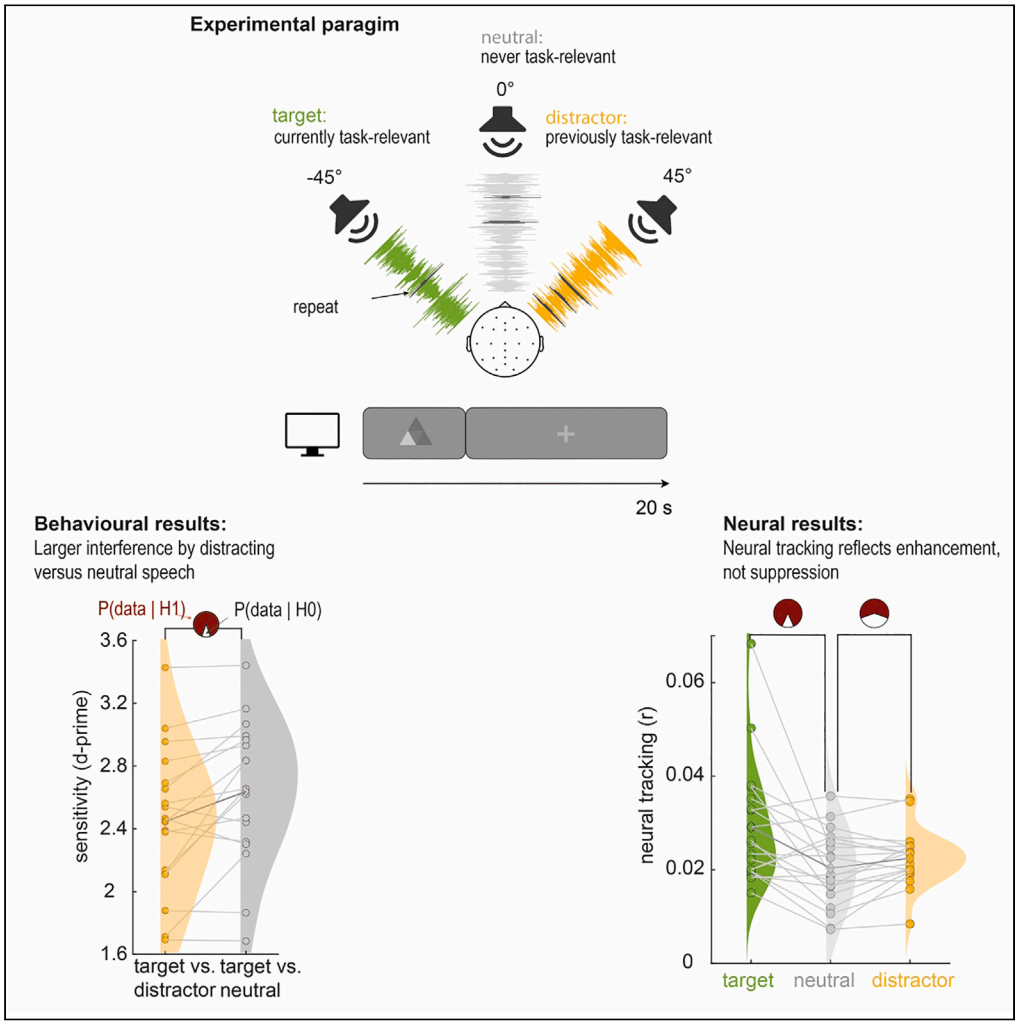
iScience

CellPress
OPEN ACCESS

Article

Target enhancement but not distractor suppression in auditory neural tracking during continuous speech



Martin Orf, Malte Wöstmann, Ronny Hannemann, Jonas Obleser

m.orf@uni-luebeck.de

Highlights

Larger interference by distracting versus neutral speech

Neural tracking reflects enhancement, not suppression

Neural tracking of the target stream explains performance

Orf et al., iScience 26, 106849
June 16, 2023
https://doi.org/10.1016/j.isci.2023.106849

iScience

CellPress
OPEN ACCESS

Article

Target enhancement but not distractor suppression in auditory neural tracking during continuous speech

Martin Orf,[1,2,4,*] Malte Wöstmann,[1,2] Ronny Hannemann,[3] and Jonas Obleser[1,2]

SUMMARY

Selective attention modulates the neural tracking of speech in auditory cortical regions. It is unclear whether this attentional modulation is dominated by enhanced target tracking, or suppression of distraction. To settle this long-standing debate, we employed an augmented electroencephalography (EEG) speech-tracking paradigm with target, distractor, and neutral streams. Concurrent target speech and distractor (i.e., sometimes relevant) speech were juxtaposed with a third, never task-relevant speech stream serving as neutral baseline. Listeners had to detect short target repeats and committed more false alarms originating from the distractor than from the neutral stream. Speech tracking revealed target enhancement but no distractor suppression below the neutral baseline. Speech tracking of the target (not distractor or neutral speech) explained single-trial accuracy in repeat detection. In sum, the enhanced neural representation of target speech is specific to processes of attentional gain for behaviorally relevant target speech rather than neural suppression of distraction.

INTRODUCTION

Selective attention refers to the neural filtering processes of prioritizing relevant objects over irrelevant distractions.[1] Typically, attentional selection is quantified by the difference in the behavioral or neural response to target versus distractor. However, such a difference can be driven by either target enhancement, distractor suppression, or a combination of the two. Here, we investigated how the mechanism of selective attention is represented in neural (electroencephalographic) activity, and we linked the trial-by-trial neural responses to behavioral responses associated with different sub-processes of attention.

In the visual domain, single-cell studies have shown that attention operates when multiple stimuli compete for access to neural representation. Distractors within a receptive field become suppressed, while attended stimuli are enhanced.[1] The mechanism of how selective attention is implemented at the level of neural networks is still in debate in attention research.[2,3] It has been argued that an often-missing, predefined baseline is needed to test whether the target exceeds the baseline (enhancement) and the distractor falls below the baseline.[4,5] In the visual modality Seidl et al.[6] had implemented such a "neutral" baseline by assigning a given class of stimuli as the never task-relevant, and therefore least distracting, category. They measured brain activity in fMRI (functional magnetic resonance imaging) in response to natural scene photographs that contained objects from a task-relevant (target) category, a task-irrelevant (distractor) category, and a never task-relevant (neutral) category. In addition, distractor suppression was linked to attentional capture. A distractor requires capturing attention initially, followed by suppression.[7–9]

Speech is one of the most salient and behaviorally relevant signals in human environments, but for a long time it was not possible to study the neural processing of time-varying natural stimuli like speech quasi-continuously. Neuroscientists thus studied attention to short, isolated events due to the need for temporally discrete event-related potentials (ERP).[10] Recently, research has begun to investigate the electrophysiology of attention to continuous speech.[11–13] Electrophysiological responses in cortical regions phase-lock to the temporal envelope of the speech signal.[14] This linear relationship is well-captured by the so-called temporal response function (TRF), which can be interpreted as a cortical impulse response, in close analogy to the conventional ERP.[15,16] The TRF can indicate a stereotypical, phase-locked brain response to various acoustic features. The most often used feature is the low-frequency temporal

[1]Department of Psychology, University of Lübeck, Lübeck, Germany

[2]Center of Brain, Behavior and Metabolism (CBBM), University of Lübeck, Lübeck, Germany

[3]WS Audiology, Erlangen, Germany and Lynge, Denmark

[4]Lead contact

*Correspondence: m.orf@uni-luebeck.de

https://doi.org/10.1016/j.isci.2023.106849

envelope, also referred to as neural speech tracking.[17] This neural speech tracking shows a robust and often-reproduced differentiation of attended versus ignored speech.[11,16,18–20] Thus, neural tracking is a feasible approach to quantify the neural processing of several speech streams at the same time to reveal the effect of attention.[11,21,22] In addition, Fiedler et al.[16] showed that late TRF components are associated with cortical tracking of ignored speech and are differently modulated for varying signal-to-noise ratios. These findings indicate that different components of the TRF are associated with different attentional processes. In sum, a hitherto underutilized advantage of this approach is its ability to delineate two potential sub-processes of attention: target enhancement vs. distractor suppression.[5]

What characterizes a distractor stream in such an experimental setup? First, the implementation of the distractor stream was based on the phenomenon of negative priming, which describes the finding that a distractor from the previous trial is harder to select on the next trial.[23–25] It is assumed that a stimulus and the response it elicits become integrated into so-called "event files" in memory.[26,27] Therefore, a specific stimulus automatically retrieves the response that was previously linked with this stimulus.[28] In this sense, the whole distractor stream in a given trial is distracting, since the same event that was previous task-relevant triggers a response, despite currently being task-irrelevant, and must be inhibited.

Second, it was shown that spatial statistical regularities influence selective attention on a longer timescale. A location that contained a distractor with a higher probability is suppressed relative to other locations. In this context, participants would learn about the location of the distractor stream and suppress it over time.[29]

In the auditory modality, Hambrook and Tata[30] investigated the mechanisms of distraction by increasing the number of distractor streams in the auditory scene. Their results suggest that distraction is not an active process but rather simply a loss of phase tracking of the target envelope. However, the attentional sub-processes target enhancement and distractor suppression have been suggested but have rarely been probed explicitly.[16,31,32] Here, we adopted the rationale of Seidl et al.[6] and implemented three auditory speech streams, a target (task-relevant) stream, distractor stream (previously task-relevant), and—critically—a neutral stream, which is never task-relevant. Larger target-vs-neutral tracking would indicate enhancement, while smaller distractor-vs-neutral tracking would indicate suppression). In the context of the auditory scence, the neutral stream can be conceived as a weaker distractor not as non-distractor. We operationalized the neutral stream as the never task-relevant stimulus. However, the neutral stream is not neutral in the strongest sense: Like the distractor stream, it was associated with the attentional background since it had to be ignored by the listener.[21] In other words, the neutral stream was more similar to the distractor stream than the target stream. Critically, it is conceivable that suppression is preceded by initial attentional capture of the distractor, indicated by larger distractor-vs-neutral tracking for early neural responses (see Figure 1B).

However, a severe disadvantage of continuous speech paradigms thus far has been their typical lack of rich behavioral data.[33] Typically, comprehension questions are asked intermittently or afterward regarding the content of the audio stream, which are insufficient to assess task-relevance of neural responses, especially during a complex continuous speech paradigm.

In the present study, we use electroencephalography (EEG) to investigate neural responses in human participants. We asked to what extent selective attention to speech is implemented in the human brain through target enhancement versus distractor suppression, and whether enhanced tracking of target speech or suppressed tracking of distraction would explain behavioral trial-by-trial indices of selective attention.

To this end, we designed a new experimental paradigm with two key advances over previous neural speech-tracking experiments (Figure 1A). First, a speech stimulus that was never relevant served as a neurally and behaviorally "neutral" baseline, against which the processing of concurrent target speech (relevant on a present trial) and distractor speech (relevant on other trials) can be contrasted.[5,6] Second, listeners had the task to continuously monitor and detect short repeats in the target stream[34] and to ignore short repeats in the distractor and neutral streams. This enabled us to contrast whether neural responses to target, neutral, or distractor speech would independently explain trial-by-trial variation in attentional performance.

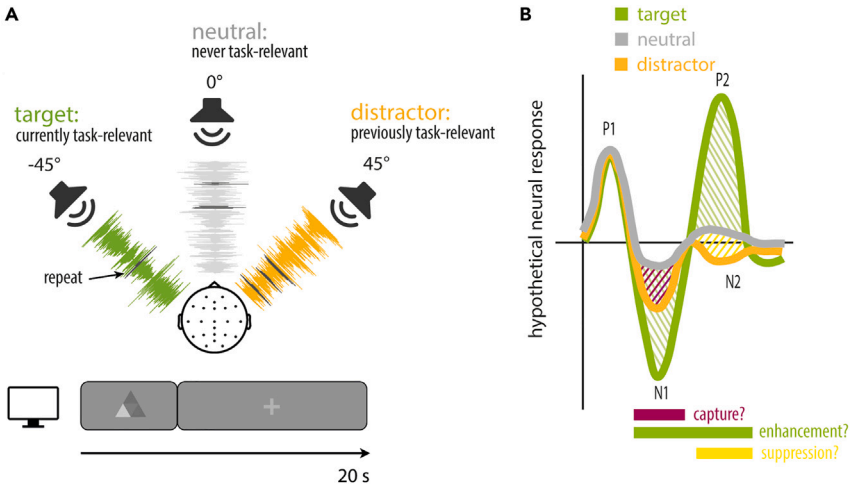


Figure 1. Experimental design and hypothetical results

(A) Simultaneously, we presented three different audio streams at different locations (−45°, 0°, 45°). Participants were instructed to attend to the cued audio stream for the duration of a trial (currently task-relevant target). In the next trial, another stream was cued, that became the target stream. The stream which was previously task-relevant became the distractor stream. During the entire experiment, the cue alternated between these two streams. The task-irrelevant (never cued) stream was defined as the neutral stream. We embedded short repeats in all three streams. Participants had to detect repeats in the target stream and had to ignore repeats in the neutral and distractor streams. Further, participants were instructed to process the content of the target audio stream.

(B) Hypothetical neural outcomes. While target enhancement (stronger target-vs.-neutral tracking; green) is expected for early and late TRF components, earlier components are expected to show neural capture by the distractor, that is, distraction (stronger distractor-vs.-neutral tracking; red) and later components are expected to show suppression (reduced distractor-vs.-neutral tracking; yellow).

RESULTS

We recorded the electroencephalogram (EEG) from 19 young, normal-hearing participants (7 male and 12 female, mean age 21.9 years, range 18–27 years). They were presented with three continuously narrated audio streams simultaneously (Figure 1A). On a trial-by-trial basis, they had to switch their attention between the same two audio streams. The to be attended audio stream was defined as the target stream, the audio stream attended in the trial before as the distractor stream, and the never task-relevant audio stream as the neutral stream. Participants had to detect any repetitions in the target stream as fast and accurately as possible and ignore the neutral and distractor streams.

Here, we analyzed behavioral data in terms of signal detection theory. We tested whether selective attention is driven by an enhancement of the target, a suppression of the distractor, or a combination of the two by investigating differential neural tracking of target versus neutral speech and distractor versus neutral speech by slow (1–8 Hz) cortical responses.

Larger repeat-evoked responses in the target stream

Overall, participants were well able to detect repetitions in the target stream (mean accuracy: 69.8% ± SEM 2.7%; response time: 735 ms ± SEM 14.1 ms), but performance was clearly not ceiling up with up to 86% (the highest score of single individual) correct responses (Figure 2A). In comparison, the false alarm rates for the neutral (false alarm rate: 2.1% ± SEM 3.4%; response time: 789 ms ± SEM 44.7 ms) and distractor streams (false alarm rate 2.9% ± SEM 3.4%; response time: 801 ms ± SEM 37.7 ms) were low. Jointly, the number of hits and false alarms indicated that participants were attending to the cued target audio streams. No significant differences in response times were observed ($t = 2.20$; $df = 30$; $p > 0.05$, for all comparisons).

We also estimated regression-based TRFs phase-locked to repeat onset (Figure 2B). TRFs to repeats in the target stream yield an auditory ERP-typical, biphasic response with an early positive deflection (0–170 ms) and a later negative deflection (170–550 ms). Topographies show β-weights with the highest magnitude for central channels. In contrast, the TRFs for the neutral and distractor streams did not show clear TRFs.

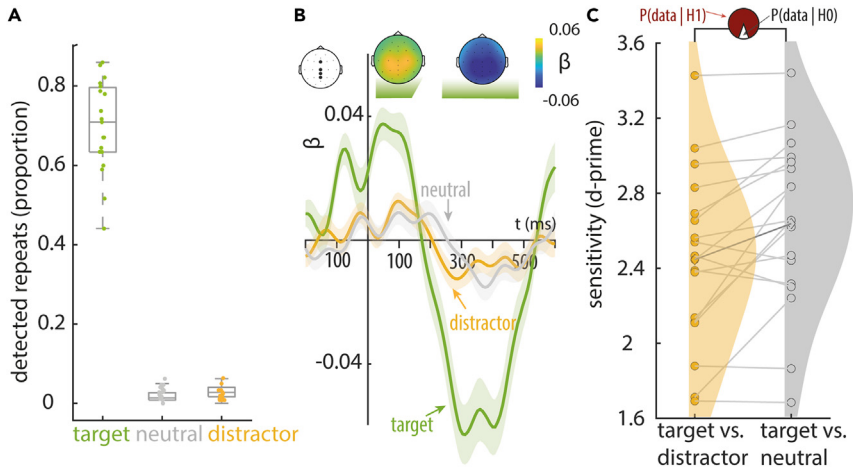


Figure 2. Behavioral results and TRFs to repeats

(A) Boxplots depict the proportion of detected repeats for the target (green), neutral (gray), and distractor stream (orange). Scatter dots depict individual subject data.

(B) TRF to repeats in the target (green), neutral (gray), and distractor stream (orange). TRFs β-weights are averaged across subjects (N = 19) and channels of interest (solid line). Shaded areas show the standard error for each time lag across subjects. Topographic maps depict β-weights for an early time window (0–100 ms) and for a later time window (300–400 ms) for the attended stream.

(C) The spaghetti plot shows the sensitivity index (d-prime) for target versus distractor streams and target versus neutral streams. Dots depict individual data, with connection lines indicating data from the same subject. Shaded areas illustrate the distribution of the data. Bayes factor visualization: probability pie charts show the ratio of the likelihood of H1(red) and H0 (white) for pairwise comparisons.

Regression based ERPs indicated a different brain response to target versus neutral repeats, but no different brain response to repeats in the neutral and distractor streams, which is in line with the observed behavior in Figure 2A. For further neural analysis, we treated the magnitude of these TRFs in all three streams as potential confounds and controlled for them statistically (for details see STAR Methods: TRFs. The fact that the participant's performance was off ceiling for detected repeats in the target stream and had a low false alarm rate in combination with no TRFs to the distractor and neutral streams indicate that the repeats did not pop out of the streams automatically. However, we label repeats "detected" in the distractor and neutral streams (false alarms) only if they are followed by a response. We cannot exclude the possibility that some repeats are detected but not followed by a response (response inhibition), even though the TRF for false alarms indicates no pop-out.

Larger interference by distracting versus neutral speech

To better understand the contrast in behavior between the neutral and distractor streams, we analyzed the behavioral data in terms of signal detection theory. Based on the hit rate and false alarm rates, two different d' could be calculated (Figure 2C). We calculated $d'_{target\text{-}distractor}$ to index the perceptual separation of target versus distractor stream, and $d'_{target\text{-}neutral}$ to index the perceptual separation of target versus neutral stream. Participants achieved a mean $d'_{target\text{–}distractor}$ of 2.46 $\pm$ 0.1 (M$\pm$SEM) and a somewhat higher mean $d'_{target\text{–}neutral}$ of 2.66 ($\pm$0.1).

A mixed model (supported by a Bayesian paired-samples t-test) with the regressor attention (target-distractor vs. target-neutral) confirmed this difference to be statistically significant (t = 3.01; df = 15; p = 0.009; BF10 = 8.1, supporting H1 over H0), indicating larger interference by the distractor than the neutral speech stream.

Morphology of neural responses to target, neutral, and distractor speech

We analyzed the neural tracking response to the target, neutral, and distractor streams by investigating the temporal, time-lagged relationship between the stimulus representation of each stream and the brain signal. This relationship is captured by an impulse response, the so-called TRF (TRF; see STAR Methods). Each component of the TRF is interpreted as a neural operation along the auditory pathway, analogous to

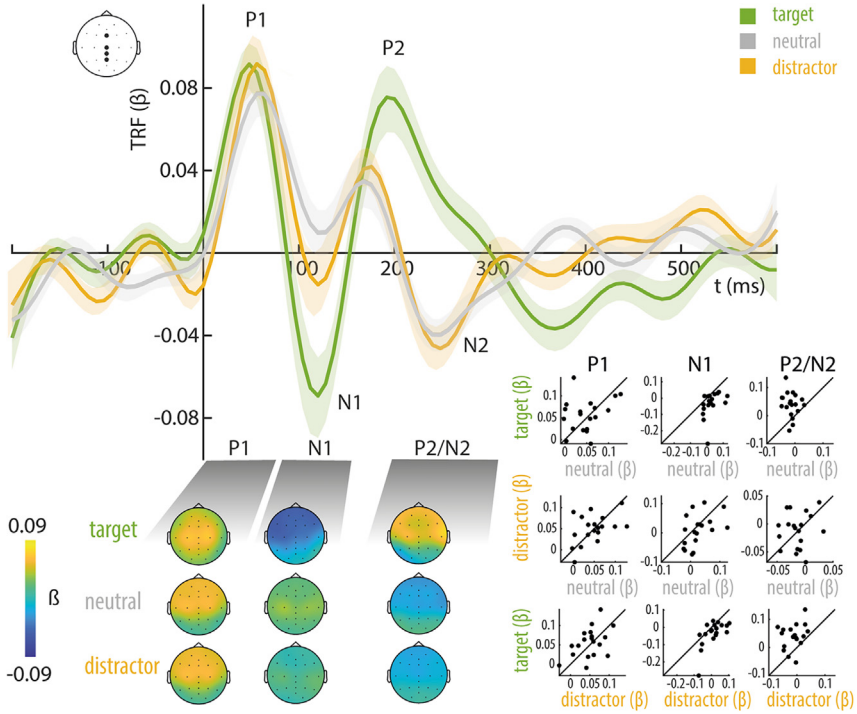


Figure 3. Temporal response functions (TRFs) of the target, neutral and distractor streams

TRF β-weights are averaged across subjects (N = 19) and channels of interest: Fz, Cz, CPz and Pz (solid lines). Shaded areas show the standard error for each time lag across subjects. Topographic maps depict β-weights for time windows of the P1, N1 and P2/N2 components for the three streams. 45°-plots show the single subject (N = 19) β-weights separately for neutral versus target, neutral versus distractor and distractor versus target for the P1, N1, and P2/N2 components.

the event-related potential.[35,36] Here, we describe differences between the TRF for target, neutral, and distractor streams, followed by a statistical analysis of the neural tracking response.

As expected, the morphology of the TRF for the target stream showed the succession of P1-N1-P2 response components, and the TRFs for the neutral and distractor streams showed the succession of the P1-N2 response components (Figure 3).

The early positive deflection P1 (0–80 ms) appeared in the TRF for the target, neutral, and distractor streams without any difference, indicating no attentional modulation. Topographies (located to fronto-central regions), latencies, and polarity of the P1 component were in line with previously observed TRFs and auditory-evoked potentials (AEPs) in the literature.

The later negative component N1 (80-150 ms) was prominent for the TRF of the target stream. The magnitude of N1 was increased (i.e., more negative) compared with the neutral and distractor streams.

The late positive deflection P2 (170–300 ms) was only present for the TRF of the target stream. In contrast, we found a negative deflection N2 in the TRF for the distractor and neutral stream in about the same time interval. This anti-polar relationship was also reported in previous studies.[11,16] However, there was no considerable difference in N2 for the TRF of the neutral stream versus the TRF of the distractor stream.

Neural tracking reflects target enhancement, not distractor suppression

Neural tracking reflects the strength of the representation of a speech stream in the EEG (see STAR Methods for details). For neural tracking, we asked whether selective attention is driven by an enhancement of the target, a suppression of the distractor, or a combination of the two. The most important finding of this study resulted from the differential neural tracking of the target and neutral streams (target enhancement; Figure 4B).

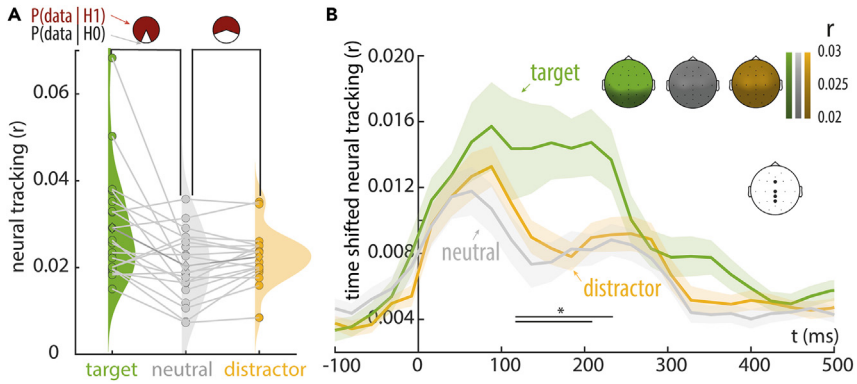


Figure 4. Neural tracking reveals target enhancement but no distractor suppression

(A) Neural tracking was computed based on the extracted TRFs and the envelopes of the attended (green), neutral (gray), and distractor streams (orange). Spaghetti plot shows single-subject data averaged across channels of interest. Connection lines between dots indicate the same subject. Bayes factor visualization: pie charts show probability of data given H1(red) and H0 (white) for pairwise comparisons. Shaded areas depict distributions of the data.

(B) Unfolding neural tracking across time lags (-100–500 ms). Solid lines show the averaged neural tracking (encoding accuracy; r) across subjects (N = 19) and channels of interest (topographic map). Shaded areas show the standard error for each time lag across subjects. Cluster permutation test revealed two significant clusters between target and neutral (136–232 ms) and between target and distractor (136–208 ms). Black bars indicate significant clusters. No significant clusters between distractor versus neutral were found. Topographic maps depict average neural tracking (r) for the three streams (0–500 ms).

Analysis of the neural tracking (0–500 ms) revealed a difference between the target and neutral stream indicated by a linear mixed model on the mean neural tracking (0–500 ms) and Bayesian t-test for target stream versus neutral stream ($t = 3.67$; $df = 32$; $p < 0.001$; $BF10 = 6.5$, supporting H1 over H0) and between the target and distractor stream ($t = 2.78$; $df = 32$; $p < 0.05$; $BF10 = 2$, weakly supporting H1 over H0). There was no significant difference in neural tracking of the distractor versus neutral stream and also the Bayes factor is not evidential ($t = 0.88$; $df = 32$; $p = 0.383$; $BF10 = 1.6$). If at all, the Bayes factor indicates the unexpected finding that the distractor stream was tracked slightly better than the neutral stream (see Figure 4). Topographies revealed the strongest neural tracking for central and frontal channels.

Lastly, we analyzed the temporally resolved dynamics of target enhancement and distractor suppression (Figure 4B). Unfolding neural tracking across time lags revealed differential tracking of the target and neutral streams. Target enhancement of encoding target versus neutral speech was signified by one cluster (136–232 ms; cluster $p = 0.0044$). We observed no significant clusters separating the neural response to neutral versus distractor stream.

Altogether, these findings indicate that neural tracking in a continuous speech tracking paradigm reflects a neural mechanism of target enhancement at the auditory cortical level, but no active distractor suppression.

Neural tracking of the target stream is associated with perceptual performance

To test the relationship between neural tracking and repeat detection performance, we modeled binary response behavior (hit vs. miss) as a linear function of neural tracking for the speech streams in the target, neutral and distractor streams using a generalized linear mixed model (GLMM; see STAR Methods for details). Further, we also controlled for the different numbers of repeats in the target stream by adding trial number as a continuous predictor into the model. We also included subject ID, the number of repeats (total experiment) and the condition-to-location assignment (neutral front, left, right) as random intercepts into the GLMM.

Neural tracking of continuous speech of the target stream displayed a positive linear relationship with participant's performance ($\beta \pm SEM = 0.077 \pm 0.029$; $z = 2.618$; $p = 0.009$). The higher the tracking accuracy of the target stream during a 20-s trial, the more likely participants detected repeats in that stream during that trial. We observed no such linear relationship in the neutral ($\beta \pm SEM = 0.017 \pm 0.029$; $z = 0.589$; $p = 0.556$) or in the distractor streams ($\beta \pm SEM = -0.023 \pm 0.029$; $z = -0.806$; $p = 0.420$; Figure 5A, left panel).

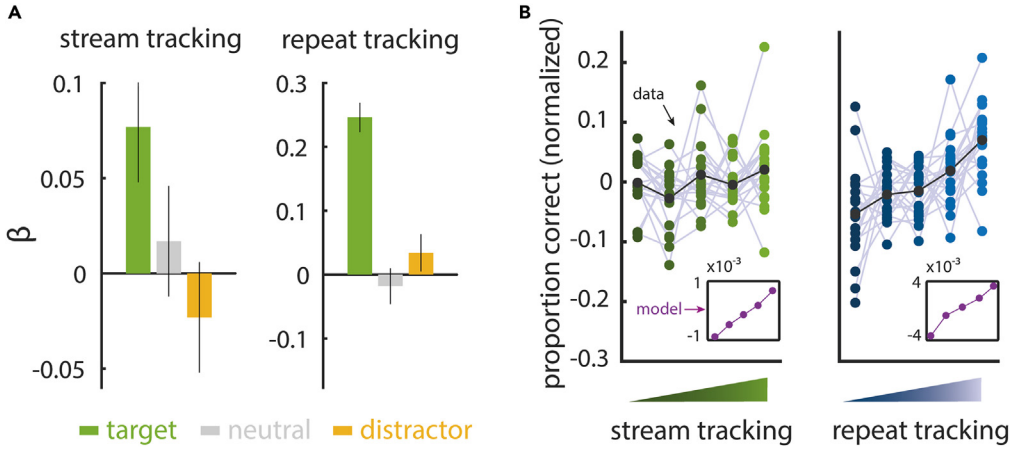


Figure 5. Brain–behavior relation

(A) Standardized estimates (fixed effects, with SE) for the prediction of binary response behavior (hit vs. miss) by speech and repeat tracking for the target (green), neutral (gray), and distractor stream (orange).

(B) Colored dots and gray lines show single subject proportion correct scores; black dots and black line show the average across (N = 19) subjects. For illustration, data were binned by stream/repeat tracking and normalization was done by subtracting the mean of single subject data across all bins from each corresponding subject data bin. Inset shows the model prediction for each bin.

Note especially that the estimates for the target stream and distractor stream pointed in opposite directions (Figure 5A, left panel). We thus, used a Wald statistic to test if the two estimates differed significantly from each other. The behavior-beneficial contribution of the neural tracking of the target stream was positive and differed significantly from (as per sign of the estimator, behavior-detrimental) neural tracking of the distractor stream (Z_{Wald} = 2.44, p = 0.015). As to be expected, the smaller differences of neutral versus target estimates (Z_{Wald} = −1.44, p = 0.147) and neutral versus distractor estimate (Z_{Wald} = 0.97, p = 0.332) proved not significant.

To control for potential confounding of the speech tracking in the target stream by the neural response to the to-be-attended repeats, we also included neural repeat tracking from all three streams in our model. Unsurprisingly, we observed a positive linear relationship between participant's performance and neural repeat tracking (β = 0.246; SE = 0.023; z = 8.235; p < 0.001) in the target stream. This shows that stronger neural responses to repeats in the target stream were associated with better behavioral detection of repeats. On the other hand, we observed no significant linear relationship between the tracking of a repeat in the neutral (β = −0.018; SE = 0.028; z = −0.644; p = 0.520) or in the distractor stream (β = 0.034; SE = 0.029; z = 1.197; p = 0.231; Figure 5A, right panel). For illustration only, we binned the data by the strength of stream and repeat tracking into five bins (Figure 5B, right panel).

Control analysis I: Listeners process the content of competing speech streams

The behavioral outcome from the comprehension questions was not of major interest to us, since the detection of repeats provides a much more reliable and finely resolved measure of behavioral performance. However, one concern we aimed to alleviate was that participants might have been only detecting repeats rather than listening to the speech content of the target stream at all; acoustic—phonologically processing the speech streams alone would probably be enough to identify the repeats. To further explore the degree to which listeners processed the speech streams semantically, 15 multiple-choice comprehension questions addressing all three streams were provided at the end of the study.

We used double iterative bootstrapping to estimate the 95% CI for the difference between the percentage of correctly answered questions and the previously determined empirical chance level of 40% (N = 9 different participants only answering the questions without exposure to the full audio books; see STAR Methods). By design, we were not able to differentiate between percentages of correctly answered questions in the target and distractor streams, as these switched their roles on a trial-by-trial basis. For instance, some questions required processing on a timescale that exceeded the trial length of 20s, which meant that some parts of the respective audiobook content belonged to the target and others to the distractor.

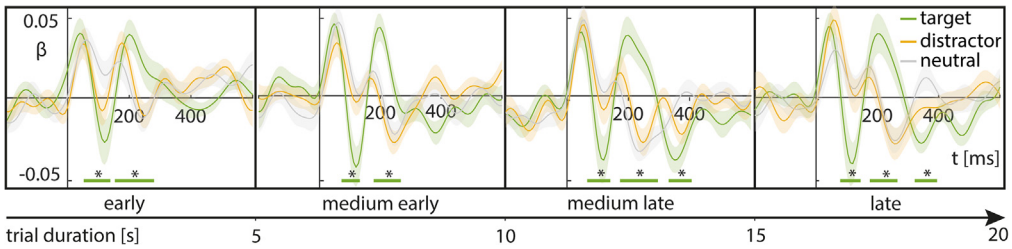


Figure 6. TRFs across trial duration

TRF β-weights are estimated in four separate 5 s time windows across the trial duration (20s), representing early to late attentional processing during the trial. TRF β-weights are averaged across subjects (N = 19) and channels of interest: Fz, Cz, CPz, and Pz (solid lines). Shaded areas show the standard error for each time lag across subjects. Cluster permutation test show significant clusters between target and neutral speech tracking in each time window (green bars). No significant clusters for distractor versus neutral speech tracking are observed.

Hence, we combined the correctly answered questions from the target and distractor streams (50 ± 2%, mean ± SEM, range: 30–67%).

This average response accuracy was significantly better than the empirical chance level (CI: 4.6–14.2% above chance). The percentage of correctly answered questions of the neutral audio stream was closer to chance (48 ± 3%, mean ± SEM, range: 27–80%), but there remained a significant if slim difference against the empirical chance level (CI: 0.9–14.6% above chance). The percentages of correctly answered questions did not differ systematically for the target/distractor stream versus the neutral stream (CI: –3 – 6.3%).

Control analysis II: Condition-to-location assignment does not confound interference by distracting speech and sub-processes of attention

In a further control analysis, we considered the possibility that the spatial condition-to-location assignment could have an indirect effect on our behavioral and neural measures. Between subjects, we varied the position of the neutral sound stream (neutral: front/left/right). The different positions of the neutral stream lead to a different assignment of the target and distractor streams. The spatial separation between the target and distractor streams was 90° when neutral was presented at 0° and 45° when neutral was presented at 45° or −45°. To control for the different spatial condition-to-location assignments, we included the factor condition-to-location assignment as a covariate in our behavioral and neural analysis.

In our behavioral analysis, we observed a significant main effect of the factor condition-to-location assignment ($F = 4.47$; $df = 15$; $p = 0.03$). This effect is mostly driven by a significant difference between the condition-to-location assignment: neutral front versus neutral right ($t = 2.96$; $df = 15$; $p = 0.01$). In other words, participants correctly detected more repeats when the neutral stream was presented in the front than the neutral stream presented on the left or right. There was no significant difference between neutral front versus neutral left ($t = 1.29$; $df = 15$; $p = 0.22$) and neutral right versus neutral left ($t = -1.71$; $df = 15$; $p = 0.11$). Importantly; however, the difference in sensitivity was independent of the spatial position of the neutral stream: There was no significant interaction between the factors attention and condition-to-location assignment ($F = 1.44$; $df = 15$; $p = 0.268$).

In our neural analysis, the main effect for the factor condition-to-location assignment was not significant ($F = 0.328$; $df = 16$; $p = 0.725$). Importantly, the differences in neural tracking were independent of the spatial position of the neutral streams. There was no significant interaction between the factors attention and condition-to-location assignment ($F = 0.88$; $df = 32$; $p = 0.482$). In sum, between-subject differences in the spatial condition-to-location assignment did not confound our results.

Control analysis III: Unfolding of neural filters (TRFs) across trial duration

To account for the possibility that attentional processes such as enhancement, capture, and suppression unfold on different time scales over the trial duration and might cancel each other out, we divided the 20-s trial into 4, non-overlapping windows of 5 s each and estimated TRFs separately for each window (Figure 6). Cluster permutation tests revealed that target enhancement is sustained across trial duration. Importantly, we found no significant clusters for the distractor-vs.-neutral contrast (i.e., no evidence for

capture or suppression). Also, a temporally more finely resolved analysis revealed no evidence for distractor capture or suppression. This analysis further supports our finding that target enhancement (i.e., attentional gain) is the dominant mechanism that modulates the neural phase-locked response to competing speech in a cocktail party scenario.

DISCUSSION

The present study aimed to test whether human auditory cortex enhances targets or suppresses distractors when implementing selective attention to continuous speech. To do so, we here have proposed a new, three-stream continuous-speech design with an embedded psychophysical task. The most important results can be summarized as follows:

First, the paradigm is feasible to delineate different sub-processes of auditory attention, separating a task-relevant target speech stream better from potentially neutral speech than from distracting speech. This finding proved robust under analyses controlling for stream location relative to the listener.

Second, the neural results suggest that attention is implemented through enhancement of the target stream. This lack of neural differentiation of tracking a distracting vs. tracking a neutral stream speaks against mechanisms of "active" or below-baseline neural suppression of distractors at the level of human auditory cortex as measured with EEG.

Third, in line with an enhancing neural attention mechanism, the momentary neural tracking of the target but not the neural tracking of other, competing streams can predict the momentary likelihood that a listener detects events in this target stream.

Neural tracking of speech implements enhancement, not suppression

As in previous studies,[11,12,16,19,36–38] we found the strongest neural tracking for the target stream, which was mainly due to enhanced N1 and P2 components of the cortical response. Notably, this improved tracking could be due to increased sensory gain, but it could also be due to more precise temporal fidelity of the target stream, or both.[39] Critically extending these previous findings by implementing a neutral, task-irrelevant "baseline stream" in a three-talker paradigm, we were able to assign these previous findings to two sub-processes of selective attention: target enhancement and distractor suppression. We found a significant difference in neural tracking between target and neutral streams but no significant difference between distractor and neutral streams.

We found that participants erroneously detected more repeats in distractor versus neutral speech, which indicates attentional capture on the behavioral level. Despite this signature of capture in behavior, we found neither suppression nor capture in the neural speech tracking response. In the visual modality, it was shown that capture and suppression go together. A distractor can capture attention, followed by suppression thereafter.[9] We have addressed this issue by analyzing different time windows along the trial. However, we found no evidence for distractor capture or suppression, analyzing early and late time windows separately. But that does not mean that suppression is not implemented on the cortical level in general. For instance, modulation of alpha oscillatory power is a potential neural mechanism that might implement distractor suppression in a scenario with competing auditory streams.[40]

Neural tracking of ignored speech is modulated by signal-to-noise ratio (SNR), hearing loss, and perceptual demand. Fiedler et al.[16] showed that SNR manipulations of ignored speech led to differential modulation of ignored speech and the resulting neural tracking. Also, hearing loss differentially affected neural tracking of attended versus ignored speech.[31] Recently, it was found that neural tracking of distracting speech in noisy auditory scenes depends on perceptual demand.[41] Here following a rationale established before in visual neuroscience,[6] we manipulated the attentional fate of ignored speech by varying listener's need to minimize or eliminate interference generated by the (previously task-relevant) distractors.

There is plenty of experimental evidence suggesting that selective attention is mainly enhancing the neural SNR, thus effectively clearing or sharpening target representations in the visual and auditory domain.[1,20,22,42–46] In line with these findings, we show that the prioritization of the neural representation of the target auditory input is mainly implemented by an enhancement of the target. In this respect, our results are also notably in line with a recent visual EEG study on attentional suppression by Gundlach

et al.[4] Also, another recent study investigated whether exogenous attention led to facilitation of attended information, suppressed unattended information, or both.[47] Both studies found that attention rather operates on target enhancement than distractor suppression.

Generally, our study adds to the unsettled debate in attention research over neural implementations of suppression. Even before the present study, evidence in the literature for distractor suppression has been mixed, with some studies speaking to[1,6,40,48] and others speaking against distractor suppression.[4,47,49]

Classical theories of attention permit some form of distractor suppression,[50,51] and there might well be distinct types of distractor suppression as endpoints to a continuum. Also, from a neurocognitive vantage point, distractor suppression does not need to be one single process and could be rather implemented via multiple neural mechanisms.

Firstly, suppression could be driven by the current intention of the observer extracting statistical regularities of certain features, such as location of a distractor over time, enabling the brain to learn to produce suppression.[5,29] In the long term (duration of the experiment), participants could learn based on statistical regularities the location (same location of distractor stream), and the voice of the talker (same voice). Secondly, in the short term (every trial), participants are cued (current intention) to attend to one stream and to suppress the distractor (negative priming). In principle, our paradigm might initialize both of these types of distractor suppression. While it is debatable whether the effect of our negative priming manipulation persists over the whole trial duration (probably decreasing over time), learning and using statistical regularities of the distractor over time should persist in the long term of the experiment. However, we found no significantly suppressed neural tracking of the distractor vs. neutral stream, which suggests that the neural speech tracking response does not implement distractor suppression. Contrary to our hypothesis, results hinted rather at a potentially stronger tracking of the distractor compared to the neutral stream although this was not a statistically robust observation in the present data. For future studies, it is nevertheless important to consider such an attentional capture of the distractor stream.[52] In addition, participants could also have left some residual attention to the distractor stream in terms of divided attention between the currently relevant target stream and the previously relevant distractor stream, which led to the potentially stronger tracking of the distractor compared to the neutral stream.[53] However, given the high hit rate for the target and the comparably low false alarm rate for the distractor stream, it appears rather unlikely that participants used divided attention as a strategy at least over the entire trial duration.

Secondly, distractor suppression can be generally divided into proactive (processing before the distractor appears) and reactive suppression (processing after the distractor has captured attention).[40,54] The amplitude of neural alpha oscillations (~10 Hz) related to top-down selective attention processes can be modulated by target- and distractor-processing. Wöstmann et al.[40] found that alpha power during the anticipation of competing tone sequences implements distractor suppression independent of target enhancement. In a behavioral study, it was shown that the intelligibility of the target is improved when the masker is a familiar voice.[55] Their findings suggest that the brain uses a prior model of the characteristics of the distractor to actively suppress the distractor. In sum, the aforemetioned results speak to a proactive implementation of distractor suppression. But neural tracking is characterized by the time-lagged neural responses which phase-lock to the stimulus. Due to this characteristic, neural tacking is rather suited to investigate reactive suppression than proactive suppression. With respect to these distinguishable sub-processes of distractor suppression, our results indicate that at least reactive suppression is absent for auditory cortex responses in a multi-talker situation.

Auditory attention exploits statistical regularities to separate distracting versus neutral speech

When considering how distracting versus neutral, task-irrelevant speech might be encoded neurally, a previous auditory study using also three streams had suggested that higher-order auditory areas provide an object-based representation for the foreground, but the background remains unsegregated.[21] At first glance, our results are broadly in line with this conclusion, but note that Puvvada and Simon had not applied any differential task manipulation to the two background speech streams, which we aimed to achieve here. The here proposed experimental paradigm aimed to strike important compromises in studying the listener's neurocognitive ability to separate target, distractor, and neutral speech.

In contrast to trial-based designs, continuous speech paradigms often lack rich behavioral data. Usually, comprehension questions regarding the content of the audio streams are asked to differentiate between attended and ignored audio streams.[16,56] Asking comprehension questions have some drawbacks. Comprehension questions usually refer to a comparable long-time range. This limits the number of questions and thus the number of behavioral data that can be extracted from the experiment. Further, in our paradigm participants had to switch their attention every 20s between two audio streams, which did not allow us to strictly assign the question to attended or ignored parts of the audio streams. Hence, it was insufficient solely to ask comprehension questions to investigate the listener's cognitive ability to separate target, distractor, and neutral speech on the behavioral level. More fine-grained behavioral data were needed ideally without losing much of the ecological validity of natural speech.

We used short repeats in the audio streams to obtain rich behavioral data. In trial-based designs, participants are asked much more frequently to respond, which also ensures a steady engagement into the listening task. Baldauf et al.[34] also embedded short repeats in auditory objects, arguing that such a detection task requires the processing of the acoustic stream at the level of auditory objects. Such a repeat detection task might thus be particularly suited to study object-based mechanisms of selective attention. Adopting this approach here, we found that participants detected much more repeats in the target (hits) than the neutral and ignored stream (false alarms).

Recall that, in our paradigm, participants had to switch attention between the same two streams while they had to ignore the never-task relevant neutral stream. Importantly, we found a significantly larger behavioral interference by distractor speech than by neutral speech, but what is the underlying mechanism? Our results suggest that the neural fate of a stream on the previous trial has the potency to make it more distracting and captures attention on the text trial. This corresponds with the concept of negative priming. Negative priming refers to the effect that the reaction to a stimulus that was previously ignored is more error-prone and slower.[25] Classical negative priming designs consist of two main components: prime (trial N) and probe (trial N+1). The prime presents a certain stimulus (or stimulus feature) as a distractor, which becomes the target in the probe trial. Negative priming has been studied in vision in a detailed manner.[57,58]

Although there are fewer studies that investigated negative priming in auditory selective attention, they reported similar results.[26] Nowadays most researchers agree that auditory negative priming (similar in vision) is explained by inhibition and retrieval theories.[26] Longer response times and higher error rates are typically observed relative to a no priming condition.[59–61] Notably, we did not present the same segments of the audio streams on two consecutive trials. Participants had to attend and ignore different segments of the audio streams in each trial, due to the ongoing structure of continuous speech. We assume that it was rather the spatial location or/and the voice that was associated with negative priming and leaked into the present trial, than the identity of the auditory stimulus. On the one hand, if a listener attended to a specific feature of an auditory object, not only this specific feature is enhanced, but all features related to the selected object.[62] On the other hand, one could argue that this also holds for features concerning negative priming and object suppression.

A more recent study varied randomly the location of the target and distractor and the speaker.[63] They demonstrated negative priming in auditory selective attention switching with the spoken material. In sum, our new paradigm has proven feasible to utilize the negative priming phenomenon to unravel listeners' separation of distractor speech versus neutral speech.

Neural tracking of target but not distractor explains performance

Continuous speech paradigms often lack rich behavioral data. But only if we unravel the precise relationship between brain and behavior can we reach a veridical understanding of cognitive processes, such as selective attention.[64] We embedded short repeats into the speech streams which served as a trial-by-trial measure for behavior. In addition, this also enabled us to predict behavior from neural responses on a single-trial level. We found that neural tracking of the target stream only predicted trial-by-trial variation in repeat detection. Our results not only provide support to the functional relevance of neural speech tracking,[65] but significantly expand this by providing an explanation for the underlying sub-processes of auditory selective attention, that is, enhancement of the target and not suppression of distractors predicts performance. In addition, this finding supports the feasibility of our new continuous speech paradigm since, we found a significant relation between the neural tracking of continuous speech and the repeat

detection behavior. Further, the finding supports our previous findings since only target enhancement predicts behavior. Indicating that the prominent process of selective attention is target enhancement rather than distractor suppression.

Limitations of the study

There are limitations regarding the operationalization of the neutral and distractor streams. First, the attentional manipulation by their respective task-relevance[6] of the distractor stream might not lead to an interference strong enough that distractor suppression was useful. Thus, it is possible that negative priming in combination with the spatial and/or spectral separation of the audio streams was insufficient to activate the need of distractor suppression in our study. Future studies could address this by varying for instance the separation between the audio streams.[41] The task may become more difficult with smaller spatial separation, which potentially activates distractor suppression.

Second, terminology can sometimes lead to conflicting theoretical inferences, as discussed recently.[66] In the context of theory-driven versus methodologically based terms, the term "neutral" initially falls under the former category, as it makes some assumptions about participants' cognitive or internal processing. At first glance, one might expect a neutral stimulus to be one that does not elicit a natural response. For example, in studies on emotion regulation, participants may be shown neutral images, such as pictures of objects or landscapes, to serve as a baseline for comparison with emotionally arousing stimuli.[67] In studies on decision-making, the term "neutral" is used by reinforcement learning theorists to describe responses from the environment that neither increase nor decrease the probability of a behavior being repeated.[68] However, in our case, we use the term "neutral" in close analogy to Seidl et al.,[6] who measured brain activity in response to photographs containing objects from a task-relevant (target) category, a task-irrelevant (distractor) category, and a never task-relevant (neutral) category. Here, the term "neutral" can be considered as a placeholder for the task-irrelevant category, and it belongs to the group of methodologically based terms. However, the term "neutral" plays a special role here, as it refers not only to the never-task-relevant stream but is also used as a control or baseline to separate target enhancement and distractor suppression. In the context of auditory scenes, the neutral stream can be also conceived as a weaker distractor rather than a non-distractor. Thus, the neutral stream is not neutral in the strongest sense since, like the distractor stream, it is associated with the attentional background, as it must be ignored by the listener.[21] Therefore, the neutral stream is more similar to the distractor stream than the target stream. In sum, there are multiple definitions of the term neutral. We used the term neutral to describe the task-irrelevant condition and as a baseline to measure target enhancement and distractor suppression. Since even a task-relevant speaker in the cocktail party is not neutral in the strongest sense, future studies are needed to investigate this in more detail. For instance, one could find more neutral, less distractor-like sound objects, such as broadband noise, but this would of course come at the price of losing some specificity in the condition comparisons being performed.

In addition, our sample size (N = 19) could have been too small to detect small distractor suppression effects. Note; however, that any such distractor-suppression effect size would need to be put in perspective to the considerable effect sizes of target enhancement we observed. So, the relative conclusion about target enhancement vs. distractor suppression would remain. Thus, the conclusion stands that target enhancement is the behaviorally and neurally more prominent sub-process of selective attention in a continuous speech paradigm.

Conclusion

In attention research, previous paradigms have rarely aimed at conclusively separating mechanisms of distractor suppression from mechanisms of target enhancement. Using a new, psychophysically augmented continuous-speech paradigm with three speech streams, our results demonstrate that the neural tracking of continuous speech reflects target enhancement, not distractor suppression. These findings call for a refinement of current models about enhanced neural responses to speech and should account for specific sub-processes of selective attention, that is, the enhancement of targets rather than the suppression of distraction.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

ACKNOWLEDGMENTS

MO is supported by a Widex-Sivantos Audiology grant (JO). MW is supported by a Deutsche Forschungsgemeinschaft DFG grant (WO 2371/1-1).

AUTHOR CONTRIBUTIONS

Conceptualization, M.O., M.W., R.H., and J.O.; methodology, M.O., M.W., and J.O.; investigation, M.O., M.W., and J.O.; writing—original draft M.O., M.W., and J.O.; writing—review & editing, M.O., M.W., R.H., and J.O.; funding acquisition, J.O. and R.H.; resources, J.O. and R.H.; supervision, M.W., R.H., and J.O.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 7, 2022
Revised: February 13, 2023
Accepted: May 5, 2023
Published: May 12, 2023

REFERENCES

1. Moore, T., and Zirnsak, M. (2017). Neural mechanisms of selective visual attention. Annu. Rev. Psychol. *68*, 47–72.
2. Schneider, D., Herbst, S.K., Klatt, L.-I., and Wöstmann, M. (2022). Target enhancement or distractor suppression? Functionally distinct alpha oscillations form the basis of attention. Eur. J. Neurosci. *55*, 3256–3265. https://doi.org/10.1111/ejn.15309.
3. van Moorselaar, D., and Slagter, H.A. (2020). Inhibition in selective attention. Ann. N. Y. Acad. Sci. *1464*, 204–221. https://doi.org/10.1111/nyas.14304.
4. Gundlach, C., Forschack, N., and Müller, M.M. (2021). Suppression of unattended features is independent of task relevance. Cereb. Cortex *32*, 2437–2446. https://doi.org/10.1093/cercor/bhab351.
5. Wöstmann, M., Störmer, V.S., Obleser, J., Addleman, D.A., Andersen, S.K., Gaspelin, N., Geng, J.J., Luck, S.J., Noonan, M.P., Slagter, H.A., et al. (2022). Ten simple rules to study distractor suppression. Prog. Neurobiol. *213*, 102269. https://doi.org/10.1016/j.pneurobio.2022.102269..
6. Seidl, K.N., Peelen, M.V., and Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. J. Neurosci. *32*, 11812–11819. https://doi.org/10.1523/JNEUROSCI.1693-12.2012.
7. Alexopoulos, T., Muller, D., Ric, F., and Marendaz, C. (2012). I, me, mine: automatic attentional capture by self-related stimuli. Eur. J. Soc. Psychol. *42*, 770–779. https://doi.org/10.1002/ejsp.1882.
8. Dalton, P., and Lavie, N. (2004). Auditory attentional capture: effects of singleton distractor sounds. J. Exp. Psychol. Hum. Percept. Perform. *30*, 180–193. https://doi.org/10.1037/0096-1523.30.1.180.
9. Gaspelin, N., and Luck, S.J. (2018). The role of inhibition in avoiding distraction by salient stimuli. Trends Cogn. Sci. *22*, 79–92. https://doi.org/10.1016/j.tics.2017.11.001.
10. Handy, T.C. (2005). Event-related Potentials: A Methods Handbook (MIT Press).
11. Ding, N., and Simon, J.Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. USA *109*, 11854–11859. https://doi.org/10.1073/pnas.1205381109.
12. Lalor, E.C., and Foxe, J.J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. Eur. J. Neurosci. *31*, 189–193. https://doi.org/10.1111/j.1460-9568.2009.07055.x.
13. Wöstmann, M., Fiedler, L., and Obleser, J. (2017). Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. Lang. Cogn. Neurosci. *32*, 855–869. https://doi.org/10.1080/23273798.2016.1262051.

14. Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron *54*, 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004.

15. Crosse, M.J., Di Liberto, G.M., Bednar, A., and Lalor, E.C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. *10*, 604.

16. Fiedler, L., Wöstmann, M., Herbst, S.K., and Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. Neuroimage *186*, 33–42. https://doi.org/10.1016/j.neuroimage.2018.10.057.

17. Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. *23*, 913–926. https://doi.org/10.1016/j.tics.2019.08.004.

18. Horton, C., D'Zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. J. Neurophysiol. *109*, 3082–3093. https://doi.org/10.1152/jn.01026.2012.

19. Kerlin, J.R., Shahin, A.J., and Miller, L.M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party. J. Neurosci. *30*, 620–628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010.

20. Mesgarani, N., and Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature *485*, 233–236. https://doi.org/10.1038/nature11020.

21. Puvvada, K.C., and Simon, J.Z. (2017). Cortical representations of speech in a multitalker auditory scene. J. Neurosci. *37*, 9189–9196. https://doi.org/10.1523/JNEUROSCI.0938-17.2017.

22. Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party. Neuron *77*, 980–991. https://doi.org/10.1016/j.neuron.2012.12.037.

23. Kristjánsson, A., and Driver, J. (2008). Priming in visual search: separating the effects of target repetition, distractor repetition and role-reversal. Vis. Res. *48*, 1217–1232. https://doi.org/10.1016/j.visres.2008.02.007.

24. Shiffrin, R.M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychol. Rev. *84*, 127–190. https://doi.org/10.1037/0033-295X.84.2.127.

25. Tipper, S.P. (1985). The negative priming effect: inhibitory priming by ignored objects. Q. J. Exp. Psychol. *37*, 571–590. https://doi.org/10.1080/14640748508400920.

26. Frings, C., Schneider, K.K., and Moeller, B. (2014). Auditory distractor processing in sequential selection tasks. Psychol. Res. *78*, 411–422. https://doi.org/10.1007/s00426-013-0527-3.

27. Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., et al. (2020). Binding and retrieval in action control (BRAC). Trends Cogn. Sci. *24*, 375–387. https://doi.org/10.1016/j.tics.2020.02.004.

28. Hommel, B. (1998). Event files: evidence for automatic integration of stimulus-response episodes. Vis. Cogn. *5*, 183–216. https://doi.org/10.1080/713756773.

29. Wang, B., and Theeuwes, J. (2018). Statistical regularities modulate attentional capture. J. Exp. Psychol. Hum. Percept. Perform. *44*, 13–17. https://doi.org/10.1037/xhp0000472.

30. Hambrook, D.A., and Tata, M.S. (2019). The effects of distractor set-size on neural tracking of attended speech. Brain Lang. *190*, 1–9. https://doi.org/10.1016/j.bandl.2018.12.005.

31. Petersen, E.B., Wöstmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. J. Neurophysiol. *117*, 18–27. https://doi.org/10.1152/jn.00527.2016.

32. Vanthornhout, J., Decruy, L., and Francart, T. (2019). Effect of task and attention on neural tracking of speech. Front. Neurosci. *13*, 977.

33. Hamilton, L.S., and Huth, A.G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. Lang. Cogn. Neurosci. *35*, 573–582. https://doi.org/10.1080/23273798.2018.1499946.

34. Marinato, G., and Baldauf, D. (2019). Object-based attention in complex, naturalistic auditory streams. Sci. Rep. *9*, 2854. https://doi.org/10.1038/s41598-019-39166-6.

35. Davis, M.H., and Johnsrude, I.S. (2003). Hierarchical processing in spoken language comprehension. J. Neurosci. *23*, 3423–3431. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003.

36. Di Liberto, G.M., O'Sullivan, J.A., and Lalor, E.C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr. Biol. *25*, 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030.

37. Har-shai Yahav, P., and Zion Golumbic, E. (2021). Linguistic processing of task-irrelevant speech at a cocktail party. Elife *10*, e65096. https://doi.org/10.7554/eLife.65096.

38. Kraus, F., Tune, S., Ruhe, A., Obleser, J., and Wöstmann, M. (2021). Unilateral acoustic degradation delays attentional separation of competing speech. Trends Hear. *25*. 23312165211013242. https://doi.org/10.1177/23312165211013242.

39. Ponjavic-Conte, K.D., Hambrook, D.A., Pavlovic, S., and Tata, M.S. (2013). Dynamics of distraction: competition among auditory streams modulates gain and disrupts inter-trial phase coherence in the human electroencephalogram. PLoS One *8*, e53953. https://doi.org/10.1371/journal.pone.0053953.

40. Wöstmann, M., Alavash, M., and Obleser, J. (2019). Alpha oscillations in the human brain implement distractor suppression independent of target selection. J. Neurosci. *39*, 9797–9805. https://doi.org/10.1523/JNEUROSCI.1954-19.2019.

41. Hausfeld, L., Shiell, M., Formisano, E., and Riecke, L. (2021). Cortical processing of distracting speech in noisy auditory scenes depends on perceptual demand. Neuroimage *228*, 117670. https://doi.org/10.1016/j.neuroimage.2020.117670.

42. Fritz, J.B., Elhilali, M., David, S.V., and Shamma, S.A. (2007). Auditory attention—focusing the searchlight on sound. Curr. Opin. Neurobiol. *17*, 437–455. https://doi.org/10.1016/j.conb.2007.07.011.

43. Gazzaley, A., Cooney, J.W., McEvoy, K., Knight, R.T., and D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. J. Cogn. Neurosci. *17*, 507–517. https://doi.org/10.1162/0898929053279522.

44. Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., and Ungerleider, L.G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron *22*, 751–761. https://doi.org/10.1016/S0896-6273(00)80734-5.

45. McAdams, C.J., and Maunsell, J.H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J. Neurosci. *19*, 431–441. https://doi.org/10.1523/JNEUROSCI.19-01-00431.1999.

46. Peelen, M.V., Fei-Fei, L., and Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature *460*, 94–97. https://doi.org/10.1038/nature08103.

47. Keefe, J.M., Pokta, E., and Störmer, V.S. (2021). Cross-modal orienting of exogenous attention results in visual-cortical facilitation, not suppression. Sci. Rep. *11*, 10237. https://doi.org/10.1038/s41598-021-89654-x.

48. Schwartz, Z.P., and David, S.V. (2018). Focal suppression of distractor sounds by selective attention in auditory cortex. Cereb. Cortex *28*, 323–339. https://doi.org/10.1093/cercor/bhx288.

49. Noonan, M.P., Crittenden, B.M., Jensen, O., and Stokes, M.G. (2018). Selective inhibition of distracting input. Behav. Brain Res. *355*, 36–47. https://doi.org/10.1016/j.bbr.2017.10.010.

50. Broadbent, D.E. (2013). Perception and Communication (Elsevier).

51. Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. J. Verb. Learn. Verb. Behav. *3*, 449–459. https://doi.org/10.1016/S0022-5371(64)80015-3.

52. Gaspelin, N., and Luck, S.J. (2019). Inhibition as a potential resolution to the attentional capture debate. Curr. Opin. Psychol. *29*, 12–18. https://doi.org/10.1016/j.copsyc.2018.10.013.

53. Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. Cogn. Psychol. *14*, 247–279. https://doi.org/10.1016/0010-0285(82)90010-X.

54. Chelazzi, L., Marini, F., Pascucci, D., and Turatto, M. (2019). Getting rid of visual distractors: the why, when, how, and where. Curr. Opin. Psychol. *29*, 135–147. https://doi.org/10.1016/j.copsyc.2019.02.004.

55. Johnsrude, I.S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H.P., and Carlyon, R.P. (2013). Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. Psychol. Sci. *24*, 1995–2004. https://doi.org/10.1177/0956797613482467.

56. Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., and Lalor, E.C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr. Biol. *28*, 803–809.e3. https://doi.org/10.1016/j.cub.2018.01.080.

57. Fox, E. (1995). Negative priming from ignored distractors in visual selection: a review. Psychon. Bull. Rev. *2*, 145–173. https://doi.org/10.3758/BF03210958.

58. May, C.P., Kane, M.J., and Hasher, L. (1995). Determinants of negative priming. Psychol. Bull. *118*, 35–54. https://doi.org/10.1037/0033-2909.118.1.35.

59. Banks, W.P., Roberts, D., and Ciranni, M. (1995). Negative priming in auditory attention. J. Exp. Psychol. Hum. Percept. Perform. *21*, 1354–1361. https://doi.org/10.1037/0096-1523.21.6.1354.

60. Mayr, S., Möller, M., and Buchner, A. (2011). Evidence of vocal and manual event files in auditory negative priming. Exp. Psychol. *58*, 353–360. https://doi.org/10.1027/1618-3169/a000102.

61. Mayr, S., and Buchner, A. (2010). Auditory negative priming endures response modality change; prime response retrieval does not. Q. J. Exp. Psychol. *63*, 653–665. https://doi.org/10.1080/17470210903067643.

62. Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. Trends Cogn. Sci. *12*, 182–186. https://doi.org/10.1016/j.tics.2008.02.003.

63. Eben, C., Koch, I., Jolicoeur, P., and Nolden, S. (2020). The persisting influence of unattended auditory information: negative priming in intentional auditory attention switching. Atten. Percept. Psychophys. *82*, 1835–1846. https://doi.org/10.3758/s13414-019-01909-y.

64. Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. Neuron *93*, 480–490. https://doi.org/10.1016/j.neuron.2016.12.041.

65. Tune, S., Alavash, M., Fiedler, L., and Obleser, J. (2021). Neural attentional-filter mechanisms of listening success in middle-aged and older individuals. Nat. Commun. *12*, 4533. https://doi.org/10.1038/s41467-021-24771-9.

66. Makov, S., Pinto, D., Har-shai Yahav, P., Miller, L.M., and Zion Golumbic, E. (2023). Unattended, distracting or irrelevant": theoretical implications of terminological choices in auditory selective attention research. Cognition *231*, 105313. https://doi.org/10.1016/j.cognition.2022.105313.

67. Ochsner, K.N., Silvers, J.A., and Buhle, J.T. (2012). Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. Ann. N. Y. Acad. Sci. *1251*, E1–E24. https://doi.org/10.1111/j.1749-6632.2012.06751.x.

68. Skinner, B.F. (1938). The Behavior of Organisms: An Experimental Analysis (Appleton-Century).

69. O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., and Lalor, E.C. (2015). Attentional selection in a cocktail party environment can Be decoded from single-trial EEG. Cereb. Cortex *25*, 1697–1706. https://doi.org/10.1093/cercor/bht355.

70. Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436. https://doi.org/10.1163/156856897X00357.

71. Kleiner, M. (2007). What's New in Psychtoolbox-3? 89.

72. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442.

73. Waschke, L., Wöstmann, M., and Obleser, J. (2017). States and traits of neural irregularity in the age-varying human brain. Sci. Rep. *7*, 17381. https://doi.org/10.1038/s41598-017-17766-4.

74. Wöstmann, M., Waschke, L., and Obleser, J. (2019). Prestimulus neural alpha power predicts confidence in discriminating identical auditory stimuli. Eur. J. Neurosci. *49*, 94–105. https://doi.org/10.1111/ejn.14226.

75. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869. https://doi.org/10.1155/2011/156869.

76. Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., and Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. J. Neural. Eng. *14*, 036020. https://doi.org/10.1088/1741-2552/aa66dd.

77. Chi, T., Ru, P., and Shamma, S.A. (2005). Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. *118*, 887–906. https://doi.org/10.1121/1.1945807.

78. Hausfeld, L., Riecke, L., Valente, G., and Formisano, E. (2018). Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. Neuroimage *181*, 617–626. https://doi.org/10.1016/j.neuroimage.2018.07.052.

79. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Despoited data		
Electroencephalography (EEG) and behavioural data of 19 human participants	This study; Open Science Framework	https://osf.io/nwaz5
Software and algorithms		
MATLAB R2018a	Mathworks, Natick, MA, USA	http://www.mathworks.com/products/matlab/ RRID:SCR_001622
Psychtoolbox 3	Brainard,[70]; Pelli,[72]	http://psychtoolbox.org/; RRID:SCR_002881

RESOURCE AVAILABILITY

Lead contact

Martin Orf (m.orf@uni-luebeck.de, Dept. of Psychology, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany, Telephone: +49 451 3101 3614).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data associated with this manuscript will be available as of the date of publication at the Open Science Framework (OSF) repository (https://osf.io/nwaz5/). The OSF project contains all data files used in this study, including raw and processed data.
- All original code associated with this manuscript will be available as of the date of publication at the Open Science Framework (OSF) repository (https://osf.io/nwaz5/).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Nineteen young adults (12 female, 7 male), aged between 18 and 27 years (Ø 21.9) participated in the present study. All participants had German as their mother tongue and reported normal hearing and no histories of neurological disorders. We did not explicitly ask for race or ancestry, but the sample was drawn from a broadly Caucasian student population. To verify normal hearing, we measured pure tone audiometry within a range of 125 to 8,000 Hz. All participants showed auditory thresholds below 20 dB for the tested frequencies. They gave written informed consent and received compensation of 10 €/hour. The study was approved by the local ethics committee of the University of Lübeck.

METHOD DETAILS

Stimulus materials and spatial cue

We presented three different narrated book texts as audio, spoken by different male, professional talkers (“Michael Kohlhaas” by Heinrich von Kleist, “Pole Poppenspäler” by Theodor Storm, and “Das Wrack” by Friedrich Gerstäcker). We chose audio streams that were fictional instead of fact-based, to minimise the impact of variations in prior knowledge on a topic and a resulting possible bias to one of the audio streams. All three audio streams overlapped in time, at an SPL (mixture) of about 65 dB(A), which matches normal conversation levels.

The following processing steps of the stimuli were done using custom written code in MATLAB (Version 2018a Mathworks Inc., Natick, MA, United States). The sound files were sampled with 44.1 kHz and a

16-bit resolution. The sound level was matched to the same long-term root-mean-square (rms) dB full scale (dBFS) between the three audio streams. Silent periods were truncated to maximally last 500 ms.[69]

We embedded short repeats in the audio streams by pseudo-randomly selecting a 400-ms segment from the original stream and repeating it directly thereafter.[34] The first repeat was presented at least two seconds after stimulus onset. Each repeat was included in the sound stream by a linear ramping and cross-fading. The linear ramping was done by using a window of 220 samples (5 ms) of the end of the to be repeated part (down ramp) and by using the first 220 samples (5 ms) of the repeat itself (up ramp). The cross-fading was done by adding the down and up ramp together.

The onset time of each repeat was drawn randomly to avoid predictability of the repeat. To avoid that repeats occurring in the different streams overlap in time, the distance between two repeat onsets was at least 2 seconds.

We further used a rms (root mean square) criterion (the rms of the repeat had to be at least the same as the rms of the stream from which the repeat was drawn) to avoid undetectable repeats of low sound intensity.The cue was presented at the center of the screen (resolution: 1920x1080, Portable HDMI Screen, Wimaxit) in front of the participant (distance: 1 m). The cue (Figure 1A) consisted of three sub-triangles that had a size of 1.3° and pointed to the three sound sources (front, left, and right). The background of the screen (RGB: 127, 127, 127), the cued sub-triangle (RGB: 204, 204, 204), and the not cued triangles (RGB: 115, 115, 115) were kept in different shades of gray to keep the contrast low. The bright triangle indicates the to-be-attended position. Since the cue and the fixation cross were presented at the same time as the auditory stimuli, we ensured that the possible interference between visual and auditory neural responses was as small as possible. To this end, the change between the fixation cross and cue was made smooth by linearly fading in and fading out (50 ms each) the cue.

Experimental setup

The experiment took place in a laboratory space with eight loudspeakers (Genelec: Speaker 8020D, Denmark) arranged in a circle with a radius of one meter. The loudspeakers were spaced at 45 degrees. A chair was placed in the middle of the radial speaker array, face-aligned to the loudspeaker at position 0°. The three audio streams were presented over the three frontmost loudspeakers (-45°, 0°, and 45° in the azimuth plane, elevation was not adjusted for participants' height, ground-to-loudspeaker distance: 1,20 m, the five remaining speakers were not used in the present experiment). In advance, participants were briefed about the experiment. Importantly, they were not briefed about the condition-to-location assignment of the streams. Each participant was asked to keep their eyes open, focus on the center of the screen, and sit as relaxed as possible. To avoid head motion, a chin rest was used. The height of the chin rest was adjusted for each participant.

Experimental procedure

We created a new experimental paradigm to investigate the underlying neural mechanism of selective attention (Figure 1). The experiment was designed using MATLAB (Version 2018a Mathworks Inc., Natick, MA, United States) and Psychophysics Toolbox extensions.[70–72] Participants were presented with three concurrent audio streams. Each trial had a total duration of 20s and started with a cue. The cue indicated which location to attend. The cue was presented for 500 ms. After the cue, a fixation cross was presented for the remaining trial duration (19.5 s). However, the auditory stimuli were presented simultaneously with the cue and the fixation cross resulting in a continuous playback of the auditory stimuli without any breaks between trials. Hence, the next trial started instantly after the trial before.

Each participant had to switch their attentional focus between the same two streams and locations. The stream at the cued location was defined as target, the stream cued in the previous trial was defined as distractor. Crucially, this left one, never task-relevant stream and location for each participant, here defined as neutral. Between participants, we implemented three condition-to-location assignments to avoid any confound with the position of the neutral stream (neutral front (0°), neutral left (-45°) and neutral right (45°). We aggregated across the three condition-to-location assignments to obtain our measures of interest, i.e., neutral tracking of target, neutral and distractor. As the position of the neutral stream, the different audio streams were almost balanced between the 19 participants (neutral front: n=7; neutral right n= 6; neutral left n= 6).

Participants had to detect short repeats in the target stream. Each trial contained 6 repeats, which were randomly partitioned in the three streams (for procedure details see section: stimulus materials and spatial cue) Before data collection, participants were familiarized with the experiment. During instruction, it was emphasized to respond as fast and accurately as possible to a repeat in the target stream, but also to listen to the content of the target stream. To familiarize participants with the repeats, we presented them a single sentence with one repeat included. They had to give oral feedback if they were able to detect the repeat. Further, we presented them with 6 training trials corresponding to the main experiment but using different audio streams. The main experiment consisted of 180 trials divided in 4 blocks, resulting in a total duration of 60 min. After each block, participants were able to take a rest. The total number of repeats was 360 per stream across the experiment.

We asked participants 15 multiple choice questions (with four possible answers, each) about the content of each audio stream at the end of the experiment. To avoid participants attending to the to-be-ignored audio stream, we did not ask the questions after every block. The order of the questions and the possible answers were randomized between participants.

Data acquisition and pre-processing

EEG was recorded using a 24 electrodes EEG-cap (Easycap, Herrsching, Germany; Ag–AgCl electrodes placed according to the 10-20 International System) connected to a SMARTING amp (mBrainTrain, Belgrade, Serbia). This is a mobile EEG system, which transfers the signal via Bluetooth to a recording computer.[73,74] EEG activity was recorded with the software Smarting Streamer (mBrainTrain, version: 3.4.2) at a sample rate of 500 Hz. During recording, electrode FCz served as online reference and impedances were kept below 20 kΩ. No data loss was reported during the sessions.

Offline, EEG preprocessing was done using MATLAB (Version 2018a Mathworks Inc., Natick, MA, United States), built-in functions, custom-written code, and the Fieldtrip-toolbox.[75] EEG-data were re-referenced to the average of the electrodes M1 and M2 (left and right mastoids) and high- and low-pass filtered between 1 and 100 Hz (two-pass Hamming window, FIR). An independent component analysis (ICA) was computed on each participants' EEG data. M1 and M2 were removed before ICA. ICA components related to eye blinks, eye movement, muscle noise, channel noise and line noise were identified by visual inspection and removed. On average, 8.37 of 22 (SD = 3.13) components were rejected. Components not associated with artifacts were back projected to the data. Clean EEG data were further processed. Hence, EEG data were low-pass filtered again at 10 Hz (two-pass Hamming window, FIR). Afterwards, EEG data were resampled to 125 Hz and segmented into epochs corresponding to the trial length of 20s.

Extraction of the speech envelope

The temporal fluctuations of speech were quantified by computing the onset envelope of each audio stream.[76] First, we computed an auditory spectrogram (128 sub-band envelopes logarithmically spaced between 90-4000 Hz) using the NSL toolbox.[77] Second, the auditory spectrogram was summed up across frequencies resulting in a broadband temporal envelope. Third, the onset envelope was obtained by computing the first derivative of this envelope and zeroing negative values to obtain the half-wave rectified first derivative. Finally, the onset envelope was down sampled to match the target sampling rate of the EEG analysis (125 Hz). Compared to the envelope, using the onset envelope shifts the envelope in time. Importantly, the TRF obtained by using the onset envelope as a regressor has the most similarity to a classical ERP.[76]

Temporal response functions (TRFs)

The deconvolution kernel or impulse response, which describes the linear mapping between an ongoing stimulus to an ongoing neural response, is called the temporal response function (TRF). We used a multiple linear regression approach to compute the TRF.[15] More precisely, we trained a forward model using the onset envelopes[16] of the target, distractor, and neutral speech to predict the recorded EEG response. In this framework, we analysed time lags between –100 and +500 ms between envelope changes and brain response.

To account for the EEG variance attributable to the detection and processing of the behaviourally relevant repeats and corresponding evoked brain responses, we also included all onsets of the repeats in the three streams and the button press in the model as nuisance regressors, represented by stick functions. The

onsets of the repeats are independent of the speech envelope regressors by design, since these were almost randomly (constraint of SNR threshold) added into the speech streams.

To prevent ill-posed problems and overfitting, we used ridge regression to estimate the TRF.[15] Lambda (λ) is the ridge parameter for regularization. We estimated the optimal ridge parameter that optimized the mapping between stimulus and response by leave-one-out cross-validation for each participant. First, the stimuli are segmented in M-trials and different ridge values ($\lambda = 2^0, 2^1, \ldots 2^{20}$) are predefined. In this approach, a separate model for each λ is calculated. Second, the trials are mixed, and each time one trial is left out. This trial is used as a test set, while the M-1 trials are used as a training set. Then, the models are averaged over the trials and convolved with the data from the matching test set to predict the neural response. This is done for every predefined λ. Computing the MSE between the predicted estimate and the original data provides a validation metric that enables to select the λ with the lowest MSE. We used the ridge value with the lowest MSE (specific for each subject) for the TRF model that jointly contained the target, distractor, and neutral onset envelopes as regressors.

TRFs were estimated based on the trials in the experiment. Participants had to switch their attention trial-wise between two of the streams. Hence, the trials enable the assignment of target, distractor, and neutral onset envelopes. The time window in which the stimulus and response are cut to estimate the TRF is referred to as a "trial." To avoid any conflicts with the cue, the first second of each trial was cut off in the EEG signal and the envelope onsets. One model was trained on 180 trials, incorporating multiple predictor variables: the onset envelope for target, distractor, and neutral streams; and the stick functions for the repeats and button presses. Resulting in a single TRF for each predictor variable that predicts a separable response component. Similar to the TRF approach, we estimated TRFs for the embedded repeats, but we modelled repeats as a stick function based on the repeat onset. Importantly, TRFs for the three streams, TRFs for repeats in the three streams, and button presses were estimated in the same model with the same regularization.

Neural tracking

Neural tracking quantifies how strongly a single stream is represented in the EEG signal. TRFs were used to predict the EEG response. The neural tracking (r) was calculated by correlating the predicted and measured EEG responses using Pearson correlation. We predicted the EEG signal on single trials using the leave-one-out cross-validation approach (see above). The r-values that resulted were averaged across trials and participants. We obtained the neural tracking accuracy over TRF time lags by using a sliding-time window of time lags (size: 48 ms, 6 samples) with an overlap of 24 ms (3 samples) for the prediction.[16,38,69,78] For every window position, the neural tracking was calculated, resulting in a time-resolved neural tracking. We used the term "stream tracking" which refers to the neural tracking of the envelope onsets, and "repeat tracking," which refers to the neural tracking of the repeat onsets. To obtain the repeat tracking, we used the same pipeline as for the speech tracking procedure (see above), with the exception that we estimated neural tracking based on the onsets of repeats (instead of the speech onset envelope), which we modelled as stick functions.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioural data analysis

We evaluated participants' behavioural performance in two ways. We analysed the proportion of detected repeats and, as a control, the proportion of correctly answered content questions.

We analyzed the detection of repeats in terms of signal detection theory. Button presses to repeats in a time window (150-1500 ms) after repeat onset were considered in this analysis. A button press following a repeat in the target stream was assigned as hit. Button presses following repeats in the distractor stream and in the neutral stream were assigned as separate types of false alarms. To differentiate between false alarms to repeats in the neutral versus distractor stream, we calculated sensitivity (d') between hit rate and false alarms to distractor repeats [$d'_{\text{target vs distractor}}$ = z(hit rate) – z(false alarm rate distractor)] and hit rate and false alarms to neutral repeats [$d'_{\text{target vs neutral}}$ = z(hit rate) – z(false alarm rate neutral)]. For this signal-detection analysis of repeats, we excluded one participant who did not respond to any repeats in the distractor stream.

A challenge in creating multiple-choice comprehension questions is to provide multiple (here: four) response options that cannot be solved based on prior knowledge or the possibility of excluding some of the response options. Hence, participants' actual guess rate might be considerably higher than the theoretical chance level of 25 %. Thus, in a pilot experiment, we presented the multiple-choice comprehension questions to N=9 different participants who had not listened to the audio streams at all. This resulted in a new, 'empirical' chance level of 40 % ($\pm$ 3.9 % S.E.M). In the following, we tested the proportion of correctly answered questions in the main experiment against this empirical chance level.

Neural data analysis

A study[16] investigated the attentional effects of neural tracking in a comparable continuous speech paradigm by recording the EEG of N = 18 participants. It is reasonable to expect that similar effect sizes will be observed in a replication of auditory attention effects with the same sample size. The present study is supposed to detect neural tracking effects with at least medium to large effect sizes (Cohen's $d \geq 0.7$) and a power of 80 % (two-sided, within-subject tests, Alpha = 0.05) for N = 18 subjects.

We also used different statistical procedures to answer different questions. To answer the main research question (outlined in Figure 1B), we used generalized mixed models (jamovi 1.6, R 4.0). This approach enables us to include and jointly model factors that potentially influence behaviour and the neural response. These included at least the factor condition-to-location assignment (neutral front, left, or right) and the subject as a random intercept to account for between-participant variability.To determine statistically significant differences in behavioural sensitivity (outcome measure), we included target versus distractor and target versus neutral as categorical predictors in the model.

To determine statistically significant differences in neural tracking (outcome measure), we included the target, neutral, and distractor streams as categorical predictors in the model. In both models, we included the factor condition-to-location assignment as a covariate and the random intercept (subject ID) into the model. Bayesian t-tests were calculated to obtain Bayes factors to quantify evidence for the null hypothesis. (JASP Team, 2022).

For quantifying the brain-behaviour relations, we used a generalized linear mixed-effects model (repeat detected or not; binomial distribution, with logit link function), since we predicted a binary outcome. The predicted outcome variable was the binary response to the detection of a single repeat in the target stream (Hit = 1; Miss = 0). We included the encoding accuracies for the target, neutral, and distractor streams as continuous, z-scored, fixed-effects predictors in our model. We assigned repeat tracking (trial-based) to each repeat within a trial. To again control for potential confounding between stream tracking and repeats, we also included repeat tracking similar to stream tracking in our model. Beside the factors condition-to-location assignment and subject as random intercepts, we also included the number of repeats during the total experiment and the number of repeats within a trial, as well as the trial number as a random intercept, into the model.

Statistical analysis on time series

We were looking for time points in time-resolved neural tracking that might differ between subjects (target enhancement: neutral vs. target, and active suppression: neutral vs. distractor). To answer this question, we used an established two-level statistical analysis, more specifically a cluster permutation test implemented in Fieldtrip.[75] Data from 22 channels was used in this analysis. As a test statistic at the single-subject level, we used one sample t-tests to test the time-resolved neural tracking to the target, neutral, and distractor as well as the neutral-target, neutral-distractor, and target-distractor difference against zero. At the group level, clusters were defined by the resulting t-values and a threshold that was set to t-values that corresponded to $p < 0.05$ for at least three neighboring electrodes. Each observed cluster is compared to 5000 clusters with a permutation distribution. The permutation distribution was generated by randomly assigning the time-resolved neural tracking data to conditions. The Monte Carlo method was used to correct for multiple comparisons. The relative number of Monte Carlo iterations in which the summed t-statistic of the observed cluster is exceeded is indicated by the cluster p-value.[79]