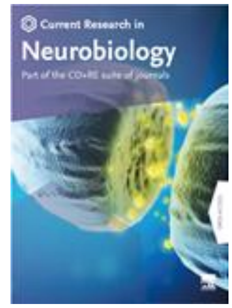
# Peer Review Overview

**Manuscript Title:** Distracting Linguistic Information Impairs Neural Tracking of Attended Speech





| | |
|---|---|
| Received | Dec 16, 2021 |
| 1st Decision | Mar 01, 2022 |
| 1st Revision Submitted | Apr 27, 2022 |
| Accepted | May 24, 2022 |


---

## 1st Decision letter

**Reference:** CRNEUR-D-21-00076
**Title:** Distracting Linguistic Information Impairs Neural Tracking of Attended Speech
**Journal:** Current Research in Neurobiology

Dear Dr Kosem,

Thank you for submitting your manuscript to Current Research in Neurobiology. The reviewers and I agreed that it can make an important contribution to our understanding of cocktail party listening.

The reviewers recommend reconsideration of your manuscript following revision. I invite you to resubmit your manuscript after addressing the comments below. Please resubmit your revised manuscript by Apr 15, 2022.

When revising your manuscript, please consider all issues mentioned in the reviewers' comments carefully: please outline in a cover letter every change made in response to their comments and provide suitable rebuttals for any comments not addressed. Please note that your revised submission may need to be re-reviewed.

Current Research in Neurobiology values your contribution and I look forward to receiving your revised manuscript.

*CRNEUR* aims to be a unique, community-led journal, as highlighted in the Editorial Introduction. As part of this vision, we will be regularly seeking input from the scientific community. We encourage you and your co-authors to take the survey as part of the editorial.

Kind regards,

Kerry Walker, DPhil
Associate Editor
Current Research in Neurobiology

## Comments from Editors and Reviewers:

This is an interesting and valuable manuscript, which makes an important contribution to the field of selective listening. The approach of using vocoded speech training to manipulate intelligibility is clever and insightful. Both reviewers and I were puzzled by the lateralization of your behavioural effects, which was present when the trained distractor speech was presented to the right ear, but not the left ear. Given this unexpected result, you should also compare their neural effects of vocoding training when distractor speech is presented to the left vs right ear. Is there a neural correlate of the behavioural lateralization? This would be strong evidence of the link between the measured behavioural and neural effects.

In addition to the reviewer's suggestions below, I would also request that you re-write your public and media statement to be more lay-audience friendly. It is currently understandable only to a specialized scientific audience. You should remove or define the jargon in the statement (e.g. vocoded speech), so that it is understandable to the general public. You might also re-examine your graphical abstract with the aim of removing some of the text, especially the text in the coloured boxes.

**Reviewer #1:**

This paper tackles the important question of how the presence of task-irrelevant speech, and its intelligibility (or not), impacts neural tracking and intelligibility of target-speech, in a cocktail-party type paradigm.

This study has two operational strengths for investigating this question: First, by using noise-vocoded speech as the task-irrelevant stimulus + unmasking it between runs, they are able to manipulate intelligibility but still maintain perfect acoustic control of the stimuli. Second, the use of an explicit repetition task for target speech provides a detailed behavioral metric for assessing 'distraction/interference', which is not always available in other studies.

The study is cleverly designed and well executed. The results nicely demonstrate the "competition" for both acoustic and linguistic processing resources in multi-talker situations. They also fit with increasing evidence that although attention can "modulate" sensory responses for target stimuli, non-target stimuli are still represented and sometime directly interfere with target-processing.

I have a few comments for the authors:

Major
1. The behavioral effect of training seems to have had a side-specific effect - with repetition of target speech impaired only if intelligible (post-training) speech was presented to the right, but not to the left ear). The authors interpret this in the context of the "right ear advantage" hypothesis. I think this point needs more fleshing out in the discussion, since it is not mirrored in the neural data. How should we interpret this discrepancy?
2. Related, the lack of a correlation between changes in brain-speech coherence and behavioral metrics is disappointing, since we would have like to attribute the behavioral "distraction" to changes in neural encoding. Can you expand on that in the discussion? The results should also be describe in full (currently only data from the left-ear condition are reported, and they are somewhat different than what is shown

in Figure S1).
3. It would be nice to see a depiction of how the effect of attention on brain-speech coherence (for target vs. distractor) was affected by the training, rather than focusing only on the response to each stimulus separately. That might bring our interesting lateralization effects as well.

Minor
1. The experiment was quite long (just for the tasks alone take < 90 minutes). Can you please add details on the total duration of the session? And also, how do you think this might have affected participants ability to maintain attention/perform the task?

2. The method for extracting the speech-envelopes is not mentioned. Are these broad-band envelopes? Also, could you give the readers a sense of the modulation-spectrum of the stimuli? This might help clarify/justify the delta/theta separation.

3. Running the risk of 'tooting our own horn' (and revealing my identity :-), I would recommend that the authors check out a recent MEG paper from our group, where we also found that intelligible distracting speech affects neural tracking of target speech (which, in that study, we found in more inferior-frontal regions; Har-shai Yahav, eLife 2021).

**Reviewer #2:**

The authors of this manuscript investigate how the linguistic information in a distractor speech signal impacts comprehension of a target speech, as well as how it impacts the neural tracking of both the attended and the distractor speech. To this end, the authors employ a conrast between two conditions in which the distractor speech is either more or less comprehensible. To avoid confounds from differences in the acoustics, they employ a paradigm in which the distractor speech (noise-vocoded speech) remains the same, but, after a training session, increases in comprehensibility. The authors observe that, after training, the target speech signal is less comprehensible, but only if the target speech signal is presented to the right, not the left, ear. The authors also employ neuroimaging to investigate the neural tracking of both the target and the distractor speech signal. They find that, after training, the tracking of the target speech decreases in the delta band but not in the theta band. The tracking of the distractor speech is, after training, reduced in the delta band but not in the theta band, both when the intelligibility of the target speech increases and in a control condition when it remains constant.

The experimental design, using noise-vocoded speech as a distractor and a training session to affect intelligibility, is interesting and should allow for valuable insights into the neural processing of both the target and the distractor speech. The data analysis is also sound. Unfortunately the results appear somewhat murky, so I am not yet sure whether there is an interesting story here.

1. The decrease in the comprehension of the target speaker after training appears only when the target speech is presented to the left ear (distractor to the right), and not in the opposite case. This finding appears odd, and makes the interpretation and connection to the neuroimaging results harder. I understand that the right-ear advantage might underlie this effect. But it also seems to me that this result might be due to a non-optimal choice of experimental conditions. The comprehension of the target speech is very high, above 90%. A change of comprehension due to a change in the intelligibility will therefore be hard to measure, since it will not cause a big deviation. If speech comprehension had

been around 50%, a change would presumable be much more noticeable. Why have the authors used such a high level of speech comprehension of the target speech, and what would happen if they operated at a lower level?

2. The neural tracking of the target speaker in the delta range is lower after training. However, to be able to relate this to behaviour, we would need to know if this is the case when the target speech is presented to the left ear (where speech comprehension decreases), for the target speech presented to the right ear (where speech comprehension remains constant), or an average over both cases. I haven't found a statement on this issue, so I assume that the authors have averaged over both cases. For the interpretation, I would like to see this analyzed separately for the two cases (target speech to the left or to the right ear). Does the neural tracking decrease only in the former and not in the latter case? If so, this would strengthen your argument that the change in the neural tracking is linked to behaviour. If not, what would this mean?

3. The delta tracking of the distractor speaker decreases after training, both for the 4-band NV speech where target comprehension can decrease as well as for the control condition (2-band NV). This result is worrying to me since it suggests that something odd might go on. More so since the tracking in the theta band remains stable. You explain this finding through potential "habituation", but I would like to see more analysis/potential explanations of this unexpected result, as well as a discussion of how it might link to the results for the neural tracking of the target speaker (shouldn't that then also be influenced by habituation)?

4. A hypothesis for which evidence has recently emerged is that the tracking of the attended and of the ignored speech stream compete for neural resources, see, e.g., Fiedler et al, Neuroimage 2019; Brodbeck et al, PLoS Biol. 2020; Keshavarzi et al, J. Neurosci 2021. How do your findings relate to this hypothesis?

Typos:

- Line 707: "grant" -> "grand"

- Line 738: full stop missing after "(right panel)"

## 1st Author Response Letter

### Response to comments from Editors and Reviewers:

This is an interesting and valuable manuscript, which makes an important contribution to the field of selective listening. The approach of using vocoded speech training to manipulate intelligibility is clever and insightful. Both reviewers and I were puzzled by the lateralization of your behavioural effects, which was present when the trained distractor speech was presented to the right ear, but not the left ear. Given this unexpected result, you should also compare their neural effects of vocoding training when distractor speech is presented to the left vs right ear. Is there a neural correlate of the behavioural lateralization? This would be strong evidence of the link between the measured behavioural and neural effects.

In addition to the reviewer's suggestions below, I would also request that you re-write your public and media statement to be more lay-audience friendly. It is currently understandable only to a specialized scientific audience. You should remove or define the jargon in the statement (e.g. vocoded speech), so that it is understandable to the general public. You might also re-examine your graphical abstract with the aim of removing some of the text, especially the text in the coloured boxes.

We thank the editor and the two reviewers for their positive feedback and their insightful comments on our manuscript. We have now discussed the lateralization of our effect both our at the behavioral and at the neural level in greater detail. Based on both reviewers' suggestions, we have performed additional analyses and have added new figures in the manuscript to describe the comparison of the neural effects when the distractor speech is presented to the left vs right ear. These new analyses suggest that the behavioral lateralization effects are linked to neural lateralization effects, and we discuss those links. We have also modified the public and media statement and the graphical abstract to make them more lay-audience friendly.

Please find the detailed response to the reviewers below.

**Comments from Reviewer 1**

This paper tackles the important question of how the presence of task-irrelevant speech, and its intelligibility (or not), impacts neural tracking and intelligibility of target-speech, in a cocktail-party type paradigm.

This study has two operational strengths for investigating this question: First, by using noise-vocoded speech as the task-irrelevant stimulus + unmasking it between runs, they are able to manipulate intelligibility but still maintain perfect acoustic control of the stimuli. Second, the use of an explicit repetition task for target speech provides a detailed behavioral metric for assessing 'distraction/interference', which is not always available in other studies.

The study is cleverly designed and well executed. The results nicely demonstrate the "competition" for both acoustic and linguistic processing resources in multi-talker situations. They also fit with increasing evidence that although attention can "modulate" sensory responses for target stimuli, non-target stimuli are still represented and sometime directly interfere with target-processing.

We thank the Reviewer 1 for their positive feedback on our manuscript, and we respond in detail to their comments below.

I have a few comments for the authors:

Major
1. The behavioral effect of training seems to have had a side-specific effect - with repetition of target speech impaired only if intelligible (post-training) speech was presented to the right, but not to the left ear). The authors interpret this in the context of the "right ear advantage" hypothesis. I think this point needs more fleshing out in the discussion, since it is not mirrored in the neural data. How should we interpret this discrepancy?

We have now discussed the lateralization of the behavioral and neural effects in more details throughout the manuscript, and have updated the results and discussion sections. Specifically, based on Reviewer 1's suggestions, we have run additional neural analyses that show lateralized effects and correspondence with behavioral data. We have now included these analyses in the manuscript (see details in the response to points 2 and 3).

2. Related, the lack of a correlation between changes in brain-speech coherence and behavioral metrics is disappointing, since we would have like to attribute the behavioral "distraction" to changes in neural encoding. Can you expand on that in the discussion? The results should also be describe in full (currently only data from the left-ear condition are reported, and they are somewhat different than what is shown in Figure S1).

We thank the Reviewer 1 for pointing out the discrepancy between the statistics in the manuscript and in Fig. S1. The reported statistics in the figure corresponded to the correlation between behavioral change and coherence change as of ($Coh_{post} – Coh_{pre}$). We now report the correlation between behavioral change with the normalized coherence change = ($Coh_{post}$ – $Coh_{pre}$) / ($Coh_{post}$ + $Coh_{pre}$) (which was originally intended). Statistics slightly changed, importantly we do now report a significant correlation between left target behavioral change and normalized coherence change to target in the left hemisphere (rho = 0.49, p =0.015). These statistics are uncorrected for multiple comparisons and should be taken with caution, but we do now report them in the manuscript l. 375-382:

*A significant correlation between the relative change of delta tracking of target speech in the left hemisphere and the behavioral reduction of reporting target speech was observed, but only when target speech was delivered to the left ear (Fig. S2, rho = 0.55, p = 0.005). This correlation was still significant after exclusion of one outlier participant (Fig. 4D, rho = 0.49, p = 0.015, uncorrected for multiple comparison testing). All other correlations were not significant (Fig. S2, left target, right hemisphere, rho = 0.22, p = 0.29; right target, left hemisphere, rho = 0.04, p = 0.83; right target, right hemisphere, rho = 0.04, p = 0.85).*

as well as in Figure 4D:

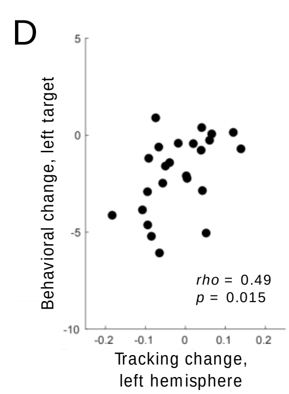


*__Figure 4__. (D) Significant inter-individual correlation (Spearman's rho) between the change in target-brain coherence in the delta range before and after training in the left hemisphere (Tracking change) and the change in target speech intelligibility before and after training when target speech was delivered into the left ear (Behavioral change). One outlier participant was excluded from the correlation (see full results in Figure S2).*

and in Figure S2:

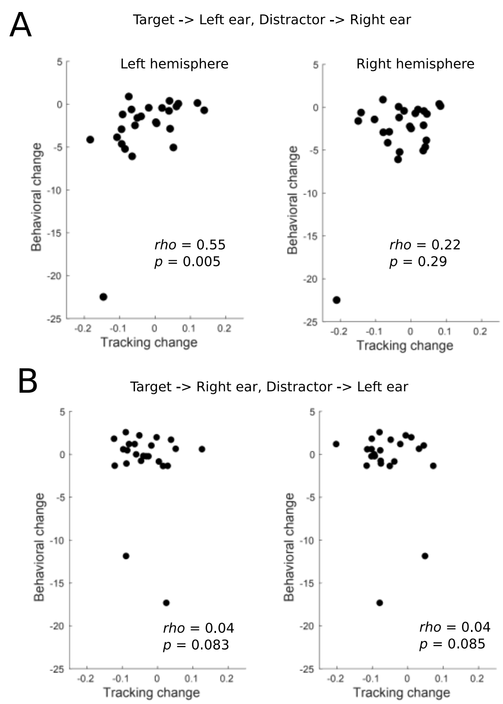



***Figure S2. Correlation between the strength of delta tracking of target speech and the target speech intelligibility**. Inter-individual correlation (Spearman's rho) between the change in target-brain coherence in the delta range before and after training (Coherence change) and the change in target speech intelligibility before and after training, when (A) target speech was delivered into left ear; (B) Target speech was delivered into right ear. Each dot corresponds to one participant.*

And we added a discussion point l.571:

*In line with this, we report a positive correlation between target speech-brain coherence in the left hemisphere and behavior in the left target condition. This correlation suggests that the decrease in perceived intelligibility is directly associated with a decrease in target speech neural tracking in the left hemisphere.*

3. It would be nice to see a depiction of how the effect of attention on brain-speech coherence (for target vs. distractor) was affected by the training, rather than focusing only on the response to each stimulus separately. That might bring our interesting lateralization effects as well.

We thank reviewer 1 for this suggestion that lead to an interesting new pattern of results. We now report figures and statistical analysis on the contrast between target speech brain coherence and distractor speech brain coherence. We show that training had an effect on this contrast, and that this effect was right lateralized. We describe the full analyses in the results section l. 430:

***Lateralization effects on the contrast between target and distracter speech-brain tracking***

*Finally, we investigated how the effect of distracter's intelligibility affected the contrast between target and distracter speech brain coherence ($Coh_{target}$ – $Coh_{distracter}$). In the delta range, we observed that this contrast is affected by the amount of noise-vocoding of the distracters (main effect of NV: $F(1, 24) = 9.17$, $p = 0.006$, $\eta2 = 0.28$). As shown previously, target speech-brain coherence is stronger than distractor's speech brain coherence, but this contrast is stronger in the presence of 2-band NV distractors as compared to when targets are in the presence of 4-band NV distractors. This suggests that more the distractor's signal is degraded, the more dissociable the neural processing of the target is from that of the distractor. As expected, we report a significant interaction between target position and hemisphere ($F(1, 24) = 21.68$, $p < 0.001$, $\eta2 = 0.47$), due do the dichotic nature of the design. We do also report a significant effect linked to the training (Fig. 8A, interaction among time, NV, and hemisphere : $F(1, 24) = 4.78$, $p = 0.04$, $\eta2 = 0.17$). Training significantly impacted the contrast between target and distracter speech in the right hemisphere: distracting and target speech-brain coherence is less distinguishable in this hemisphere after training for the 4 band-NV condition (Fig. 8B). This effect was not dependent on the side of target position (Fig. 8A, interaction among time, NV, hemisphere, and side $F(1, 24 < 1)$.*

*In the theta range, the contrast between target and distractor's speech brain coherence was only significantly impacted by the interaction between target side and hemisphere ($F(1, 24) = 33.06$, $p < 0.001$, $\eta2 = 0.58$) and by the interaction between target side, NV and hemisphere ($F(1, 24) = 15.25$, $p < 0.001$, $\eta2 = 0.39$): the contrast between target and distractor's speech was bigger contralateral to target's side, and the type of NV distractors influenced more strongly speech-brain coherence on the ipsilateral hemisphere to target presentation. No significant effect linked to training were observed.*

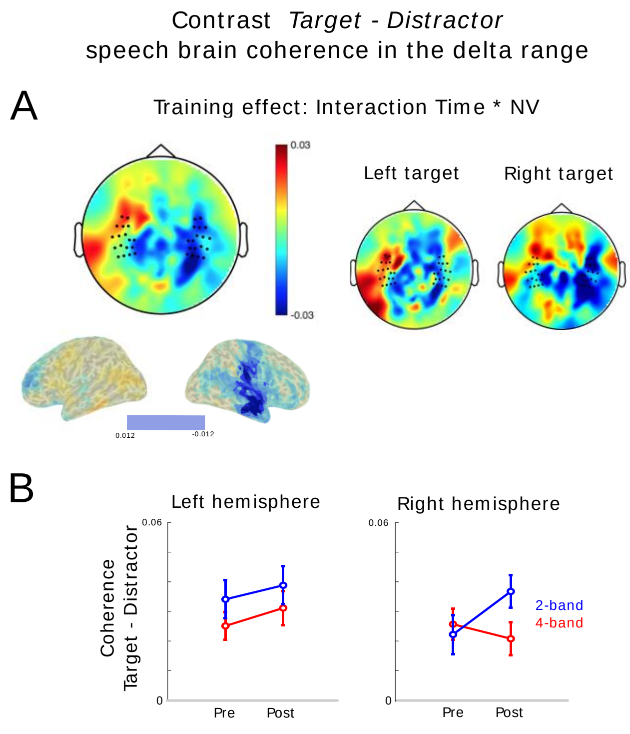



***Figure 8. Contrast between tracking of target speech and of distracting speech in the delta range*** *(A) Topographies and source reconstruction of interaction between factors time and NV $(Coh_{Post} - Coh_{Pre})_{4band}$ -$(Coh_{Post} - Coh_{Pre})_{2band}$. A significant interaction time*NV*hemisphere was observed. Training significantly impacted the contrast between target and distracter speech in the right hemisphere: distracting and target speech-brain coherence is less distinguishable in this hemisphere after training for the 4 band-NV condition. This effect was not dependent on the side of target position. (B) Delta tracking of the contract between target and distracting speech averaged across selected channels in each hemisphere, when in competition with distracting 4-band NV speech (red) or 2-band NV speech (blue). Error bars indicated standard error of the mean.*

We interpret these new findings with regards to the behavioral lateralization effects in the discussion section l. 575:

*Additionally, we report that the contrast in neural tracking of target and distractor speech is also lateralized and sensitive to the intelligibility of the distracting signal. The more the distracting signal becomes intelligible, the more reduced the contrast is between target and distractor speech-brain coherence in the right hemisphere. This means that processing of the target could interfere more strongly with that of the distracter in the right hemisphere. This interference could therefore be more impactful on behavior when the target is presented on the left ear because acoustical processing of the target would primarily take place in the right auditory cortex.*

Minor

1. The experiment was quite long (just for the tasks alone take < 90 minutes). Can you please add details on the total duration of the session? And also, how do you think this might have affected participants ability to maintain attention/perform the task?

The total duration of the experiment was 90 minutes (30 for each cocktail party task, and 30 minute training session including pauses), which is a common experimental duration. Details on the durations of the experiment are given l.153 and 165:

*Each dichotic listening task was 30 min long.*

*The training phase took 30 min including pauses. In total, the experiment was 90 minutes long.*

Participants may have had difficulty in maintaining attention throughout the whole experiment. This is why we designed we added the control 2-band condition to control for these potential attentional effects. This point is now made explicit in manuscript l. 166:

*Considering the duration of the experiment, we specifically incorporated the 2-band NV speech condition in the experimental design to control for potential changes in task performance and attentional maintenance throughout the experiment.*

2. The method for extracting the speech-envelopes is not mentioned. Are these broad-band envelopes? Also, could you give the readers a sense of the modulation-spectrum of the stimuli? This might help clarify/justify the delta/theta separation.

We used broad-band speech envelopes for speech-brain coherence analyses. Details on the methods are given l. 217:

*Broad-band speech envelopes were computed by band-pass filtering the acoustic waveforms (fourth-order Butterworth filter with [250 Hz – 4000 Hz] cut-off frequencies), and by computing the absolute value of the Hilbert transform of the filtered signal (Fig. 3A).*

We also show the average envelope power spectra in figure 3A :

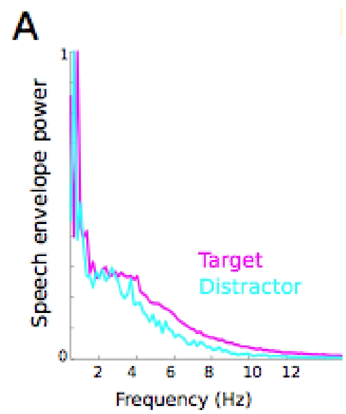



***Figure 3.*** *(A) Average power spectra of the speech envelopes of targets (magenta) and distractors (cyan).*

3. Running the risk of 'tooting our own horn' (and revealing my identity :-), I would recommend that the authors check out a recent MEG paper from our group, where we also found that intelligible distracting speech affects neural tracking of target speech (which, in that study, we found in more inferior-frontal regions; Har-shai Yahav, eLife 2021).

We thank Reviewer 1 for suggesting this very interesting and relevant study. We have now discussed these findings in the manuscript l. 505:

*A recent report further supports this hypothesis in cocktail party settings (Har-Shai Yahav & Zion-Golumbic, 2021): in this study, target speech was either in competition with structured speech (speech with phrasal structure) or with unstructured speech (words were presented in random order). Results show that competition with structured speech led to a decrease in neural tracking of target speech in the delta range.*

l.510:

*In line with previous reports, we show that the brain concurrently tracks the slow dynamics of both attended and ignored speech (Brodbeck et al., 2020; Ding and Simon, 2012a; Fiedler et al., 2019; Keshavarzi et al., 2021; Riecke et al., 2019; Zion Golumbic et al., 2013; Har-Shai Yahav & Zion-Golumbic, 2021), reinforcing the hypothesis that both signals compete for neural resources.*

l.529:

*In particular in studies investigating the cocktail party effect, the changes in neural tracking associated with attention are usually assumed to relate to the linguistic processing while this is not explicitly tested, as neural tracking is rarely compared to speech comprehension performance. When intelligibility scores are measured, they do not always correlate with the neural tracking measures of attended and/or distracting speech (Har-Shai Yahav & Zion-Golumbic, 2021). Therefore, in multitalker environments, speech-brain tracking could both reflect speech-specific processing and domain-general temporal* *attention mechanisms (Schroeder & Lakatos, 2009) that would influence speech parsing and hence comprehension (Kösem et al., 2018; 2020; van Bree et al., 2021).*

And l. 541:

*Har-Shai Yahav and colleagues (2021) observed that brain activity could track the phrasal structure of distracting speech. However, other recent findings have failed to find this effect, and argue that the neural tracking of larger linguistic structures (e.g. words that requires the sequential grouping of two syllables) requires attention (Ding et al., 2018). In both studies, unstructured and structured speech stimuli were made by manipulating the speech signals. Even if this manipulation makes it possible to control for many acoustical factors it could still cause small changes in the temporal structure of the envelope that could impact the neural tracking response (Kösem et al., 2016). In our findings, where we kept the acoustics of the distractors identical across conditions, the absence of an intelligibility effect in the neural tracking of the distractor speech could also suggest that linguistic processing of the distractor stops before the syllables or words in the distractor are grouped sequentially.*

**Comments from Reviewer 2**

The authors of this manuscript investigate how the linguistic information in a distractor speech signal impacts comprehension of a target speech, as well as how it impacts the neural tracking of both the attended and the distractor speech. To this end, the authors employ a conrast between two conditions in which the distractor speech is either more or less comprehensible. To avoid confounds from differences in the acoustics, they employ a paradigm in which the distractor speech (noise-vocoded speech) remains the same, but, after a training session, increases in comprehensibility. The authors observe that, after training, the target speech signal is less comprehensible, but only if the target speech signal is presented to the right, not the left, ear. The authors also employ neuroimaging to investigate the neural tracking of both the target and the distractor speech signal. They find that, after training, the tracking of the target speech decreases in the delta band but not in the theta band. The tracking of the distractor speech is, after training, reduced in the delta band but not in the theta band, both when the intelligibility of the target speech increases and in a control condition when it remains constant.

The experimental design, using noise-vocoded speech as a distractor and a training session to affect intelligibility, is interesting and should allow for valuable insights into the neural processing of both the target and the distractor speech. The data analysis is also sound. Unfortunately the results appear somewhat murky, so I am not yet sure whether there is an interesting story here.

We thank the reviewer for their positive feedback on the experimental design and data analysis, and for the valuable comments that helped us improve the manuscript. New analyses and discussion points have now been included, in particular to clarify the link between the neural and behavioral data. We believe that these clarifications help to emphasize how our findings are informative about how, through effects on neural tracking, distracting linguistic input influences the intelligibility of attended speech.

1. The decrease in the comprehension of the target speaker after training appears only when the target speech is presented to the left ear (distractor to the right), and not in the opposite case. This finding appears odd, and makes the interpretation and connection to the neuroimaging results harder. I understand that the right-ear advantage might underlie this effect. But it also seems to me that this result

might be due to a non-optimal choice of experimental conditions. The comprehension of the target speech is very high, above 90%. A change of comprehension due to a change in the intelligibility will therefore be hard to measure, since it will not cause a big deviation. If speech comprehension had been around 50%, a change would presumable be much more noticeable. Why have the authors used such a high level of speech comprehension of the target speech, and what would happen if they operated at a lower level?

As discussed in the original version of the manuscript, we do interpret the lateralization of the behavioral effects as a consequence of the right ear advantage, which is a well-documented phenomenon in the literature (Berlin et al., 1973; Brancucci et al., 2004; Della Penna et al., 2007; Ding and Simon, 2012b; Hiscock and Kinsbourne, 2011). We agree that the reported behavioral effect is small in our study. We are nevertheless confident in these results, as they are the direct replication of a previous study performed in our lab (Dai et al., 2017, JASA). The observed high performances in our design are directly linked to the fact that we used a dichotic listening task: when target and distracting speech are spatially separated performances are usually much higher than when both signals are presented in the two ears (Brungart, 2001, JASA). The choice of dichotic listening was motivated because we wanted to alleviate any effect of energetic masking, and really capture effects that could be linked to linguistic processing. This choice has both advantages and drawbacks. The motivation of this design is discussed l. 581:

*Yet, the laterality of the reported effects cannot fully be tested considering our dichotic listening design, which was originally chosen to focus on informational masking effects and alleviate the energetic masking of the distractors. Dichotic listening also implies complete peripheral separation of the target and distractor, which is rare in natural listening. Distractor and target speech could be presented in the two ears to overcome these two limitations, and we do expect that the current findings can be replicated in diotic listening environments.*

2. The neural tracking of the target speaker in the delta range is lower after training. However, to be able to relate this to behaviour, we would need to know if this is the case when the target speech is presented to the left ear (where speech comprehension decreases), for the target speech presented to the right ear (where speech comprehension remains constant), or an average over both cases. I haven't found a statement on this issue, so I assume that the authors have averaged over both cases. For the interpretation, I would like to see this analyzed separately for the two cases (target speech to the left or to the right ear). Does the neural tracking decrease only in the former and not in the latter case? If so, this would strengthen your argument that the change in the neural tracking is linked to behaviour. If not, what would this mean?

We did consider the ear of target presentation as a factor when performing the speech-brain coherence statistical analyses. This has been clarified in the results section l.343:

*Specifically, we expected the neural analysis of target speech (as reflected by speech-brain tracking) to be more impaired in the presence of an intelligible distractor. Hence, we predicted speech-brain coherence to target speech to become weaker after training, and this only for the 4-band NV distractor condition as the effect of training was limited to this type of distracting signal. We also investigated whether this effect was dependent of the side of target of presentation, and whether it was differentially observed in each hemisphere of interest. We therefore analyzed speech-brain coherence using a four-way repeated measure*

*ANOVA with the following factors: distractor's noise-vocoding type (NV: 4-band, 2-band), time (pre-training, post-training), side (left target, right target), and hemisphere (left, right).*

We did not observe a significant effect of the side of target presentation, this has been clarified l. 359:

*Yet, the effect of distractor's intelligibility was not significantly affected by the side of target speech presentation, nor by the hemisphere of interest as no other significant interactions were found (Fig. S1, interaction among hemisphere, NV and time: ($F(1, 24) = 2.72$, $p = 0.11$, $\eta 2 = 0.10$); interaction among side, NV and time: ($F(1, 24) = 0.32$, $p = 0.58$, $\eta 2 = 0.01$); interaction among hemisphere, side, NV and time: ($F(1, 24) = 3.29$, $p = 0.08$, $\eta 2 = 0.12$)). Unlike the behavioral results, where we only observed a significant effect for targets presented on the left ear, the decrease in delta tracking of target speech was observed regardless of the ear of presentation, in both hemispheres.*

and we added Figure S1 to illustrate the lack of effect:

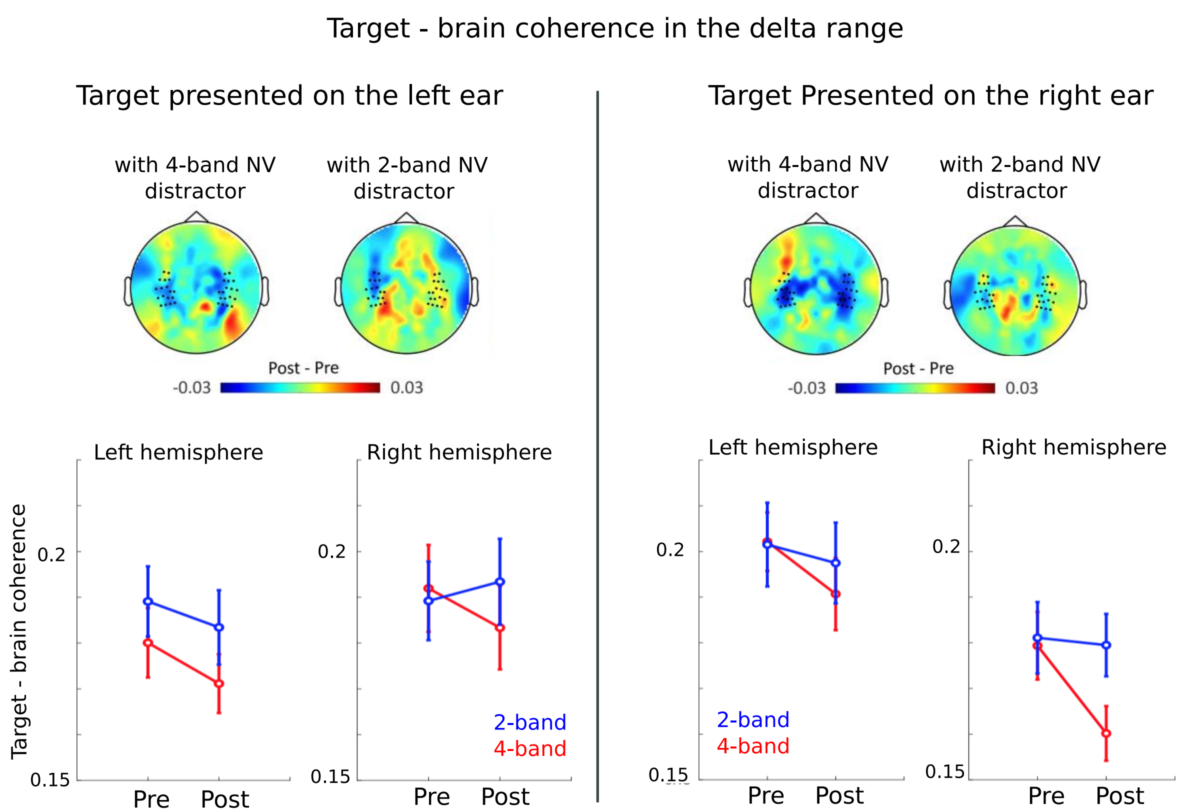



***Figure S1: Delta tracking of target speech reduced when the distracting NV speech gained intelligibility, irrespective of the ear of target speech presentation**. Left panel: Target presented on the left ear (and distracter on the right ear); Right panel: target presented on the right ear (and distracter on the left ear). Topographies represent delta tracking changes (Post-training minus Pre-training) in the presence of 4-band (trained) and 2-band (untrained) NV distracting speech. Bottom figures represent delta tracking of target speech averaged across selected channels in each hemisphere, when in competition with distracting 4-band NV speech (red) or 2-band NV speech (blue). Error bars indicated standard error of the mean.*

However, following Reviewer 1's suggestions, we performed new analyses and now present lateralization effects in neural data that could be linked to behavior (see response to Reviewer 1's points 2 and 3).

3. The delta tracking of the distractor speaker decreases after training, both for the 4-band NV speech where target comprehension can decrease as well as for the control condition (2-band NV). This result is worrying to me since it suggests that something odd might go on. More so since the tracking in the theta band remains stable. You explain this finding through potential "habituation", but I would like to see more analysis/potential explanations of this unexpected result, as well as a discussion of how it might link to the results for the neural tracking of the target speaker (shouldn't that then also be influenced by habituation)?

We do not think this result is odd, neural tracking of sensory rhythms is highly sensitive to attention/ habituation/ and or learning. Participants may have had difficulty in maintaining attention throughout the whole experiment, or may have learned to better ignore the distracting speech with time. The fact that it is restricted to delta frequencies is in line with the role of these dynamics in speech processing and that these dynamics are strongly present in the acoustic speech signal. We have clarified this point in the discussion section l. 405:

*The reduction of delta tracking of distracting speech can thus not be attributed to the training or the degree of intelligibility of the distractor, but may relate to habituation effects: over time, participants may have learned to pay less attention to the distracting stimuli. Therefore, the decrease in tracking after training could reflect overall temporal attention effects, unrelated to linguistic processing (Schroeder & Lakatos, 2009).*

4. A hypothesis for which evidence has recently emerged is that the tracking of the attended and of the ignored speech stream compete for neural resources, see, e.g., Fiedler et al, Neuroimage 2019; Brodbeck et al, PLoS Biol. 2020; Keshavarzi et al, J. Neurosci 2021. How do your findings relate to this hypothesis?

*In line with previous reports, we show that the brain concurrently tracks the slow dynamics of both attended and ignored speech (Brodbeck et al., 2020; Ding and Simon, 2012a; Fiedler et al., 2019; Keshavarzi et al., 2021; Zion Golumbic et al., 2013; Har-Shai Yahav & Zion-Golumbic, 2021), reinforcing the hypothesis that both signals compete for neural resources.*

Typos:

- Line 707: "grant" -> "grand"
- Line 738: full stop missing after "(right panel)"

They have been corrected.

# Accept Letter

Dear Dr Kosem,

Thank you for submitting your manuscript to Current Research in Neurobiology.

I am pleased to inform you that your manuscript has been accepted for publication.

My comments, and any reviewer comments, are below.

Your accepted manuscript will now be transferred to our production department. We will create a proof which you will be asked to check, and you will also be asked to complete a number of online forms required for publication. If we need additional information from you during the production process, we will contact you directly.

We appreciate and value your contribution to Current Research in Neurobiology. We regularly invite authors of recently published manuscript to participate in the peer review process. If you were not already part of the journal's reviewer pool, you have now been added to it. We look forward to your continued participation in our journal, and we hope you will consider us again for future submissions.

*CRNEUR* aims to be a unique, community-led journal, as highlighted in the Editorial Introduction. As part of this vision, we will be regularly seeking input from the scientific community and encourage you and your co-authors to take the survey.

Kind regards,

Kerry Walker, DPhil
Associate Editor
Current Research in Neurobiology

Editor and Reviewer comments:

The authors have done a thorough job responding to reviewer suggestions, including the addition of new analyses in their manuscript. These revisions have substaintially improved the quality of the manuscript overall, as well as the insights we gain from the results. In particular, they show a neural lateralization that mirrors the lateralization effect observed behaviourally. I am satisfied that the authors have addressed the important concerns raised by the reviewers, and I am delighted to accept this manuscript for publication as a research article in Current Research in Neurobiology in its current form. I offer my congratulations to the authors on this interesting research.

*-------- End of Review Comments --------*