Neural Speech-Tracking During Selective Attention and Attention-Switching Between Concurrent Talkers: a spatially realistic audiovisual study

Paz Har-shai Yahav, Adi Korisky, Renana Vaknin Harel, Eshed Rabinovitch, Elana Zion Golumbic

The Gonda Center for Multidisciplinary Brain Research, Bar Ilan University, Israel

Corresponding author: Prof. Elana Zion-Golumbic

The Leslie and Susan Gonda (Goldschmied) Multidisciplinary Brain Research Center

Bar- Ilan University, Bldg 901, rm 412

Ramat-Gan 5290002, Israel.

Tel: +972 3 7384430

Email: elana.zion-golumbic@biu.ac.il

Abstract

Selective attention to a talker in multi-talker scenarios, is associated with enhanced neural speech-tracking of the target talker, relative to competing non-target talkers. This response is proposed to reflect the prioritization and preferential processing of the target speech in auditory and language related regions. Here we investigated whether this neural-bias for tracking target speech, persists also after listeners actively switch their attention to another target talker. Using a spatially-realistic audiovisual design, we studied responses at both the group and individual level.

Participants watched a video lecture (target speech) on a screen in front of them, which they were instructed to pay attention to and answered comprehension questions about its content. Concurrently, audio from an additional lecture, by a different talker, was played though a loudspeaker to their left, which they were told to ignore (non-target speech). Importantly, in the middle of the experiment, the lectures serving as target and non-target switched roles: the lecture that was formerly 'non-target' was presented as a video on the central screen and became the 'target', and vice versa. We compared behavioral performance and neural speech-tracking of both lectures before and after the switch, to test how effectively participants switched their attention and whether there are carry-over effects from before the switch.

At the group level, we found no significant differences in performance or in the neural-bias towards the target talker before vs. after the switch. This implies that, generally speaking, selective attention was just as effective after the switch as it was before. However, analysis of individual level data revealed a more complex and diverse pattern. We found that the neural group level results were driven by only 25% of participants, who showed consistent neural-bias towards the target talker both before and after the switch. However, 50% of participants showed neural-bias towards the target talker only in one half of the experiment, and an additional 25% showed no difference in the neural representation of the target vs. non-target speech, before or after the switch. We further show that these individual differences cannot be 'trivially' explained by poor signal quality, nor are they associated with behavioral performance. Rather, they suggest that neural-bias towards the target talker is not a prerequisite for selective attention performance, and in many cases target and non-target speech are co-represented in the neural signal. Whether this variability reflects the employment of different listening strategies or insufficient sensitivity of the neural-bias speech-tracking metric remains to be explored in future studies. However, these results challenge the assumptions that modulation of the speech-tracking response can be used as a one-to-one proxy for selective attention to speech in multi-talker contexts.

Introduction

Effectively directing attention to a single talker in multi-talker situations is a challenging task. Selective attention can facilitate the prioritization of processing a target talker amidst competing speech, and has been associated at the neural level with enhanced neural tracking of the target speech, compared to non-target speech (Ding et al., 2012; Fiedler et al., 2019; Har-shai Yahav & Zion Golumbic, 2021; Kerlin et al., 2010; O'Sullivan et al., 2015; Zion Golumbic, Ding, et al., 2013). This neural-bias for target speech has been demonstrated in numerous EEG and MEG studies of selective attention to speech, typically using a dichotic listening paradigm, where different speech stimuli are presented simultaneously to each ear, and participants are instructed to focus their attention on only one of them (Aydelott et al., 2015; Bentin et al., 1995; Brodbeck et al., 2020; Cherry, 1953; Makov et al., 2022). However, this paradigm assumes that selective attention is consistently directed towards one talker, whereas many real-life situations require individuals to switch their attention from one talker to another. For example, when watching the news on TV, if another person in the room begins a conversation, the listener may shift their attention towards them and away from the speech of the TV news anchor, which will be relegated to the 'perceptual background' as it is no longer the focus of attention. Attention-switching between talkers has not been studied as extensively as selective attention, but has been shown to affect speech processing and reduce intelligibility (Lin & Carlile, 2015; Teoh & Lalor, 2019; Uhrig et al., 2022) and recall of specific details (Best et al., 2008). Critically, little is known about how effectively listeners disengage from a speech source they had previously paid attention to, once it is no longer relevant to them, and how attention-switching affects neural speech-tracking of target and non-target speech.

Three different hypotheses can be considered regarding the representation of concurrent speech following an attention-switch. One possibility is that the former-target speech is still highly salient to the listeners and therefore continues to be well represented and processed by the brain, even though it is no longer behaviorally relevant to them. This hypothesis suggests that after attention-switching, listeners employ more of a 'divided attention' listening strategy, relative to before the switch (Agmon et al., 2021; Kaufman & Zion Golumbic, 2023; Lambez et al., 2020; Pinto et al., 2023). A second possibility is that the attention-switch is highly effective and listeners are able to 'tune out' and suppress speech by the former-target talker, just as well as they suppressed the non-target talker before the switch. According to this hypothesis, similar effects of selective attention should be observed before and after the attention-switch. A third possibility, driven by theories of voice familiarity and predictive coding, suggests that selective attention to the new target talker may actually *improve* after the attention-switch, since the voice and semantic context of the formerly-attended talker are more predictable and familiar to the listener, which can facilitate its more

efficient suppression (Gordon et al., 2019; Har-shai Yahav et al., 2023; Heilbron & Chait, 2018; Johnsrude et al., 2013).

In the current study, we tested these hypotheses, by comparing neural speech-tracking to target and non-target speech, before and after attention-switching. In a spatially-realistic audiovisual selective attention study, we recorded EEG from participants as they watched a video lecture on a screen in front of them (target talker), while the audio from another lecture was played from a loudspeaker to their left (non-target talker). Critically, in the middle of the experiment the two talkers switched roles, such that the lecture by the talker who had been non-target in the first half, was presented as a video lecture and became the target in the second half, whereas the lecture by the target talker from the first half of the experiment was shifted as audio-only to the loudspeaker on the left and became the non-target in the second half. Using a linear approach for speech-tracking analysis and auditory attention decoding (Crosse et al., 2016; Zion Golumbic, Ding, et al., 2013), we investigated how the neural representation of the concurrent speech stimuli was modulated by their status as target/non-target, and whether this was affected once they switched roles. Moreover, our choice to use natural audiovisual speech and to present audio in a spatially-realistic free-field manner, increases the ecological validity of this study, relative to past studies that mostly use audio-only speech presented through headphones (Brown et al., 2023; Freyman et al., 2001; Keidser et al., 2020; Ross et al., 2007; Shavit-Cohen & Zion Golumbic, 2019; Tye-Mmurray et al., 2016; Uhrig et al., 2022).

Materials and Methods

Participants

We collected data from 24 adult volunteers (16 female, 8 male), ranging in age between 19 and 34 (M=23.83, SD=±3.42). All participants were fluent Hebrew speakers with self-reported normal hearing and no history of psychiatric or neurological disorders. The study was approved by the IRB ethics board of Bar-Ilan University, and participants gave their written informed consent prior to the experiment. Participants were either paid or received course credit for participating in the experiment. Data from one participant was excluded from all analysis due to technical issues during EEG recording, therefore all further analyses are reported for N=23.

Speech stimuli

The stimuli consisted of two 20-minutes video recordings of a public lecture on popular science topics, one delivered by a male talker and the other by a female talker. Each video recording included the lecturer as well as the slides accompanying the talk. Both talkers gave their approval to use these materials for research purposes. Lecture videos were segmented, edited and volume equated using Matlab (The MathWorks), Filmora software (filmora.wondershare.net), Audacity software (version 3.2.1, www.audacityteam.org) and FFMPEG (www.ffmpeg.org). Lectures were cut into 63 segments ranging between 22 to 40 seconds each (lecture 1: M=31.83, SD=±4.2, lecture 2: M=30.68, SD=2.24). The varied lengths were necessary to ensure that the segments did not cut-off the lecture mid-sentence or mid-thought. Ultimately, the experiment included only 42 segments from each lecture, with 21 un-used segments from the middle of the lecture (see experimental procedure).

Experimental procedure

The experiment was programmed and presented to participants using OpenSesame software (https://osdoc.cogsci.nl). Participants were seated on a comfortable chair in a sound attenuated booth and were instructed to keep as still as possible. Participants viewed a video lecture, presented segment-by-segment, on a computer monitor in front of them with the lecture audio presented through a loudspeaker placed behind the monitor. They were instructed to pay attention to the video and after every three segments, were asked three multiple-choice comprehension questions, regarding the content of the recent segments (one question per segment). Participants received feedback regarding the correctness of their responses. They indicated via button press when they were ready to continue to the next segment.

In addition to the target lecture, the audio of an additional non-target lecture, was presented through a loudspeaker placed on their left side. The non-target speech was also presented segment by segment, but each one began 3 seconds after the onset of the target speech, and included a volume ramp-up of 2 seconds (Figure 1B). Both loudspeakers were placed in the same distance, approximately 95cm from participants' head. Critically, in the middle of the experiment, after 21 trials, the stimuli switched such that the non-target lecture that had been presented from a loudspeaker on the left in the first half, was presented as the target video lecture in the second half, whereas the lecture that had been the target in the first half, in the second half was presented as audio-only from a loudspeaker on the left, and was the non-target (Figure 1A). Participants were not informed in advance about the switch.

Importantly, in both halves of the experiment the speech presented as video (target talker) was always the first 21 segments of each lecture, to ensure that participants could follow and understand its content, whereas the speech presented as audio-only (non-target talker) consisted of the last 21 segments of each lecture. Accordingly, the lecture segments presented after the switch were never a direct continuation of the content presented before the switch, but were still part of the same lecture. Within each half, however, lecture segments were presented in sequential order. The order of the starting lecture (male/female talker) was counter balanced across participants. Audio from on-ear microphones and eye movements were also recorded during the experiment, but their analysis is outside the scope of this study.

EEG data acquisition

Electroencephalography (EEG) was recorded using a 64 Active-Two system (BioSemi; sampling rate: 2048 Hz) with Ag-AgCl electrodes, placed according to the 10–20 system. Two external electrodes were placed on the mastoids and served as reference channels. Electrooculographic (EOG) signals were simultaneously measured by 4 additional electrodes, located above and beneath the right eye and on the external side of both eyes.

EEG preprocessing

EEG preprocessing and analysis were performed using the Matlab-based FieldTrip toolbox (Oostenveld et al., 2011), as well as custom written scripts. Raw data was first re-referenced to the linked left and right mastoids and was bandpass filtered between 0.5-40Hz. Data were then visually inspected and gross artifacts (that were not eye-movements) were removed. Independent component analysis (ICA) was performed to identify and remove components associated with horizontal or vertical eye-movements as well as heartbeats. Any remaining noisy electrodes that

exhibited either extreme high-frequency activity or low-frequency drifts, were replaced with the weighted average of their neighbors using an interpolation procedure. The clean data was then segmented into trials and divided to the two halves of the experiment – before and after attention-switching.

Behavioral data analysis

Behavioral data consisted of accuracy on the comprehension questions asked about each segment. These values were averaged across trials, separately for each half of the experiment, and for each participant. We used a two-tailed paired t-test to evaluate whether accuracy rates differed significantly before and after the attention-switch.

EEG data analysis

Reliable Components Analysis (RCA)

Clean segmented data was spatially filtered using the RCA Matlab toolbox (Dmochowski et al., 2012). This method focuses on extracting the most reliable information, facilitating more effective data evaluation and hypothesis testing. Using RCA, we were able to simplify our further analyses by reducing data dimensionality. This is achieved by condensing multivariate data into a set of components, each of which represents a combination of sensors and exhibits distinctive spatial and temporal characteristics. Therefore, further encoding analyses (see speech-tracking analysis) were conducted on the resulted data from the first RCA component (Figure 3A).

Importantly, the RCA algorithm requires an equal amount of trials across all participants, and therefore 4 additional participants, that were missing data recordings from 1 trial were removed, resulting with N=19 for the RCA analysis only.

Speech-tracking analysis

To estimate the neural response to the talkers in the different noise scenarios we performed speech-tracking analysis, using both an encoding and a decoding approach. We estimated linear Temporal Response Functions (TRFs) using the mTRF MATLAB toolbox (Crosse et al., 2016), which constitutes a linear transfer function describing the relationship between a particular feature of the stimulus (S) and the neural response (R) recorded when hearing it.

The S used here was the broadband envelope of the speech stimuli presented in each trial, which was extracted using an equally spaced filterbank between 100-10,000 Hz based on Liberman's cochlear frequency map. The narrowband filtered signals were summed across bands after taking the absolute value of the Hilbert transform for each one, resulting in a broadband envelope signal.

The R used here was the continuous cleaned EEG data, bandpass filtered between 1 and 20 Hz. S and R were aligned in time and down-sampled to 100 Hz for computational efficiency. The 5 first seconds of each trial were excluded from data analysis, due to the differences in onset and ramp-up period (Figure 1B). Note that the first trial subsequent to the attention-switch was deliberately omitted from data analysis. this exclusion aimed to preclude potential attentional ambiguity and confounding effects associated with the immediate adjustment period following the switch, resulting in 20 trials per half.

Encoding and decoding models were run and optimized separately for each half of the experiment. Encoding TRFs were calculated over time lags ranging from −150 (pre-stimulus) to 450 ms, and the decoding analysis used time lags of 0 to 400 ms.

Leave-one-out cross validation protocol was used to assess the TRF predictive power. In each iteration, 90% of the trials are selected to train the model (train set), which was then used to predict either the neural response at each electrode (encoding) or the speech envelope (decoding) in the left-out trial (test set). The goodness of fit (predictive power) of the encoding model was determined by calculating the Pearson correlation between the predicted and actual neural response at each sensor. Similarly, the goodness of fit of the decoding model was determined by calculating the Pearson correlation between the predicted and actual speech envelope (reconstruction accuracy). To prevent overfitting of the model, a ridge parameter was chosen as part of the cross-validation process (λ -predictive power). This parameter significantly influences the shape and amplitude of the TRF and therefore, rather than choosing a different λ for each participant (which would limit group level analyses), a common λ value was selected for all participants, based on results from previous studies (Har-shai Yahav et al., 2023; Kaufman & Zion Golumbic, 2023). For the encoding model this value was λ=1000 and for the decoding model this value was λ=500. These values were chosen using a similar procedure as that discussed by Kaufman et al. (2023), namely: a range of λ values was tested and the value that yielded the highest average predictive power, across all channels and participants, was chosen as the common optimal λ. All results reported here use these ridge parameters. Importantly, encoding models was calculated on RCA (reduced dimensions) data, whereas decoding analysis was performed on the full (un-reduced) data.

Group level statistics

First, we evaluated both encoding and decoding models at the group level. All comparisons were conducted for the first and second halves separately. Specifically, we assessed the statistical differences of the predictive power separately for target and non-target speech, using a two-tailed

paired t-test. Note that predictive power was calculated from the reduced RCA data, and therefore resulted in a single value per participant. Then, encoding models were compared using a temporal cluster-based permutation test ($p<0.05$ corrected for multiple comparisons; FieldTrip toolbox). Lastly, to test for statistical differences between decoding accuracies, we ran a two-tailed paired t-test.

Individual level statistics

In addition to the group level analysis, we also aimed to assess the degree of which the effect of selective attention was present in individual participants. In this analysis we used only the decoding approach, rather than including multiple comparisons over sensors and time (i.e. the encoding model) and was performed separately for each half of the experiment. For each participant we calculated a selective attention-bias index, which is the difference in the reconstruction accuracy for the target and non-target talkers, similar to the approach used in Kaufman et al. (2023). To account for individual differences in SNR and to obtain a data-driven estimation of chance-level, we used 100 permutations to create an "attention-agnostic" null distribution of bias indices. Specifically, in each permutation the two speech stimuli were randomly re-labeled so that the speech included in each regressor was 50% target and 50% non-target. Two decoders were trained on the data using these attention-agnostic labelings, and the reconstruction accuracy for each regressor was estimated as well as the difference between them (attention-agnostic bias index). The real attention-bias index of each participant was normalized (z-scored) relative to this null distribution. If the selective attention instructions truly lead to enhanced neural representation of the target talker relative to the non-target one, this would be reflected in a positive normalized attention-bias index that exceeded a z-value threshold. However, if the normalized attention-bias index falls within this null distribution (beneath threshold), this would suggest that the difference between target and non-target speech is not larger than could be achieved by chance. This also enabled us to compare between the two halves of the experiment separately for each participant and thus test for selective attention-bias differences before and after attention-switching at the individual level. We chose a relatively permissive threshold in order to observe the individual differences in an exploratory manner, rather than a tight statistical approach. Therefore, a participant showing z-score>1 was considered to have a selective attention-bias. Accordingly, participants were divided into subgroups, based on whether they had an attention-bias or not, in each half of the experiment (z-score threshold=1), resulting in four subgroups: attention-bias in both halves, no attention-bias, attention-bias in first half only, attention-bias in second half only.

Additionally, to eliminate the possibility that a lack of attention-bias resulted from poor speech-tracking or low signal-to-noise ratio (SNR), we compared the reconstruction accuracies averaged across both speech stimuli (target and non-target) and compared them between the four attention-bias subgroups, using a 1-way ANOVA (JASP-Team, 2022). Moreover, Pearson correlations between these decoding accuracies and attention-bias indices were calculated as well, separately for each half of the experiment.

Lastly, we evaluated behavioral performance across the four subgroups in order to eliminate the possibility that a neural attention-bias is not a prerequisite for adequately comprehending the target speech (1-way ANOVA). Pearson correlations between behavioral performance and attention-bias indices were evaluated as well, separately for each half of the experiment.

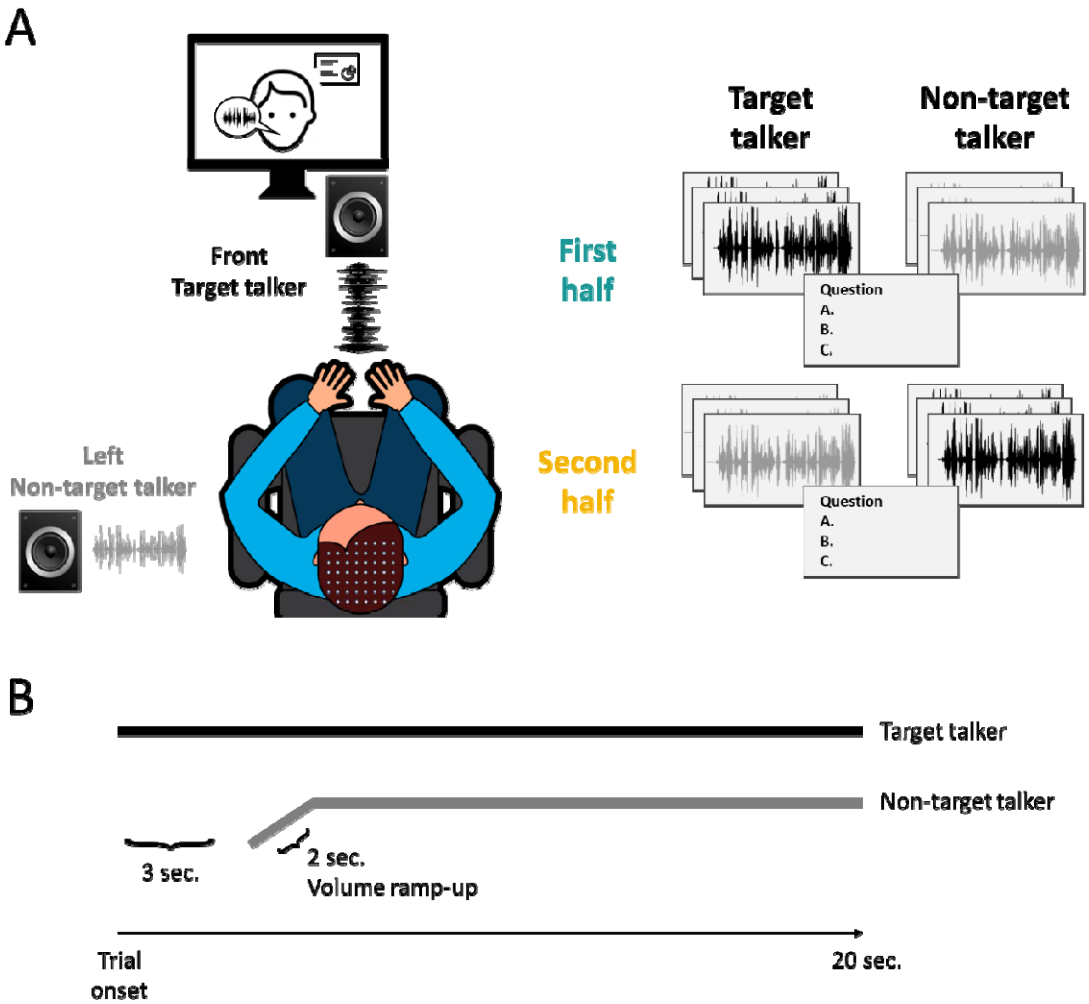

Figure 1 Experimental setup. A. Two lectures were presented simultaneously, with one lecture (target talker) displayed on the screen and audio emitted through the front loudspeaker. The other lecture (non-target talker) was played audio-only through the left loudspeaker. Participants were instructed to focus their attention on the lecture presented on the screen. Critically, in the middle of the experiment, the stimuli switched such that the

lecture that was played from the side and had been non-target in the first half, was presented as a video in the second half and became the target talker, whereas the target talker from the first half was presented in the second half from a loudspeaker on the side and was non-target. Participants answered comprehension questions regarding the target lecture after every three trials. B. Single trial illustration. Target speech began at trial onset and non-target speech began 3 seconds after target speech onset and included a 2-second volume ramp-up.

Results

Behavioral data

Accuracy on answering comprehension questions about the content of the lecture was significantly above chance [M=0.845, SD=±0.076; t(22)=32.178, $p<10^{-20}$]. No significant differences in performance were observed between the first half (M=0.846, SD=±0.09) and second half (M=0.856, SD=±0.1) of the experiment [t(22)=-0.42, p=0.68], as shown in Figure 2.

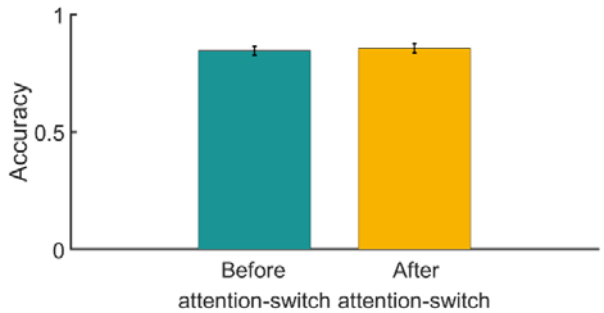

Figure 2 Behavioral results. Averaged accuracy rates across trials and participants, for multiple-choice comprehension questions, separately for the first (green) and second (yellow) half of the experiment. Error-bars denote SEM across participants.

Speech-tracking analysis

Group level

Figure 3 shows the results of the encoding approach. TRF models were estimated for target and non-target speech, separately for data before and after the switch (dimension-reduced using RCA analysis). The RCA component that was used in our analyses followed the common centro-frontal topography characteristic of auditory responses in EEG (Figure 3A). As expected, TRFs for target speech had larger amplitudes than for non-target speech. However, no significant differences were found for either stimuli between the TRF amplitudes before vs. after the switch (Figure 3B). This was evident in the predictive power of the model as well, that was significantly higher for target vs. non-target speech before attention-switch only [before attention-switch: t(18)=2.78, p=0.012; after attention-switch: t(18)= 0.56, p=0.58], but did not differ significantly between the first and second half of the experiment, for either target and non-target speech (Figure 3C).

Similar results were found using the decoding approach (Figure 3D). Confirming the effect of selective attention, decoding accuracies were significantly higher for target vs. non-target speech, in both halves of the experiment [before attention-switch: t(22)=3.24, p=0.0038; after attention-switch: t(22)= 4.434, $p<10^{-4}$]. However, no significant differences between the two halves of the experiment were found in decoding accuracies of the target (p=0.86) and the non-target speech (p=0.75), nor was there a significant interaction (p=0.87).

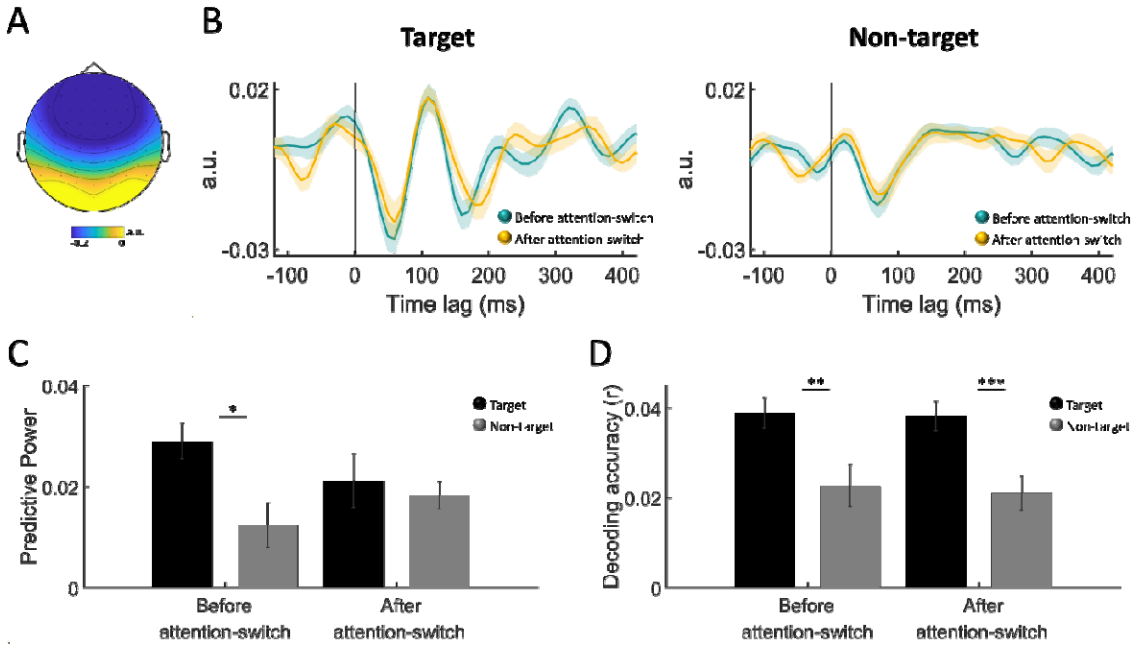

Figure 3 Group level TRF. A. Topography plot of the RCA component used to reduce data dimensionality. B. TRF encoding models (RCA data) for the first half (green) and second half (yellow) of the experiment, plotted separately for target and non-target speech. Shaded highlights denote SEM across participants. C. The predictive power for the first half and second half of the experiment, for both target and non-target speech. Error-bars denote SEM across participants. Asterisks indicate significant effects. D. Decoding accuracies for the first half and second half of the experiment, for both target and non-target speech. Error-bars denote SEM across participants. Asterisks indicate significant effects.

Individual level

We used a data-driven approach to examine the speech-tracking responses at the level of individual participants, and its modulation by experimental conditions, in attempt to gain more fine-grained insights into the interaction between instructed task-relevance and neural tracking of concurrent speech. For each participant we calculated a selective attention-bias index, normalized relative to a null distribution (Figure 4). We chose a conservative threshold of z=1 as a cutoff to quantify which participants exhibited significant attention-bias towards the target speech, in each half of the experiment. This analysis showed that in each half, only ~50% of participants showed evidence for attention-bias (N=12 in the first half and N=11 in the second half). Moreover, only N=5 participants (~25% of the sample) showed consistent attention-bias in BOTH halves of the experiment. Based on this analysis, we divide participants into four subgroups: Those showing attention-bias in both halves of the experiment (N=5), no attention-bias in either half (N=6), attention-bias in first half only (N=7), and attention-bias in second half only (N=5). For the subgroup who did show attention-bias in both halves of the experiment, no difference was found in the magnitude of attention-bias between the first and second half of the experiment [t(4)=-0.62, p=0.57], in line with the reported group level effects. However, the great variability in attention-bias

patterns across participants and the observation that most participants do not actually show a consistent attention-bias for the target speech, implies that group level results are driven by a small subset of participants and do not capture the underlying diversity in how the brain encodes concurrent speech.

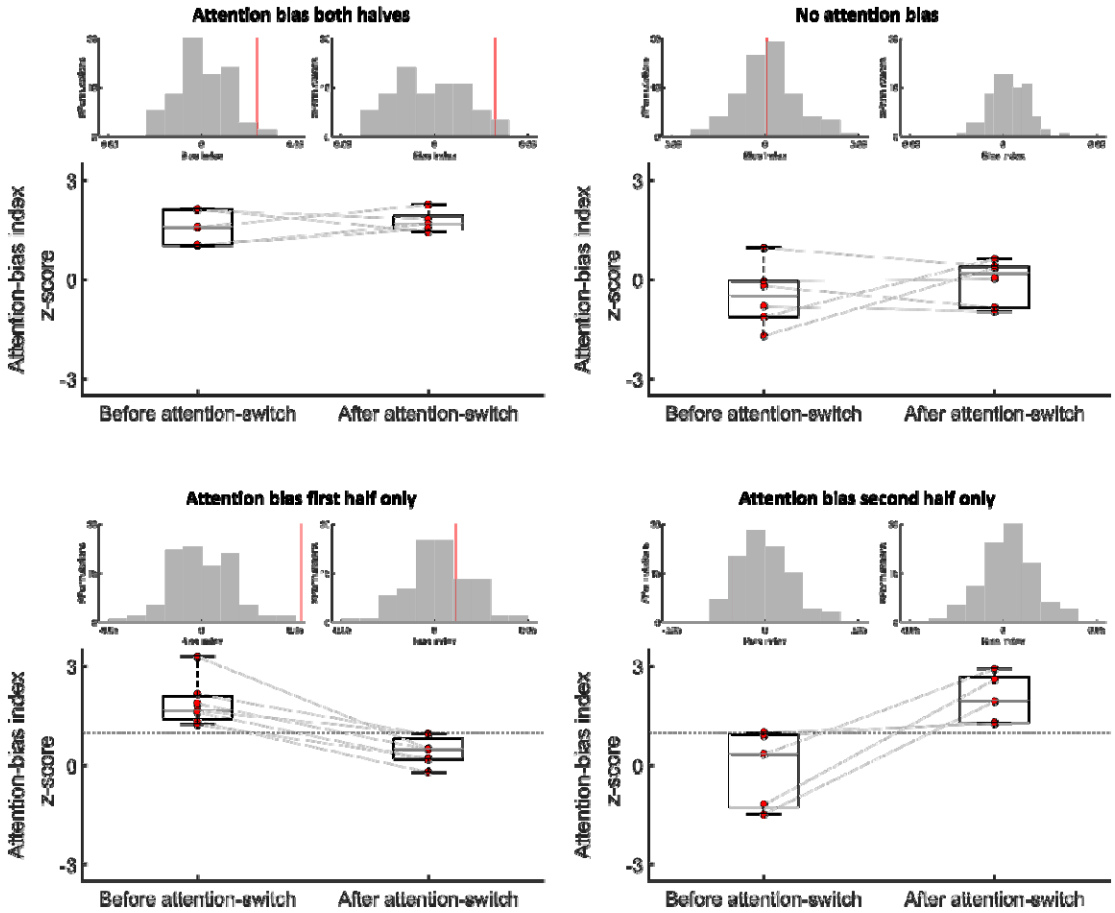

Figure 4 Selective attention-bias index – individual analysis. Attention-bias index calculated separately for each participant and normalized relative to a null distribution (100-repetition permutation test). This resulted in an attention-agnostic distribution of bias indices for each participant (example-subject shown at the top of each panel, real bias-index marked in red line). Participants were divided into four subgroups depending on whether they showed a selective attention-bias above threshold (z>1, marked in dashed line) or not, before and after attention-switching. The normalized selective attention-bias index for all participants within each group are presented in the four panels. Red circles denote individual values and boxes represent the 25–75$^{\%tile}$.

We conducted two follow-up analyses to better understand the diversity in neural attention-bias across participants. First, we examined whether the lack of attention-bias to target speech was simply a matter of poor signal quality or inadequate speech-tracking. To this end, we tested whether the overall decoding accuracy (averaged across target and non-target speech) was lower in participants who showed reduced attention-bias. If this was the case, it would support a trivial

explanation for the diversity across participants. However, no correlations were found between the overall decoding accuracies and the attention-bias indices in either half of the experiment [before switch: r=-0.37, p=0.08; after switch: r=-0.125, p=0.57; Figure 5 top]. Similarly, no differences were found in the overall decoding accuracies across the four subgroups of participants, in either the first or second half of the experiment [F(3)=0.417, p=0.74; F(3)=1.008, p=0.41, respectively; Figure 5 bottom]. This finding rules out the interpretation that a lack of attention-bias is a result of poor neural speech-tracking, and suggests instead that the two concurrent speech stimuli are equally co-represented in the neural response, perhaps reflecting a more 'distributed attention' listening strategy, as suggested by Kaufman and Zion Golumbic (2023).

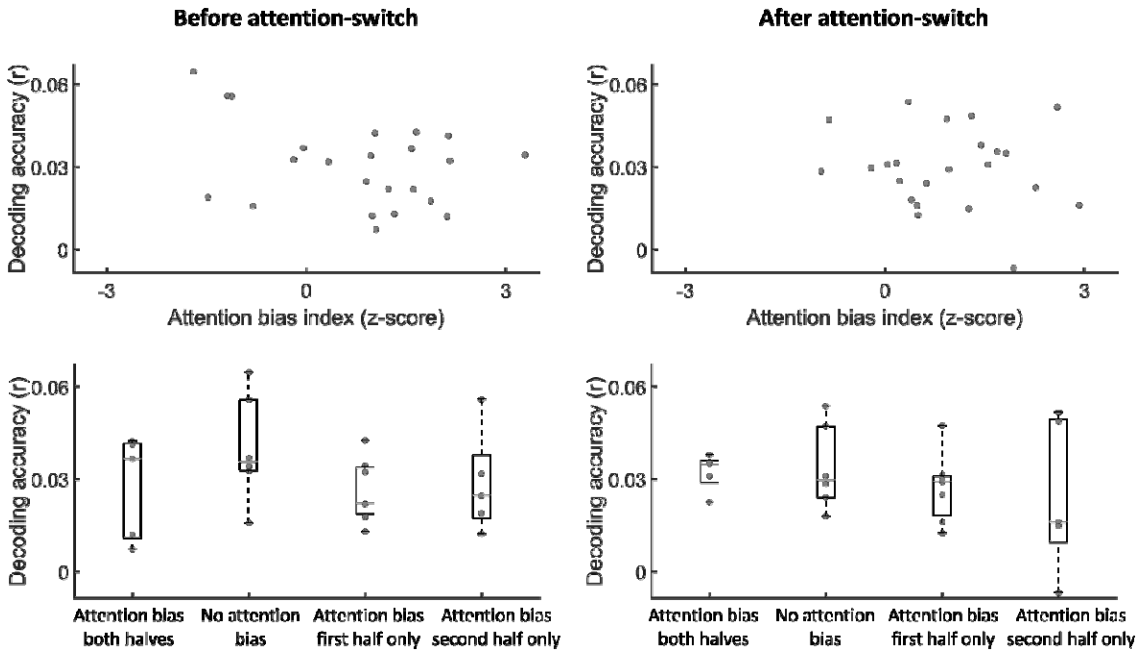

Figure 5 Speech decoding accuracies vs. attention-bias index. Top, Scatter plots showing the decoding accuracies and attention-bias indices for all participants, separately for the first and second half of the experiment, before and after the attention-switch. No significant correlations were found, in either half. Bottom, decoding accuracies (averaged for both target and non-target talkers), in each of the four subgroups of participants, shown separately for the first and second half of the experiment, before and after the attention-switch. Grey circles denote values of individual participants and boxes represent the 25–75$^{\%tile}$. No significant differences were found between decoding accuracies across subgroups.

Last, we tested whether task performance corresponds to the neural attention-bias metric, and whether individuals who display lower attention-bias also exhibit poorer performance. However, here too, we did not find significant correlations between the neural attention-bias indices and performance, in either half of the experiment [first half: r=-0.07, p=0.75; second half: r=0.116, p=0.56; Figure 6 top]. Moreover, no significant differences in performance were observed between the four subgroups, in either the first or second half of the experiment [F(3)=1.118, p=0.37;

F(3)=0.49, p=0.69, respectively; Figure 6 bottom]. These results further imply that neural attention-bias is not necessary in order to perform well on this selective attention task, as good performance could be achieved even in participants in whom both talkers were equally represented in the neural response.

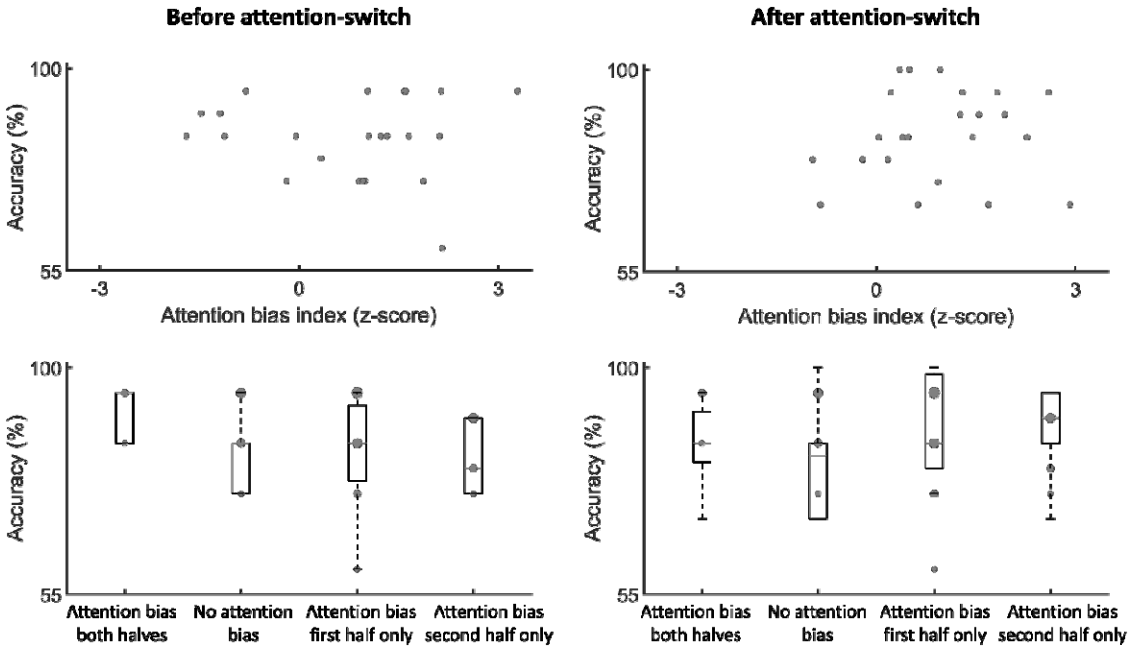

Figure 6 Performance on task vs. attention-bias index. Top, Scatter plots showing the accuracy levels and attention-bias indices for all participants, separately for the first and second half of the experiment, before and after the attention-switch. No significant correlations were found, in either half. Bottom, accuracy on answering comprehension questions about the target lecture, in each of the four subgroups of participants, shown separately for the first and second half of the experiment, before and after the attention-switch. Grey circles denote values of individual participants, and circle sizes represent the frequency of each value, with larger markers indicating higher occurrence. Boxes represent the 25–75$^{\%tile}$. No significant differences were found between accuracies across subgroups.

Discussion

The stated goal of the current study was to study how neural tracking of concurrent speech is affected by selective attention and by attention-switching. Analyzing the data at both the group and individual level revealed a complex and diverse pattern, suggesting no simple monolithic answer, and raising questions more broadly about the modulation of speech-tracking by selective attention. At the group level we replicate the well-established effect of selective attention, showing neural attention-bias and preferential neural tracking of the target talker, using both encoding and decoding approaches (Brodbeck, Presacco, et al., 2018; Har-shai Yahav & Zion Golumbic, 2021; Kaufman & Zion Golumbic, 2023; Niesen et al., 2019). This extends previous results to a spatially-realistic free-field setting and audiovisual listening condition. Importantly, at the group level, similar effects of selective attention were found before and after the attention-switch, and similar level of behavioral performance, suggesting good ability to disengage from the previous-target and re-align attention to the new target lecture after the switch.

However, a closer examination of individual level results revealed a more diverse pattern. When quantifying the attention-bias index for each participant, we found that the group level results were driven primarily by a subgroup of participants (~25%) who showed consistent neural attentional-bias towards the target speech, both before and after the switch. However, the majority of participants did not show significant neural-bias toward the target talker in at least one half of the experiment, and ~25% of participants showed no neural-bias in either half. Importantly, decoding accuracy was not lower in participants without attention-bias vs. those who exhibited consistent attention-bias in their neural response, suggesting that this variability across participants cannot be trivially explained by poor SNR or the inadequacy of the speech-tracking analysis. Rather, these findings suggest a more heterogeneous pattern of effects of selective attention across individuals, that is lost when focusing only on group averages. Below, we discuss the implications of these results for making inferences about selective attention, attention-switching and listening strategies from EEG-based speech-tracking data.

Group level effects: Highly effective selective attention and attention-switching

When designing this study, we considered three different hypotheses regarding the way that neural tracking of concurrent speech could be affected by switching roles between the target and non-target talkers: 1) following the switch the former-target talker speech remains highly salient to the listener and is more difficult to suppress, which would result in reduced neural-bias in the second vs. first half of the experiment; 2) listeners effectively switch attention to the new target talker, resulting in similar neural-bias before and after the switch; and 3) selective attention to the new target speech actually improves after the attention-switch, and the new non-target (formerly-

target) talker is easier to suppress due to its familiarity, resulting in an increase in neural-bias in the second half.

As noted, the group level results support the second hypothesis, as we found similarly robust effects of selective attention on the neural tracking of concurrent speech and equally good behavioral performance, before and after the attention-switch. This pattern suggests that, the dominant effect across the group is that of highly effective attention-switching, whereby speech by a previously-attended talker can successfully be suppressed and de-prioritized when it becomes irrelevant to the listener's behavioral goals. This finding stands in contrast to some previous findings regrading attention-switching, where performance was shown to adversely affect speech processing and behavior performance. For example, a recent study similar to ours, found that a previously-attended stream was more interfering in a behavioral task compared to a consistently task-irrelevant stream (Orf et al., 2023). Furthermore, other studies have demonstrated reduced speech processing, decreased intelligibility, and impaired recall of specific details, resulting from attention-switching between talkers (Best et al., 2008; Getzmann et al., 2017; Lin & Carlile, 2015; Teoh & Lalor, 2019; Uhrig et al., 2022). In contrast, the current experimental design probes more long-term effects of a previous focus attention. Here, the listener gains substantial familiarity with the voice and semantic context of one talker, features that are likely used to guide selective attention in the first half, and then is abruptly required to switch attention and disregard these same features and focus attention elsewhere. The finding that, at the group level, this switching did not seem to carry a behavioral or neural processing cost is testimony to the high adaptability and flexibility of the attentional system, which does not 'get stuck' on processing previously relevant features but updates its preferences according to changing task demands (Agmon et al., 2021; Kaufman & Zion Golumbic, 2023; Kiesel et al., 2010; Koch & Lawo, 2014).

Individual level effects: Variable listening strategies and dissociation between neural attention-bias and performance

In the majority of cognitive neuroscience research, only group level results are reported. However, following previous work from our group, showing substantial differences in the neural-bias to concurrent speech between individuals (Kaufman & Zion Golumbic, 2023), and given the many efforts to use neural-bias as a marker for selective attention in individuals (Fallahnezhad et al., 2023; Henshaw & Ferguson, 2013; Kidd, 2017; O'Sullivan et al., 2015), we felt it was important to delve deeper into the data of individual participants. Using a data-driven analysis developed by Kaufman & Zion Golumbic (2023), we quantified a neural attention-bias index for each participant, in each half of the experiment. Since group level results consistently show a neural-bias for target speech, we were surprised to find that only ~25% of participants showed this effect in both halves of the

experiment, ~50% of participants showed this effect in only one half of the experiment, and ~25% of participants showed no effect of selective attention whatsoever in their neural response. These results resonate with previous findings reported by Kaufman & Zion Golumbic (2023), where significant modulation of speech-tracking by selective attention was observed in only ~30% of participants, and trial-by-trial analysis suggested that neural-bias fluctuates between concurrent talkers (see also Ding et al., 2012; Ding & Simon, 2012; Fuglsang et al., 2017; Rosenkranz et al., 2021).

How should we interpret this versatility in individual level neural tracking results? One possibility could have been that the low neural-bias observed in some participants is a result of poor EEG signal quality which does not allow estimation of a sufficiently robust speech-tracking response. This possibility was ruled out in the current data, since we show that speech decoding performance can be just as good in individuals with high and low neural attentional-bias, with no correlation between the two measures. Another possibility could have been that participants with low neural-bias simply didn't perform the task and/or didn't really pay attention to the designated target talker. However, this interpretation also does not fit the current results, since we find similarly good performance in individuals with high and low neural attention-bias, with no correlation between them. Taken together, these observations challenge some key assumptions regarding the modulation of neural speech-tracking by selective attention, showing that although it is often observed, it is not a prerequisite/necessary component for successful performance on a selective attention task.

We offer two additional possible interpretations for these results. First, speech-tracking as quantified here as the envelope-following response, may not capture the complete extent of selective attention. Indeed, previous studies have shown a dissociation between the cortical regions that co-represent concurrent speech and those where attention selectivity is more prominent (Brodbeck, Elliot Hong, et al., 2018; Brodbeck, Presacco, et al., 2018; Har-shai Yahav & Zion Golumbic, 2021; Zion Golumbic, Ding, et al., 2013). Similarly, recent work has attempted to separate between neural tracking of acoustic- and linguistic/semantic- features of speech, suggesting that selective attention primarily affects the former (Brodbeck, Elliot Hong, et al., 2018; Ding et al., 2018; Lachter et al., 2004), although this is not always the case (Har-shai Yahav & Zion Golumbic, 2021; Parmentier, 2008; Parmentier et al., 2018; Vachon et al., 2019). Therefore, it is possible that by using more detailed characterization of the speech stimuli, a larger proportion of participants would have shown evidence for neural attention-bias. However, a second alternative explanation involves the notion of diverse listening strategies. Although in selective attention paradigms participants are instructed to pay attention only to one talker, we cannot control how they actually distributed their

attention nor do we have direct insight into the listening strategies employed (Makov et al., 2022). Moreover, since it is possible to understand the content of more than one concurrent speech stream (Agmon et al., 2021; Kaufman & Zion Golumbic, 2023; Vanthornhout et al., 2019), and some individuals are particularly apt at divided attention (Colflesh & Conway, 2007; Lambez et al., 2020), this may underly the decoupling between manifestations of neural attention-bias and performance. Therefore, it is highly possible that the diverse patterns of neural-biases (or the lack thereof) across individuals and across the two halves of the experiment reflect genuine differences in listening strategies and the way participants chose to prioritize the concurrent speech. Going back to the main research question posed in this experiment, according to this perspective, we would propose that the four subgroups of participants identified here represent **four different listening strategies** employed by different individuals: 1) Highly effective selective attention and attention-switching (manifest also in the group level results); 2) Highly effective selective attention in the first half, which deteriorates in the second half, either due to the difficultly to suppress a previous-target talker or due to fatigue; 3) Ability to co-represent both talkers and still maintain sufficient understanding of the target talker in order to answer comprehension questions; and 4) same as 3 in the first half, but with suppression of the non-target speech after the switch, due to its familiarity/redundancy. Since the current experiment was not designed specifically to test these hypotheses at the level of individual participants, additional research is required to substantiate the notion of different listening strategies. We hope that future follow up research will be able to shed additional light on the nature of individual differences in neural attention-bias, and will take into consideration the possibility of different listening strategies rather than assuming a uniform and fully selective listening strategy across all individuals (Makov et al., 2022).

Towards more ecological investigations of attention to speech

In this study we incorporated several design features aimed at enhancing its ecological validity. First, we chose to use natural audiovisual speech as the target stimulus, since this is the more prevalent mode of speech processing in real-life situations, and several studies have shown that having congruent visual cues enhances neural tracking of target speech (Crosse et al., 2015; Fu et al., 2019; Zion Golumbic, Cogan, et al., 2013). Second, the speech itself was taken from recordings of a real-life lecture, which more faithfully represents the type of speech that the brain deals with on a daily basis, as opposed to scripted materials read by a trained actor, such as audiobooks, used in many studies of selective attention (Agmon et al., 2023). Realistic speech can contain disfluencies, repetitions and redundancies that would not be expected in scripted materials, and typically has more variable speech rate (Goldman-Eisler, 1961; Miller et al., 1984), features that may dramatically

change the way that individuals listen and selectively attend to this type of speech. Last, the audio of both target and non-target speech was presented free-field through loudspeakers from different spatial locations, rather than through earphones, to better emulate the spatial acoustics of natural environments.

In moving towards more ecological circumstances and using natural continuous speech, we face an experimental challenge of obtaining sufficiently sensitive behavior measures without disrupting listening with an ongoing task. Therefore, we attribute the lack of a behavioral effect, primarily to the insensitivity of the behavioral measures used here, that consisted of asking multiple-choice question after each three 20s long narratives.

Conclusions

The current study focused on neural speech-tracking during selective attention and attention-switching in a multi-talker scenario and demonstrated a complex and diverse pattern at both the group and individual levels. While group level analyses showed no significant differences in behavioral performance or neural attention-bias toward the target talker before and after the attention-switch, individual-level data revealed a different picture. Only 25% of participants exhibited consistent neural attention-bias in both halves of the experiment, challenging the assumption that modulation of the speech-tracking response serves as a one-to-one indication for selective attention in multi-talker contexts. The results propose the existence of different listening strategies employed by participants, that cannot be trivially explained by signal quality or behavioral performance. Our work emphasizes the need for a more nuanced understanding of the neural mechanisms underlying selective attention in complex listening environments. Future studies are needed to explore individual differences in attentional strategies and consider the ecological validity of experimental designs to better capture the intricacies of real-world listening scenarios. Overall, these insights contribute to a more comprehensive understanding of how the brain navigates and adapts to dynamic changes in auditory attention.

References

Agmon, G., Har-Shai Yahav, P., Ben-Shachar, M., & Zion Golumbic, E. (2021). Attention to Speech: Mapping Distributed and Selective Attention Systems. *Cerebral Cortex*. https://doi.org/10.1093/CERCOR/BHAB446

Agmon, G., Jaeger, M., Tsarfaty, R., Bleichner, M. G., & Golumbic, E. Z. (2023). "Um…, It's Really Difficult to… Um… Speak Fluently": Neural Tracking of Spontaneous Speech. *Neurobiology of Language*, *4*(3), 435–454. https://doi.org/10.1162/NOL_A_00109

Aydelott, J., Jamaluddin, Z., & Nixon Pearce, S. (2015). Semantic processing of unattended speech in dichotic listening. *The Journal of the Acoustical Society of America*, *138*(2), 964–975. https://doi.org/10.1121/1.4927410

Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and unattended words in dichotic listening: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 54–67. https://doi.org/10.1037/0096-1523.21.1.54

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences*, *105*(35), 13174–13178. https://doi.org/10.1073/PNAS.0803718105

Brodbeck, C., Elliot Hong, L., & Simon, J. Z. (2018). Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Current Biology*. https://doi.org/10.1016/j.cub.2018.10.042

Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLOS Biology*, *18*(10), e3000883. https://doi.org/10.1371/journal.pbio.3000883

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*, *172*(September 2017), 162–174. https://doi.org/10.1016/j.neuroimage.2018.01.042

Brown, A., Pinto, D., Burgart, K., Zvilichovsky, Y., & Zion-Golumbic, E. (2023). Neurophysiological Evidence for Semantic Processing of Irrelevant Speech and Own-Name Detection in a Virtual Café. *Journal of Neuroscience*, *43*(27), 5045–5056. https://doi.org/10.1523/JNEUROSCI.1731-22.2023

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, *14*(4), 699–703. https://doi.org/10.3758/BF03196824

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *Journal of Neuroscience*, *35*(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli.

Frontiers in Human Neuroscience, *10*(November), 1–14. https://doi.org/10.3389/fnhum.2016.00604

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. *Psychological Review*, *70*(1), 80–90. https://doi.org/10.1037/h0039515

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention Is Required for Knowledge-Based Sequential Grouping: Insights from the Integration of Syllables into Words. *The Journal of Neuroscience*, *38*(5), 1178–1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(29), 11854–11859. https://doi.org/10.1073/pnas.1205381109

Ding, N., Simon, J. Z., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J Neurophysiol*, *107*, 78–89. https://doi.org/10.1152/jn.00297.2011

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Frontiers in Human Neuroscience*, *MAY 2012*. https://doi.org/10.3389/fnhum.2012.00112

Fallahnezhad, T., Pourbakht, A., & Toufan, R. (2023). The Effect of Computer-Based Auditory Training on Speech-in-Noise Perception in Adults: a Systematic Review and Meta-Analysis. *Indian Journal of Otolaryngology and Head and Neck Surgery*, *75*(4), 4198–4211. https://doi.org/10.1007/S12070-023-03920-0/TABLES/4

Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage*, *186*(August 2018), 33–42. https://doi.org/10.1016/j.neuroimage.2018.10.057

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112–2122. https://doi.org/10.1121/1.1354984

Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, *156*(December 2016), 435–444. https://doi.org/10.1016/j.neuroimage.2017.04.026

Fu, Z., Wu, X., & Chen, J. (2019). Congruent audiovisual speech enhances auditory attention decoding with EEG. *Journal of Neural Engineering*, *16*(6), 066033. https://doi.org/10.1088/1741-2552/AB4340

Getzmann, S., Jasny, J., & Falkenstein, M. (2017). Switching of auditory attention in “cocktail-party” listening: ERP evidence of cueing effects in younger and older adults. *Brain and Cognition*, *111*, 1–12. https://doi.org/10.1016/J.BANDC.2016.09.006

Goldman-Eisler, F. (1961). The Significance of Changes in the Rate of Articulation. *Http://Dx.Doi.Org/10.1177/002383096100400305*, *4*(3), 171–174. https://doi.org/10.1177/002383096100400305

Gordon, N., Tsuchiya, N., Koenig-Robert, R., & Hohwy, J. (2019). Expectation and attention increase the integration of top-down and bottom-up signals in perception through different pathways. *PLoS Biology*, *17*(4). https://doi.org/10.1371/journal.pbio.3000233

Har-shai Yahav, P., Sharaabi, A., & Zion Golumbic, E. (2023). The effect of voice familiarity on attention to speech in a cocktail party scenario. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhad475

Har-shai Yahav, P., & Zion Golumbic, E. (2021). Linguistic processing of task-irrelevant speech at a Cocktail Party. *ELife*, *10*. https://doi.org/10.7554/eLife.65096

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? In *Neuroscience* (Vol. 389, pp. 54–73). Elsevier Ltd. https://doi.org/10.1016/j.neuroscience.2017.07.061

Henshaw, H., & Ferguson, M. A. (2013). Efficacy of Individual Computer-Based Auditory Training for People with Hearing Loss: A Systematic Review of the Evidence. *PLoS ONE*, *8*(5), e62836. https://doi.org/10.1371/journal.pone.0062836

JASP-Team. (2022). *JASP (Version 0.16.3) [Computer software]*.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, *24*(10), 1995–2004. https://doi.org/10.1177/0956797613482467

Kaufman, M., & Zion Golumbic, E. (2023). Listening to two speakers: Capacity and tradeoffs in neural speech tracking during Selective and Distributed Attention. *NeuroImage*, *270*, 119984. https://doi.org/10.1016/J.NEUROIMAGE.2023.119984

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., Carpenter, M. G., Grimm, G., Hohmann, V., Holube, I., Launer, S., Lunner, T., Mehra, R., Rapport, F., Slaney, M., & Smeds, K. (2020). The Quest for Ecological Validity in Hearing Science: What It Is, Why It Matters, and How to Advance It. *Ear and Hearing*, *41*. https://journals.lww.com/ear-hearing/fulltext/2020/11001/the_quest_for_ecological_validity_in_hearing.2.aspx

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a Cocktail Party. *The Journal of Neuroscience*, *30*, 620–628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Kidd, G. (2017). Enhancing Auditory Selective Attention Using a Visually Guided Hearing Aid. *Journal of Speech, Language, and Hearing Research: JSLHR*, *60*(10), 3027. https://doi.org/10.1044/2017_JSLHR-H-17-0071

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching--a review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/A0019842

Koch, I., & Lawo, V. (2014). Exploring temporal dissipation of attention settings in auditory task switching. *Attention, Perception & Psychophysics*, *76*(1), 73–80. https://doi.org/10.3758/S13414-013-0571-5

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. *Psychological Review*, *111*(4), 880–913. https://doi.org/10.1037/0033-295X.111.4.880

Lambez, B., Agmon, G., Har-Shai Yahav, P., Rassovsky, Y., & Zion Golumbic, E. (2020). Paying attention to speech: The role of working memory capacity and professional experience. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-020-02091-2

Lin, G., & Carlile, S. (2015). Costs of switching auditory spatial attention in following conversational turn-taking. *Frontiers in Neuroscience*, *9*(MAR), 136588. https://doi.org/10.3389/FNINS.2015.00124/BIBTEX

Makov, S., Pinto, D., Har-shai Yahav, P., Miller, L. M., & Zion Golumbic, E. (2022). "Unattended, distracting or irrelevant": Theoretical implications of terminological choices in auditory selective attention research. *Cognition*, 105313. https://doi.org/10.1016/J.COGNITION.2022.105313

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*(4), 215–225. https://doi.org/10.1159/000261728/MACHINEREADABLECITATION/RIS

Niesen, M., Bourguignon, M., Vander Ghinst, M., Bertels, J., Wens, V., Choufani, G., Hassid, S., Goldman, S., & De Tiège, X. (2019). Cortical processing of hierarchical linguistic structures in adverse auditory situations. *Frontiers in Neuroscience*, *13*. https://doi.org/10.3389/conf.fnins.2019.96.00052

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Orf, M., Wö, M., Hannemann, R., & Obleser, J. (2023). Target enhancement but not distractor suppression in auditory neural tracking during continuous speech. *ISCIENCE*, *26*, 106849. https://doi.org/10.1016/j.isci.2023.106849

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

Parmentier, F. B. R. (2008). Towards a cognitive model of distraction by auditory novelty: The role of involuntary attention capture and semantic processing. *Cognition*, *109*(3), 345–362. https://doi.org/10.1016/j.cognition.2008.09.005

Parmentier, F. B. R., Pacheco-Unguetti, A. P., & Valero, S. (2018). Food words distract the hungry: Evidence of involuntary semantic processing of task-irrelevant but biologically-relevant unexpected auditory words. *PLoS ONE*, *13*(1), 1–17. https://doi.org/10.1371/journal.pone.0190644

Pinto, D., Kaufman, M., Brown, A., & Zion Golumbic, E. (2023). An ecological investigation of the capacity to follow simultaneous speech and preferential detection of ones' own name. *Cerebral Cortex*, *33*(9), 5361–5374. https://doi.org/10.1093/CERCOR/BHAC424

Rosenkranz, M., Holtze, B., Jaeger, M., & Debener, S. (2021). EEG-Based Intersubject Correlations Reflect Selective Attention in a Competing Speaker Scenario. *Frontiers in Neuroscience*, *15*, 685774. https://doi.org/10.3389/FNINS.2021.685774/BIBTEX

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., & Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: Deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophrenia Research*, *97*(1), 173–183. https://doi.org/https://doi.org/10.1016/j.schres.2007.08.008

Shavit-Cohen, K., & Zion Golumbic, E. (2019). The Dynamics of Attention Shifts Among Concurrent Speech in a Naturalistic Multi-speaker Virtual Environment. *Frontiers in Human Neuroscience*, *13*, 386. https://doi.org/10.3389/fnhum.2019.00386

Teoh, E. S., & Lalor, E. C. (2019). EEG decoding of the target speaker in a cocktail party scenario: Considerations regarding dynamic switching of talker location. *Journal of Neural Engineering*, *16*(3). https://doi.org/10.1088/1741-2552/ab0cf1

Tye-Mmurray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging*, *31*(4), 380–389. https://doi.org/10.1037/pag0000094

Uhrig, S., Perkis, A., Möller, S., Svensson, U. P., & Behne, D. M. (2022). Effects of Spatial Speech Presentation on Listener Response Strategy for Talker-Identification. *Frontiers in Neuroscience*, *15*, 730744. https://doi.org/10.3389/FNINS.2021.730744/BIBTEX

Vachon, F., Marsh, J. E., & Labonté, K. (2019). The Automaticity of Semantic Processing Revisited: Auditory Distraction by a Categorical Deviation. *Journal of Experimental Psychology: General*, *149*(7), 1360–1397. https://doi.org/10.1037/xge0000714

Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of Task and Attention on Neural Tracking of Speech. *Frontiers in Neuroscience*, *13*, 977. https://doi.org/10.3389/fnins.2019.00977

Zion Golumbic, E. M., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "Cocktail Party." *Journal of Neuroscience*, *33*(4). https://doi.org/10.1523/JNEUROSCI.3675-12.2013

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., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037