



## OPEN ACCESS

## EDITED BY

Celine Richard,  
St. Jude Children's Research Hospital,  
United States

## REVIEWED BY

Brittany Jaekel,  
Starkey Hearing Technologies,  
United States  
Tom Gajęcki,  
Hannover Medical School, Germany

## \*CORRESPONDENCE

Andrew Dimitrijevic  
andrew.dimitrijevic@sunnybrook.ca

## SPECIALTY SECTION

This article was submitted to  
Sensory Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

RECEIVED 13 September 2022

ACCEPTED 21 October 2022

PUBLISHED 07 November 2022

## CITATION

Xiu B, Paul BT, Chen JM, Le TN, Lin VY  
and Dimitrijevic A (2022) Neural  
responses to naturalistic audiovisual  
speech are related to listening  
demand in cochlear implant users.  
*Front. Hum. Neurosci.* 16:1043499.  
doi: 10.3389/fnhum.2022.1043499

## COPYRIGHT

© 2022 Xiu, Paul, Chen, Le, Lin and  
Dimitrijevic. This is an open-access  
article distributed under the terms of  
the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution  
or reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Neural responses to naturalistic audiovisual speech are related to listening demand in cochlear implant users

Bowen Xiu<sup>1</sup>, Brandon T. Paul<sup>2</sup>, Joseph M. Chen<sup>3,4</sup>,  
Trung N. Le<sup>3,4</sup>, Vincent Y. Lin<sup>3,4</sup> and Andrew Dimitrijevic<sup>1,3,4\*</sup>

<sup>1</sup>Evaluative Clinical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, Canada,

<sup>2</sup>Department of Psychology, Toronto Metropolitan University, Toronto, ON, Canada, <sup>3</sup>Department of Otolaryngology – Head and Neck Surgery, Sunnybrook Health Sciences Centre, Toronto, ON, Canada, <sup>4</sup>Department of Otolaryngology – Head and Neck Surgery, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

There is a weak relationship between clinical and self-reported speech perception outcomes in cochlear implant (CI) listeners. Such poor correspondence may be due to differences in clinical and “real-world” listening environments and stimuli. Speech in the real world is often accompanied by visual cues, background environmental noise, and is generally in a conversational context, all factors that could affect listening demand. Thus, our objectives were to determine if brain responses to naturalistic speech could index speech perception and listening demand in CI users. Accordingly, we recorded high-density electroencephalogram (EEG) while CI users listened/watched a naturalistic stimulus (i.e., the television show, “The Office”). We used continuous EEG to quantify “speech neural tracking” (i.e., TRFs, temporal response functions) to the show’s soundtrack and 8–12 Hz (alpha) brain rhythms commonly related to listening effort. Background noise at three different signal-to-noise ratios (SNRs), +5, +10, and +15 dB were presented to vary the difficulty of following the television show, mimicking a natural noisy environment. The task also included an audio-only (no video) condition. After each condition, participants subjectively rated listening demand and the degree of words and conversations they felt they understood. Fifteen CI users reported progressively higher degrees of listening demand and less words and conversation with increasing background noise. Listening demand and conversation understanding in the audio-only condition was comparable to that of the highest noise condition (+5 dB). Increasing background noise affected speech neural tracking at a group level, in addition to eliciting strong individual differences. Mixed effect modeling showed that listening demand and conversation understanding were correlated to early cortical speech tracking, such that high demand and low conversation understanding occurred with lower amplitude TRFs. In the high noise condition, greater listening demand was negatively correlated to parietal alpha power, where higher demand was related to lower alpha power. No significant correlations were observed between TRF/alpha and

clinical speech perception scores. These results are similar to previous findings showing little relationship between clinical speech perception and quality-of-life in CI users. However, physiological responses to complex natural speech may provide an objective measure of aspects of quality-of-life measures like self-perceived listening demand.

#### KEYWORDS

speech tracking, EEG, listening in noise, attention, temporal response function, movies, naturalistic stimuli, cochlear implant

## Introduction

Cochlear implants (CI) can successfully restore hearing for many individuals with profound-to-severe sensorineural hearing loss. Despite increases in post-implantation speech perception scores and quality-of-life (QoL) for most CI users (e.g., [Dillon et al., 2018](#); [Fabie et al., 2018](#)), not all individuals achieve a favorable level of speech performance or QoL (e.g., [Holden et al., 2013](#); [Capretta and Moberly, 2016](#)). Studies examining post-implantation outcomes have focused on etiology of hearing loss, duration of deafness, age at implantation, the amount of residual hearing, and device differences (e.g., [Blamey et al., 2012](#); [Lazard et al., 2012](#); [James et al., 2019](#); [Kurz et al., 2019](#)). However, these factors only account for 10–22% of the outcome variability ([Blamey et al., 2012](#); [Lazard et al., 2012](#)). Furthermore, the correlation between clinical speech perception test and subjective QoL questionnaire outcomes appears to be weak to moderate at best, with some studies reporting no significant relationship (e.g., [Ramakers et al., 2017](#); [McRackan et al., 2018](#); [Thompson et al., 2020](#)). Accordingly, QoL does not necessarily improve after cochlear implantation even if a desired level of speech understanding is reached. These observations motivated the development of clinical testing procedures that are sensitive to post-implant changes to QoL.

Clinical speech tests may fail to reliably predict QoL changes because current clinical testing materials do not fully capture aspects of real-world listening. Clinical tests take place with participants centered in a sound-attenuated booth with minimal reverberation where speech and background noise stimuli are presented at fixed, equidistant positions around the azimuth ([Dunn et al., 2010](#); [Minimum Speech Test Battery \(MSTB\) For Adult Cochlear Implant Users, 2011](#); [Spahr et al., 2012](#)). Additionally, the properties of speech and background noise stimuli themselves are often constant in terms of loudness and content. In contrast, listening in everyday life often occurs in dynamic complex environments where speech is often accompanied by fluctuating background noise, competing speech sources, and with varying access to visual (speech) input that influences speech comprehension.

CI users often report that visual articulatory cues (i.e., lip, tongue, and teeth movements that accompany speech) are missing from clinical testing ([Dorman et al., 2016](#)). Visual articulatory cues are correlated with properties of the auditory speech signal and provide information that assists CI users in interpreting ambiguous speech, as demonstrated through the McGurk effect ([Rouger et al., 2008](#); [Stropahl et al., 2017](#)). Results from neuroimaging studies suggest that visual articulatory cues increase selective attention, speech intelligibility, and neural speech tracking (the entrainment of the neural response to the attended speech stimuli) compared to audio-only speech-in-noise tasks, especially in individuals with hearing loss ([Zion Golumbic et al., 2013](#); [O'Sullivan et al., 2019](#); [Puschmann et al., 2019](#)). On the other hand, background noise and reverberation negatively affects word recognition and listening effort ([Picou et al., 2016](#)). The quality of neural speech tracking has also been shown to decrease in the presence of noise or competing sound sources; background noise has been demonstrated to attenuate the brain response around 100 ms after a change in the speech envelope amplitude ([Ding and Simon, 2012](#); [Fuglsang et al., 2017](#); [Petersen et al., 2017](#)).

In these instances, the burden of degraded speech and environmental distractions can increase the overall cognitive demand required to adequately comprehend speech ([Rönnberg et al., 2010](#); [Picou et al., 2016](#)). Listening to speech in noise can impose a strong demand on cognitive resources due to increased engagement of working memory ([Rönnberg et al., 2010, 2013](#)) and attention processes ([Shinn-Cunningham and Best, 2008](#)). In complex listening environments where the sound quality of speech is reduced, listeners must apportion their limited cognitive resources toward attending to the targeted speech source, in addition to processing and storing auditory information ([Rönnberg et al., 2013](#); [Pichora-Fuller et al., 2016](#); [Peelle, 2018](#)). This is especially concerning for individuals with hearing loss who on average report significantly increased listening effort and fatigue as scored through subjective measures and indirect measures of brain activity ([McGarrigle et al., 2014](#)). Along with the innate loss of spectral resolution due to the CI processor, the demand of comprehending degraded speech places additional strain on the cognitive resources of

CI users and can lead to decreased motivation and effort in engaging in social situations (Pichora-Fuller et al., 2016; Castellanos et al., 2018). Perceived cognitive demand and listening effort has been found to be a primary factor that is detrimental to QoL amongst individuals with hearing loss, especially in CI users (Hughes et al., 2018; Holman et al., 2019).

Current understanding of neural processes underlying the management of cognitive resources when listening in complex environments is limited. Including neural speech tracking, various methods have been devised to measure cognitive demand at a neural level. Alpha power (power of neural oscillations between 8 and 12 Hz) as a potential neural marker of effort and demand has been studied extensively (Foxy and Snyder, 2011; Klimesch, 2012; Obleser et al., 2012; Strauß et al., 2014; Wostmann et al., 2015; Petersen et al., 2015; Dimitrijevic et al., 2019; Magosso et al., 2019; Price et al., 2019; Hauswald et al., 2020; Hjortkjær et al., 2020; Paul et al., 2021). Increased alpha power in adverse listening conditions is hypothesized to represent the suppression of neural responses in brain regions unrelated to the task at hand (Payne and Sekuler, 2014; Strauß et al., 2014; Wilsch et al., 2015) and may reflect a conscious effort to ignore task-irrelevant stimuli (Petersen et al., 2015). Indeed, an increase in parietal alpha power has been observed during working memory tasks in individuals with hearing loss when both background noise and task difficulty increases (Petersen et al., 2015; Paul et al., 2021). A positive relationship between listening effort and alpha power in the left inferior frontal gyrus (IFG) was also observed in CI users performing a digits speech-in-noise task, suggesting that language networks are involved in self-perceived listening effort (Dimitrijevic et al., 2019). When combined with visual stimuli, an increase in alpha power is seen in the parieto-occipital region when individuals attend to audio that is incongruent to the presented visual cues, suggesting a suppression of the unrelated incoming visual stimuli (O'Sullivan et al., 2019). Therefore, alpha power potentially reflects a gating mechanism toward incoming sensory stimuli in a manner relevant to task goals and may serve as a marker of cognitive demand during listening in noisy environments.

While previous literature links alpha power and speech tracking to cognitive demand, demand appears to affect speech tracking, cognitive engagement, and alpha power differently, with speech envelope coherence demonstrating an inverted-U pattern while alpha power declines overall (Hauswald et al., 2020). Furthermore, it is unclear whether neural speech tracking and alpha power reflect self-perceived cognitive demand compared to current clinical speech perception measures. Therefore, the current objectives of the study are to first measure neural speech tracking in CI users using real-world speech, and then evaluate the effects of increasing background noise and the addition of visual cues on the quality of neural speech tracking. Then, we will quantify the relationship of neural speech tracking to the subjective self-reported level of cognitive demand, self-perceived word understanding, and self-perceived conversation understanding. Self-perceived

word understanding and conversation understanding will be measured separately in the planned analyses, since the number of perceived words may not always provide enough context to understand a conversation. Based on previous research regarding effort-related alpha oscillations, we conducted a secondary analysis of alpha band power over the superior-parietal and left IFG regions. We hypothesize that:

1. Self-rated cognitive demand during speech listening will decrease with the presence of visual cues but increase in response to background noise levels.
2. The degree of speech tracking will be reduced by increasing background noise levels and enhanced by the presence of visual cues. Similarly, the presence of visual cues will result in a decrease in alpha power.
3. Neural speech tracking and alpha power will be related to subjective self-reports of cognitive/mental demands and words/conversation understanding.

## Materials and methods

### Participants

Participant inclusion criteria consisted of adult CI users between the ages of 18–80 that have native or bilingual fluency in English, with at least 1 year of experience with their implants. As the current objective is to quantify the relationship between clinical speech perception tests and self-perceived cognitive demand in a general CI user population, potential participants were not excluded based on etiology of hearing loss, age of hearing loss onset, performance on clinical speech perception tests, or sidedness of implantation (Table 1). Eighteen adult CI users were recruited from Sunnybrook Health Sciences Centre Department of Otolaryngology. Three participants were excluded due to poor electroencephalogram (EEG) recording quality and unavailable clinical hearing scores, leaving a final participant sample of fifteen subjects (9 males, 6 females). The remaining 15 participants were between 36 and 74 years of age ( $M = 59.1$ ,  $SD = 12.0$ ; Table 1). The participant sample varied in terms of device implantation, and included three unilateral, six bimodal (hearing aid and CI), and six bilateral CI users (Table 1). Unilateral and bimodal users were on average 55.2 years old ( $SD = 12.4$ ) on the date of implant activation, while the average age of bilateral CI users on the implant activation date for the ear they used for the study was 50.5 years old ( $SD = 13.5$ ). All methods and protocols used in the current study were approved by the Research Ethics Board at Sunnybrook Health Sciences Centre (REB #474-2016) in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. All participants provided written informed consent and received monetary compensation and full reimbursement for parking at the hospital campus for their time and participation.

TABLE 1 Demographics of cochlear implant listeners included in the current study.

ID	Age	Sex	Implant side	Ear used	CI use (year)	CI brand	Processor	Strategy	Etiology
CI01	62	F	Right	Right	5	Advanced Bionics	Harmony	HiRes Optima S	Unknown
CI02	36	M	Bilateral	Right	6	MED-EL	SONNET	FS4	Cogan's Syndrome
CI03*	53	F	Right*	Right	8	MED-EL	OPUS 2	FS4-P	Unknown
CI04*	62	M	Left*	Left	2	Advanced Bionics	Naída Q90	HiRes Optima S	Hereditary
CI05*	74	F	Right*	Right	5	MED-EL	OPUS 2	FS4-P	Hereditary
CI06	63	M	Bilateral	Left	4	Advanced Bionics	Naída Q90	HiRes Optima S	Meniere's disease
CI07	62	M	Bilateral	Left	2	MED-EL	SONNET	FS4-P	Unknown
CI08	60	M	Bilateral	Left	16	Advanced Bionics	Naída Q70	Unknown	Hereditary
CI09*	44	F	Left*	Left	1	MED-EL	SONNET	FS4-P	Unknown
CI10	71	M	Bilateral	Right	5	MED-EL	OPUS 2	FS4-P	Unknown
CI11	60	F	Bilateral	Right	11	Cochlear	Freedom	ACE	Congenital
CI12	38	F	Left	Left	1	MED-EL	SONNET	FS4-P	Meningitis
CI13	73	M	Left	Left	8	MED-EL	RONDO	FSP	Meniere's disease
CI14*	72	M	Left*	Left	1	MED-EL	SONNET	FS4-P	Unknown
CI15*	56	M	Right*	Right	6	MED-EL	SONNET	Unknown	Unknown

\*Participant used a hearing aid on the non-implanted side in daily life.

## Clinical speech perception scores

Clinical speech perception was evaluated using the AzBio Sentence Test (Spahr et al., 2012). During testing, the 20-sentence lists were presented in quiet, and in noise at an SNR of +5 dB. Testing took place during clinical visits and were administered by audiologists as needed. During testing, participants were seated in a sound-isolated room with minimal reverberation, 1 m away from a loudspeaker positioned directly in front of the participant at approximately the level of a typical listener's head (Minimum Speech Test Battery (MSTB) For Adult Cochlear Implant Users, 2011). Participants were required to repeat back the sentence heard, and each sentence was scored as the number of words correctly reported. Percent scores are calculated for each sentence based on the percentage of words correct within the presented sentence. Sentence scores were averaged and used as the estimate of speech perception ability within the listening situation. For the current study, the most recent post-implantation test scores relative to the study date for the tested ear of participant (or both ears if individual ear scores were unavailable) were collected (Table 2). The average AzBio scores were 86.9 for listening in quiet ( $SD = 8.51$ ), and 60.2 for listening in noise ( $SD = 26.6$ ).

## Experimental setup

### Listening environment

The continuous listening paradigm took place in a sound-attenuated and electrically shielded booth with minimal reverberation. Participants were seated at the center of the booth, surrounded by a circular ring array of eight speakers

located at  $0^\circ$ ,  $\pm 45^\circ$ ,  $\pm 90^\circ$ ,  $\pm 135^\circ$ , and  $180^\circ$ . The speaker ring array was raised 1 m from the ground, with each speaker cone located 0.80 m away from the center of the ring. A computer monitor was placed 0.80 m directly in front of the listener's head, under the speaker located at  $0^\circ$  azimuth. Audio stimuli were played from the loudspeaker at  $0^\circ$  azimuth located directly in front of the listener, while multi-talker background noise was played from surrounding speakers.

### Audiovisual stimuli

The stimuli used for the paradigm were 15-min audio and video segments of episodes one to four from the first season of *The Office* television show. The show was chosen as it features naturalistic dialogue with little to no music, as well as its use in previous neuroimaging studies (e.g., Byrge et al., 2015; Pantelis et al., 2015). In addition, the opening sequence containing the theme song of the show was removed to minimize non-speech sounds. To investigate cognitive demand in complex listening conditions, segments of the show were presented in conditions that varied in multi-talker babble noise levels. The babble noise soundtrack used for the current study was adapted from the four-talker babble noise used in the Quick Speech in Noise (QuickSIN) test (Killion et al., 2004).

Prior to the start of the paradigm, bilateral and bimodal CI participants were asked to remove the cochlear implant or hearing aid that was contralateral to the ear chosen for testing. In the case of bilateral participants, the ear chosen for testing was the ear that was first implanted with a CI. Participants were asked to pay attention to the television show, attending to the audio and video stimuli presented from the speaker and monitor located directly in front of the participant at the  $0^\circ$  azimuth. While participants attended to the target stimuli, babble noise was simultaneously presented from the seven

TABLE 2 AzBio performance of CI listeners in quiet and in noise (SNR + 5 dB) for the ear used during listening task.

ID	Listening side	Tested ear	Score in quiet (%)	Score in noise (%)
CI01	Right	Right	88	22
CI02	Right	Right	97	95
CI03*	Right	Right	74	29
CI04*	Left	Left	71	20
CI05*	Right	Both	92	50
CI06	Left	Left	95	56
CI07	Left	Both	80	31
CI08	Left	Left	90	69
CI09*	Left	Left	79	60
CI10	Right	Right	95	95
CI11	Right	Right	83	70
CI12	Left	Left	93	94
CI13	Left	Left	79	49
CI14*	Left	Left	91	74
CI15*	Right	Right	96	89

\*Participant used a hearing aid on the non-implanted side in daily life. The side used during the study is listed as Listening Side. Listening scores for both ears are reported for two participants as their scores for individual ears were not available.

speakers located at  $\pm 45^\circ$ ,  $\pm 90^\circ$ ,  $\pm 135^\circ$ , and  $180^\circ$  azimuth. Each episode was presented in blocks of 5-min segments, with each segment varying in background noise level. Audio-video segments were presented at 65 dB SPL, while the babble noise was varied to achieve the desired signal-to-noise ratio (SNR). In the Low-Noise condition, the SNR of the audio-video segment to the background babble noise was + 15 dB. Audio-video segments were presented in the Moderate-Noise and High-Noise conditions at +10 and +5 dB SNR, respectively. Finally, to account for the benefits of visual cues for speech tracking, segments of only audio were presented in the Audio-Only condition at SNR +15 dB, with the video being replaced by a visual crosshair that served as a visual fixation point. Three runs for each condition were performed for a total of 15-min per condition, with the order of runs and conditions randomized for each participant to avoid bias. The presentation of video and audio segments were processed in MATLAB 2009b (The MathWorks, Natick, MA, USA) and controlled by a Tucker Davis Technologies (TDT, Alachua, FL, USA) RX8 Processor. After each episode, participants were asked to answer nine simple open-response questions on the events that occurred in the episode to ensure that they attended to the stimuli, averaging to 79.63% correct ( $SD = 19.53\%$ ).

### Assessment of self-perceived listening effort

In order to measure self-reported listening effort, participants responded to the first section of the NASA-Task Load Index (Hart and Staveland, 1988) after each 5-min run. The NASA-Task Load Index (NASA-TLX) is a multi-scale self-report that rates an individual's perceived workload for a task (Hart and Staveland, 1988). The dimensions consist

of six ordinal subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. For the current study, the scales for Physical Demand and Temporal demand were removed. The Effort subscale was merged with the Mental Demand subscale in order to align with the concept of listening effort as defined as the mental energy an individual perceives they need to meet external demands (Pichora-Fuller et al., 2016; Peelle, 2018). Specifically, participants responded to the question "How mentally demanding was the task?" on a 21-point scale to gauge cognitive demand, with the endpoints on the left and right being "Low" and "Very High," respectively. Additionally, subscales assessing the perceived percentage of words understood and the percentage of conversation understood were added.

### Electrophysiological recording

Electroencephalogram data were recorded using a 64-channel antiCHamp Brain Products recording system (Brain Products GmbH, Inc., Munich, Germany) at a sampling rate of 2,000 Hz. EEG caps were fitted onto the participants so that the electrode corresponding to the typical Cz location according to the 10–20 international system was at the vertex of the skull, as determined using the point of intersection between nasion-to-inion and tragus-to-tragus midpoints. The EEG online reference was a separate reference electrode located slightly anterior to the vertex on the midline, while the ground electrode was located on the midline at the midpoint between the nasion and the vertex. Electrodes in proximity to or overlapping the CI magnet and coil on the side of the tested ear were not used; this ranged from 1 to 3 electrodes across all participants.

### Electroencephalogram preprocessing

Electroencephalogram recordings were imported into EEGLAB (Delorme and Makeig, 2004) for MATLAB 2019a (The MathWorks, Natick, MA, USA) before being filtered with a Finite Impulse Response bandpass filter from 0.3 to 40 Hz. Stimuli audio segments were simultaneously recorded along with the EEG in an auxiliary channel to maximize the temporal alignment of the audio stimuli to the EEG data. The audio soundtrack was then aligned offline to the recorded audio stream by calculating the cross-covariance to find the corresponding starting time point of the soundtrack itself. Speech envelopes were extracted by calculating the absolute value of the Hilbert transformation of the aligned audio soundtrack. The extracted envelope was then downsampled to the EEG recording sampling rate and appended to the EEG data for later analyses. After alignment, the EEG and audio data were both downsampled to 250 Hz and concatenated together across runs and conditions for each participant. Noisy channels and durations of extremely noisy data were removed prior to artifact removal. Independent Component Analysis (ICA) was conducted on the concatenated data to identify

stereotypical physiological artifacts (e.g., eye blinks, oculomotor movements, and cardiac activity) and technical artifacts (e.g., electrode pop and line noise). Components containing these artifacts were manually removed based on the visual inspection of component topographies. An average of 6.53 (10.36%;  $SD = 1.13$ ) components were removed across all participants. The recordings were then separated based on study condition for each participant, and the previously removed channels were interpolated using the spherical splines of neighboring channels (Perrin et al., 1989).

### Cochlear implant artifact suppression

In the presence of auditory stimuli, the electrical stimulation and radio-frequency signals of CIs impart electrical stimulation artifacts into the EEG recording (Wagner et al., 2018). While ICA has been previously used to identify CI artifacts in EEG recordings involving auditory evoked potentials (Gilley et al., 2006; Miller and Zhang, 2014), the nature of CI artifacts makes their reduction more challenging, especially for continuous tasks. In general, ICA separates statistically independent components from the mixed signal. In the case of artifact reduction in MEG and EEG data, temporal ICA separates signal components based on their temporal independence. Artifacts that are not temporally aligned to the stimulus onset (e.g., eye blinks), are therefore separated from sources that are temporally aligned (e.g., event-related potentials). CI artifacts, however, are temporally aligned to the sound onset in a continuous EEG recording, and have a similar morphology to the sound envelope (Mc Laughlin et al., 2013).

Second-order blind identification (SOBI) was applied (Belouchrani et al., 1997) to reduce CI-related artifacts as we have used previously (Paul et al., 2020, 2021). As the name suggests, SOBI is a blind source separation technique that uses second-order statistics in order to separate temporally correlated sources (Belouchrani et al., 1997). SOBI was performed on the continuous EEG data for each participant, and components of suspected CI artifacts were visually identified based on the topographical centroid and implant side used during the listening paradigm. Between 0 and 2 (~2.51%) components corresponding to the artifact centroids were removed for each participant ( $M = 1.58$ ,  $SD = 0.62$ ). After CI artifact suppression, the cleaned EEG and aligned audio stimuli were subsequently imported into Fieldtrip (Oostenveld et al., 2011) and re-referenced to an average reference before further processing to prepare for temporal response function (TRF) and alpha power analysis.

## Data analysis

### Temporal response function calculation

Temporal response functions (TRFs) were estimated from EEG data and the speech envelope of the audio signal using

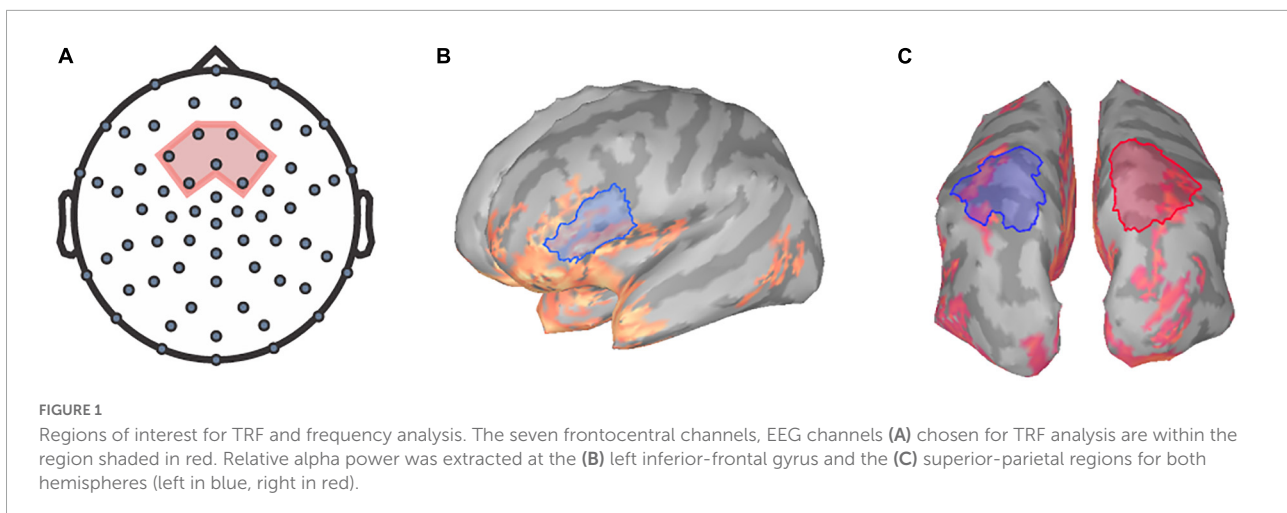
the mTRF Toolbox 2.0 (Crosse et al., 2016) in MATLAB. The processed audio signal and the EEG signal were filtered by a 1–20 Hz bandpass, 2nd order, zero-phase filter. The filtered EEG and audio signals were subsequently z-scored for each participant before TRF calculations. Integration window of time lags between  $-100$  and  $500$  ms were chosen for analysis. The *mTRFcrossval* function was used to optimize the regularization parameter that reduces the overfitting of the speech envelope to the EEG data. The *mTRFcrossval* function performs a leave-one out  $N$ -fold cross-validation, where for each participant TRFs are estimated for all trials except one. The estimated encoding model is then used to predict the EEG signal for the trial that was omitted, and the Pearson's correlation coefficient is calculated to assess the correlation between the predicted and actual EEG signals. This process was repeated for regularization parameters ranging from  $2^{-5}$  to  $2^{15}$ .

While some studies suggest the use of uniquely optimized regularization parameters for each participant dataset to maximize the TRF estimation (Crosse et al., 2021), this can potentially overfit the TRFs to remnant CI artifacts in CI user recordings. Thus, the correlation coefficients were then averaged across all subjects, channels, and conditions for each regularization parameter, and the value associated with the highest correlation value was chosen as the optimal parameter to use. The continuous EEG data was therefore epoched into 60-s trials prior to TRF calculation, and  $2^{11}$  was selected as the regularization parameter. TRF epochs were averaged within each condition for every participant, generating four sets of TRFs per participant for each listening condition.

### Neural speech tracking analysis

Prior to analysis, TRFs for each channel were baseline corrected by subtracting the mean of the values between the baseline period ( $-100$  to  $-8$  ms) from each point in the TRF waveform using Brainstorm (Tadel et al., 2011). Previous literature has identified that speech tracking component amplitudes reaching their maximum magnitudes within the frontocentral region (e.g., Ding and Simon, 2012; Fiedler et al., 2019). Thus, we focused on the frontocentral scalp region as the region of interest (ROI) at the sensor level. Specifically, the TRFs for seven frontocentral EEG sensors within the chosen ROI were averaged for each participant per condition to generate a mean TRF for further analysis (Figure 1A).

To identify the time windows of interest containing the relevant speech components, the babble noise audio was used to create a comparison standard with similar spectral but a temporally different speech envelope. As with the stimuli soundtrack, the absolute value of the Hilbert transformation was calculated for the babble comparison soundtrack to extract the amplitude envelope. The comparison soundtrack was then scaled to the root-mean-squared amplitude of attended soundtrack amplitude before being subjected to the same TRF estimation procedure. TRF differences between the stimuli TRF



and the babble noise TRF at each time lag were first compared using two-tailed paired *t*-tests. *T*-test values were then corrected using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). TRF differences at time lags were considered significantly different if the adjusted *p*-values were less than the alpha criterion of 0.05. Component TRF peak amplitudes and peak latencies were then extracted by finding the point of the local maxima/minima of the component within the identified time lag windows.

### Alpha power

The involvement of the IFG in effortful speech listening has been examined in various studies (e.g., Alain et al., 2018; Price et al., 2019). We have previously shown that neural tracking and alpha power within the left IFG has been observed to predict successful trials for speech perception and be correlated with listening effort (Dimitrijevic et al., 2019). Additionally, previous literature has indicated that parietal alpha oscillations may reflect cross-modal attentional modulation (e.g., Misselhorn et al., 2019). Relative alpha power was extracted from EEG data using Brainstorm following the methods outlined in Niso et al. (2019). The preprocessed spontaneous EEG recordings were first imported into Brainstorm, and source activations were calculated using standardized low-resolution electromagnetic tomography (sLORETA) modeling (Pascual-Marqui, 2002). sLORETA models provide estimates of the power and location of the neural generators that underlie electrophysiological processes. Boundary element model (BEM) head models were created using the OpenMEEG plugin prior to sLORETA modeling. sLORETA models were z-scored, and alpha power was computed using the Brainstorm resting-state pipeline as outlined in Niso et al. (2019).

The power spectral densities (PSDs) of each voxel of the sLORETA models were estimated from 0 to 125 Hz by applying Welch's method using a window of 1 s at 50% overlap to obtain a 1.0 Hz frequency resolution. Spectrum normalization was then

applied by dividing the PSDs by the total power. PSDs were then averaged across all participants and sides for each condition to identify ROIs for alpha analysis. Alpha power was extracted over the left IFG ROI, a 38.74 cm<sup>2</sup> region encompassing the pars triangularis and the pars opercularis (Figure 1B) as defined by the Desikan-Killiany Atlas (Desikan et al., 2006). Additionally, ROIs encompassing the superior-parietal cortices (Desikan et al., 2006) were also defined as a 57.32 cm<sup>2</sup> area within the left posterior parietal region and a 57.45 cm<sup>2</sup> area within the right posterior parietal region (Figure 1C). Mean relative alpha power was calculated by averaging alpha power values between 8 and 12 Hz and then dividing by the total power. The relative alpha power was then extracted over all ROIs for each participant per condition.

### Statistical analysis

All statistical analyses were performed in R (R Core Team, 2021) and MATLAB 2019a (MATLAB, 2019). One-way repeated measures analysis of variance (RM-ANOVA) with the main effect of Condition (Low-Noise, Moderate-Noise, High-Noise, Audio-Only) were used to compare behavioral scores. Two-way factorial RM-ANOVAs with the main effect of Condition and TRF Component (TRF<sub>100</sub>, TRF<sub>200</sub>, TRF<sub>350</sub>) were performed to compare the effect of background noise and visual cues on TRF component amplitudes. Two-way RM-ANOVAs with the main effects of Condition and Listening Side (Ipsilateral to implant, Contralateral to implant) and a paired *t*-test were performed to compare relative alpha power for the superior-parietal and IFG ROIs, respectively. In cases where sphericity is violated, Greenhouse-Geisser corrections were applied to *p*-values (*p*<sub>GG</sub>). For all tests, the alpha criterion was set at 0.05, and *p*-values (*p*<sub>adj</sub>) were adjusted for multiple comparisons using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Generalized eta squared ( $\eta^2_G$ ) values are

also reported alongside ANOVA test statistics to indicate effect size.

Linear mixed models were applied to predict listening demand, perceived percentage of words understood, and perceived percentage of the conversation understood from the fixed effects of TRF component or relative alpha power, listening condition, and subject age using the R package *lme4()* v1.1.29 (Bates et al., 2015). The relationship between listening demand and neural variables were analyzed separately for TRF components and relative alpha power. TRF component amplitudes, alpha power values, and age were treated as continuous variables and z-scored prior to modeling. Although by-subject random effects and condition were included as random intercepts and random slopes respectively to account for the variability explained by the aforementioned factors, random effects are not interpreted due to being jointly unidentifiable as each subject was statistically measured once per condition. Additionally, the by-subject random slope for condition was removed from the perceived percentage of words models due to non-convergence. The correlation between the random intercept for subject and the by-subject random slope for condition was removed for the perceived percentage of conversation models due to convergence issues as well. Fixed effects were subsequently analyzed using ANOVAs with degrees of freedom adjusted by the Satterthwaite method (Luke, 2017), and all *post-hoc* analyses were performed through pairwise comparisons of the estimated marginal means. As there were no AzBio score matching the SNR + 15 and + 10 dB conditions, two-tailed partial Pearson correlations while controlling for age were used to correlate neural measures and AzBio scores. Correlations were controlled for age due to the effects of aging on hearing thresholds. The relationship between perceived percentage of conversation understood and AzBio scores were also explored using a two-tailed partial Spearman correlation while controlling for age. For the partial correlations, the alpha criterion was set at 0.05 and *p*-values ( $p_{adj}$ ) were adjusted for multiple comparisons using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

## Results

### Self-reported listening demand, percentage of words, and conversation understood

Average self-reported mental demand ratings during listening in the audiovisual conditions were 6.78 ( $SD = 2.58$ ) for the Low-Noise (LN) condition, 11.21 ( $SD = 3.20$ ) for the Moderate-Noise (MN) condition, and 15.44 ( $SD = 2.61$ ) for the High-Noise (HN) condition. The average mental demand rating for the Audio-Only (AO) condition was 14.62 ( $SD = 2.63$ ).

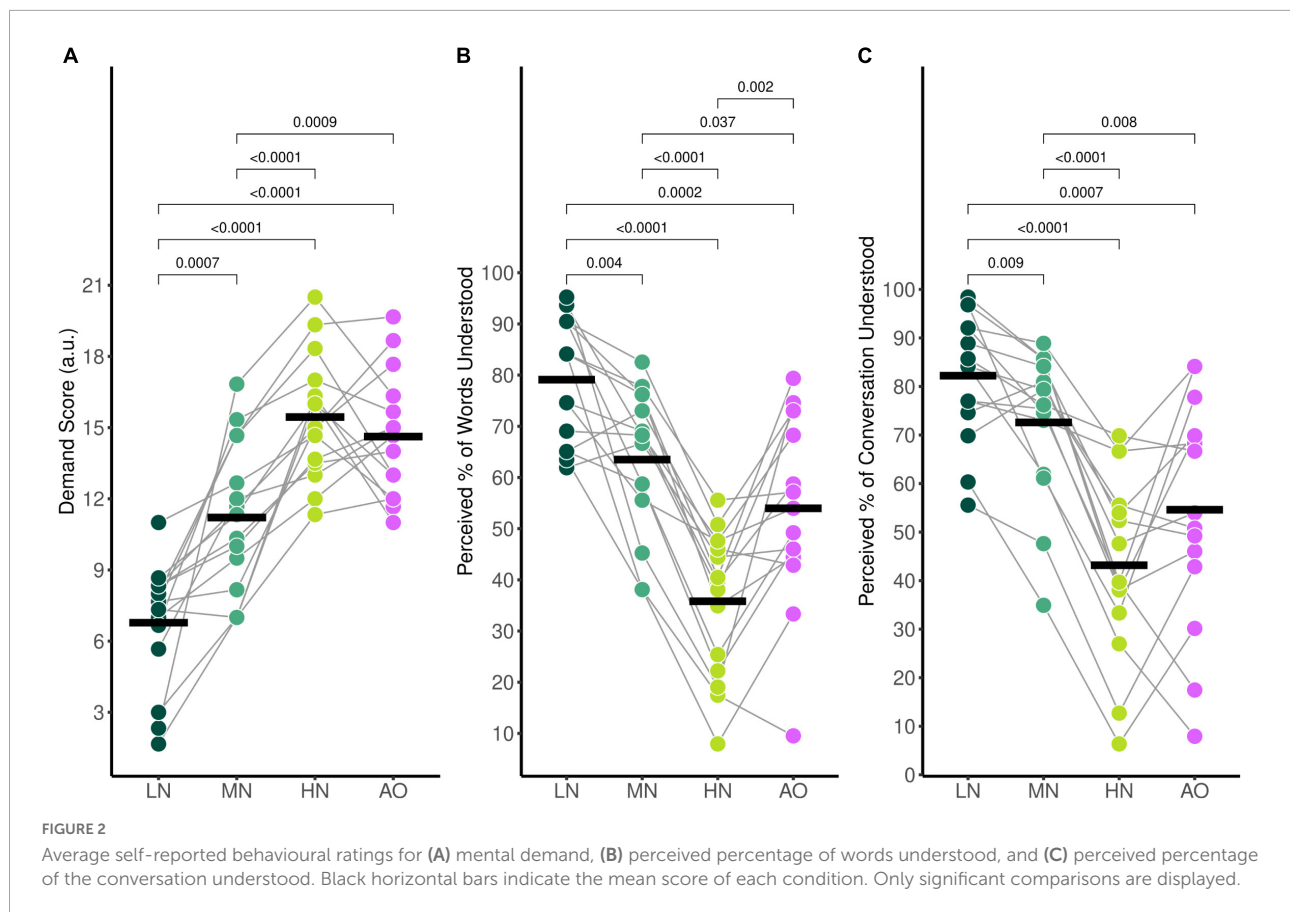
The ANOVA model returned an effect of condition on mental demand ratings [ $F_{(3,42)} = 40.83, p < 0.001, \eta^2_G = 0.620$ ].

As expected, self-reported mental demand ratings increased as background babble noise levels increased during the audiovisual listening paradigm (Figure 2A). *Post-hoc* analysis using paired *t*-tests revealed that the average demand rating was significantly higher in the HN condition compared to the MN condition ( $p_{adj} < 0.001$ ), and the LN condition ( $p_{adj} < 0.001$ ). Additionally, the MN condition demand rating was also statistically significantly higher than for LN ( $p_{adj} = 0.007$ ). To compare the effects of visual cues on listening effort, the background babble noise level for the LN and AO conditions were both set at SNR +15 dB. *Post-hoc* analysis indicates that the average demand rating for AO was significantly higher than LN ( $p_{adj} < 0.001$ ) despite taking place in the same background noise level. Interestingly, while the AO demand rating was statistically significantly higher compared to the MN rating ( $p_{adj} = 0.009$ ) as well, it was not significantly different compared to the HN demand rating ( $p_{adj} = 0.322$ ) despite the 10 dB difference in SNR between the two listening conditions.

The average percentage of words understood followed the same trends (Figure 2B); participants reported 79.1% ( $SD = 12.0\%$ ) of words understood for the LN condition, 63.5% ( $SD = 14.1\%$ ) for the MN condition, 35.8% ( $SD = 14.1\%$ ) for the HN condition, and 54.0% ( $SD = 18.5\%$ ) for the AO condition. There was a significant effect of condition [ $F_{(3,42)} = 36.67, p_{GG} < 0.001, \eta^2_G = 0.543$ ], with all conditions significantly differing as shown by *post-hoc* pairwise analysis. The percentage of words understood in the LN condition was significantly higher compared to the MN ( $p_{adj} = 0.004$ ), HN ( $p_{adj} < 0.001$ ), and AO ( $p_{adj} < 0.001$ ) condition. The MN percentage rating was significantly higher than the HN ( $p_{adj} < 0.001$ ) and AO ( $p_{adj} = 0.037$ ) ratings as well. The percentage of words understood in the HN condition was significantly lower compared to the percentage understood in the AO condition ( $p_{adj} = 0.002$ ).

Perceived percentage of conversation understood also decreased parametrically as background noise increased (Figure 2C). Participants reported that they understood an estimated 82.2% ( $SD = 13.1\%$ ) of the conversation for the LN condition, 72.6% ( $SD = 15.2\%$ ) for the MN condition, 43.2% ( $SD = 18.7\%$ ) for the HN condition, and 54.6% ( $SD = 23.1\%$ ) for the AO condition. The ANOVA model revealed a significant effect of condition [ $F_{(2,10,29.36)} = 27.65, p_{GG} < 0.001, \eta^2_G = 0.435$ ]. *Post-hoc* analysis revealed that the perceived percentage of the conversation understood differed across all conditions. The percentage rating in the LN condition was significantly higher compared to the MN ( $p_{adj} = 0.009$ ), HN ( $p_{adj} < 0.001$ ), and AO conditions ( $p_{adj} = 0.007$ ). The rating in the MN condition was also higher than the HN ( $p_{adj} < 0.001$ ) and AO condition ( $p_{adj} = 0.079$ ). Similar to mental demand ratings, the percentage of conversation understood also did





not significantly differ between the HN and AO conditions ( $p_{adj} = 0.051$ ).

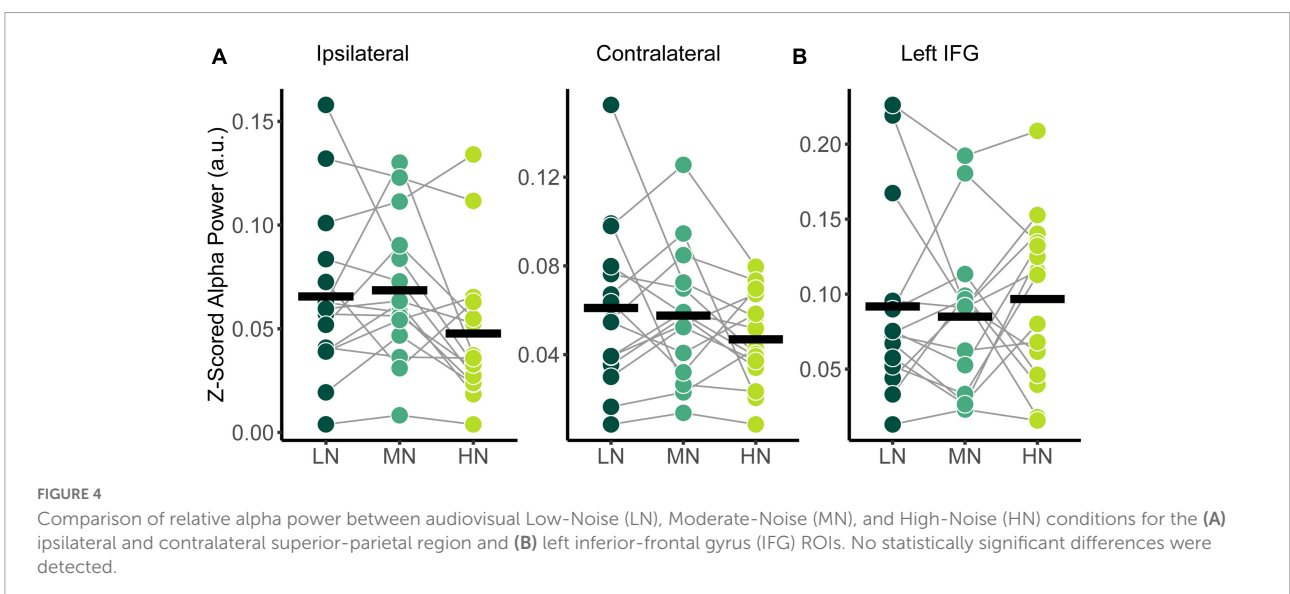
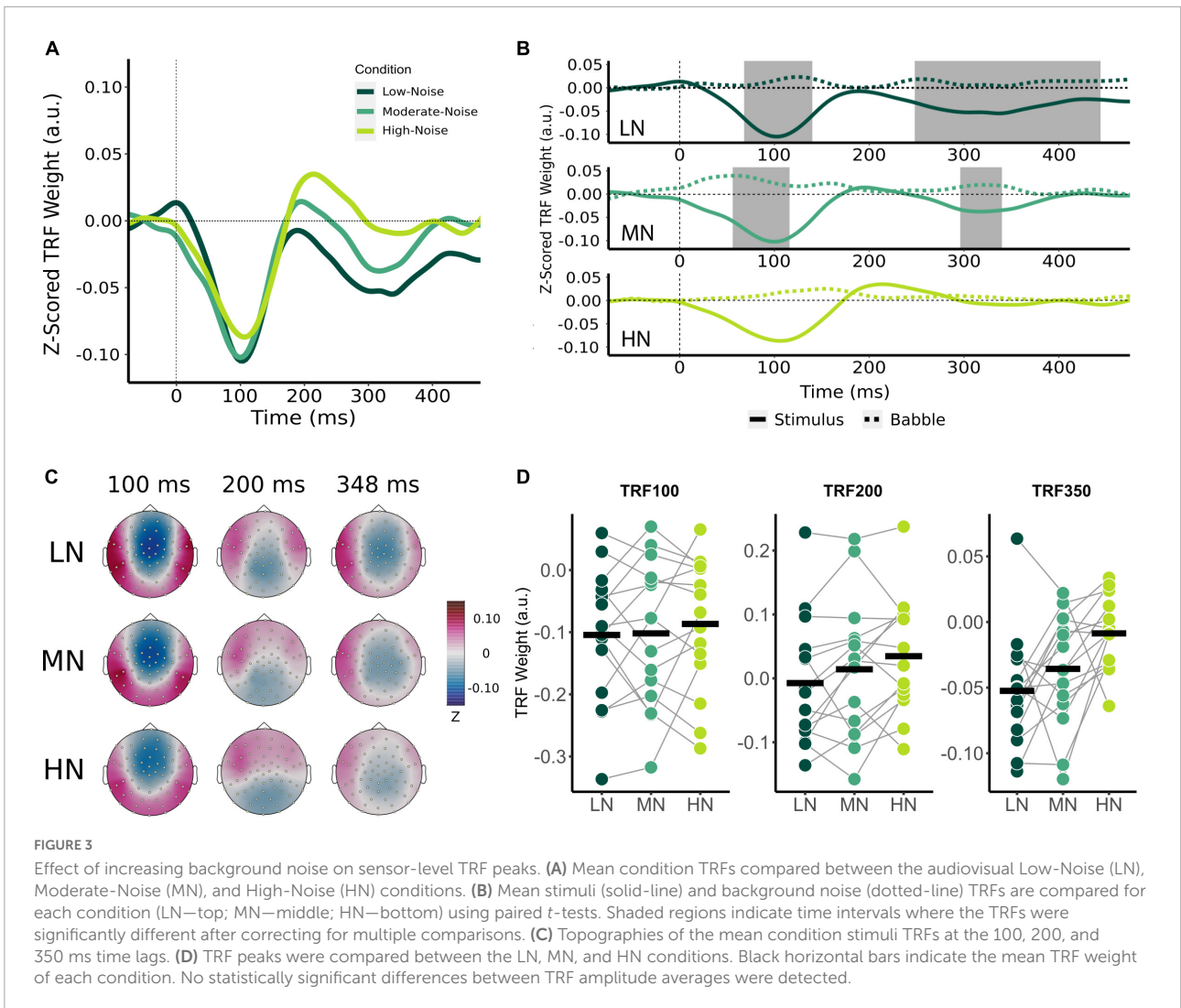
## Effect of increasing background noise on neural measures

### Effect of background noise on audiovisual-driven TRF components

Comparison of audiovisual-driven stimuli and babble noise TRFs revealed two significant time lag windows containing negative TRF components for the LN and MN conditions as indicated by the shaded regions in **Figure 3B**, where stimuli TRFs differed from babble noise TRFs around the 100 ms time lag and the 350 ms time lag (herein referred to as TRF<sub>100</sub> and TRF<sub>350</sub> respectively). Visual inspection of the TRF at frontocentral sensors also suggested a TRF component at the 200 ms time lag (TRF<sub>200</sub>; **Figures 3A,C**), but this did not significantly differ from the babble noise TRFs in each condition. Nonetheless, the TRF<sub>200</sub> was submitted to statistical analysis to compare across latencies. Based on these responses, an 8 ms time window surrounding the grand mean component peaks were chosen for analysis for the TRF<sub>100</sub> and TRF<sub>200</sub>. A 48 ms time window surrounding the TRF<sub>350</sub> component

peaks was chosen due to the previously observed wider time lag window width. The average component peak amplitudes were subsequently calculated by averaging the mean stimuli TRF values within the chosen time lag intervals for each participant, and compared between conditions using a two-way RM-ANOVA with the main effect of Condition (Low-Noise, Moderate-Noise, High-Noise) and TRF Component (TRF<sub>100</sub>, TRF<sub>200</sub>, TRF<sub>350</sub>).

Initially, mean TRF waveforms of the audiovisual conditions appear to differ in average amplitude for all three speech tracking components (**Figures 3A,D**). TRF<sub>100</sub> amplitude seemed to decrease with increasing SNR, with mean amplitudes being  $-0.104$  ( $SD = 0.106$ ) for LN,  $-0.102$  ( $SD = 0.116$ ) for MN, and  $-0.087$  ( $SD = 0.107$ ) for the High-Noise condition. While the mean TRF<sub>200</sub> component was not significantly different from the babble TRF, TRF<sub>200</sub> amplitude initially appeared to increase as SNR decreased. The TRF<sub>200</sub> peak amplitudes were as follows:  $-0.007$  ( $SD = 0.099$ ) for LN,  $0.014$  ( $SD = 0.108$ ) for MN, and  $0.035$  ( $SD = 0.090$ ) for HN. TRF<sub>350</sub> peak amplitudes followed a similar trend to the TRF<sub>100</sub> component, in that amplitudes decreased as listening condition SNR increased: TRF<sub>350</sub> peak amplitudes were  $-0.052$  ( $SD = 0.046$ ) for LN,  $-0.058$  ( $SD = 0.051$ ) for MN, and  $-0.040$  ( $SD = 0.032$ ) for HN. While the ANOVA model revealed no significant



interaction effect between condition and TRF component [ $F_{(2,91,40.75)} = 0.468$ ,  $p_{GG} = 0.701$ ,  $\eta^2_G = 0.003$ ], the main effect of Condition was statistically significant [ $F_{(1,78,24.95)} = 3.58$ ,  $p_{GG} = 0.048$ ,  $\eta^2_G = 0.027$ ]. *Post-hoc* analysis however revealed no significant differences in general TRF component amplitude between conditions (all  $p_{adj} > 0.05$ ).

Beyond the results of the previous analysis, the large variance in TRF amplitudes between participants and the inclusion of the TRF<sub>200</sub> component (despite the non-significant difference between stimuli and babble-noise TRF<sub>200</sub> amplitudes) into the omnibus analysis may have lowered its statistical power. Thus, exploratory one-way RM-ANOVAs with the main effect of Condition were conducted separately for each TRF component. The ANOVAs revealed non-significant main effects for the TRF<sub>100</sub> [ $F_{(2,28)} = 0.347$ ,  $p = 0.710$ ,  $\eta^2_G = 0.005$ ] and TRF<sub>200</sub> components [ $F_{(2,28)} = 3.03$ ,  $p = 0.064$ ,  $\eta^2_G = 0.031$ ]. TRF<sub>350</sub> component amplitude was found to significantly differ between listening conditions [ $F_{(1,84,25.82)} = 5.47$ ,  $p_{GG} = 0.012$ ,  $\eta^2_G = 0.193$ ]. *Post-hoc* analysis revealed that HN TRF<sub>350</sub> amplitude was significantly lower than in LN ( $p_{adj} = 0.032$ ), but MN TRF<sub>350</sub> did not significantly differ from LN ( $p_{adj} = 0.234$ ) and HN ( $p_{adj} = 0.051$ ).

### Effect of background noise on relative alpha power

The power spectral densities for each participant were estimated from the sLORETA models of the spontaneous EEG recordings. Subsequently, the relative alpha power over superior-parietal cortices (Figure 1A) and the left IFG (Figure 1B) were compared between conditions. Relative alpha power at the superior-parietal cortices was analyzed with a two-way RM-ANOVA with the main effects of Condition (Low-Noise, Moderate-Noise, High-Noise) and Listening Side (Ipsilateral, Contralateral), while one-way RM-ANOVA with the main effect of Condition (Low-Noise, Moderate-Noise, High-Noise) was performed for the left IFG. Initially, there appeared to be a parametric increase in relative alpha power at the left IFG as listening condition demand increased. Subsequent analysis, however, revealed no statistically significant interaction effect between Condition and Listening Side for relative alpha power for all ROIs (all  $p > 0.078$ ; Figure 4). Furthermore, the main effect of Condition and Listening Side were also not statistically significant (all  $p > 0.230$ ).

## Effects of the presence of visual cues on neural measures

### Effect of visual cues on TRF components

Like audiovisual-driven TRFs, the audio-only TRF at the frontocentral sensors also contained two significant time lag windows around the 100 and 250 ms time lag (Figures 5B,C). As with the movie TRFs, an 8 ms time window was chosen for

analysis for the TRF<sub>100</sub> and TRF<sub>200</sub> and a 48 ms time window was chosen for TRF<sub>350</sub>.

Average component peak amplitudes were compared between LN and AO conditions using paired *t*-tests. Like the audiovisual conditions, TRF component amplitudes appeared to change parametrically in the more demanding AO condition (Figure 5D). Compared to LN, which had component amplitudes of  $-0.104$  ( $SD = 0.106$ ) for TRF<sub>100</sub>,  $-0.007$  ( $SD = 0.099$ ) for TRF<sub>200</sub>, and  $-0.052$  ( $SD = 0.046$ ) for TRF<sub>350</sub>, AO TRF components amplitudes were slightly greater. The AO TRF<sub>100</sub>, TRF<sub>200</sub>, and TRF<sub>350</sub> amplitudes were  $-0.113$  ( $SD = 0.132$ ),  $0.002$  ( $SD = 0.102$ ), and  $-0.059$  ( $SD = 0.053$ ) respectively. Similar to the analysis for background noise, no significant interaction effect between condition and TRF component was observed [ $F_{(2,28)} = 0.573$ ,  $p = 0.570$ ,  $\eta^2_G = 0.002$ ]. The main effect of Condition was also not significant [ $F_{(1,14)} = 0.022$ ,  $p = 0.885$ ,  $\eta^2_G = 0.0001$ ].

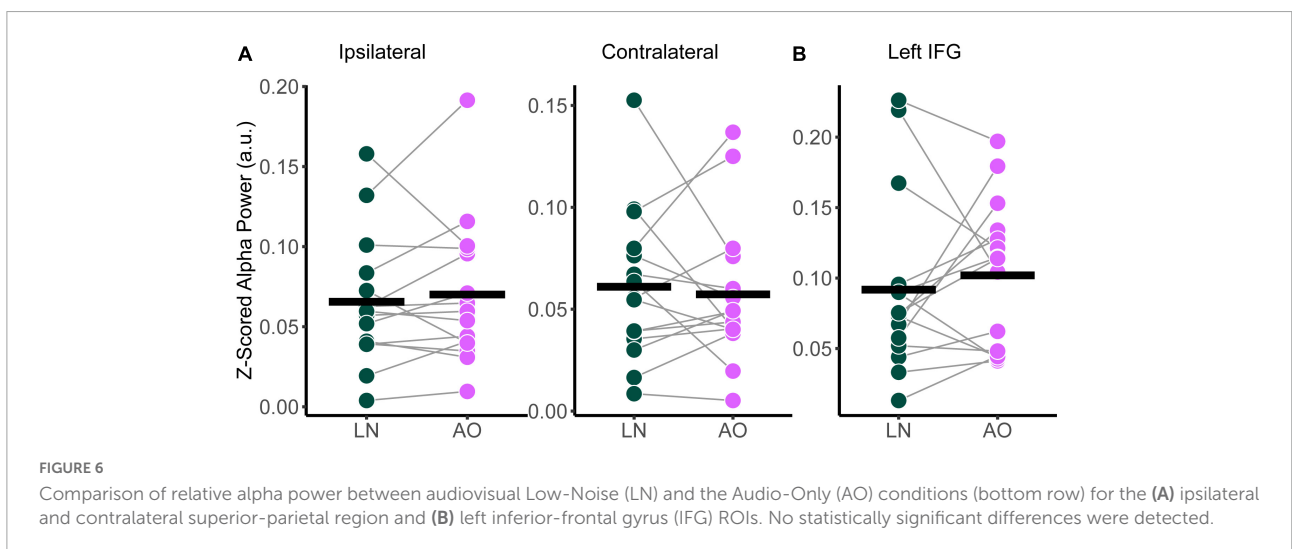
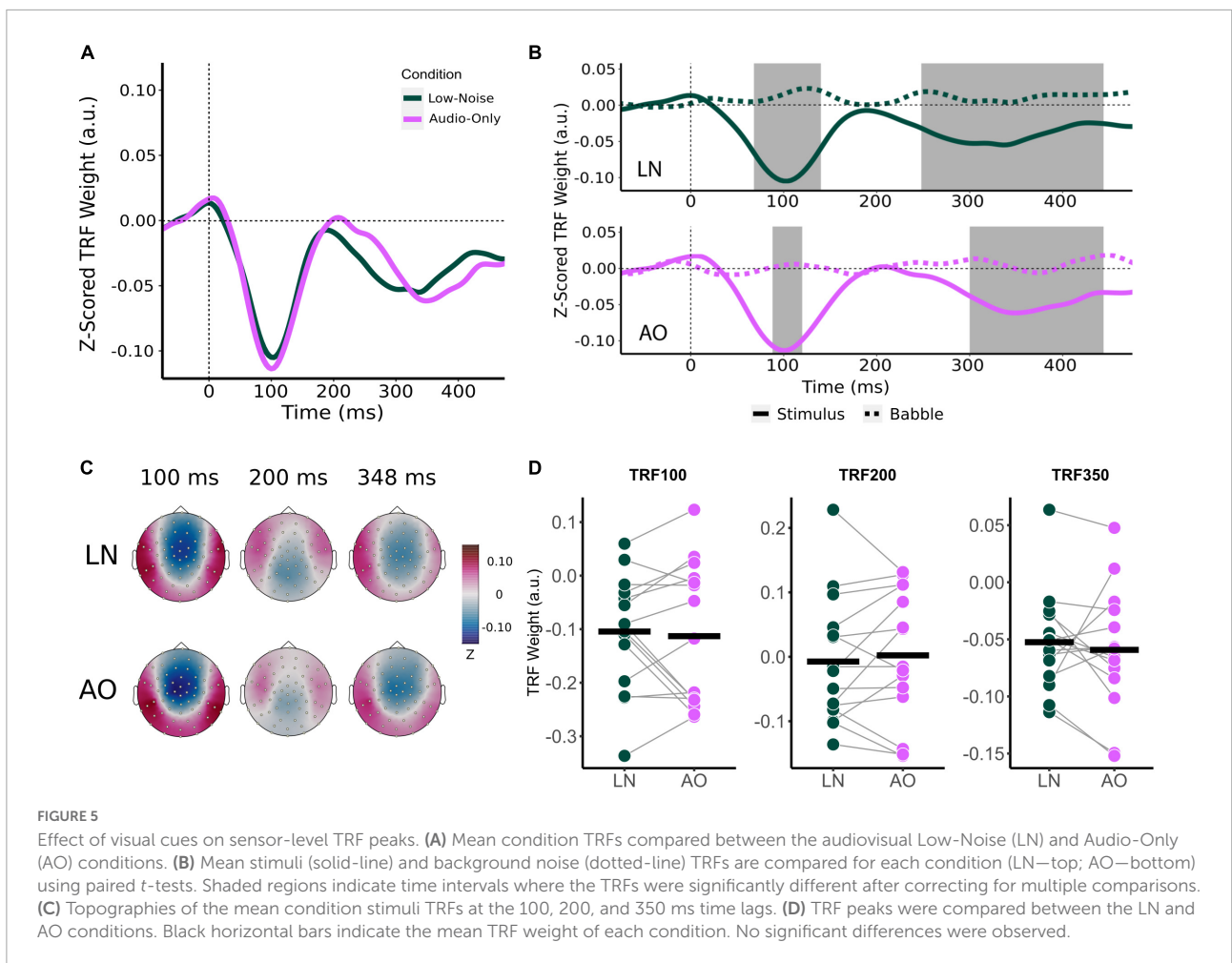
### Effect of visual cues on relative alpha power

Relative alpha power was analyzed with a two-way RM-ANOVA with the main effects of Condition (Low-Noise, Audio-Only) and Listening Side (Ipsilateral, Contralateral) for the superior-parietal cortices. A paired *t*-test was performed for the left IFG. As with the audiovisual condition comparisons, analyses revealed no statistically significant interaction effect between Condition and Listening Side for relative alpha power the superior-parietal ROI [ $F_{(1,14)} = 1.39$ ,  $p = 0.258$ ,  $\eta^2_G = 0.003$ ; Figure 6A] and no significant effect of Condition at the left IFG [ $t(14) = -0.674$ ,  $p = 0.512$ ; Figure 6B], despite the apparent parametric change in alpha with condition demand. Furthermore, the main effects of Condition and Listening Side were also not statistically significant for the superior-parietal region (all  $p > 0.218$ ).

## Relationship between neural measures, subjective listening demand, perceived percentage of conversation understood, and AzBio scores

Across all linear mixed models involving TRFs, the Audio-Only condition was set as the reference category to investigate the relationship between changes speech tracking components and behavioral measures as an effect of adding visual cues and increasing background noise levels. The fixed effect of condition was significant across all models (all  $p < 0.001$ ), supporting the previous analysis on differences in subjective behavioral scores between listening conditions. Thus, the fixed effect of condition was not analyzed further owing to the previous analyses performed.

Mixed effects modeling was performed on self-reported listening demand, speech perception during listening, and TRF component amplitude while controlling for participant



age. Modeling indicates that demand scores increase as negative-going TRF component magnitudes attenuate and age increases. The analysis reported significant fixed effects of TRF<sub>100</sub> amplitude [standardized  $\beta = 1.961$ , SE = 0.419,

$F_{(1,31.4)} = 5.461$ ,  $p < 0.001$ ] and age [standardized  $\beta = 0.646$ , SE = 0.267,  $F_{(1,7.7719)} = 5.847$ ,  $p = 0.043$ ]. No other fixed effects aside from condition reached significance (all  $p > 0.110$ ). Regarding the percentage of words understood, no significant

fixed effects other than condition was detected (all  $p > 0.069$ ). In contrast, analysis of the mixed model on perceived percentage of conversations understood and TRF component amplitudes while accounting for age also indicates a significant fixed effect of TRF<sub>100</sub> amplitude [standardized  $\beta = -1.418$ , SE = 1.058,  $F(1,19.797) = 11.361$ ,  $p < 0.001$ ] as well as an interaction effect between condition and TRF<sub>200</sub> amplitude [ $F(3,13.071) = 3.428$ ,  $p = 0.049$ ]. However, there was no specific interaction effects between TRF components and condition on listening demand (all  $p > 0.090$ ).

Similar models were fitted for relative alpha power at the superior-parietal and left IFG ROIs. Since there was no significant main effect of Listening Side, the superior-parietal relative alpha power at the ipsilateral and contralateral hemispheres were first averaged for each participant to create a composite measure. ANOVA on the fixed effects revealed a significant interaction between superior-parietal relative alpha power and condition [ $F(3,13.364) = 3.457$ ,  $p = 0.047$ ], specifically in the High-Noise condition (standardized  $\beta = -1.992$ , SE = 0.701,  $p = 0.014$ ). As with the TRF model, the fixed effect of age was also significant [standardized  $\beta = 0.6460$ , SE = 0.267,  $F(1,7.7719) = 5.846$ ,  $p = 0.043$ ]. For the left IFG, ANOVA on the model fixed effects reported no significant terms (all  $p > 0.056$ ). The regression lines and the 95% confidence intervals of z-scored TRF<sub>100</sub> amplitude against the respective self-report scores are plotted in **Figures 7A,B**. Weaker TRF<sub>100</sub> components displaying increasingly positive amplitudes (as TRF<sub>100</sub> is a negative component) are correlated with higher demand scores (**Figure 7A**). In tandem, weaker TRF<sub>100</sub> components are also correlated with a lower perceived percentage of the conversation understood (**Figure 7B**). **Figure 7C** shows the mean superior-parietal alpha power is plotted against demand for the High-Noise condition along with the regression line and the 95% confidence interval to show the association between increasing alpha power in the superior-parietal region and decreasing demand ratings.

Two-tailed partial Pearson correlations between TRF component amplitudes and AzBio In-Quiet and In-Noise scores revealed no significant correlations (all  $p_{adj} > 0.160$ ). Similarly, no significant correlations were found between relative alpha power and AzBio scores across all ROIs (all  $p_{adj} > 0.246$ ). Correlations between AzBio In-Quiet and In-Noise scores and perceived percentage of conversation understood also revealed that while the percentage rating correlated with AzBio In-Quiet and In-Noise scores in the AO condition prior to correcting for multiple comparisons, no correlations survived corrections for multiple comparisons (all  $p_{adj} > 0.075$ ).

In summary, TRF<sub>100</sub> amplitude was correlated to listening demand and proportion of the movie conversation understood and superior-parietal alpha power in the high-noise condition was related to listening demand. No significant correlations were observed for TRF component amplitude and alpha power with clinical AzBio speech perception scores.

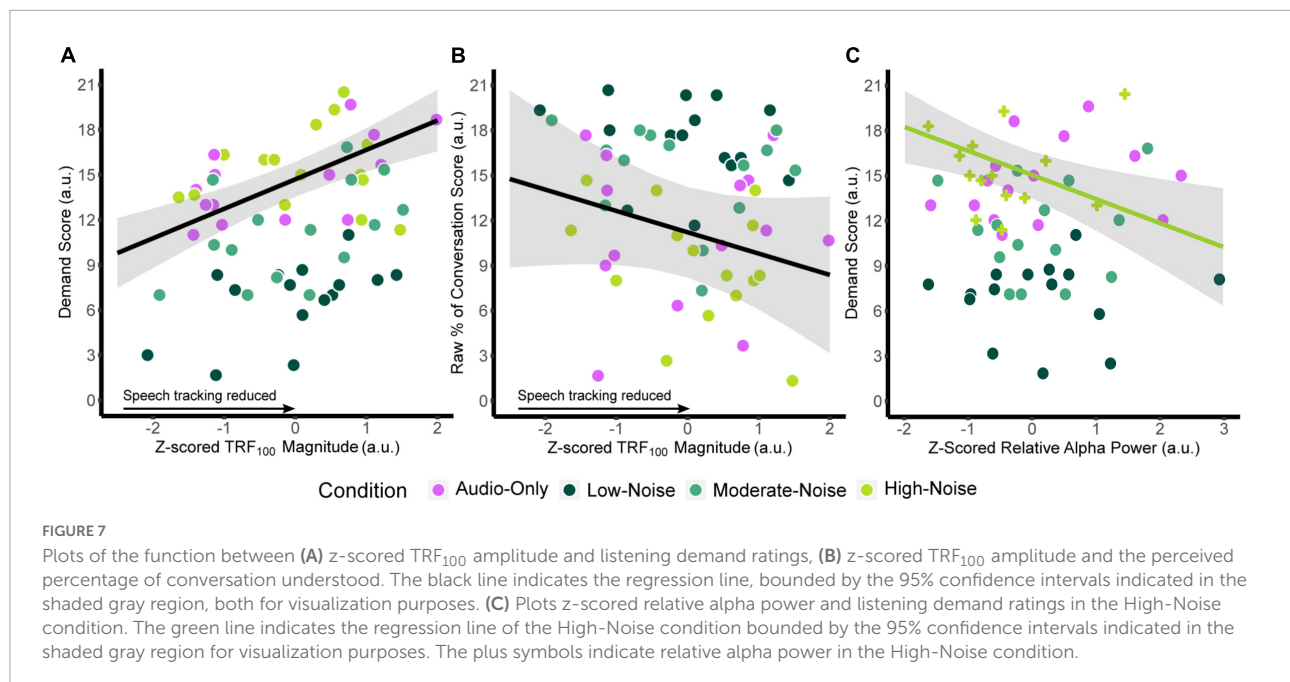
## Discussion

The current study investigated neural speech tracking and alpha power and the degree to which these neural measures are reflected in the self-reported mental demand of listening in CI users when attending to naturalistic stimuli. We found: (1) On average, subjects reported higher demand ratings and lower word/conversation understanding as the noise levels increased. Compared to the audiovisual condition, participants presented with audio-only stimuli reported demand and conversation understanding scores that were comparable to the high noise condition, even though the babble noise levels were the same as the low noise condition. (2) Neural tracking/alpha power differences: The neural tracking data showed no significant parametric change in TRF components with increasing babble noise. No differences in neural tracking were observed when the movie visuals were removed. Alpha power did not show parametric changes with noise masker level or with access to visual cues. (3) Relationships to behavior: Neural tracking TRF<sub>100</sub> was related to self-reported demand and conversations understood, while superior parietal alpha power was only related to self-reported demand.

### Self-reported behavioral scores change concurrently with background babble noise levels, and is offset by the presence of visual cues

In the current study, the Effort subscale of the NASA-TLX was subsumed under the Mental Demand subscale to coincide with the definition of mental effort as “the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task” according to the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016). In line with previous literature on listening effort and speech intelligibility, self-reported mental demand ratings of CI users attending to audiovisual speech stimuli significantly increased concurrently with background noise levels, supported by similar decreases in perceived percentage of words and conversation understood. Increasing background noise has been shown to require more self-perceived listening effort as indicated by significantly increasing mean NASA-TLX scores of NH listeners attending to conversations and monologs (Peng and Wang, 2019). Accordingly, mean NASA-TLX scores of NH listeners also increases as the spectral resolution of speech (and thereby speech intelligibility) decreased in CI simulations (Pals et al., 2013).

Interestingly, mental demand ratings in the Audio-Only condition with low background noise levels (SNR + 15 dB) were comparable to the demand ratings when attending to audiovisual stimuli in the High-Noise condition (SNR + 5 dB). Additionally, while the perceived percentage of words



understood significantly differed between the HN and AO conditions, the percentage of conversation understood did not. When looking specifically at the percentage of words and conversation understood, participants on average reported a  $\sim 20\%$  decrease in understood words in the High-Noise condition compared to the Audio-Only condition. These results indicate that although CI users may perceive fewer words in a SNR + 5 dB listening condition with visual cues, their speech perception and comprehension ability is at a similar level to that of listening without visuals in a lower noise condition. Visual speech cues have been shown to improve the ability of CI users to recognize speech (Dorman et al., 2016). The benefit of visual speech cues on speech intelligibility therefore possibly reduces the mental demand required for CI users to attend to speech in noise. However, prior research on the effect of visual speech cues on subjective listening demand are divergent. For instance, NH individuals reported more difficulty in understanding audiovisual speech stimuli in which the talker wore a non-transparent face mask, compared to when no face mask was worn (Yi et al., 2021). Fraser et al. (2010) observed that NH listeners were more accurate and reported less subjective effort during the audiovisual speech recognition task compared to the audio-only task in equally noisy conditions. In contrast, Brown and Strand (2019) reported that listening effort toward speech in NH listeners did not differ due to the presence of visual facial cues during a dual-task paradigm in hard listening conditions (SNR + 5 dB). Reaction time was instead significantly slower in the audiovisual condition compared to the audio-only condition in easy listening conditions at SNR + 10 dB, indicating that a greater listening effort was required (Brown and Strand, 2019).

The increase in subjective demand in the Audio-Only condition may be due to the degree to which visual speech cues influences speech perception for CI users, compared to NH listeners. CI users have been shown to gravitate toward and rely more on visual speech cues, especially in situations with poor speech intelligibility (e.g., Mastrantuono et al., 2017; Wang et al., 2020). In fact, individuals with hearing loss have been previously observed to be innately biased toward visual speech cues compared to NH listeners (Stropahl et al., 2017). Furthermore, individuals with hearing loss and CI users display cross-modal recruitment of auditory regions by visual stimuli (e.g., Nishimura et al., 1999; Winn et al., 2013; Anderson et al., 2017). Anderson et al. (2017) has shown that the post-operative increase in cross-modal superior temporal cortex activation due to visual speech cues was positively correlated with increases in speech understanding. Thus, while CI users demonstrate greater listening effort compared to NH listeners during everyday listening, it is possible that they also receive greater benefit from visual speech cues in the reduction of listening demand.

## Influence of background noise and visual cues on TRF components

The current study produced speech tracking waveforms with components that have a similar morphology to potentials commonly identified as part of the cortical auditory evoked potential (i.e., N1, P2, N2). Audiovisual-driven TRFs at frontocentral sensors in conditions of increasing background noise resembling N1-P2 cortical auditory evoked potentials (Figure 3A) has been observed in past studies

(Hjortkær et al., 2020; Verschueren et al., 2020). The audio-driven TRF also resemble an N1-P2 cortical auditory evoked potential (Figure 5A). Despite the similarities between the polarity and latencies of the TRF<sub>100</sub>, TRF<sub>200</sub>, and TRF<sub>350</sub> components to the N100, P200, and N200 event-related potentials respectively, we will not refer to these responses as the classic event-related-potentials due to differences in response characteristics as described in previous literature (Broderick et al., 2018; Reetzke et al., 2021). Recent evidence suggests that TRFs are sensitive to stimulus properties, and are more representative of low-level speech encoding (Broderick et al., 2022; Prinsloo and Lalor, 2022). Accordingly, weaker speech tracking would be indicative of either a deficit in the basic neural processing of speech acoustics, or poor encoding of low-level speech features (Broderick et al., 2022; Prinsloo and Lalor, 2022).

### General effect of increasing background noise levels on TRF components during audiovisual listening in cochlear implant users

Previous work with NH listeners has shown that background noise and attention both modulate neural speech tracking (Ding and Simon, 2012; Petersen et al., 2017; Verschueren et al., 2020). The magnitude of TRF components were found to decrease as background noise levels increase/SNR decrease (Petersen et al., 2017; Verschueren et al., 2020), even if speech intelligibility remained comparable (Verschueren et al., 2020). Increasing background noise has also been shown to also increase the magnitude of a positive speech tracking component around 200 ms time lag in individuals with hearing loss (Petersen et al., 2017). Concerning the attentional modulation of speech tracking, hearing loss has been shown to negatively affect attentional modulation. Greater degrees of hearing loss were associated with changes in the tracking of the ignored speech stream, leading to smaller differences between the neural tracking of attended and ignored speech streams (Petersen et al., 2017).

Although a general effect of background noise on TRFs was detected, the difference in specific component amplitudes between listening conditions was non-significant. Individual differences between CI users may explain the lack of difference in group-level TRF component amplitudes despite the main effect of increasing background noise, reflecting the large variability in CI outcomes reported in past studies (Blamey et al., 2012; Lazard et al., 2012). Some factors to be considered are the type of CI user (i.e., unilateral, bilateral, bimodal), CI device limitations, and individual differences in effort and attention. The present study contains unilateral, bimodal, and bilateral CI users who may have different subjective listening experiences, and thus different effort and demand requirements. For instance, bilateral CI users have reported a lower degree of listening effort compared to unilateral CI users (Hughes and Galvin, 2013; Schnabl et al., 2015;

Kocak Erdem and Ciprut, 2019). The neural response of CI users may also have been limited by individual differences related to the user device, such as CI distortions to the speech envelope, and thus leading to no detectable changes in TRF amplitude at a group level that corresponded with listening effort and attention. Interestingly, TRF<sub>200</sub> components were not significantly different from the babble noise TRF waveform in all listening conditions in contrast to the TRF<sub>100</sub> and TRF<sub>350</sub> components compared to previous studies with NH listeners (e.g., Ding and Simon, 2012; Broderick et al., 2018). This perhaps is another indicator that TRF components represent different aspects of low-level speech encoding, and that specific encoding processes reflected in the TRF<sub>200</sub> component are impaired in CI users.

It is possible that the effect of background noise on speech tracking is driven by the TRF<sub>350</sub> component, as exploratory analysis suggested that the late component magnitude was lower in the High-Noise condition compared to the Low-Noise condition. Recently, Paul et al. (2020) reported that bilateral CI users demonstrate stronger speech separation as reflected in later speech tracking differences around 250 ms between attended and ignored speech in a concurrent digit-stream task, while speech tracking toward the attended speech stream was stronger in the early component around 150 ms for NH listeners. Although CI users' TRFs in the current study demonstrated differentiation of speech from background noise, the later negative component initially appeared to be enhanced for stimuli TRFs, contrary to the stronger late cortical representation of the ignored speech stream described by Paul et al. (2020). Furthermore, the mean late negative component seen in the present study peaked 100 ms later at ~350 ms, possibly as a result of using continuous speech stimuli instead of number sequences since TRF latency has been linked to speech prosody (Verschueren et al., 2020). Additionally, both TRF<sub>100</sub> and TRF<sub>350</sub> were significantly different from the babble TRF for the High-Noise listening condition, suggesting that cortical speech differentiation was intact for low and moderate noise conditions. Speech envelope TRFs have also been shown to remain relatively stable in quiet conditions, despite modulations of speech comprehension via the scrambling the words presented in the speech narrative stimuli (Broderick et al., 2022). Late cortical differentiation of speech may therefore be more sensitive to background noise levels in specific regard to low-level speech encoding, as the TRF<sub>350</sub> component was not statistically significantly different from the babble noise TRF in the High-Noise condition.

### Speech tracking components did not differ for audiovisual manipulations in low background noise scenarios

Despite the reported benefits of visual speech cues for speech perception in noise, speech tracking component amplitudes did not differ between the Low-Noise and Audio-Only conditions.

The lack of difference is potentially explained by a ceiling effect regarding speech tracking. As the SNR in both audiovisual Low-Noise and Audio-Only condition was relatively high (+15 dB), participants may have been able to track the attended speech stream at the same level of performance for both conditions, despite the increased subjective listening demand in the Audio-Only condition. The benefits of visual cues then present itself in listening situations with lower SNRs, when CI users cannot adequately rely on acoustic speech cues to identify and maintain separate speech streams. However, previous research also suggest that in audiovisual conditions, CI users favor a top-down attentional modulation process when confronted with incongruent visual information, compared to NH listeners that adopt a more bottom-up process (Song et al., 2015). Since the CI users of the current study were not instructed to close their eyes, the irrelevant visual environment in the Audio-Only condition could influence the attentional modulation process by serving as a distractor.

### Early speech tracking component amplitudes are correlated with self-reported demand and percentage of conversation understood, but not the percentage of words understood

The current study found that the strength of early TRF components explained mental demand ratings; more positive (i.e., weaker) TRF<sub>100</sub> amplitudes are associated with increased perceived listening demand, regardless of background and audiovisual manipulations. Weaker TRF<sub>100</sub> amplitudes are also correlated with a decreased perceived percentage of conversation understood, but not the perceived percentage of words understood, reflecting recent findings by Broderick et al. (2022) where word recognition remained high in participants despite the scrambling of the speech narrative stimuli. The non-significant relationship observed for TRF<sub>200</sub> and TRF<sub>350</sub> components has also been reported in previous literature. Müller et al. (2019) found no significant correlations between listening effort and later component amplitudes in NH listeners after measuring listening effort subjectively through self-reports and objectively using cross-correlations between the speech envelope and EEG signals.

These results suggest that differences in speech encoding between CI users may be strongly reflected in TRF<sub>100</sub> amplitude. Despite the large variance in CI user characteristics that may have masked the effects of background noise and visual cues on speech tracking at a group level, the relationships between TRF<sub>100</sub> amplitude, mental demand, and conversation understanding remained statistically significant. TRFs being a visual representative of the quality of low-level speech encoding (Broderick et al., 2022; Prinsloo and Lalor, 2022), can serve as an indicator of deficits related to upstream processes in auditory

processing, that is, the relaying of acoustic information from the periphery auditory system to the central auditory system (i.e., downstream) in CI users. Poor encoding of the speech stimuli, whether due to poor encoding of low-level acoustics or deficits in basic auditory processing, would thereby lead to an increase in mental demand when interpreting speech signals, in turn resulting in poor comprehension. Increasing listening demand may also force CI users to compensate by relying on visual speech cues, which may not fully recoup the effects of degradation in auditory information. Alternatively, the concurrent increase in TRF<sub>100</sub> and demand ratings may represent a potential ceiling effect related to speech intelligibility (e.g., Drennan and Lalor, 2019; Verschuere et al., 2020). Whether due to low background noise or information from visual cues, participants may not have had significant issues understanding the stimuli for the neuromodulatory effects on speech tracking to significantly presents itself in the SNR + 15 dB listening conditions, despite their perception of increased demand and decreased understanding.

The non-significant correlation between AzBio scores and self-perceived percentage of conversation understood is similar to the weak correlation between clinical speech perception tests and subjective QoL outcomes seen in previous studies (e.g., Ramakers et al., 2017; McRackan et al., 2018; Thompson et al., 2020). For instance, Ramakers et al. (2017) reported weak to moderate correlations between Utrecht Sentence Test with Adaptive Randomized Roving sentences performance and the Speech, Spatial, and Qualities of Hearing Scale (SSQ) (Gatehouse and Noble, 2004) Speech scale ( $r = -0.36$ ,  $p = 0.0429$ ) and Nijmegen Cochlear Implant Questionnaire (NCIQ) (Hinderink et al., 2000) Advanced Sound Perception Domain ( $r = -0.47$ ,  $p = 0.0214$ ). Similarly, Thompson et al. (2020) reported that AzBio in noise (SNR 0 dB) scores do not correlate with the Speech-in-Speech Context subscale within the SSQ Speech scale. In turn, a meta-analysis by McRackan et al. (2018) pooled correlation values from studies utilizing tests such as the SSQ and the NCIQ. The pooled correlation values ranged from negligible to low for sentence recognition in quiet ( $r = 0.219$  [0.118–0.316]) and in noise ( $r = 0.238$  [–0.054 to 0.493]) (McRackan et al., 2018). These results again highlight the complex role of visual cues during listening, as clinical hearing tests are strictly auditory in nature.

### Influence of background noise and visual cues on alpha power

Alpha power is theorized to increase in neural regions involved in processing information irrelevant to the task at hand (Strauß et al., 2014), possibly as a reflection of the inhibition of distracting cortical networks (Jensen and Mazaheri, 2010; Petersen et al., 2015). Changes in alpha power during listening differ depending on cortical regions of interest as well as



how the sounds are presented analyzed (i.e., long continuous speech sentences or short duration sounds in event-related paradigms, e.g., see Hauswald et al., 2020). Previous studies have demonstrated that alpha power increases concurrently with task difficulty (Wöstmann et al., 2017), with increases in alpha power at the IFG in tandem with increases in listening effort (Dimitrijevic et al., 2019). Parietal alpha power was also found to increase as listening difficulty and attention increases due to increasing background noise or vocoded speech (Dimitrijevic et al., 2017; Wöstmann et al., 2017). In direct contrast, parietal and frontal alpha power has been shown to decrease as the difficulty of the listening condition increases, whether due to increasing background noise (Seifi Ala et al., 2020), decreasing target stimuli intelligibility (Fiedler et al., 2021), or vocoded speech (Hauswald et al., 2020).

Our results partially conflict with previous research, as alpha power itself at the superior-parietal and left IFG ROIs did not significantly change as an effect of increasing background noise or the presence of visual cues. One interpretation is that the effects of increasing background noise are potentially attenuated by the benefit that visual cues confer for attending to speech, reducing the required resources needed to suppress irrelevant background information. The modulation of parietal alpha power by SNR has been demonstrated for purely auditory tasks (e.g., Dimitrijevic et al., 2017; Wöstmann et al., 2017; Seifi Ala et al., 2020; Paul et al., 2021). The lack of difference in alpha power for the audiovisual listening conditions may therefore be an indicator of audiovisual advantage during sustained listening. At low noise levels in the Audio-Only and Low-Noise condition, background noise affects subjective listening demand, but possibly not enough to warrant neural differences in the suppression of background noise. Superior-parietal alpha power also did not differ between the side ipsilateral and contralateral to the CI side used during the listening task. Increased alpha power has also been reported for the brain hemisphere ipsilateral to the attended stimuli for listening tasks involving spatial attention and suppression of irrelevant stimuli located on the contralateral side (e.g., Foxe and Snyder, 2011; Thorpe et al., 2012; Paul et al., 2020). The target stimuli for the current study were located directly in front of the participants and did not change location. Thus, no lateralization of alpha power was expected. Stimulus material may also influence how alpha power changes as a result of degrading speech clarity. Hauswald et al. (2020) found that when using continuous stimuli parietal alpha power decreased as speech degradation increased, in contrast to studies that used short stimuli such as words or digits (Obleser and Weisz, 2012; Wostmann et al., 2015; Dimitrijevic et al., 2019). The combination of stimulus choice, the benefit of visual cues, and low background noise in some conditions may then have contributed to masking the effect of decreasing alpha power due to decreased speech comprehension across all listening conditions. Alternatively, alpha power as

measured in this study may not relate to the conditions of our task design.

Regarding listening demand, a decrease in superior-parietal alpha power was associated with an increase in demand scores, but only in the audiovisual High-Noise condition. Continuous listening in an environment with high background noise would be especially difficult for individuals with hearing difficulty, perhaps to a point where CI users begin to favor visual speech cues over auditory cues to compensate for decreased speech intelligibility. Here, alpha power may represent a release from inhibition (Jensen and Mazaheri, 2010) regarding multisensory integration processes that occur in the parietal region (e.g., Rouger et al., 2008; Misselhorn et al., 2019), similar to the inverted-U pattern displayed for alpha power in high effort listening scenarios (Hauswald et al., 2020; Paul et al., 2021; Ryan et al., 2022). Any association between alpha power and demand scores for conditions with lower background noise may be complicated by CI users' bias toward visual cues (Stropahl et al., 2017), in that CI users overestimate the benefit that visual cues confer on speech intelligibility, and thereby inherently rely more on visual cues despite the audibility of the speech stimuli. However, it is difficult to determine if this bias is purely subjective based on alpha power alone, due to the cross-modal plasticity observed in individuals with hearing loss (Lazard et al., 2010; Anderson et al., 2017) as well as the aforementioned multisensory integration. Since no visual speech cues were present in the Audio-Only condition, audiovisual integration of visual and auditory speech cues was not necessary, which would lead to the suppression of parietal multisensory integration.

Differences between the effort required for a task (i.e., listening effort) and the difficulty of the task itself (i.e., mental demand) may be reflected in the objective and subjective measures respectively, and thus influencing both TRF components and alpha power. If CI users did not attempt to change their listening effort to match the difficulty/demand of the task throughout all listening conditions, then the differences in both TRF amplitude and alpha power related to listening demand between conditions would be minimized, and thus not detected due to having a small effect size. As the participants were only briefly assessed on stimuli content to gauge attention, passive listening may also have taken place during the task if lapses in attention occurred. Factors beyond attention also influence both speech tracking and listening effort. Even though attentional processes affect listening effort, motivation and fatigue are also confounding factors for both concepts of listening effort and listening demand. Motivation has been found to generally increase subjective listening effort and fatigue when attending to speech in noise, in addition to influencing strategies for coping with effortful tasks (Picou and Ricketts, 2014). Cognitive overload due to task difficulty may also result in participants disengaging from the task itself, thereby producing behavior that is similar to "low effort" (Wu et al., 2016). In addition, familiarity with the stimuli may also play a role in

attention, in that familiarity with a voice in a complex listening environment can assist listeners to either attend to or ignore said voice (Johnsrude et al., 2013). When sentences is spoken concurrently by two voices during a word identification task, error rates were found to be higher if both voices were novel, compared to when a familiar voice was the target stimuli or the informational masker (Johnsrude et al., 2013).

## Conclusion

Increasing background noise increased subjective mental demand when listening in noise, while visual speech cues appeared to decrease listening demand in low background noise levels. Stimuli TRF component amplitudes were significantly different from babble noise TRFs (TRF<sub>100</sub> and TRF<sub>350</sub> in the low and medium noise conditions) implicating potential difficulties in separating speech from the background for high noise. The significant relationship between early speech tracking components, alpha power, and subjective listening demand suggests that perhaps visual cues play a more complex role for CI users in higher-noise conditions. Relative alpha power however did not appear to change accordingly with background noise in contrast to previous literature. Nevertheless, factors such as study design, motivation, and fatigue may confound the results of neural measures for individuals with CIs compared to NH listeners, and thus warrant further investigation.

Overall, the results of this study suggest that naturalistic multi-talker scenarios such as everyday movie stimuli, can be used to study listening demand in CI users. Neural tracking of continuous naturalistic speech in noise was successfully measured in CI users using TRFs. Meaningful information about the CI user listening experience and the influence of auditory environmental factors can be extracted from brain responses using continuous “ecological” stimuli such as a normal conversation. These findings contribute to the development of naturalistic objective measures of listening effort that may provide more insight into the mental effort and fatigue of CI users in everyday listening. The method of speech tracking estimation used in the current study also does not require an active response from participants, and thus has the potential to be integrated in assessments for individuals that cannot respond to conventional task-based measures. This approach may be especially useful in pediatric populations where behavioral measures of demand and long EEG recordings are difficult to perform.

## Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Board at Sunnybrook Health Sciences Centre (REB #474-2016) in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

BX performed data analyses, interpretation, and wrote the manuscript as part of a Master of Science degree thesis that was subsequently adapted for publication. BP and AD contributed to the conceptualization of the work, provided supervision, data analysis, and revised the manuscript before publication. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by the Mason Scientific Discovery Fund and the Harry Barberian Scholarship. The funding organizations had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the decision to submit the article for publication; or in the preparation, review, or approval of the article.

## Acknowledgments

This research article was adapted from a dissertation as part of a Master of Science degree at the Institute of Medical Science at University of Toronto (Xiu, 2022).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alain, C., Du, Y., Bernstein, L. J., Barten, T., and Banai, K. (2018). Listening under difficult conditions: an activation likelihood estimation meta-analysis. *Hum. Brain Mapp.* 39, 2695–2709. doi: 10.1002/hbm.24031
- Anderson, C. A., Wiggins, I. M., Kitterick, P. T., and Hartley, D. E. H. (2017). Adaptive benefit of cross-modal plasticity following cochlear implantation in deaf adults. *Proc. Natl. Acad. Sci. U S A.* 114, 10256–10261. doi: 10.1073/pnas.1704785114
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* 45, 434–444. doi: 10.1109/78.554307
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.2307/2346101
- Blamey, P., Artieres, F., Baskent, D., Bergeron, F., Beynon, A., Burke, E., et al. (2012). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiol. Neurotol.* 18, 36–47. doi: 10.1159/000343189
- Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.* 28, 803–809.e3. doi: 10.1016/j.cub.2018.01.080
- Broderick, M. P., Zuk, N. J., Anderson, A. J., and Lalor, E. C. (2022). More than words: neurophysiological correlates of semantic dissimilarity depend on comprehension of the speech narrative. *Eur. J. Neurosci.* 56, 5201–5214. doi: 10.1111/ejn.15805
- Brown, V. A., and Strand, J. F. (2019). About face: seeing the talker improves spoken word recognition but increases listening effort. *J. Cogn.* 2:44. doi: 10.5334/joc.89
- Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R., and Kennedy, D. P. (2015). Idiosyncratic brain activation patterns are associated with poor social comprehension in Autism. *J. Neurosci.* 35, 5837–5850. doi: 10.1523/JNEUROSCI.5182-14.2015
- Capretta, N. R., and Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *Laryngoscope* 126, 699–706. doi: 10.1002/lary.25525
- Castellanos, I., Kronenberger, W. G., and Pisoni, D. B. (2018). Psychosocial outcomes in long-term cochlear implant users. *Ear Hear.* 39, 527–539. doi: 10.1097/AUD.0000000000000504
- Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10:604. doi: 10.3389/fnhum.2016.00604
- Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., and Lalor, E. C. (2021). Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research. *Front. Neurosci.* 15:705621. doi: 10.3389/fnins.2021.705621
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Dillon, M. T., Buss, E., Rooth, M. A., King, E. R., Deres, E. J., Buchman, C. A., et al. (2018). Effect of cochlear implantation on quality of life in adults with unilateral hearing loss. *Audiol. Neurotol.* 22, 259–271. doi: 10.1159/000484079
- Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Front. Hum. Neurosci.* 11:88. doi: 10.3389/fnhum.2017.00088
- Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Sci. Rep.* 9:11278. doi: 10.1038/s41598-019-47643-1
- Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U S A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Dorman, M. F., Liss, J., Wang, S., Berisha, V., Ludwig, C., and Natale, S. C. (2016). Experiments on auditory-visual perception of sentences by users of unilateral, bimodal, and bilateral cochlear implants. *J. Speech, Lang. Hear. Res.* 59, 1505–1519. doi: 10.1044/2016\_JSLHR-H-15-0312
- Drennan, D. P., and Lalor, E. C. (2019). Cortical tracking of complex sound envelopes: modeling the changes in response with intensity. *eNeuro* 6:ENEURO.0082-19.2019.
- Dunn, C. C., Noble, W., Tyler, R. S., Kordus, M., Gantz, B. J., and Ji, H. (2010). Bilateral and unilateral cochlear implant users compared on speech perception in noise. *Ear Hear.* 31, 296–298. doi: 10.1097/AUD.0b013e3181c12383
- Fabie, J. E., Keller, R. G., Hatch, J. L., Holcomb, M. A., Camposo, E. L., Lambert, P. R., et al. (2018). Evaluation of outcome variability associated with lateral wall, mid-scalar, and perimodiolar electrode arrays when controlling for preoperative patient characteristics. *Otol. Neurotol.* 39, 1122–1128. doi: 10.1097/MAO.0000000000001951
- Fiedler, L., Seif, Ala, T., Graversen, C., Alickovic, E., Lunner, T., and Wendt, D. (2021). Hearing aid noise reduction lowers the sustained listening effort during continuous speech in noise—a combined pupillometry and EEG study. *Ear Hear.* 42, 1590–1601. doi: 10.1097/AUD.0000000000001050
- Fiedler, L., Wöstmann, M., Herbst, S. K., and Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage* 186, 33–42. doi: 10.1016/j.neuroimage.2018.10.057
- Foxe, J. J., and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2:154. doi: 10.3389/fpsyg.2011.00154
- Fraser, S., Gagné, J.-P., Alepins, M., and Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J. Speech Lang. Hear. Res.* 53, 18–33. doi: 10.1044/1092-4388(2009)08-0140
- Fuglsang, S. A., Dau, T., and Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage* 156, 435–444. doi: 10.1016/j.neuroimage.2017.04.026
- Gatehouse, S., and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int. J. Audiol.* 43, 85–99. doi: 10.1080/14992020400050014
- Gilley, P. M., Sharma, A., Dorman, M., Finley, C. C., Panch, A. S., and Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clin. Neurophysiol.* 117, 1772–1782. doi: 10.1016/j.clinph.2006.04.018
- Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183.
- Hauswald, A., Keitel, A., Chen, Y., Rösch, S., and Weisz, N. (2020). Degradation levels of continuous speech affect neural speech tracking and alpha power differently. *Eur. J. Neurosci.* 55, 3288–3302. doi: 10.1111/ejn.14912
- Hinderink, J. B., Krabbe, P. F. M., and Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen Cochlear Implant Questionnaire. *Otolaryngol. Neck Surg.* 123, 756–765. doi: 10.1067/mhn.2000.108203
- Hjortkjær, J., Märcher-Rørsted, J., Fuglsang, S. A., and Dau, T. (2020). Cortical oscillations and entrainment in speech processing during working memory load. *Eur. J. Neurosci.* 51, 1279–1289. doi: 10.1111/ejn.13855
- Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear.* 34, 342–360. doi: 10.1097/AUD.0b013e3182741aa7
- Holman, J. A., Drummond, A., Hughes, S. E., and Naylor, G. (2019). Hearing impairment and daily-life fatigue: a qualitative study. *Int. J. Audiol.* 58, 408–416. doi: 10.1080/14992027.2019.1597284
- Hughes, K. C., and Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int.* 14, 121–129. doi: 10.1179/1754762812Y.0000000000
- Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., and Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: a grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear.* 39, 922–934. doi: 10.1097/AUD.0000000000000553
- James, C. J., Karoui, C., Laborde, M.-L., Lepage, B., Molinier, C. -É, Tartayre, M., et al. (2019). Early sentence recognition in adult cochlear implant users. *Ear Hear.* 40, 905–917. doi: 10.1097/AUD.0000000000000670
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186

- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., and Carlyon, R. P. (2013). Swinging at a cocktail party. *Psychol. Sci.* 24, 1995–2004. doi: 10.1177/0956797613482467
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116, 2395–2405. doi: 10.1121/1.1784440
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007
- Kocak Erdem, B., and Ciprut, A. (2019). Evaluation of speech, spatial perception and hearing quality in unilateral, bimodal and bilateral cochlear implant users. *Turkish Arch. Otorhinolaryngol.* 57, 149–153. doi: 10.5152/tao.2019.4105
- Kurz, A., Grubenbecher, M., Rak, K., Hagen, R., and Kühn, H. (2019). The impact of etiology and duration of deafness on speech perception outcomes in SSD patients. *Eur. Arch. Oto-Rhino-Laryngol.* 276, 3317–3325. doi: 10.1007/s00405-019-05644-w
- Lazard, D. S., Lee, H. J., Gaebler, M., Kell, C. A., Truy, E., and Giraud, A. L. (2010). Phonological processing in post-lingual deafness and cochlear implant outcome. *Neuroimage* 49, 3443–3451. doi: 10.1016/j.neuroimage.2009.11.013
- Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., et al. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time. *PLoS One* 7:e48739. doi: 10.1371/journal.pone.0048739
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y
- Magosso, E., De Crescenzo, F., Ricci, G., Piastra, S., and Ursino, M. (2019). EEG alpha power is modulated by attentional changes during cognitive tasks and virtual reality immersion. *Comput. Intell. Neurosci.* 2019, 1–18. doi: 10.1155/2019/7051079
- Mastrantuono, E., Saldaña, D., and Rodríguez-Ortiz, I. R. (2017). An eye tracking study on the perception and comprehension of unimodal and bimodal linguistic inputs by deaf adolescents. *Front. Psychol.* 8:1044. doi: 10.3389/fpsyg.2017.01044
- MATLAB (2019). *version 9.6.0.1072779 (R2019a)*. Natick, MA: The MathWorks Inc.
- Mc Laughlin, M., Lopez Valdes, A., Reilly, R. B., and Zeng, F.-G. (2013). Cochlear implant artifact attenuation in late auditory evoked potentials: a single channel approach. *Hear. Res.* 302, 84–95. doi: 10.1016/j.heares.2013.05.006
- McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., et al. (2014). Listening effort and fatigue: what exactly are we measuring? a british society of audiology cognition in hearing special interest group 'white paper'. *Int. J. Audiol.* 53, 433–445. doi: 10.3109/14992027.2014.890296
- McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., et al. (2018). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope* 128, 982–990. doi: 10.1002/lary.26738
- Miller, S., and Zhang, Y. (2014). Validation of the cochlear implant artifact correction tool for auditory electrophysiology. *Neurosci. Lett.* 577, 51–55. doi: 10.1016/j.neulet.2014.06.007
- Minimum Speech Test Battery For Adult Cochlear Implant Users (2011). *Advanced Bionics LLC, Cochlear Americas, MED-EL Corporation*. Available at: <http://www.auditorypotential.com/MSTBfiles/MSTBManual2011-06-20%20.pdf> (accessed December 15, 2021).
- Misselhorn, J., Friese, U., and Engel, A. K. (2019). Frontal and parietal alpha oscillations reflect attentional modulation of cross-modal matching. *Sci. Rep.* 9:5030. doi: 10.1038/s41598-019-41636-w
- Müller, J. A., Wendt, D., Kollmeier, B., Debener, S., and Brand, T. (2019). Effect of speech rate on neural tracking of speech. *Front. Psychol.* 10:449. doi: 10.3389/fpsyg.2019.00449
- Nishimura, H., Hashikawa, K., Doi, K., Iwaki, T., Watanabe, Y., Kusuoka, H., et al. (1999). Sign language 'heard' in the auditory cortex. *Nature* 397:116. doi: 10.1038/16376
- Niso, G., Tadel, F., Bock, E., Cousineau, M., Santos, A., and Baillet, S. (2019). Brainstorm pipeline analysis of resting-state data from the open MEG archive. *Front. Neurosci.* 13:284. doi: 10.3389/fnins.2019.00284
- Obleser, J., and Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb. Cortex* 22, 2466–2477. doi: 10.1093/cercor/bhr325
- Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383.
- Oostenvelde, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- O'Sullivan, A. E., Lim, C. Y., and Lalor, E. C. (2019). Look at me when I'm talking to you: selective attention at a multisensory cocktail party can be decoded using stimulus reconstruction and alpha power modulations. *Eur. J. Neurosci.* 50, 3282–3295. doi: 10.1111/ejn.14425
- Pals, C., Sarampalis, A., and Başkent, D. (2013). Listening effort with cochlear implant simulations. *J. Speech, Lang. Hear. Res.* 56, 1075–1084.
- Pantelis, P. C., Byrge, L., Tysza, J. M., Adolphs, R., and Kennedy, D. P. (2015). A specific hypoactivation of right temporo-parietal junction/posterior superior temporal sulcus in response to socially awkward situations in autism. *Soc. Cogn. Affect. Neurosci.* 10, 1348–1356. doi: 10.1093/scan/nsv021
- Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.
- Paul, B. T., Chen, J., Le, T., Lin, V., and Dimitrijevic, A. (2021). Cortical alpha oscillations in cochlear implant users reflect subjective listening effort during speech-in-noise perception. *PLoS One* 16:e0254162. doi: 10.1371/journal.pone.0254162
- Paul, B. T., Uzelac, M., Chan, E., and Dimitrijevic, A. (2020). Poor early cortical differentiation of speech predicts perceptual difficulties of severely hearing-impaired listeners in multi-talker environments. *Sci. Rep.* 10:6141. doi: 10.1038/s41598-020-63103-7
- Payne, L., and Sekuler, R. (2014). The importance of ignoring. *Curr. Dir. Psychol. Sci.* 23, 171–177. doi: 10.1177/0963721414529145
- Peelle, J. E. (2018). Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000494
- Peng, Z. E., and Wang, L. M. (2019). Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during english speech perception. *J. Speech Lang. Hear. Res.* 62, 1068–1081.
- Perrin, F., Pernier, J., Bertrand, O., and Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187. doi: 10.1016/0013-4694(89)90180-6
- Petersen, E. B., Wöstmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *J. Neurophysiol.* 117, 18–27. doi: 10.1152/jn.00527.2016
- Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., and Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front. Psychol.* 6:177. doi: 10.3389/fpsyg.2015.00177
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the Framework for Understanding Effortful Listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312
- Picou, E. M., Gordon, J., and Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear Hear.* 37, 1–13. doi: 10.1097/AUD.0000000000000222
- Picou, E. M., and Ricketts, T. A. (2014). Increasing motivation changes subjective reports of listening effort and choice of coping strategy. *Int. J. Audiol.* 53, 418–426. doi: 10.3109/14992027.2014.880814
- Price, C. N., Alain, C., and Bidelman, G. M. (2019). Auditory-frontal channeling in  $\alpha$  and  $\beta$  bands is altered by age-related hearing loss and relates to speech perception in noise. *Neuroscience* 423, 18–28. doi: 10.1016/j.neuroscience.2019.10.044
- Prinsloo, K. D., and Lalor, E. C. (2022). General auditory and speech-specific contributions to cortical envelope tracking revealed using auditory chimeras. *J. Neurosci.* 42, 7782–7798. doi: 10.1523/JNEUROSCI.2735-20.2022
- Puschmann, S., Daeglau, M., Stropahl, M., Mirkovic, B., Rosemann, S., Thiel, C. M., et al. (2019). Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise. *Neuroimage* 196, 261–268. doi: 10.1016/j.neuroimage.2019.04.017
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Ramakers, G. G. J., Smulders, Y. E., van Zon, A., Van Zanten, G. A., Grolman, W., and Stegeman, I. (2017). Correlation between subjective and objective hearing tests after unilateral and bilateral cochlear implantation. *BMC Ear Nose Throat Disord.* 17:10. doi: 10.1186/s12901-017-0043-y

- Reetzke, R., Gnanateja, G. N., and Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain Lang.* 213, 104891. doi: 10.1016/j.bandl.2020.104891
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031
- Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. (2010). When cognition kicks in: working memory and speech understanding in noise. *Noise Heal.* 12:263. doi: 10.4103/1463-1741.70505
- Rouger, J., Fraysse, B., Deguine, O., and Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Res.* 1188, 87–99. doi: 10.1016/j.brainres.2007.10.049
- Ryan, D. B., Eckert, M. A., Sellers, E. W., Schairer, K. S., McBee, M. T., Jones, M. R., et al. (2022). Impact of effortful word recognition on supportive neural systems measured by alpha and theta power. *Ear Hear.* 43, 1549–1562. doi: 10.1097/AUD.0000000000001211
- Schnabl, J., Bumann, B., Rehbein, M., Müller, O., Seidler, H., Wolf-Magele, A., et al. (2015). Höranstrengung bei Cochlea-Implantaten [Listening effort with cochlear implants: unilateral versus bilateral use]. *HNO* 63, 546–551. doi: 10.1007/s00106-015-0020-y
- Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. M., and Lunner, T. (2020). An exploratory study of EEG alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech. *PLoS One* 15:e0235782. doi: 10.1371/journal.pone.0235782
- Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306
- Song, J.-J., Lee, H.-J., Kang, H., Lee, D. S., Chang, S. O., and Oh, S. H. (2015). Effects of congruent and incongruent visual cues on speech perception and brain activity in cochlear implant users. *Brain Struct. Funct.* 220, 1109–1125. doi: 10.1007/s00429-013-0704-6
- Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., et al. (2012). Development and validation of the AzBio sentence lists. *Ear Hear.* 33, 112–117. doi: 10.1097/AUD.0b013e31822c2549
- Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8:350. doi: 10.3389/fnhum.2014.00350
- Stropahl, M., Schellhardt, S., and Debener, S. (2017). McGurk stimuli for the investigation of multisensory integration in cochlear implant users: the Oldenburg Audio Visual Speech Stimuli (OLAVS). *Psychon. Bull. Rev.* 24, 863–872. doi: 10.3758/s13423-016-1148-9
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011, 1–13. doi: 10.1155/2011/879716
- Thompson, N. J., Dillon, M. T., Buss, E., Rooth, M. A., King, E. R., Buckner, A. L., et al. (2020). Subjective benefits of bimodal listening in cochlear implant recipients with asymmetric hearing loss. *Otolaryngol. Neck Surg.* 162, 933–941. doi: 10.1177/0194599820911716
- Thorpe, S., D'Zmura, M., and Srinivasan, R. (2012). Lateralization of frequency-specific networks for covert spatial attention to auditory stimuli. *Brain Topogr.* 25, 39–54. doi: 10.1007/s10548-011-0186-x
- Verschueren, E., Vanthornhout, J., and Francart, T. (2020). The effect of stimulus choice on an EEG-Based objective measure of speech intelligibility. *Ear Hear.* 41, 1586–1597. doi: 10.1097/AUD.0000000000000875
- Wagner, L., Maurits, N., Maat, B., Baskent, D., and Wagner, A. E. (2018). The cochlear implant EEG artifact recorded from an artificial brain for complex acoustic stimuli. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 392–399. doi: 10.1109/TNSRE.2018.2789780
- Wang, J., Zhu, Y., Chen, Y., Mamat, A., Yu, M., Zhang, J., et al. (2020). An eye-tracking study on audiovisual speech perception strategies adopted by normal-hearing and deaf adults under different language familiarities. *J. Speech Lang. Hear. Res.* 63, 2245–2254. doi: 10.1044/2020\_JSLHR-19-00223
- Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., and Obleser, J. (2015). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cereb. Cortex* 25, 1938–1946. doi: 10.1093/cercor/bhu004
- Winn, M. B., Rhone, A. E., Chatterjee, M., and Idsardi, W. J. (2013). The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Front. Psychol.* 4:824. doi: 10.3389/fpsyg.2013.00824
- Wostmann, M., Herrmann, B., Wilsch, A., and Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J. Neurosci.* 35, 1458–1467. doi: 10.1523/JNEUROSCI.3250-14.2015
- Wöstmann, M., Lim, S.-J., and Obleser, J. (2017). The human neural alpha response to speech is a proxy of attentional control. *Cereb. Cortex* 27, 3307–3317. doi: 10.1093/cercor/bhx074
- Wu, Y.-H., Stangl, E., Zhang, X., Perkins, J., and Eilers, E. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear Hear.* 37, 660–670. doi: 10.1097/AUD.0000000000000335
- Xiu, B. (2022). *Electrophysiological correlates of subjective cognitive demand in cochlear implant users during listening in noise*. Ph.D. thesis. Toronto, ON: University of Toronto.
- Yi, H., Pingsterhaus, A., and Song, W. (2021). Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. *Front. Psychol.* 12:682677. doi: 10.3389/fpsyg.2021.682677
- Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a “Cocktail Party.”. *J. Neurosci.* 33, 1417–1426. doi: 10.1523/JNEUROSCI.3675-12.2013