Dear Chris Chambers,

Thank you for your favorable reply and the opportunity to revise and resubmit our manuscript. We answered the questions of the reviewers and adjusted the manuscript accordingly. We are very much looking forward to your further evaluation.

Sincerely and on behalf of the co-authors,

Lisa Reisinger

Response to Reviewer 1 (Will Sedley):

This manuscript details research background, rationale, questions, hypotheses and methods for a study that has completed data collection, but not yet commenced data analysis. It focuses on the timely question of whether differences in how individuals form sensory predictions is an important determinant of whether they develop tinnitus following hearing loss.

Overall, I think this is a very strong submission in a number of regards, and I have no significant concerns.

We appreciate your favorable evaluation and constructive input.

- Line 67: "a highly predictive trigger" might be better phrased as "the main risk factor" for a few reasons (avoiding multiple uses of 'predictive' throughout the manuscript, and acknowledging a possible differences between risk factors, which are likely long-term, and triggers, which may be short-term and transient)

We appreciate your suggestion and agree that “main risk factor” is a better phrasing to emphasize the long-term impact of hearing loss on tinnitus. We therefore changed the sentence.

Page 3:

“Hearing loss has been identified as **a main risk factor** for tinnitus.”

- Lines 114-117: The argument is made that the previous finding of stronger anticipatory predictions in people with (compared to without) tinnitus is interpreted as indicating that these predictive tendencies are a risk factor for tinnitus. This is a reasonable preferred explanation, but other reasonable possibilities include tinnitus being the cause of altered predictive tendencies, and also of their being a third factor that is responsible for both predictive tendencies and tinnitus development.

We agree with you that these are reasonable explanations as well. More decisive evidence would need to implement longitudinal studies. Since these are difficult to implement in humans,

research efforts on suitable animal models would be very valuable. In humans, at this stage, it makes sense to test predictions in cross-sectional studies first that would be in line with our “preferred explanation“. However, we added other explanations in our introduction as well. Further, we will target this topic in more detail in the discussion of the final report and sketch possible approaches to further address this question.

Page 6:
“In a previous work (Partyka et al., 2019), we proposed that, given the tendency to predict auditory events, individuals with stronger prediction tendencies are more vulnerable to developing tinnitus (this is similar to the strong prior hypothesis developed by Corlett et al., 2019). **However, using a cross-sectional design alternative explanations cannot be excluded with certainty, such as tinnitus being the cause of altered prediction tendencies or that there is a third variable being responsible for predictions and tinnitus development. Adjudicating research** would require longitudinal studies **in humans or animals**. As such research is **challenging to implement, especially in humans**, we first focussed on finding group differences between individuals with and without tinnitus.”

- Line 274: 'within the range of hearing' might be better phrased as 'within a region of normal audiometric thresholds'

We agree that your proposed phrasing is more concrete and we changed the manuscript accordingly.

Page 12:
“We reduced the carrier frequencies to a maximum of 1000Hz to further ensure that the sounds provided were within **a region of normal audiometric thresholds**.“

I also have one larger point, though it is more of a suggestion for the authors to consider, rather than anything needing to necessarily be incorporated into this manuscript. The use of the time-generalised classifier to reveal anticipated stimuli is clearly very strong. However, the majority of studies examining stimulus-related predictions in tinnitus use some version of the mismatch negativity (MMN) paradigm. Therefore, to facilitate comparison of the results of this study to other studies of predictions in tinnitus, I wonder whether the authors might also perform some kind of equivalent to an MMN analysis of these data: i.e. a straightforward analysis based on the evoked field waveform itself. Whilst there are not straightforward 'standards' and 'deviants' here, it should still be possible to compare physically identical stimuli which differ according to how unexpected they were based on auditory sequence properties from that block (and whether or not they are a repetition of the preceding stimulus, as a more trivial factor to account for).

Thank you for your suggestions, we appreciate it. We agree with you that these kinds of analyses are interesting and would yield the opportunity to integrate results with other studies (Sedley et al., 2019; Weisz et al., 2004). Especially, targeting repetitions, which are more unexpected in the ordered condition could be an interesting avenue. However, we decided to not implement it for this manuscript, as we want to place the focus on anticipatory predictions (or pre-activations of sensory templates) than on post-stimulus prediction errors as in MMN analyses and quantify the extent to which results are influenced by hearing loss as main "confound". Adding the MMN analysis would make the manuscript more difficult to follow in our view. It would be however interesting as a topic for a separate manuscript that could cross-reference to the decoding approach wherever adequate. As a note: we would of course share the data upon publication, so that other groups could also test their ideas on this data-set.

Response to Reviewer 2 (Pia Brinkmann):

I enjoyed the opportunity to review the submitted and interesting proposal 'Is enhanced predictive engagement in tinnitus independent of hearing loss?'

We appreciate your favorable evaluation and constructive input.

Major issues:

Title: It is not known if persons with tinnitus experience enhanced predictive engage for auditory input. This proposal tries to replicate findings from yet to be published study (i.e., the preprint from Partyka et al., 2019). Therefore, the current title could be misleading as is and should be re-formulated or more precise.

Thank you for your input regarding the title of the manuscript. We rephrased it to make it more precise and to remove misleading assumptions about enhanced predictive engagement in tinnitus. We agree with you that this has not yet been shown in a published work.

Title:
"Registered Report: Are anticipatory predictions enhanced in tinnitus and independent of hearing loss?"

Introduction:

The introduction lacks depth of relevant literature. Only the altered-gain hypothesis and the Bayesian inference framework are addressed. However, there are other models that also try to explain increased spontaneous activity in the central nervous system, such as the noise cancellation model (Rauschecker et al., 2010) or the thalamocortical dysrhythmia hypothesis (Llinas et al, 1999 or De Ridder et al, 2015).

We agree that there are several models and hypotheses that aim to explain tinnitus development and we added the two mentioned theories to the manuscript.

Page 4-5:

“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). **In this regard, the thalamocortical dysrhythmia hypothesis proposes as well that tinnitus development is a consequence of altered neural thalamo-cortical coherence. Findings concerning this hypothesis state that tinnitus is both related to enhanced theta, delta and gamma-band activity in the auditory cortex as well as decreased connectivity between the thalamic medial geniculate body and auditory regions (Brinkmann et al., 2021; De Ridder et al., 2015; Llinas et al., 1999).**

Other potential explanations for tinnitus perception are for instance noise cancellation models (Rauschecker et al., 2010). The noise cancellation model states that due to cochlear lesions and resulting neuroplastic reorganization, hyperactivity in auditory pathways generates or enables acute tinnitus. Normally, noise cancellation mechanisms in the limbic system start identifying and inhibiting the wrong sound signal but in cases of dysfunctions in the limbic system and especially in the anterior cingulate cortex, noise signals persist consciously as tinnitus in the auditory system. Permanent dysfunctions lead to cortical reorganizations which result in chronic phantom sound perceptions (Rauschecker et al., 2010; Song et al., 2015). More recent work also states alterations in a more general cognitive network including prefrontal, limbic, and subcortical structures which lead to the chronicity of tinnitus (Lan et al., 2022).

Apart from a significant shortage of data bridging animal and human research in these different frameworks, empirical support in humans is weak, difficult to replicate, and marked by strong interindividual variability (Eggermont and Roberts, 2015; Elgohyen et al., 2015)**. Beyond the lack of solid evidence, the models face further explanatory challenges** (Sedley, 2019).”

References:

Brinkmann, P., Kotz, S. A., Smit, J. V., Janssen, M. L., & Schwartze, M. (2021). Auditory thalamus dysfunction and pathophysiology in tinnitus: a predictive network hypothesis. *Brain Structure and Function*, *226*(6), 1659-1676. doi: https://doi.org/10.1007/s00429-021-02284-x.

De Ridder, D., Vanneste, S., Langguth, B., & Llinas, R. (2015). Thalamocortical dysrhythmia: a theoretical update in tinnitus. *Frontiers in neurology*, *6*, 124. doi: https://doi.org/10.3389/fneur.2015.00124.

Lan, L., Chen, Y. C., Lu, L., Xu, J. J., Yin, X., Wu, Y., & Cai, Y. (2022). Topological features of limbic dysfunction in chronicity of tinnitus with intact hearing: New hypothesis for 'noise-cancellation'mechanism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *113*, 110459. doi: https://doi.org/10.1016/j.pnpbp.2021.110459.

Llinás, R. R., Ribary, U., Jeanmonod, D., Kronberg, E., & Mitra, P. P. (1999). Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proceedings of the National Academy of Sciences*, *96*(26), 15222-15227. doi: https://doi.org/10.1073/pnas.96.26.15222.

Rauschecker, J. P., Leaver, A. M., & Mühlau, M. (2010). Tuning out the noise: limbic-auditory interactions in tinnitus. *Neuron*, *66*(6), 819-826. doi: https://doi.org/10.1016/j.neuron.2010.04.032.

Song, J. J., Vanneste, S., & De Ridder, D. (2015). Dysfunctional noise cancelling of the rostral anterior cingulate cortex in tinnitus patients. *PloS one*, *10*(4), e0123538. doi: https://doi.org/10.1371/journal.pone.0123538.

I-171-172; H3: It is not clear why you expect no influence of tinnitus distress on anticipatory processing. If H1 turns out to be false, do you still attempt to confirm H3? In the design table (Table 1) it is not clear why H3 is linked to H1 and not H2. In addition, it is not clear in the 'Analysis plan' in Table 1 that you will look at pre-stimulus mean decoding accuracies.

Thank you for pointing that out. We based this hypothesis on our previous findings (Partyka et al., 2019) where tinnitus distress did not correlate with anticipatory predictions. We have no rationale for changing the direction of the hypothesis in this manuscript, since we aim to replicate the findings of our previous work. In case we cannot replicate the findings of enhanced pre-activations in individuals with tinnitus (H1), we will still pursue to draw conclusions regarding H3. We think that it will still be interesting to take a look at the influence of tinnitus distress even though we were not able to confirm H1. However, we changed the design table to make it more clear how H3 is especially related to H1. We also tried to make the section "Analysis plan" more precise.

Table 1. Design Table

Question	Hypothesis	Sampling Plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes
Do individuals with tinnitus show different regularity-dependent pre-activations of carrier-frequency-specific information compared to a control group without tinnitus?	H0: No. H1: Yes. Referring to our previous results, we expect higher regularity-dependent pre-activations in the tinnitus group.	See section *Sampling Plan*	**Mean** decoding accuracies in the pre-stimulus interval will be compared between groups, using a cluster-based permutation t-test.	We base our decision on the minimum requirement of an effect size of d=0.75 with a certainty of 95% and an alpha-level at 0.05.	H0: Finding no group differences would contradict our previous results (Partyka et al., 2019) and highlight discrepancies between study designs. Either differences in the stimuli or in the sample might be responsible for such results. In the latter case we would not be able to exclude influences of hearing loss on the results. H1: Similarly, to our previous results, we would cautiously interpret stronger regularity-dependent pre-activations as a sign of increased vulnerability to developing tinnitus. This is in line with reports about auditory hallucinations and links to strong priors.
Are tone-carrier frequencies processed normally in individuals with	H0: No. H1: Yes. As previous results	See section *Sampling Plan*	**Mean** decoding accuracies in the post-stimulus interval will be compared	We base our decision on the minimum requirement of an effect size of d=0.75 with a certainty of 95%	H0: Deviations in normal tone-carrier-frequency processing in individuals with tinnitus contradict our previous results. It is important to extract differences in the study designs and to filter out the variables that might influence results.

Question	Hypothesis	Sampling Plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes
tinnitus?	suggest, there are no differences in processing of different tone-carrier frequencies between individuals with tinnitus and without.		between groups using a cluster-based permutation t-test.	and an alpha-level at 0.05.	H1: We interpret normal tone-carrier-frequency processing in individuals with tinnitus as in-line with previous findings, indicating no abnormal tonotopic representations in individuals with tinnitus.
If individuals with tinnitus show different regularity-dependent pre-activations of carrier-frequency-	H0: No. H1: Yes. These effects are explained exclusively by tinnitus and not by	See section *Sampling Plan*	Mean decoding accuracies **in the pre-stimulus interval** will be correlated with a mean value of subjective	We will decide based on the significance of the correlation.	H0: Correlations between the effects and subjective tinnitus distress would suggest the importance of the current tinnitus state. We would suggest longitudinal studies to further investigate the influence of tinnitus characteristics on regularity-dependent pre-activations of carrier-frequency-specific information. H1: Similar to our previous results, we interpret

Question	Hypothesis	Sampling Plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes
specific information, are these effects not driven by any influence of subjective tinnitus distress?	confounds like tinnitus distress.		tinnitus distress.		independence of tinnitus distress and the effects as a sign of individual predispositions to tinnitus development and resulting regularity-dependent pre-activations of carrier-frequency-specific information, which are not correlated to the current characteristics of tinnitus but more likely temporally stable “trait-like” features.

- In addition, why did you choose to use the mean scores? In a possible further analysis, it would be interesting to inspect subscales of the TQ (i.e., intrusiveness of sound and/or auditory issues).
- There is a possibility that people experiencing louder tinnitus have increased TQ scores and therefore, according to your H1, enhanced anticipatory predictions. Are you planning or have you collected data on the loudness or the pitch of the tinnitus?

We recognized an imprecision in the manuscript, thanks to your comment. In our online study, we assessed tinnitus not by using the TQ but by implementing the short version Mini-TQ. The short version included solely 12 questions of the subscales emotional distress, cognitive distress and sleep disturbances. We changed the manuscript accordingly. The rationale of choosing the Mini-TQ instead of the TQ was to restrict the duration of the online study. We did not expect our results regarding tinnitus distress to be limited since the Mini-TQ was assessed to have no psychometric disadvantages compared to the TQ (Hiller & Goebel, 2004). However, we unfortunately do not have the possibility to inspect your suggested subscales. Therefore, we will solely compute an overall mean score since we do not expect different effects between the included subscales.

We also appreciate your comment regarding tinnitus loudness and we agree with you that these would be interesting aspects to analyze. In our previous work (Partyka et al., 2019), tinnitus loudness was assessed and did not influence the effect However, we are unfortunately not able to replicate these specific findings, since we did not collect data on tinnitus loudness and pitch. We recognize that this is a limitation of our study design and will highlight the topic in the discussion section of the final manuscript.

Page 11:

“The online study included demographic information as well as questionnaires covering tinnitus (**German short version of Tinnitus Questionnaire, Mini-TQ** (Goebel and Hiller, 1992)) and hearing characteristics (German version of the Speech, Spatial and Qualities of Hearing Scale, SSQ (Kiessling et al., 2012)), along with an online hearing test (Shoebox, Ottawa, Canada). **The Mini-TQ includes subscales targeting emotional distress, cognitive distress and sleep disturbances which we will use to draw conclusions about the impact of tinnitus distress.**“

Page 18:

“We will analyze H3 by extracting individual values **of the short version of the Tinnitus Questionnaire (Mini-TQ)** and calculating the mean subjective tinnitus distress for each individual of the tinnitus group.”

References:

Hiller, W., & Goebel, G. (2004). Rapid assessment of tinnitus-related psychological distress using the Mini-TQ. *Int J Audiol*, *43*(10), 600-604.

Participants:

l-247-250: The matching of hearing loss seems reasonable. However, it is not clear how you define hearing impairment. Which thresholds did you use?

We agree with you that this part is missing some precision. We categorized hearing loss as a hearing threshold above 30 dB in at least one frequency and added this in the manuscript to make it more clear.

Page 11:

“We performed standardized pure-tone audiometry for frequencies from .125 to 8kHz in all participants using an Interacoustic AS608 audiometer to characterize hearing status. **Hearing loss was defined by a hearing threshold above 30 dB in at least one frequency.“**

MEG data acquisition and preprocessing:

l-348-359: I am not an expert on machine learning models, but how do you justify to use the same data for training and testing data sets?

We appreciate your comment and apologize for the imprecise wording that implied testing and training on the same data sets in the random condition. We rephrased it to emphasize the importance of the cross-validation that we will include. Additionally, training will be performed on the post-stimulus interval of the random tones, while we will test on the pre-stimulus interval.

Page 16:

“When testing on the ordered condition, we will not perform any cross-validation, as our approach already consists of cross-decoding. **For testing on the random tones, we will perform a 5-fold cross-validation. It is further important to specify that we will train on the post-stimulus interval and test on the pre-stimulus interval of the random tones.** We will construct two time-generalization matrices: one for each condition.“

Stimuli and experimental procedure:

To improve a better understanding of the experimental design, please specify if the number of conditions was also balanced. You describe that per block 1500 stimuli are presented, while 500 stimuli belong to one condition. It would be possible that in block 1 you presented random (500 stimuli), ordered (500 stimuli), random (500 stimuli) and repeat the same pattern in block 2, which means that you would have a different number of data points per condition entering the analysis. Please specify.

Thank you for pointing out the imprecision in the manuscript. The number of conditions was also balanced. When block 1 consisted of random (500 stimuli), ordered (500 stimuli) and random (500 stimuli), block 2 started accordingly with an ordered sequence (500 stimuli), followed by random (500 stimuli) and ordered (500 stimuli). Therefore, we collected the same amount of data points for each condition (1500 stimuli per condition in total). We changed the paragraph to make it more clear.

Page 13:

"These groups of 500 stimuli with a particular entropy condition were presented in random order within each of the two blocks. **To balance the number of conditions, one of the two blocks started with a random condition (500 stimuli), followed by an ordered sequence (500 stimuli) and ended with a random condition (500 stimuli). For the other block, sounds started accordingly in an ordered condition, followed by random sounds and a second sequence of ordered sounds. Therefore, data collection comprised 1500 stimuli of each condition.** The experiment was written using the MATLAB-based (version 9.1 The MathWorks, Natick, Massachusetts, U.S.A) Psychophysics Toolbox (Brainard, 1997)."

Minor issues:

Abstract:

l-36-37: Consider reformulating 'varies from random to ordered', as there are 2 conditions and no steps in between them.

We agree with you and changed the phrasing to make it more precise that we included two distinguishable conditions.

Page 2:
"This registered report aims to close this gap. We will use an established passive-listening paradigm, in which the regularity (i.e. predictability) of pure-tone sequences **is either random or ordered**."

l-46-48: Do you refer to chronic tinnitus here? Please explain or clarify this further. It is always difficult to speak of ‘consequences’ or ‘causations’ as it is extremely hard to establish causation – it might be better to refrain from such terminology.

Thank you for your comment. We agree with you that we are not able to establish causation and we rephrased it.

Page 2:

“This would lay the foundation for any later works that need to disentangle whether dysregulated predictive processes **are a side product of tinnitus or rather pose a risk factor for developing tinnitus.”**

Introduction:

l-59-61: According to Jarach et al., 2022, the prevalence in older participants is up to 24%, I would be more precise here.

Thank you for your suggestion, we acknowledge the missing precision in reporting the prevalence. We changed it accordingly.

Page 3:

“Approximately 10-15% of the young to middle-aged adult population experience tinnitus as a common auditory phantom perception, with greater prevalence of 24% in older adults (Henry et al., 2020; Jarach et al., 2022).”

l-67: Please add references to the sentence ‘Hearing loss has been identified as a highly predictive trigger for tinnitus. ‘. On what literature do you base this statement?

This sentence was also pointed out by Reviewer 1 as well and we rephrased it. Additionally, we added literature to justify the statement.

Page 3:

“What neural mechanisms contribute to the generation of tinnitus remains unresolved. Hearing loss has been identified as **a main risk factor** for tinnitus **(Kim et al., 2015)**. Indeed, for 75-80% of people with tinnitus, objective audiometric testing indicates hearing loss (Wallhäusser-Franke et al., 2017).“

References:

Kim, H. J., Lee, H. J., An, S. Y., Sim, S., Park, B., Kim, S. W., ... & Choi, H. G. (2015). Analysis of the prevalence and associated risk factors of tinnitus in adults. *PloS one*, *10*(5), e0127578. doi: https://doi.org/10.1371/journal.pone.0127578.

l-105-106: How could the model explain tinnitus when opposed to hyperactivity in the auditory system, please specify.

Thank you for your feedback. We rephrased the statement to make it more clear that we refer back to the inconsistent findings regarding the "altered gain" framework.

Page 5:
"This conceptual model bridges several explanatory gaps: **for example, the inconsistent findings in humans regarding the "altered gain" view which states enhanced neural activity in the auditory pathway. The Bayesian inference framework could, therefore**, explain the experience of tinnitus in lieu of any increase in neural activity in the auditory system."

Statistical analysis:

l-371: Please add the exact ms for the pre-stimulus interval. Are you using -400 – 0 ms? The same holds for the post-stimulus interval for H2 (i.e., l382 – 396).

Thank you for pointing this out. We added the exact ms to the according intervals.

Page 17:

"For this, we will consider the pre-stimulus interval **(-400 to 0 ms)** to perform cluster-based permutation t-tests.
Then, considering the post-stimulus interval **(0 to 500 ms)** for statistical analysis, we will be able to draw conclusions about H2, regarding normal carrier-frequency processing in the tinnitus and control groups. "

Response to Reviewer 3 (Emilie Cardon):

This Stage 1 registered report details a proposed research protocol to investigate anticipatory auditory predictions in tinnitus patients compared to age-, gender- and hearing level-matched control subjects. The authors plan to analyze already available MEG data from 80 subjects in total, using both an experimental design and analytical pipeline they have already utilized in

previous studies. In this protocol, the research questions are well-define, address important questions in the field of tinnitus research, and correspond clearly to the proposed methodology. In addition, I would also like to commend the authors for the truly excellent overview provided in the introductory section.

We appreciate your favorable evaluation and constructive input.

L200 and following, *Sampling plan*: The sample size calculation yielded a minimum number of 80 participants. This calculation was based on an expected effect size of 0.75. I agree with the authors that a dataset including 40 tinnitus patients and 40 control subjects is larger than average in the field. However, it is not clear to me whether the expected effect size is solely a theoretical estimate or if it is based on their earlier findings (Partyka et al. 2019). Would the authors be able to provide the effect size obtained in their earlier work, so that the reader might more readily determine whether this effect size justified? Or would this not be feasible due to the differences in analytical methods used in both studies?

Thank you very much for your feedback. Our effect size was more of a theoretical estimate since we could not sufficiently base it on our previous findings due differences between the two studies - as you mentioned as well. Additionally, our main rationale was practical limitations in the participant recruitment and lab capacities. As you acknowledge as well, our dataset is larger than average in the field, therefore, we reason that we included a suitable sample size to address our research questions. In our manuscript, we also pointed out that with respect to clinical relevance, solely strong effects are worth pursuing because of the difficult circumstances in data collection. Therefore, we based our power analysis on a high effect size. To make it more clear, we added a phrase to the paragraph to highlight the theoretical basis of our estimate.

Page 10:

“Next, we target a clinical topic for which recruiting participants is more challenging since we are looking for specific characteristics in volunteers. Additionally, finding suitable controls for our strict matching procedure is time consuming as well since lab capacities are restricted. With respect to clinical relevance, solely strong effects are worth pursuing because of the difficult circumstances in data collection. **Therefore, our power analysis was based on a theoretical estimate of a medium to large effect size (*d*=0.75). Using G*Power (Faul et al. 2009), we calculated an estimated sample size for a one-sided t-test, expecting a true effect of *d*=0.75. We used a power of 0.95 and a one-sided α-error probability of 0.05, which resulted in a required total sample size of at least 80 participants.** This is also in line with our previous arguments warranting a sample of 40 individuals with tinnitus and 40 controls.

At the date of the Stage 1 submission, data collection of 80 participants was already completed. **With our analyses, we therefore aim for a smallest estimated effect size of interest of *d*=0.75 to be found with a power of 95 % and a conventional alpha at 5 %.**“

Moreover, the authors clearly state that the required sample size is *at least* 80. This means that data from each of the currently included 80 subjects will need to be utilized in order to answer the research questions. However, it might be the case that some of these data are of insufficient quality to be included in the final data analyses. I understand that the authors have not yet observed the data (and agree with their corresponding assessment of their registered report as Level 3). Nevertheless, is there any way to guarantee the usability of the entire dataset? Are there any quality control checks, perhaps already performed by independent researchers, that these data have passed before subjects were included in the dataset?

We agree with you that we can not guarantee the usability of the entire data since data has not been analyzed yet (i.e., also no preprocessing etc.). During the measurements, no technical issues or difficulties occurred, which was reported by the technical assistant at the lab. Therefore, we do not expect data to be of insufficient quality also as the paradigm is now routinely used in the lab (see also Schubert et al., 2023; Schubert et al., 2023, bioRxiv). However, we do acknowledge your concern and recalculated the power analysis with a smaller sample. In case we are solely able to include 35 participants per group in our analyses (which is still a reasonably high sample size in this field), the calculated effect size rises up to d=0.8. Since we expect strong effects in this clinical sample, we conclude that our analyses are still reasonable, even with a data loss of up to 10 participants.

Line 225 and following, *Participants*: How did the authors perform the matching procedure based on hearing level? Specifically, were subjects matched based on their 'hearing status' (i.e. the different categories explained in L247-250), or based on pure tone averages at certain frequencies? If subjects are matched based on 'hearing status', there might still be important differences in hearing thresholds between both groups (for example, thresholds for participants with 'high-frequency hearing loss' might still differ substantially). Would the authors expect such potential differences to influence the results? As excluding potential confounding effects of hearing loss is crucial, something which the authors also stress at different points throughout the protocol, I would strongly recommend at least the inclusion of audiometric data in the final report. This would allow to judge whether there are any systematic quantitative differences in hearing levels between both groups. If such differences would exist, I would urge the authors to examine whether they had any effect on the final results.

We appreciate your questions and will happily provide more details about the matching procedure. We matched the subjects not solely on their "hearing status" but aimed for a best fit over frequencies. However, we acknowledge that we can not aim for a perfect fit between individuals with tinnitus and controls. In order to quantify the matching procedure, we calculated

independent t-tests between the two groups for the right and left ear separately. The results were integrated in the according paragraph and did not show any significant differences between groups. Therefore, we do not expect an influence on the results. However, we plan to include audiometric data in the final report and will address the issue in the discussion section.

Additionally, thanks to your comment, we re-checked the classification of the hearing status and found some errors in the assignment of the tinnitus group. We corrected the paragraph.

Page 11-12:

“Four individuals with tinnitus did not show any audiometric peculiarity; four of the participants showed unilateral hearing impairments; 26 volunteers had high-frequency hearing loss; and six individuals were hearing impaired over most frequencies. The control group was recruited afterward in order to match the distribution of the tinnitus group by age, gender and hearing status. **Accordingly, we aimed to find the best possible match that our data allowed for between individuals with tinnitus and control subjects regarding the results of the audiometry. Using independent t-tests, we found no differences in hearing status between groups for the left (*t*=-1.32, *p*=.192) and right ear (*t*=-1.27, *p*=.212).** Control subjects were age-matched to each tinnitus participant by a +/-2-year criterion, choosing the closest match when more than one subject was suitable. All participants provided written informed consent before participating.“

Moreover, tinnitus patients often concurrently experience psychological complaints, such as elevated levels of anxiety and/or depression. For some studies examining neural activity in this patient population, these factors are also considered to have potential confounding effects. Do the authors expect these potential concurrent complaints to affect their results? If so, are the authors planning on taking any precautions to exclude potential confounding effects due to elevated psychological distress?

Thank you for your input, we agree with you that psychological distress is a very relevant topic and can not be neglected regarding its potential confounding effects. In our hypothesis H3 we target tinnitus distress as well as its relation to our expected effects (H1) and aim to highlight herein the importance of psychological effects on neural activity. We added an explanation of the tinnitus questionnaire we used to make the subscales and the inclusion of psychological aspects more clear. In Hiller and Goebel (2004) the Mini-TQ was assessed and the 12 questions of the subscales emotional distress, cognitive distress and sleep disturbances were included. In our opinion these subscales represent a sufficiently broad overview of psychological distress to analyze the mean results in terms of their impact on our effects.

Page 11:

“The online study included demographic information as well as questionnaires covering tinnitus (**German short version of Tinnitus Questionnaire, Mini-TQ** (Goebel and Hiller, 1992)) and hearing characteristics (German version of the Speech, Spatial and Qualities of Hearing Scale, SSQ (Kiessling et al., 2012)), along with an online hearing test (Shoebox, Ottawa, Canada). **The Mini-TQ includes subscales targeting emotional distress, cognitive distress and sleep disturbances which we will use to draw conclusions about the impact of tinnitus distress (Hiller & Goebel, 2004). “**

L256 and following, *Stimuli and experimental procedure*: Please add the sound intensity of the auditory stimuli that were provided to the participants.

Thank you for pointing this out. We added the procedure to determine sound intensity at the according paragraph.

Page 13:

“Each tone lasted 100 ms, tapered at both ends with 5 ms linearly ascending/descending periods, and we presented the sounds at a constant 3Hz stimulation rate. **Sound intensity was individually determined by presenting a short audio sequence to the participants and adjusting the loudness according to an individual pleasant volume.**“

L306 and following, *MEG data acquisition and preprocessing*: The authors do not mention the removal of bad trials or bad sensors (and subsequent interpolation) from the data. Are these steps not a planned element of the processing pipeline? If so, I would like the authors to point out why they choose not to remove bad sensors and/or trials in the protocol. If not, please add the necessary information about how this removal will be performed.

We agree with you that removal of bad trials and bad sensors is an important step in the processing pipeline. The removal of bad sensors is implemented in the Maxfilter algorithm that we will use as a first step on the data. After the preprocessing steps, we will run an ICA to detect ECG and EOG components and cut the data into epochs. On that basis, we will implement the decoding analysis. Therefore, we aim to stick to the procedure of our previous work (Partyka et al., 2019).

Page 15:

“We plan to use a signal-space separation algorithm (SSS (Taulu and Kajola, 2005)) implemented in the Maxfilter program (version 2.2.15) to reduce external noise from the MEG signal (mainly 16.6Hz, and 50Hz-plus harmonics) and to realign data of different measurement blocks to a common standard-head position (“-trans default” Maxfilter parameter), based on the head position measured at the beginning of each block (Cichy and Pantazis, 2017).

Additionally, the Maxfilter algorithm will detect bad channels, remove and interpolate the data."

The authors describe the removal of ICA components containing unwanted artefacts. I would suggest that they report the number of components removed for each experimental group in their final report, in order to ensure that the number of removed components does not substantially differ across groups.

We acknowledge your input and we will report the number of removed components in the final report.

Page 15:

"Next, we will inspect the ICA components visually to detect and remove unwanted artifacts, such as eye blinks and movements, heartbeats and 16 ⅔ Hz artifacts (the level of German/Austrian train power supply). **We will report the number of removed components for each group to highlight whether the number of components differed substantially across groups."**

Would it be feasible to blind the involved researchers to the group to which the data belong (tinnitus vs. control) during MEG preprocessing and analysis? Is this planned and, if so, could this be added to the protocol?

Thank you for your input. We agree with you that blinding the researchers is good practice to reduce biases. For the analyses, information regarding the group will have to be added to the subject files explicitly in order to correctly assign the files to the tinnitus or the control group. We will bring in an independent researcher to assign the files to the groups before the involved researchers observe them. Thereby we will be able to blind the researchers until analyses are finished, which data belongs to the tinnitus vs. the control group.

Page 9:

"Further analyses of the participant characteristics are necessary to link information regarding tinnitus to the participant codes and corresponding files. These characteristics are with a colleague and have not been accessed by the authors yet. **In order to further blind the researchers during the analyses, the subject files will be assigned to two groups (tinnitus vs. control) without the involved researchers knowing which group represents which condition. The information will not be passed to the involved researchers until the analyses are completed.**"