

Thank you for pointing that out. We clarified our number of participants in the abstract.

Line 41-43:

"Analyses encompassed data from 80 participants. 40 participants with tinnitus and 40 control subjects without tinnitus were not only matched for age and gender, but importantly also in terms of hearing loss."

Introduction:

Lines 85-90: Mention some other work (e.g. Adjamian et al. 2012) finding no differences in resting-state delta or gamma between tinnitus and hearing-matched controls.

Thank you for your input, we added references for other work as well.

Line 85-89:

"In humans, resting-state M/EEG studies reported divergent patterns, especially in the delta, alpha and gamma frequency band ranges within and beyond auditory regions (de Ridder et al., 2011; van der Loo et al., 2009; Weisz et al., 2005; for an opposite finding see e.g., Adjamian et al., 2012; Zobay and Adjamian, 2015)."

Results:

Line 650: Please make it clear how hearing loss was controlled for in this statement (e.g. by removing the effect of hearing loss by linear regression)?

We appreciate your feedback and rephrased the sentence to clarify our approach.

Line 654-658:

"For this purpose, we used logistic regression to analyse the presence of tinnitus as a dependent variable, and (mean) hearing loss and (mean) difference in decoding accuracy between the ordered and random tones in the pre-stimulus interval as independent variables, controlling for hearing loss by including mean hearing ability as a covariate."

Discussion:

Line 687: 'indication' should be 'indicating'

Line 692: A full stop and space is missing

Thank you for your careful evaluation of our manuscript, we corrected the errors.

The discussion spends a lot of words repeating the main findings, and could perhaps be streamlined somewhat, without losing content or clarity.

We appreciate your feedback and removed a repetition of our main findings (see line 717ff.). Deleting more parts of our discussion would in our opinion, however, diminish clarity since we focus on different aspects of our findings in different sections of the discussion.

Reviewer 2:

2A. Whether the data are able to test the authors' proposed hypotheses (or answer the proposed research question) by passing the approved outcome-neutral criteria, such as absence of floor and ceiling effects or success of positive controls or other quality checks.

Yes, the data are able to test the author's proposed hypotheses. For example, to strengthen their results for H2, they added equivalence testing (p. 17, l. 446). Regarding H3, only 31 participants out of 40 filled in the Mini-TQ and they state that they "correlated the individual tinnitus distress values with the mean decoding accuracy of each individual in the previously analyzed pre-stimulus interval." (p. 17, ll. 450-452). For the reader, it is unclear if you matched the Mini-TQ data for these specific 31 subjects. Could you clarify this and/or add some more detail here?

We appreciate your feedback. We added a sentence to clarify that the subjects without Mini-TQ data were excluded from this analysis.

Line 461-464:

"Importantly, information regarding tinnitus distress was not available for all 40 tinnitus subjects but solely 31 subjects were included in this analysis. We therefore excluded the nine subjects that did not complete the Mini-TQ for this specific analysis."

2B. Whether the introduction, rationale and stated hypotheses (where applicable) are the same as the approved Stage 1 submission. This can be readily assessed by referring to the **tracked-changes manuscript supplied by the authors**.

Yes, the introduction, rationale and hypothesis were identical to the approved Stage 1 after inspecting the tracked-changes manuscript. The only changes made were due to the tense used.

Thank you for your careful evaluation of our manuscript regarding the changes we made to our approved Stage 1 submission.

2C. Whether the authors adhered precisely to the registered study procedures.

Overall yes, some small adjustments were made, i.e.., p.17, l. 431, for the time window assessed.

Thank you for your feedback. We hope that we made our adjustments as clear and comprehensible as possible.

2D. Where applicable, whether any unregistered exploratory analyses are justified, methodologically sound, and informative.

Indeed, there were exploratory analyses added, see the section “Exploratory results”. This was done to compare the current results with the results of Partyka et al. (2019). In my opinion, these elaborations add to the quality of the paper.

Thank you for your constructive feedback on our addition of exploratory analyses.

2E. Whether the authors’ conclusions are justified given the evidence.

Yes, the conclusions are justified.

We appreciate your careful evaluation of our discussion section.

Reviewer 3:

In this stage 2 registered report, participants with tinnitus are shown to display relatively enhanced tone frequency specific pre-activation (i.e. larger differences in decoding accuracy) compared to matched tinnitus-free controls. The authors have carefully complied with their registered Stage 1 study design. Analytical methods have been performed as described in the Stage 1 protocol. Furthermore, the authors have presented their results completely and transparently.

Throughout this report, I have identified some minor issues where information may be lacking or interpretation of the findings might be improved upon. I have outlined these items below. Additionally, I recommend a careful readthrough to identify some very minor grammatical errors (some of which I have also addressed below).

Thank you for your favorable and constructive feedback on our Stage 2 report.

Results

Figure 4 and the corresponding results section (L502 onwards): Overall, it seems that the primary outcome of this report (the difference in decoding accuracy) shows quite a bit of between-subject variability. Even at the time point with the most pronounced group differences (panel B of Fig. 4), these differences in decoding accuracy show a significant amount of variation, with the distributions for both groups (Tinnitus vs. No Tinnitus) overlapping considerably. Could the authors address this considerable variability and offer some potential explanations for the observed variation? Are there parameters present in the current dataset that are potentially

associated with the difference in decoding accuracy, and that could offer some more insight? Or do the authors hypothesize that this between-subject variability is largely driven by factors that were not investigated in the context of this study?

Thank you for pointing that out. In our opinion this between-subject variability is driven by factors that are not known yet. Since we carefully matched our groups in terms of hearing loss, we confidently exclude this factor as a potential confound, however, tinnitus is experienced highly individually and there are possibly other factors that influence the predictive processing in participants with and without tinnitus. We acknowledge this peculiarity in our discussion.

Line 770-776:

"Interestingly, neural prediction scores showed high between-subjects variability (see Figure 5B). Since we ruled out hearing loss as a potential confound, other - yet unknown - factors seem to influence neural prediction scores in both tinnitus and controls. Although our current design does not include other variables to analyze interindividual differences, this phenomenon opens up possibilities for future work to further investigate this variability. "

L560 "Together, with the result reported by Partyka et al. (2019), our results strongly support the notion that unspecific distress due to tinnitus is not a good explanation for tinnitus": This seems to be a typo – could it be that the authors meant that "unspecific distress due to tinnitus is not a good explanation for the identified differences in decoding accuracy", for example?

We appreciate your careful evaluation of our manuscript and agree with you on that. We corrected the sentence accordingly.

Line 570-573:

"Together, with the result reported by Partyka et al. (2019), our results strongly support the notion that unspecific distress due to tinnitus is not a good explanation for the identified differences in decoding accuracy."

L646 and following: Here, tinnitus presence is significantly predicted by the mean difference in decoding accuracy. Could, potentially, the opposite be done, i.e. predict the mean difference in decoding accuracy by group (Tinnitus vs. No Tinnitus)? Would this give us some insight into how much of the between-subject variability in this difference in decoding accuracy (see also my comment above) is explained by the presence of tinnitus, versus the proportion of the variance that would have to be explained by other factors? It might be interesting to have such insights in order to compare the current results to other established differences between tinnitus patients and controls based on evoked MEG or EEG that currently exist in the literature.

Thank you for your constructive feedback. We analysed the opposite model as well and included it in the corresponding section. The group was revealed as a significant predictor for the difference in decoding accuracy, strengthening our claim that tinnitus influences predictive processes in the brain.

Line 665-669:

"Reversing this model, we also analysed whether the presence of tinnitus can predict differences in decoding accuracy. This model was significant as well (b=0.003, SE=0.001, p=.018), indicating a statistically relevant influence of tinnitus on the differences in decoding accuracies and further supporting our previous findings."

Discussion

L686 What would this "neural prediction score" entail specifically? Do the authors propose a concrete numerical score, e.g. the values that are shown in Figure 4B, to be used in future applications? Or is this a broader concept?

We appreciate your input. Since the values in Figure 4B showed high between-subject variability, we do not want to propose this concrete numerical score as appropriate for future applications. Moreover, as you already suggested, we aim to introduce the "neural prediction score" as a broader concept to initiate further investigations that specifically target this phenomenon. Since

we do not understand the underlying factors entirely, determining our calculated difference values as concrete scores appears not justifiable with our current state of knowledge. We specified this in the according paragraph.

Line 703-706:

“We will refer to this tone frequency specific pre-activation as a “neural prediction score” in tinnitus, indicating deviations between random and predictable sounds that are specifically found in tinnitus patients. By now, this “neural prediction score” specifies merely a broader concept than a concrete numerical score. Influencing factors and underlying mechanisms are still not fully understood and it is therefore not applicable yet to determine a concrete value to quantify neural predictions.”

L687 “indication” does not make sense grammatically here, did the authors meant to write “indicating”?

Thank you for pointing that out, we corrected the sentence accordingly.

Line 700-703:

“We will refer to this tone frequency specific pre-activation as a “neural prediction score” in tinnitus, indicating deviations between random and predictable sounds that are specifically found in tinnitus patients.”

L746 “As they follow the onset the subsequent tone”: I think an ‘of’ may be missing here.

Thank you for your careful evaluation, we added an “of” to the text.