Attentional disengagement during external and internal distractions reduces neural speech tracking in background noise

Yue Ren[1*], M. Eric Cui[1,2], Björn Herrmann[1,2,*]

[1] Rotman Research Institute, Baycrest Academy for Research and Education
North York, ON, Canada

[2] Department of Psychology, University of Toronto
Toronto, ON, Canada

*Correspondence concerning this article should be addressed to Yue Ren and Björn Herrmann, Rotman Research Institute, Baycrest, 3560 Bathurst St, North York, ON, M6A 2E1, Canada. E-mail: yren@research.baycrest.org and bherrmann@research.baycrest.org

Keywords: Electroencephalography (EEG), naturalist speech encoding, background noise, attention, temporal response function

Conflict of interest statement: The author declares no competing interests.

Author contributions: YR: Methodology, data curation, formal analysis, visualization, writing - original draft, writing - review and editing, project administration. **MEC**: Data curation, formal analysis, writing - original draft, writing - review and editing, project administration. **BH**: Conceptualization, methodology, formal analysis, writing - original draft, writing - review and editing, supervision, funding acquisition.

Acknowledgements: We thank Tiffany Lao and Saba Junaid for their support with stimuli creation and data collection. This study has been supported by BH's Canada Research Chair program (CRC-2023-00383) and BH's Natural Sciences and Engineering Research Council of Canada (Discovery Grant: RGPIN-2021-02602).

Word count: Abstract (235), Significance Statement (107), Introduction (834), Discussion (1630)

Abstract

Within-situation disengagement – the mental withdrawal during conversations in acoustically challenging environments – is a common experience of older people with hearing difficulties. Yet, most research on the neural mechanisms of attentional disengagement from speech listening has focused on the distraction by one competing speaker, whereas within-situation disengagement is often characterized by distraction towards external visual stimuli or internal thoughts and occurs in situations with ambient, multi-talker background masking. Across three electroencephalography (EEG) experiments, the current study examined how disengagement due to external and internal distractions affect the neural tracking of speech masked by different levels of multi-talker babble (speech in quiet, +6 dB, and -3 dB SNR). We observed early (<200 ms), enhanced neural responses to the speech envelope for speech masked by background babble compared to speech in quiet (Experiments 1-3), suggesting stochastic facilitation. Importantly, neural tracking of the speech envelope was reduced when individuals were distracted by a visual-stimulus stream (Experiment 2) and by internal thought and imagination (Experiment 3). There were some indices suggesting the greatest disengagement-related decline in neural speech tracking occurs for the most difficult speech-masking condition, but this was not consistent across all measures. The current data show that disengagement due to external and internal distractions yield decreases in neural speech tracking, potentially suggesting a common neural pathway through which gain is downregulated in auditory cortex. These results indicate that disengagement from listening can be identified through non-invasive neural measures.

Significance

Many older adults with hearing difficulties mentally “tune out” during conversations in noisy environments, yet the neural mechanisms underlying this within-situation disengagement remain poorly understood. Across three electroencephalography experiments, we examined how external (visual) and internal (thought-based) distractions influence neural tracking of speech masked by multi-talker babble. We observed that attentional disengagement – whether induced by external stimuli or internal thoughts – reduced the brain’s tracking of the speech envelope. These findings demonstrate that listening disengagement can be objectively identified through neural measures and suggest a shared neural pathway through which both external and internal distractions down-regulate auditory gain, providing new insight into attentional control in challenging listening conditions.

Introduction

Everyday communication, such as conversations in cafés or restaurants, often involves listeners extracting meaning from a target talker amid multiple competing voices and background sounds. The increased cognitive resources, such as sustained attention and working memory, required to maintain comprehension under such conditions makes listening effortful, especially for older adults (Pichora-Fuller et al., 2016; Peelle, 2018; Herrmann and Johnsrude, 2020). When effort persists within a listening situation or the listening situation becomes overly taxing, listeners may temporarily withdraw attention and “tune out” (Heffernan et al., 2016; Herrmann and Johnsrude, 2020; McNeice et al., 2024; Sarant et al., 2024). This within-situation disengagement is a common response to sustained cognitive demand during listening (Heffernan et al., 2016; Herrmann and Johnsrude, 2020). Repeated experiences of within-situation disengagement may promote social withdrawal and eventually reduced social participation and connections (Motala et al., 2024). Identifying and characterizing markers of this disengagement could provide a powerful window into how attention fluctuates during listening and towards detecting disengagement in real-time.

Individuals may disengage from speech listening for multiple reasons. Externally, attention may be captured, by various factors, such as salient acoustic events, sustained visual distractions, or competing speech inputs, by which the processing of a target stimulus is disrupted (Lavie, 2005; Lavie and de Fockert, 2005; Shinn-Cunningham, 2008). Internally, attention can drift away from the external environment as listeners engage in self-generated thought – often characterized as mind wandering – during which cognitive resources are reallocated toward memories, future plans, or imaginings (Smallwood and Schooler, 2006; Eastwood et al., 2012; Smallwood and Andrews-Hanna, 2013; Ibaceta et al., 2024). Such internal diversion reflects a shift in the balance between perceptual and reflective processing: when perceptual demands exceed available capacity or motivational value, attention may retreat inward. Recent behavioral evidence indicates that this form of disengagement within a situation becomes more frequent when individuals listen to speech in the presence of background noise compared to when listening in quiet (Steindorf et al., 2023). Thus, external distraction and internal drift represent two pathways to the same functional outcome of temporarily uncoupling attention from speech, but whether such disengagement leaves distinct neural signatures and whether such signatures depend on listening demands is unclear.

One increasingly common approach to investigate the encoding of continuous speech in the brain involves regressing relevant features of speech against signals recorded using electro- or magnetoencephalography (EEG; MEG; Aiken and Picton, 2008; Lalor and Foxe, 2010; Ding and

Simon, 2012; Hertrich et al., 2012). This approach, formalized as temporal response function (TRF) modeling (Crosse et al., 2016; Crosse et al., 2021), is often used to quantify how neural activity tracks the temporal dynamics of the speech envelope, due to its importance for speech intelligibility (Ding et al., 2014; Vanthornhout et al., 2018; Lesenfants et al., 2019). The neural tracking of the speech envelope can be decreased when speech is masked by background sound (Ding and Simon, 2013; Synigal et al., 2023; Herrmann, 2025a), although this seems to depend on the specific TRF metric and type of background masker, sometimes even leading to noise-related increases that suggest stochastic resonance in speech-relevant brain regions (Herrmann, 2025c, a). Critically, a robust pattern across many studies is that the cortical tracking of the envelope of attended speech is stronger than the tracking of the envelope of speech that is ignored, reflecting selective amplification of behaviorally relevant neural signals (Kerlin et al., 2010; Ding and Simon, 2012; Power et al., 2012; Presacco et al., 2016; Fiedler et al., 2019; Fuglsang et al., 2020; Kraus et al., 2021; Tune et al., 2021; Orf et al., 2025). Yet, most prior research on the effects of attention on speech tracking has relied on paradigms where a single talker competes with the target speech, whereas communication also often occurs in contexts with diffuse multi-talker background babble. Individuals may be particularly susceptible to attentional inward drifts under such ambient background noise (Steindorf et al., 2023). However, whether neural speech tracking is sensitive to attentional disengagement due to internal or external distraction under such masking conditions remains unknown.

In the present study, we investigate how external and internal distractions affect the neural tracking of continuous speech across varying levels of speech clarity (clear speech and speech in background noise at two levels). Using EEG and TRF analyses, we first characterize the dependence of neural speech tracking on speech clarity (Experiment 1) and then investigate how disengagement – arising from external visual distraction (Experiment 2) or internal thought and imagination (Experiment 3) – affects this neural tracking under varying speech-clarity levels. We hypothesize that attentional disengagement – whether caused by external or internal distraction – will reduce the neural encoding of speech as reflected by diminished TRF metrics, and that this reduction may be strongest for speech masked by background noise. If external and internal distraction yield comparable decreases in neural speech tracking, this would suggest that both forms of disengagement converge on a shared neurocognitive mechanism, such as a down-regulation of auditory-cortical gain, potentially reflecting a common pathway through which attention decouples from sensory processing during effortful listening.

General Methods

Participants

Participants were aged 20-34 years, native English speakers or individuals who grew up in English-speaking countries (mostly Canada) and had been speaking English since early childhood (before age 6). All participants reported normal hearing abilities and no severe psychological or neurological disorders. Specific participant demographics are detailed for each experiment individually below.

Written informed consent was obtained prior to participation, and participants were compensated at a rate of $15 per hour. The study adhered to the Declaration of Helsinki, the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2-2014), and received ethical approval from the Research Ethics Board of the Rotman Research Institute at Baycrest Academy for Research and Education.

Story materials and stimuli

In each of the three experiments, participants listened to ~1.5-min stories. Stories were generated using OpenAI's GPT-3.5 (OpenAI et al., 2023) based on researcher-provided prompts and manually reviewed to ensure unique character names, absence of non-English terms, proper use of punctuation, and avoidance of generic openings (e.g., "once upon a time"). Stories were revised manually or rewritten using GPT-3.5 as needed. Each story was on a different topic (e.g., high school reunion, friends on a trip; Panela et al., 2024; Herrmann, 2025c). For each story, GPT-3.5 was also used to generate five comprehension questions in a multiple-choice format, with four options per question. Comprehension questions and answers were manually edited wherever needed to ensure accuracy (Panela et al., 2024; Herrmann, 2025c).

Audio files were synthesized using Google's AI-based speech synthesizer using the male voice en-US-Neural2-J (https://cloud.google.com/text-to-speech/docs/list-voices-and-types). Previous research has demonstrated that AI-generated speech is perceived as highly naturalistic and yields comparable speech-comprehension effects to human speech (Herrmann, 2023b, 2025b).

In each experiment, participants listened to 24 stories presented in 8 blocks (3 stories per block). Stories were presented under one of three listening conditions: clear speech (i.e., speech in quiet) or speech masked by 12-talker babble +6 dB SNR, or -3 dB SNR. 12-talker babble simulates a crowded listening environment while preventing the identification of individual words in the masker (Bilger, 1984; Bilger et al., 1984; Wilson et al., 2012). The +6 dB and -3 dB SNR levels were

chosen to create challenging listening conditions that correspond to approximately 90% and 55% speech intelligibility (Irsik et al., 2022). All stories and story-noise mixtures were normalized to the same root-mean-square (RMS) amplitude and presented at approximately 65 dB SPL.

Acoustic presentation environment

Data collection took place in a sound-attenuating booth. Auditory stimuli were delivered via Sennheiser HD 25-SP II headphones, with audio output controlled by an RME Fireface UCX external sound card. Stimuli were presented using Psychtoolbox (v3.0.14) in MATLAB 2021b (MathWorks Inc.) on a Lenovo T480 laptop running Microsoft Windows 7. Visual stimuli were projected into the booth via screen mirroring.

Electroencephalography (EEG) recording and preprocessing

EEG data were recorded from 32 scalp electrodes and the left and right mastoids using a BioSemi system (Amsterdam, The Netherlands). The online sampling rate was set to 1024 Hz, with an low-pass filter of 208 Hz. Electrodes were referenced online using a monopolar reference feedback loop, which connected a driven passive sensor and a common-mode-sense (CMS) active sensor, both positioned posteriorly on the scalp.

Preprocessing was conducted in MATLAB (MathWorks, Inc.). Data were first notch-filtered to suppress 60 Hz power line noise using an elliptic filter. Signals were then re-referenced to the averaged mastoids to enhance the signal-to-noise ratio for auditory responses at fronto-central electrodes (Ruhnau et al., 2012; Herrmann, 2023a). Additional filtering included a 0.7 Hz high-pass filter (2449 samples, Hann window) and a 22 Hz low-pass filter (211 samples, Kaiser window). EEG data were segmented into time series locked to story onset and downsampled to 512 Hz. Artifact removal was performed using Independent Component Analysis (ICA) in FieldTrip MATLAB toolbox to eliminate blinks, eye movements, and muscle artifacts (Makeig et al., 1995; Oostenveld et al., 2011). After ICA, further artifact rejection was applied by zeroing segments where EEG amplitude varied by more than 80 μV within 0.2 s in any channel (Dmochowski et al., 2012; Cohen and Parra, 2016). Finally, data were low-pass filtered at 10 Hz (251 points, Kaiser window) to focus on low-frequency neural signals most sensitive to acoustic features (Di Liberto et al., 2015; Zuk et al., 2021). The preprocessed EEG data were used for subsequent analyses of neural responses to auditory stimuli.

Calculation of amplitude-onset envelopes

For each clear speech signal, a cochleogram was generated using a simple auditory-periphery model consisting of 30 auditory filters (McDermott and Simoncelli, 2011). The amplitude envelope obtained for each auditory filter was compressed by a factor of 0.6 to simulate inner ear compression (McDermott and Simoncelli, 2011). The amplitude envelopes were then averaged across auditory filters and low-pass filtered using a 40-Hz Butterworth filter. The first derivative was calculated and all negative values were set to zero to obtain the amplitude-onset envelope (Hertrich et al., 2012; Tune et al., 2021; Yasmin et al., 2023; Panela et al., 2024). The amplitude-onset envelope was computed since it elicits strong neural speech tracking (Hertrich et al., 2012). The amplitude-onset envelope was down-sampled to match the EEG sampling rate.

Analysis of neural tracking of speech

To investigate neural speech tracking, we adopted Temporal Response Functions (TRFs) and assessed EEG prediction accuracy (Crosse et al., 2016; Crosse et al., 2021). The TRF is a system identification model that employs regularized linear regression to relate sensory stimulus features with neural responses. Specifically, we utilized the Multivariate Temporal Response Function (mTRF) Toolbox in MATLAB (Crosse et al., 2016; Crosse et al., 2021) to map the amplitude-onset envelope of speech onto EEG signals (forward model). To mitigate overfitting, ridge regularization was applied, with the ridge parameter (λ) set to 10, a value informed by prior research to avoid extreme value selection during cross-validation and optimize computational efficiency (Fiedler et al., 2017; Fiedler et al., 2019; Yasmin et al., 2023; Panela et al., 2024).

For each story, 100 randomly selected 25-s EEG snippets and their corresponding time series of the amplitude-onset envelope were extracted for model training and testing using leave-one-sample-out cross-validation. On each iteration, one snippet's EEG data and the corresponding time series of the amplitude-onset envelope were reserved as the test set, while the remaining, non-overlapping data were used for training. The model was trained using time lags ranging from 0 to 400 ms, employing the *mTRFcrossval* and *mTRFtrain* functions. To assess prediction accuracy, the acoustic feature time series of the test dataset was convolved with the TRF weights to generate a predicted EEG time series, separately for each of the 100 iterations. EEG prediction accuracy was calculated as the Pearson correlation between the predicted and actual EEG data, computing using the *mTRFpredict* function. EEG prediction accuracy was averaged across the leave-one-out iterations.

To investigate the neural-tracking response amplitude at a more nuanced timescale, TRFs for each training dataset were calculated for lags ranging from -0.15–0.55 s. Baseline correction was performed by subtracting the mean signal across -0.15–0 s from the TRF data at each time point. The weights were averaged across the 100 training datasets.

For each participant, TRF weights and EEG prediction accuracy were averaged across stories, separately for the three different speech-clarity conditions and averaged across the frontal-central electrode cluster (F3, F4, Fz, FC1, FC2, Cz, FC5, FC6), which is sensitive to auditory cortical activity (Näätänen and Picton, 1987; Picton et al., 2003).

Time courses of TRF weights exhibit positive and negative deflections (peaks), akin to traditional event-related potentials (Luck, 2014; Crosse et al., 2016; Crosse et al., 2021). To examine response magnitude differences across speech-clarity levels and attention conditions, we focused on the P1-N1 (first major positive peak minus first major negative peak) and P2-N1 amplitude differences. The amplitudes of P1, N1, and P2 components were determined by local extrema (maxima for P1 and P2, minima for N1) within a 100-ms time window centered on each component based on averaged data across all participants in each experiment. Latencies for P1, N1, and P2 were identified as the time points corresponding to these extrema. For each participant, amplitudes were calculated within a 30-ms window centered on the averaged latencies of each component. The P1-N1 and P2-N1 amplitude differences were computed by subtracting N1 amplitudes from P1 and P2, respectively.

Statistical analyses

All statistical analyses were carried out using MATLAB (MathWorks) and JASP software (version 0.19.2). Details about the specific tests used are provided in the relevant sections for each experiment.

Experiment 1: Influences of speech clarity on neural speech tracking

Experiment 1 measured speech-evoked EEG activity under focused listening across speech-clarity conditions to establish the baseline signal changes underlying distraction effects tested in Experiments 2 and 3.

Method and materials

Participants

Twenty-four younger adults participated in Experiment 1 (mean = 25.2 years; 14 females, 10 males).

Stimuli and procedure

Twenty-four stories, each lasting between 109 and 133 seconds, were generated for Experiment 1. The stories were presented under one of three speech-clarity conditions: clear speech (no noise), +6 dB SNR, or -3 dB SNR (12-talker babble). The experiment was divided into 8 blocks, with each block containing 3 stories with the 3 speech-clarity conditions. Assignment of speech-clarity conditions to specific stories was counter-balanced across participants. During each trial, participants listened attentively to the story while their EEG signals were recorded. After each story, participants answered five multiple-choice comprehension questions, each with four response options (chance level = 25%). Additionally, participants provided a self-reported gist rating on a scale from 1 to 9, with higher ratings indicating a better understanding of the story's gist. Gist ratings correlate highly with speech intelligibility as measured using word report scores (Davis and Johnsrude, 2003; Ritz et al., 2022). The exact wording for the gist rating was: “I understood the gist of the story. Please rate this statement independently of how you felt about the other stories”. Gist ratings were linearly transformed to ranging from 0 to 1 in order to facilitate interpretation similar to proportions (Mathiesen et al., 2024; Herrmann, 2025c). Participants underwent a practice block prior to the main experimental procedures.

Analysis

The effect of Speech Clarity (clear, +6 dB, -3 dB SNR) on comprehension accuracy, gist rating, EEG prediction accuracy, P1-N1 amplitude, and P2-N1 amplitude were evaluated in separate one-way repeated-measures analysis of variance (rmANOVA). Post-hoc pairwise comparisons were performed using the Holm correction for multiple comparisons (Holm, 1979).

Results

Both comprehension accuracy and gist ratings (Figure 1A) were affected by speech clarity (comprehension: $F_{2, 46} = 11.986$, $p = 6.5 \cdot 10^{-5}$, $\eta_p^2 = .343$; gist: $F_{2, 46} = 34.466$, $p = 7.1 \cdot 10^{-10}$, $\eta_p^2 = .600$). Performance dropped only at the lowest clarity (-3 dB SNR): pair-wise comparisons showed lower scores at -3 dB than at clear and at +6 dB for both behavioural measures (all $p_{Holm} < .01$). Clear and +6 dB did not differ significantly on either measure (all $p_{Holm} > .05$).

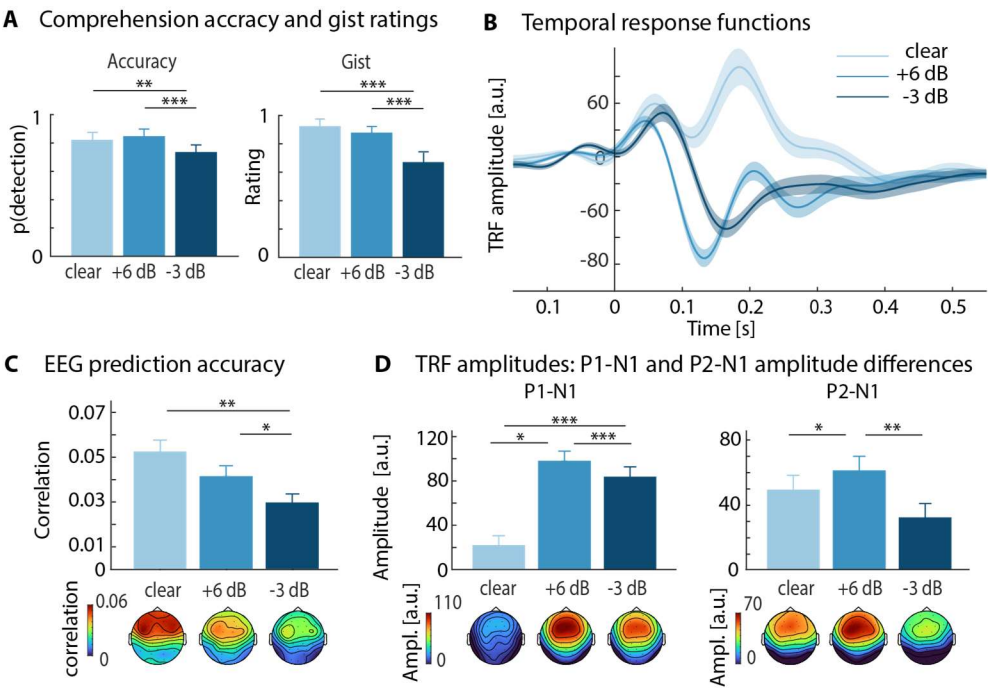

Figure 1: Results for Experiment 1. A: Comprehension scores (left), gist ratings (right) across different speech-clarity conditions. **B:** Temporal response functions (TRF) waveforms. **C:** EEG prediction accuracy and corresponding topographies. **D:** P1-N1 (left) and P2-N1 (right) amplitude and topographies for each speech-clarity condition. Error bars denote the standard error of the mean. *p< .05, **p< .01, ***p< .001 (Holm-corrected where applicable).

EEG prediction accuracy (Figure 1C) showed a significant main effect of speech clarity ($F_{2, 46}$ = 9.376, p = $3.8 \cdot 10^{-4}$, η_p^2 = .269). The values were lower at -3 dB SNR compared to clear (t_{23} = 3.882, p_{Holm} = .002, d = .843) and +6 dB (t_{23} = 2.843, p_{Holm} = .018, d = .471), whereas clear and +6 dB did not differ (t_{23} = 1.966, p_{Holm} = .062, d = .409). Scalp topographies revealed highest EEG prediction accuracy at frontal-central electrodes that are sensitive to auditory cortex activity.

The speech-clarity effect for the P1-N1 amplitude ($F_{2, 46}$ = 52.108, p = $1.5 \cdot 10^{-12}$, η_p^2 = .694) revealed a non-monotonic pattern: responses were largest at +6 dB and remained elevated at -3 dB relative to clear (all p_{Holm} < .05). For the P2-N1 amplitude the effect of speech clarity ($F_{2, 46}$ = 8.110, p = $9.6 \cdot 10^{-4}$, η_p^2 = .261) showed larger amplitudes at +6 dB compared to -3 dB and clear (both p_{Holm} < .05), whereas -3 dB and clear did not differ significantly (p_{Holm} = .110; Figures 1B, D).

In sum, the results of Experiment 1 shows that background babble impairs speech comprehension, gist rating, and overall EEG prediction accuracy, with declines observed only at the lowest speech clarity (-3 dB). Early TRF responses (both P1-N1 and P2-N1 components) were selectively enhanced under moderate noise – peaking at + 6 dB – consistent with recent work suggesting stochastic resonance phenomena in neural speech tracking and for more simple

sounds (Parbery-Clark et al., 2011; Alain et al., 2014; Herrmann, 2025c, a). With the foundation of Experiment 1, Experiment 2 examines how disengagement from listening due to visual distraction affects neural speech tracking, and whether disengagement effects are affected by speech clarity.

Experiment 2: Influences of visual distraction on neural speech tracking

Method and materials

Participants

Twenty-four younger adults participated in Experiment 2 (mean = 26 years; 14 females, 10 males).

Stimuli and procedure

We generated a new set of 24 stories (102–141 s each) for Experiment 2. Similar to Experiment 1, the session comprised 8 blocks, with each block containing 3 stories, one for each speech-clarity condition (clear, +6 dB, and -3 dB SNR). In four blocks, participants attended to the speech (‘attend’ condition), whereas in the other four blocks, participants ignored the speech (‘ignore’ condition). Assignment of speech clarity and attention conditions to specific stories was counter-balanced across participants. Blocks alternated in attention conditions, and the starting attention condition was counter-balanced across participants.

In ‘Attend’ blocks, tasks mirrored those of Experiment 1. That is, participants were instructed to listen attentively to the stories, followed by comprehension questions and a gist rating. In ‘Ignore’ blocks, participants were asked to ignore the stories and instead perform a visual 1-back task. The visual 1-back task utilized images of white digits (0 to 9) on a black background, sourced from the MNIST Handwritten Digit Classification Dataset (Herrmann 2024; Deng et al. 2012). Each digit image was displayed for 0.25 seconds, followed by a 0.25-s black screen before the next image appeared. Visual stimuli thus repeat as a rate of 2 Hz. The digit stream included a digit repetition – albeit a different image sample – every 6 to 12 digits. Participants were asked to press a keyboard button as soon as they detected a repetition. Hit rates and response times were recorded. Additionally, participants rated the degree to which they were able to ignore the speech using a 9-point scale (1 = strongly disagree, 9 = strongly agree; again linearly transformed to range from 0-1). No comprehension questions or gist ratings were used in the ‘Ignore’ blocks to avoid incentivizing participants to listen to the stories.

Participants underwent a practice block prior to the main experimental procedures to familiarize them with the tasks.

Analysis

For each participant, comprehension accuracy and gist ratings from 'attend' blocks were averaged across the four stories per speech-clarity condition (clear, +6 dB, -3 dB). Reaction times, visual detection accuracy, and able-to-ignore ratings from 'ignore' blocks were also averaged across the four stories per speech-clarity condition. For each behavioral measure, we calculated a one-way rmANOVA with the within-participant factor Speech Clarity (clear, +6 dB, -3 dB). Note that behavioral measures did not comprise the Attention factor.

Because Experiment 2 comprised a repetitive visual stream of digits occurring at 2 Hz, this provided an opportunity to neurally investigate attention to the visual stream. To this end, EEG data for each story / visual steam were divided into 7-s epochs (shifted by 0.5 s across the story duration) that were time-locked to a visual stimulus. A fast Fourier transform (FFT) was calculated for each epoch. Inter-trial phase coherence (ITPC; Lachaux et al., 1999) was then calculated across epochs to obtain a measure of neural synchronization with the visual stream. ITPC was used here instead of evoked amplitude because ITPC spectra do not show 1/f patterns that amplitude spectra exhibit (Herrmann et al., 2019; Keitel et al., 2019; Irsik et al., 2021), but which can complicate analyses (Donoghue et al., 2020; Donoghue et al., 2022). Mathematically, ITPC and evoked amplitude are otherwise closely related. Synchronization was calculated as the average across ITPC at the 2-Hz visual repetition rate and its harmonics up to 10 Hz (Irsik et al., 2021). Taking into account the harmonics is needed because data are not sinusoidal in the time domain. ITPC was averaged across 3 occipital channels (O1, Oz, O2) to focus on visual-cortex activity. For the statistical analysis, ITPC was averaged across stories, separately for each speech-clarity condition (clear, +6 dB, -3 dB) and attention condition (Attend, Ignore), and a rmANOVA with the two factors was calculated.

For the neural speech tracking analysis, EEG prediction accuracy, P1-N1 amplitude, and P2-N1 amplitude were evaluated in separate rmANOVAs with the factors Speech Clarity (clear, +6 dB, -3 dB) and Attention (Attend, Ignore). Post-hoc pairwise comparisons were performed using the Holm correction for multiple comparisons (Holm, 1979).

Results

Behavioral performance during 'attend' speech blocks are shown in Figure 2A. Speech Clarity affected both comprehension accuracy ($F_{2, 46} = 11.279$, $p = 1{\cdot}10^{-4}$, $\eta_p^2 = .329$) and gist ratings ($F_{2, 46} = 47.964$, $p = 5.6{\cdot}10^{-12}$, $\eta_p^2 = .676$). Similar to Experiment 1, performance was lower in the -3 dB condition compared to the clear and +6 dB conditions for both measures (both $p_{Holm} < .01$). For gist rating, +6 dB was also lower than clear ($t_{23} = 3.174$, $p_{Holm} = .004$, $d = .444$), whereas for comprehension accuracy, clear and +6 dB did no differ ($t_{23} = .554$, $p_{Holm} = .585$, $d = .103$). During 'ignore' blocks (Figure 2B), speech clarity did not affect visual-detection accuracy ($F_{2, 46} = .506$, $p = .606$, $\eta_p^2 = .022$) or reaction times ($F_{2, 46} = 2.855$, $p = .068$, $\eta_p^2 = .110$). In contrast, 'able-to-ignore' ratings indicate that participants found the stories at -3 dB easier to ignore compared to clear and +6 dB (both $p_{Holm} < .05$; effect of Speech Clarity: $F_{2, 46} = 7.828$, $p = .001$, $\eta_p^2 = .254$).

Consistent with the periodic visual stream, ITPC showed clear peaks at 2 Hz and harmonics (Figure 2C). Peak ITPC (2–10 Hz harmonics averaged; Figure 2D) was higher when participants ignored than attended to the speech ($F_{1, 23} = 78.844$, $p = 6.8{\cdot}10^{-9}$, $\eta_p^2 = .774$). There was no effect of Speech Clarity ($F_{2, 46} = .882$, $p = .421$, $\eta_p^2 = .037$), and no interaction ($F_{2, 46} = 1.676$, $p = .198$, $\eta_p^2 = .068$). These data show that participants indeed attended to the visual digit stream more when they ignored relative when they attended to the speech.

EEG prediction accuracy was lower for -3 dB than at Clear ($t_{23} = 4.137$, $p_{Holm} = 8{\cdot}10^{-4}$, $d = .882$) and +6 dB ($t_{23} = 4.665$, $p_{Holm} = 3.2{\cdot}10^{-4}$, $d = .689$), with no difference between Clear and +6 dB ($t_{23} = 1.001$, $p_{Holm} = .327$, $d = .193$; effect of Speech Clarity: $F_{2, 46} = 12.368$, $p = 5{\cdot}10^{-5}$, $\eta_p^2 = .350$ Figure 2F). EEG prediction accuracy was greater for the attend than ignore condition (effect of Attention: $F_{1, 23} = 7.244$, $p = .013$, $\eta_p^2 = .240$), although the significant interaction ($F_{2, 46} = 3.561$, $p = .036$, $\eta_p^2 = .134$) shows that this difference was specific for the -3 dB SNR condition ($t_{23} = 3.346$, $p_{Holm} = .031$, $d = .703$).

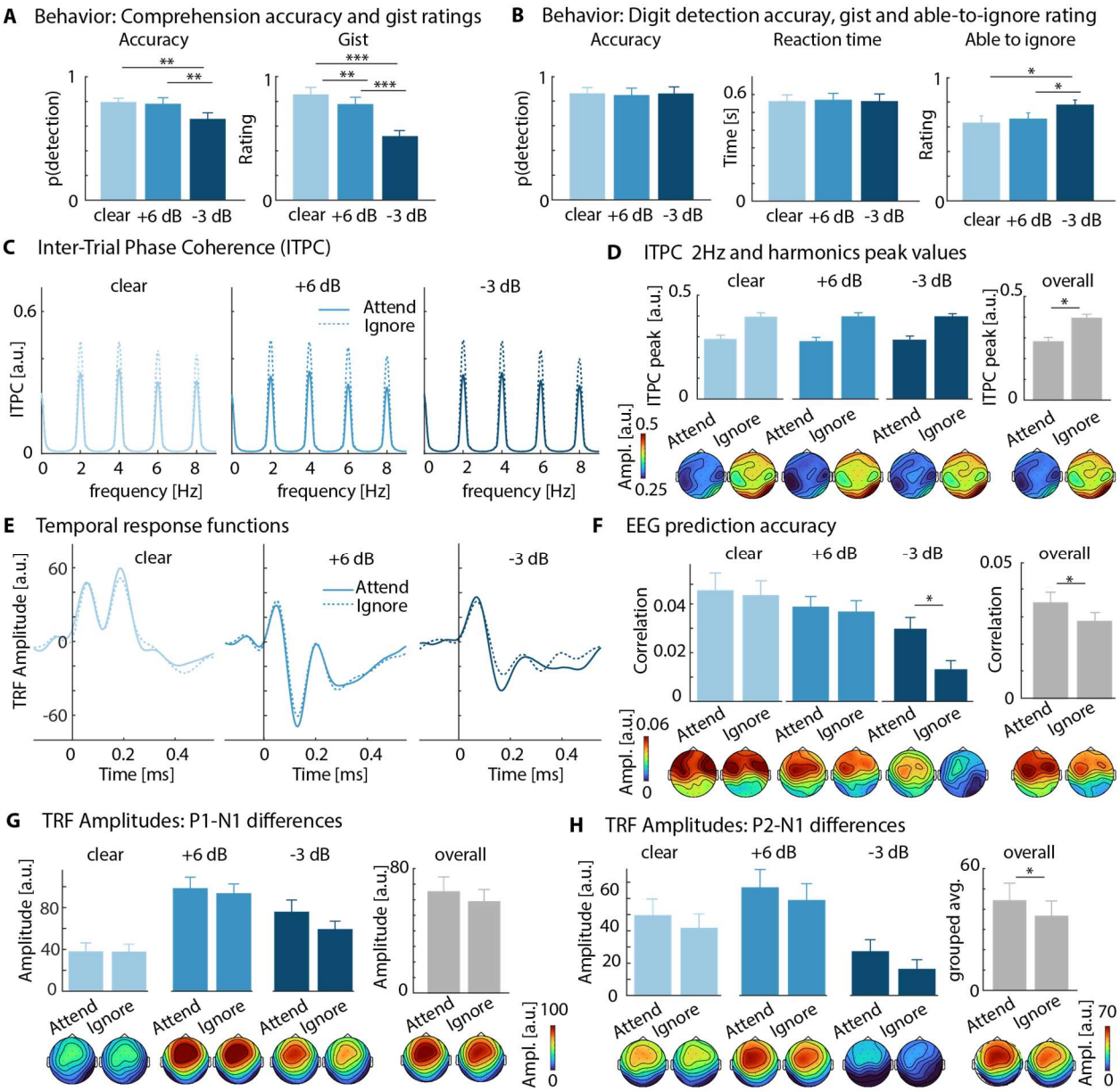

Figure 2: Results for Experiment 2. A: Comprehension scores and gist ratings across three speech-clarity levels during 'attend' blocks. **B:** Detection accuracy and reaction times (RT) for the visual 1-back task, and ratings of the ability to ignore the spoken stories during visual distraction blocks. **C:** ITPC showing peaks at 2 Hz and harmonics, reflecting neural synchronization with the 2-Hz visual stimulus stream. **D:** Peak ITPC (2-10 Hz averaged) for three speech-clarity conditions separated by attention blocks. **E:** Temporal response functions (TRFs) for 'attend' (solid) and 'ignore' (dashed) conditions for the different speech-clarity conditions. **F:** Mean EEG prediction accuracy and corresponding topographies. **G:** Mean P1-N1 amplitudes and corresponding topographies. **H:** Mean P2-N1 amplitudes and corresponding topographies. Error bars denote the standard error of the mean. *p< .05, **p< .01, ***p< .001 (Holm-corrected where applicable).

The P1-N1 amplitude was smallest for clear speech, largest for +6 dB SNR, and in-between for -3 dB (all pairwise $p_{Holm} < .001$; effect of Speech Clarity: $F_{2, 46} = 33.186$, $p = 1.2 \cdot 10^{-9}$, $\eta_p^2 = .591$).

There was no effect of Attention ($F_{2, 46}$ = 3.008, p = .096, η_p^2 = .116), but the interaction between Speech Clarity and Attention was significant ($F_{2, 46}$ = 3.726, p = .032, η_p^2 = .139). Although not surviving Holm correction, the interaction was driven by the attention effect at -3 dB SNR as indicated by the uncorrected comparison (t_{23} = 2.471, p = .021, p_{Holm} = .102, d = .504). Attention effects for other the two speech-clarity levels were not significant (both effects, p_{Holm} and p > .50).

The P2-N1 amplitude was larger at +6 dB than at clear (t_{23} = 2.648, p_{Holm} = .014, d = .382) and -3 dB (t_{23} = 6.342, p_{Holm} = $5.4 \cdot 10^{-6}$, d = .947; effects of Speech Clarity: $F_{2, 46}$ = 19.481, p = $7.4 \cdot 10^{-7}$, η_p^2 = .459). P2-N1 amplitudes were larger when participants attended to the speech compared to when they ignored it (effect of Attention: t_{23} = 2.104, p_{Holm} = .047, d = .194). There was no interaction ($F_{2, 46}$ = 3.008, p = .096, η_p^2 = .116).

In sum, disengagement from speech listening due to visual distraction reduced signatures of neural speech tracking (EEG prediction accuracy and TRF component amplitudes), most prominently for speech masked at -3 dB SNR. This result was mirrored by participants also reporting greater ease ignoring speech at -3 dB than at +6 dB and clear speech. Experiment 2 demonstrates that neural speech tracking is sensitive to disengagement from listening due external distraction, such that the amplitude-envelope of the speech is encoded less strongly. Disengagement from listening in everyday life not only occurs because of external distraction, but also because of internal distraction when a listener's mind is wandering off. Experiment 3 was designed to examine whether disengagement due to internal distractions affect neural speech tracking.

Experiment 3: Influences of internal imagination on neural speech tracking

Method and materials

Participants

Twenty-four younger adults were recruited for Experiment 3 (mean = 26.3 years, SD = 4.37; 13 females, 10 males, 1 non-binary).

Stimuli and procedure

A new set of 24 stories was generated, each lasting between 111 and 139 seconds. The experiment was divided into 8 blocks, , with each block containing 3 stories, one for each speech-clarity condition (clear, +6 dB, and -3 dB SNR). In four blocks, participants attended to the speech

('attend' condition), whereas in the other four blocks, participants ignored the speech ('ignore' condition). Assignment of speech clarity and attention conditions to specific stories was counter-balanced across participants. Blocks with different attention conditions alternated in the experimental session and starting attention condition in the session was counter-balanced across participants. Participants underwent a practice block prior to the main experimental procedures to familiarize them with the tasks.

In the 'attend' blocks, participants attentively listen to the stories, answered comprehension questions, and rating gist understanding (similar to Experiments 1 and 2). In 'ignore' blocks, participants were asked to disregard the spoken stories and engage in an 'imagination' task, where they mentally wandered off. During the imagination task, participants were provided with a verbal instruction outlining a daily scenario accompanied by three concrete cues (e.g., scenario: public transport commute; cues: seats, tapping cards, and bus driver). After a brief preparation period, participants were instructed to immerse themselves in the imagined scenario as soon as the auditory stimulus began. They were encouraged to use internal narratives and mental imagery to facilitate the task, and to freely mentally jump to any other scenarios that comes to mind as much as they wanted. This task aimed to mirror disengagement in daily life where people may think about concrete daily activities or interactions (Smallwood and Andrews-Hanna, 2013; Ibaceta et al., 2024), although here facilitated by task instructions. Following each story, participants evaluated their experience using a 9-point scale (1 = strongly disagree, 9 = strongly agree) across four statements: whether they were able to ignore the speech, mentally wander off, visualize the given scenario, and would be able retell the story. No comprehension questions or gist ratings were used in the 'ignore' blocks to avoid that participants attend to the stories. Ratings were linearly transformed to range from 0-1 to facilitate interpretability (Mathiesen et al., 2024; Herrmann, 2025c).

For each behavioral measure, a one-way rmANOVA with the within-participant factor Speech Clarity was calculated. Behavioral measures did not comprise the Attention factor to avoid that participants are biased towards the irrelevant dimension. For each EEG measure (EEG prediction accuracy, P1-N1 amplitude, and P2-N1 amplitude), a two-way rmANOVA was conducted using the with-participant factors Speech Clarity (clear, +6 dB, and -3 dB SNR) and Attention (Attend, Ignore).

Results

Speech clarity affected both comprehension accuracy ($F_{2, 46} = 20.647$, $p = 4 \cdot 10^{-7}$, $\eta_p^2 = .473$) and gist ratings ($F_{2, 46} = 40.847$, $p = 7.5 \cdot 10^{-11}$, $\eta_p^2 = .637$). Both behavioral measures were lower at -3 dB compared to the clear (both $p_{Holm} < .001$) and +6 dB (both $p_{Holm} < .001$), with no difference between Clear and +6 dB (both $p_{Holm} > .05$; Figure 3A).

Behavioral ratings related to the imagination task are shown in Figure 3B. Participants rated it to be easier to ignore stories (effect of Speech Clarity: $F_{2, 46} = 18.591$, $p = 1.2 \cdot 10^{-6}$, $\eta_p^2 = .447$; $p_{Holm} < .05$ for all pairwise comparisons) and to be less able to retell stories with decreasing speech clarity (effect of Speech Clarity: $F_{2, 46} = 26.878$, $p = 1.9 \cdot 10^{-8}$, $\eta_p^2 = .539$; $p_{Holm} < .05$ for all comparisons). The ability to mentally visualize was rated higher at -3 dB compared to Clear ($t_{23} = 4.236$, $p_{Holm} = 9.4 \cdot 10^{-4}$, $d = .552$) and +6 dB ($t_{23} = 3.636$, $p_{Holm} = .003$, $d = .364$; effect of Speech Clarity: $F_{2, 46} = 11.030$, $p = 1.2 \cdot 10^{-4}$, $\eta_p^2 = .324$), but speech-clarity had no effect on participants' ability to wander off mentally ($F_{2, 46} = .215$, $p = .807$, $\eta_p^2 = .009$). Overall, the ratings suggest that participants could well imagine mentally and ignore the speech, such that they felt relatively unable to retell the story.

EEG prediction accuracy decreased with decreasing speech clarity ($F_{2, 46} = 26.774$, $p = 1.9 \cdot 10^{-8}$, $\eta_p^2 = .538$; $p_{Holm} < .05$ for all pairwise comparisons; Figure 3D). Critically, EEG prediction accuracy was lower during imagination (ignore) than attentive listening (effect of Attention: $F_{1, 23} = 7.405$, $p = .012$, $\eta_p^2 = .244$), with no Speech Interaction × Attention interaction ($F_{2, 46} = .091$, $p = .913$, $\eta_p^2 = .004$).

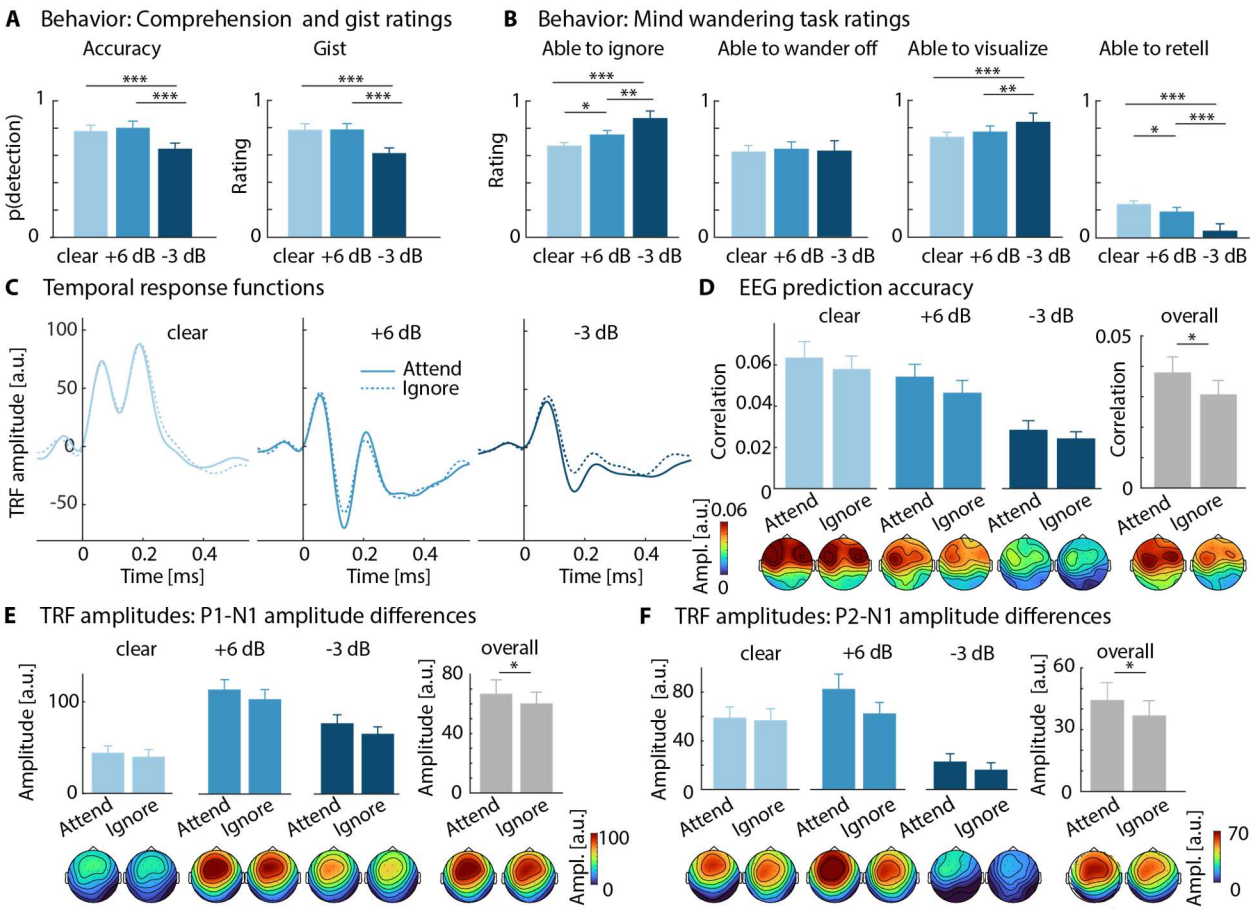

Figure 3: Results for Experiment 3. A: Comprehension scores and gist ratings across three speech-clarity conditions during attentive listening blocks. **B:** Self-evaluations of the ability 'to ignore the spoken stories', 'to be able to wander off', 'to be able to visualize the given scenery', 'to be able to retell the spoken stories' during imagination blocks. **C:** Temporal response functions (TRFs) during Attend (solid) and Ignore (dashed) across Speech Clarity. **D:** EEG prediction accuracy and corresponding topographies. **E:** P1-N1 amplitudes and corresponding topographies. **F:** P2-N1 amplitudes and corresponding topographies. Error bars denote the standard error of the mean. *p< .05, **p< .01, ***p< .001 (Holm-corrected where applicable).

Speech Clarity modulated the P1-N1 amplitudes (main effect of Speech Clarity: $F_{2, 46} = 40.196$, p $8 \cdot 10^{-11}$, $\eta_p^2 = .636$), as amplitudes were largest at +6 dB and smallest in the Clear condition (all pairwise $p_{Holm} < .001$). Mental imagination reduced P1-N1 amplitudes compared to attentive speech listening ($F_{1, 23} = 4.862$, p = .038, $\eta_p^2 = .175$). The interaction was not significant ($F_{2, 46} = .585$, p = .561, $\eta_p^2 = .025$).

P2-N1 amplitudes also varied with Speech Clarity ($F_{2, 46} = 29.536$, $p = 5.6 \cdot 10^{-9}$, $\eta_p^2 = .562$), such that amplitudes were largest at +6 dB and smallest in the Clear condition (all pairwise $p_{Holm} < .05$). Mental imagination reduced P2-N1 amplitude compared to attentive speech listening ($F_{1, 23} = 5.837$, p = .024, $\eta_p^2 = .202$). The interaction was not significant ($F_{2, 46} = 1.516$, p = .230, $\eta_p^2 = .062$).

In sum, Experiment 3 shows, consistent with the findings from Experiments 1 and 2, that background babble can enhance early neural responses to the amplitude-envelope of speech (see also Yasmin et al., 2023; Panela et al., 2024; Herrmann, 2025c). Experiment 3 results further indicate that participants can ignore speech well while engaging in a mental imagination task. Critically, Experiment 3 extends the external distraction effect in Experiment 2 by showing that internal distraction (imagination) reduces the encoding of the amplitude-envelope of speech, indicated by all three signatures of neural speech tracking.

Comparison of attention effects between external and internal distraction

In order to directly compare the attention effects between external and internal distraction, we first calculated an rmANOVA for the participants' ratings of their ability to ignore speech when they performed the visual distraction or internal imagination task. Rating scores were used as dependent measure, with the within-participant factor Speech Clarity (clear, +6 dB, -3 dB) and the between-participants factor Distraction Type (external, internal; Experiments 2 and 3). The rmANOVA showed higher speech ignoring ratings for -3 dB than for the clear (t_{47} = 6.125, p_{Holm} = $5.6 \cdot 10^{-9}$, d = .892) and +6 dB conditions (t_{47} = 4.598, p_{Holm} = $6.7 \cdot 10^{-5}$, d = .600), and for +6 dB than clear speech (t_{47} = 2.553, p_{Holm} = 0.014, d = .292; effect of Speech Clarity: $F_{2, 92}$ = 24.187, p = $3.6 \cdot 10^{-9}$, η_p^2 = .345). However, there was no effect of Distraction Type ($F_{2, 92}$ = .735, p = .482, η_p^2 = .016), nor an interaction ($F_{1, 46}$ = 2.768, p = .103, η_p^2 = .057), indicating that participants were able similarly ignore speech under external and internal distraction.

Next, we calculated an rmANOVA – separately for the three TRF metrics – using the within-participants factors Speech Clarity (clear, +6 dB, -3 dB) and Attention (Attend, Ignore), and the between-participants factor Distraction Type (external, internal; Experiments 2 and 3). If there were any differences in the attention effect between distraction types, we would expect a significant interaction involving Attention and Distraction Type. In no analysis did we observe a significant interaction involving Attention and Distraction Type (for all p > 0.05), suggesting that the attention effects did not differ between the type of distraction. The interaction closest to significance was the Speech Clarity by Attention by Distraction Type interaction for the EEG prediction accuracy ($F_{2, 92}$ = 2.611, p = .079, η_p^2 = .054), which is consistent with the visual differences between distraction types of the attention effect for the -3 dB (Figures 2F and 3D). But even limiting the rmANOVA to the -3 dB condition did not lead to a significant Attention by Distraction Type interaction ($F_{1, 46}$ = 3.325, p = .075, η_p^2 = .067). A Bayesian independent-samples t-test of the attention effect between distraction types revealed a BF10 = 1.091, suggesting ambiguity about a difference.

Discussion

Across three EEG experiments, we investigated how disengagement from listening due to external and internal distractions alter the neural tracking of speech under different levels of background noise. We observed early (<200 ms), enhanced neural responses to the speech envelope for speech masked by background babble compared to speech in quiet (Experiments 1-3), suggesting stochastic facilitation. Critically, neural tracking of the speech envelope was reduced when individuals were distracted by a visual-stimulus stream (Experiment 2) or by internal thought and imagination (Experiment 3). Overall, the current study shows that disengagement from listening – towards external or internal spaces – that many people experience in crowded, noisy situations is associated with reduced neural encoding of speech.

Background babble altered neural speech tracking

Neural speech tracking was assessed using a temporal responses function (TRF) approach (Crosse et al., 2016; Crosse et al., 2021). We focused on different TRF metrics: EEG prediction accuracy reflects a global, cross-validated fit that pools information across time (here 0-400 ms), whereas P1-N1 and P2-N1 amplitudes reflect time-resolved indices of early sensory processing.

Across all three experiments, we observed a decline in EEG prediction accuracy with decreasing speech clarity (Figures 1-3), with the most prominent decline for the -3 dB SNR condition. These results mirror those of story-comprehension accuracy and ratings of gist understanding (Figures 1-3). These results are consistent with previous research showing a close link between EEG prediction accuracy and speech-perception measures (Ding and Simon, 2013; Vanthornhout et al., 2018; Lesenfants et al., 2019; Synigal et al., 2023).

For early responses (P1-N and P2-N1 amplitudes) of the TRF time sources, we observed a different picture. P1-N1 and P2-N2 amplitudes for speech at moderate (+6 dB SNR) and difficult (-3 dB SNR) levels of background noise were increased relative to clear speech. This increase may be surprising, given the reduced speech comprehension in the presence of background noise. However, there is a variety of recent works observing enhanced early speech-tracking responses when speech is presented in babble noise (Yasmin et al., 2023; Panela et al., 2024; Herrmann, 2025c, a). The work is also in line with other research using simpler sounds such as tone bursts, syllables, and high-frequency sound modulations, showing noise-enhanced early sensory responses relative to sounds in the absence of background noise (Alain et al., 2009; Ward et al., 2010; Parbery-Clark et al., 2011; Alain et al., 2012; Alain et al., 2014; Shukla and Bidelman, 2021).

Recent research specifically examined the noise-related enhancement of early neural tracking responses due to background noise (Herrmann, 2025c, a). The work shows that the enhancement does not depend on individuals attending to the speech (consistent with the current study), that it is present for very high SNRs (>30 dB) for which speech is highly intelligible, and that it is strongest for multi-talker babble (but present also for other background noises). The results from this research argue against a listening effort or attentional focus mechanism, but instead suggest stochastic resonance/facilitation (Stocks, 2000; Stein et al., 2005; McDonnell and Abbott, 2009; McDonnell and Ward, 2011) such that noise in the auditory neural pathway introduced by the acoustic background noise increases sensitivity to the speech envelope. That is, EEG signals reflects the synchronized activity of over 10,000 neurons (Niedermeyer and da Silva, 2005), but some neurons may receive insufficient synaptic input to fire when speech in presented in quiet. Noise in the neural pathway can push additional neurons beyond their firing threshold, because the noise increases the stimulus-related neuronal activity for brief periods (Ward et al., 2002; Moss et al., 2004; Ward et al., 2010). But regardless of the specific underlying mechanisms, the current study shows noise-related enhancements in neural speech tracking across all three experiments.

External and internal distraction reduces neural speech tracking

The main purpose of the current study was to examine how disengagement from speech listening due to external and internal distraction affects the neural encoding of speech as indicated by the temporal response function metrics. We show that EEG prediction accuracy, P1-N1 amplitude, and P2-N1 amplitude – and thus the neural representation of the speech envelope – decrease when individuals disengage from listening and attend to an external visual stream or to internal thoughts and imagination.

Participants were able to disengage from listening in both the external and internal distraction tasks. The enhanced inter-trial phase coherence when participants attended to the visual stream compared to attentively listening to the speech is consistent with previous work on the effects of attention on visual-steady state responses (Müller and Hillyard, 2000; Joon Kim et al., 2007; Andersen et al., 2011; Keitel et al., 2017; Keitel et al., 2019). Participants further stated that they could ignore the stories well. During the imagination task, ratings of being able to ignore the story were high and of being able to retell a story were low. Participants also stated that they could visualize mentally well. Subjective reports are common measures of mental attentional drift (Giambra, 1995; Smallwood and Schooler, 2006; Stawarczyk et al., 2011; Weinstein, 2018; Robison et al., 2019; Villena-González and Cosmelli, 2020; Safati et al., 2024), because such

internal changes resist objective assessments. The reported subjective experiences point to mental disengagement from listening during both distraction tasks. We did not ask participants specific comprehension questions about the stories when they attended to the visual stream or internal thoughts, because we wanted to avoid biasing them towards divided attention.

Previous work on attentional disengagement from speech listening has mainly focused on speech in the presence of one competing speaker, finding – consistent with the current results – that the tracking of the speech envelope of a target speaker is reduced when listeners disengage from listening to it (Kerlin et al., 2010; Ding and Simon, 2012; Power et al., 2012; Presacco et al., 2016; Fiedler et al., 2019; Hambrook and Tata, 2019; Fuglsang et al., 2020; Kraus et al., 2021; Tune et al., 2021; Straetmans et al., 2024; Orf et al., 2025). Much fewer works have shown attention effects when individuals attend to visual distractors while ignoring speech (Vanthornhout et al., 2019; for divided attention see Xie et al., 2023), or under ambient speech masking conditions, such as multi-talker babble. Such conditions, however, reflect common scenarios in which listeners disengage in daily life, making them prime targets for understanding altered speech encoding.

Importantly, the current work is first to show reduced speech encoding when individuals disengage from speech listening and engage in internal thought and imagination. This is critical, because internally oriented disengagement from listening is common among older people with hearing difficulties when the listening effort needed to comprehend speech exceeds the available resources and/or perceived conversational benefit (Heffernan et al., 2016; Pichora-Fuller et al., 2016; Herrmann and Johnsrude, 2020). We used an intentional imagination task over a spontaneous mind wandering task as a proof on concept. The processes underlying intentional imagination and spontaneous mind wandering may be similar (Villena-González and Cosmelli, 2020; Safati et al., 2024), although others have argued that dissociating between intentional and unintentional task-unrelated thought may be beneficial (Forster and Lavie, 2009; Seli et al., 2016a; Seli et al., 2016b). Critically, the reduced speech tracking due to internal thought highlights neurophysiological feedback mechanisms, supressing auditory cortex activity in the light of mental imagination.

Our data provide little evidence for differences in attentional suppression of neural speech tracking by external compared to internal distractions. This might suggest that both forms of disengagement converge on a common pathway through which attention decouples from sensory processing. Although visual distraction necessarily involves visual sensory areas that internal thoughts may not, attentional focus to external stimuli and internal imagination likely both involve

pre-frontal and possibly parietal networks that feedback to auditory cortex (Corbetta and Shulman, 2002; Buschman and Miller, 2007; Braga et al., 2013; Puschmann et al., 2024). This, in turn, could result in a similar down-regulation of auditory-cortical gain and reductions in neural speech tracking for both external and internal distractions.

Nevertheless, the were a few surface-level differences between external and internal distractions that may require additional research. Based on behavioral mind-wandering work (Steindorf et al., 2023) and consistent with behavioral ratings in the current study (Figures 2B and 3B), we had expected the effects of attention on neural speech tracking to be greatest under the most challenging listening conditions (i.e., -3 dB SNR). This was observed for some TRF metrics during external visual distraction (e.g., Figure 2F), but not during internal distraction, for which we only observed a main effect of attention, but no interaction with speech clarity. Direct comparisons between distraction types did not show any significant differences, although there was ambiguity as indicated by the Bayes factor about a difference in distraction type at -3 dB SNR. This may potentially indicate that internal thought and imagination can lead to a blanket form of disengagement from the acoustic surroundings, whereas a distracting visual stream leaves more capacity for speech encoding in quiet, but very little when speech is presented in noise. It is worth noting, however, that the current experimental design was not specifically suited for such direct interpretations due to its between-participant nature. This is thus an area for future research.

Conclusions

The current study investigated how disengagement from speech listening due to distractions by external visual stimuli or internal thought and imagination alters the neural encoding of speech under different degrees of background masking babble. We show that background babble enhances early neural responses to the speech envelope, consistent with recent reports of stochastic resonance. Critically, we show that external and internal distractions both lead to reduced neural tracking of the amplitude-onset envelope of speech, pointing to a shared neural pathway that feeds back to and supresses auditory cortex. The current study provides a neurophysiological marker that indexes forms of disengagement from listening that are common among older adults with hearing challenges in their daily lives.

Declaration of conflict of interest

None.

Data availability

The data will be made available to others upon reasonable request.

Acknowledgements

We thank Tiffany Lao and Saba Junaid for their support with stimuli creation and data collection. This study has been supported by BH's Canada Research Chair program (CRC-2023-00383) and BH's Natural Sciences and Engineering Research Council of Canada (Discovery Grant: RGPIN-2021-02602).

References

Aiken SJ, Picton TW (2008) Human Cortical Responses to the Speech Envelope. Ear & Hearing 29:139-157. doi: 10.1097/aud.0b013e31816453dc.

Alain C, McDonald K, Van Roon P (2012) Effects of age and background noise on processing a mistuned harmonic in an otherwise periodic complex sound. Hearing Research 283:126-135. doi: 10.1016/j.heares.2011.10.007.

Alain C, Roye A, Salloum C (2014) Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. Frontiers in Systems Neuroscience 8:Article 8. doi: 10.3389/fnsys.2014.00008.

Alain C, Quan J, McDonald K, Van Roon P (2009) Noise-induced increase in human auditory evoked neuromagnetic fields. European Journal of Neuroscience 30:132-142. doi: 10.1111/j.1460-9568.2009.06792.x.

Andersen SK, Müller MM, Hillyard SA (2011) Tracking the allocation of attention in visual scenes with steady-state evoked potentials. In: Cognitive neuroscience of attention (Posner MI, ed), pp 197-216. USA: The Guilford Press.

Bilger RC (1984) Manual for the clinical use of the revised SPIN Test. Champaign, IL, USA: The University of Illinois.

Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C (1984) Standardization of a test of speech perception in noise. Journal of Speech, Language, and Hearing Research 27:32-48. doi: 10.1044/jshr.2701.32.

Braga RM, Wilson LR, Sharp DJ, Wise RJS, Leech R (2013) Separable networks for top-down attention to auditory non-spatial and visuospatial modalities. NeuroImage 74:77-86. doi: https://doi.org/10.1016/j.neuroimage.2013.02.023.

Buschman TJ, Miller EK (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science 315:1860-1862. doi: 10.1126/science.1138071.

Cohen SS, Parra LC (2016) Memorable audiovisual narratives synchronize sensory and supramodal neural responses. eNeuro 3:e0203. doi: 10.1523/ENEURO.0203-16.2016.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience 3:201-215. doi: 10.1038/nrn755.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. Frontiers in Human Neuroscience 10:604. doi: 10.3389/fnhum.2016.00604.

Crosse MJ, Zuk NJ, Di Liberto GM, Nidiffer AR, Molholm S, Lalor EC (2021) Linear modeling of neurophysiological responses to speech and other continuous stimuli: Methodological considerations for applied research. Frontiers in Neuroscience 15. doi: 10.3389/fnins.2021.705621.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. The Journal of Neuroscience 23:3423-3431. doi: 10.1523/JNEUROSCI.23-08-03423.2003.

Di Liberto Giovanni M, O'Sullivan James A, Lalor Edmund C (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. Current Biology 25:2457-2465. doi: 10.1016/j.cub.2015.08.030.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences 109:11854-11859. doi: 10.1073/pnas.1205381109.

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. The Journal of Neuroscience 33:5728-5735. doi: 10.1523/JNEUROSCI.5297-12.2013.

Ding N, Chatterjee M, Simon JZ (2014) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. NeuroImage 88:41-46. doi: 10.1016/j.neuroimage.2013.10.054.

Dmochowski JP, Sajda P, Dias J, Parra LC (2012) Correlated components of ongoing EEG point to emotionally laden attention – a possible marker of engagement? Frontiers in Human Neuroscience 6:Article 112. doi: 10.3389/fnhum.2012.00112.

Donoghue T, Schaworonkow N, Voytek B (2022) Methodological considerations for studying neural oscillations. European Journal of Neuroscience 55:3502-3527. doi: https://doi.org/10.1111/ejn.15361.

Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, Voytek B (2020) Parameterizing neural power spectra into periodic and aperiodic components. Nature Neuroscience 23:1655-1665. doi: 10.1038/s41593-020-00744-x.

Eastwood JD, Frischen A, Fenske MJ, Smilek D (2012) The Unengaged Mind: Defining Boredom in Terms of Attention. Perspectives on Psychological Science 7:482-495. doi.

Fiedler L, Wöstmann M, Herbst SK, Obleser J (2019) Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. Neuroimage 186:33-42. doi: 10.1016/j.neuroimage.2018.10.057.

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

Forster S, Lavie N (2009) Harnessing the wandering mind: The role of perceptual load. Cognition 111:345-355. doi: https://doi.org/10.1016/j.cognition.2009.02.006.

Fuglsang SA, Märcher-Rørsted J, Dau T, Hjortkjær J (2020) Effects of Sensorineural Hearing Loss on Cortical Synchronization to Competing Speech during Selective Attention. The Journal of Neuroscience 40:2562-2572. doi.

Giambra LM (1995) A Laboratory Method for Investigating Influences on Switching Attention to Task-Unrelated Imagery and Thought. Consciousness and Cognition 4:1-21. doi: https://doi.org/10.1006/ccog.1995.1001.

Hambrook DA, Tata MS (2019) The effects of distractor set-size on neural tracking of attended speech. Brain and Language 190:1-9. doi: https://doi.org/10.1016/j.bandl.2018.12.005.

Heffernan E, Coulson NS, Henshaw H, Barry JG, Ferguson MA (2016) Understanding the psychosocial experiences of adults with mild-moderate hearing loss: An application of Leventhal's self-regulatory model. International Journal of Audiology 55:S3-S12. doi.

Herrmann B (2023a) Neural responses to repeated noise structure in sounds are invariant to temporal interruptions. Auditory Perception & Cognition 6:330-352. doi.

Herrmann B (2023b) The perception of artificial-intelligence (AI) based synthesized speech in younger and older adults. International Journal of Speech Technology 26:395-415. doi: 10.1007/s10772-023-10027-y.

Herrmann B (2025a) Age-related differences in the impact of background noise on neural speech tracking. Neurobiology of Aging 153:10-20. doi: 10.1101/2025.01.09.632269.

Herrmann B (2025b) Language-agnostic, automated assessment of listeners' speech recall using large language models. Trends in Hearing 29:1-22. doi: 10.1177/23312165251347131.

Herrmann B (2025c) Enhanced neural speech tracking through noise indicates stochastic resonance in humans. eLife 13:RP100830. doi: 10.7554/eLife.100830.3.

Herrmann B, Johnsrude IS (2020) A model of listening engagement (MoLE). Hearing Research 397:108016. doi: 10.31234/osf.io/y3cr9.

Herrmann B, Buckland C, Johnsrude IS (2019) Neural signatures of temporal regularity processing in sounds differ between younger and older adults. Neurobiology of Aging 83:73-85. doi: https://doi.org/10.1101/522375

Hertrich I, Dietrich S, Trouvain J, Moos A, Ackermann H (2012) Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49:322-334. doi: 10.1111/j.1469-8986.2011.01314.x.

Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 6:65-70. doi: http://www.jstor.org/stable/4615733.

Ibaceta M, Holman D, Niven K (2024) Mind-wandering contents and characteristics: an exploratory study comparing between work and non-work contexts. European Journal of Work and Organizational Psychology 33:503-516. doi: 10.1080/1359432X.2024.2328373.

Irsik VC, Johnsrude IS, Herrmann B (2022) Neural activity during story listening is synchronized across individuals despite acoustic masking. Journal of Cognitive Neuroscience 34:933-950. doi: 10.1162/jocn_a_01842.

Irsik VC, Almanaseer A, Johnsrude IS, Herrmann B (2021) Cortical responses to the amplitude envelopes of sounds change with age. The Journal of Neuroscience 41:5045-5055. doi: 10.1523/JNEUROSCI.2715-20.2021.

Joon Kim Y, Grabowecky M, Paller KA, Muthu K, Suzuki S (2007) Attention induces synchronization-based response gain in steady-state visual evoked potentials. Nature Neuroscience 10:117-125. doi: 10.1038/nn1821.

Keitel C, Thut G, Gross J (2017) Visual cortex responses reflect temporal structure of continuous quasi-rhythmic sensory stimulation. NeuroImage 146:58-70. doi.

Keitel C, Keitel A, Benwell CSY, Daube C, Thut G, Gross J (2019) Stimulus-Driven Brain Rhythms within the Alpha Band: The Attentional-Modulation Conundrum. The Journal of Neuroscience 39:3119. doi: 10.1523/JNEUROSCI.1633-18.2019.

Kerlin JR, Shahin AJ, Miller LM (2010) Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party”. The Journal of Neuroscience 30:620-628. doi.

Kraus F, Tune S, Ruhe A, Obleser J, Wöstmann M (2021) Unilateral acoustic degradation delays attentional separation of competing speech. Trends in Hearing 25:23312165211013242. doi: 10.1177/23312165211013242.

Lachaux J-P, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring Phase Synchrony in Brain Signals. Human Brain Mapping 8:194-208. doi.

Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European Journal of Neuroscience 31:189-193. doi: 10.1111/j.1460-9568.2009.07055.x.

Lavie N (2005) Distracted and confused?: Selective attention under load. Trends in Cognitive Sciences 9:75-82. doi.

Lavie N, de Fockert J (2005) The role of working memory in attentional capture. Psychonomic Bulletin & Review 12:669-674. doi.

Lesenfants D, Vanthornhout J, Verschueren E, Decruy L, Francart T (2019) Predicting individual speech intelligibility from the cortical tracking of acoustic- and phonetic-level speech representations. Hearing Research 380:1-9. doi: 10.1016/j.heares.2019.05.006.

Luck SJ (2014) An Introduction to the Event-Related Potential Technique. USA: MIT Press.

Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1995) Independent component analysis of electroencephalographic data. In: Advances in Neural Information Processing Systems (Touretzky D, Mozer M, Hasselmo M, eds), pp 145-151. Cambridge, MA, USA: MIT Press.

Mathiesen SL, Van Hedger SC, Irsik VC, Bain MM, Johnsrude IS, Herrmann B (2024) Exploring age differences in absorption and enjoyment during story listening. Psychology International 6:667-684. doi: 10.3390/psycholint6020041.

McDermott Josh H, Simoncelli Eero P (2011) Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. Neuron 71:926-940. doi: 10.1016/j.neuron.2011.06.032.

McDonnell MD, Abbott D (2009) What Is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. PLOS Computational Biology 5:e1000348. doi: 10.1371/journal.pcbi.1000348.

McDonnell MD, Ward LM (2011) The benefits of noise in neural systems: bridging theory and experiment. Nature Reviews Neuroscience 12:415-425. doi: 10.1038/nrn3061.

McNeice Z, Tomlin D, Timmer B, Short CE, Nixon G, Galvin K (2024) A scoping review exploring how adults self-describe and communicate about the listening difficulties they experience. International Journal of Audiology 63:163-170. doi: 10.1080/14992027.2023.2245136.

Moss F, Ward LM, Sannita WG (2004) Stochastic resonance and sensory information processing: a tutorial and review of application. Clinical Neurophysiology 115:267-281. doi: 10.1016/j.clinph.2003.09.014.

Motala A, Johnsrude IS, Herrmann B (2024) A longitudinal framework to describe the relation between age-related hearing loss and social isolation. Trends in Hearing 28:1-18. doi: 10.1177/23312165241236041.

Müller MM, Hillyard S (2000) Concurrent recording of steady-state and transient event-related potentials as indices of visual-spatial selective attention. Clinical Neurophysiology 111:1544-1552. doi: https://doi.org/10.1016/S1388-2457(00)00371-0.

Näätänen R, Picton TW (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375-425. doi: 10.1111/j.1469-8986.1987.tb00311.x.

Niedermeyer E, da Silva FHL (2005) Electroencephalography: Basic principles, clinical applications, and related fields: Lippincott Williams & Wilkins.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience 2011:Article ID 156869. doi: 10.1155/2011/156869.

OpenAI et al. (2023) GPT-4 Technical Report. arXiv. doi: 10.48550/arXiv.2303.08774.

Orf M, Hannemann R, Obleser J (2025) Does amplitude compression help or hinder attentional neural speech tracking? The Journal of Neuroscience 45:e0238242024. doi: 10.1523/JNEUROSCI.0238-24.2024.

Panela RA, Copelli F, Herrmann B (2024) Reliability and generalizability of neural speech tracking in younger and older adults. Neurobiology of Aging 134:165-180. doi: 10.1016/j.neurobiolaging.2023.11.007.

Parbery-Clark A, Marmel F, Bair J, Kraus N (2011) What subcortical–cortical relationships tell us about processing speech in noise. European Journal of Neuroscience 33:549-557. doi: 10.1111/j.1460-9568.2010.07546.x.

Peelle JE (2018) Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. Ear & Hearing 39:204-214. doi: 10.1097/aud.0000000000000494.

Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips NA, Richter M, Rudner M, Sommers MS, Tremblay KL, Wingfield A (2016) Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). Ear & Hearing 37 Suppl 1:5S-27S. doi: 10.1097/aud.0000000000000312.

Picton TW, John SM, Dimitrijevic A, Purcell DW (2003) Human auditory steady-state responses. International Journal of Audiology 42:177-219. doi: 10.3109/14992020309101316.

Power AJ, Foxe JJ, Forde E-J, Reilly RB, Lalor EC (2012) At what time is the cocktail party? A late locus of selective attention to natural speech. European Journal of Neuroscience 35:1497-1503. doi: https://doi.org/10.1111/j.1460-9568.2012.08060.x.

Presacco A, Simon JZ, Anderson S (2016) Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. Journal of Neurophysiology 116:2346-2355. doi: 10.1152/jn.00372.2016.

Puschmann S, Regev M, Fakhar K, Zatorre RJ, Thiel CM (2024) Attention-Driven Modulation of Auditory Cortex Activity during Selective Listening in a Multispeaker Setting. The Journal of Neuroscience 44:e1157232023. doi: 10.1523/JNEUROSCI.1157-23.2023.

Ritz H, Wild CJ, Johnsrude IS (2022) Parametric cognitive load reveals hidden costs in the neural processing of perfectly intelligible degraded speech. The Journal of Neuroscience 42:4619-4628. doi: 10.1523/JNEUROSCI.1777-21.2022.

Robison MK, Miller AL, Unsworth N (2019) Examining the effects of probe frequency, response options, and framing within the thought-probe method. Behavior Research Methods 51:398-408. doi: 10.3758/s13428-019-01212-6.

Ruhnau P, Herrmann B, Schröger E (2012) Finding the right control: The mismatch negativity under investigation. Clinical Neurophysiology 123:507-512. doi: 10.1016/j.clinph.2011.07.035.

Safati AB, Carr TH, Lowe CJ, Smilek D (2024) Mind wandering on command. Frontiers in Psychology Volume 15 - 2024. doi.

Sarant J, Lemke U, Giroud N, Scherpiet S, Weinstein B (2024) Promoting hearing and cognitive health in audiologic rehabilitation for the well-being of older adults. International Journal of Audiology 63:761-771. doi: 10.1080/14992027.2023.2260099.

Seli P, Risko EF, Smilek D (2016a) On the Necessity of Distinguishing Between Unintentional and Intentional Mind Wandering. Psychological Science 27:685-691. doi: 10.1177/0956797616634068.

Seli P, Risko EF, Smilek D, Schacter DL (2016b) Mind-Wandering With and Without Intention. Trends in Cognitive Sciences 20:605-617. doi: 10.1016/j.tics.2016.05.010.

Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends in Cognitive Sciences 12:182-186. doi.

Shukla B, Bidelman GM (2021) Enhanced brainstem phase-locking in low-level noise reveals stochastic resonance in the frequency-following response (FFR). Brain Research 1771:147643. doi: 10.1016/j.brainres.2021.147643.

Smallwood J, Schooler JW (2006) The Restless Mind. Psychological Bulletin 132:946-958. doi.

Smallwood J, Andrews-Hanna J (2013) Not all minds that wander are lost: the importance of a balanced perspective on the mind-wandering state. Frontiers in Psychology Volume 4 - 2013. doi.

Stawarczyk D, Majerus S, Maj M, Van der Linden M, D'Argembeau A (2011) Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. Acta Psychologica 136:370-381. doi: https://doi.org/10.1016/j.actpsy.2011.01.002.

Stein RB, Gossen ER, Jones KE (2005) Neuronal variability: noise or part of the signal? Nature Reviews Neuroscience 6:389-397. doi: 10.1038/nrn1668.

Steindorf L, Pink S, Rummel J, Smallwood J (2023) When there is noise on Sherlock Holmes: mind wandering increases with perceptual processing difficulty during reading and listening. Cognitive Research: Principles and Implications 8:31. doi.

Stocks NG (2000) Suprathreshold stochastic resonance in multilevel threshold systems. Physical Review Letters 84:2310-2313. doi: 10.1103/PhysRevLett.84.2310.

Straetmans L, Adiloglu K, Debener S (2024) Neural speech tracking and auditory attention decoding in everyday life. Frontiers in Human Neuroscience Volume 18 - 2024. doi.

Synigal SR, Anderson AJ, Lalor EC (2023) Electrophysiological indices of hierarchical speech processing differentially reflect the comprehension of speech in noise. BioRxiv. doi: 10.1101/2023.03.30.534927.

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

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

Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T (2018) Speech intelligibility predicted from neural entrainment of the speech envelope. Journal of the Association for Research in Otolaryngology 19:181-191. doi: 10.1007/s10162-018-0654-z.

Villena-González M, Cosmelli D (2020) Chapter 5 - Imagination and mind wandering: two sides of the same coin? A brain dynamics perspective. In: Creativity and the Wandering Mind (Preiss DD, Cosmelli D, Kaufman JC, eds), pp 93-120: Academic Press.

Ward LM, Neiman A, Moss F (2002) Stochastic resonance in psychophysics and in animal behavior. Biological Cybernetics 87:91-101. doi: 10.1007/s00422-002-0328-z.

Ward LM, MacLean SE, Kirschner A (2010) Stochastic resonance modulates neural synchronization within and between cortical sources. PLoS ONE 5:e14371. doi: 10.1371/journal.pone.0014371.

Weinstein Y (2018) Mind-wandering, how do I measure thee with probes? Let me count the ways. Behavior Research Methods 50:642-661. doi: 10.3758/s13428-017-0891-9.

Wilson RH, McArdle RA, Watts KL, Smith SL (2012) The revised speech perception in noise test (R-SPIN) in a multiple signal-to-noise ratio paradigm. Journal of the American Academy of Audiology 23:590-605. doi: 10.3766/jaaa.23.7.9.

Xie Z, Brodbeck C, Chandrasekaran B (2023) Cortical Tracking of Continuous Speech Under Bimodal Divided Attention. Neurobiology of Language 4:318-343. doi: 10.1162/nol_a_00100.

Yasmin S, Irsik VC, Johnsrude IS, Herrmann B (2023) The effects of speech masking on neural tracking of acoustic and semantic features of natural speech. Neuropsychologia 186:108584. doi: 10.1016/j.neuropsychologia.2023.108584.

Zuk NJ, Murphy JW, Reilly RB, Lalor EC (2021) Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies. PLOS Computational Biology 17:e1009358. doi: 10.1371/journal.pcbi.1009358.